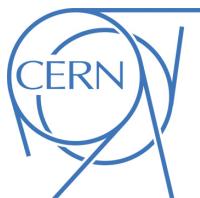




ATLAS Note

ANA-EXOT-2018-46-INT1

29th October 2019



Draft version 0.2

1

2

3

4

5

Search for Dark Matter produced in association with a Standard Model Higgs boson decaying to b -quarks with 139 fb^{-1} of pp collision data with the ATLAS detector

6

7

8

9

10

11

12

Abreu Juliao Ochoa De Castro, Maria Ines^a, Argyropoulos, Spyridon^b,
Brooijmans, Gustaaf^a, Burr, Jonathan Thomas^c, Chan, Chen-hsun^d, Chen, Jue^a,
Gadow, Paul Philipp^e, Fabiani, Veronica^f, Filthaut, Frank^f, Frost, James^g,
Ghosh, Anindya^b, Guest, Daniel^h, Junggeburth, Johannes Josef^e, Kortner,
Sandra^e, Lorenz, Jeanetteⁱ, Mallik, Usha^b, Matic, Andreaⁱ, Peters, Krisztian^c,
Pham, Tuan Minh^d, Pottgen, Ruth^j, Rieck, Patrick^e, Skorda, Eleni^k, Von Ahnen,
Janik^c, Wu, Sau Lan^d, Zhou, Chen^d

^a*Nevis Laboratory, Columbia University*

^b*University of Iowa*

^c*DESY, Hamburg and Zeuthen*

^d*Department of Physics, University of Wisconsin*

^e*Max-Planck-Institut fuer Physik*

^f*Radboud University Nijmegen and NIKHEF*

^g*University of Oxford, Particle Physics*

^h*University of California, Irvine*

ⁱ*Fakultaet fuer Physik, Ludwig-Maximilians-Universitaet Muenchen*

^j*Lund University (SE)*

^k*Lunds universitet, Fysiska institutionen*

13

14

15

16

17

18

19

20

21

22

23

24 This note documents a search for dark matter candidates produced in association with a
25 Standard Model Higgs boson in the $b\bar{b}$ decay channel. The search utilises a dataset of pp
26 collisions at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 139 fb^{-1} , recorded
27 with the ATLAS detector. The results are interpreted in the context of the 2-Higgs doublet
28 models with an extra vector or pseudoscalar mediator that mediates the interactions of the
29 dark matter particles with the Standard Model.

© 2019 CERN for the benefit of the ATLAS Collaboration.
30 Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

31 Contents

32	0 Version history [Dan, Spyros]	5
33	1 Executive Summary [Dan, Spyros]	6
34	1.1 Motivation	6
35	1.2 Previous results	7
36	1.3 Target	7
37	1.3.1 Signal models	7
38	1.3.2 Brief description of the analysis	8
39	1.3.3 Analysis improvements with respect to previous iteration	9
40	1.3.4 Blinding strategy	10
41	1.3.5 Milestones	10
42	2 Data and MC samples [Andrea, Spyros]	11
43	2.1 Data sample [Andrea]	11
44	2.2 Trigger	11
45	2.2.1 E_T^{miss} triggers in 0 lepton channel	13
46	2.2.2 Trigger efficiency measurements	13
47	2.2.3 Systematic uncertainties	15
48	2.2.4 Discussion and comparison to previous analysis	15
49	2.3 MC samples [Spyros]	17
50	2.3.1 2HDM+a signal	17
51	2.3.2 Z' 2HDM signal	18
52	2.3.3 V+jets	18
53	2.3.4 $t\bar{t}$	19
54	2.3.5 Single top	20
55	2.3.6 Diboson	21
56	2.3.7 SM $Vh(b\bar{b})$	21
57	2.3.8 $t\bar{t} + Z/h$	21
58	3 Object definitions [Andrea, Veronica]	22
59	3.1 Electrons	22
60	3.2 Muons	22
61	3.3 τ -leptons	23
62	3.4 Small-R jets	23
63	3.5 Large-R jets	24
64	3.6 Variable-radius track jets	24
65	3.7 b -tagging	25
66	3.8 Missing transverse momentum E_T^{miss}	25
67	3.9 Object-based E_T^{miss} significance S	26
68	3.10 Overlap removal	26
69	4 Event selection [Andrea, Anindya, Jay]	26
70	4.1 Event cleaning	27
71	4.2 Signal regions	27
72	4.2.1 Common event selection	27

Not reviewed, for internal circulation only

73	4.2.2	Resolved Region	28
74	4.2.3	Merged Region	29
75	4.3	Control regions	30
76	4.3.1	1-lepton channel	30
77	4.3.2	2-lepton channel	31
78	5	Systematics [Spyros, Andrea, Eleni, Veronica]	33
79	5.1	Experimental systematics [Andrea]	33
80	5.2	Theory systematics [Spyros, Eleni, Veronica]	36
81	5.2.1	Theoretical uncertainty sources	36
82	5.2.2	Propagation of theory uncertainties to likelihood fit	36
83	5.2.3	Top	37
84	5.2.4	$V+jets$	37
85	5.2.5	Signal	37
86	5.2.6	Others	37
87	5.2.7	Derivation of the theory uncertainties at truth level	37
88	6	Pre-fit results [Anindya]	39
89	6.1	0-lepton [Anindya]	39
90	6.2	1-lepton [Anindya]	41
91	6.3	2-lepton [Anindya]	44
92	6.3.1	2-muon channel	44
93	6.3.2	2-electron channel	46
94	7	Statistical analysis [Andrea, Spyros, Dan]	48
95	7.1	Likelihood definition	48
96	7.2	Fit variables and event categories	48
97	7.3	Treatment of nuisance parameters [Andrea, Spyros]	50
98	7.4	Results [Andrea, Dan, Spyros]	50
99	7.5	Discussion [Dan, Spyros]	50
100	Appendices		60
101	A	Optimisation of cut-based analysis [Andrea]	60
102	B	Removal of non-collision background with tight jet cleaning [Philipp]	61
103	C	Optimizaiton using Boosted Decision Tree [Jay]	62
104	C.1	Preslections	63
105	C.2	BDT training	63
106	C.3	Categorization optimization	64
107	C.4	Statistical analysis results	65
108	C.5	Performance without the pseudo continuous b-tagging scores of the first two leading (track) jets in the training	67
110	D	Higgs tagging [Jue, Ines, Dan]	74

Not reviewed, for internal circulation only

111	E MV2 vs DL1r [Veronica, Janik]	74
112	F Pre-fit results [Anindya]	75
113	F.1 0-lepton [Anindya]	75
114	F.2 1-lepton [Anindya]	83
115	F.3 2-lepton [Anindya]	91
116	G Studies of V+jets flavour composition [Anindya]	107
117	G.1 V+jets flavour labelling	107
118	G.2 How to merge V+jets flavours	107
119	G.2.1 Merging strategy	108
120	H Implemntation of E_T^{miss} filtered $t\bar{t}$, B-filtered Diboson VZ($Z \rightarrow b\bar{b}$) and $Z\nu\nu$+jets PTV sliced samples [Anindya]	112
121	H.1 Merging inclusive with E_T^{miss} filtered $t\bar{t}$ samples	112
122	H.2 Merging $Z\nu\nu$ +jets MAX(HT,PTV) with $Z\nu\nu$ +jets $p_T(V)$ sliced samples	114
123	H.3 Merging inclusive diboson VZ($Z \rightarrow q\bar{q}$) with B-filtered diboson VZ($Z \rightarrow b\bar{b}$) samples	116
125	I Resolved/Merged combination strategy [Jon, James]	118
126	J Photon CR [Eleni]	118
127	K Truth-level studies of signals [Anindya, Andrea, Spyros]	118
128	K.1 Comparison of ggF and b-associated production	118
129	K.2 Cross-sections, widths, BR etc as a function of $m_A, m_a, \tan\beta, \sin\theta$	118
130	L Merging of V+jets, $t\bar{t}$ and diboson samples [Anindya]	118

0 Version history [Dan, Spyros]

- 132 **Version 0.2** Implementation of comments from JDM conveners on v0.1
- 133 **Version 0.1**
- 134 First version aiming to provide the basic details of the analysis leading to an EB request.
- 135 **New items** (with respect to previous analysis iteration [1])
- 136 • inclusion of 2018 dataset
 - 137 • optimised/consolidated analysis selection (Sec. 4)
 - 138 • interpretation in the context of the 2HDM+a model (Sec. 1.3.1, 2.3.1)
 - 139 • derivation of theory uncertainties in Mono-H($b\bar{b}$) phase space (Sec. 5.2)
 - 140 • optimisation studies (as documented in Appendices)
- 141 **Items planned for next iterations**
- 142 • final derivations including new jet collections

- 143 • switching to particle-flow jet collection
- 144 • inclusion of detector systematics
- 145 • derivation of theory uncertainties
- 146 • blinded fit to complete dataset excluding Higgs peak
- 147 • extended Z' 2HDM and 2HDM+a signal grids with increased granularity, harmonised settings with
- 148 CMS (Z' 2HDM) and ME-weights for $\tan\beta$, $\sin\theta$ scans (2HDM+a)
- 149 • expansion of INT note

150 **Framework tags for derivations, analysis framework, fit**

- 151 • Derivation tags: p3840 - p3842 for MC, all sample lists [here](#)
- 152 • Analysis framework: Using XAMPPmonoH-00-11-00
- 153 • WSMaker tag: N/A

154 **1 Executive Summary [Dan, Spyros]**

155 This note documents a search for Dark Matter (DM) candidates produced in association with a Standard
 156 Model (SM) Higgs boson h decaying to b -jets, utilising the complete Run-2 dataset of pp collisions,
 157 corresponding to an integrated luminosity of 139 fb^{-1} .

158 The previous iteration of the analysis, corresponding to a dataset of 80 fb^{-1} was published as a preliminary
 159 result in [1]. The corresponding supporting note can be found in [2].

160 **1.1 Motivation**

161 Multiple astrophysical observations [3, 4] indicate that a large fraction of the matter density of the universe
 162 is in the form of DM, i.e. matter that does not interact via electromagnetic interactions. Its nature is yet
 163 unknown and the Standard Model of particle physics (SM) does not provide any suitable DM candidates,
 164 i.e. stable, neutral and weakly interacting massive particles. The search for DM candidates is actively
 165 being pursued in direct and indirect detection experiments as well as in collider experiments.

166 Since DM particles would interact weakly with SM matter, if they are produced at the LHC, they would
 167 leave a missing transverse energy signature in the detector. For this reason one has to rely on the presence
 168 of an additional visible object X produced in association with the DM particles, which is used to identify
 169 DM-like events. The visible object can come from initial-state QCD, weak or electromagnetic radiation,
 170 which give rise to the so-called mono-jet, mono- Z/W , and mono-photon final states.

171 A particularly interesting alternative is for the visible object to be a SM Higgs boson, which leads to the
 172 so-called mono-Higgs signature. The particularity of the mono-Higgs signature is that since the Higgs
 173 couplings to initial state quarks and gluons are Yukawa suppressed, the Higgs is predominantly produced
 174 in the final state, therefore unlike the other mono- X signatures, mono-Higgs can probe directly the DM/SM
 175 interaction vertex [5]. Furthermore, several model suggest that DM could be connected to the electroweak
 176 symmetry breaking [6, 7], or even that DM might only couple to SM through the Higgs sector (Higgs

¹⁷⁷ portal models) [8]. This renders the searches for DM through Higgs signatures a compelling alternative
¹⁷⁸ to the LHC DM search programme.

¹⁷⁹ 1.2 Previous results

¹⁸⁰ DM searches have been performed in Run 1 and with the first half of the Run 2 ATLAS data in a variety
¹⁸¹ of models, as described in [9].

¹⁸² If the mass of the DM candidate is smaller than half of the SM Higgs-boson mass the SM Higgs boson can
¹⁸³ directly decay to DM particles, thereby increasing the invisible SM Higgs width. ATLAS has performed
¹⁸⁴ a combination of searches with $H \rightarrow$ invisible signatures [10], yielding constraints on the WIMP-
¹⁸⁵ nucleon cross-section which are much tighter than those obtained from direct detection experiments. If
¹⁸⁶ $m_{\text{DM}} > m_h/2$, a direct decay of the Higgs boson to DM particles is not possible, so for $m \simeq 60$ GeV,
¹⁸⁷ mono-Higgs signatures provide the strongest constraints on the coupling of DM particles to the Higgs
¹⁸⁸ sector.

¹⁸⁹ The mono-Higgs searches can be performed in several channels, according to the decays of the SM Higgs
¹⁹⁰ boson. Previous results using the $h \rightarrow b\bar{b}$ final state have been presented by ATLAS in [1, 11] and CMS
¹⁹¹ in [12].

¹⁹² A comprehensive review of recent ATLAS results on DM searches can be found in [9].

¹⁹³ 1.3 Target

¹⁹⁴ 1.3.1 Signal models

¹⁹⁵ The analysis uses two simplified benchmark models for the interpretation of the search results: the Z' -
¹⁹⁶ 2HDM model and the 2HDM+ a model. Both models correspond to an extended Higgs sector with two
¹⁹⁷ complex scalar doublets. Different incarnations of the 2HDM models exist with: (i) one doublet coupling
¹⁹⁸ only to bosons and the second only to fermions (Type-I), (ii) one doublet coupling to isospin $-1/2$ fermions
¹⁹⁹ and the second coupling to isospin $+1/2$ fermions (Type-II), (iii) one doublet coupling only to charged
²⁰⁰ leptons and the second doublet coupling to fermions (lepton-specific), (iv) same as Type-II but with the
²⁰¹ doublet coupling to charged fermions inverted (flipped) [13]. These models have 5 physical Higgs bosons
²⁰² in the spectrum: two CP-even neutral states h (SM-like) and H , with $m_H > m_h$, two charged states H^\pm
²⁰³ and a CP-odd neutral state A . These particles introduce 7 free parameters: the masses of the 5 Higgs
²⁰⁴ bosons, the ratio of the vacuum expectation values of the two Higgs doublets ($\tan \beta$), and the mixing angle
²⁰⁵ α which diagonalises the mass matrix of the neutral CP-even Higgs bosons.

²⁰⁶ The Z' -2HDM model extends the 2HDM models with an additional $U(1)_{Z'}$ gauge symmetry that is
²⁰⁷ spontaneously broken by a SM singlet scalar above the electroweak symmetry breaking scale, thus
²⁰⁸ generating a mass for the Z' boson [14]. The coupling constant of the Z' boson $g_{Z'}$ as well as its mass
²⁰⁹ $m_{Z'}$ are free parameters. The $h + E_T^{\text{miss}}$ signature arises in this model via the process $pp \rightarrow Z' \rightarrow Ah$,
²¹⁰ with the A boson further decaying into a $\chi\bar{\chi}$ pair of fermionic DM particles, as shown in Fig. 1

²¹¹ The 2HDM+ a models further extend the 2HDM models by including an additional pseudoscalar singlet,
²¹² that mediates the interactions between SM and fermionic DM. This singlet mixes with the CP-odd com-
²¹³ ponent of the 2HDM Higgs doublets, with the mixing angle θ and the mass of the additional pseudoscalar

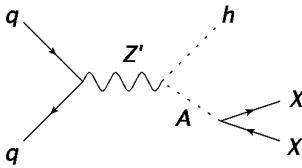
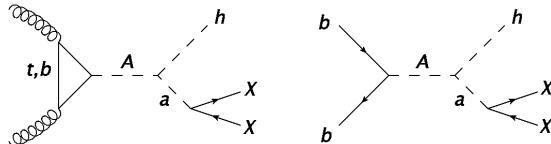


Figure 1: Feynman diagram corresponding to the production of the $h + \chi\bar{\chi}$ signature in the Z' 2HDM model.

214 a being free parameters of the model [15]. The production of the $h + \chi\bar{\chi}$ final state can proceed through
215 two different channels: the gluon-gluon fusion (ggF) and the b -initiated production, as shown in Fig. 2.



216 In the Type-II 2HDM models, the $q\bar{q}A$ coupling scales like $-\cot\beta$ for up-type quarks and $\tan\beta$ for
217 down-type quarks. Therefore coupling to bottom quarks is preferred than coupling to top quarks for high
218 $\tan\beta$ values and as a result the b -initiated contribution dominates for $\tan\beta \gtrsim 0.5$ (see App. K for more
219 details). The coupling Aha scales like $\cos\theta\sin\theta$.

220 For pseudoscalar extensions of the Type-I and lepton-specific 2HDM models, the couplings of the SM-like
221 Higgs boson to bottom quarks are suppressed by $1/\tan\beta$, where $\tan\beta$ is the ratio of the vacuum expectation
222 value of the two Higgs doublets, and therefore these models are irrelevant for the mono-H($b\bar{b}$) search,
223 while similar sensitivity is expected between the Type-II and flipped 2HDM models [16]. For this reason
224 the simplified DM models outlined above are based on Type-II 2HDM models.

225 1.3.2 Brief description of the analysis

226 The analysis uses final states with b -jets and missing transverse energy (E_T^{miss}) to search for the production
227 of DM candidates in association with a SM Higgs boson. Since the E_T^{miss} is correlated with the transverse
228 momentum of the Higgs boson, at low E_T^{miss} , the b -jets from the Higgs decay are well separated and can be
229 reconstructed as small-radius jets. As the E_T^{miss} increases, the b -jets from the Higgs decay begin to merge
230 in a single jet, and therefore reconstructing the Higgs-boson candidate with a large-radius jet becomes
231 necessary. The analysis employs both reconstruction techniques, referred to respectively as “resolved”
232 and “merged”, with the resolved analysis used for $E_T^{\text{miss}} < 500$ GeV and the merged analysis used for
233 $E_T^{\text{miss}} \geq 500$ GeV.

234 The analysis consists of three channels: 0, 1 and 2-leptons, where the 0-lepton channel defines the “Signal
235 Region” (SR) of the search and the 1 and 2-lepton regions are included as “Control Regions” (CR1,
236 CR2) in order to constrain the major backgrounds, top-pair production ($t\bar{t}$) and W, Z bosons produced
237 in association with jets. A profile likelihood fit is performed to extract the upper limit on the signal
238 cross-section, which is then used to extract exclusion limits on the $(m_{Z'}, m_A)$ or (m_a, m_A) , m_a , $\tan\beta$ planes
239 for the Z' 2HDM and 2HDM+a benchmark signals respectively, as shown in Fig. 3. A detailed description
240 of the fit model can be found in Sec. 7.

