

# ZH, H → γγ<sub>D</sub> analysis

24 February 2021

Federica Piazza - 2<sup>nd</sup> year PhD student  
Università degli studi e INFN, Milano



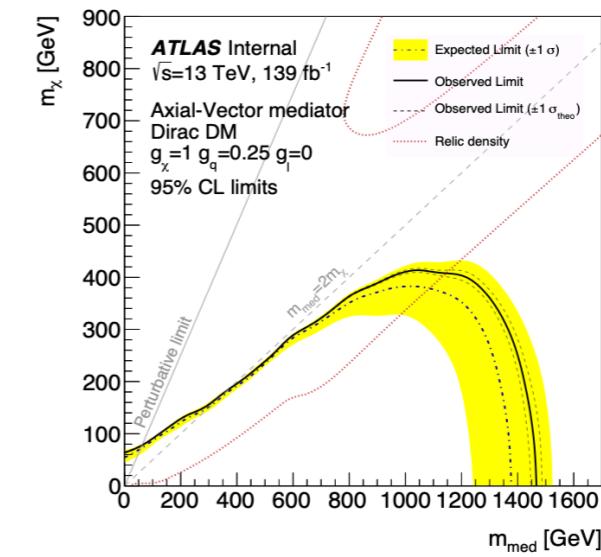
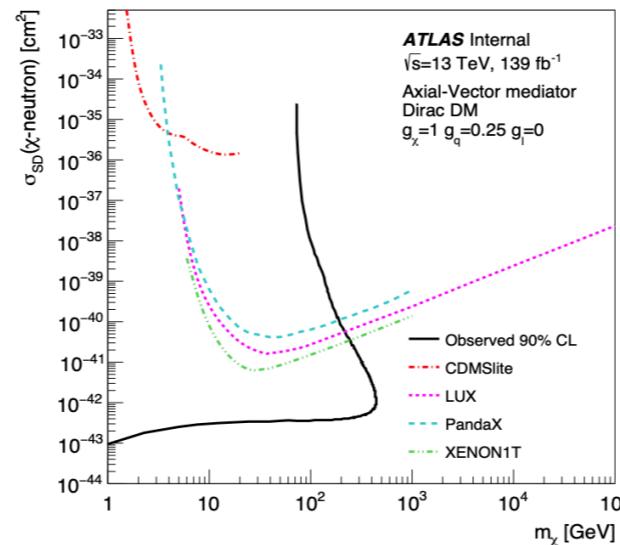
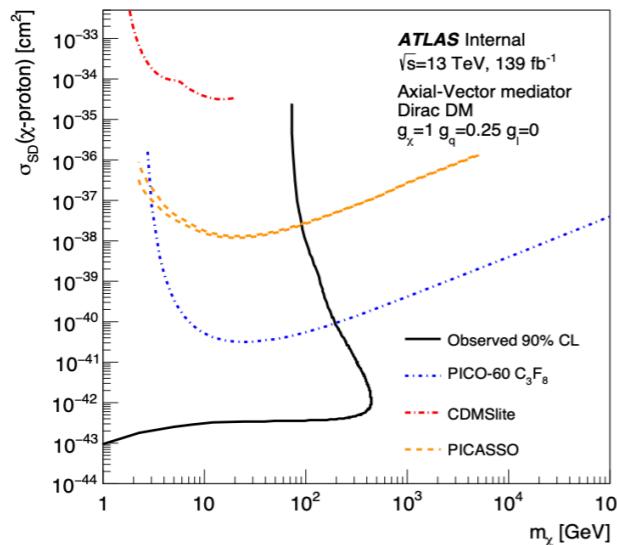
UNIVERSITÀ  
DEGLI STUDI  
DI MILANO



# My PhD activities

## Physics Analysis

- Focus on Dark-Matter related analyses:
  - **Mono-photon** analysis in the context of Exotics, JDM subgroup (master thesis + 1<sup>st</sup> PhD year)
    - $\gamma+\text{MET}$  final state
    - ▶ Paper accepted by JHEP ([arxiv](#))



- **Dark-photon analysis** in the context of HDBS, HLRS subgroup (joined the analysis team in july 2020)
  - $Z(\text{II})H \rightarrow \gamma\gamma D$
  - ▶ New analysis in ATLAS (CMS [results](#) already published)

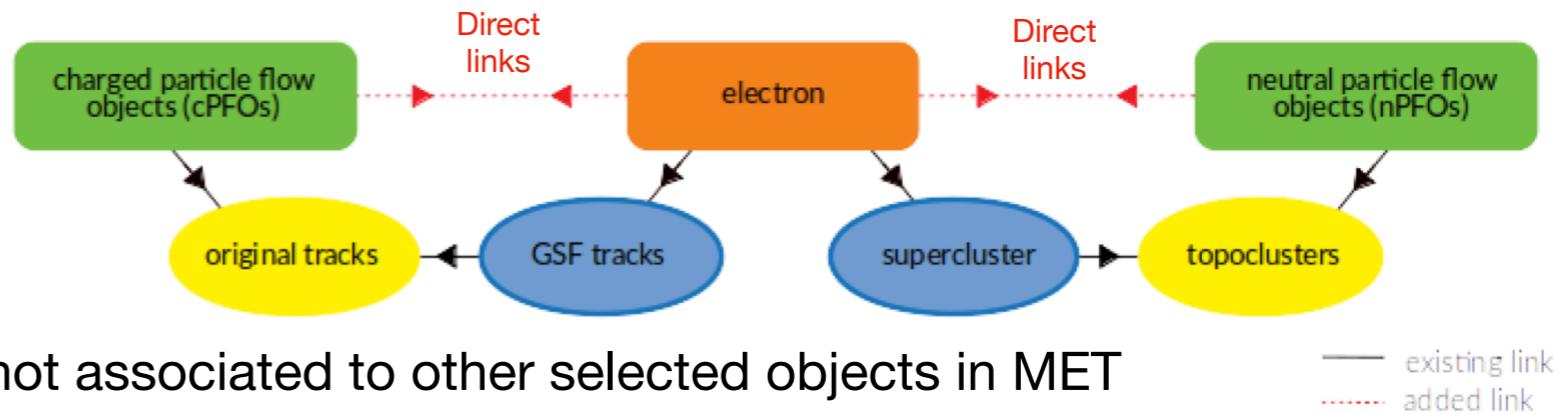
# My PhD activities

## Performance

- Performance studies in the context of JetETmiss, MET+PU subgroup: implementation and preliminary validation and performance studies for a MET reconstruction based on Global Particle Flow objects (QT project. End date: 16/12/20 )

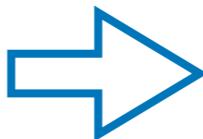
- Two main tasks:

1. Association Map based on **direct links** between PFOs and leptons/photons
2. Building Jet Term from PFOs not associated to other selected objects in MET



### Current PFlow MET

- **PFO association** to  $e/\mu/\gamma$  through  $\Delta R$  or clusters/tracks matching
- **JetTerm** from standard jet collection + overlap removal with other objects



### GPF MET

- **PFO association** to  $e/\mu/\gamma$  using direct links
- **JetTerm** from new “Overlap Removed (OR)” jet collection. No need for overlap removal

- Future plans:
  - Finalize integration of links between FlowElement and egamma/tau/muons in Association Maps
  - Finalize OR MET reconstruction (handling jet-muon overlap)
  - Large scale validation

# Dark-photon analysis

- **Glance entry:** [ANA-HDBS-2019-13](#)
- **Twiki page:** [link](#)
- **Analysis team**
  - BNL: Ketevi Assamagan
  - **Milano:** Leonardo Carminati, Marcello Fanti, Davide Mungo, Federica Piazza, Silvia Resconi + 3 bachelor students (Dario Pullia, Denise Tantucci, Andrea Mitta)
  - Nikhef: Stefano Manzoni
  - Osaka: Hajime Nanjo
  - Rabat: Hassnae El Jarrari, Yahya Tayalati
  - Standford: Stanislava Sevova, Lauren Tompkins, Rocky Garg + 2 Summer Students: Blanca Nino, Elyssa Hofgard
  - Taipei AS: Rachid Mazini

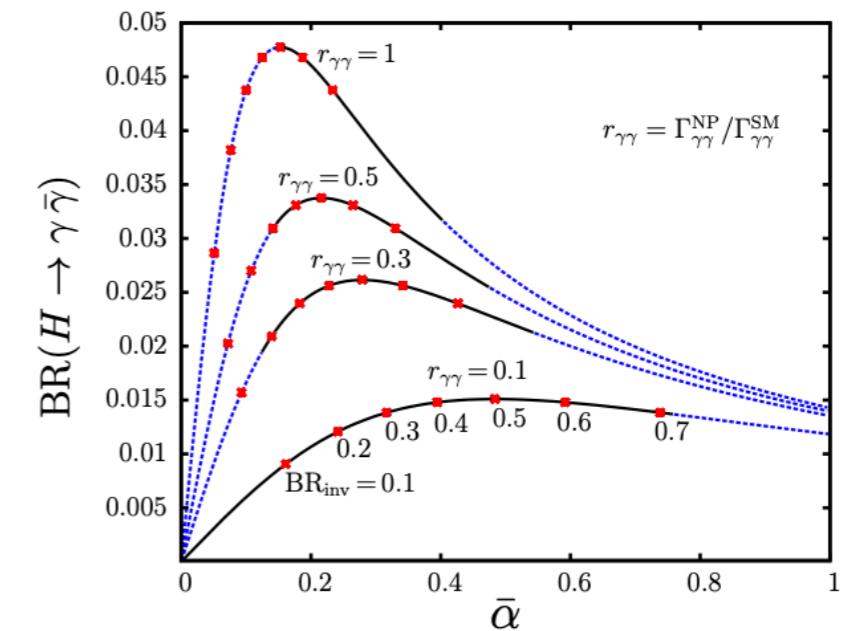
# Motivation

- Dark-photon is predicted as a massless (or light) gauge boson of a new unbroken U(1) group
  - Mediator of long range forces in the dark-sector, but also possible DM candidate
  - Can help explaining
    - small-scale structure formation
    - light-DM annihilation in asymmetric DM scenarios
    - Flavor hierarchy problem
- Why massless or light?
  - Given the unbroken U(1) symmetry, massive dark-photon leads to tree-level mixing with SM photons => strong constraints
  - Massless dark-photon => on-shell dark photons fully decoupled from the SM sector at tree level. Coupling with SM sector through higher-dimensional interactions via messenger exchange

# Phenomenology

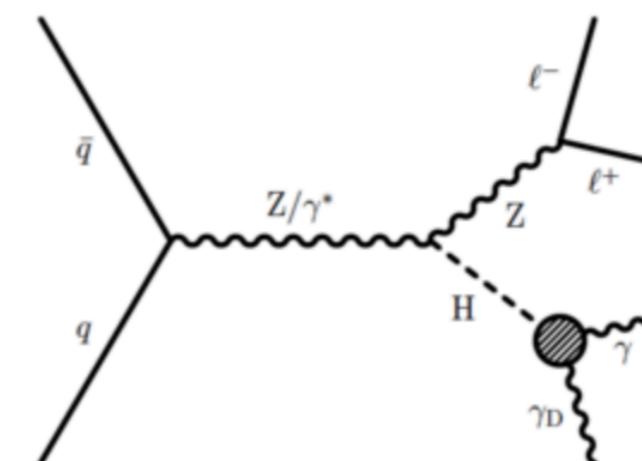
- Dark-photon from Higgs decay:
  - One loop decay
  - BSM BR up to 5% allowed by present constraints  
(Biswas, Gabrielli, Heikinheimo, Mele (2016))
- Different production modes can be explored:
  - VBF (analysis ongoing in ATLAS) link
  - ggH: potentially high sensitivity, but missing viable triggers for Run-2 (mT distribution spoilt)
  - **ZH (this analysis)**

Biswas, Gabrielli, Heikinheimo, Mele (2014)



## ZH, H→yyD

- Lower production cross-section
- But clean final state:  
two leptons from Z decay + one photon +  $E_T^{\text{miss}}$ 
  - VV+y only irreducible bkg (subdominant)
  - Dominant background contributions from Zy+jets and Zjets processes (fake  $E_T^{\text{miss}}$ )



# MC samples and Ntuple production

- **MC samples** (HIGG2D1 derivations)
  - Mainly Sherpa NLO
  - Powheg+Pythia8 for single top, Vt, Vtll, ttbar, Higgs related processes, signal
  - MadGraph+Pythia8 for ttV
  - Using  $H \rightarrow \gamma Gr$  for the signal
    - Graviton behaves like dark-photon
    - $m_{Gr} = 0, 1, 10, 20, 30, 40 \text{ GeV}$
- **Framework:**
  - “mini-Ntuple” production using [STAnalysisCode framework](#) (based on SUSYTools)
  - “Micro-Ntuple” production and post processing (plots, cutflows, ...) using [STPostProcessing framework](#)
  - Release 21.2.151

# Event selection

## Preselections

- GRL, Detector event cleaning, PV selection, Loose Jet Cleaning (MiniNtuple skim)
- At least one baseline electron or muon (MiniNtuple skim)
- Overlap removal between Vy and Vjets MC
- Trigger (single/double lepton)

## SR selections (optimization studies by E. Hofgard presented during last update)

- 2 opposite sign muons/electrons with  $p_T^{\text{lep}1} > 26 \text{ GeV}$  and  $p_T^{\text{lep}2} > 10 \text{ GeV}$
- 1 photon with  $p_T^\gamma > 25 \text{ GeV}$
- $\text{Njet} \leq 2$  and  $\text{Nbjet} = 0$
- $76 < m_{\parallel} < 116 \text{ GeV}$
- $m_{\parallel\gamma} > 100 \text{ GeV}$
- MET  $> 60 \text{ GeV}$

# Background composition in SR

## Background processes

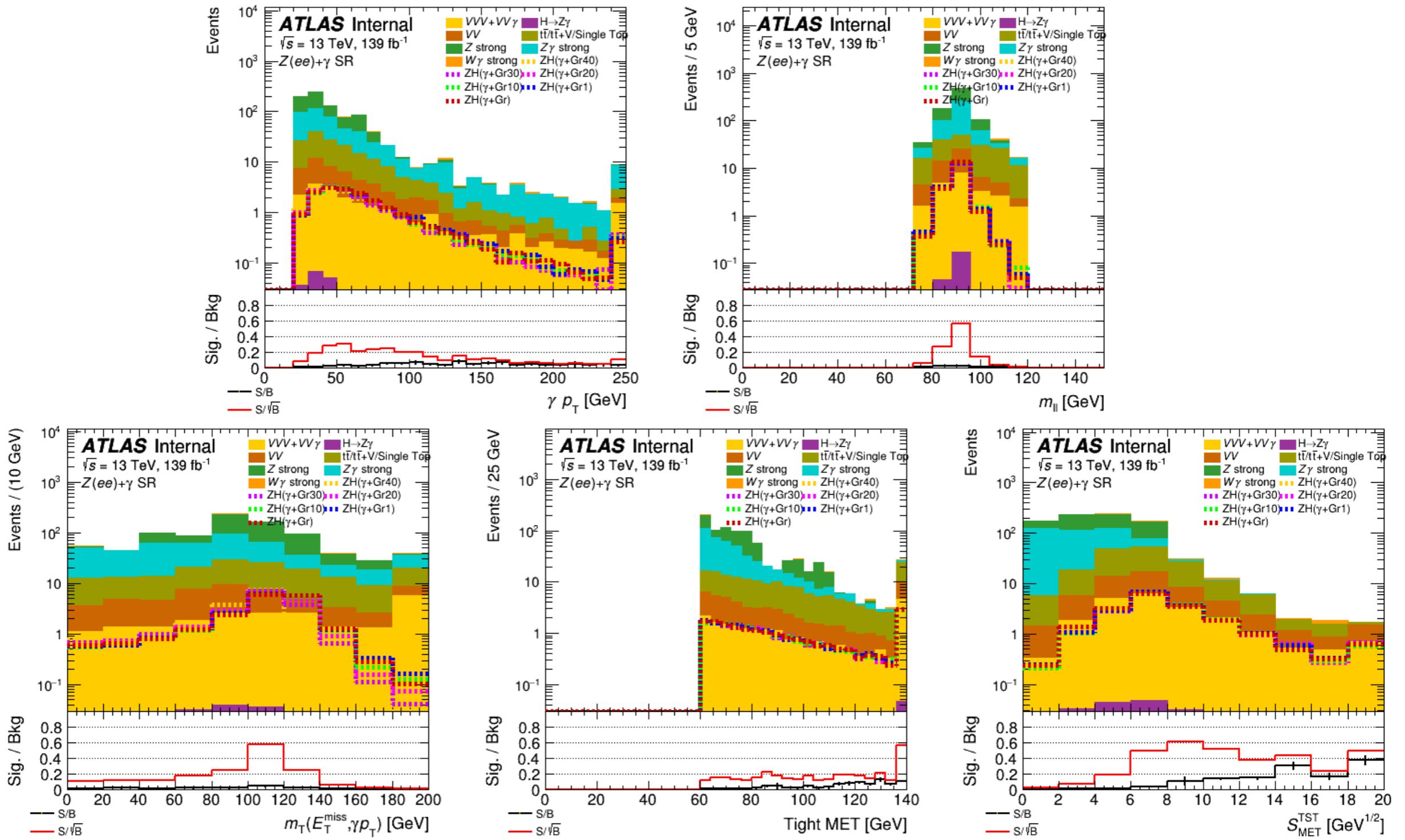
- Zgam:** Z $\gamma$ +jets => fake MET (fakeMET ABCD method)
- Zqcd:** Z+jets => fake MET + jet faking photon ( $j\rightarrow\gamma$  ABCD method)
- Top:** single Top + ttbar + Vtll + ttV + ttbarVV => non resonant background
- VV:** ggVV, VV ewk => jet faking photons ( $j\rightarrow\gamma$  ABCD) and electron faking photon ( $e\rightarrow\gamma$  fake rate from Zee)
- VV (VVy) => irreducible background (subdominant)**
- HZy:** ttHZy + VH $\gamma$ y + VBF $\gamma$ y + ggHZy (low contribution)

<b>mc16a</b>	HyGr	Zgam	Zqcd	Top	W	VV	Wgam	HZy	bkgs
<b>ee channel</b>	5.19 $\pm 0.12$	89 $\pm 13$	56 $\pm 19$	36.1 $\pm 1.6$	9.77 $\pm 0.53$	6.29 $\pm 0.37$	1.91 $\pm 0.96$	0.0689 $\pm 0.0019$	198 $\pm 23$
<b><math>\mu\mu</math> channel</b>	6.49 $\pm 0.14$	176 $\pm 21$	153 $\pm 35$	38.4 $\pm 1.7$	9.51 $\pm 0.60$	7.07 $\pm 0.43$	0.40 $\pm 0.29$	0.0885 $\pm 0.0022$	385 $\pm 41$

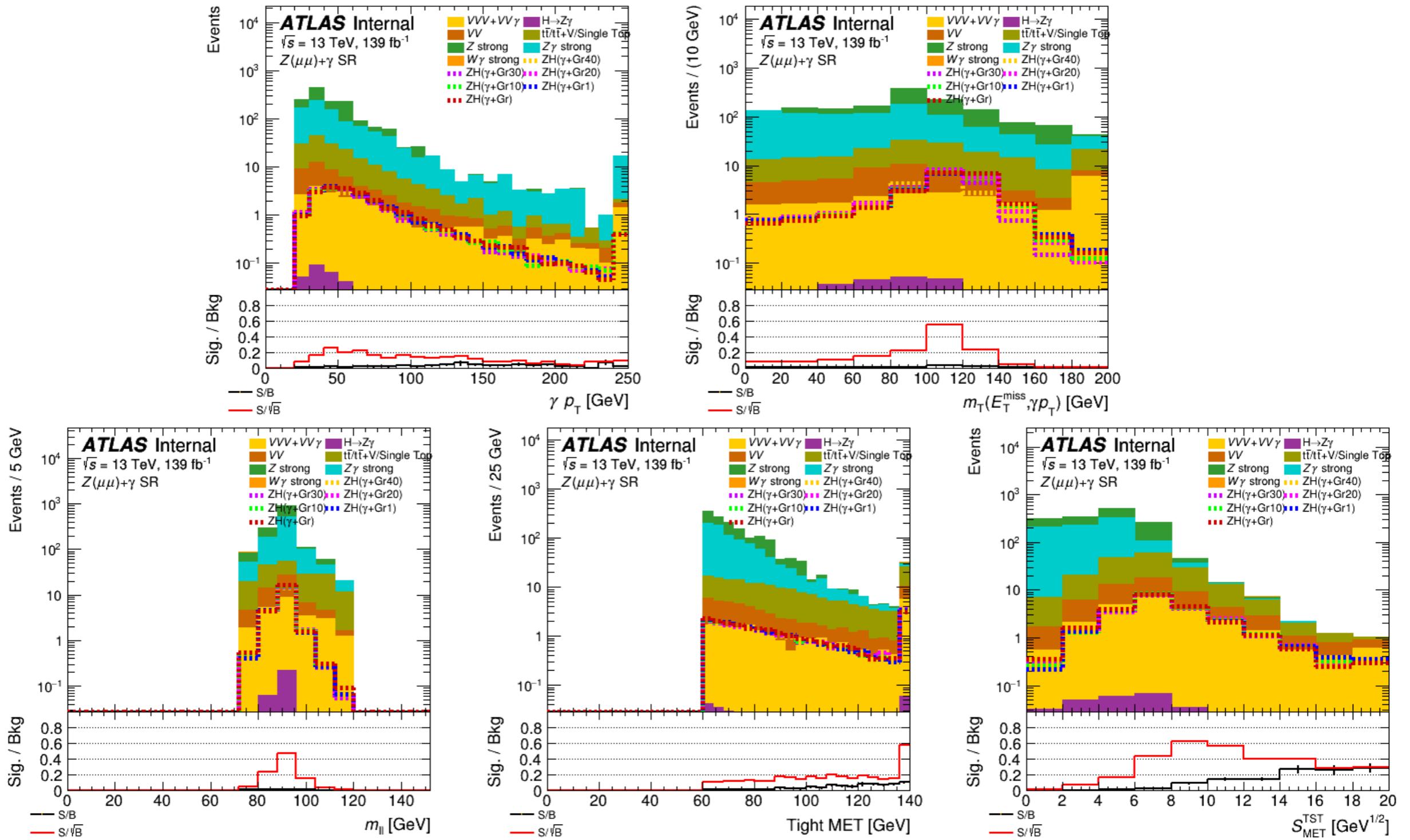
<b>mc16d</b>	HyGr	Zgam	Zqcd	Top	W	VV	Wgam	HZy	bkgs
<b>ee channel</b>	5.86 $\pm 0.13$	95 $\pm 12$	56 $\pm 29$	35.6 $\pm 1.6$	10.32 $\pm 0.44$	6.49 $\pm 0.41$	4.5 $\pm 2.4$	0.0762 $\pm 0.0020$	209 $\pm 31$
<b><math>\mu\mu</math> channel</b>	7.25 $\pm 0.14$	256 $\pm 21$	216 $\pm 47$	37.3 $\pm 1.6$	11.46 $\pm 0.47$	6.88 $\pm 0.39$	0.70 $\pm 0.41$	0.1007 $\pm 0.0024$	528 $\pm 51$

<b>mc16e</b>	HyGr	Zgam	Zqcd	Top	W	VV	Wgam	HZy	bkgs
<b>ee channel</b>	8.12 $\pm 0.15$	134 $\pm 13$	260 $\pm 53$	47.9 $\pm 1.8$	15.62 $\pm 0.68$	9.04 $\pm 0.48$	4.4 $\pm 1.5$	0.0943 $\pm 0.0023$	471 $\pm 55$
<b><math>\mu\mu</math> channel</b>	9.47 $\pm 0.17$	318 $\pm 20$	200 $\pm 55$	55.1 $\pm 2.0$	14.42 $\pm 0.65$	9.76 $\pm 0.46$	-1.7 $\pm 1.8$	0.1275 $\pm 0.0027$	596 $\pm 58$