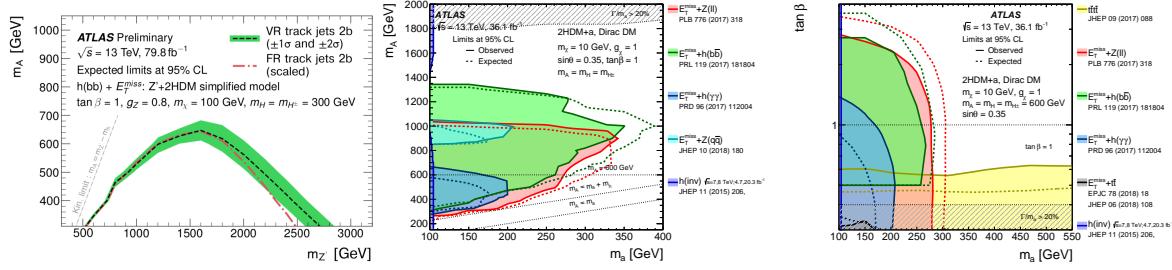


Figure 3: Exclusion contours on the $(m_{Z'}, m_A)$ plane for the Z' -2HDM model (left) and on the (m_a, m_A) (middle) and $(m_a, \tan \beta)$ (right) planes for the 2HDM+ a model, taken from [15].

1.3.3 Analysis improvements with respect to previous iteration

Several improvements have been considered with respect to the previous iteration of the analysis [1, 2]. Some of these have not been implemented in the current analysis but are documented in the appendices for future reference.

The improvements, in terms of the objects or the analysis strategy, which have already been incorporated in the analysis are

- **inclusion of 2018 dataset**
- **usage of particle-flow jets**
- **selection optimisation:** a comprehensive optimisation procedure was used to optimise and consolidate the selection used to define the signal and control regions, as described in App. A. This resulted in an increase of the expected sensitivity but also a simplification of the analysis, as a result of inefficient analysis cuts being dropped.
- **signal models:** the previous result [1] only included an interpretation in the context of the Z' -2HDM model. In this iteration the 2HDM+ a model is also included and for the first time dedicated studies of the region with $> 2 b$ -jets, sensitive to the high $\tan \beta$ region is
- **theory uncertainties:** these were previously taken from the search for the production of the SM Higgs boson in the $VH(b\bar{b})$ final state. In this iteration dedicated studies were performed to derive the theory uncertainties in the Mono-H($b\bar{b}$) phase-space.

Additional improvements of the analysis strategy, which are currently under study and might be included in the final result include:

- **muon-in-jet correction:** for b -jets that contain leptonic b -decays, including the muon momentum to the b -jet momentum leads to an improved resolution for the mass of the Higgs candidate and is therefore expected to lead to an improvement in the expected sensitivity. Initial studies on the performance of the muon-in-jet correction have shown an improvement of the $m(b\bar{b})$ resolution. The corrected $m(b\bar{b})$ will be propagated to the fit inputs and the decision on whether to fit the $m(b\bar{b})$ distribution before or after the muon-in-jet correction will be studied in the context of the PL fit.
- **3+ tag/2_{in}1+_{out} tag. region:** in the 2HDM+ a model, the b -initiated production of the $h + E_T^{\text{miss}}$ signature leads to more than 2 b -jets in the final state, due to $g \rightarrow b\bar{b}$ splitting in the initial state.

269 Since the b -initiated production dominates for $\tan\beta > 5$, including these regions in the fit is expected
 270 to enhance the sensitivity in the high $\tan\beta$ region. This will be studied in the context of the PL fit.

271 1.3.4 Blinding strategy

272 In order to avoid biasing the analysis towards signal-like effects that may appear in the data, data events
 273 with a reconstructed Higgs candidate mass with $70 \leq m_H \leq 140$ GeV (“Higgs window”) and with at least
 274 1 b -tagged jet are blinded (in both the resolved and merged regions). Fits to data are also performed using
 275 a background-only hypothesis.

276 The proposed unblinding procedure, is as follows:

- 277 • **Stage 0:** initially a background-only fit to data excluding the Higgs window is performed, in order
 278 to investigate the behaviour of the fit model (pulls, correlations). The regions included in this fit
 279 will be the 1 and 2-lepton control regions. The possibility of including the 0-lepton m_H sidebands
 280 is also considered, however this might only be used as a validation region.
- 281 • **Stage 1:** at this stage the whole m_H spectrum in the control and signal regions will be used in the
 282 profile likelihood fit. Either a background-only fit or an unconditional $s + b$ fit will be performed,
 283 however in the latter case the fitted signal strength and the Higgs window in the post-fit plots will
 284 remain blinded.
- 285 • **Stage 2:** this corresponds to the completely unblinded result, where an unconditional $s + b$ fit is
 286 performed to the complete m_H spectrum in order to extract the upper limits on the signal cross-
 287 section.

288 1.3.5 Milestones

289 A tentative analysis schedule towards up to the first circulation of the paper draft is presented in Fig. 4.
 290 The major deciding factors for a timely delivery of the analysis results are:

- 291 – the production of a bug-free set of derivations including the particle-flow jet collection - estimated
 292 time: 2 months
- 293 – obtaining a stable preliminary fit setup for “stage-0 unblinding”, using the full dataset and including
 294 the full suite of detector systematics, but likely using the theory systematics from the previous
 295 iteration [1] - estimated time: 2-3 months
- 296 – the derivation of the theory uncertainties - estimated time: 3-4 months
- 297 – the production of the final set of fit inputs based on the final round of derivations - estimated time:
 298 2 months
- 299 – establishing the final fit model, based on the final fit inputs and including the new theory systematics
 300 - estimated time: 2-3 months

301 Based on these estimated delivery times, a set of milestones have been put in place as follows:

- 302 • Stage-0 unblinding (end of December): the goal is to judge whether the analysis can proceed to a
 303 background-only fit to the whole $m_{b\bar{b}}$ spectrum in the signal region

Not reviewed, for internal circulation only

- Unblinding approval (end of February): the goal is to judge whether the analysis can do an unblinded unconditional fit to extract the observed upper limits on the cross-section
- Exotics approval (beginning of April): the goal is to decide whether the analysis and paper draft are mature enough to proceed to ATLAS circulation
- 1st circulation of paper draft (beginning of May): at this point it will also be decided if a CONF-note conversion will be done in order to show the results to the summer conferences

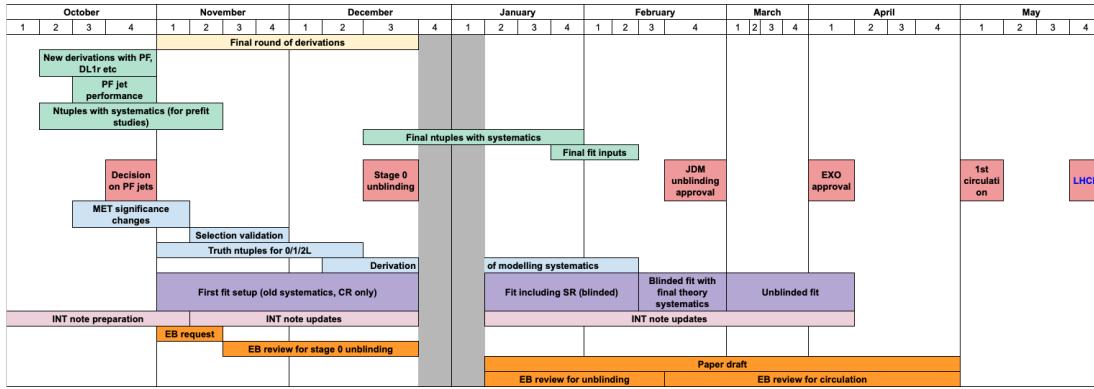


Figure 4: Analysis schedule. Milestones are indicated in red.

310 2 Data and MC samples [Andrea, Spyros]

311 2.1 Data sample [Andrea]

312 This search makes use of proton–proton collisions at a centre-of-mass energy of 13 TeV recorded between
 313 2015 and 2017. The following Good-Run-Lists are used:

- data15_13TeV.periodAllYear_DetStatus-v89-pro21-02_Uncertain_PHYS_StandardGRL_All_Good_25ns.xml
- data16_13TeV.periodAllYear_DetStatus-v89-pro21-01_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml
- data17_13TeV.periodAllYear_DetStatus-v99-pro22-01_Uncertain_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml
- data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Uncertain_PHYS_StandardGRL_All_Good_25ns_Triggerno17e33prim.xml

318 The resulting dataset corresponds to an integrated luminosity of 3.2 fb⁻¹, 33.0 fb⁻¹, 44.3 fb⁻¹ and
 319 58.5 fb⁻¹ for each Good-Run-Lists, respectively. The total integrated luminosity is 139.0 fb⁻¹. The
 320 proton bunch gap amounts to 25 ns.

321 2.2 Trigger

322 Events in the 0 and 1 lepton regions are selected using E_T^{miss} triggers. Events in the 2 lepton region are
 323 selected with unprescaled single lepton triggers. Their thresholds are determined by requiring lowest
 324 unprescaled single-lepton triggers [17]. The full list of the triggers used in all channels can be found in
 325 Table 1.

Not reviewed, for internal circulation only

Period	0 lepton	1 lepton	2 lepton + E_T^{miss} trigger SF measurement
2015	HLT_XE70_MHT	HLT_XE70_MHT	$\text{HLT}_{\text{E}24\text{-LHMEDIUM}}\text{L1EM20VH}$ OR $\text{HLT}_{\text{E}60\text{-LHMEDIUM}}$ OR $\text{HLT}_{\text{E}120\text{-LHLOOSE}}$ OR $\text{HLT}_{\text{MU}20\text{-ILOOSE}}\text{L1MU15}$ OR $\text{HLT}_{\text{MU}50}$
2016 (A)	HLT_XE90_MHT_L1XE50	HLT_XE90_MHT_L1XE50	$\text{HLT}_{\text{E}60\text{-LHMEDIUM}}\text{NOD0}$ OR $\text{HLT}_{\text{E}140\text{-LHLOOSE}}\text{NOD0}$ OR $\text{HLT}_{\text{MU}40}$ OR $\text{HLT}_{\text{MU}50}$
2016 (B-D3)	HLT_XE90_MHT_L1XE50	HLT_XE90_MHT_L1XE50	$\text{HLT}_{\text{E}60\text{-LHMEDIUM}}\text{NOD0}$ OR $\text{HLT}_{\text{E}140\text{-LHLOOSE}}\text{NOD0}$ OR $\text{HLT}_{\text{MU}24\text{-IVARMEDIUM}}$ OR $\text{HLT}_{\text{MU}50}$
2016 (D4-E3)	HLT_XE110_MHT_L1XE50	HLT_XE110_MHT_L1XE50	$\text{HLT}_{\text{E}26\text{-LHTIGHT}}\text{NOD0_IVARLOOSE}$ OR $\text{HLT}_{\text{E}60\text{-LHMEDIUM}}\text{NOD0}$ OR $\text{HLT}_{\text{E}140\text{-LHLOOSE}}\text{NOD0}$ OR $\text{HLT}_{\text{MU}24\text{-IVARMEDIUM}}$ OR $\text{HLT}_{\text{MU}26\text{-IVARMEDIUM}}$ OR $\text{HLT}_{\text{MU}50}$
2016 (F1)	HLT_XE110_MHT_L1XE50	HLT_XE110_MHT_L1XE50	$\text{HLT}_{\text{E}26\text{-LHTIGHT}}\text{NOD0_IVARLOOSE}$ OR $\text{HLT}_{\text{E}60\text{-LHMEDIUM}}\text{NOD0}$ OR $\text{HLT}_{\text{E}140\text{-LHLOOSE}}\text{NOD0}$ OR $\text{HLT}_{\text{MU}26\text{-IVARMEDIUM}}$ OR $\text{HLT}_{\text{MU}50}$
2016 (F2-)	HLT_XE110_MHT_L1XE50	HLT_XE110_MHT_L1XE50	$\text{HLT}_{\text{E}26\text{-LHTIGHT}}\text{NOD0_IVARLOOSE}$ OR $\text{HLT}_{\text{E}60\text{-LHMEDIUM}}\text{NOD0}$ OR $\text{HLT}_{\text{E}140\text{-LHLOOSE}}\text{NOD0}$ OR $\text{HLT}_{\text{MU}26\text{-IVARMEDIUM}}$ OR $\text{HLT}_{\text{MU}50}$
2017	HLT_XE110_PUFIT_L1XE55	HLT_XE110_PUFIT_L1XE55	OR $\text{HLT}_{\text{E}60\text{-LHMEDIUM}}\text{NOD0}$ OR $\text{HLT}_{\text{E}140\text{-LHLOOSE}}\text{NOD0}$ OR $\text{HLT}_{\text{MU}26\text{-IVARMEDIUM}}$ OR $\text{HLT}_{\text{MU}50}$
2018 (B-C5)	HLT_XE110_PUFIT_70_L1XE55	HLT_XE110_PUFIT_70_L1XE55	OR $\text{HLT}_{\text{E}60\text{-LHMEDIUM}}\text{NOD0}$ OR $\text{HLT}_{\text{E}140\text{-LHLOOSE}}\text{NOD0}$ OR $\text{HLT}_{\text{MU}26\text{-IVARMEDIUM}}$ OR $\text{HLT}_{\text{MU}50}$
2018 (C5-)	HLT_XE110_PUFIT_65_L1XE55	HLT_XE110_PUFIT_65_L1XE55	OR $\text{HLT}_{\text{E}60\text{-LHMEDIUM}}\text{NOD0}$ OR $\text{HLT}_{\text{E}140\text{-LHLOOSE}}\text{NOD0}$ OR $\text{HLT}_{\text{MU}26\text{-IVARMEDIUM}}$ OR $\text{HLT}_{\text{MU}50}$

Table 1: E_T^{miss} and single-lepton triggers used in the analysis.

326 **2.2.1 E_T^{miss} triggers in 0 lepton channel**

327 The following E_T^{miss} triggers were used by ATLAS in Run 2:

- 328 • HLT_XE70_MHT seeded by L1_XE50 for 2015 data (3.2 fb^{-1}).
- 329 • HLT_XE90_MHT_L1XE50 for 2016 data (periods A-D3, 6.1 fb^{-1})
- 330 • HLT_XE100_MHT_L1XE50 for 2016 data (periods D4-, 27.1 fb^{-1}).
- 331 • HLT_XE110_PUFIT_L1XE55 for 2017 data (44.3 fb^{-1}).
- 332 • HLT_XE110_PUFIT_XE70_L1XE50 for 2017 data (58.5 fb^{-1}).

333 **TODo: Update luminosities for 2016 triggers - changed since CONF due to new GRLs**

334 To achieve a good sensitivity of the analysis, also to Mono- h models which typically result in relatively low
 335 E_T^{miss} values, a value for the offline E_T^{miss} cut is chosen where these triggers are not fully efficient, namely
 336 $E_T^{\text{miss}} > 150 \text{ GeV}$. However, the trigger turn-on curve is not well modeled in MC. Thus it is mandatory
 337 to measure trigger efficiencies so that the simulated trigger response can be corrected to match the data.
 338 Therefore, trigger efficiencies are measured in both data and MC from a single-muon measurement region,
 339 and scale factors are calculated to correct turn-ons in MC to those in data in the zero-lepton signal region
 340 and the single-muon control region.

341 In the High Level Trigger (HLT) system E_T^{miss} is reconstructed using only calorimeter information only.
 342 Hence muons are barely taken into account in the reconstruction of the online E_T^{miss} . Therefore, the E_T^{miss}
 343 seen by the trigger for events with a muon corresponds to the offline E_T^{miss} obtained when the contribution
 344 of the muon is not included in the E_T^{miss} calculation. This is denoted by $E_{T,\mu \text{ invis.}}^{\text{miss}}$. Because of this, events
 345 collected with the muon trigger can be used to measure the trigger turn-on curve of the E_T^{miss} trigger. The
 346 single muon-region is orthogonal to the signal region, which features a lepton veto, and is dominated by
 347 $W(\mu, \nu) + \text{jets}$ and $t\bar{t}$ production.

348 **2.2.2 Trigger efficiency measurements**

349 The efficiencies of E_T^{miss} triggers are derived in a single-muon measurement region. The single muon
 350 triggers used can be found in Table 1. The event selection is almost equivalent to the selection in the
 351 resolved regime, as described in Section 4, except that the cut on E_T^{miss} is dropped. Furthermore, the
 352 efficiencies are calculated inclusively in the number b -jets, including events with 0 b -jets as well. This
 353 allows for a larger statistics in the measurement region.

354 The E_T^{miss} trigger efficiency is defined by:

$$\text{efficiency} = \frac{\# \text{Events passed selection AND } E_T^{\text{miss}} \text{ trigger requirement}}{\# \text{Events passed selection}} \quad (1)$$

355 The efficiencies are calculated for each E_T^{miss} trigger separately within the data-taking periods in which
 356 it was used, and separately for data and MC. To mimic the E_T^{miss} topology on trigger level the trigger
 357 efficiencies are measured as a function of $E_{T,\mu \text{ invis.}}^{\text{miss}}$. The resulting efficiency curves for each E_T^{miss} trigger
 358 are shown in Figure 5.