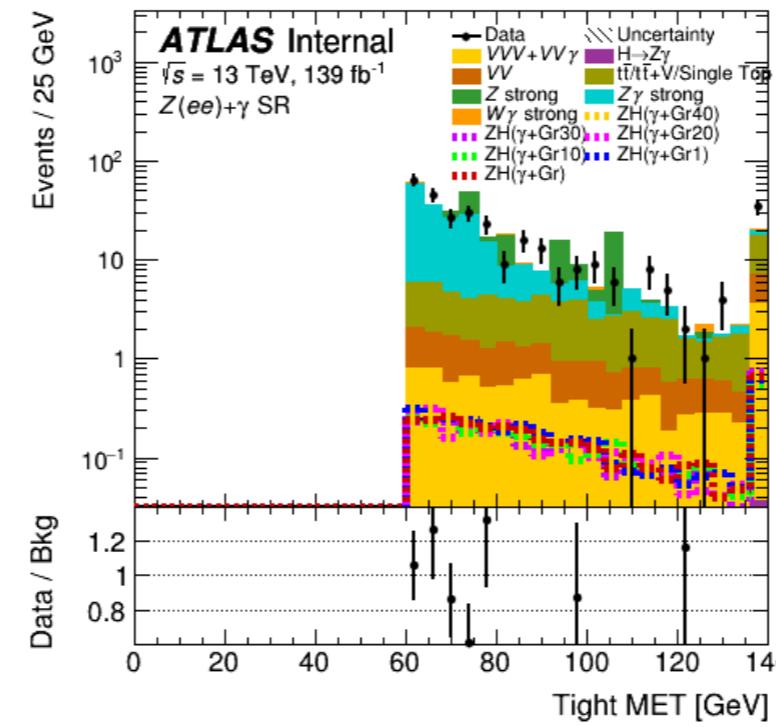
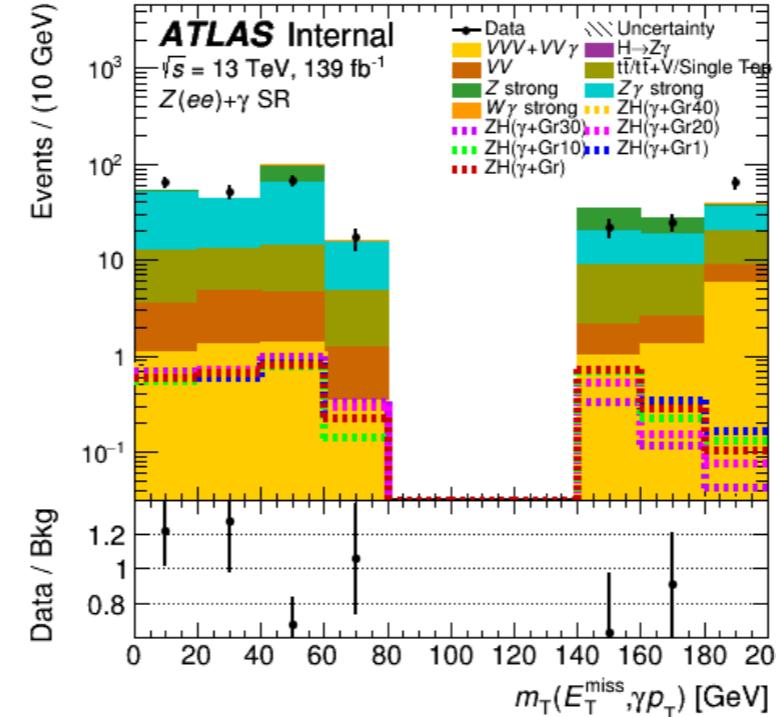
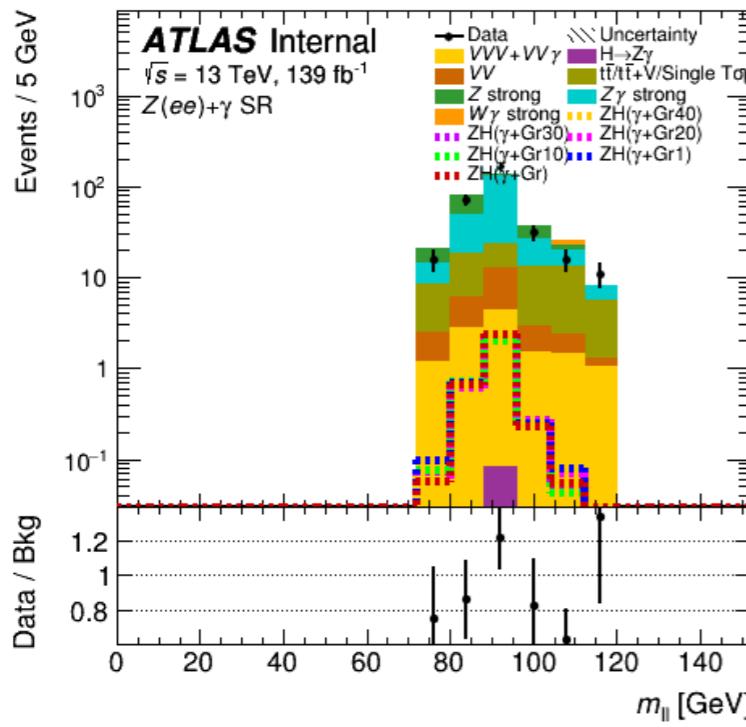
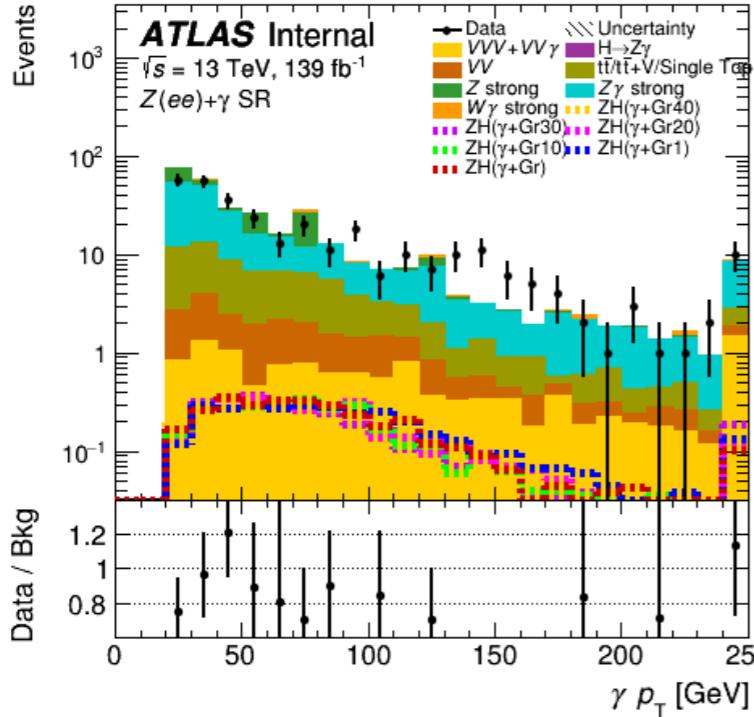
# Distributions in SR: ee-channel



# Distributions in SR: uu-channel

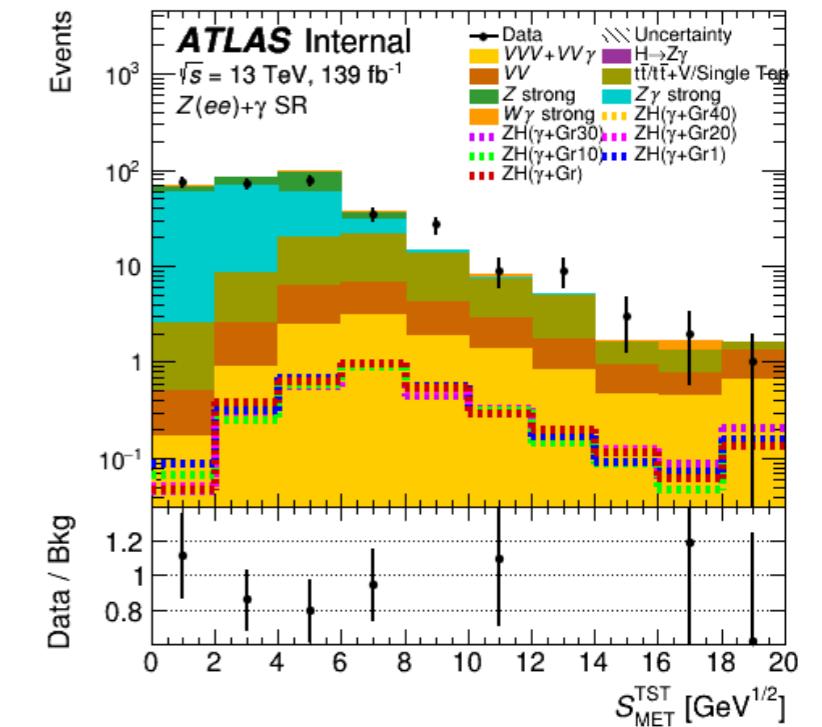
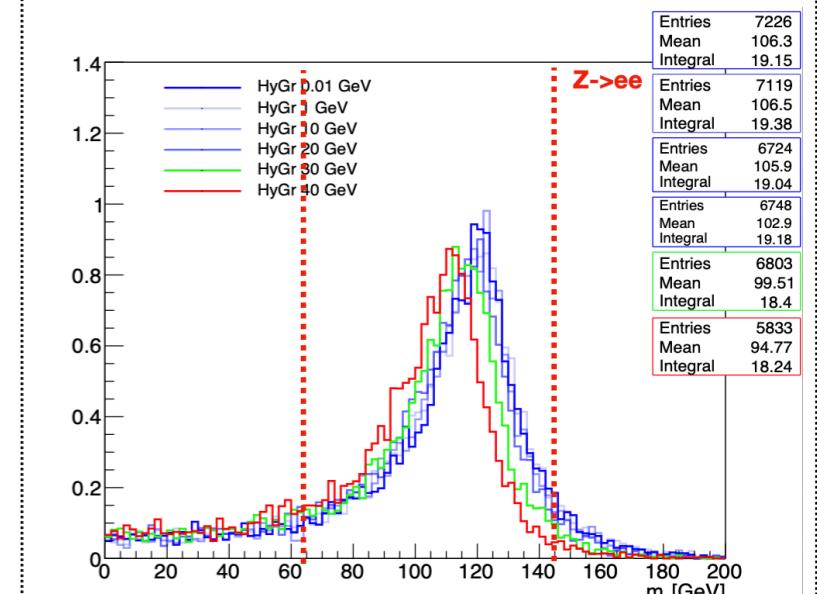


# Data/MC in blinded SR: ee-channel

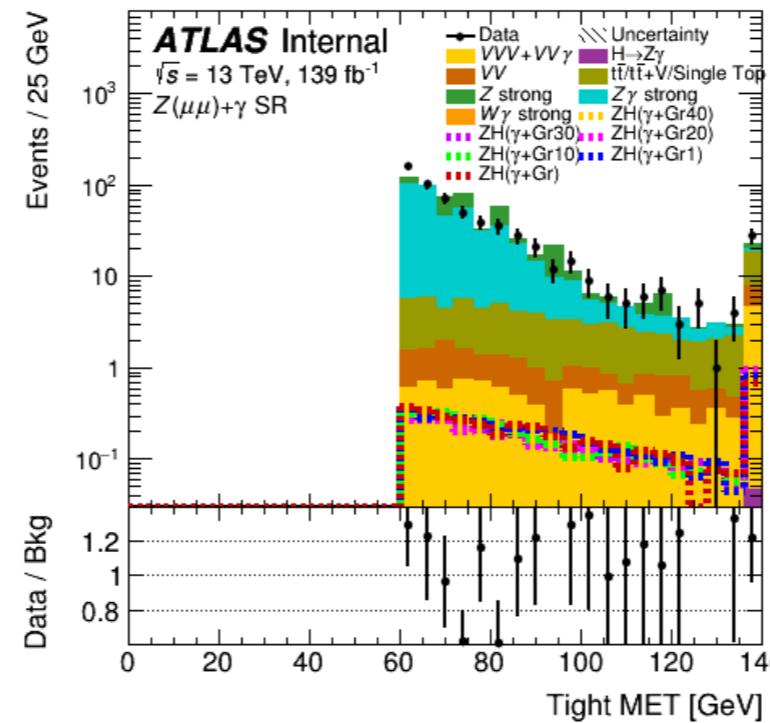
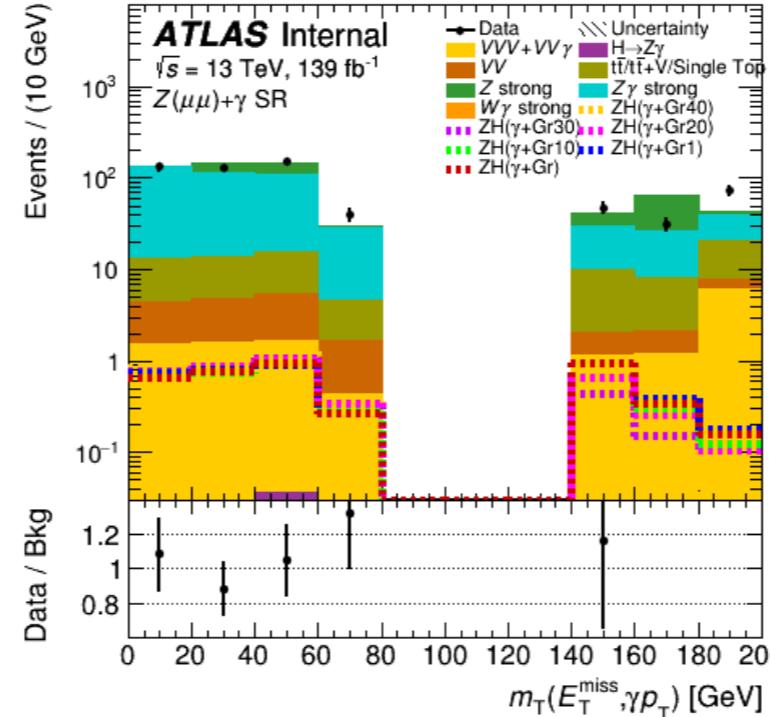
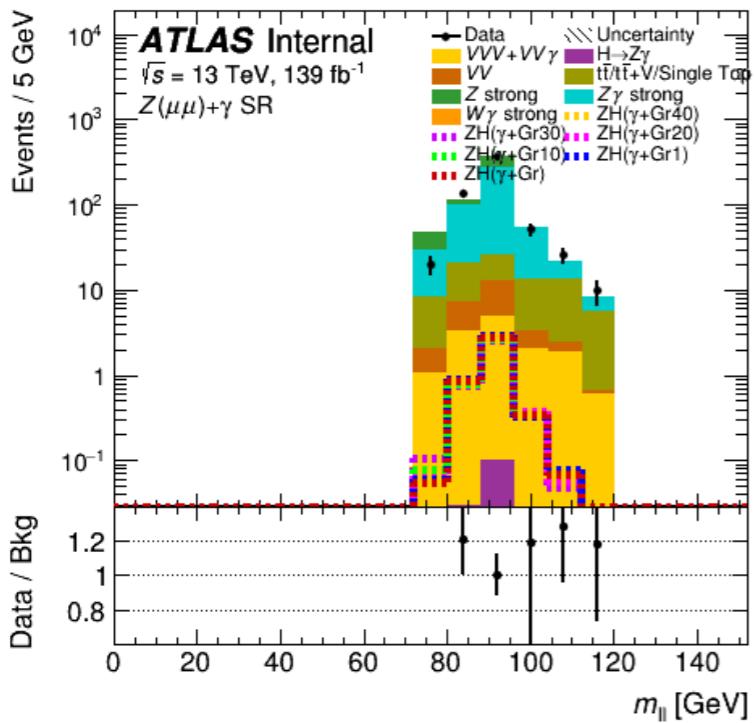
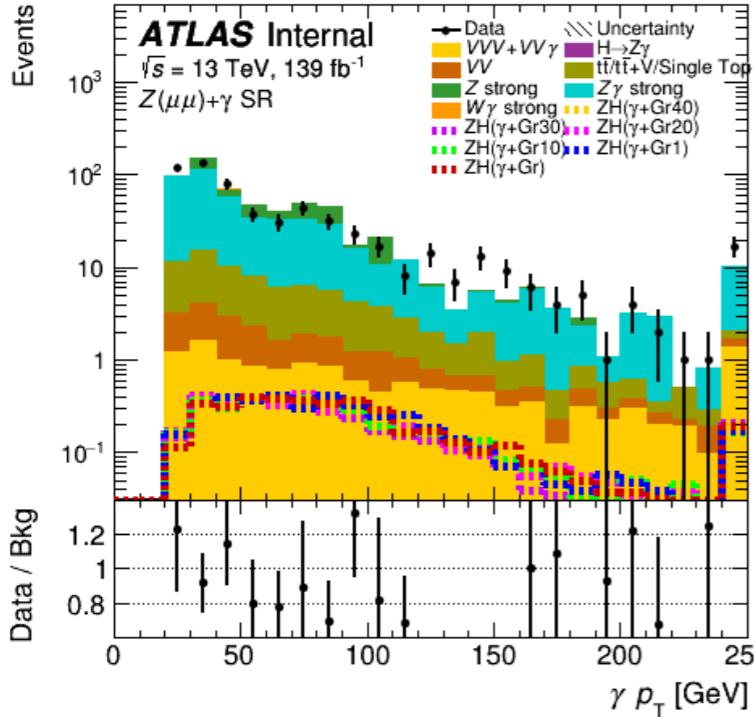


**Blinding**

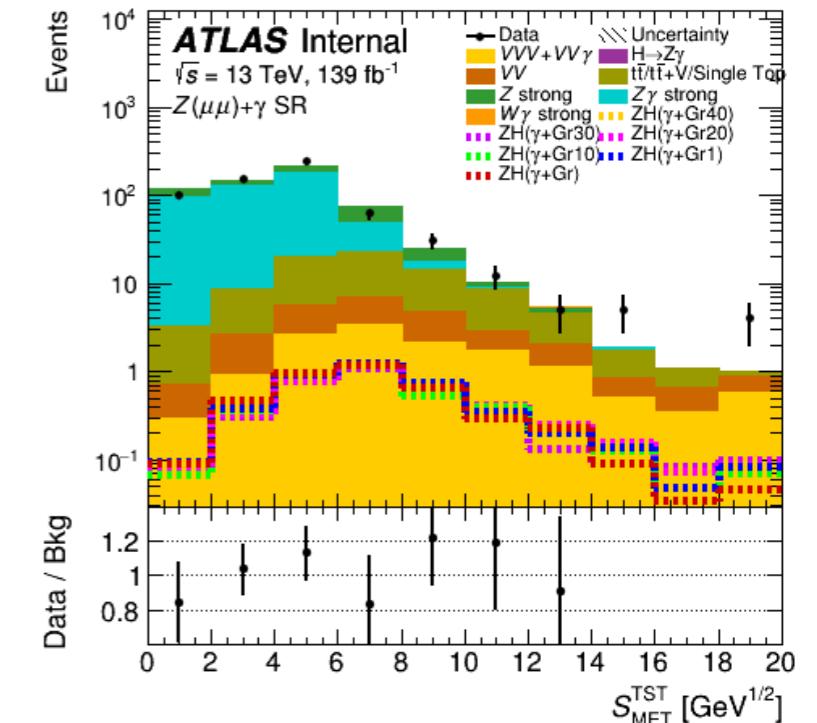
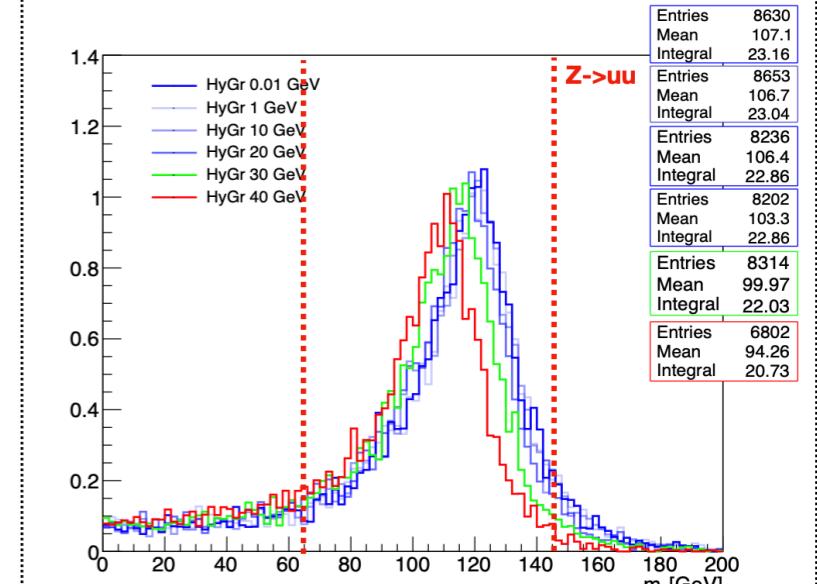
$m_T < 65 \text{ GeV} \parallel m_T > 145 \text{ GeV}$   
 $\Rightarrow$  about 82% signal rejected



# Data/MC in blinded SR: uu-channel



**Blinding**  
 $m_T < 65 \text{ GeV} \parallel m_T > 145 \text{ GeV}$   
 $\Rightarrow$  about 82% signal rejected

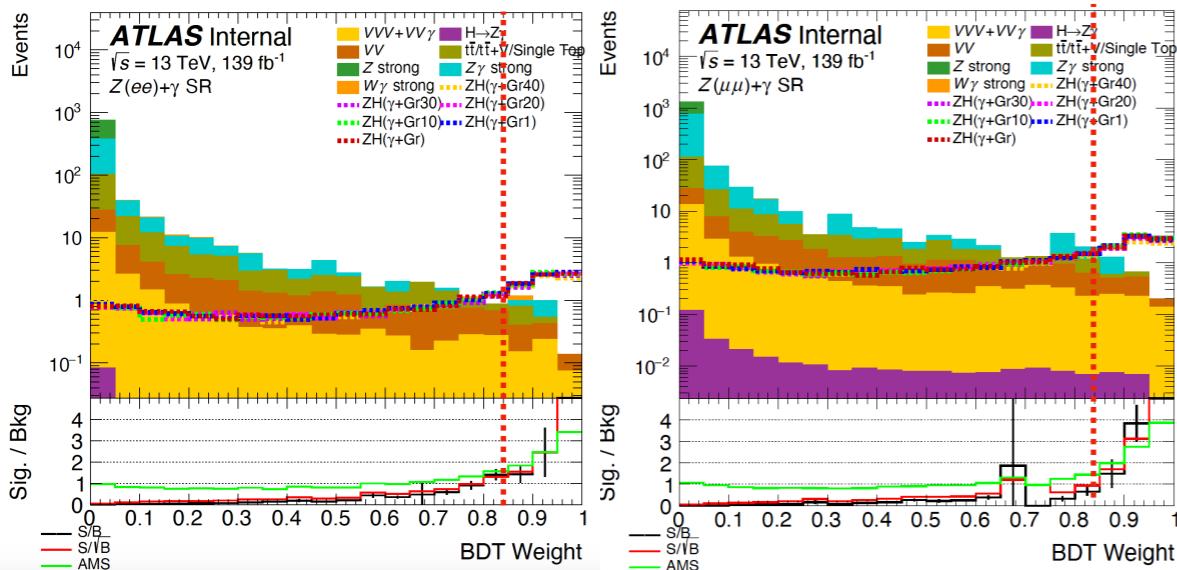


# SR optimization with Machine Learning

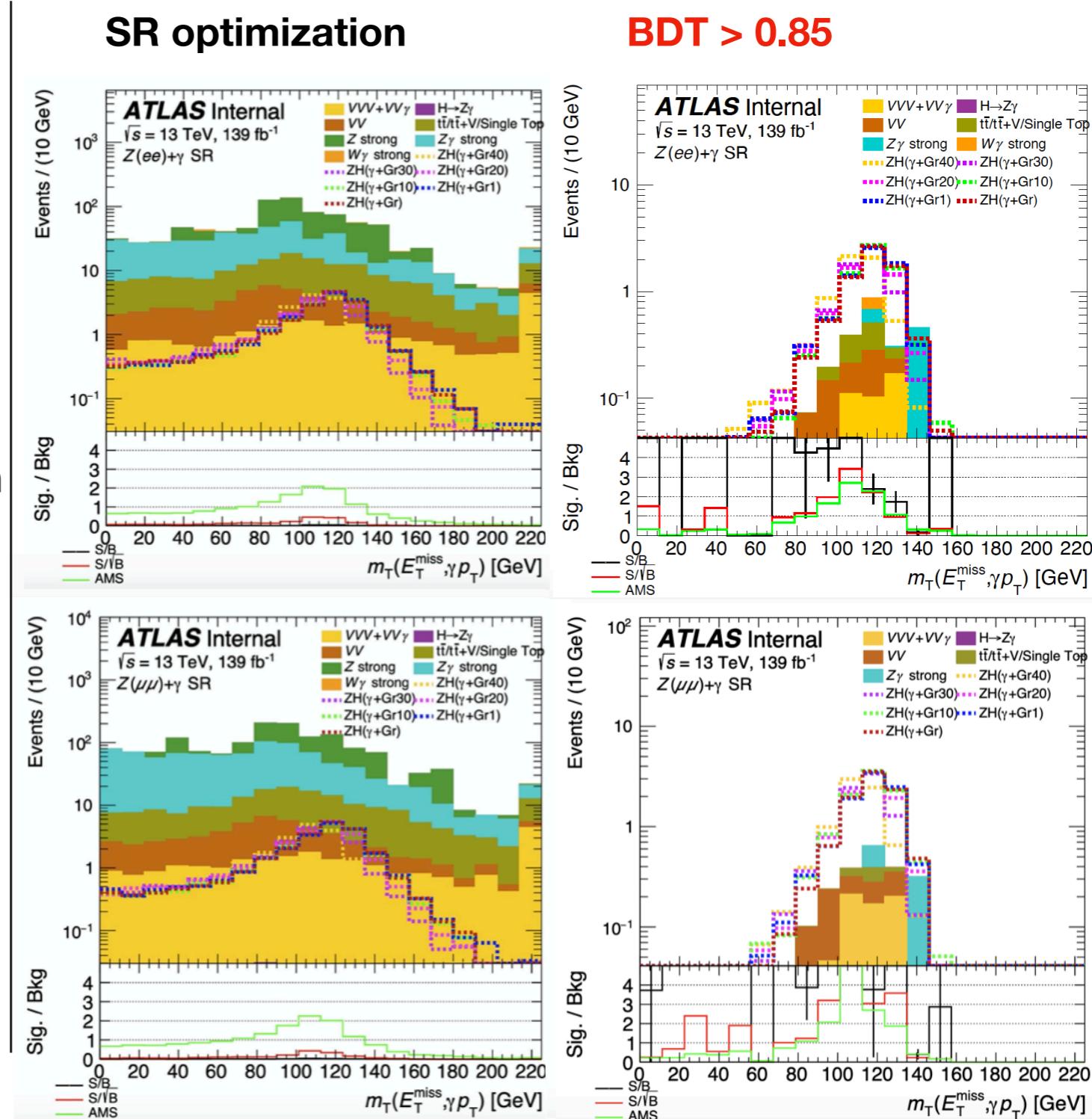
## Impact of BDT cut in SR

- Input variables: met\_tight\_tst\_et , met\_tight\_tst\_phi , mT , ph\_pt , dphi\_mety\_ll , AbsPt , Ptll , mllg , lep1pt , lep2pt , mll , metsig\_tst , Ptllg ,dphi\_met\_ph
- Hyperparameter tuning performed with [hpogrid](#)

Optimisation of cut on BDT weight based on AMS metric:  $\sqrt{2[(s + b)\ln(1 + s/b) - s]}$