Not reviewed, for internal circulation only

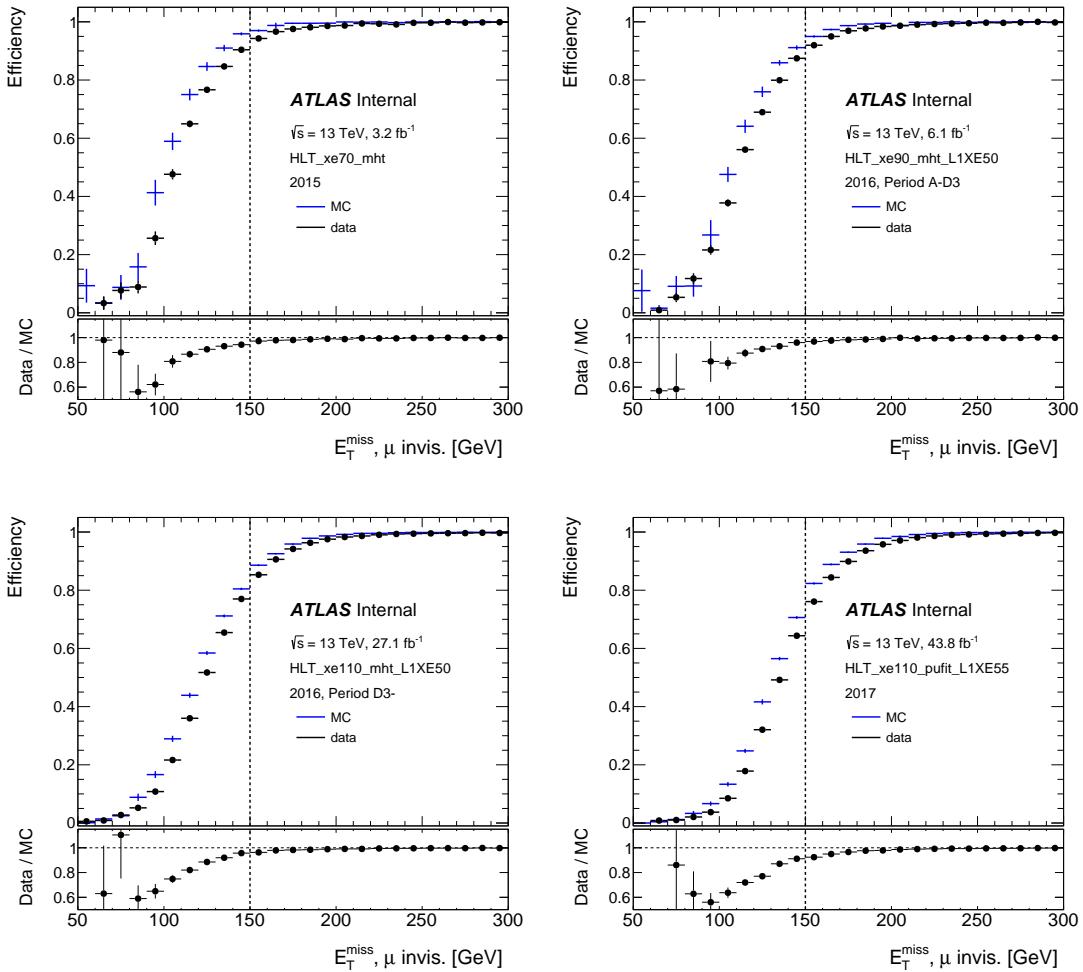


Figure 5: Measured trigger efficiencies as a function of offline $E_{T,\mu \text{ invis.}}^{\text{miss}}$ in data and MC for the E_T^{miss} triggers used in 2015–2018. The plots are shown for 0,1 and 2 tags together. The lower panels provide the ratio of data and MC events (the scale factor).

359 Scale factors (SF) are defined as the ratio of E_T^{miss} trigger efficiencies for data and MC:

$$\text{SF} = \frac{\text{Efficiency}_{\mu}^{\text{data}}}{\text{Efficiency}_{\mu}^{\text{MC}}} \quad (2)$$

360 This is done in E_T^{miss} bins of 10 GeV. To calculate the data-driven corrections for the MC trigger turn-on
361 curves, the scale factors are fitted for each E_T^{miss} trigger starting in the range $100 \text{ GeV} < E_{T,\mu \text{ invis.}}^{\text{miss}} <$
362 300 GeV using the following fit function:

$$f(x) = p_0 \cdot \left[1 + \text{erf} \left(\frac{x - p_1}{\sqrt{2}p_2} \right) \right] + p_3 \quad (3)$$

363 where $x = E_{T,\mu \text{ invis.}}^{\text{miss}}$. The correction factor applied to the MC in the zero-lepton signal region and
364 one-lepton control region for a is given by evaluating $f(E_T^{\text{miss}})$ or $f(E_{T,\mu \text{ invis.}}^{\text{miss}})$, respectively. The scale

Not reviewed, for internal circulation only

365 factors measured for the different E_T^{miss} are shown in Figure 6 together with the fitted SF curves. The
 366 hatched band shows the statistical uncertainty of the SF fit curve, defined as the 1σ fit uncertainty. As a
 367 validation, the data and MC efficiencies are compared once more in Figure 7, after the correction factors
 368 are applied to the simulation. Good agreement is observed.

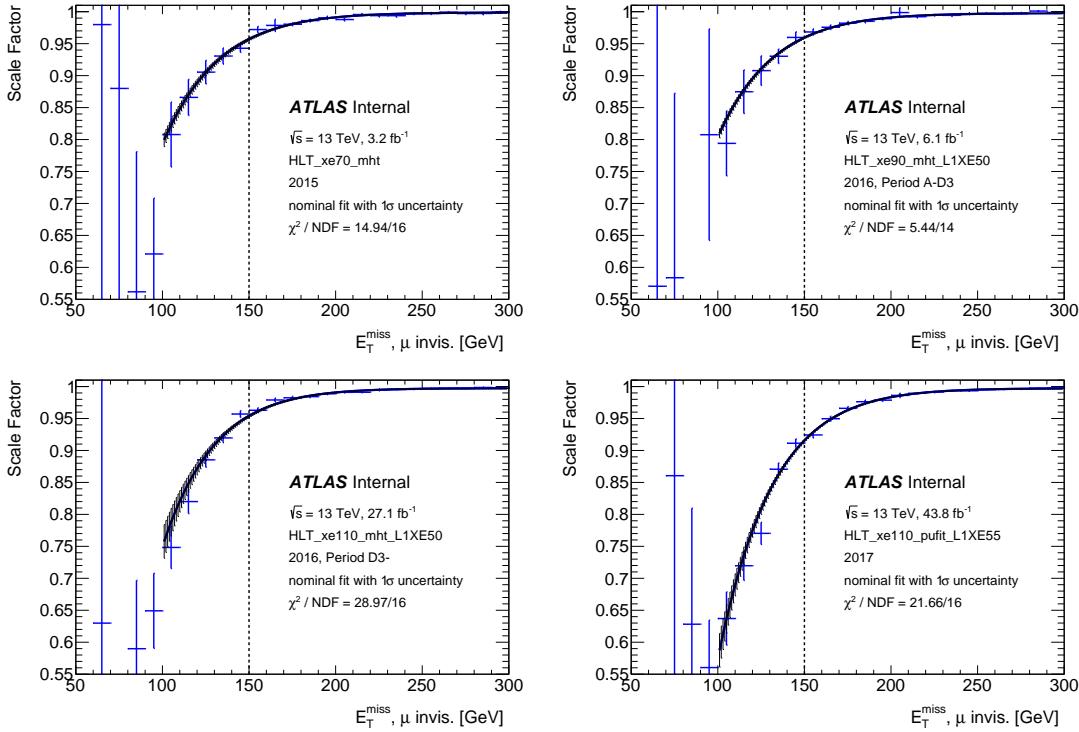


Figure 6: Measured scale factors as a function of offline E_T^{miss} , $\mu_{\text{invis.}}$ for the E_T^{miss} triggers used in 2015-2010. The scale factors were derived for 0,1 and 2 tags together. The hatched band shows the 1σ fit uncertainty.

369 2.2.3 Systematic uncertainties

370 As the nominal SF are measured inclusively in the b -jet multiplicity, the SF difference from events with
 371 b -jets is treated as a source of systematic uncertainty. It is defined as the difference between the nominal
 372 SF and the SF in events with ≥ 1 b -jets. This effectively accounts for the fact that the background
 373 composition in the signal region changes with the number of b -jets: In 0 b -tag events, $W+\text{jets}$ is the
 374 dominant background process in the signal region, while with higher number of b -jets $t\bar{t}$ production
 375 dominates. The resulting SF and their fit curves are shown in Figure 8

376 2.2.4 Discussion and comparison to previous analysis

377 Instead of the 4-parameter fit function in Equation 2, in the previous analysis iteration a 2-parameter fit
 378 function was used [1]:

Not reviewed, for internal circulation only

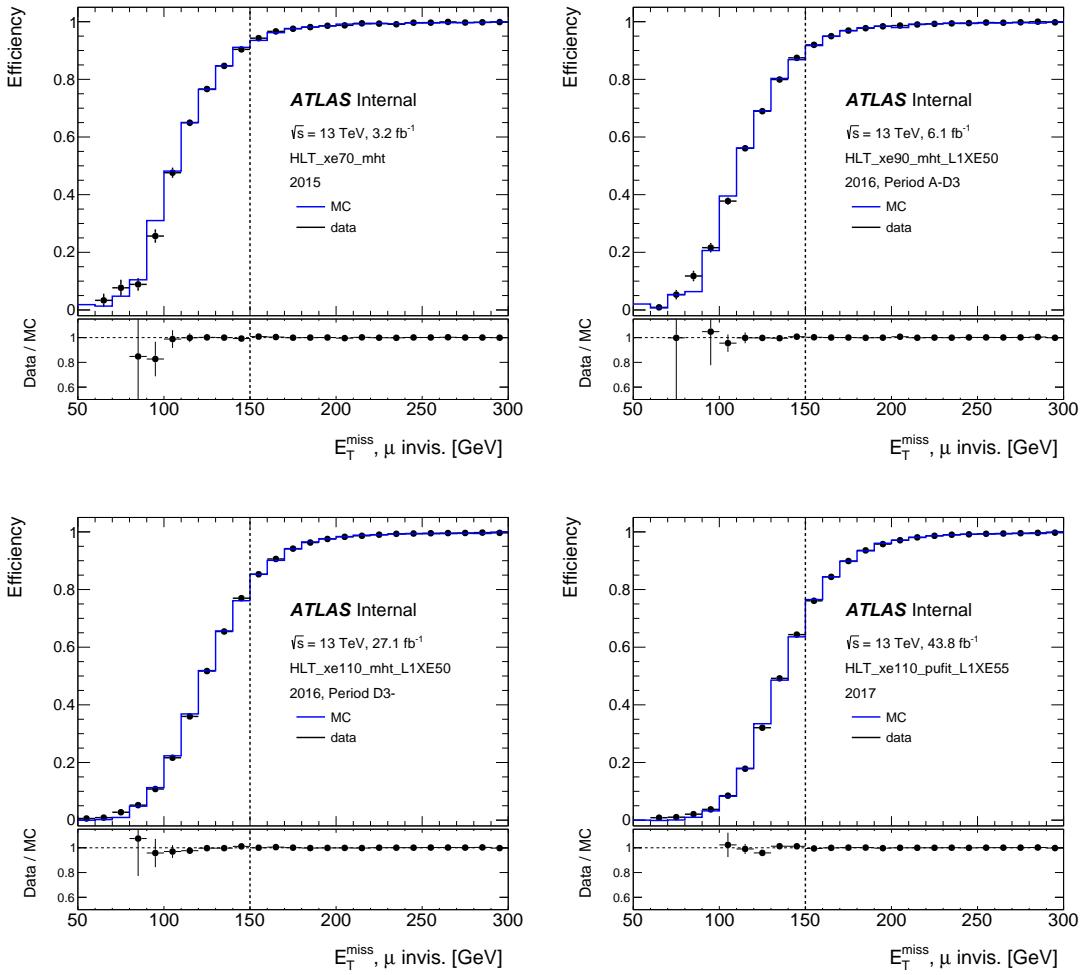


Figure 7: Validation plots showing E_T^{miss} trigger efficiencies and scale factors as function of offline $E_{T,\mu \text{ invis.}}^{\text{miss}}$ after applying scale factor corrections for $E_{T,\mu \text{ invis.}}^{\text{miss}} > 100 \text{ GeV}$ to the MC in the full single lepton region. Good agreement is observed between data and MC.

$$f(x) = 0.5 \cdot \left[1 + \text{erf} \left(\frac{x - p_0}{\sqrt{2}p_1} \right) \right] \quad (4)$$

With this function the SF curve was parametrized sufficiently well for $E_T^{\text{miss}} > 150 \text{ GeV}$, but did not match the curve well for lower E_T^{miss} values. The 4-parameter function, on the other hand, describes the SF curve well over the full E_T^{miss} range. Moreover, in Ref. [1] a E_T^{miss} binning of 2 GeV was used. For this iteration a broader binning of 10 GeV was chosen. Together with the new fit function, much better fit results are achieved. Furthermore, with the new approach it is possible to also sufficiently model the SF in events with $1 \geq b\text{-tags}$. Therefore, the nominal SF could be measured in events with $b\text{-tags}$ as well, which is closer to the event selection in the signal. This needs to be confirmed once the SF are updated for the new event selection as in Section 4.

Not reviewed, for internal circulation only

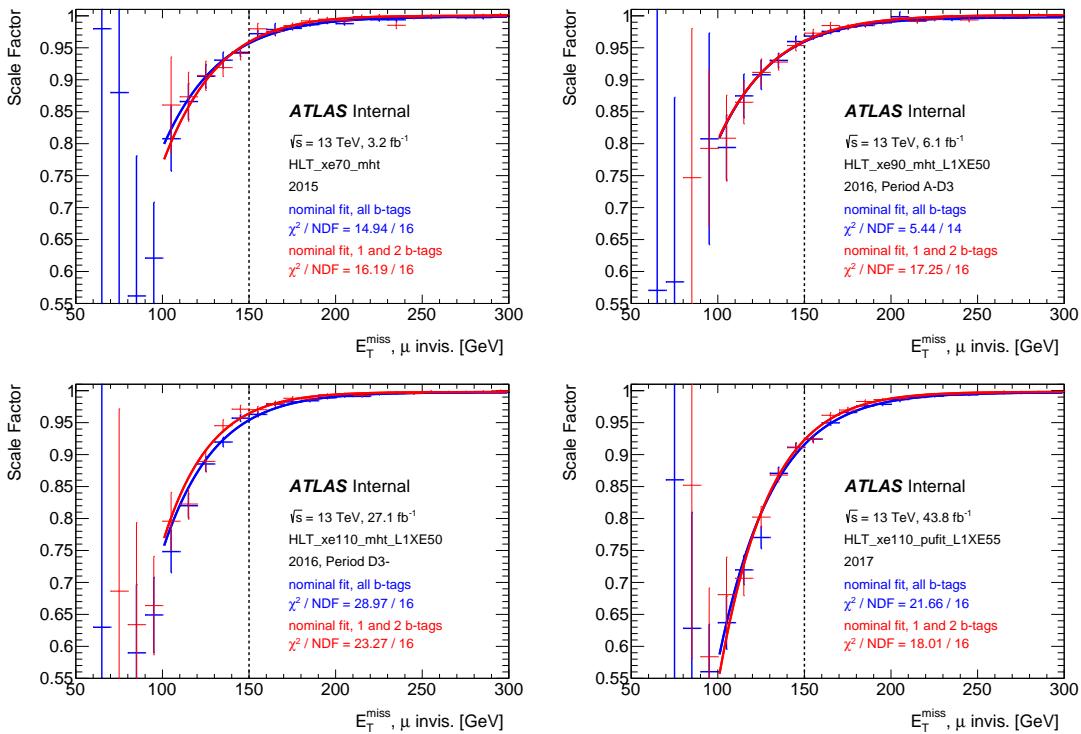


Figure 8: Measured scale factors as a function of offline $E_{T,\mu_{\text{invis.}}}^{\text{miss}}$ for the E_T^{miss} triggers used in 2015–2018 for any number of b -tags (shown in blue) as well as for 1 and 2 b -tags (shown in red). The latter one is used to estimate the systematic uncertainty on the nominal scale factors.

387 update plots with new event selection and also for 2018 trigger once new derivations are available

388 2.3 MC samples [Spyros]

389 2.3.1 2HDM+a signal

390 Simulated events corresponding to the $pp \rightarrow h\chi\bar{\chi}$ process were generated with **MADGRAPH5_AMC@NLO**
 391 (**MG5_AMC**) 2.6.1 [18] based on the UFO model developed in [15] at leading-order (LO) accuracy using
 392 the NNPDF 3.0 NLO PDF set with $\alpha_s = 0.118$ [19]. The signal samples are generated separately for the
 393 loop-induced gluon–gluon fusion (ggF) process and the b -initiated production. The former are generated
 394 using the 4-flavour scheme, thereby including bottom quarks in the loop diagrams, whereas the b -initiated
 395 process is calculated using the 5-flavour scheme. Both processes include resonant contributions where
 396 $A \rightarrow ah$, but also non-resonant contributions as well as decays like $a \rightarrow Ah$, as illustrated in Fig. 9. The
 397 phenomenology of the two processes is quite different and is studied in detail in Appendix K.

398 The samples are generated using the following parameters, following the recommendations of [15]:

- 399 • $m_A = m_H = m_{H^\pm}$
- 400 • $\lambda_1 = \lambda_2 = \lambda_3 = 3$
- 401 • $m_\chi = 10 \text{ GeV}$

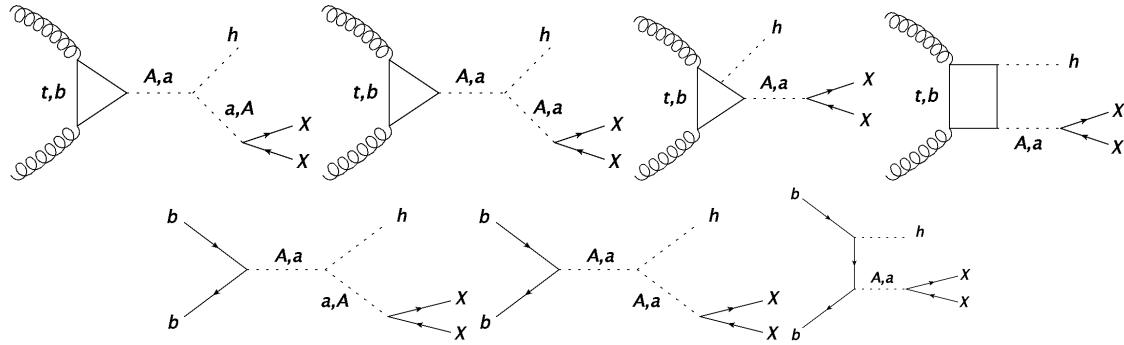


Figure 9: Representative Feynman diagrams corresponding to the simulated signal samples for the gluon–gluon fusion process (top) and the b -initiated process (bottom).

402 • $g_\chi = 1$

403 • $\sin(\beta - \alpha) = 1$ (alignment limit)

404 The samples corresponding to the ggF process are generated with $\tan \beta = 1$, while the ones corresponding
 405 to the b -initiated production are generated with $\tan \beta = 10$. For both samples the value of $\sin \theta$ is fixed to
 406 0.35, however for the interpretation plots a scan over several $\tan \beta$ and $\sin \theta$ values is performed.

407 The parton shower and hadronisation were simulated with PYTHIA 8.230 [20] using the A14 set [21] of
 408 tuned parameters (“tune”) together with the NNPDF 2.3 LO PDF set [22]. Higgs boson decays into $b\bar{b}$
 409 pairs were also simulated with PYTHIA 8.230 with a branching fraction fixed to the SM prediction.

410 2.3.2 Z' 2HDM signal

411 To be updated when we eventually have the new signal grid.

412 Simulated events corresponding to the $pp \rightarrow Z' \rightarrow hA(\chi\bar{\chi})$ process were generated with MAD-
 413 GRAPH5_AMC@NLO (MG5_AMC) 2.2.3 [18] at leading-order (LO) accuracy using the NNPDF 2.3
 414 NLO PDF set with $\alpha_s = 0.119$ [23]. The 4-flavour scheme was used for the calculation of the matrix
 415 elements. Several samples were generated for different values of $m(Z')$ and $m(A)$. The masses of the
 416 additional Higgs bosons were fixed to $m(H) = m(H^\pm) = 300$ GeV and the mass of the DM candidates was
 417 fixed to $m(\chi) = 100$ GeV. The coupling of Z' to quarks was fixed to $g_Z = 0.1$ and the ratio of the vacuum
 418 expectation value of the two Higgs doublets was fixed to $\tan \beta = 3$.

419 The parton shower and hadronisation were simulated with PYTHIA 8.186 [24] using the A14 set [21] of
 420 tuned parameters (“tune”) together with the NNPDF 2.3 LO PDF set [22]. Higgs boson decays into $b\bar{b}$
 421 pairs were also simulated with PYTHIA 8.186 with a branching fraction fixed to the SM prediction.

422 2.3.3 V+jets

423 QCD V+ jets production is simulated with the SHERPA v2.2 [25] generator. In this setup, NLO-accurate
 424 matrix elements for up to two jets, and LO-accurate matrix elements for up to four jets are calculated with
 425 the Comix [26] and OpenLoops [27, 28] libraries. They are matched with the SHERPA parton shower [29]

426 using the MEPS@NLO prescription [30–33]. Samples are generated using the NNPDF3.0nnlo set [19],
 427 along with the dedicated set of tuned parton-shower parameters developed by the SHERPA authors.

428 The samples are split according to whether they contain a B hadron or no B with $p_T > 5$ GeV and
 429 $|\eta| < 2.9$, a C hadron with $p_T > 4$ GeV and $|\eta| < 3$ (with the filtered samples called Bfilter, CFilterBVeto,
 430 CVetoBVeto respectively). They are further split by either:

- 431 • using the transverse momentum of the V boson produced by Sherpa ($p_T(V)$), or
 - 432 • using the max $p_T(V), H_T$, where H_T is the scalar sum of the p_T of the vector boson and the jets.
- 433 The two slices are combined using the procedure described in App. L.

434 2.3.4 $t\bar{t}$

435 The production of $t\bar{t}$ events is modelled using the PowHEGBox [34–37] v2 generator which provides matrix
 436 elements at next-to-leading order (NLO) in the strong coupling constant α_S with the NNPDF3.0NLO [19]
 437 parton distribution function (PDF) and the h_{damp} parameter¹ set to 1.5 m_{top} [38]. The functional form of
 438 the renormalisation and factorisation scale is set to the default scale $\sqrt{m_{\text{top}}^2 + p_T^2}$. The events are interfaced
 439 with PYTHIA8.230 [20] for the parton shower and hadronisation, using the A14 set of tuned parameters [21]
 440 and the NNPDF23LO PDF set. The decays of bottom and charm hadrons are simulated using the EvtGen
 441 v1.6.0 program [39].

442 The NLO $t\bar{t}$ inclusive production cross section is corrected to the theory prediction at next-to-next-to-
 443 leading order (NNLO) in QCD including the resummation of next-to-next-to-leading logarithmic (NNLL)
 444 soft-gluon terms calculated using Top++2.0 [40–46]. For proton–proton collisions at a centre-of-mass
 445 energy of $\sqrt{s} = 13$ TeV, this cross section corresponds to $\sigma(t\bar{t})_{\text{NNLO+NNLL}} = 832 \pm 51$ fb using a top-quark
 446 mass of $m_{\text{top}} = 172.5$ GeV. The uncertainties on the cross-section due to PDF and α_s are calculated using
 447 the PDF4LHC prescription [47] with the MSTW2008 68% CL NNLO [48, 49], CT10 NNLO [50, 51]
 448 and NNPDF2.3 5f FFN [23] PDF sets, and are added in quadrature to the scale uncertainty.

449 The events are filtered to select dilepton or semi-leptonic $t\bar{t}$ decays.

450 Two sets of samples are used in the analysis

- 451 • an inclusive $t\bar{t}$ sample
- 452 • a set of E_T^{miss} -filtered $t\bar{t}$ samples, which was not used in the previous iteration of the analysis [1],
 453 and was specifically developed to reduce the MC statistical uncertainties in the high- E_T^{miss} region.

454 The two sets of samples are combined using the procedure described in App. L.

455 Single-top tW associated production is modelled using the PowHEGBox [35–37, 52] v2 generator at NLO
 456 in QCD in the five flavour scheme with the NNPDF3.0NLO [19] parton distribution function (PDF) set.
 457 The diagram removal scheme [53] was employed to handle the interference with $t\bar{t}$ production [38]. The
 458 events are interfaced with PYTHIA8.230 [20] using the A14 tune [21] and the NNPDF23LO PDF set.

¹ The h_{damp} parameter controls the transverse momentum p_T of the first additional emission beyond the leading-order Feynman diagram in the parton shower and therefore regulates the high- p_T emission against which the $t\bar{t}$ system recoils.

459 2.3.5 Single top

460 Single-top tW associated production is modelled using the PowHEGBox [35–37, 52] v2 generator which
 461 provides matrix elements at next-to-leading order (NLO) in the strong coupling constant α_S in the five
 462 flavour scheme with the NNPDF3.0NLO [19] parton distribution function (PDF) set. The functional form
 463 of the renormalisation and factorisation scale is set to the default scale, which is equal to the top-quark
 464 mass. The diagram removal scheme [53] was employed to handle the interference with $t\bar{t}$ production [38].
 465 The events are interfaced with PYTHIA8.230 [20] using the A14 tune [21] and the NNPDF23LO PDF set.
 466 The decays of bottom and charm hadrons are simulated using the EvtGEN v1.6.0 program [39].

467 The inclusive cross section is corrected to the theory prediction calculated at NLO in QCD with NNLL
 468 soft gluon corrections [54, 55]. For proton–proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV,
 469 this cross section corresponds to $\sigma(tW)_{NLO+NNLL} = 71.7 \pm 3.8$ pb, using a top-quark mass of $m_{top} =$
 470 172.5 GeV. The uncertainty on the cross-section due to PDF is calculated using the MSTW2008 90%
 471 CL [48, 49] NNLO PDF set, and is added in quadrature to the scale uncertainty.

472 Single-top t-channel production is modelled using the PowHEGBox [35–37, 56] v2 generator which
 473 provides matrix elements at next-to-leading order (NLO) in the strong coupling constant α_S in the four
 474 flavour scheme with the NNPDF3.0NLOnf4 [19] parton distribution function (PDF) set. The functional
 475 form of the renormalisation and factorisation scale is set to $\sqrt{m_b^2 + p_{T,b}^2}$ following the recommendation
 476 of [56]. Top quarks are decayed at LO using MADSPIN [57, 58] to preserve all spin correlations. The
 477 events are interfaced with PYTHIA8.230 [20] using the A14 tune [21] and the NNPDF23LO PDF set. The
 478 decays of bottom and charm hadrons are simulated using the EvtGEN v1.6.0 program [39].

479 The inclusive cross section is corrected to the theory prediction calculated at NLO in QCD with Hathor
 480 v2.1 [54, 55]. For proton–proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, this cross section
 481 corresponds to $\sigma(t, \text{tchan})_{NLO} = 136.02^{+5.40}_{-4.57}$ pb ($\sigma(\bar{t}, \text{tchan})_{NLO} = 80.95^{+4.06}_{-3.61}$ pb) for single-top (single-
 482 anti-top) production, using a top-quark mass of $m_{top} = 172.5$ GeV. The uncertainties on the cross-section
 483 due to PDF and α_s are calculated using the PDF4LHC prescription [47] with the MSTW2008 68% CL
 484 NLO [48, 49], CT10 NLO [50] and NNPDF2.3 NLO [23] PDF sets, and are added in quadrature to the
 485 scale uncertainty.

486 Single-top s-channel production is modelled using the PowHEGBox [35–37, 59] v2 generator which
 487 provides matrix elements at next-to-leading order (NLO) in the strong coupling constant α_S in the five
 488 flavour scheme with the NNPDF3.0NLO [19] parton distribution function (PDF) set. The functional form
 489 of the renormalisation and factorisation scale is set to the default scale, which is equal to the top quark
 490 mass. The events are interfaced with PYTHIA8.230 [20] using the A14 tune [21] and the NNPDF23LO PDF
 491 set. The decays of bottom and charm hadrons are simulated using the EvtGEN v1.6.0 program [39].

492 The inclusive cross section is corrected to the theory prediction calculated at NLO in QCD with Hathor
 493 v2.1 [54, 55]. For proton–proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV, this cross section
 494 corresponds to $\sigma(t, \text{schan})_{NLO} = 6.35^{+0.23}_{-0.20}$ pb ($\sigma(\bar{t}, \text{schan})_{NLO} = 3.97^{+0.19}_{-0.17}$ pb) for single-top (single-
 495 anti-top) production, using a top-quark mass of $m_{top} = 172.5$ GeV. The uncertainties on the cross-section
 496 due to PDF and α_s are calculated using the PDF4LHC prescription [47] with the MSTW2008 68% CL
 497 NLO [48, 49], CT10 NLO [50] and NNPDF2.3 NLO [23] PDF sets, and are added in quadrature to the
 498 scale uncertainty.

499 **2.3.6 Diboson**

500 Semileptonically decaying diboson samples are simulated with the SHERPA v2.2 [25] generator. In this
 501 setup multiple matrix elements are matched and merged with the SHERPA parton shower based on Catani-
 502 Seymour dipole [26, 29] using the MEPS@NLO prescription [30–33]. The virtual QCD correction for
 503 matrix elements at NLO accuracy are provided by the OPENLOOPS library [27, 28]. library [28]. Samples
 504 are generated using the NNPDF3.0nnlo set [19], along with the dedicated set of tuned parton-shower
 505 parameters developed by the SHERPA authors.

506 Loop-induced diboson samples are simulated with the SHERPA v2.2 [25] generator. In this setup multiple
 507 matrix elements are matched and merged with the SHERPA parton shower based on Catani-Seymour
 508 dipole [26, 29] using the MEPS@NLO prescription [30–33]. The virtual QCD correction for matrix
 509 elements at NLO accuracy are provided by the OPENLOOPS library [27, 28]. Samples are generated using
 510 the NNPDF3.0nnlo set [19], along with the dedicated set of tuned parton-shower parameters developed
 511 by the SHERPA authors.

512 The loop-induced diboson samples have not been used in the previous iteration of the analysis [1].

513 **2.3.7 SM $Vh(b\bar{b})$**

514 The production of $q\bar{q} \rightarrow Wh \rightarrow \ell\nu b\bar{b}$ and $q\bar{q} \rightarrow Zh \rightarrow \ell\ell b\bar{b}$ events is modelled using the PowHEGBox
 515 v2 generator [37] using the PowHEG MiNLO procedure [60, 61] with the NNPDF3.0NLO [19] PDF set.
 516 The events are interfaced with PYTHIA8.212 [20] using the AZNLO tune [62] and the CTEQ6L1 [63]
 517 PDF set. The cross-section is calculated at NNLO (QCD) and NLO (EW) accuracy [64–70].

518 The loop-induced $gg \rightarrow Zh \rightarrow \ell b\bar{b}, \nu\bar{\nu} b\bar{b}$ process is modelled using the PowHEGBox v2 generator
 519 [37] with the NNPDF3.0NLO [19] PDF set. Parton showering and hadronisation are provided by PY-
 520 THIA8.212 [20] with the same settings as the one used for the $q\bar{q}$ process. The cross-section is calculated
 521 at NLO+NLL accuracy in QCD [71–73].

522 **2.3.8 $t\bar{t} + Z/h$**

523 The production of $t\bar{t}H$ events is modelled using the PowHEGBox [34–37, 74] generator at NLO with the
 524 NNPDF3.0NLO [19] PDF set. The events are interfaced with PYTHIA8.230 [20] using the A14 tune [21]
 525 and the NNPDF2.3LO [19] PDF set.

526 The cross sections are calculated at NLO QCD and NLO EW accuracy using MADGRAPH5_AMC@NLO
 527 and reported in Reference [75]. The predicted values at 13 TeV are 507^{+35}_{-50} fb, where the uncertainties are
 528 from variations of renormalization and factorization scales as well as α_S variations.

529 The production of $t\bar{t}V$ events is modelled using the MADGRAPH5_AMC@NLO v2.3.3 [18] generator
 530 which provides matrix elements at next-to-leading order (NLO) in the strong coupling constant α_S with
 531 the NNPDF3.0NLO [19] parton distribution function (PDF). The functional form of the renormalization
 532 and factorization scale is set to the default scale $0.5 \times \sum_i \sqrt{m_i^2 + p_{T,i}^2}$, where the sum runs over all the
 533 particles generated from the matrix element calculation. Top quarks are decayed at LO using MADSPIN [57,
 534 58] to preserve all spin correlations. The events are interfaced with PYTHIA8.210 [20] for the parton shower

and hadronisation, using the A14 set of tuned parameters [21] and the NNPDF23LO [19] PDF set. The decays of bottom and charm hadrons are simulated using the EvtGEN v1.2.0 program [39].

The cross sections are calculated at NLO QCD and NLO EW accuracies using `MADGRAPH5_AMC@NLO` and reported in Reference [75]. The $t\bar{t}Z$ cross section is further supplemented with an off-shell (down to 5 GeV) correction calculated in Reference [76]. The predicted values at 13 TeV are $0.88^{+0.09}_{-0.11}$ pb and $0.60^{+0.08}_{-0.07}$ pb for $t\bar{t}Z$ and $t\bar{t}W$ respectively, where the uncertainties are from variations of renormalization and factorization scales as well as α_S variations.

3 Object definitions [Andrea, Veronica]

This section defines the various physics objects used to construct signal and control region and describes their reconstruction. The objects in the analysis are calibrated using the Combined Performance tools of the release `AthAnalysis 21.2.85`.

3.1 Electrons

Electrons are reconstructed from an electromagnetic cluster and identified using various likelihood-based criteria, which include e.g. selections on the shower profiles, the track quality, and the presence of high-threshold TRT hits. Details about the reconstruction process can be found in [77]. The different calibration steps are described in [78].

Table 2: Electron selection criteria.

Feature	Criterion
Pseudorapidity range	$ \eta < 2.47$
Transverse momentum	$p_T > 7 \text{ GeV}$
Track to vertex association	$ d_0^{\text{BL}}(\sigma) < 5$ $ \Delta z_0^{\text{BL}} \sin \theta < 0.5 \text{ mm}$
Identification	<code>FCLoose</code>
Isolation	<code>LooseTrackOnly / FCHighPtCaloOnly</code>

3.2 Muons

The muon reconstruction is performed with different algorithms, which make use of the combined information from the Inner Detector (ID), the calorimeter and the Muon Spectrometer [79].

Muons are categorized into *baseline* and *signal* muons. For baseline objects looser selection criteria are applied than for signal objects, which leads to a larger acceptance. This is for example useful for the calculation of the missing transverse momentum E_T^{miss} . In particular, this allows for a high efficiency of the lepton veto in the signal region. Signal objects, on the other hand, have a higher purity than baseline objects. Therefore they are used to select the muons in the single-muon and dilepton control regions.

Table 3: Muon selection criteria.

Feature	Baseline criterion	Signal criterion
Selection working point	Loose	Medium
Isolation working point	FCLoose	FCTightTrackOnly
Momentum calibration	Sagitta correction [used]	Sagitta correction [used]
p_T Cut	7 GeV	7 GeV
$ \eta $ cut	< 2.7	< 2.5
d_0 significance cut	3	3
z_0 cut	0.5 mm	0.5 mm

559 3.3 τ -leptons

560 τ -leptons are reconstructed by using ID and calorimeter information. Their identification is based on
 561 Boosted Decision Trees which use as input variables e.g. the depth of the cluster shower in the calorimeter
 562 or the number of associated tracks [80]. τ -leptons are vetoed in all signal and control region and therefore
 563 the loose working point is used for their reconstruction.

Table 4: Tau selection criteria.

Feature	Criterion
Pseudorapidity range	$ \eta < 2.5$
Track selection	1 or 3 tracks
Charge	$ Q = 1$
Tau energy scale	MVA TES
Transverse momentum	$p_T > 20 \text{ GeV}$
Jet rejection	BDT-based (Loose)
Electron rejection	BDT-based
Muon rejection	Via overlap removal in $\Delta R < 0.2$ and $p_T > 2 \text{ GeV}$. Muons must not be Calo-tagged

564 3.4 Small-R jets

565 Small-R jets, i.e. jets with a radius of $R = 0.4$, are reconstructed using the anti- k_t algorithm [81] with
 566 a distance parameter of $R = 0.4$. The anti- k_t algorithm uses topological clusters from calorimeter cells
 567 as input. These are formed by the topological cell clustering algorithm described in Ref. [82]. The jets
 568 are required to pass the selections of the BadTight jet cleaning working point [83] in order to suppress
 569 the non-collisional background (NCB). More details about the jet cleaning and the NCB can be found
 570 in Appendix B. Small-R jets are calibrated according to the steps described in Ref. [84], which include
 571 corrections to the jet energy scale (JES) and resolution (JER).