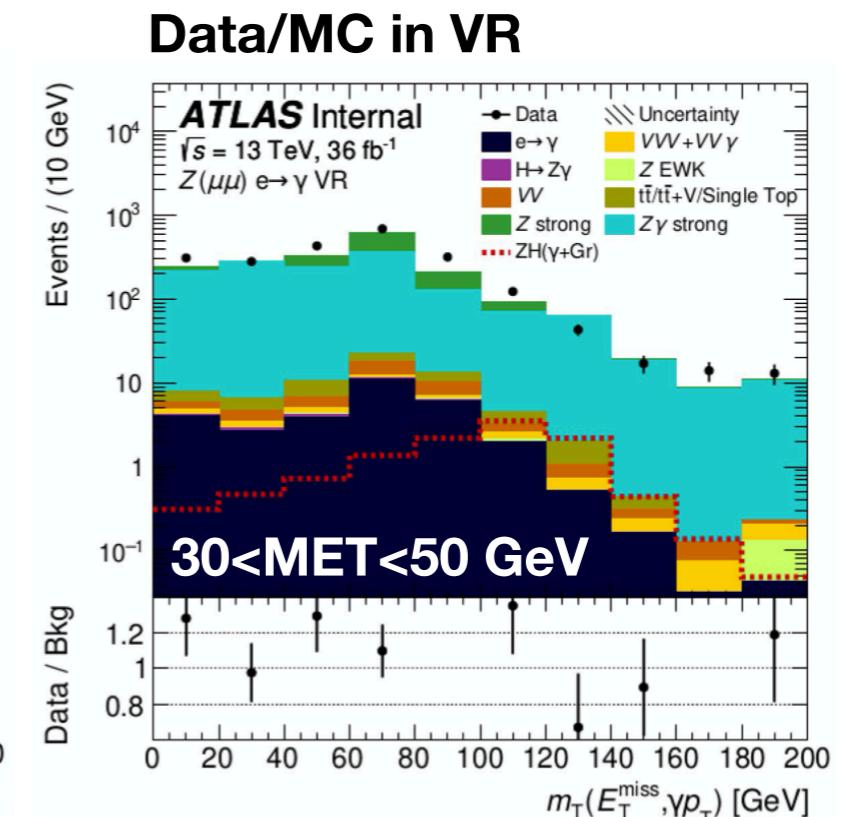
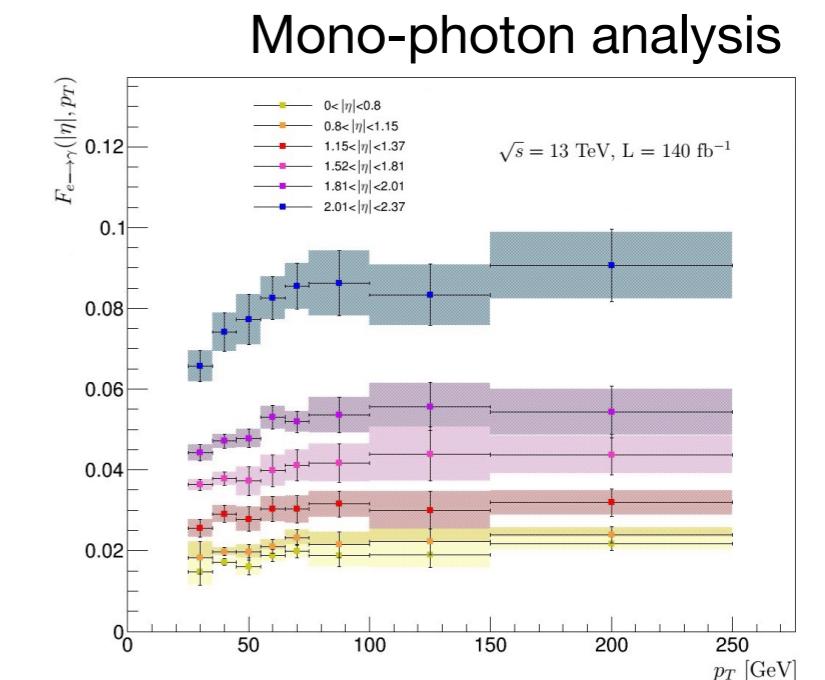
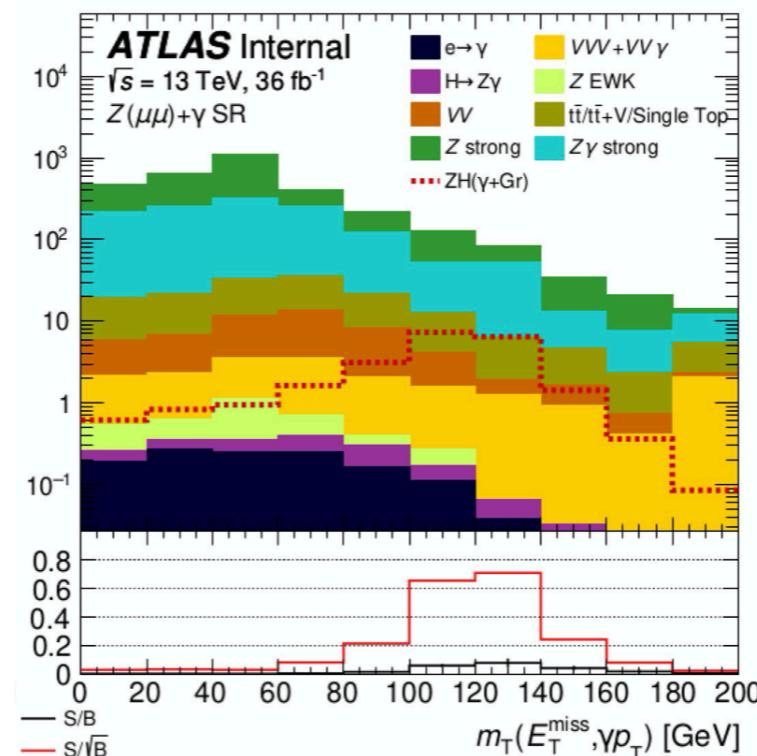
**Work in progress**



# Electrons-faking-photons

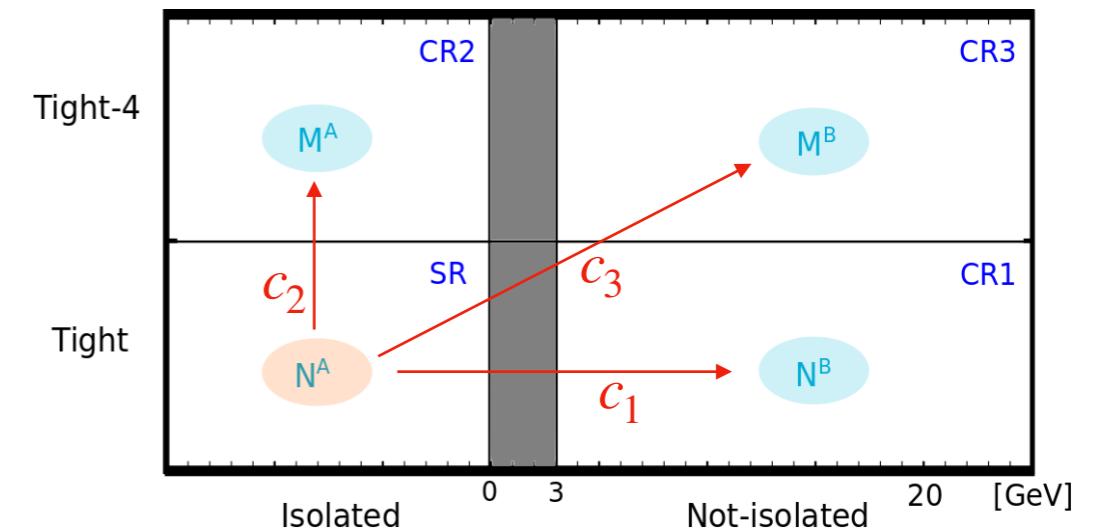
- Mainly  $W(e\nu)Z(l\bar{l})$  events with the electron from  $W$  decay misidentified as a photon
- Using  $e$ -to- $\gamma$  fake rate from mono-photon analysis
- Data yields in the  $WZ$  CR (defined as the SR, but requiring an electron in place of the signal photon) rescaled by fake-rate
- Similar procedure will be applied to all the analysis CRs
- Will be bachelor thesis project of Andrea Mitta

Preliminary results (B. Nino, S. Sevova), using previous version of the ntuples (AthAnalysis 21.2.123)



# Jets-faking-photons

- Mainly from  $Z+jets$  events, with one jet misidentified as a photon
- ABDC method, based on photon isolation and identification



- Number of jets-faking-photons:  $N_{j \rightarrow \gamma} = (1 - P)N_A$  with  $P = N_A^{sig}/N_A$

$$N_A^{sig} = N_A - R_{MC} \frac{(N_B - c_1 N_A^{sig})(M_A - c_2 N_A^{sig})}{M_B - c_3 N_A^{sig}}$$

- The  $c_i$  and  $R_{MC}$  coefficients are evaluated from MC, and allow to take into account deviations from the basic assumptions of the method (uncorrelated variables and negligible signal leakage in the B,C and D regions )
  - $c_i = N_i/N_A$  : signal (real-photon) leakage in the i-th CR, estimated from  $Z+jets$  MC
  - $R_{MC} = (N_A M_B)/(N_B M_A)$  : correlation factor between tight and isolated regions, estimated from  $Z+jets$  MC
- Study ongoing

# Fake MET background

## A new ABCD method

(Due to mismeasured jets in Z $\gamma$ +jets process)

- Defining an ABCD method
  - Different pair of variables and cut values tested
  - MET significance or MET as a first variable (MET significance variable allows good discrimination between fake MET and real MET)
  - Several second variables considered. Most promising from preliminary studies:

$$\Delta\phi(E_T^{\text{miss}}, p_T^{\parallel\gamma})$$

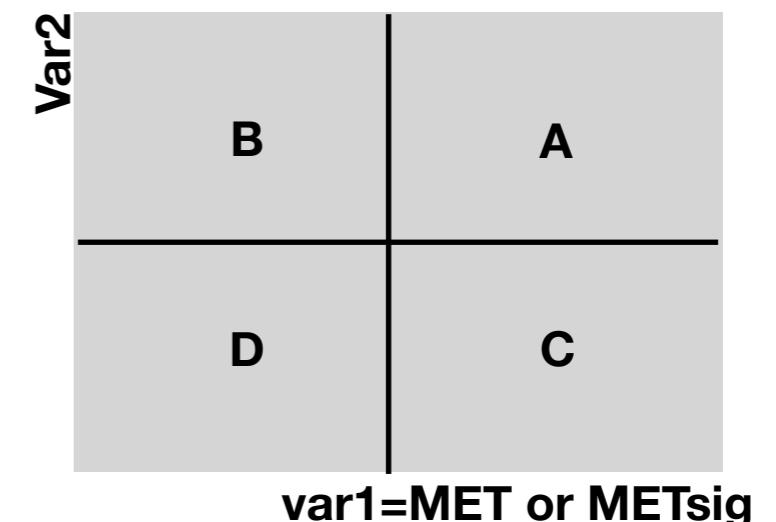
$$\Delta\phi(E_T^{\text{miss}} + p_T^\gamma, p_T^{\parallel})$$

$$\Delta\phi(|E_T^{\text{miss}} + p_T^{\parallel\gamma}|, \text{nearest}(E_T^{\text{miss}}, p_T^{\parallel\gamma}))$$

$$\Delta\phi(E_T^{\text{miss}}, \text{nearest obj.})$$

$$|E_T^{\text{miss}} / p_T^{\parallel\gamma}|$$

$$|E_T^{\text{miss}} + p_T^{\parallel\gamma}| / E_T^{\text{miss}}$$



- Further investigation based on the following criteria:
  - R stability: non dependence of R coefficient on the choice of cut values
  - R close to 1 (uncorrelated variables)
  - High signal efficiency in SR and low leakage of dark-photon and real MET events in the CRs
  - Good statistics in CRs

$$\text{Low } \chi^2_{V1V2} = \sum_{ij=\text{cut}_1\text{cut}_2} \left( \frac{R_{ij} - \langle R \rangle}{\sigma(R_{ij})} \right)^2$$

High  $(RV_{ij})_{V1V2} = \left( \frac{\epsilon_{ij}}{\sigma_{R_{ij}} \times (R_{ij} + 1/R_{ij})} \right)_{V1V2}$

# Fake MET background

## Choice of variables

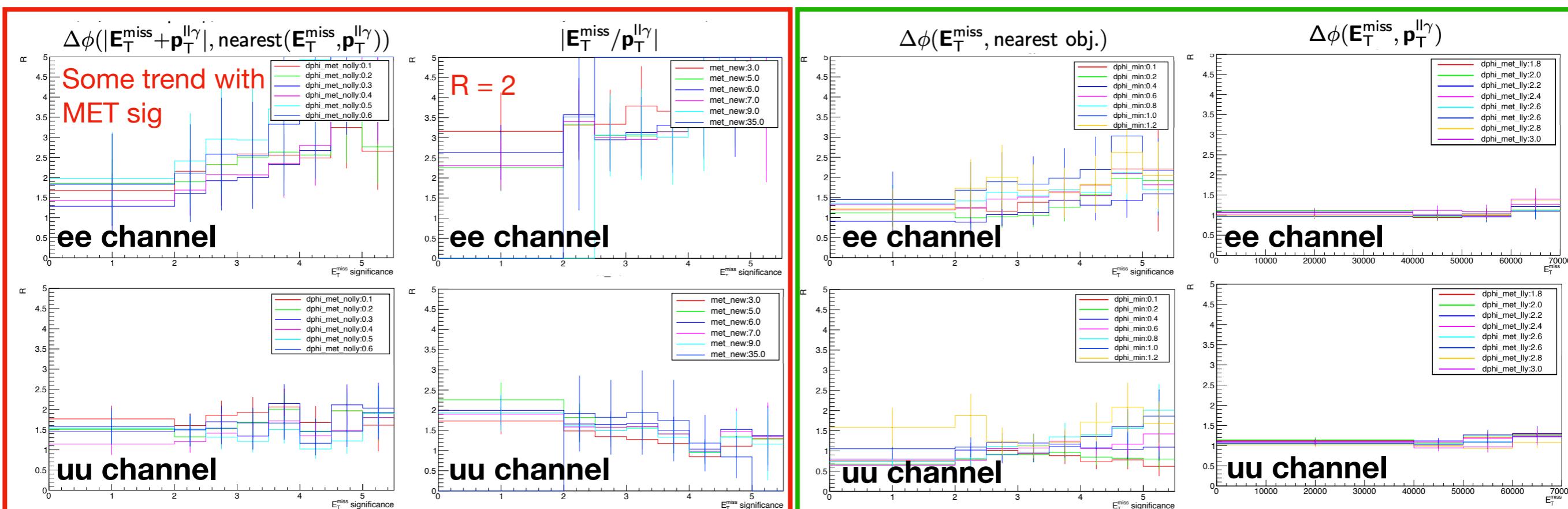
Stability tested scanning different cut values on var1 and var2, with both MET and MET significance as var1

var1	var2	mean	ee	uu
met_tight_tst_et	dphi_met_lly	0.62	0.49	0.76
metsig_tst	dphi_met_nolly	0.66	0.76	0.56
metsig_tst	met_over_ptlly	0.71	0.50	0.93
metsig_tst	dphi_min	0.83	0.61	1.06
metsig_tst	mlly	0.88	0.73	1.04
metsig_tst	dphi_met_lly	0.94	1.12	0.76
metsig_tst	dphi_mety_ll	1.07	1.02	1.12
met_tight_tst_et	dphi_min	1.14	1.36	0.92
met_tight_tst_et	dphi_mety_ll	1.25	0.49	2.01
met_tight_tst_et	dphi_met_nolly	2.36	3.30	1.43
met_tight_tst_et	met_over_ptlly	2.44	3.25	1.64

chi2 in ee-channel, uu-channel and the mean of the two reported in columns ee, uu, and mean respectively

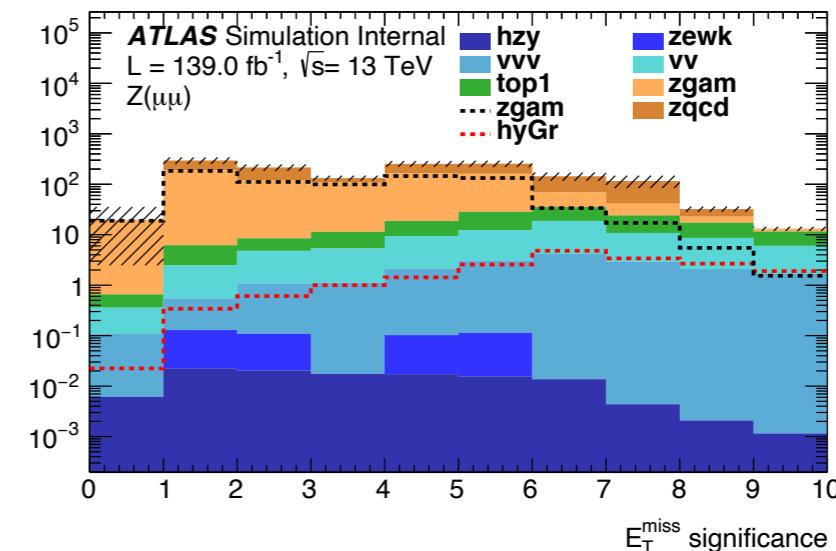
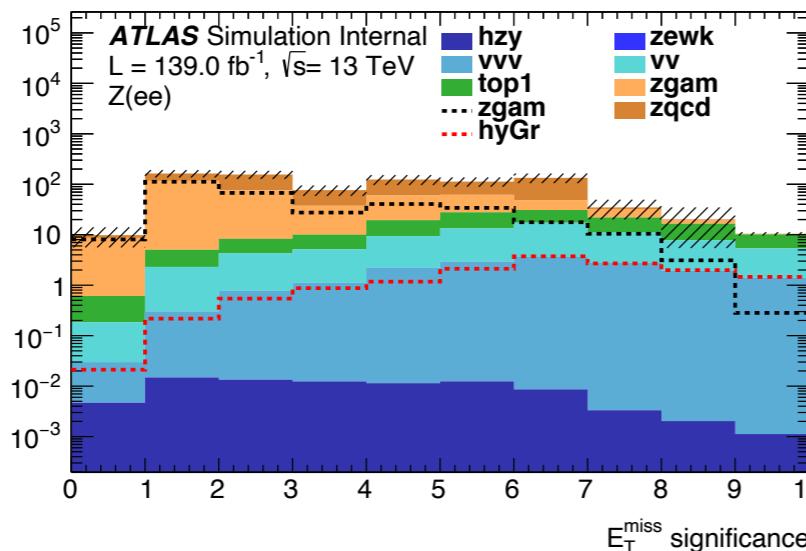
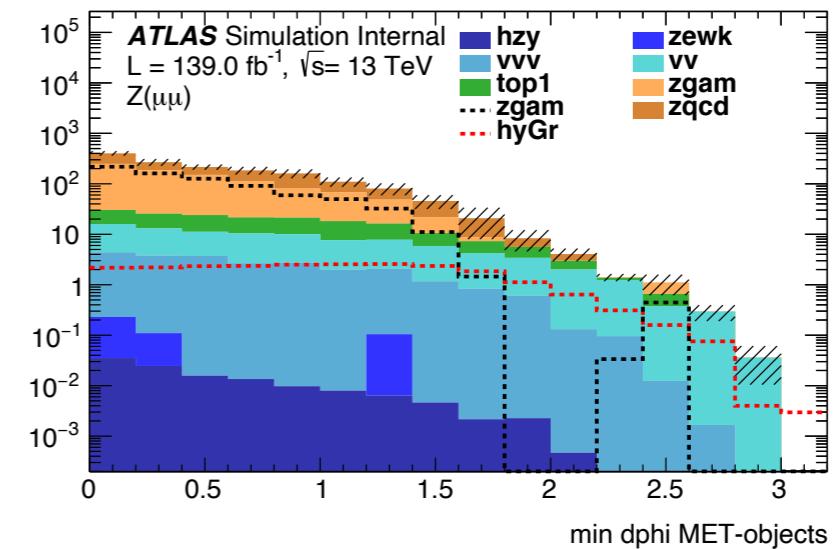
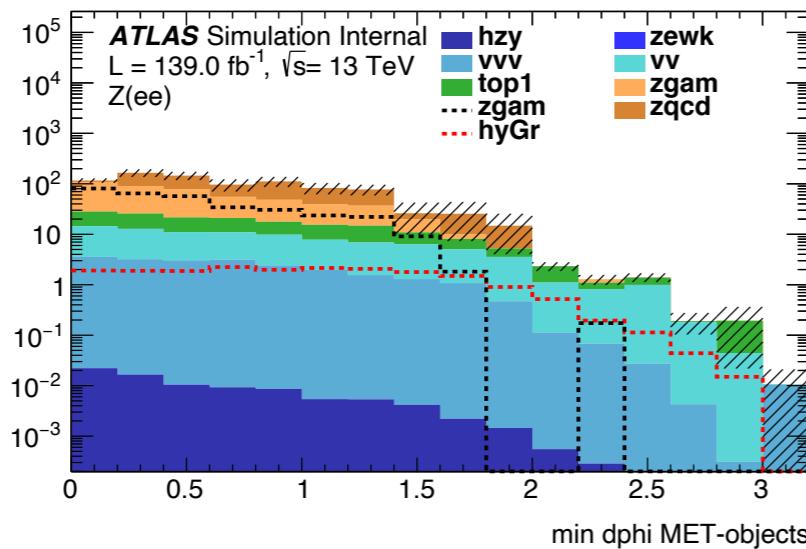
dphi_met_lly	$\Delta\phi(\mathbf{E}_T^{\text{miss}}, \mathbf{p}_T^{\parallel\gamma})$
dphi_mety_ll	$\Delta\phi(\mathbf{E}_T^{\text{miss}} + \mathbf{p}_T^{\gamma}, \mathbf{p}_T^{\parallel\gamma})$
dphi_met_nolly	$\Delta\phi( \mathbf{E}_T^{\text{miss}} + \mathbf{p}_T^{\parallel\gamma} , \text{nearest}(\mathbf{E}_T^{\text{miss}}, \mathbf{p}_T^{\parallel\gamma}))$
dphi_min	$\Delta\phi(\mathbf{E}_T^{\text{miss}}, \text{nearest obj.})$
met_over_ptlly	$ \mathbf{E}_T^{\text{miss}} / \mathbf{p}_T^{\parallel\gamma} $
met_nolly	$ \mathbf{E}_T^{\text{miss}} + \mathbf{p}_T^{\parallel\gamma}  / \mathbf{E}_T^{\text{miss}}$

Best choices based on R stability



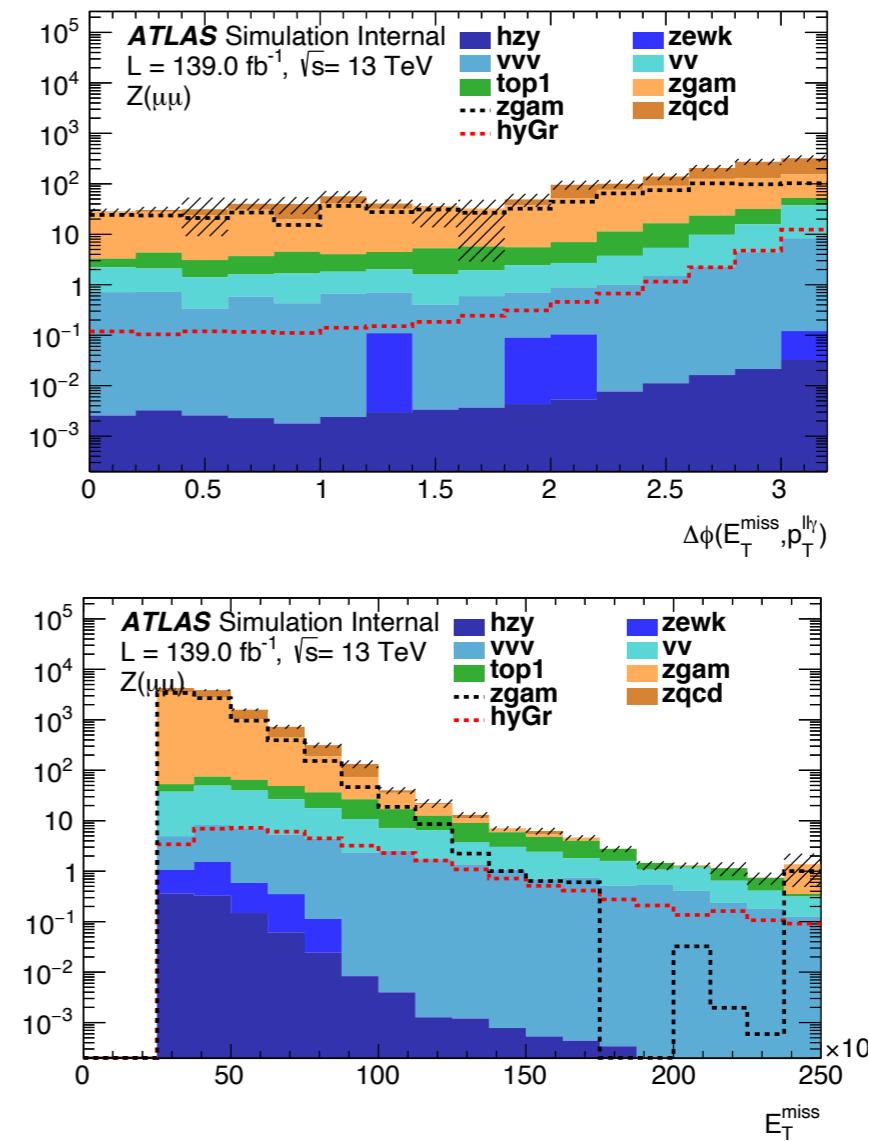
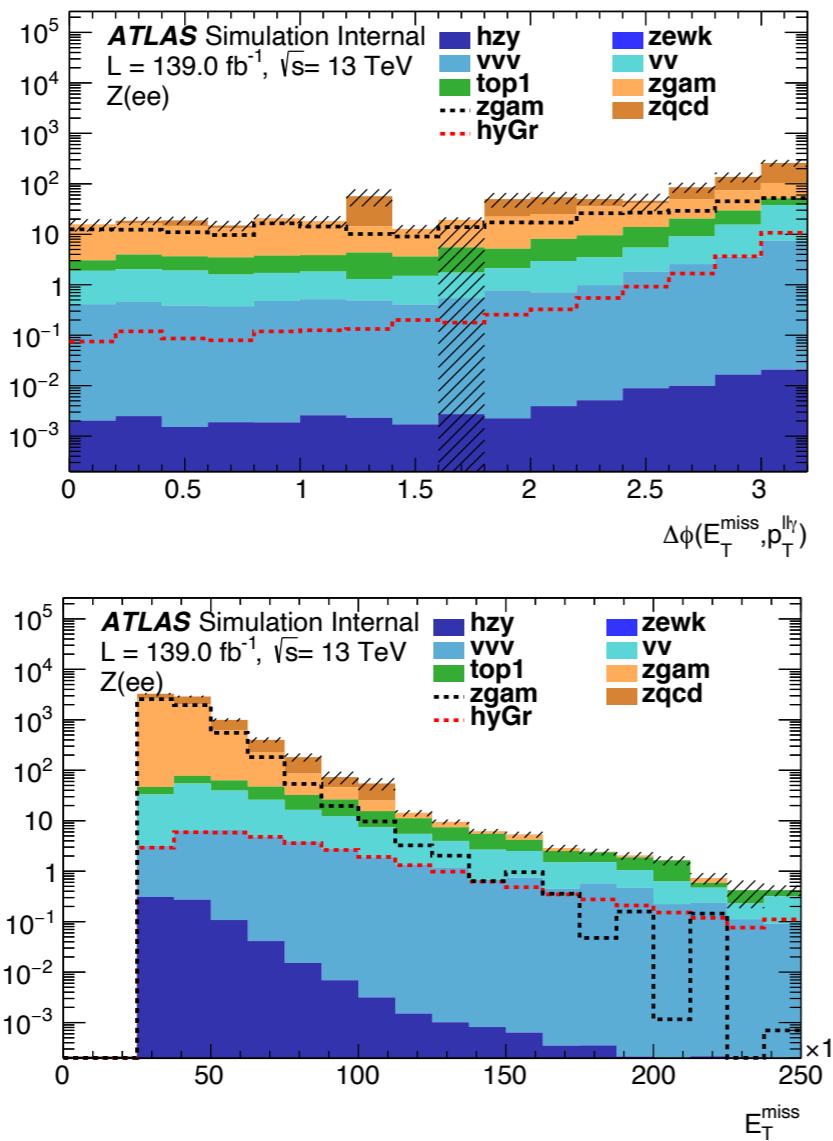
# Fake MET background