572 Small-R jets are divided into *central* and *forward* jets. Central jets have $|\eta| < 2.5$. Only this jet collection
 573 is used to reconstruct the two candidate jets from the $H \rightarrow b\bar{b}$ decay. Jets from pileup interactions are
 574 removed using the jet-vertex-tagger (JVT) algorithms [85]. Forward jets are jets with $2.5 < |\eta| < 4.5$.

575 **NOTE:** We are currently studying the performance of PFlow jets, so this section may change later.

Table 5: Reconstruction criteria for central small-R jets.

Feature	Criterion
Algorithm	Anti- k_t
R-parameter	0.4
Input constituent	EMTopo
CalibArea tag	00-04-82
Calibration configuration	JES_MC16Recommendation_Consolidated_EMTopo_Apr2019_Rel21.config
Calibration sequence (Data)	JetArea_Residual_EtaJES_GSC_Insitu
Calibration sequence (MC)	JetArea_Residual_EtaJES_GSC
Selection requirements	
Observable	Requirement
Jet cleaning	TightBad
p_T	> 20 GeV (central) / > 30 GeV (forward)
$ \eta $	< 2.5 (central) / 2.5 < $ \eta $ < 4.5 (forward)
JVT	Medium working point, applied only to central jets with p_T < 120 GeV

Not reviewed, for internal circulation only

576 3.5 Large-R jets

577 Large-radius jets are used to reconstruct boosted $H \rightarrow b\bar{b}$ decays. These jets are built from topological
 578 clusters, which are calibrated to the hadronic scale using the local hadronic cell weighting (LCW)
 579 scheme [82]. They are reconstructed using the anti- k_t algorithm with a distance parameter of $R = 1.0$. To
 580 suppress pile-up effects the jets are groomed with a trimming procedure [86]. This removes regions of the
 581 large-R jet that have a small relative contribution to the total jet p_T . Finally, the groomed jets are calibrated
 582 to the JES and jet mass scale (JMS) following the techniques described in Ref. [87]. The flavor content of
 583 the large-R jet is derived through the associated variable-radius track jets, as described below.

584 3.6 Variable-radius track jets

585 The Variable Radius (VR) track jets are reconstructed from the ID tracks using the anti- k_t algorithm.
 586 These tracks need to satisfy different conditions on the number of hits in the semiconductor tracker (SCT)
 587 and the pixel detectors [88]. The resulting VR track jets are required to have $p_T > 0.5$ GeV, $|\eta| < 2.5$ and
 588 $z_0 \sin \theta < 3$ mm. They are calibrated within dileptonic $t\bar{t}$ events following the procedures described in
 589 Ref. [89, 90]

590 The main feature of the VR track jets is the p_T dependence of the jet radius [91]:

$$R(p_T) = \frac{30 \text{ GeV}}{p_T} \quad (5)$$

591 Two additional parameters $R_{\min} = 0.02$ and $R_{\max} = 0.4$ are used to set, respectively, a lower and an
 592 upper cut on the effective track jet size [88]. As the radius decreases with the track jet p_T , the VR track
 593 jets allow for a larger detection efficiency of boosted $H \rightarrow b\bar{b}$ decays compared to the previously used
 594 fixed-radius track jets. This significantly enhances the sensitivity to very boosted $H \rightarrow b\bar{b}$ signal models,

Table 6: Large- R jet reconstruction criteria.

Feature	Criterion
Algorithm	anti- k_t
R-parameter	1.0
Input constituent	LCTopo
Grooming algorithm	Trimming
Subjet p_T fraction for trimming	0.05
R_{trim}	0.2
CalibArea tag	00-04-82
Calibration configuration MC	JES_MC16recommendation_FatJet_Trimmed_JMS_comb_170ct2018.config
Calibration configuration data	JES_MC16recommendation_FatJet_Trimmed_JMS_comb_3April2019.config
Calibration sequence (Data)	EtaJES_JMS_In situ
Calibration sequence (MC)	EtaJES_JMS
Selection requirements	
Observable	Requirement
p_T	> 200 GeV
$ \eta $	< 2

595 as demonstrated in Ref. [1]. To identify the flavor content of the calibrated large- R jet the VR track are
 596 matched to the initial ungroomed large- R jet via the ghost-association method [92]-[93].

597 3.7 b -tagging

598 In the resolved region central jets are used for b -tagging. In the merged region the b -tagging is performed
 599 on the leading two associated VR track jets. Thereby the jet is assigned as having either 0, 1, or 2
 600 b -tags. For both jet collection the MV2c10 discriminant with the 77% b -tagging efficiency working point
 601 is used[94].

Table 7: b -tagging selection criteria.

Feature	Criterion
Jet collection	AntiKt4EMTopo / AntiKtVR30Rmax4Rmin02
Algorithm	MV2c10
Operating point	Eff = 77
CDI	2017-21-13TeV-MC16-CDI-2019-07-30_v1

602 3.8 Missing transverse momentum E_T^{miss}

603 The missing transverse momentum (E_T^{miss}) is defined as the magnitude of the missing transverse mo-
 604 mentum vector \vec{E}_T^{miss} , which is calculated using the fully calibrated physics objects described above [95].
 605 Furthermore, ID tracks, which are not attributed to any reconstructed objects, are included through a track

soft term (TST). No τ -leptons or photons are included in the \vec{E}_T^{miss} calculation in the analysis. Jets with $|\eta| > 2.4$ are required to have $p_T > 30 \text{ GeV}$, corresponding to the criteria of the tight E_T^{miss} working point.

Table 8: E_T^{miss} reconstruction criteria.

Parameter	Value
Algorithm	Calo-based
Soft term	Track-based (TST)
MET operating point	Tight

3.9 Object-based E_T^{miss} significance \mathcal{S}

The objects-based E_T^{miss} significance (\mathcal{S}) is used to discriminate events with genuine E_T^{miss} , coming by neutrinos or possible exotics particles, from these events, in which E_T^{miss} is caused by measurement deficiencies and resolution effects. Therefore it is a powerful variable to suppress the QCD-background, as was shown in Ref. [1]. \mathcal{S} takes into account the expected resolutions of all the objects that enter the E_T^{miss} reconstruction and their directional correlations.

\mathcal{S} is defined by:

$$\mathcal{S} = \frac{\left| \overrightarrow{E}_T^{\text{miss}} \right|^2}{\sigma_L^2 (1 - \rho_{LT}^2)}. \quad (6)$$

where σ_L^2 , σ_T^2 are the total variances in the longitudinal and transverse directions to \vec{E}_T^{miss} respectively, and ρ_{LT} is the correlation factor of the longitudinal L and transverse T measurements [96].

3.10 Overlap removal

The reconstruction of the same energy deposits as multiple objects is resolved using the standard overlap removal tools, `AssociationUtils`, documented [here](#)

ΔR is calculated using rapidity by default.

4 Event selection [Andrea, Anindya, Jay]

This section presents the event selection cuts used to define signal and control regions, based on the object definitions provided in the previous section. A summary of the signal region selection is provided in Tab. 9 and further details are provided in the following.

Reject	Against	Criteria
Electron	Electron	shared track, $p_{T,1} < p_{T,2}$
Tau	Electron	$\Delta R < 0.2$
Tau	Muon	$\Delta R < 0.2$
Muon	Electron	is Calo-Muon and shared ID track
Electron	Muon	shared ID track
Jet	Electron	$[\Delta R < 0.2 / \text{Not a } b\text{-jet and } \Delta R < 0.2]$
Electron	Jet	$[\Delta R < 0.4 / \Delta R < \min(0.4, 0.04 + 10 \text{ GeV}/p_T(e))/\text{None}]$
Jet	Muon	$[\text{NumTrack} < 3 \text{ and (ghost-associated or } \Delta R < 0.2) / \text{not a } b\text{-jet and NumTrack} < 3 \text{ and (ghost-associated or } \Delta R < 0.2)]$
Muon	Jet	$[\Delta R < 0.4 / \Delta R < \min(0.4, 0.04 + 10 \text{ GeV}/p_T(\mu))/\text{None}]$
Jet	Tau	$\Delta R < 0.2$
Fat-jet	Electron	$\Delta R < 1.0$

626 4.1 Event cleaning

627 All events are subjected to the standard ATLAS event cleaning procedure, following the [recommendations](#)
 628 of the DataPrep group.

629 Data events must be part of the runs listed in the GRLs as described in Section 2. A veto is applied to the
 630 following bad or corrupted events:

- 631 • LAr noise burst and data corruption (`xAOD::EventInfo::LAr`),
- 632 • Tile corrupted events (`xAOD::EventInfo::Tile`),
- 633 • events affected by the SCT recovery procedure for single event upsets (`xAOD::EventInfo::SCT`),
- 634 • incomplete events (`xAOD::EventInfo::Core`).

635 Events are required to have a primary vertex with at least two associated tracks. The primary vertex is
 636 selected as the one with the largest $\sum p_T^2$, where the sum is over all tracks with transverse momentum
 637 $p_T > 0.4 \text{ GeV}$ that are associated with the vertex.

638 Debug stream events have not been included. Checks have been done to remove duplicate events.

639 4.2 Signal regions

640 4.2.1 Common event selection

641 There are a number of basic requirements applied in the event selection of the resolved as well as the
 642 merged signal region:

643 **Loose/Bad jet veto** Veto events if any jet fails the `TIGHTBADJET` jet cleaning requirement to remove
 644 non-collisional background.

645 **E_T^{miss} proxy > 150GeV** In the signal region a cut of E_T^{miss} proxy > 150GeV is applied. In the control
 646 regions $E_{\text{T, lep. invis.}}^{\text{miss}}$ is used as a proxy for this cut. In the 1-lepton channel his is the E_T^{miss} value
 647 obtained when the muon contribution is not included in the E_T^{miss} calculation, while in the 2-lepton
 648 channel the two signal electrons or muons are not included in the calculation of E_T^{miss} .

649 **Light lepton and tau veto** Any events which contain a baseline electron, baseline muon or a τ -lepton,
 650 as defined in Section 3, are rejected.

651 **Extended tau veto** To remove events, in which the τ -leptons failed to be identified, an *extended τ veto*
 652 is applied. This variable further reduces the $W+\text{jets}$, single-top and $t\bar{t}$ background. Its calculation
 653 uses the reconstructed jets as input (small-R jets in the resolved channel and large-R jets in the
 654 merged channel). A jet is considered as a tau candidate if it fulfills two conditions: (1) The track
 655 multiplicity in the jet cone is no lower than 1 nor larger than 4. (2) The angular separation between
 656 the jet and E_T^{miss} is $\Delta\phi \leq 22.5^\circ$. The cut on the track multiplicity in the jet cone makes sure that the
 657 hadronic decay products in the jet are compatible with the τ -lepton decay products, namely with
 658 charged pions. The tracks considered are associated with the primary vertex and have a p_T greater
 659 than 1 GeV. The cut on the angular separation between jet and E_T^{miss} makes sure that the τ -candidate
 660 comes from a W -boson.

661 **Minimum $\Delta\phi(E_T^{\text{miss}}, \text{Central + Forward jets}_{1,2,3}) > 20^\circ$** After the baseline selections, there is still
 662 a large number of multijet background events where sizable E_T^{miss} is caused either by poor jet energy
 663 measurements causing an energy imbalance or by semileptonic hadron decays. To reduce this
 664 background, a cut on the variable *Minimum $\Delta\phi(E_T^{\text{miss}}, \text{Central + Forward jets}_{1,2,3})$* is introduced.
 665 To construct this variable, a set of jets, referred to as *Central+Forward* jets is used. This set of
 666 jets is formed starting with the *Central* jets, described in Section 3.4, followed by the *Forward* jets.
 667 Within each category, the jets are ordered in decreasing transverse momentum. Using this set of
 668 jets, the anti-QCD selection is: *Minimum $\Delta\phi(E_T^{\text{miss}}, \text{Central + Forward jets}_{1,2,3}) > 20^\circ$* .
 669 In the case of a multijet event in which one jet is mis-measured, the E_T^{miss} will generally point in the
 670 direction of the jet. Therefore, it will azimuthally point in the same direction as the opposing jet
 671 and so multijet events are expected to reside at low values of this observable. Only the first three
 672 jets of the full set of *Central+Forward* jets (up to three total jets) are considered in this calculation.

673 Following these baseline selections, events are divided into the resolved and merged regions using a single
 674 selection on E_T^{miss} . For events with $E_T^{\text{miss}} < 500$ GeV, events are considered to be in the resolved regime,
 675 while for $E_T^{\text{miss}} > 500$ GeV, events are considered to be in the merged regime.

676 Only events with at least one b -tag are considered in the event selection. The b -tagging strategy depends
 677 on the jet topology being resolved or merged. In case of the resolved regime all small-R central jets (at
 678 least 2 small-R jets are required in the event) are used for b -tagging. In case of the merged analysis only
 679 the two leading track-jets ghost associated to the leading large-R jet of each event is used (see Section 3
 680 and [92]).

681 4.2.2 Resolved Region

682 The two jets forming the Higgs boson candidate are reconstructed using the *b-tagged-light* jet collection:
 683 From the set of jets described in Section 3, central jets that are b -tagged are considered first, with central
 684 non- b -tagged jets considered next. In all cases, the jets within each of the categories are ordered in
 685 decreasing transverse momentum. The first two out of this set of jets (referred to as j_1 and j_2 below) are

686 used to form the reconstructed Higgs candidate, referred to as H_{reco} . Therefore, H_{reco} may be a bb or bj
 687 candidate in the case that it is composed of two or one b -tagged jets, respectively.

688 H_{reco} is required to have a p_{T} of at least 100 GeV, where $p_{\text{T}}(H_{\text{reco}})$ is defined as the vectorial sum of the
 689 p_{T} vectors of j_1 and j_2 . For a signal process, $p_{\text{T}}(H_{\text{reco}})$ is expected to increase with $E_{\text{T}}^{\text{miss}}$, i.e. with the
 690 p_{T} of the DM pair. The signal sensitivity can therefore be enhanced by increasing the $p_{\text{T}}(H_{\text{reco}})$ threshold
 691 for higher $E_{\text{T}}^{\text{miss}}$. Hence for the third $E_{\text{T}}^{\text{miss}}$ bin in the analysis with $350 \text{ GeV} < E_{\text{T}}^{\text{miss}} < 500 \text{ GeV}$ a cut of
 692 $p_{\text{T}}(H_{\text{reco}}) > 300 \text{ GeV}$ is used.

693 In the resolved region dominant background process is $t\bar{t}$ production. To reduce this background, the
 694 variables $m_T^{b,\min}$ and $m_T^{b,\max}$ are used. These are defined by:

$$m_T^{b,\min/\max} = \sqrt{2p_T^{b,\min/\max} E_{\text{T}}^{\text{miss}} (1 - \cos \Delta\phi(b_{\min/\max}, E_{\text{T}}^{\text{miss}}))} \quad (7)$$

695 where b_{\min} and b_{\max} are defined as the b -jet, which is closest to $\vec{E}_{\text{T}}^{\text{miss}}$. More details on these variables
 696 can be found in Appendix A.

697 In the signal region resolved event selection a requirement based on the object-based $E_{\text{T}}^{\text{miss}}$ significance,
 698 $S > 12$ is added to the event selection in order to suppress multijet background processes.

699 As described in Appendix A, different signal models predict on average a different number of central,
 700 small-R jets. For the Z'-2HDM model a tight cut on the maximum allowed number of jets in the event
 701 would lead to a significant increase in sensitivity. On the other hand, this would reduce the sensitivity
 702 for the ggF induced signal models. To maximize the sensitivity for a large variety of signal models, the
 703 signal region can be further binned into different jet multiplicities. The studies in Appendix A indicate
 704 that this is a promising way to achieve an overall good sensitivity, though it needs to be tested within the
 705 fit setup, which binning choice is possible once the full statistical and systematic uncertainties have been
 706 included.

707 The final discriminant variable in the fit is the invariant mass of the two Higgs candidate jets m_{jj} . In the
 708 previous iteration of the analysis $50 \text{ GeV} < m_{jj} < 280 \text{ GeV}$ was used as fit range. However, also this will
 709 be re-evaluated within the fit setup.

710 4.2.3 Merged Region

711 The events comprising the merged region are selected by applying the common event selections and
 712 increasing the $E_{\text{T}}^{\text{miss}}$ cut to 500 GeV. This selection ensures that resolved and merged event sets are
 713 disjoint. Furthermore, the contamination of multijet events in this merged regime is negligible. In
 714 addition, at least one large-R jet is required with at least two ghost associated variable-radius track-jets.
 715 The remaining set of events is divided into exclusive categories based on the number of b -tagged track
 716 jets. In this case, because of the Higgs tagging prescription, the division is not identical to the resolved
 717 region. Instead, only the two highest p_{T} track jets associated to the large-R jet are examined to assign the
 718 number of b -tags to the Higgs boson candidate. Apart from that the event is allowed to have additional
 719 b -tagged track jets outside the large-R jet. This particularly increases the sensitivity to the $b\bar{b}$ -induced
 720 2HDM+a signal models.

721 In the merged region the final discriminant variable in the fit is the mass of the two Higgs candidate jet
722 m_J . As in the resolved region, the optimal fit range needs to be studied within the fit setup. In the previous
723 analysis iteration the range $50 \text{ GeV} < m_J < 270 \text{ GeV}$ was used.