## Variable distributions for METsig-based ABCD



# Fake MET background

## Variable distributions for MET-based ABCD



# Fake MET background

## Including METsig cut in MET-based ABCD

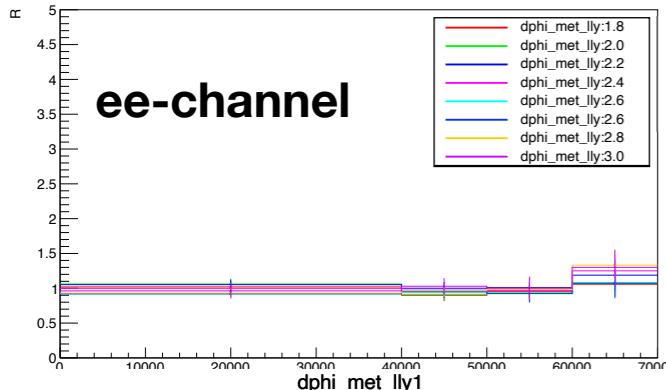
- MET significance cut improves sensitivity thanks to high discrimination between real MET and fake MET
- Trying to include a MET significance cut in high MET regions, in the MET-based ABCD

var1	var2	metsig	mean	ee	uu
met_tight_tst_et	dphi_met_lly	0.0	0.62	0.49	0.76
met_tight_tst_et	dphi_met_lly	3.0	0.92	0.70	1.13
met_tight_tst_et	dphi_met_lly	4.0	1.78	1.22	2.33

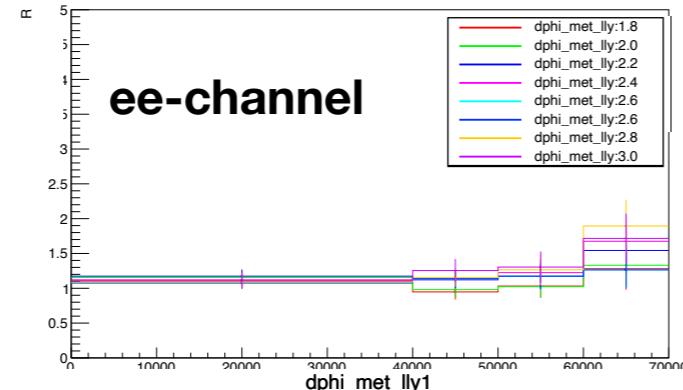


$$dphi_{met\_lly} = \Delta\phi(E_T^{\text{miss}}, p_T^{\gamma})$$

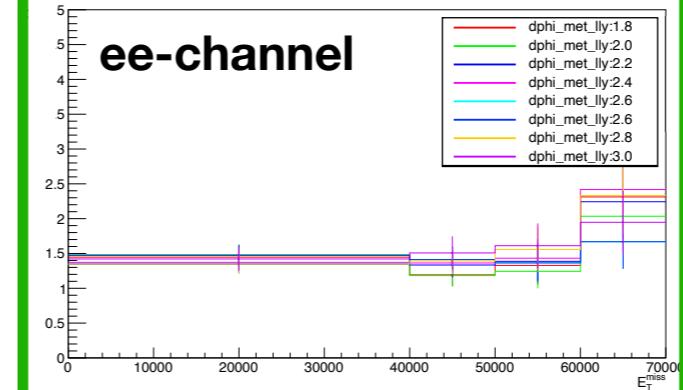
No MET sig cut



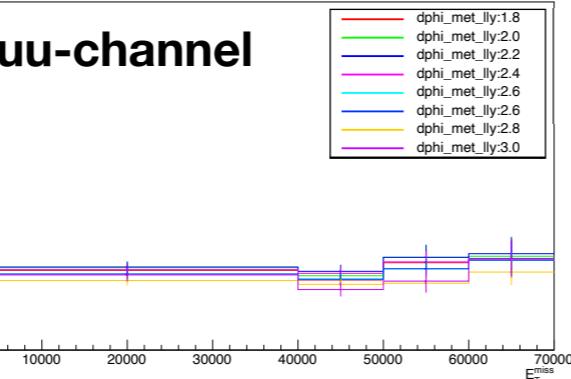
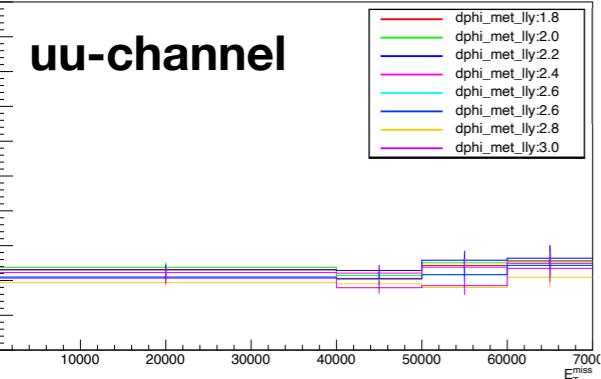
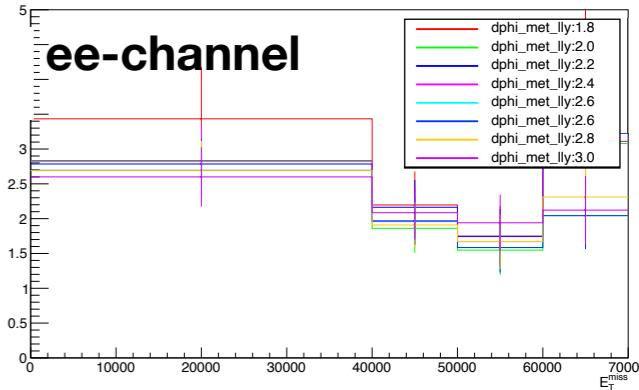
MET sig > 2



MET sig > 3



MET sig > 4



# Fake MET background

## Best cut values

Considering best MET-based and best METsig-based ABCD

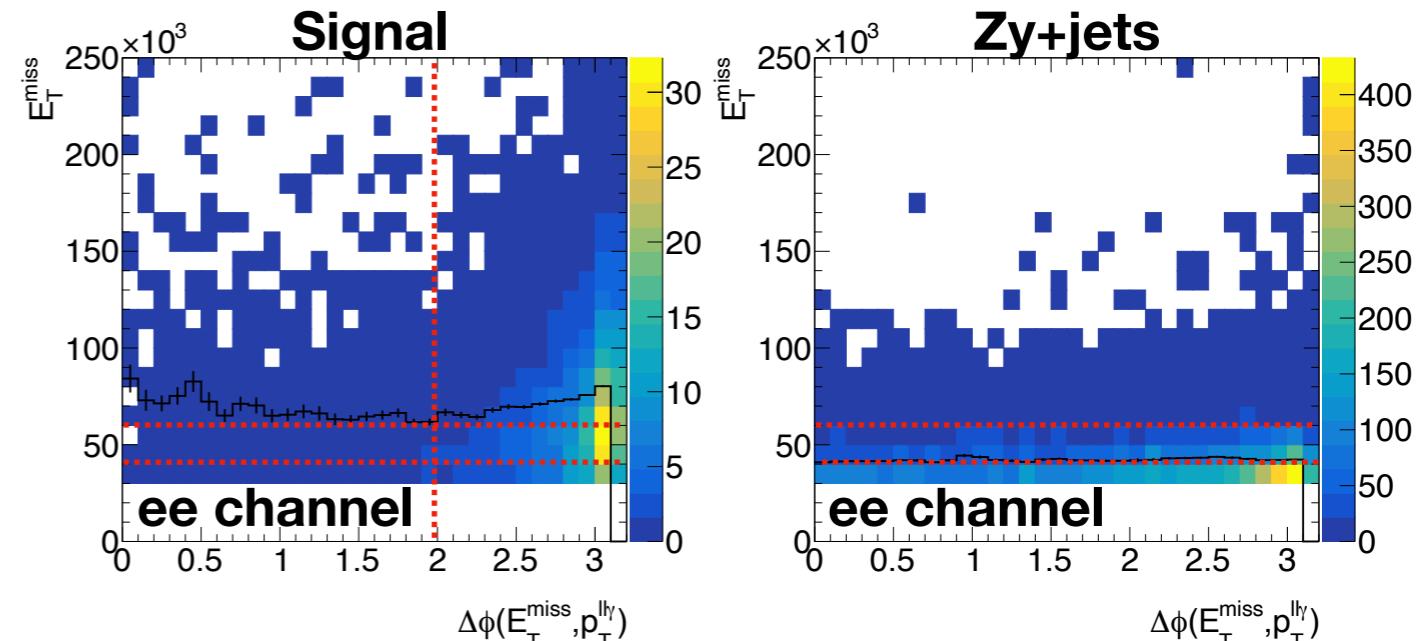
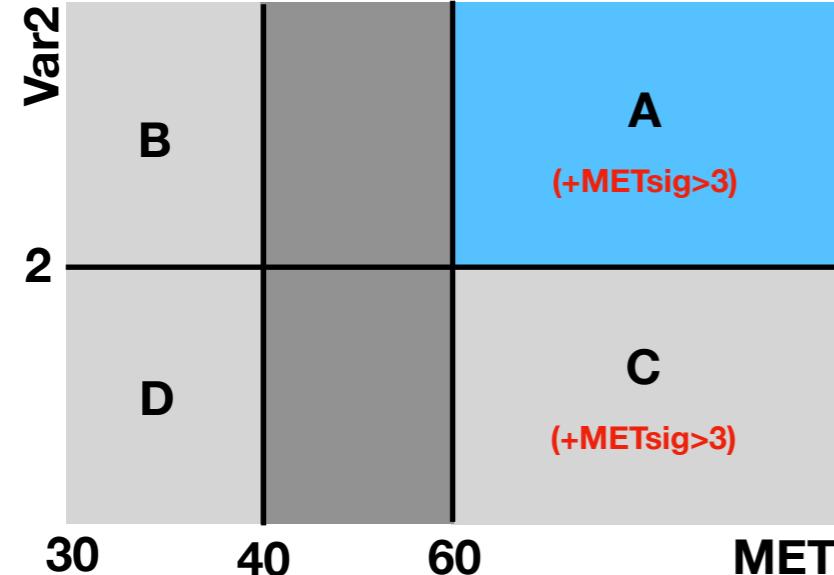
ee-channel	var1	var2	metsig	cut1	cut2	R
	met_tight_tst_et	dphi_met_lly	3.0	40000.0	2.0	1.28 +\/- 0.25
	metsig_tst	dphi_min	3.0	3.0	0.2	1.02 +\/- 0.27
	metsig_tst	dphi_min	4.0	4.0	0.2	1.26 +\/- 0.35
uu-channel	var1	var2	metsig	cut1	cut2	R
	met_tight_tst_et	dphi_met_lly	3.0	40000.0	2.0	1.23 +\/- 0.17
	metsig_tst	dphi_min	3.0	3.0	0.2	0.92 +\/- 0.19
	metsig_tst	dphi_min	4.0	4.0	0.2	0.96 +\/- 0.20

METsig-dphi min						
sample	V1	V2	ee-channel		uu-channel	
			nA	nB	nC	nD
hyGr	4	0.2	318.4+/-4.3	26.9+/-1.2	32.0+/-1.4	6.2+/-0.6
zgam	4	0.2	83.8+/-11.3	158.3+/-15.8	23.9+/-4.2	56.8+/-7.9
zqcd	4	0.2	202.3+/-49.1	161.7+/-39.5	2.1+/-5.4	5.4+/-5.6
realMET	4	0.2	127.7+/-2.7	18.0+/-1.0	22.2+/-1.1	6.0+/-0.6

MET-dphi met lly						
sample	V1	V2	ee-channel		uu-channel	
			nA	nB	nC	nD
hyGr	40	2	342.2+/-4.4	69.9+/-2.0	25.6+/-1.2	10.5+/-0.8
zgam	40	2	91.1+/-11.8	1955.2+/-56.0	44.1+/-6.4	1212.1+/-45.2
zqcd	40	2	216.5+/-50.1	739.3+/-97.0	27.0+/-14.7	199.3+/-41.8
realMET	40	2	123.9+/-2.6	44.5+/-1.3	36.1+/-1.5	19.2+/-1.0

# Fake MET background

## ABCD regions definition



- Add a gap region for  $40 < \text{MET} < 60$  GeV (reduce signal leakage in B and D regions, leaves room for VR)
- MET cut: 40 GeV
- $\text{dphi}(\text{MET}, \text{l}^{\gamma})$  cut: 2
- MET significance  $> 3$  in high MET regions

$R = 1.28 +/- 0.25$  in ee channel  
 $R = 1.23 +/- 0.17$  in uu channel

### ee channel

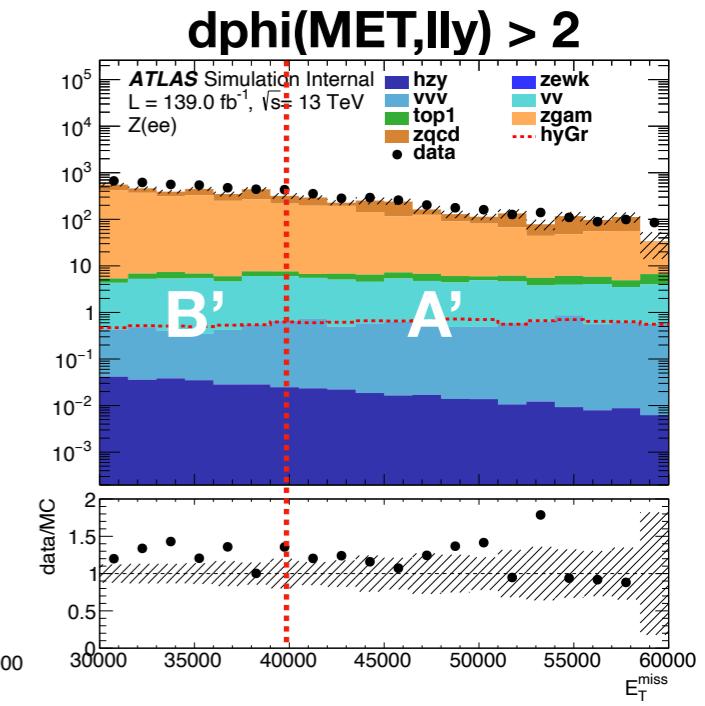
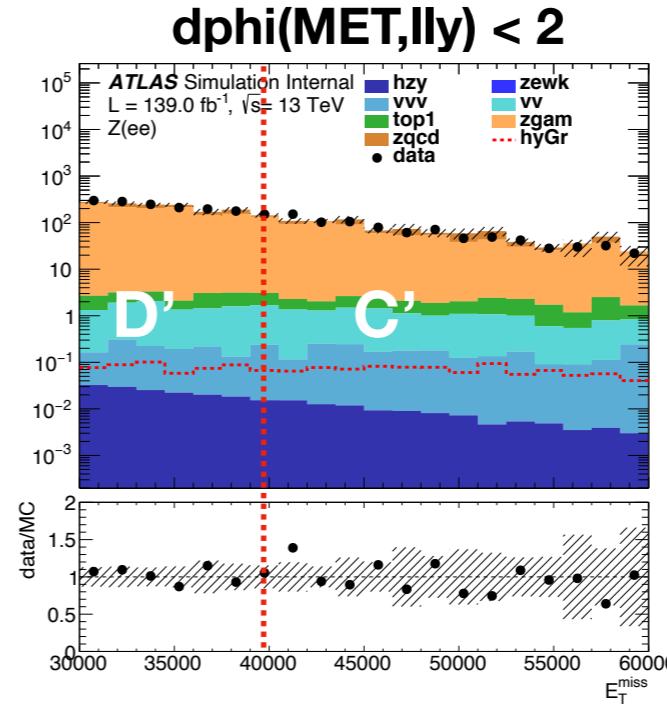
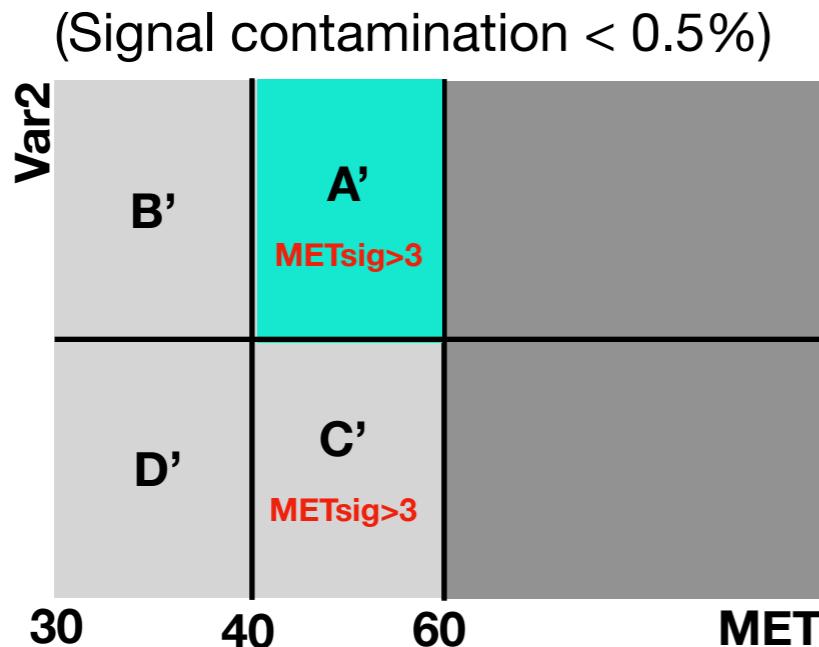
Sample	nA	nB	nC	nD
hyGr	$17.11 +/- 0.22$	$3.49 +/- 0.10$	$1.28 +/- 0.06$	$0.52 +/- 0.04$
zgam	$91.1 +/- 11.8$	$1955.2 +/- 56.1$	$44.1 +/- 6.4$	$1212.1 +/- 45.2$
zqcd	$216.5 +/- 50.1$	$739.3 +/- 97.0$	$27.0 +/- 14.7$	$199.3 +/- 41.8$
realMET	$123.9 +/- 2.6$	$44.5 +/- 1.3$	$36.1 +/- 1.5$	$19.2 +/- 1.0$

### $\mu\mu$ channel

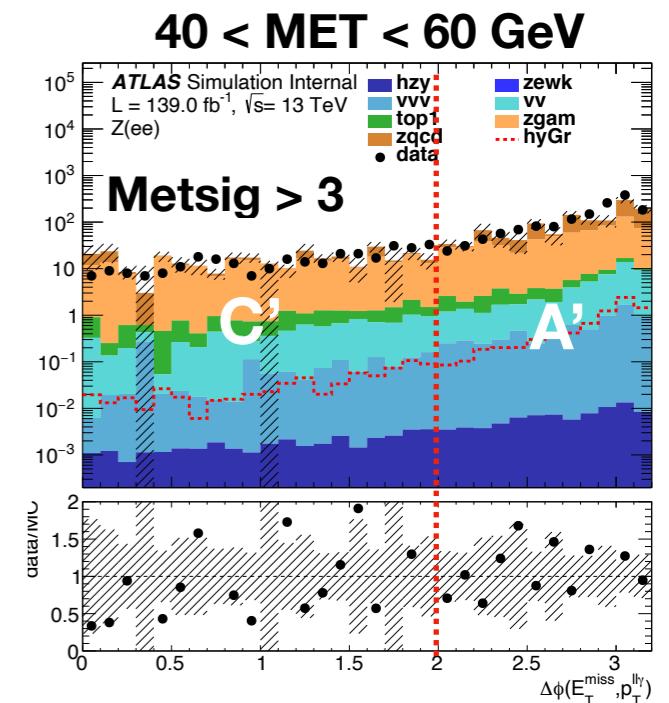
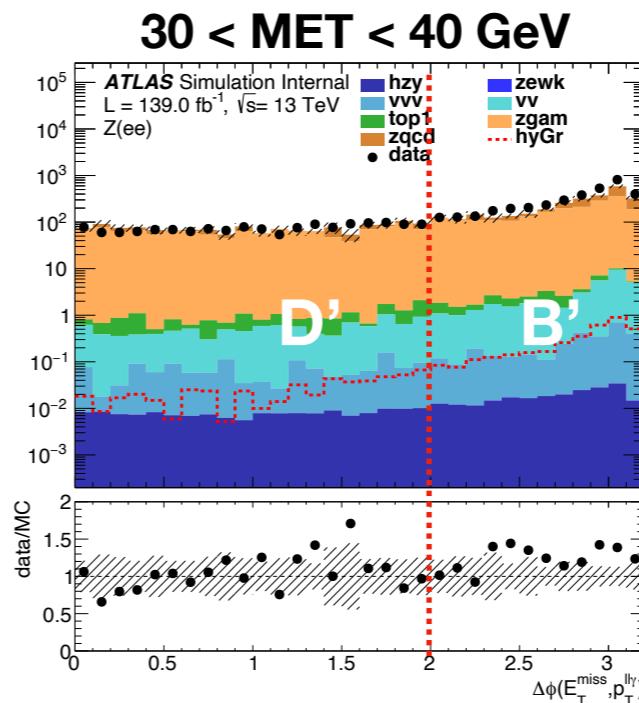
Sample	nA	nB	nC	nD
hyGr	$20.76 +/- 0.25$	$4.03 +/- 0.11$	$1.48 +/- 0.07$	$0.68 +/- 0.04$
zgam	$274.7 +/- 20.5$	$2407.9 +/- 64.5$	$161.7 +/- 18.6$	$1741.0 +/- 49.9$
zqcd	$335.1 +/- 59.0$	$1079.1 +/- 113.2$	$32.7 +/- 15.1$	$156.4 +/- 50.3$
realMET	$131.7 +/- 2.8$	$49.6 +/- 1.5$	$39.3 +/- 1.5$	$19.3 +/- 1.0$