724 The selections of the resolved and merged region are summarized in Tab. 9.

Resolved	Merged
lowest unprescaled E_T^{miss} trigger	
veto on baseline light leptons and τ -leptons	
extended τ -veto	
$E_T^{\text{miss}} > 150 \text{ GeV}$	
$\min \Delta\Phi(E_T^{\text{miss}}, \text{Central + Forward jets}_{1,2,3}) > 20^\circ$	
$E_T^{\text{miss}} < 500 \text{ GeV}$	$E_T^{\text{miss}} > 500 \text{ GeV}$
$N(\text{central small-R jets}) \geq 2$	$N(\text{central large-R jets}) \geq 1$
$N(b\text{-tagged small-R jets}) \geq 1$	$N(b\text{-tagged associated track jets}) \geq 1$
$S > 12$	—
$p_T(jj) > 100 \text{ GeV}$ if $E_T^{\text{miss}} < 350 \text{ GeV}$	—
$p_T(jj) > 300 \text{ GeV}$ if $E_T^{\text{miss}} \geq 350 \text{ GeV}$	—
$m_T^{b,\min} > 170 \text{ GeV}$	—
$m_T^{b,\max} > 200 \text{ GeV}$	—
$50 \text{ GeV} < m_{bb} < 280 \text{ GeV}$	$50 \text{ GeV} < m_J < 270 \text{ GeV}$
b -tag requirement on small-R jets	track jets

Table 9: Summary of the resolved and merged event selection applied in the 0-lepton channel. S denotes the object based E_T^{miss} significance.

725 4.3 Control regions

726 Control regions are defined to constrain various background contributions. These control regions are
727 orthogonal to the signal regions and target $W/Z+\text{jets}$ and $t\bar{t}$ processes which constitute the dominant
728 backgrounds in the signal regions. They are formed by selecting disjoint events based on the number of
729 leptons.

730 4.3.1 1-lepton channel

731 In the signal region the $W+\text{jets}$ background is mostly composed of events in which the W -boson decays
732 leptonically. The $t\bar{t}$ background in the signal region originates almost only from semileptonic $t\bar{t}$ decays.

Not reviewed, for internal circulation only

733 Events with leptons can enter the signal region if the leptons fail to be identified or if they are outside
 734 the detector acceptance. The largest contribution comes from decays involving τ -leptons, as shown in
 735 Figure 10. Figure 11 shows the distribution of different kinematic variables for the dominant decay modes.
 736 From these it can be seen that shapes of the event variables agree well within the statistical uncertainties.
 737 Therefore the $W+jets$ and $t\bar{t}$ backgrounds can be estimated using a 1-lepton control region, in which the
 738 lepton could have any flavor in principle.

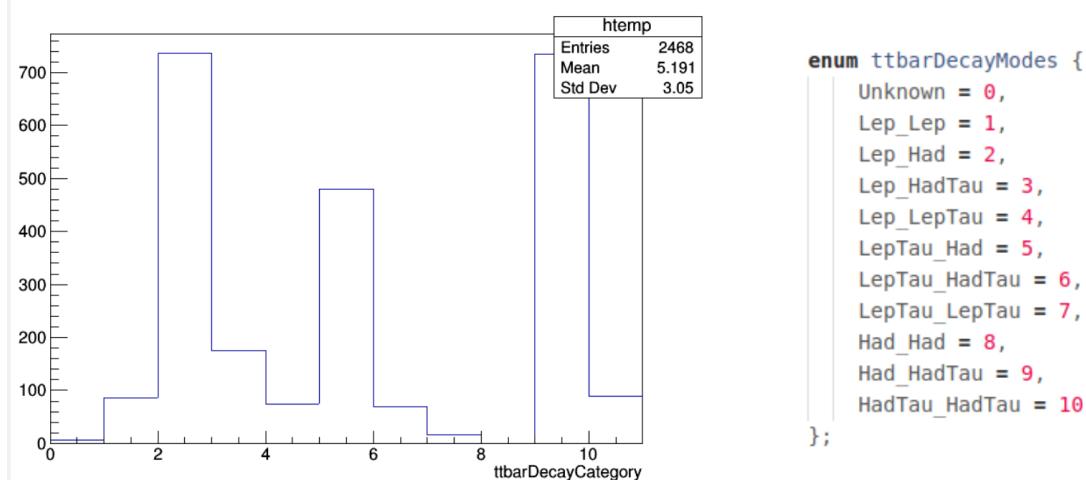


Figure 10: Decay categories for $t\bar{t}$ events in the signal region. The classification of the decay processes was performed by examining the truth particle information within the reconstructed signal region events. The three dominant decay modes are the semileptonic decays with either a light lepton, a leptonically or a hadronically decaying τ -lepton. (This is just a preliminary plot - will be replaced by a better looking one.)

739 A 1-muon control region is chosen to constrain these backgrounds. Using muons instead of electrons or
 740 τ -leptons has the advantage that the signal region kinematics can be imitated very well, as single-muon
 741 events can be selected using the same E_T^{miss} triggers² as for the signal region. Events in this control region
 742 are required to have exactly one signal muon and no additional baseline leptons. To mimic events in the
 743 signal region, in which the muon is not detected, $E_{T,\text{lep,invis}}^{\text{miss}}$ is used in the 1-muon region instead of E_T^{miss} .
 744 Also all other E_T^{miss} dependent variables, like the E_T^{miss} significance and $m_T^{b,\text{min/max}}$ are calculated with
 745 $E_{T,\text{lep,invis}}^{\text{miss}}$. Apart from that the same event selection as in the signal region is used, as described above.

746 4.3.2 2-lepton channel

747 A 2-lepton control region is used to estimate the $Z+jets$ background. In the signal region the $Z(\rightarrow \nu\nu) +$
 748 jets production leads to a significant amount of background events, which have the same decay topology
 749 as $Z(\rightarrow \ell\ell) + \text{jets}$, because the momentum of the Z boson does not depend on its decay products. Hence
 750 the normalisation of $Z(\rightarrow \nu\nu) + \text{jets}$ events can be estimated with the help of a $Z(\rightarrow \ell\ell) + \text{jets}$ control
 751 region.

² Muons passing through the inner detector deposit little energy in the calorimeters. This leads to the muon's momentum contributing to the trigger level calculation of E_T^{miss} , which is entirely calorimeter based. Therefore, it is possible to trigger on these set of events containing one muon with a E_T^{miss} trigger.

Not reviewed, for internal circulation only

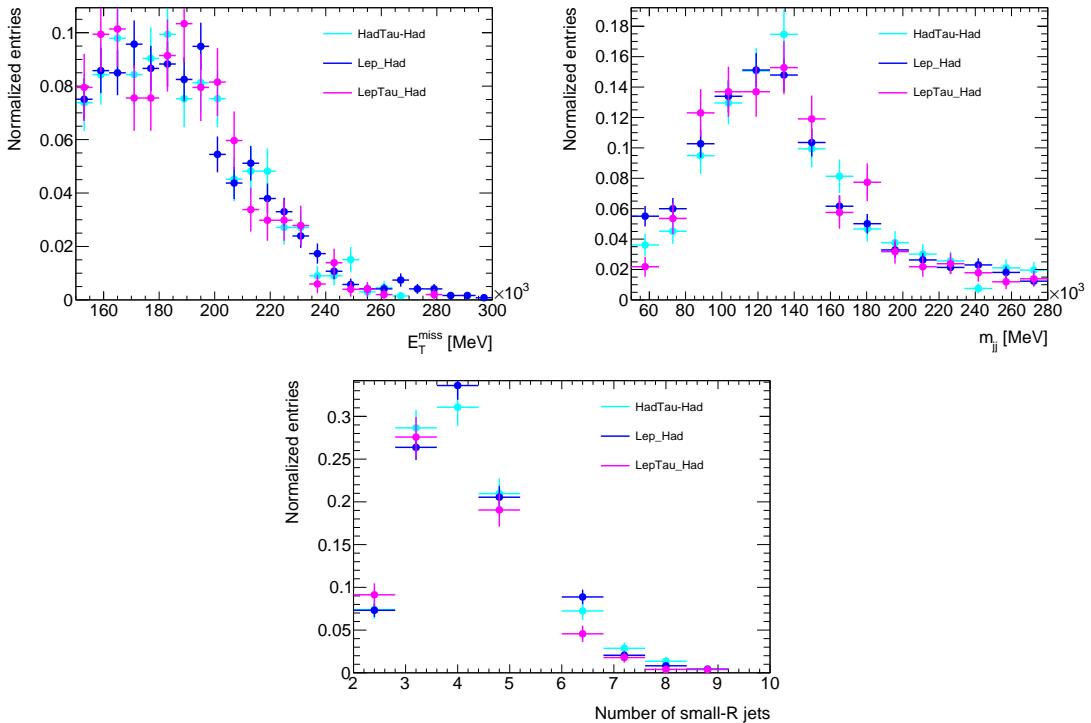


Figure 11: Normalised distributions of E_T^{miss} , the Higgs boson candidate mass and the number of small-R jets for $t\bar{t}$ events. The distributions are shown separately for the three semileptonic decay modes, which are the dominant decays modes in the signal region. The shapes of the distributions agree well within the statistical uncertainties.

Events are triggered with the lowest unprescaled single lepton triggers, see Tab. 1, and by requiring exactly 2 electrons or 2 muons. Events with further leptons are discarded. At least one of the two leptons must satisfy the signal lepton requirements and have a sufficiently high p_T to fire the trigger. The signal muons are required to have $p_T > 25 \text{ GeV}$, while signal electrons need to have $p_T > 27 \text{ GeV}$. To increase the acceptance of $Z + \text{jets}$ events the second lepton does not need to be a signal lepton. The selected leptons are required to be of opposite charge and to have an invariant mass fulfilling $|m_Z - m_{\ell\ell}| < 10 \text{ GeV}$. This suppresses backgrounds without a resonant lepton-pair around the Z -boson mass, such as $t\bar{t}$ events.

As described in Section 3, the E_T^{miss} significance \mathcal{S} is a powerful variable to distinguish events with genuine E_T^{miss} from events with "fake" E_T^{miss} . In $Z(\rightarrow \nu\nu) + \text{jets}$ events no real E_T^{miss} is expected, while for $t\bar{t}$ events real E_T^{miss} is present in the event due to the two escaping neutrinos. Therefore, to further reduce the $t\bar{t}$ background a cut of $\mathcal{S} < 5$ is applied.

For the remaining event selection the same cuts as in the signal region are applied, with the only difference that in the cuts on E_T^{miss} and all E_T^{miss} dependent variables $E_{T, \text{lep, invis.}}^{\text{miss}}$ is used. Thereby the kinematics of $Z(\rightarrow \nu\nu) + \text{jets}$ events can be imitated. In the previous iteration of the analysis $p_T(\ell\ell)$ was used as E_T^{miss} proxy. To be kinematically more similar to the signal region, the $p_T(\ell\ell)$ binning was replaced by a binning in $E_{T, \text{lep, invis.}}^{\text{miss}}$. More details on this and additional information on the construction of the control regions can be found in Appendix A.

769 5 Systematics [Spyros, Andrea, Eleni, Veronica]

770 Systematic uncertainties arise from the reconstruction of the various physics objects and from theoretical
 771 uncertainties affecting the predictions for both the backgrounds and signals. These uncertainties manifest
 772 themselves as uncertainties both in the overall yield and shape of the final observable m_h .

773 5.1 Experimental systematics [Andrea]

774 A summary of the experimental systematic uncertainties taken into account in this analysis is given in
 775 Table 10 and Table 11, along with the shorthand name of the uncertainty used throughout the analysis.
 776 Table 10 describes the uncertainties related to the luminosity, leptons, E_T^{miss} , triggers and pile-up effects
 777 and Table 11 summarizes the jet and b -tagging uncertainties. In the following, a brief description of the
 778 source and estimation of the uncertainties is given.

779 The E_T^{miss} trigger response in simulations is subject to uncertainties which are estimated in the context of
 780 the dedicated calibration discussed in Section 2.2. Tools provided by the combined performance groups
 781 are used to estimate the systematic uncertainties for electrons, muons, jets, E_T^{miss} and the lepton triggers.

782 The CategoryReduction_SimpleJER scheme is used to derive uncertainties affecting the jet energy scale
 783 (JES) of and resolution (JER) small-R jets. This is the recommended set of systematic uncertainties which
 784 is recommended by the CDM group and allows for combinations with other Dark Matter related searches.
 785 It contains 29 nuisance parameters connected to the JES and 8 nuisance parameters for the JER. b -tagging
 786 uncertainties are represented by eigenvector variations using the medium reduction scheme, leading to 4,
 787 3, and 6 eigenvector variation nuisance parameters for b, c and light-flavour jets, respectively.

Systematic uncertainty	Short description
Event	
Luminosity	uncertainty on the total integrated luminosity
Electrons	
EL_EFF_Trigger_TOTAL_1NPCOR_PLUS_UNCOR	trigger efficiency uncertainty
EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	reconstruction efficiency uncertainty
EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR	ID efficiency uncertainty
EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR	isolation efficiency uncertainty
EG_SCALE_ALL	energy scale uncertainty
EG_RESOLUTION_ALL	energy resolution uncertainty
Muons	
mu20_iloose_L1MU15_OR_HLT_mu40_MUON_EFF_Trig	trigger efficiency uncertainties (2 muon selection)
mu24_ivarmed_OR_HLT_mu40_MU_EFF_TrigStat	
mu24_ivarmed_OR_HLT_mu50_MU_EFF_TrigStat	
mu26_ivarmed_OR_HLT_mu50_MU_EFF_TrigStat	
MUON_EFF_RECO_STAT	reconstruction uncertainty for $p_T > 15$ GeV
MUON_EFF_RECO_SYS	
MUON_EFF_RECO_STAT_LOWPT	reconstruction and ID efficiency uncertainty for $p_T < 15$ GeV
MUON_EFF_RECO_SYS_LOWPT	
MUON_ISO_STAT	isolation efficiency uncertainty
MUON_ISO_SYS	
MUON_TTVA_STAT	track-to-vertex association efficiency uncertainty
MUON_TTVA_SYS	
MUONS SCALE	energy scale uncertainty
MUONS_SAGITTA_RHO	variations in the scale of the momentum (charge dependent)
MUONS_SAGITTA_RESBIAS	variations in the scale of the momentum (charge dependent)
MUONS_ID	energy resolution uncertainty from inner detector
MUONS_MS	energy resolution uncertainty from muon system
E_T^{miss}-Trigger and E_T^{miss}-Terms	
METTrigStat	trigger efficiency uncertainty
METTrigSyst	
MET_SoftTrk_ResoPerp	track-based soft term related to transversal resolution uncertainty
MET_SoftTrk_ResoPara	track-based soft term related to longitudinal resolution uncertainty
MET_SoftTrk_Scale	track-based soft term related to longitudinal scale uncertainty
MET_JetTrk_Scale	track MET scale uncertainty due to tracks in jets
PRW_DATASF	uncertainty on data SF used for the computation of pileup reweighting

Table 10: Qualitative summary of the experimental systematic uncertainties considered in this analysis.

Systematic uncertainty	Short description
Small-R Jets	
JET_EtaIntercalibration_Modelling	η -intercalibration: MC generator modelling uncertainty
JET_EtaIntercalibration_TotalStat	η -intercalibration: statistical uncertainty
JET_EtaIntercalibration_NonClosure_highE	η -intercalibration: non-closure uncertainty of jet response, high energy component
JET_EtaIntercalibration_NonClosure_negEta	η -intercalibration: non-closure uncertainty of jet response, negative η component
JET_EtaIntercalibration_NonClosure_posEta	η -intercalibration: non-closure uncertainty of jet response, positive η component
JET_Pileup_OffsetMu	Pileup: Offset, term for number of interactions per crossing μ
JET_Pileup_OffsetNPV	Pileup: Offset, term for number of primary vertices
JET_Pileup_PtTerm	Pileup: Offset, p_T term
JET_Pileup_RhoTopology	Pileup: Offset, ρ topology uncertainty on jet areas
JET_Flavor_Composition	Flavor composition uncertainty
JET_Flavor_Response	Flavor response uncertainty (dominated by gluon response)
JET_PunchThrough_MC16	Punch-through correction uncertainty
JET_EffectiveNP_Statistical	Statistical components of effective jet energy scale uncertainties; split into 6 components
JET_EffectiveNP_Modelling	Modelling components of effective jet energy scale uncertainties; split into 4 components
JET_EffectiveNP_Detector	Detector components of effective jet energy scale uncertainties; split into 2 components
JET_EffectiveNP_Mixed	Effective jet energy scale uncertainties coming from various sources; split into 3 components
JET_SingleParticle_HighPt	Uncertainty related to high p_T jets
JET_RelativeNonClosure_MC16	Closure of the calibration, relative to MC12a
JET_BJES_Response	Jet energy scale uncertainty for b -jets
JET_JER_DataVsMC_MC16	Nuisance parameter covering when jet energy resolution in data smaller than resolution in MC
JET_JER_EffectiveNP	Effective jet energy resolution uncertainty; split into 6 components
FT_EFF_EIGEN_B	b -tagging efficiency uncertainties ("BTAG_MEDIUM):
FT_EFF_EIGEN_C	
FT_EFF_EIGEN_L	
FT_EFF_EIGEN_extrapolation	b -tagging efficiency uncertainty on the extrapolation on high p_T -jets
FT_EFF_EIGEN_extrapolation_from_charm	b -tagging efficiency uncertainty on τ -jets
Large-R Jets	
JET_EtaIntercalibration_Modelling	η -intercalibration: MC generator modelling and method uncertainty
JET_EtaIntercalibration_R10_TotalStat	η -intercalibration: statistical uncertainty
JET_Flavor_Composition	Flavor composition uncertainty
JET_Flavor_Response	Flavor response uncertainty (dominated by gluon response)
JET_EffectiveNP_R10_Statistical	Statistical components of effective jet energy scale uncertainties; split into 6 components
JET_EffectiveNP_R10_Modelling	Modelling components of effective jet energy scale uncertainties; split into 4 components
JET_EffectiveNP_R10_Detector	Detector components of effective jet energy scale uncertainties; split into 2 components
JET_EffectiveNP_R10_Mixed	Effective jet energy scale uncertainties coming from various sources; split into 3 components
JET_SingleParticle_HighPt	Uncertainty related to high p_T jets (for R=0.4)
JET_CombMass_Baseline	Baseline uncertainty of the jet mass scale accounting for data-MC differences
JET_CombMass_Modelling	Modelling uncertainty of the jet mass scale accounting for different MC generators
JET_CombMass_Tracking	Uncertainty of the jet mass scale accounting for tracking variations; 3 variations in total
Track-Jets	
FT_EFF_EIGEN_B	b -tagging efficiency uncertainties ("BTAG_MEDIUM):
FT_EFF_EIGEN_C	
FT_EFF_EIGEN_L	
FT_EFF_EIGEN_extrapolation	b -tagging efficiency uncertainty on the extrapolation on high p_T -jets
FT_EFF_EIGEN_extrapolation_from_charm	b -tagging efficiency uncertainty on τ -jets

Table 11: Qualitative summary of the experimental systematic uncertainties considered in this analysis.