# Fake MET background

## Preliminary validation: data/MC in low MET region



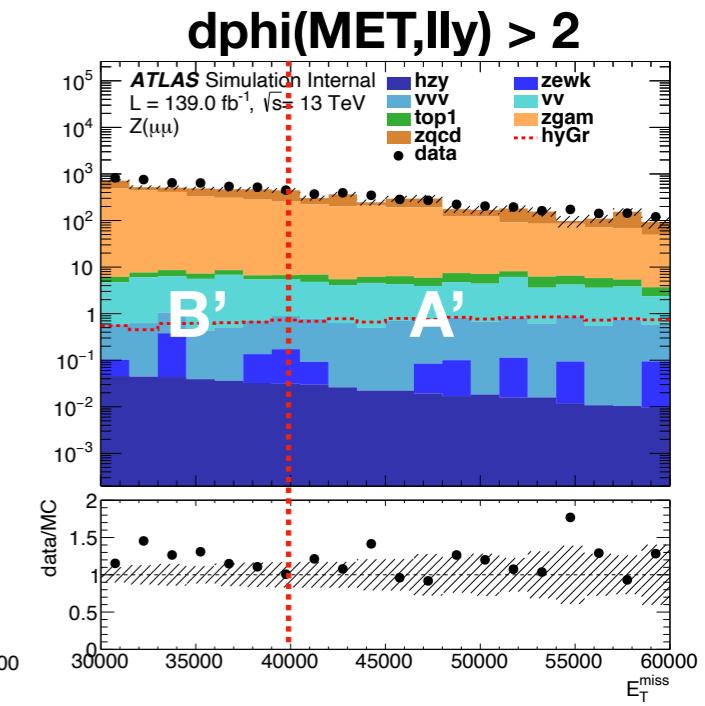
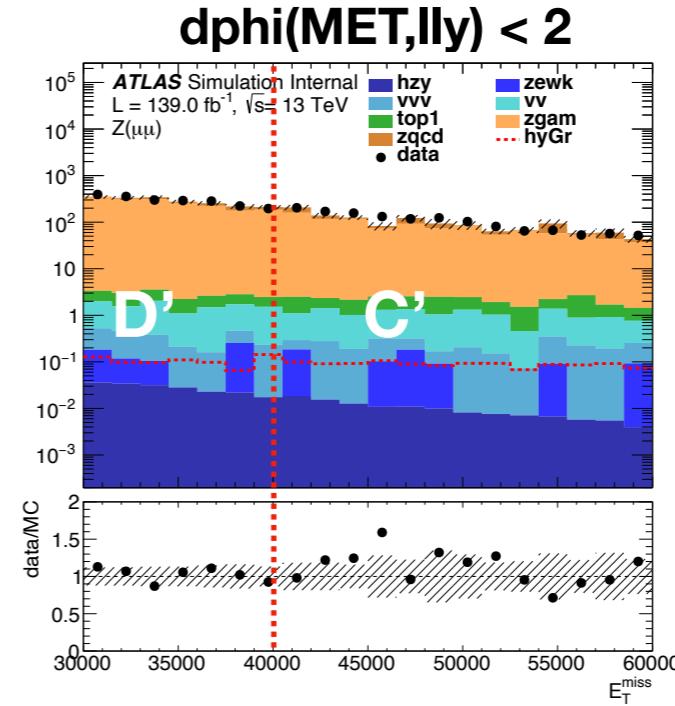
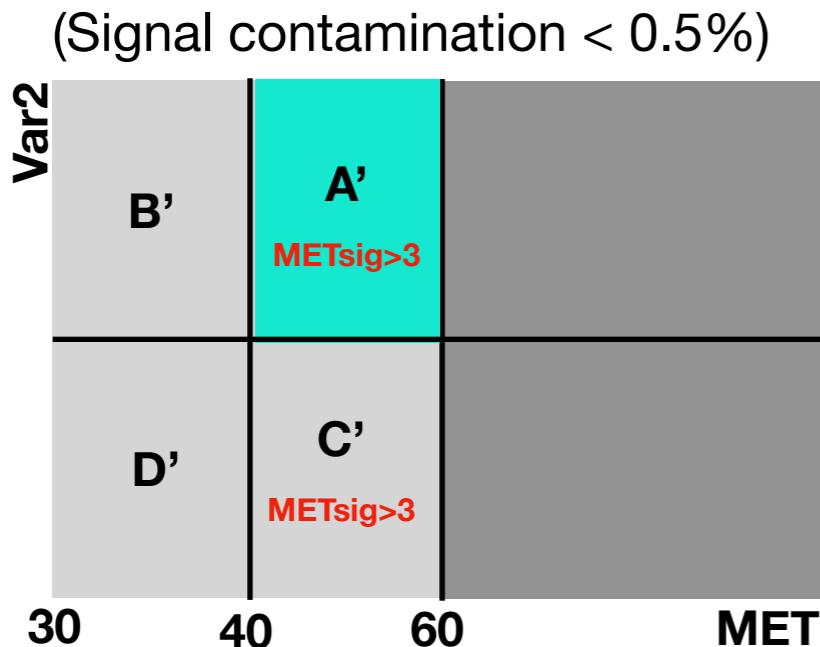
Sample	nA	nB	nC	nD
hyGr	7.6 +/- 0.1	3.5 +/- 0.1	0.7 +/- 0.0	0.5 +/- 0.0
zgam	555.3 +/- 33.0	1954.8 +/- 56.0	226.1 +/- 18.8	1211.7 +/- 45.2
zqcd	564.2 +/- 85.5	739.3 +/- 97.0	46.3 +/- 25.0	199.3 +/- 41.8
totalMC	1191.6 +/- 91.7	2739.1 +/- 112.1	292.6 +/- 31.4	1430.6 +/- 61.5
totalData	1474.0 +/- 38.4	3620.0 +/- 60.2	308.0 +/- 17.5	1513.0 +/- 38.9



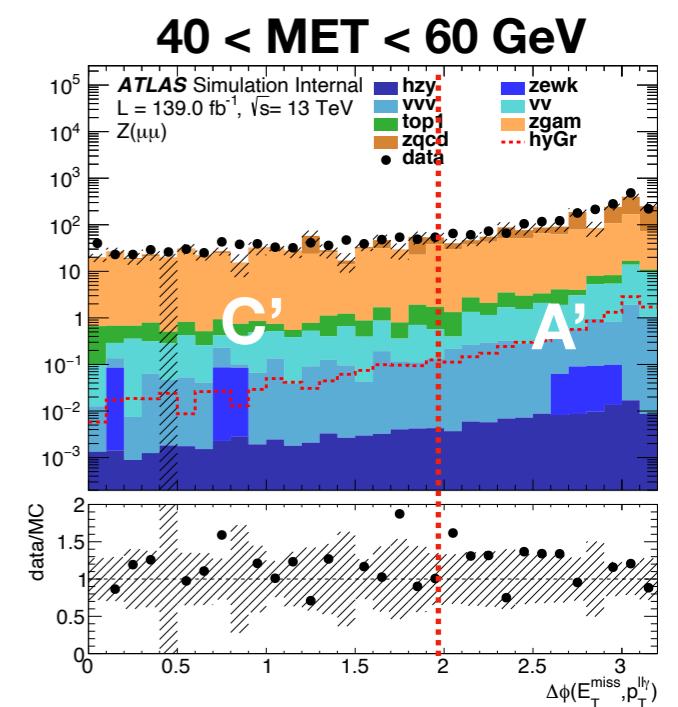
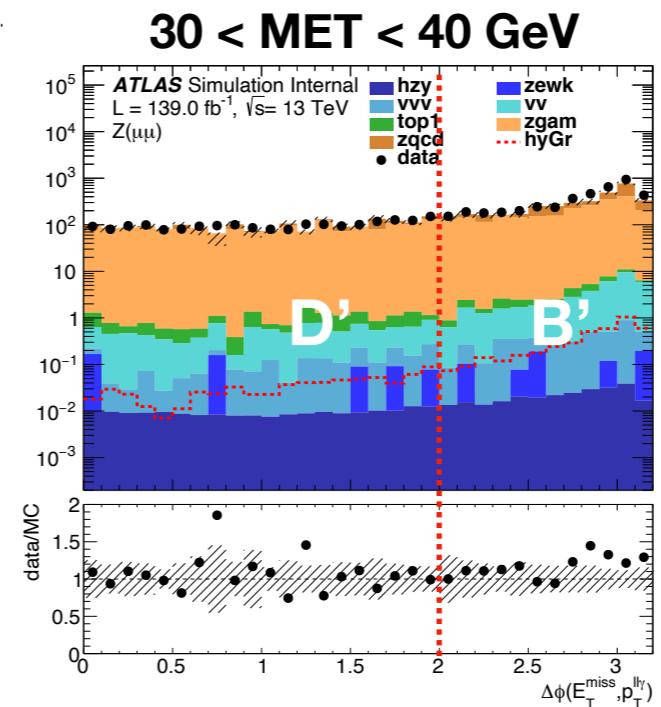
- Good data/MC agreement in low MET regions ( $\text{MET} < 60$ )

# Fake MET background

## Preliminary validation: data/MC in low MET region



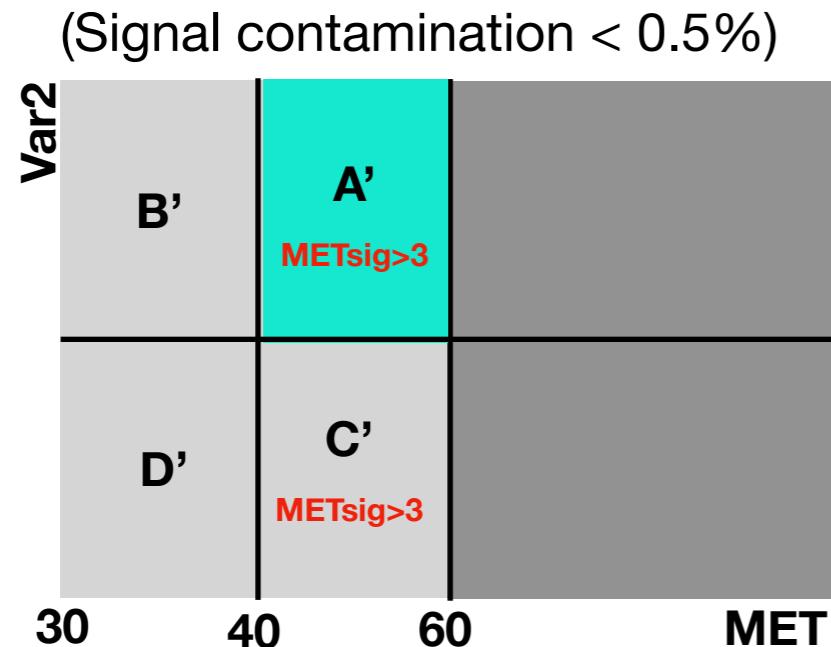
Sample	nA	nB	nC	nD
hyGr	9.1+/-0.2	4.0+/-0.1	0.9+/-0.1	0.7+/-0.0
zgam	783.8+/-34.7	2407.6+/-64.5	484.6+/-29.9	1740.8+/-49.9
zqcd	798.7+/-105.2	1079.1+/-113.2	101.8+/-32.4	156.4+/-50.3
totalMC	1650.5+/-110.8	3536.7+/-130.3	605.3+/-44.1	1916.7+/-70.9
totalData	1986.0+/-44.6	4236.0+/-65.1	749.0+/-27.4	1979.0+/-44.5



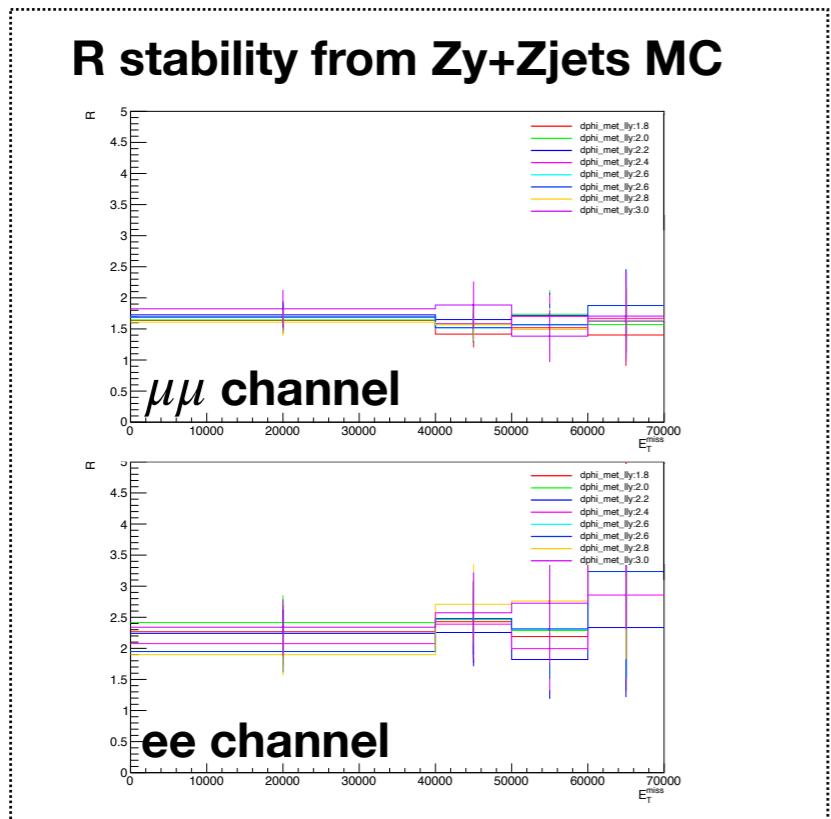
- Good data/MC agreement in low MET regions (MET < 60)

# Fake MET background

## Preliminary validation: results in VR



- All regions dominated by Zy+jets and Z+jets background
- In situ estimates for other backgrounds still not available. Preliminary tests with temporary solutions
  - => R<sub>ZyZjet</sub>: evaluating R from Zy+jets AND Z+jets (both are characterized by sizeable fake MET)
  - => R<sub>Zy</sub> : R from Zy only
- For background not used in R estimation, MC yields in each region subtracted from data



### $\mu\mu$ channel

	R <sub>mc</sub>	R <sub>data</sub>	MC	Data	ABCD
R <sub>ZyZjet</sub>	1.468 +/- 0.170	1.230 +/- 0.113	1582 +/- 111	1918 +/- 66	2291 +/- 329
R <sub>Zy</sub>	1.169 +/- 0.100	1.034 +/- 0.170	784 +/- 35	1120 +/- 124	1266 +/- 188

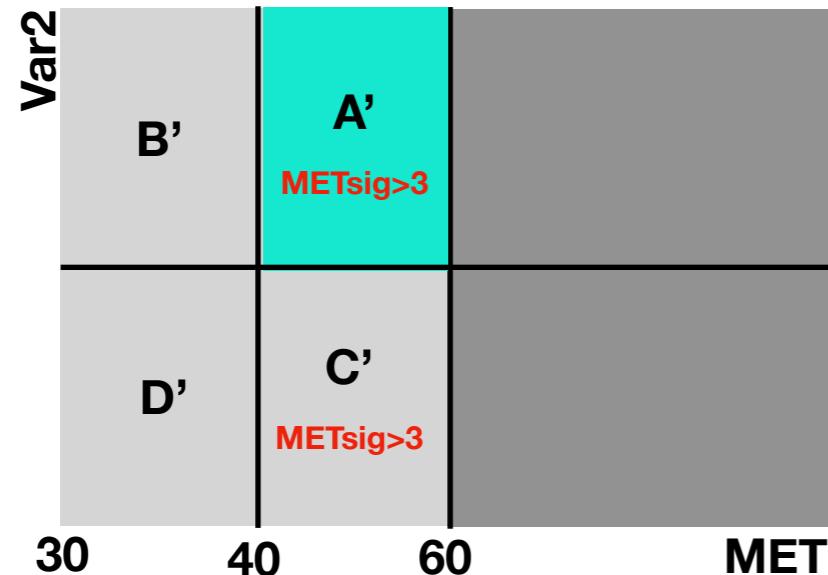
### ee channel

	R <sub>mc</sub>	R <sub>data</sub>	MC	Data	ABCD
R <sub>ZyZjet</sub>	2.153 +/- 0.330	2.035 +/- 0.269	1120 +/- 92	1402 +/- 60	1483 +/- 293
R <sub>Zy</sub>	1.523 +/- 0.172	1.583 +/- 0.356	555 +/- 33	838 +/- 105	806 +/- 176

=> R from data is consistent with R from MC  
 => ABCD estimates consistent with observed data

# Fake MET background

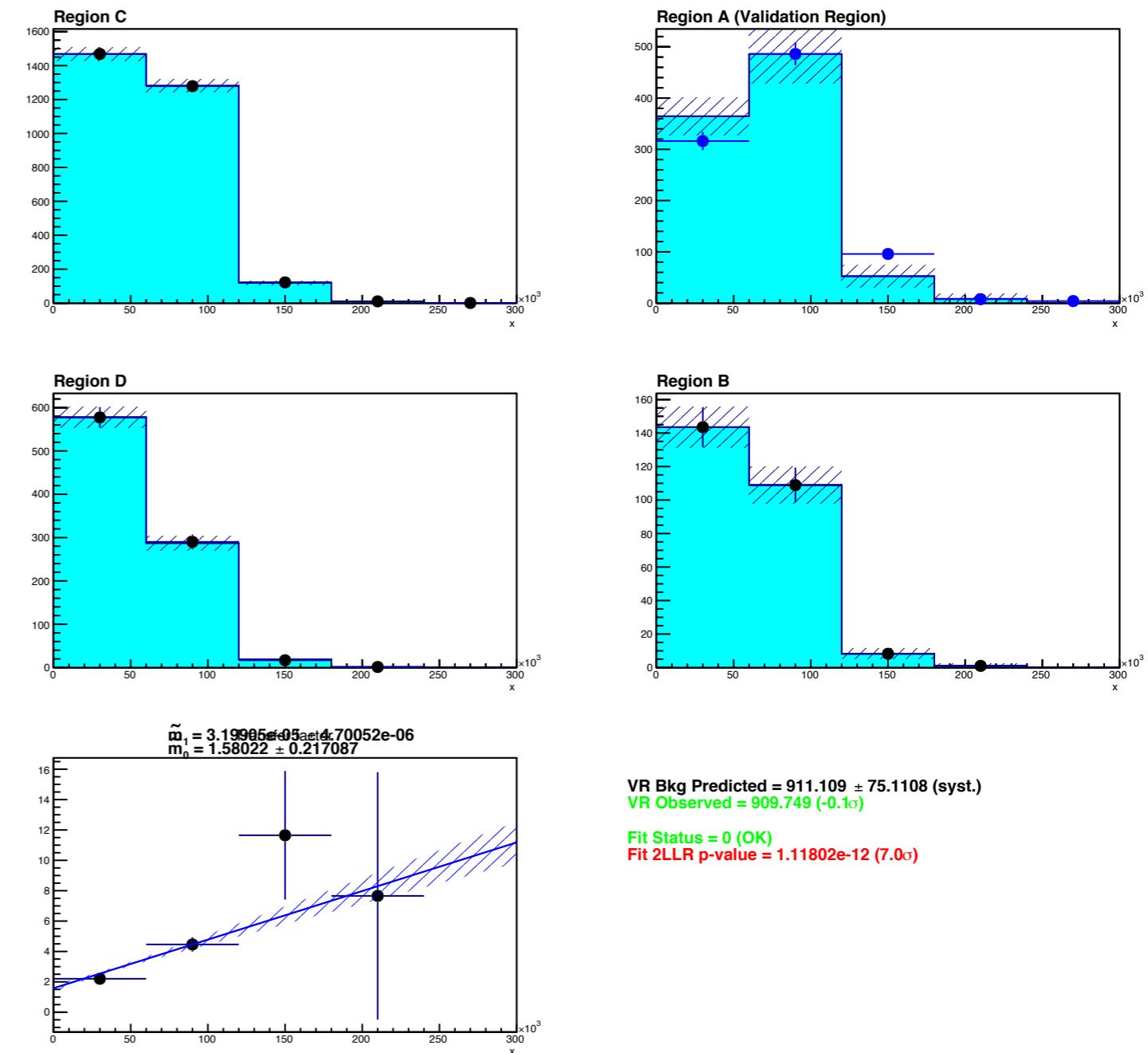
## TRooABCD, ee channel



mT distribution in region A' from simultaneous fit on the B,C,D regions with TRooABCD

- mT histograms in each B,C,D region as input
- Using data in VR, and **subtracting Zjets from MC** (first approximation)
- Correct for  $R = (nA \ nD) / (nC \ nB)$  by scaling the histogram in region D by  $1/R$  (with R estimated from Zy+jet MC)

## Rzy, subtracted Zjet



Only statistical uncertainties

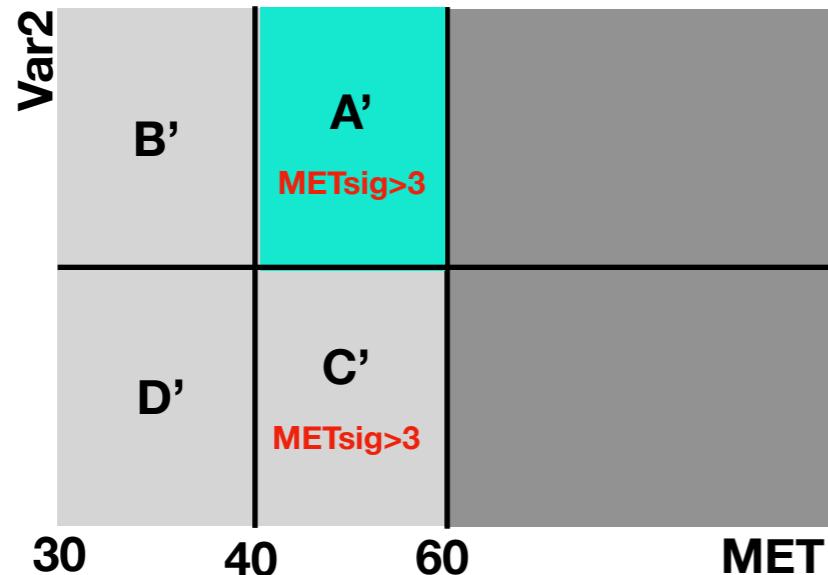
$$nA(x) = (m_0 + m_1(x)) \times nB(x)$$

$$nC(x) = (m_0 + m_1(x)) \times nD(x)$$

( x labels the bin )

# Fake MET background

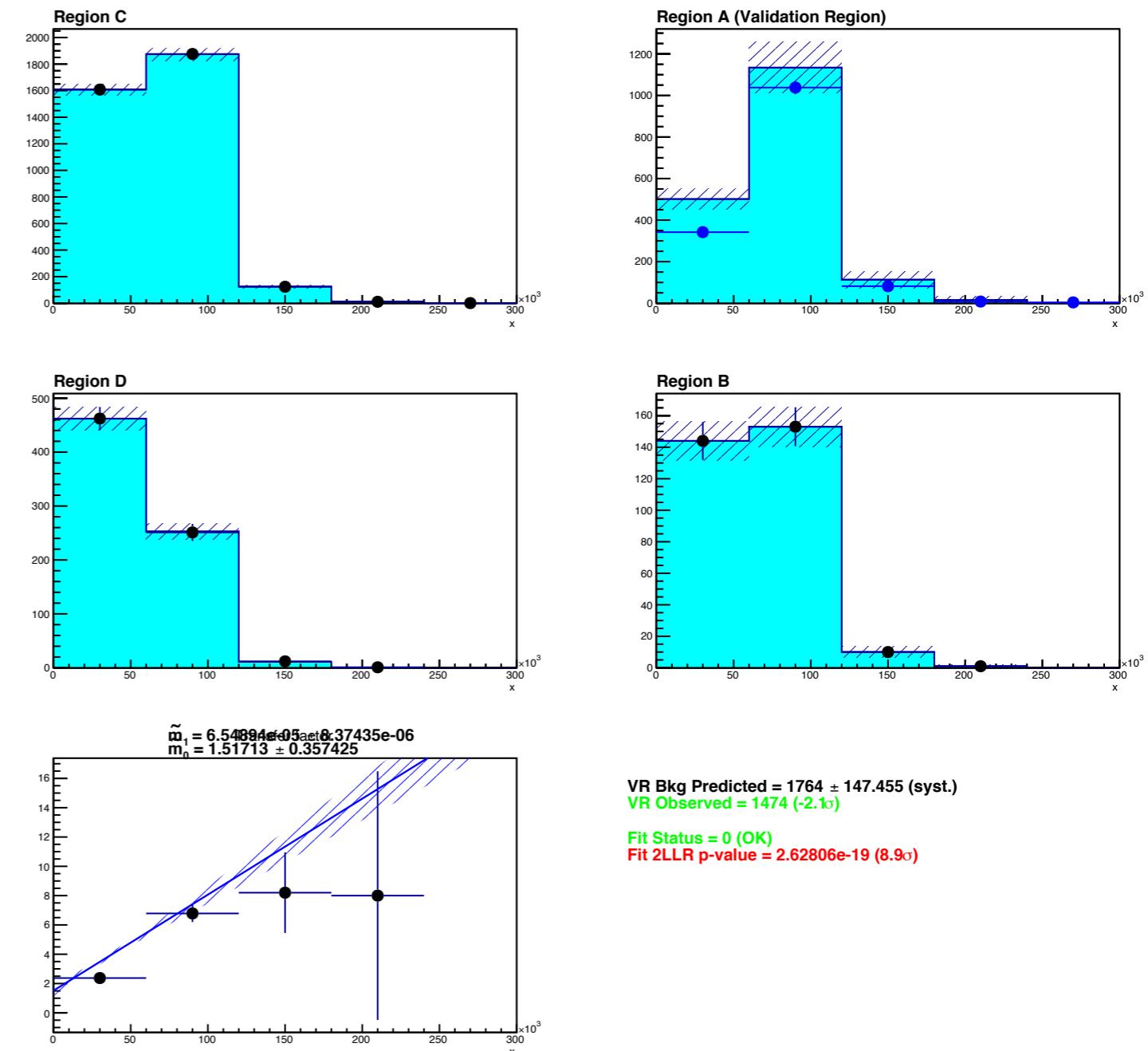
TRooABCD, ee channel



mT distribution in region A' from simultaneous fit on the B,C,D regions with TRooABCD

- mT histograms in each B,C,D region as input
- Using data in VR
- Correct for  $R = (nA \ nD) / (nC \ nB)$  by scaling the histogram in region D by  $1/R$  (with **R estimated from Zy+jet and Zjet MC**)

Rzy+zjet



Only statistical uncertainties

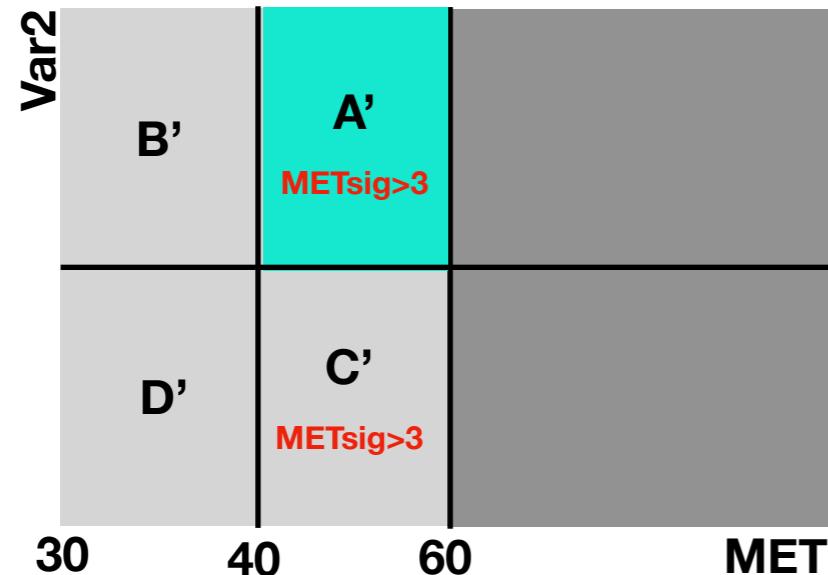
$$nA(x) = (m_0 + m_1(x)) \times nB(x)$$

$$nC(x) = (m_0 + m_1(x)) \times nD(x)$$

( x labels the bin )