⁷⁸⁸ **5.2 Theory systematics [Spyros, Eleni, Veronica]**

⁷⁸⁹ **5.2.1 Theoretical uncertainty sources**

⁷⁹⁰ Additional systematic uncertainties arising from the MC modelling of the background and signal processes
⁷⁹¹ are also considered in the analysis. These can arise from the following sources:

- ⁷⁹² • **missing higher orders in the calculation of the inclusive matrix elements:** for all processes the
⁷⁹³ calculation of the cross-section relies on a perturbative expansion of the scattering matrix, which
⁷⁹⁴ is truncated at a certain order. The effect of the missing higher orders is estimated by varying the
⁷⁹⁵ renormalisation and factorisation scales (μ_R and μ_F) independently by a factor of 2, excluding the
⁷⁹⁶ $(\mu_R, \mu_F) = (\frac{1}{2}, 2), (2, \frac{1}{2}) \times \mu_{\text{central}}$ variations, which may lead to large logarithms.
- ⁷⁹⁷ • **Uncertainties from the choice of PDFs and α_s :** which in turn arise from uncertainties in the
⁷⁹⁸ experimental measurements that are used to determine the PDF sets used in each calculation,
⁷⁹⁹ uncertainties from the choice of the functional form used in the PDF fits, and uncertainties associated
⁸⁰⁰ to the experimental determination of α_s . These are estimated using the PDF4LHC prescription [97]
- ⁸⁰¹ • **merging-scale uncertainties:** for samples generated by merging matrix elements (ME) corres-
⁸⁰² ponding to different multiplicities, e.g. $V+1,2,\dots$ jets, an uncertainty related to the choice of the
⁸⁰³ merging scale, i.e. the scale that separates soft from hard jets, is evaluated by varying the merging
⁸⁰⁴ scale by a factor of 2 up and down
- ⁸⁰⁵ • **resummation scale uncertainties:** for SHERPA samples an additional uncertainty related to the
⁸⁰⁶ energy cut-off for the integration of MC counterterms in the parton shower (PS) [98].
- ⁸⁰⁷ • **matching uncertainties:** for samples generated using a NLO matrix element and matched to a
⁸⁰⁸ parton shower, a comparison between a Powheg and a MG5aMC sample can probe uncertainties
⁸⁰⁹ related to the ME/PS matching procedure
- ⁸¹⁰ • **parton shower/hadronisation uncertainties:** uncertainties related to algorithmic or parametric
⁸¹¹ differences in the modelling of the PS and hadronisation can be assessed by comparing samples
⁸¹² generated with different showering/hadronisation (SHG) generators, typically PYTHIA 8 with HER-
⁸¹³ WIG 7.
- ⁸¹⁴ • **eigentune uncertainties:** are related to uncertainties in the choice of the free parameters that are
⁸¹⁵ used in the SHG programmes, derived so as to encompass the data used in the ATLAS tuning
⁸¹⁶ programme [21]
- ⁸¹⁷ • **other implementation-specific uncertainties:** e.g. the variation of the h_{damp} scale in the Powheg
⁸¹⁸ samples.

⁸¹⁹ **5.2.2 Propagation of theory uncertainties to likelihood fit**

⁸²⁰ The effect of the aforementioned theoretical uncertainty sources are grouped by their effect on the MC
⁸²¹ templates that are given as input to the likelihood fit. There are four types of uncertainties considered in
⁸²² the fit, as follows:

- 823 A) **inclusive cross-section uncertainties**: are uncertainties on the cross-section used for the normalisation of the MC samples. These are implemented in the fit using Gaussian priors that affect the normalisation of a given samples in all regions in a correlated manner. These uncertainties are applied only on the samples whose normalisation is not freely floating in the fit.
- 827 B) **shape uncertainties**: are uncertainties on the shape of the variables used in the fit (muon charge for
828 the 1-lepton region and $m(b\bar{b})$ for the 0-lepton region, as described in Sec. 7). These are estimated
829 by comparing the shape of the fitted variables for the nominal MC samples and the alternative
830 samples that probe the uncertainties outlined in Sec. 5.2.1.
- 831 C) **relative acceptance uncertainties**: the theory uncertainties can also alter the shape of the ob-
832 servables used to separate the different fit regions. These shape differences induce normalisa-
833 tion/acceptance differences between the regions that are used in the fit. For example differences
834 in the E_T^{miss} shape can induce relative acceptance differences between the adjacent E_T^{miss} bins and
835 differences in the lepton p_T spectrum can induce relative acceptance differences between the 0 and
836 1-lepton channels. These are included in the fit as Gaussian priors whose magnitude is estimated
837 by

$$\sigma_{\text{accept}} = \sqrt{\sum_i \left(1 - \frac{N_A^{\text{alt},i}}{N_B^{\text{alt},i}} \right) \left(\frac{N_A^{\text{nom}}}{N_B^{\text{nom}}} \right)^2}, \quad (8)$$

838 where i runs over all alternative MC generators considered for a given process and $(A, B) = (\text{SR,CR1}),$
839 $(\text{CR1,CR2}), (E_T^{\text{miss}} \text{ bin1}, E_T^{\text{miss}} \text{ bin2}), (E_T^{\text{miss}} \text{ bin2}, E_T^{\text{miss}} \text{ bin3}), (\text{resolved, merged})$. Since the
840 uncertainty is relative between region A and region B it is only applied on region B in the fit.

- 841 D) **others**: an additional uncertainty on the flavour composition is assigned on the W and $Z+\text{heavy}$
842 flavour components (Zhf, Whf), which consist of bb, cc, bc, bl , in order to allow the cc, bc, bl
843 components to vary individually from the total Zhf, Whf normalisation which is freely floating in
844 the fit.

845 The samples used for the derivation of the above uncertainties are described [here](#).

846 **Todo:** transfer systematic samples from presentation to the subsections below

847 5.2.3 Top

848 5.2.4 V+jets

849 5.2.5 Signal

850 5.2.6 Others

851 5.2.7 Derivation of the theory uncertainties at truth level

852 Since several of the alternative MC samples needed for the estimation of the theory uncertainties have not
853 been run through detector simulation but were only available at truth level, a framework to process these
854 samples has been developed.

855 The framework consists of two steps: (i) production of ntuples from EVNT files (handled by [vhbbtruthframework-](#)
856 [21](#)) and (ii) application of event selection (handled by [MonoHbbTruthFramework_NtupleToHist](#)).

857 The truth-level object reconstruction and analysis selection is designed to mimic as closely as possible the
858 event selection applied at reconstruction level. The derivation of the theory systematics in the analysis
859 phase-space consists therefore in applying the truth-level selection on all of the alternative MC samples
860 and comparing the resulting differences in shape or normalisation as described in Sec. [5.2.2](#).

861 Todo: add plots from validation studies

Not reviewed, for internal circulation only

862 6 Pre-fit results [Anindya]

863 This section describes the comparisons between data and background MC for various variables for both
 864 Resolved and Merged topology in 0,1 and 2-lepton channels for different b-tag categories. The variables are
 865 compared with full Run2 dataset with full event selection applied and before any fitting of the background
 866 model to the data. A blinding is applied around the Higgs mass [70-140] GeV for the 0-lepton channel.

867 6.1 0-lepton [Anindya]

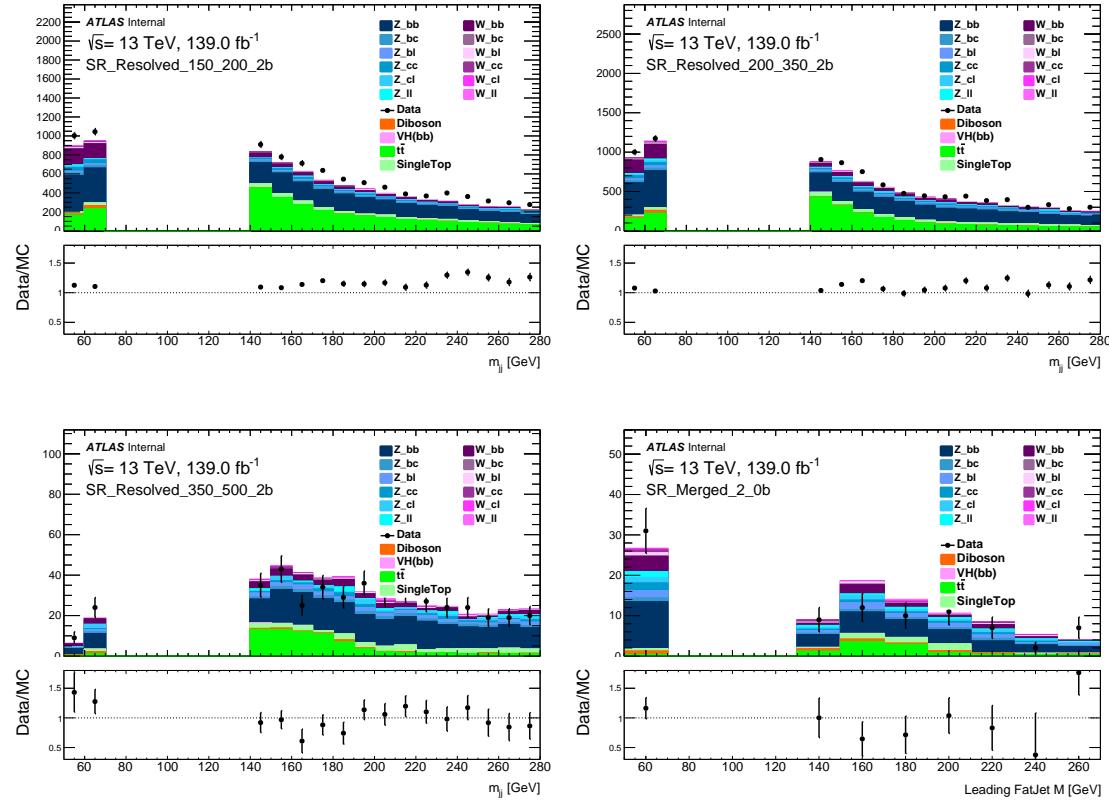


Figure 12: Higgs candidate mass spectra in the different E_T^{miss} regions with 2 b-tagged jets in the 0-lepton channel.

Not reviewed, for internal circulation only

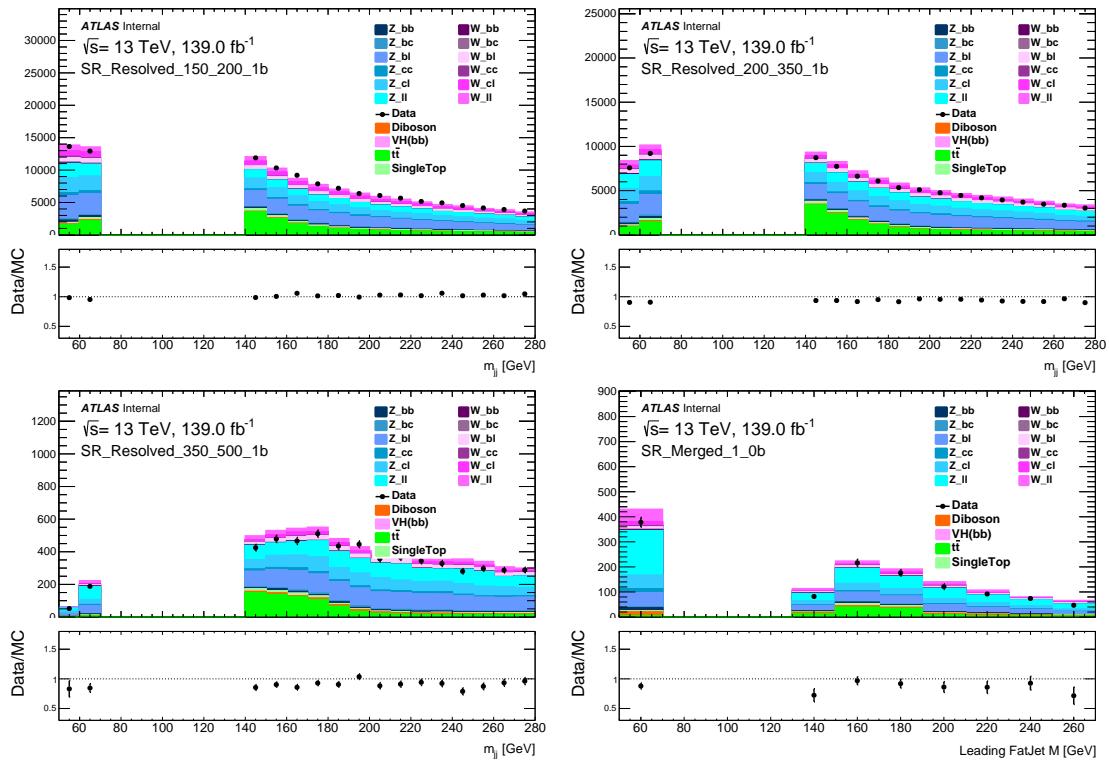
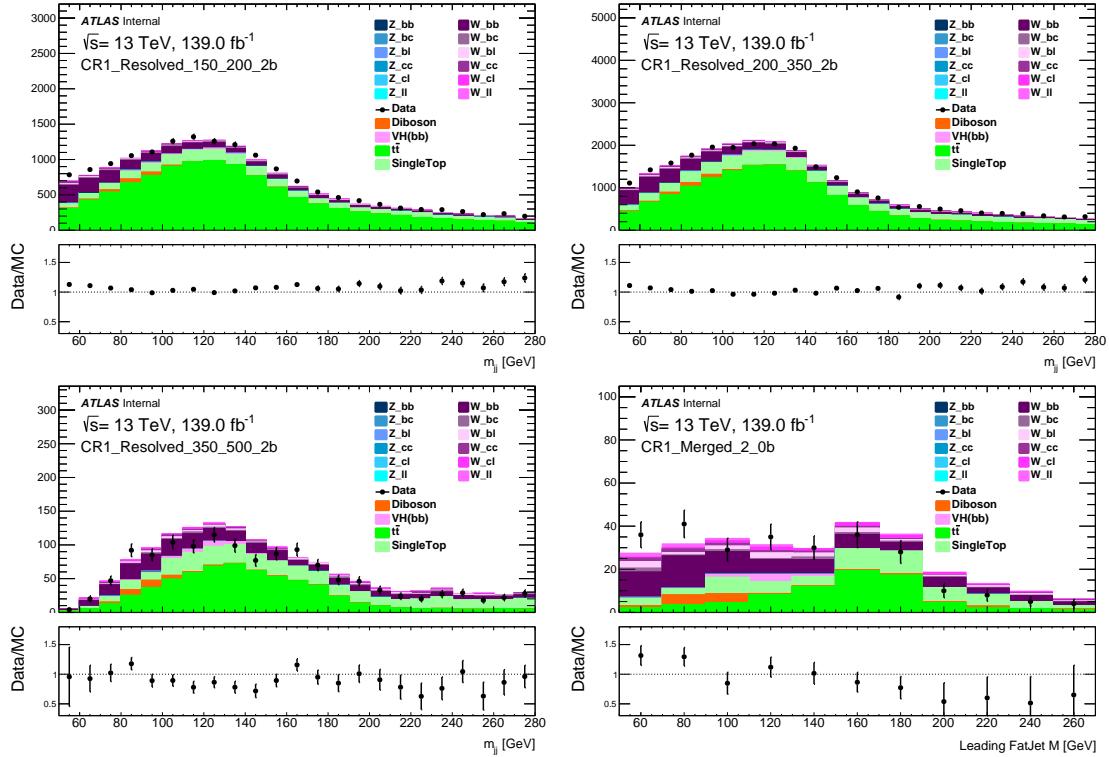


Figure 13: Higgs candidate mass spectra in the different E_T^{miss} regions with 1 b-tagged jet in the 0-lepton channel.

869 6.2 1-lepton [Anindya]

Not reviewed, for internal circulation only

Figure 14: Higgs candidate mass spectra in the different E_T^{miss} regions with 2 b-tagged jet in the 1-lepton channel.

Not reviewed, for internal circulation only

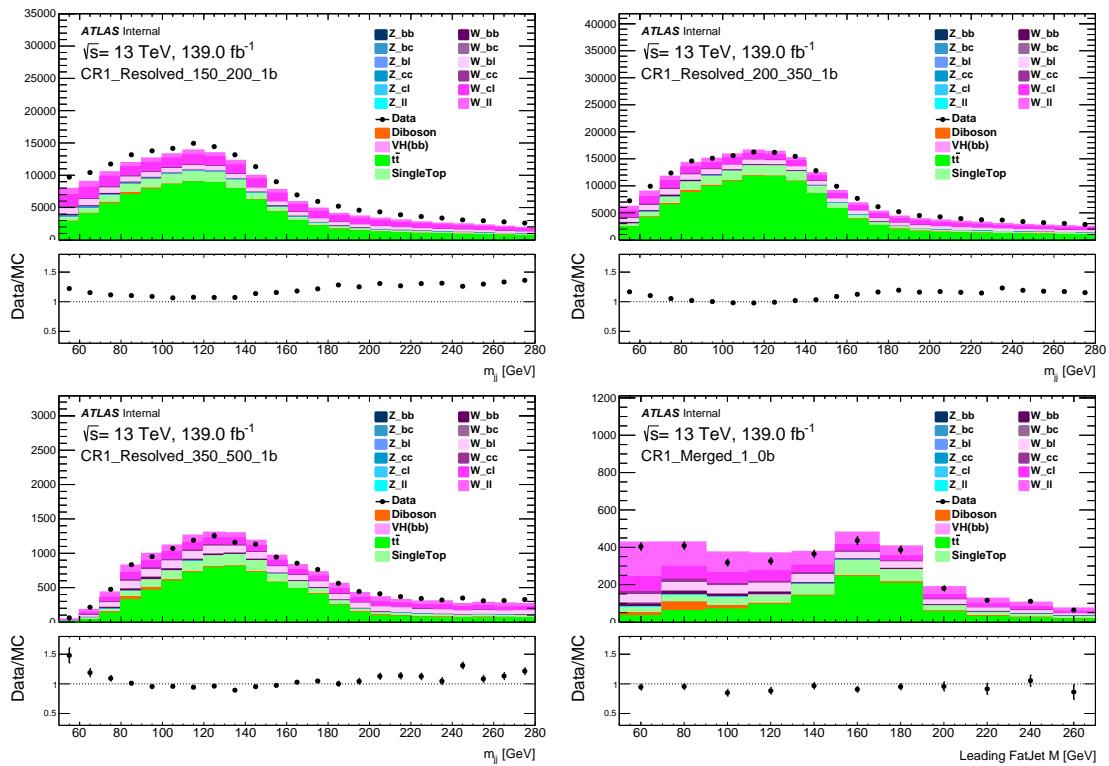


Figure 15: Higgs candidate mass spectra in the different E_T^{miss} regions with 1 b-tagged jet in the 1-lepton channel.