# Fake MET background

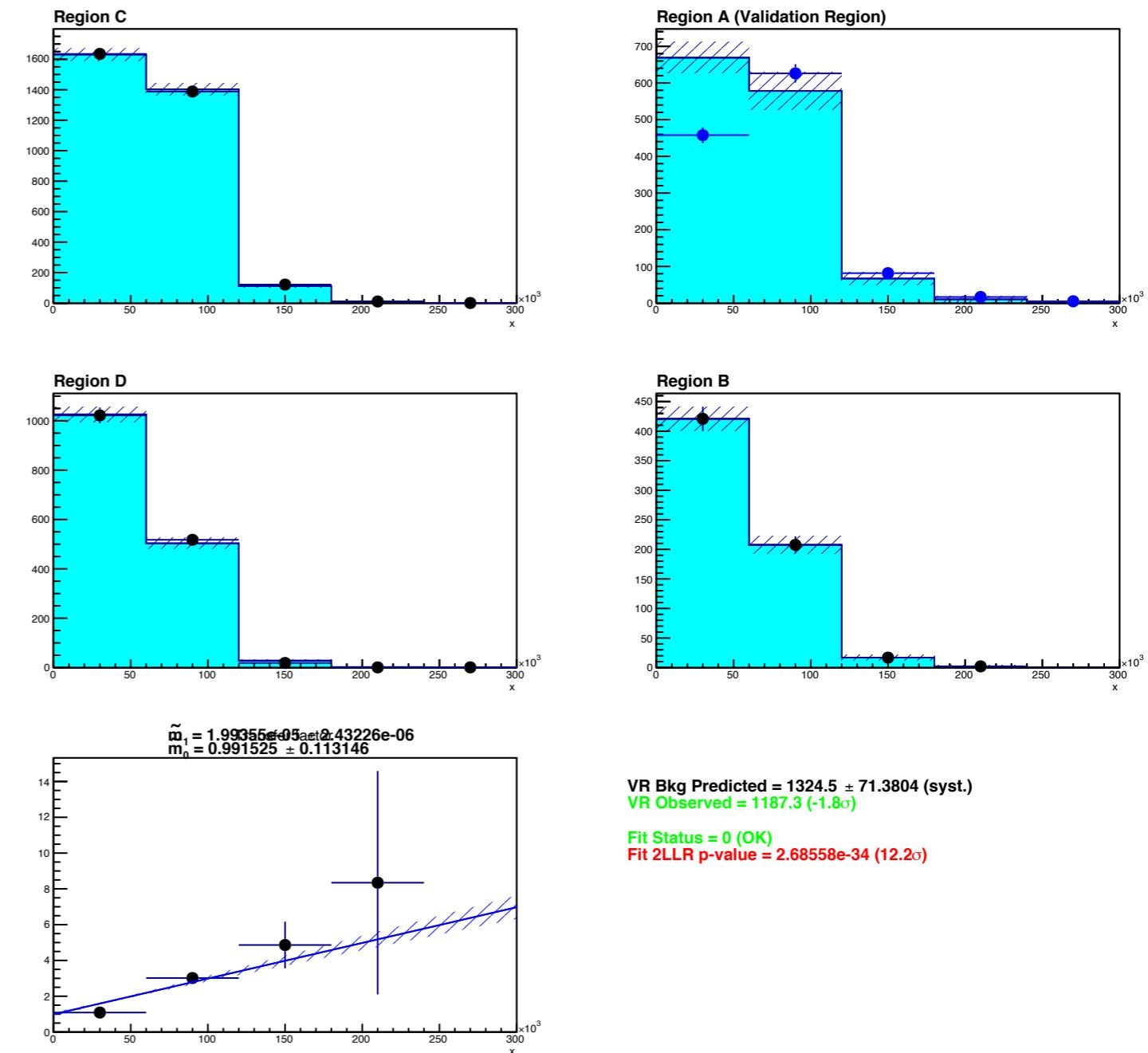
## TRooABCD, uu channel



mT distribution in region A' from simultaneous fit on the B,C,D regions with TRooABCD

- mT histograms in each B,C,D region as input
- Using data in VR, and **subtracting Zjets from MC** (first approximation)
- Correct for  $R = (nA \ nD) / (nC \ nB)$  by scaling the histogram in region D by  $1/R$  (with R estimated from Zy+jet MC)

## Rzy, subtracted Zjet



Only statistical uncertainties

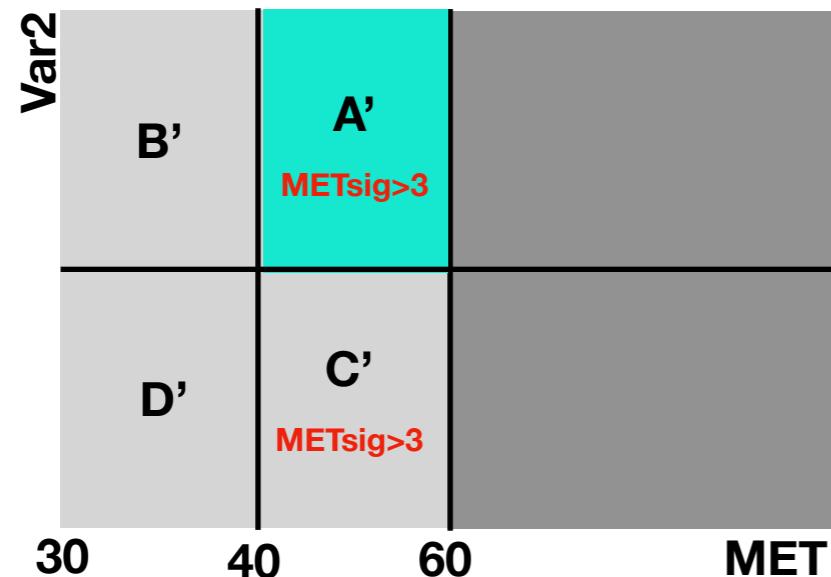
$$nA(x) = (m_0 + m_1(x)) \times nB(x)$$

$$nC(x) = (m_0 + m_1(x)) \times nD(x)$$

( x labels the bin )

# Fake MET background

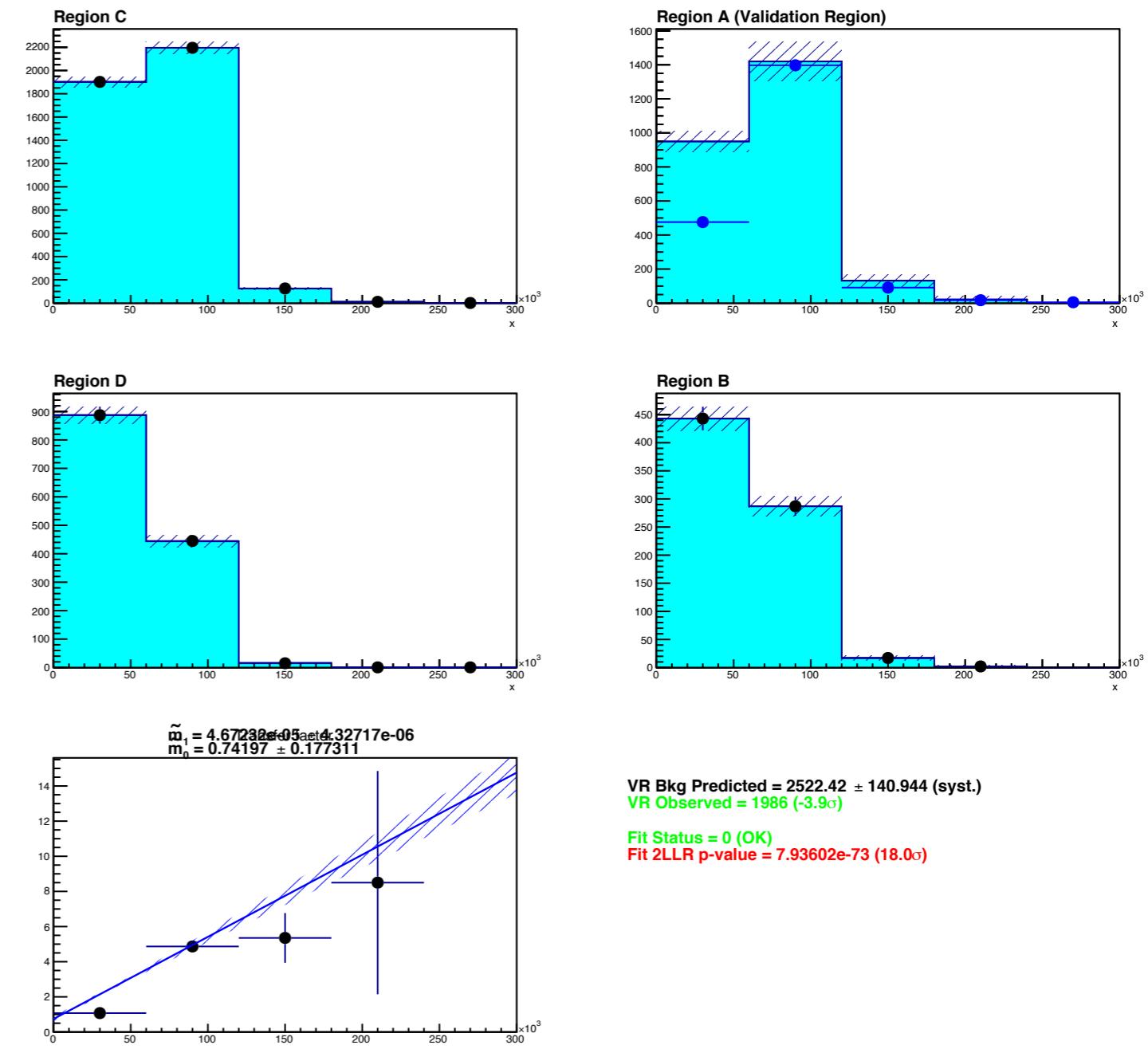
## TRooABCD, uu channel



mT distribution in region A' from simultaneous fit on the B,C,D regions with TRooABCD

- mT histograms in each B,C,D region as input
- Using data in VR
- Correct for  $R = (n_A n_D) / (n_C n_B)$  by scaling the histogram in region D by  $1/R$  (with **R estimated from Zy+jet and Zjet MC**)

**Rzy+zjet**



Only statistical uncertainties

$$n_A(x) = (m_0 + m_1(x)) \times n_B(x)$$

$$n_C(x) = (m_0 + m_1(x)) \times n_D(x)$$

(x labels the bin)

# Summary

- ZH, H->yyD analysis ongoing for the first time in ATLAS
- Challenging analysis, due to background estimation strongly relying on data-driven techniques
  - Dominant background from fake MET and jets-faking-photons, while low irreducible background
  - Defined background estimation strategies (taking advantage of experience gained in monophoton analysis for in-situ background):
    - Electron-faking-photon: scaling data in WZ CR by electron fake-rate (from mono-photon analysis)
    - Jet-faking-photon: ABCD based on photon iso and ID
    - **Fake MET: ABCD based on MET and dphi(MET,yll)**
      - New method has been defined => choice of variables, regions definition, VR
      - Preliminary validation show promising results
      - Room for further optimization
  - Still need to study ttbar, VV, VVV, Higgs related backgrounds
  - Fitting strategy still to be defined (mT or BDT as discriminating variable)
  - Internal note in preparation

# Backup slides

# Objects selection

## Electrons

$p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.47$

$|\Delta z_0 \sin\theta| < 0.5 \text{ mm}$

- Baseline  
LooseAndBLayerLLH ID

- Selected

$|d_0(\sigma)| < 5$

MediumLLH ID and FCLoose isolation

## Photons

$p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.37$

- Baseline  
Loose ID
- Selected

$p_T > 25 \text{ GeV}$

Tight ID and FixedCutTight isolation

## Muons

$p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.7$

$|\Delta z_0 \sin\theta| < 0.5 \text{ mm}$

- Baseline  
Loose ID

- Selected

$|d_0(\sigma)| < 3$

Medium ID and FLoose\_VarRad isolation

## Jets

PFlow jets

$p_T > 25 \text{ GeV}$ ,  $|\eta| < 4.5$

Medium JVT WP

$E_T^{\text{miss}}$ : Tight WP



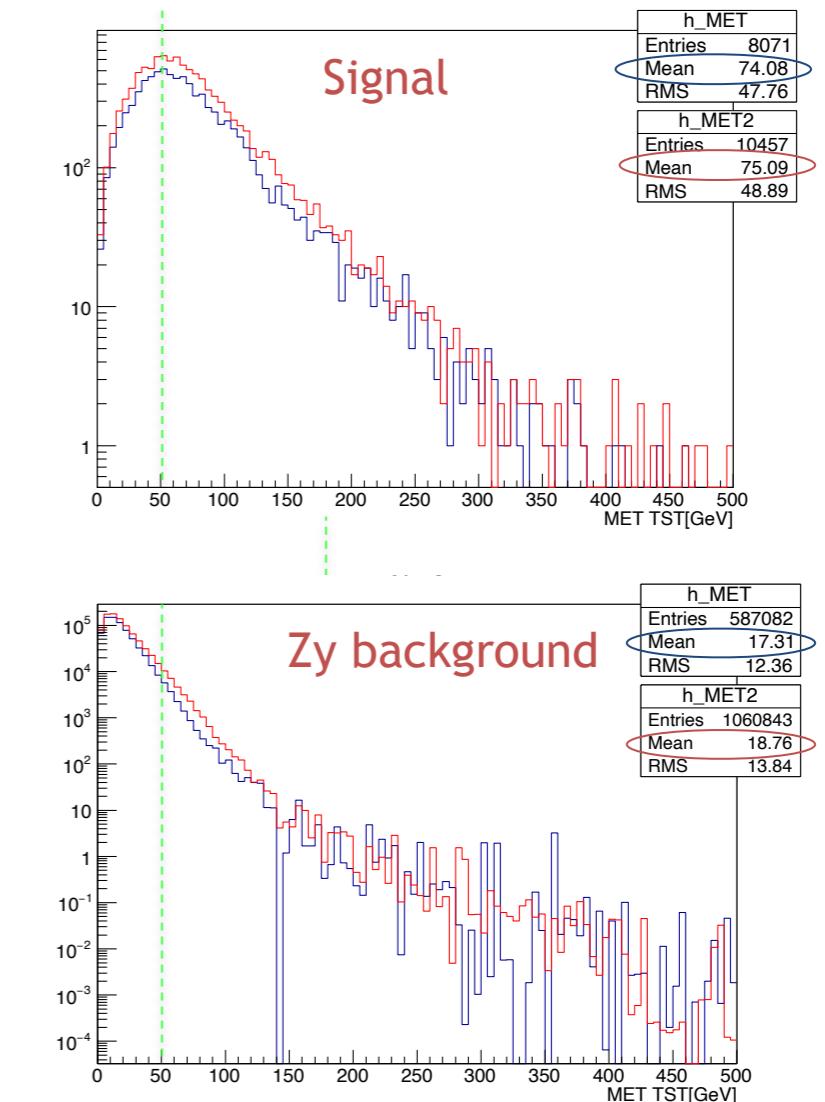




# ee VS $\mu\mu$ channel discrepancy

S. Resconi

	HyGr	Zgam	Zqcd		HyGr	Zgam	Zqcd
Input	16.95 +/- 0.22	4070 +/- 77	3570 +/- 210		16.95 +/- 0.22	4070 +/- 77	3570 +/- 210
CutMergeExt	16.95 +/- 0.22	3680 +/- 75	3570 +/- 210		16.95 +/- 0.22	3680 +/- 75	3570 +/- 210
CutMCOOverlap	16.95 +/- 0.22	3597 +/- 74	980 +/- 110		16.95 +/- 0.22	3597 +/- 74	980 +/- 110
CutTrig	16.69 +/- 0.22	3401 +/- 72	970 +/- 110		16.69 +/- 0.22	3401 +/- 72	970 +/- 110
CutJetClean	16.69 +/- 0.22	3401 +/- 72	970 +/- 110		16.69 +/- 0.22	3401 +/- 72	970 +/- 110
CutChannel	7.37 +/- 0.14	1162 +/- 41	373 +/- 68		9.24 +/- 0.17	2204 +/- 58	596 +/- 82
CutNjet	7.37 +/- 0.14	1162 +/- 41	373 +/- 68		9.24 +/- 0.17	2204 +/- 58	596 +/- 82
CutMet	7.37 +/- 0.14	1162 +/- 41	373 +/- 68		9.24 +/- 0.17	2204 +/- 58	596 +/- 82
CutNumPho	7.20 +/- 0.14	1110 +/- 40	375 +/- 67		8.98 +/- 0.16	2137 +/- 58	592 +/- 82
CutNumEle	7.20 +/- 0.14	1110 +/- 40	375 +/- 67		8.98 +/- 0.16	2137 +/- 58	592 +/- 82
CutVetoExtraLep	7.19 +/- 0.14	1103 +/- 40	374 +/- 67		8.98 +/- 0.16	2134 +/- 58	594 +/- 82
CutL0Pt	7.15 +/- 0.14	1059 +/- 39	375 +/- 67		8.98 +/- 0.16	2127 +/- 57	594 +/- 82
CutL1Pt	7.15 +/- 0.14	1059 +/- 39	375 +/- 67		8.98 +/- 0.16	2127 +/- 57	594 +/- 82
CutVetoBjets	7.15 +/- 0.14	1059 +/- 39	375 +/- 67		8.98 +/- 0.16	2127 +/- 57	594 +/- 82
CutHighMet	5.72 +/- 0.13	434 +/- 25	111 +/- 40		7.18 +/- 0.15	929 +/- 37	261 +/- 55
CutMass	5.44 +/- 0.12	157 +/- 15	110 +/- 40		6.83 +/- 0.14	331 +/- 26	224 +/- 53
CutPhPt	5.20 +/- 0.12	90 +/- 13	56 +/- 19		6.49 +/- 0.14	181 +/- 21	153 +/- 35
CutMllg	5.19 +/- 0.12	89 +/- 13	56 +/- 19		6.49 +/- 0.14	176 +/- 21	153 +/- 35

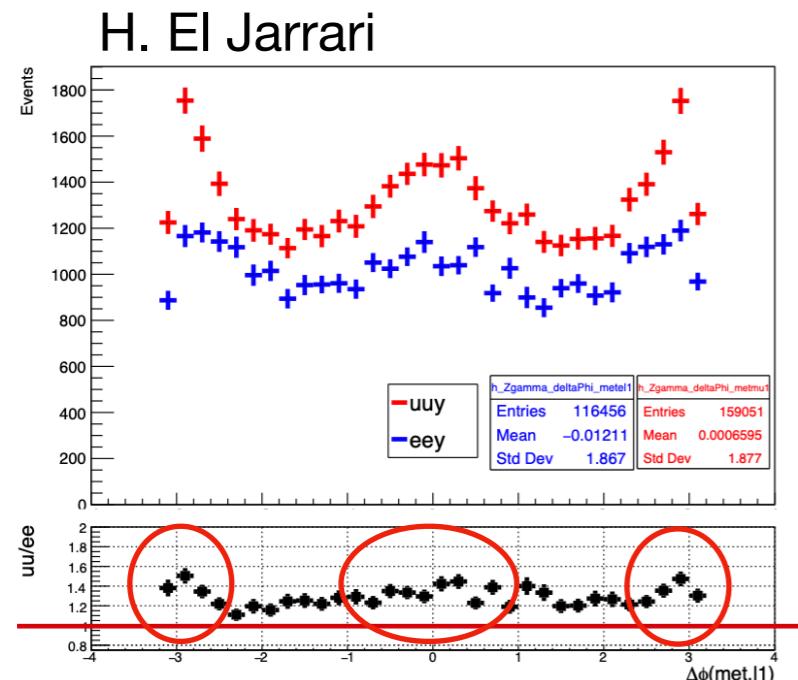
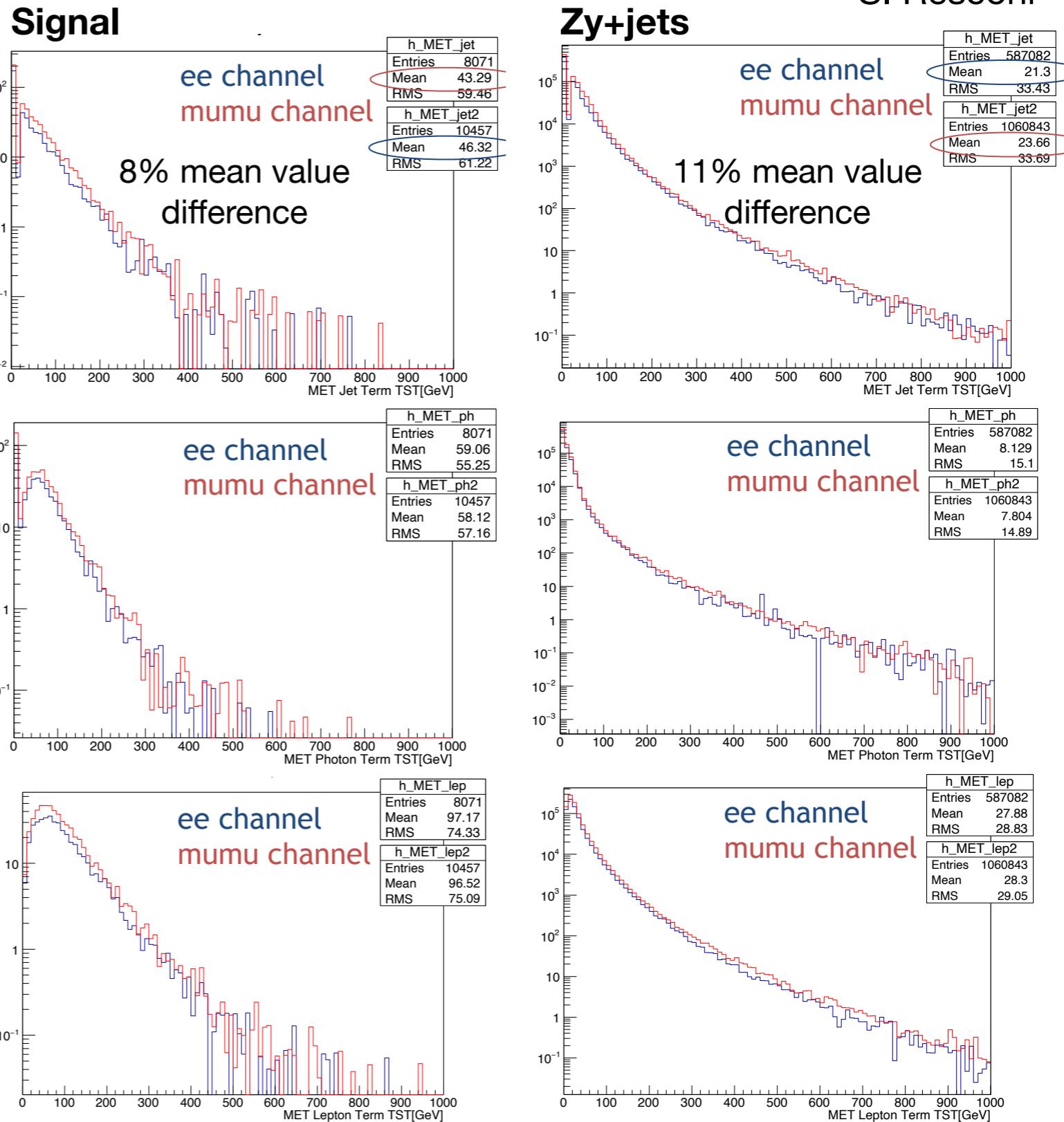


- Higher differences in Zgam and Zqcd wrt signal
- Arising from MET distribution (MET>50 GeV cut included in “Input” yields)
- Discrepancy investigated removing MET > 50 preselections
  - Requiring 2 opposite charge leptons:
    - Signal uu/ee=1.32
    - Zy uu/ee=1.25
  - 2 opposite charge leptons && MET > 50
    - Signal uu/ee = 1.32
    - Zy. uu/ee = 2.04
- MET experts have been contacted
  - Agree some differences are expected due to different e/jet and mu/jet OR treatment in MET
  - muon/pfjet bugfix in OR is included in SusyTools but suggestion to check the impact of the flag UseMuonPFlowBugfix in METMaker (default = False) => Need to implement this fix in SUSYTools

# ee VS $\mu\mu$ channel discrepancy

S. Resconi

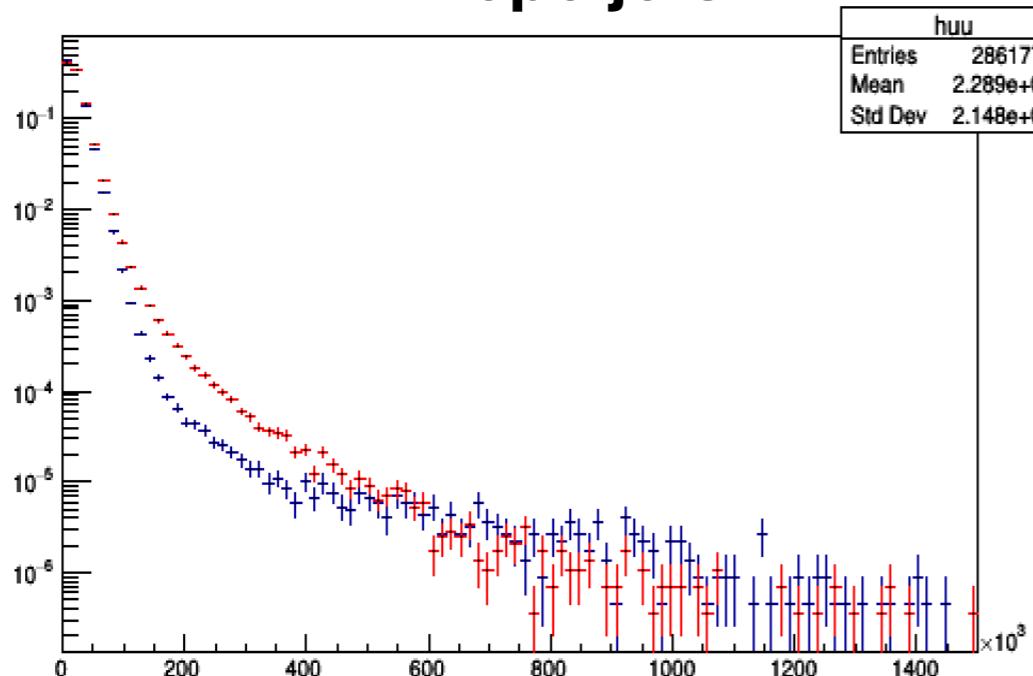
- Main difference in jet term (8% mean value difference in signal, 11% in Zy). Can be related to different treatment of muon-jet and electron-jet overlaps
- Higher impact on global MET in bkg wrt signal, due to higher importance of Jet Term relative to the other MET terms
  - Background: Jet Term value similar to Lepton Term and higher than Photon Term
  - Signal: Jet Term subdominant



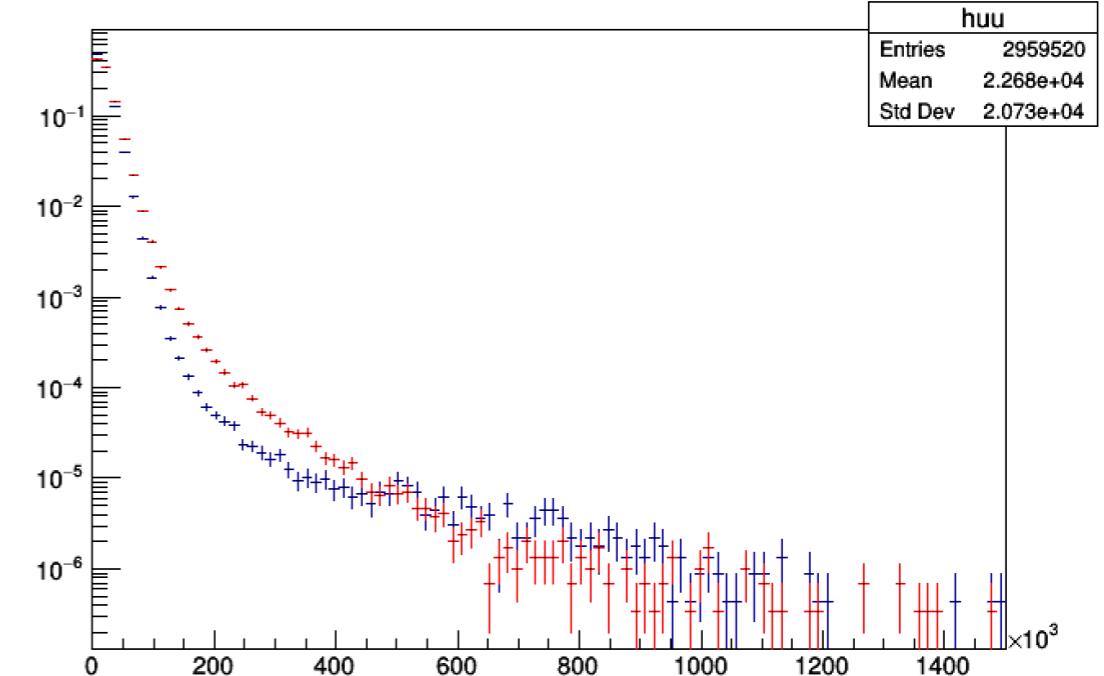
# ee VS $\mu\mu$ channel discrepancy

Check behaviour with EMTopo jets (no muon/pfjet bug) to understand if the bug can be the reason for the differences)

**EMTopo jets**



**PFlow jets**

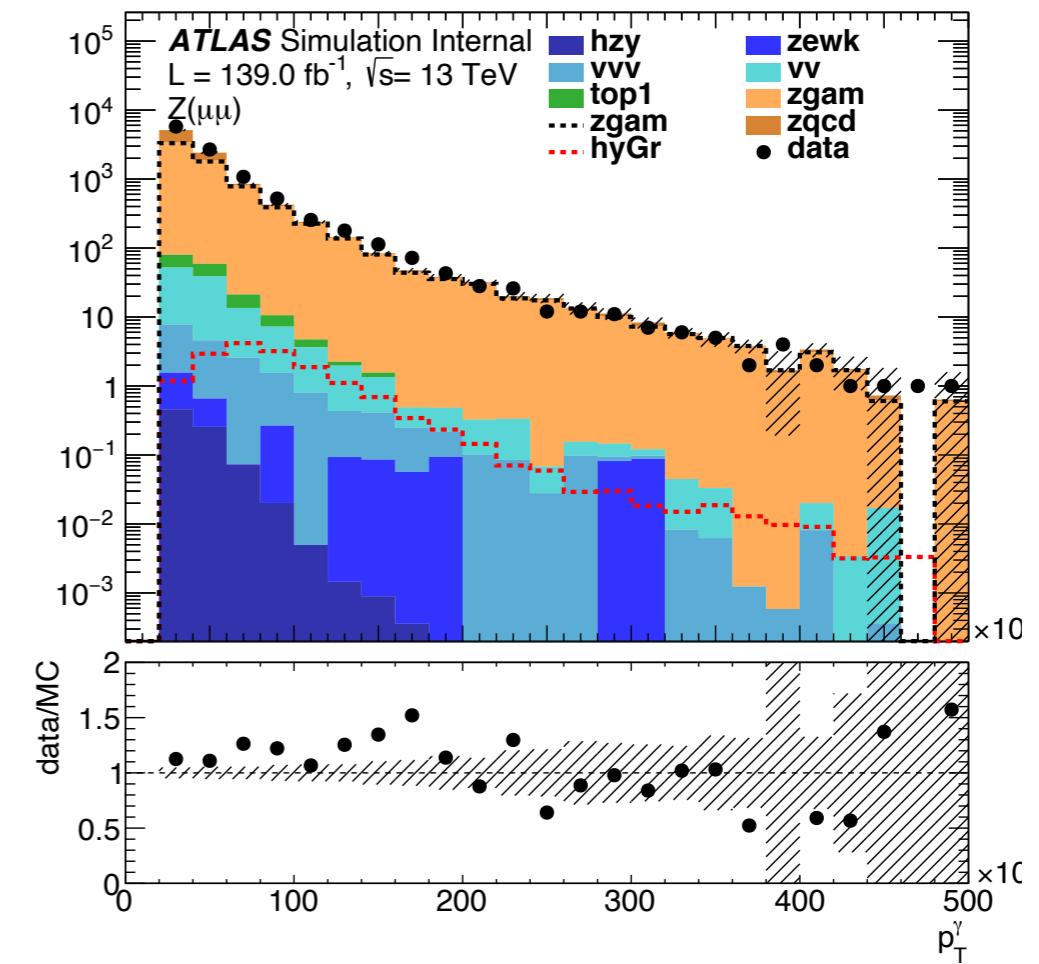
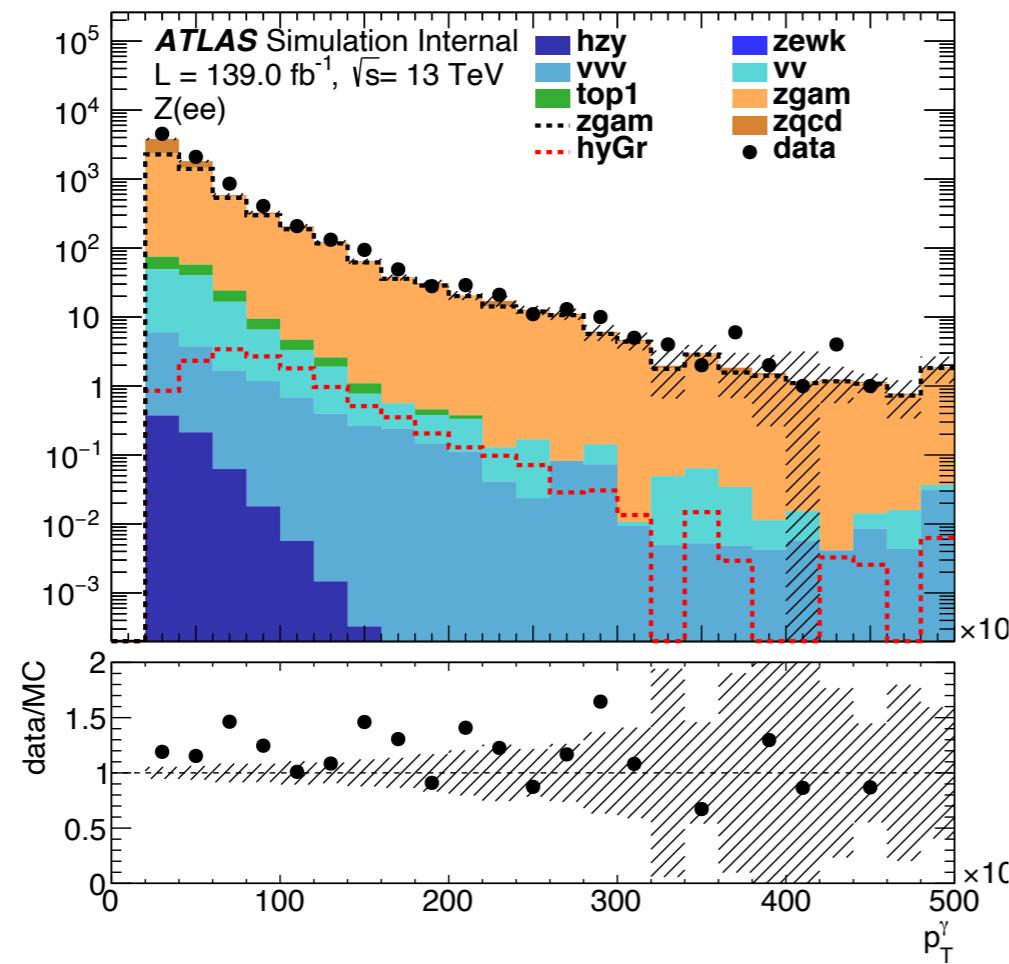


MET > 50: uu/ee = 1.37

MET > 50: uu/ee = 1.67

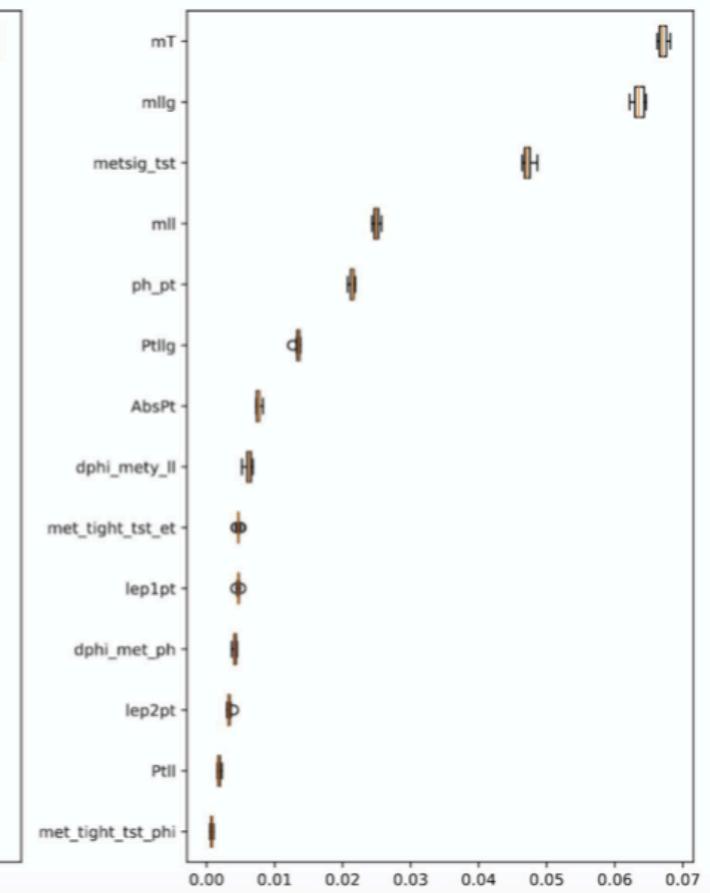
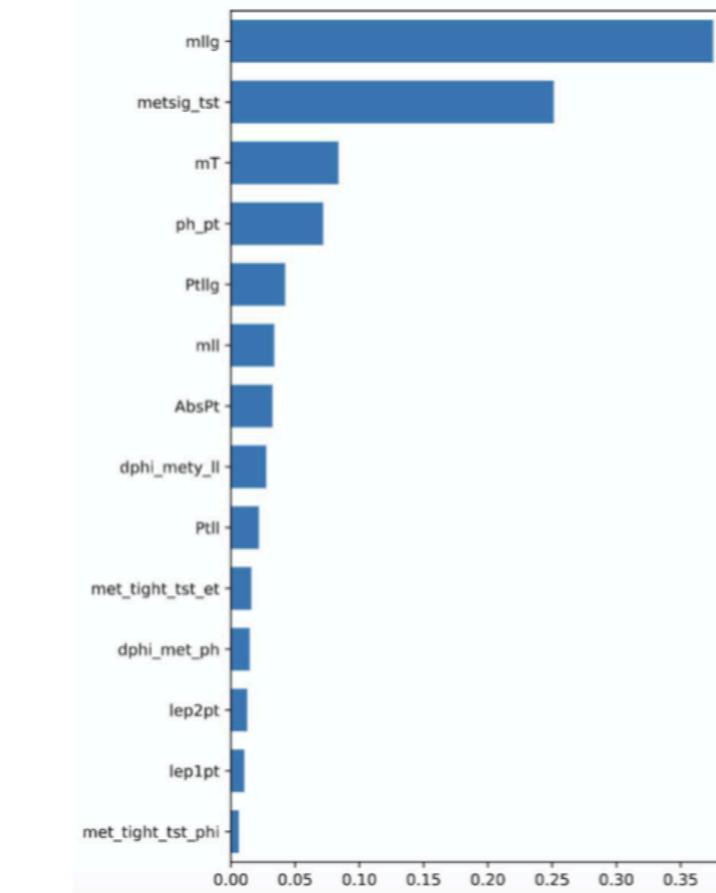
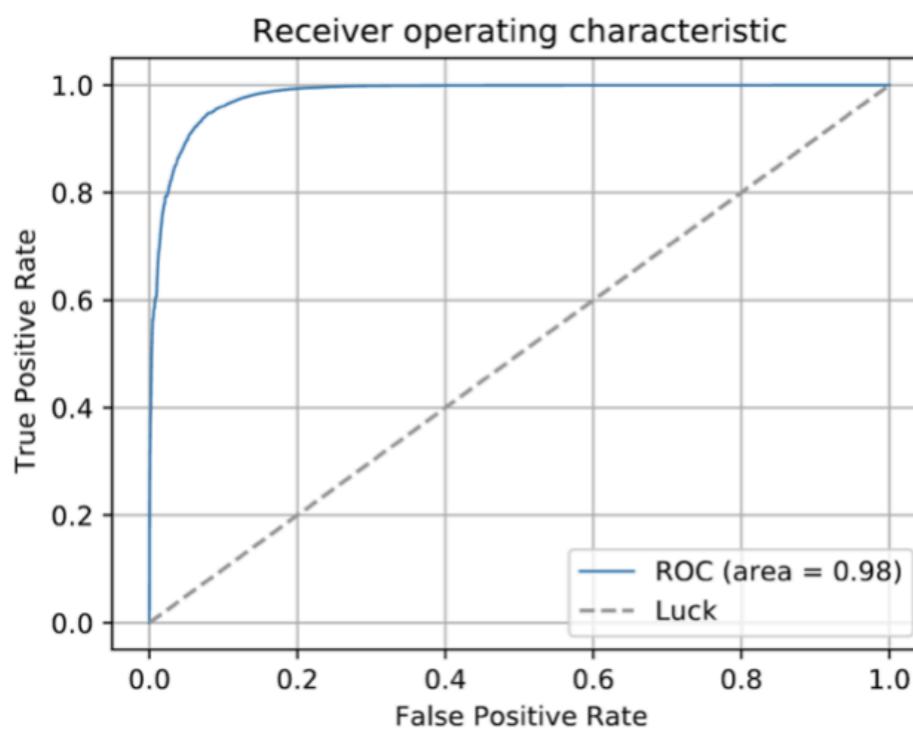
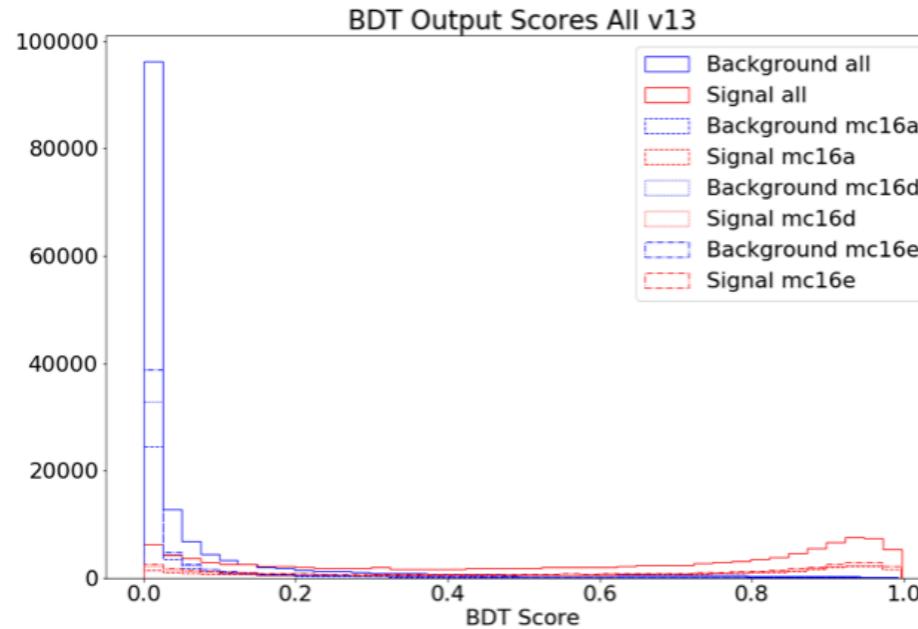
Higher difference with PFlow than EMTopo. Might be due to muon/pfjet bug.  
Worth testing the bug fix in MET

# Photon pT in low MET region: data/MC



# Machine Learning studies

## BDT results

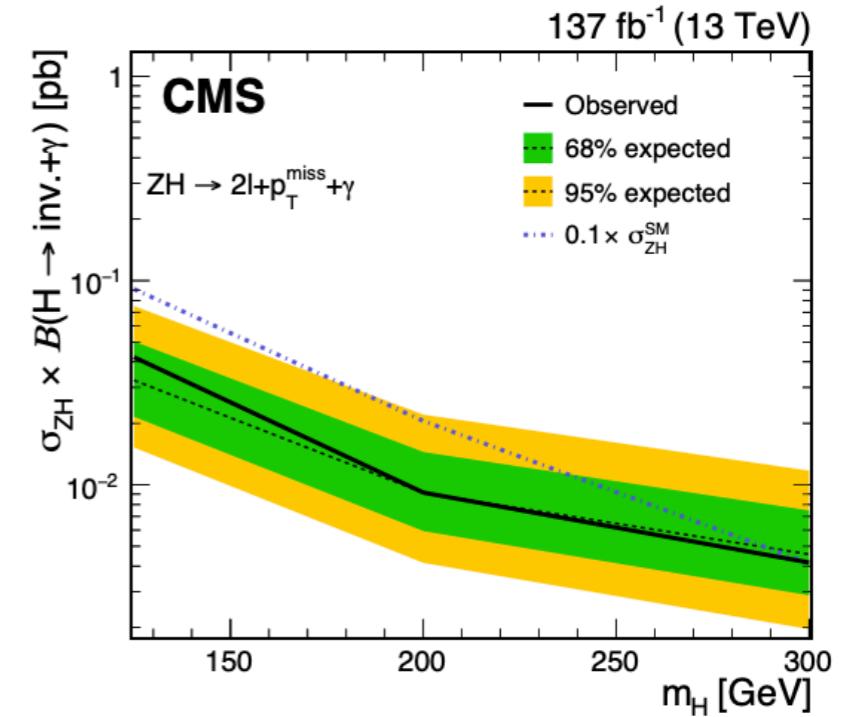
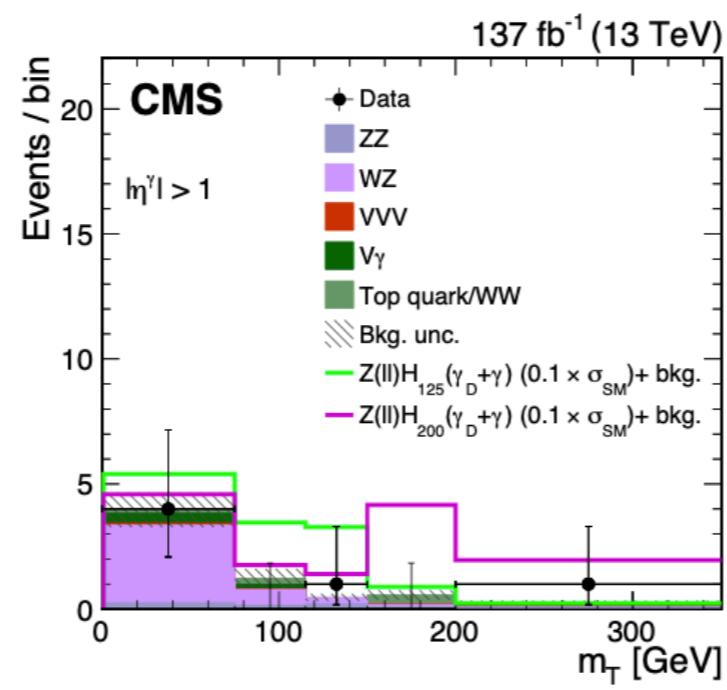
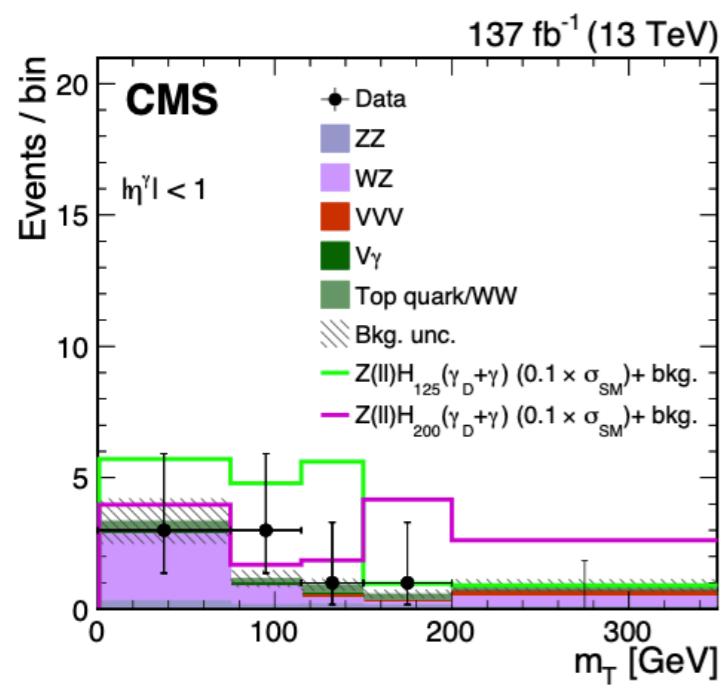


**Permutation Feature Importance:** Measure the importance of a feature by calculating the increase in the model's prediction error after permuting the feature

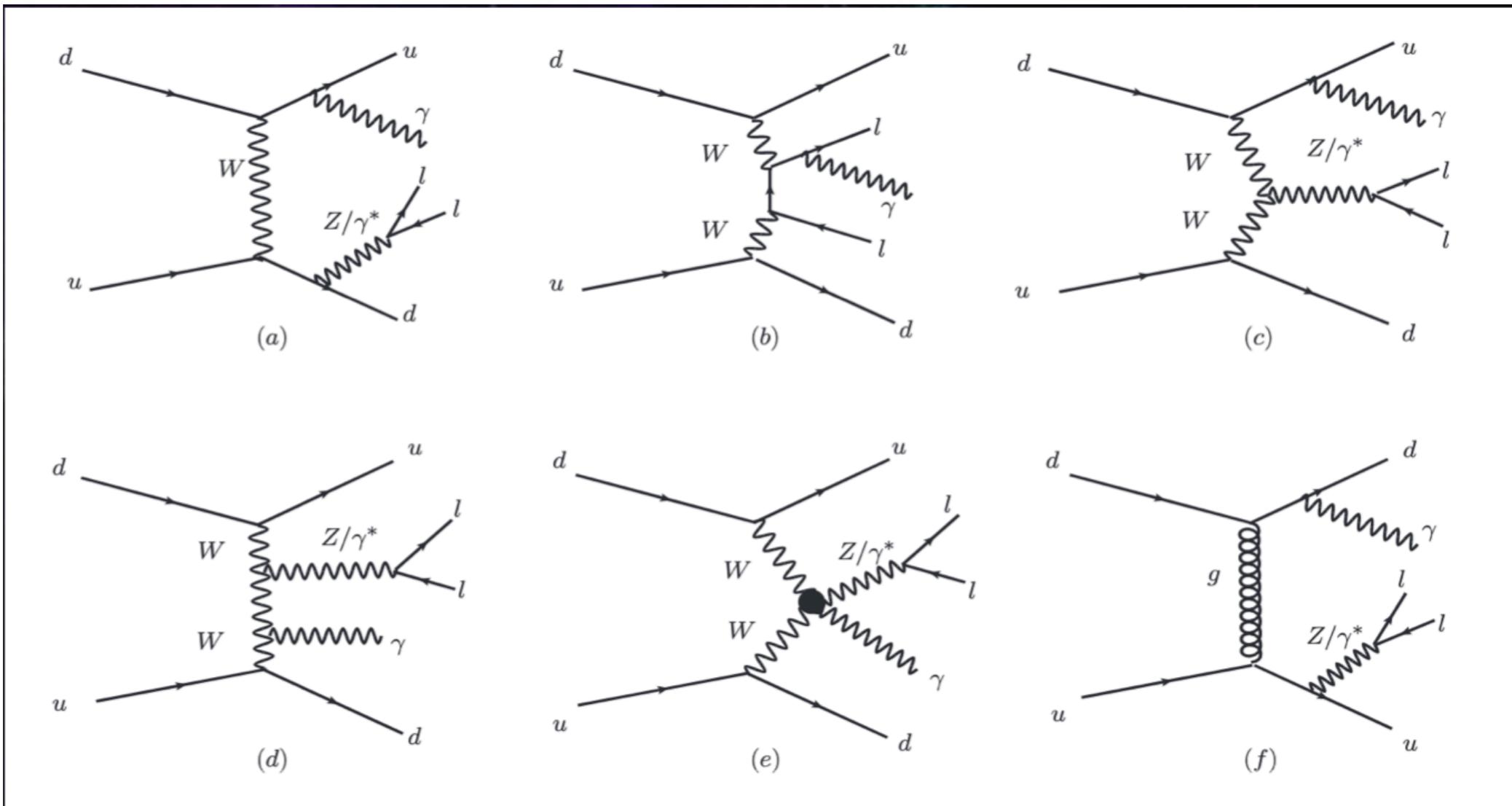


# CMS results

Process	Yield
Data	14
Nonresonant	$2.4 \pm 1.1$
WZ	$8.1 \pm 2.0$
ZZ	$1.5 \pm 0.3$
Z $\gamma$	$0.7 \pm 0.7$
Other	$0.6 \pm 0.3$
Total background	$13.3 \pm 3.8$
ZH <sub>125</sub> (product of acceptance and efficiency)	$17.9 \pm 1.2$ ( $2.13 \pm 0.14\%$ )
ZH <sub>200</sub> (product of acceptance and efficiency)	$12.3 \pm 0.8$ ( $6.48 \pm 0.42\%$ )
ZH <sub>300</sub> (product of acceptance and efficiency)	$3.9 \pm 0.2$ ( $10.20 \pm 0.51\%$ )

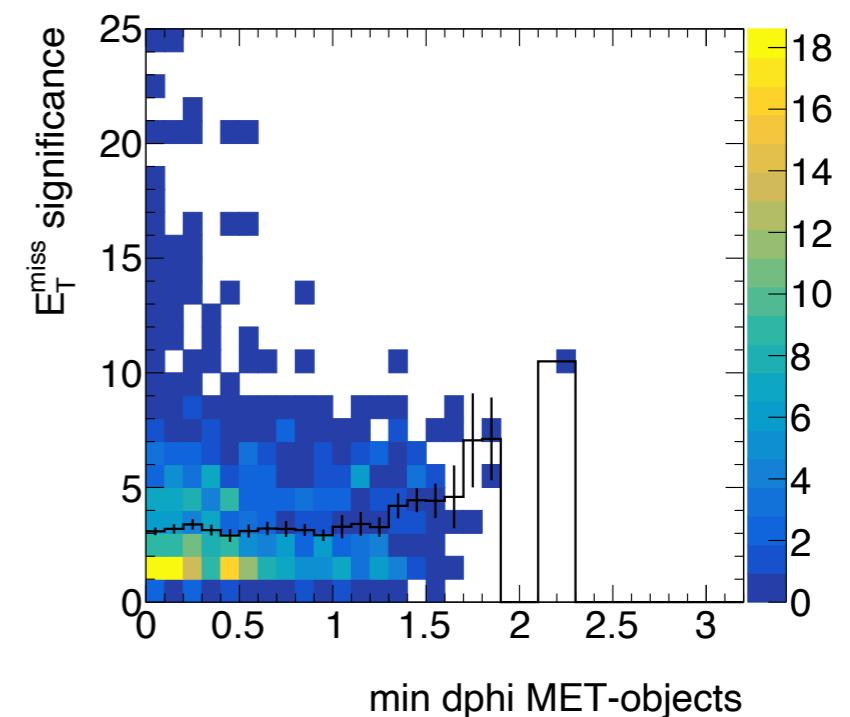
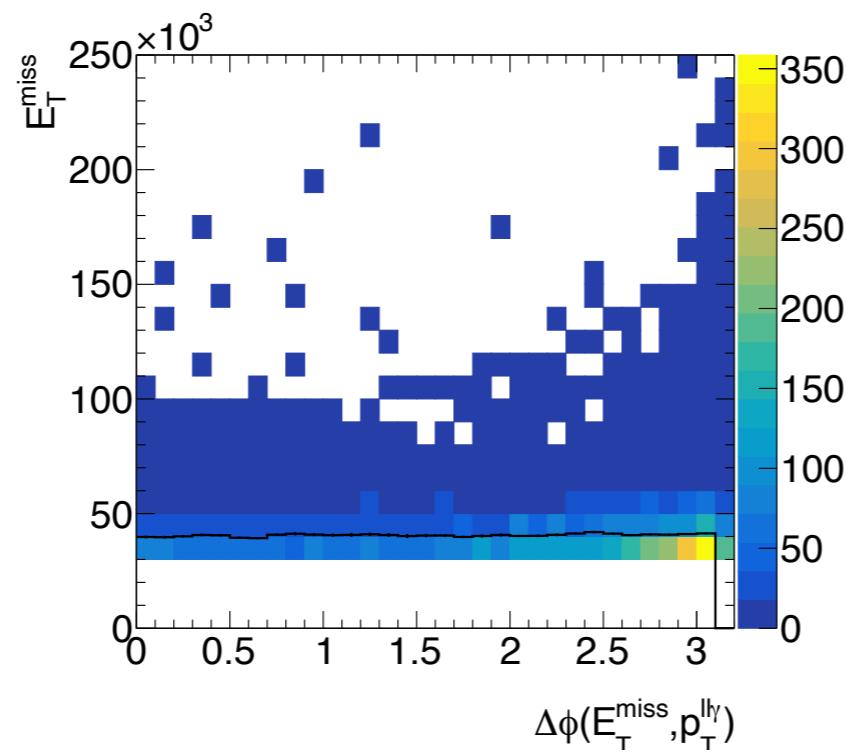


# Zy+jets background

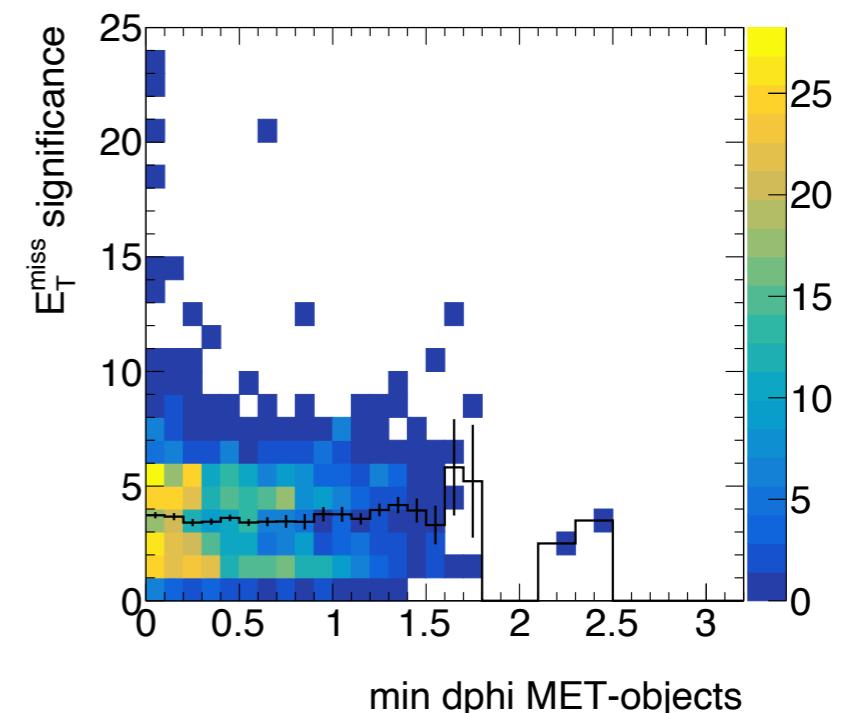
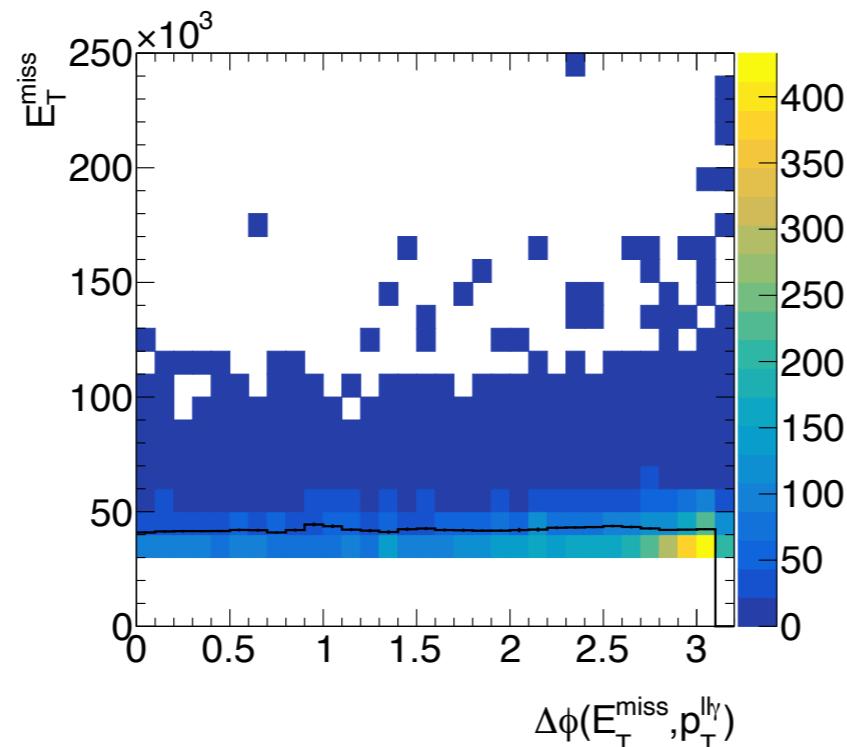


# Variable correlation

**ee-channel**

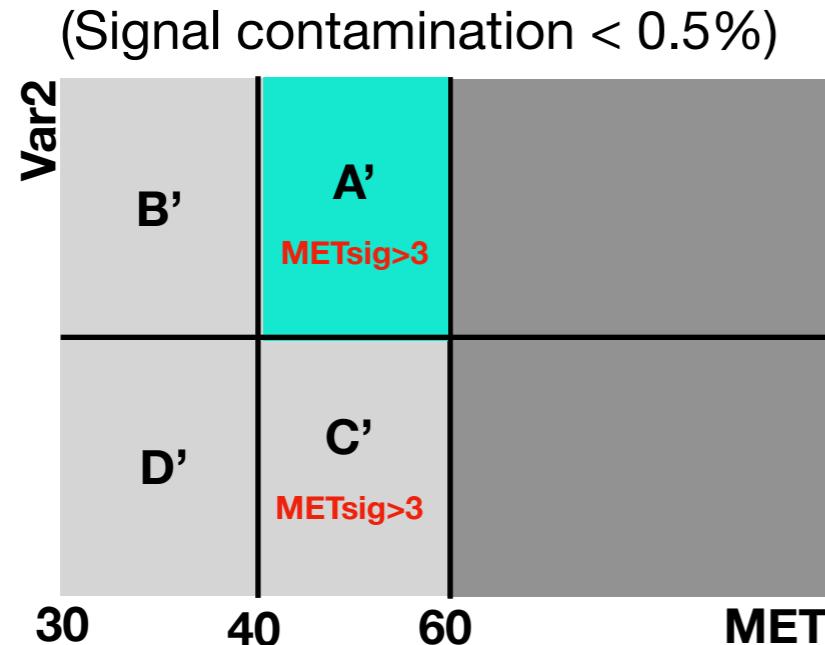


**uu-channel**



# Fake MET background

## MET-based ABCD with VS without METsig cut in VR



- All regions dominated by Zy+jets and Z+jets background
- In situ estimates for other backgrounds still not available. Preliminary tests with temporary solutions  
=> R<sub>zyzjet</sub>: evaluating R from Zy+jets AND Z+jets (both are characterized by sizeable fake MET)
- => R<sub>zy</sub> : R from Zy only  
For background not used in R estimation, MC yields in each region subtracted from data

$\mu\mu$ channel	Data in VR	Rmc	Rdata	MC	Data	ABCD
R <sub>zyzjet</sub> + metsig	1.468 $\pm$ 0.170	1.230 $\pm$ 0.113	1582 $\pm$ 111	1918 $\pm$ 66	2291 $\pm$ 329	
R <sub>zyzjet</sub>	1.118 $\pm$ 0.098	1.019 $\pm$ 0.079	2646 $\pm$ 133	3088 $\pm$ 109	3388 $\pm$ 377	
R <sub>zy</sub> + metsig	1.169 $\pm$ 0.100	1.034 $\pm$ 0.170	784 $\pm$ 35	1120 $\pm$ 124	1266 $\pm$ 188	
R <sub>zy</sub>	1.154 $\pm$ 0.077	1.039 $\pm$ 0.130	1705 $\pm$ 66	2146 $\pm$ 159	2384 $\pm$ 289	

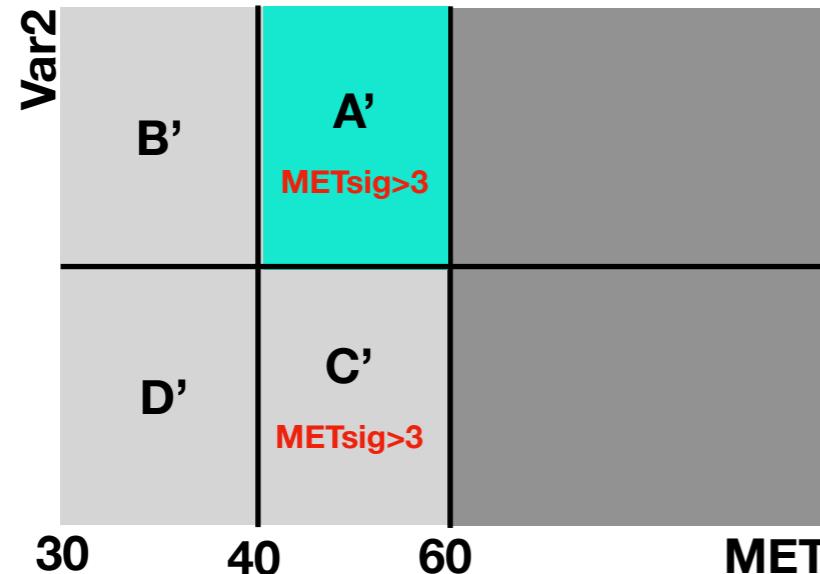
  

ee channel	Data in VR	Rmc	Rdata	MC	Data	ABCD
R <sub>zyzjet</sub> + metsig	2.153 $\pm$ 0.330	2.035 $\pm$ 0.269	1120 $\pm$ 92	1402 $\pm$ 60	1483 $\pm$ 293	
R <sub>zyzjet</sub>	1.267 $\pm$ 0.128	1.202 $\pm$ 0.110	1976 $\pm$ 106	2418 $\pm$ 80	2547 $\pm$ 338	
R <sub>zy</sub> + metsig	1.523 $\pm$ 0.172	1.583 $\pm$ 0.356	555 $\pm$ 33	838 $\pm$ 105	806 $\pm$ 176	
R <sub>zy</sub>	1.080 $\pm$ 0.082	1.058 $\pm$ 0.156	1175 $\pm$ 44	1617 $\pm$ 125	1651 $\pm$ 242	

=> Discrepancy between R with and without METsig cut, but R from data is consistent with R from MC  
 => ABCD estimates consistent with observed data

# Fake MET background

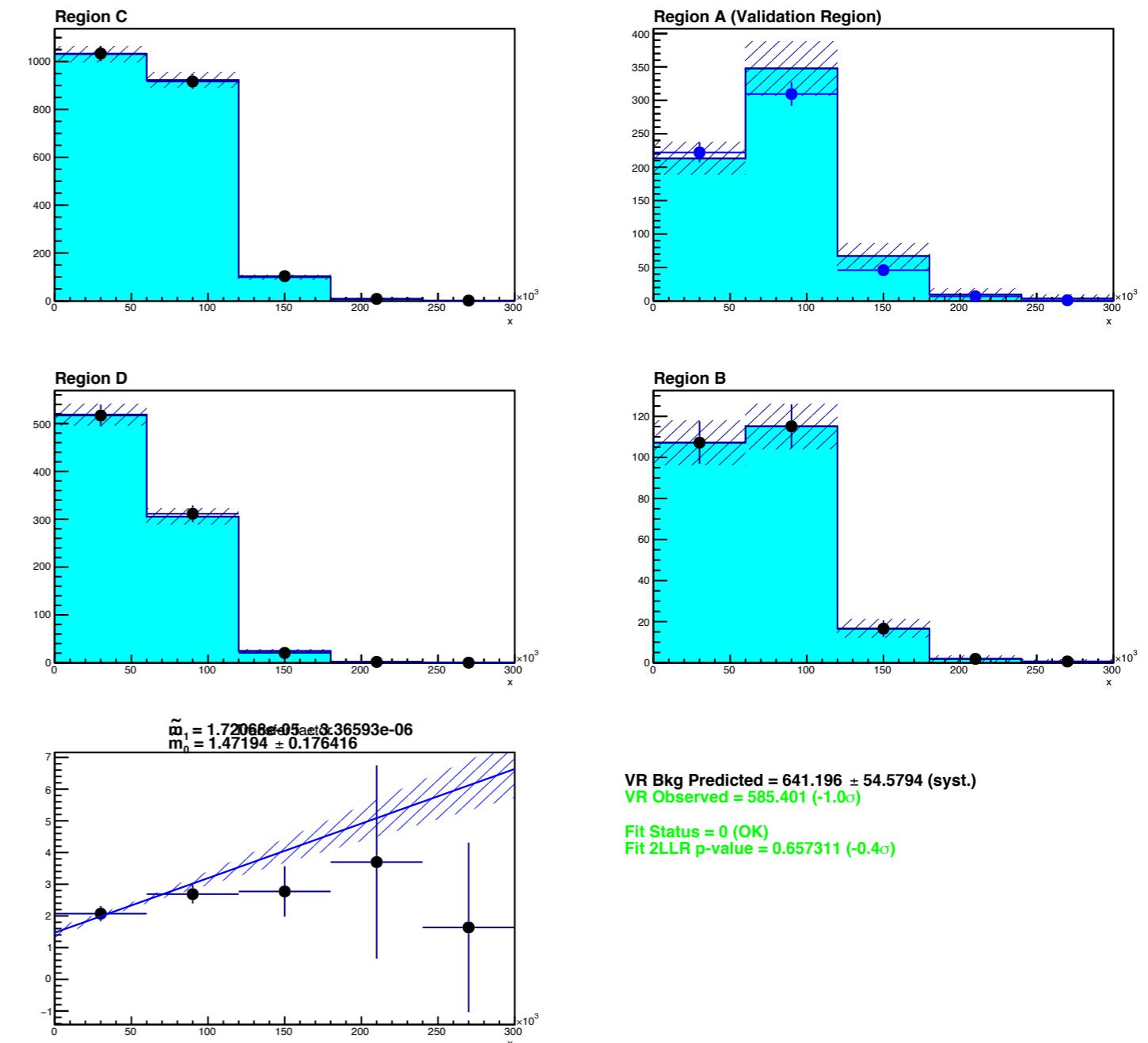
## TRooABCD, ee channel



$mT$  distribution in region A' from simultaneous fit on the B,C,D regions with TRooABCD

- $mT$  histograms in each B,C,D region as input
- Using data in VR, and subtracting Zjets from MC (first approximation)
- Correct for  $R = (nA \ nD) / (nC \ nB)$  by scaling the histogram in region D by  $1/R$  (with R estimated from Zy+jet MC)

## Zy MC sample



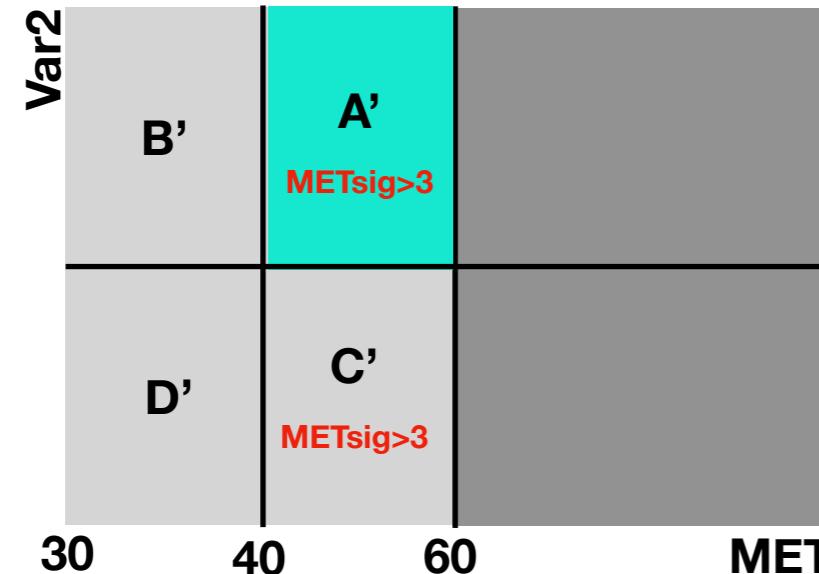
Only statistical uncertainties

$$\begin{aligned} nA(x) &= (m_0 + m_1(x)) \times nB(x) \\ nC(x) &= (m_0 + m_1(x)) \times nD(x) \end{aligned}$$

( $x$  labels the bin)

# Fake MET background

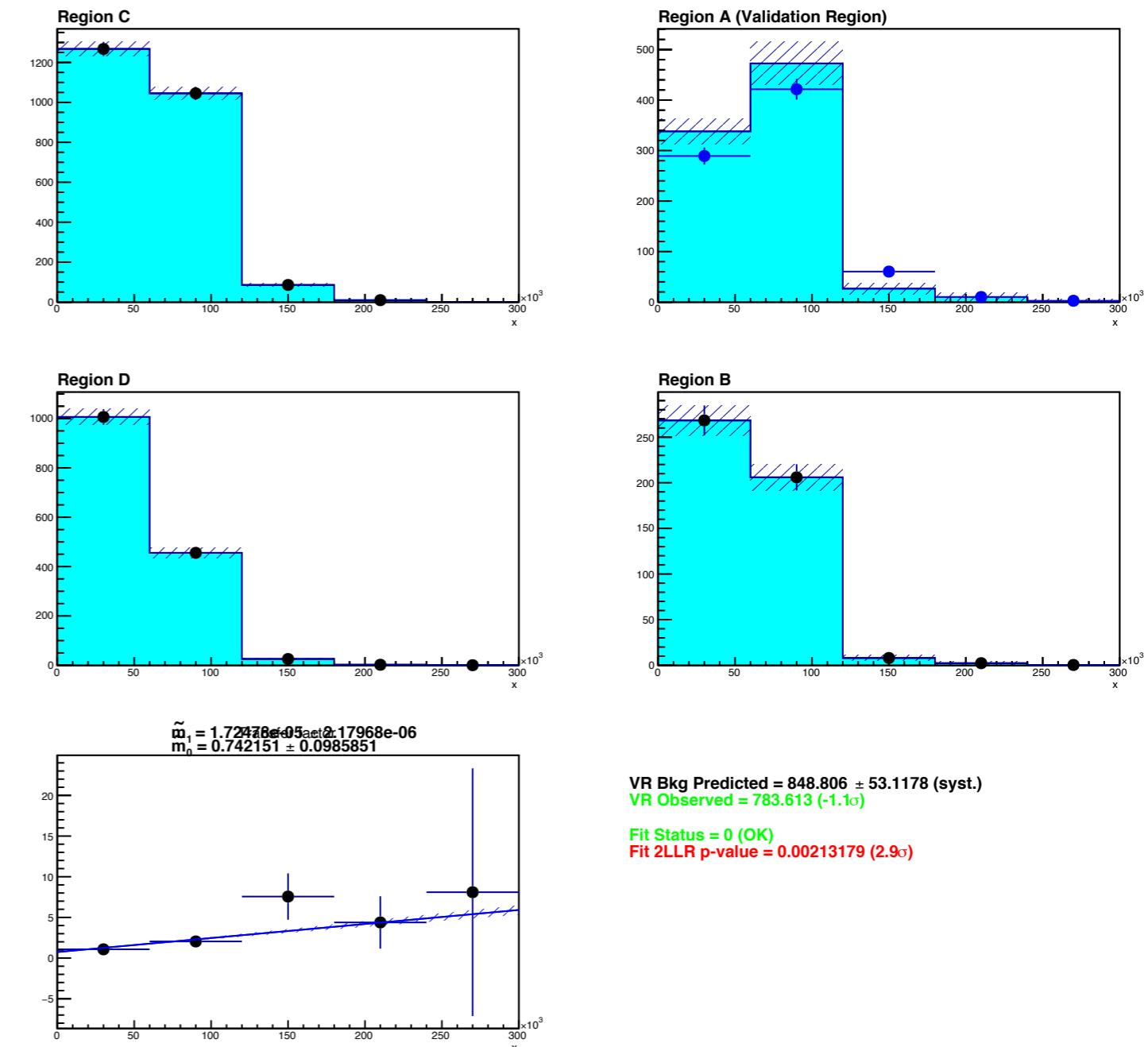
## TRooABCD, uu channel



$mT$  distribution in region  $A'$  from simultaneous fit on the  $B,C,D$  regions with TRooABCD

- $mT$  histograms in each  $B,C,D$  region as input
- Using data in VR, and subtracting Zjets from MC (first approximation)
- Correct for  $R=(nA\ nD)/(nC\ nB)$  by scaling the histogram in region D by  $1/R$  (with R estimated from Zy+jet MC)

## Zy MC sample



Only statistical uncertainties

$$\begin{aligned} nA(x) &= (m_0 + m_1(x)) \times nB(x) \\ nC(x) &= (m_0 + m_1(x)) \times nD(x) \end{aligned}$$

( $x$  labels the bin)