Not reviewed, for internal circulation only

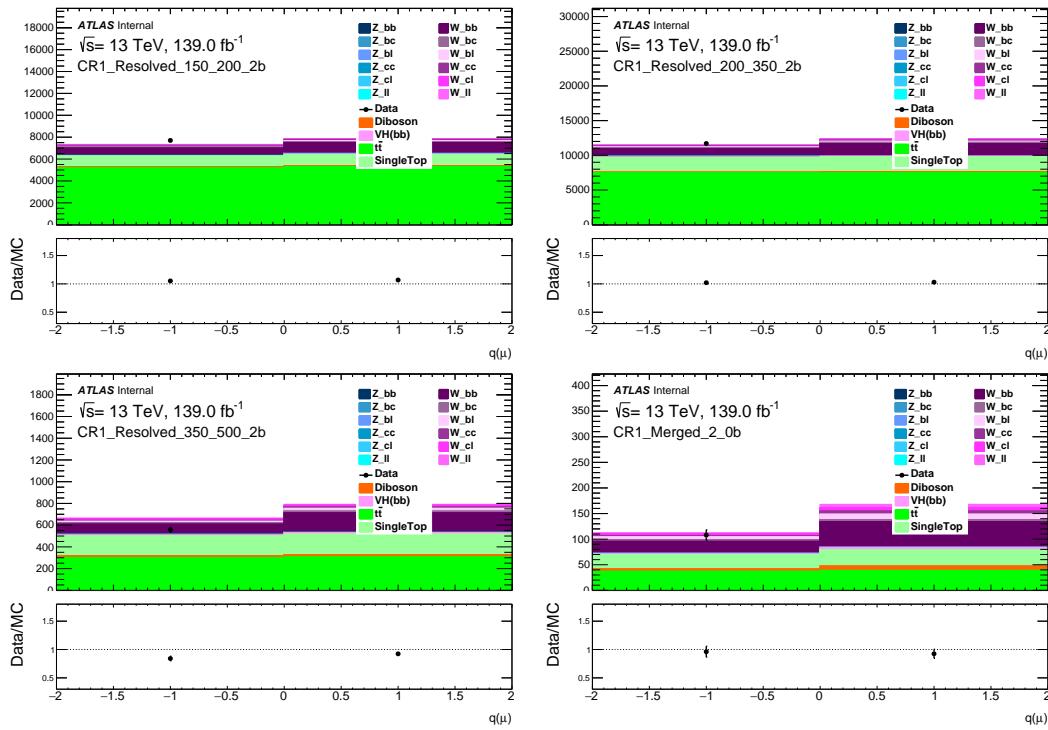


Figure 16: Muon charge distribution in the different E_T^{miss} regions with 2 b-tagged jet in the 1-lepton channel.

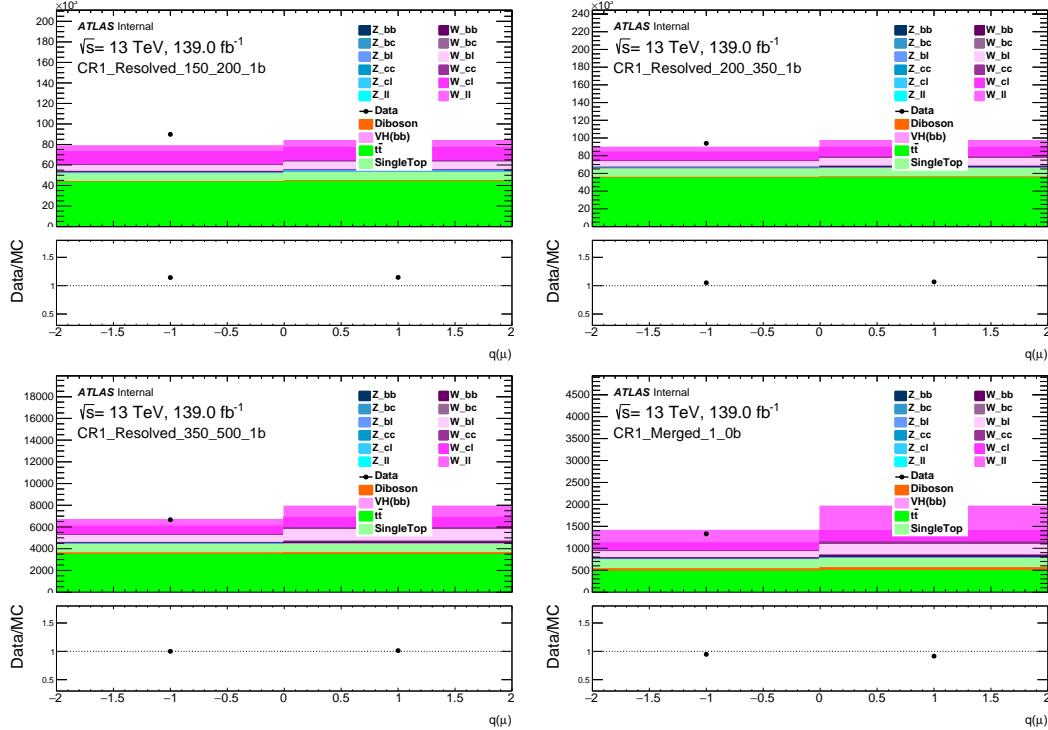
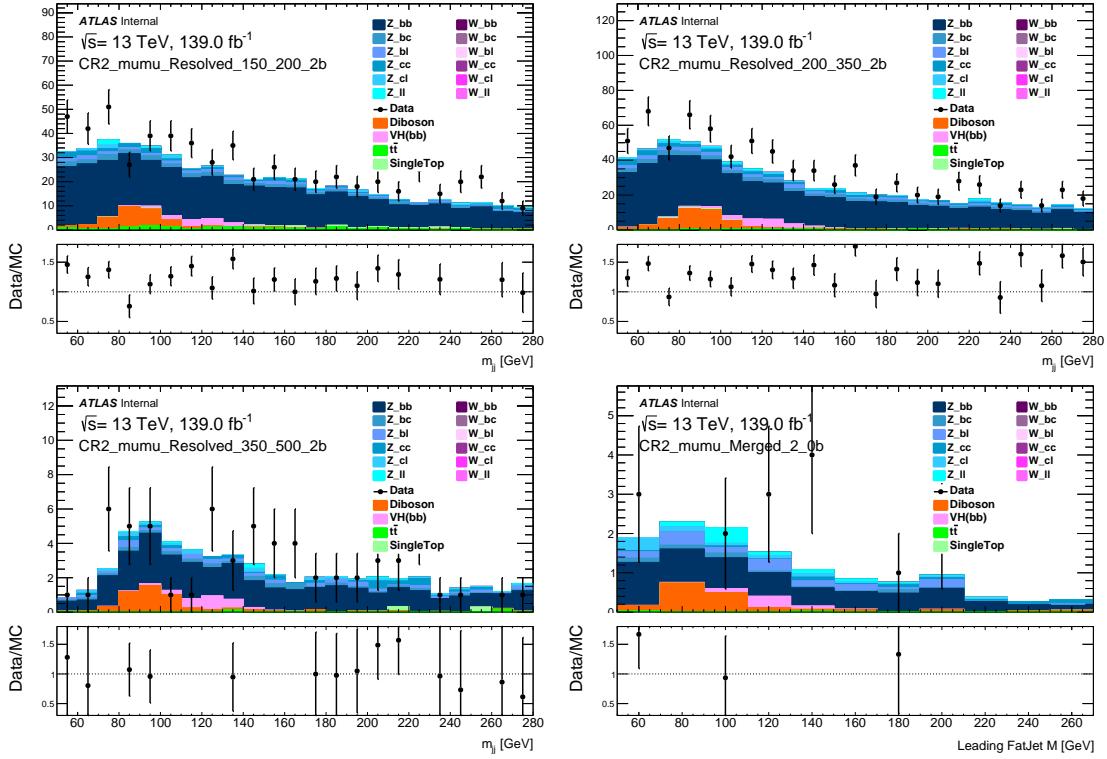


Figure 17: Muon charge distribution in the different E_T^{miss} regions with 1 b-tagged jet in the 1-lepton channel.

870 **6.3 2-lepton [Anindya]**871 **6.3.1 2-muon channel**

Not reviewed, for internal circulation only

Figure 18: Higgs candidate mass spectra in the different E_T^{miss} regions with 2 b-tagged jet in the 2-muon channel.

Not reviewed, for internal circulation only

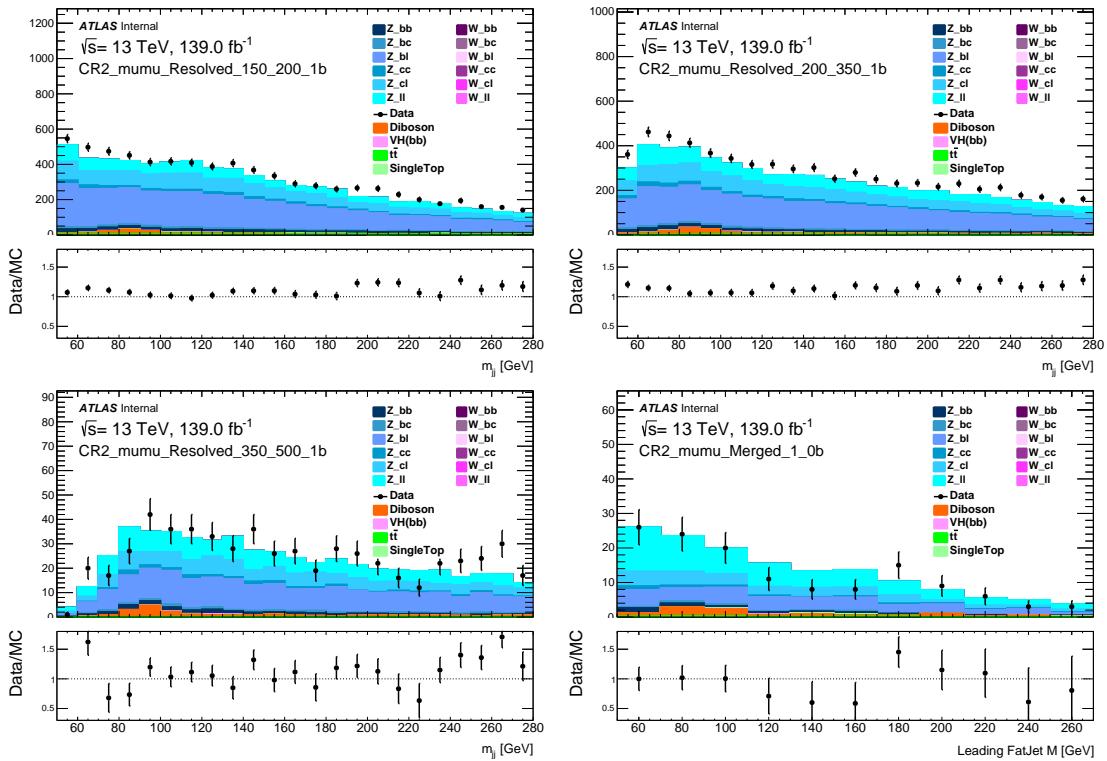


Figure 19: Higgs candidate mass spectra in the different E_T^{miss} regions with 1 b-tagged jet in the 2-muon channel.

872 **6.3.2 2-electron channel**

Not reviewed, for internal circulation only

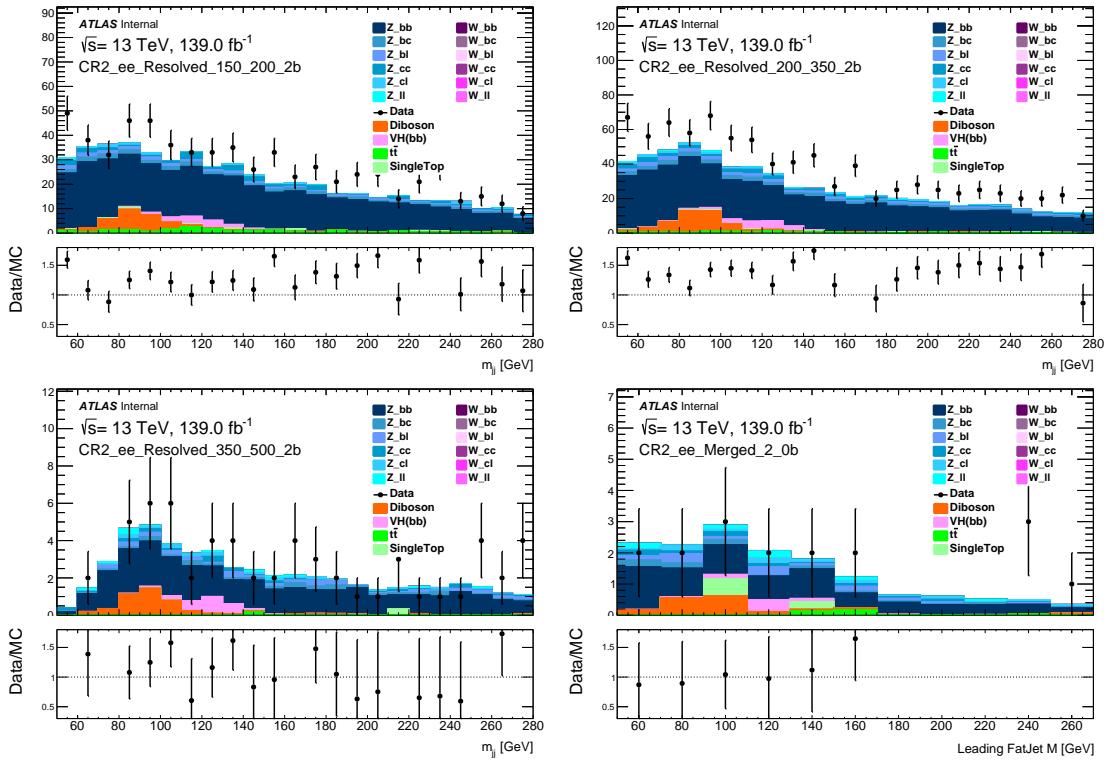


Figure 20: Higgs candidate mass spectra in the different E_T^{miss} regions with 2 b-tagged jet in the 2-electron channel.

Not reviewed, for internal circulation only

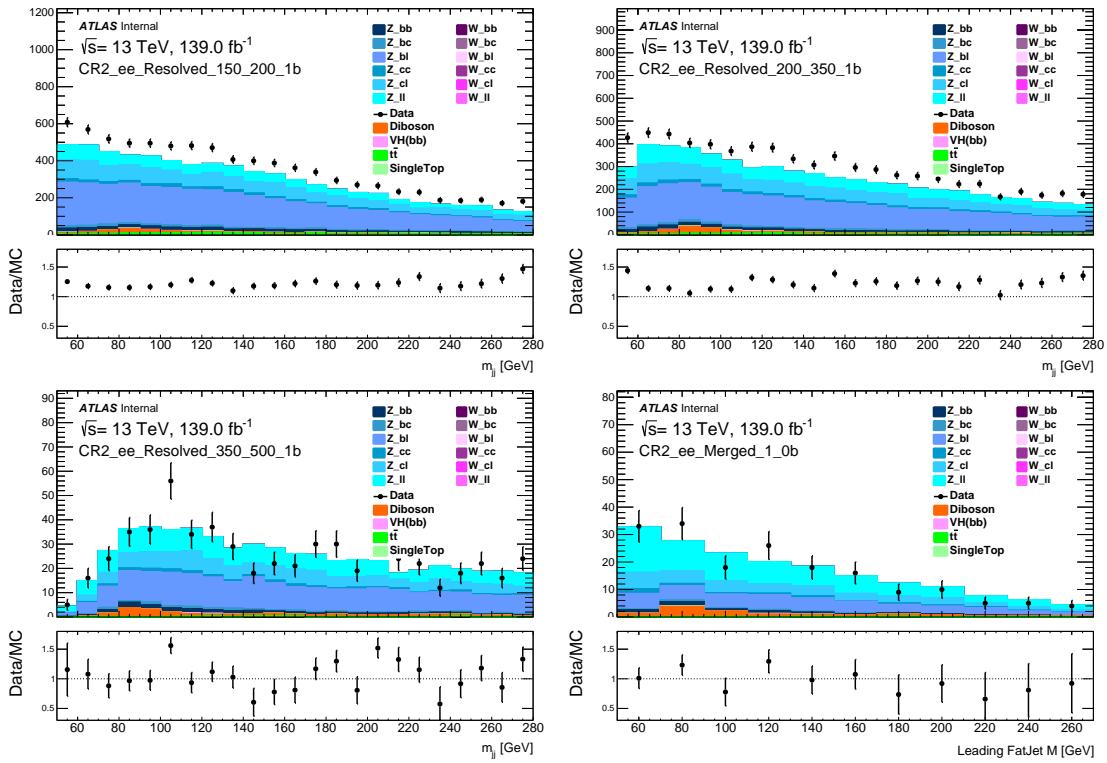


Figure 21: Higgs candidate mass spectra in the different E_T^{miss} regions with 1 b-tagged jet in the 2-muon channel.

Data/MC comparison plots for other variables are documented in appendix F.

873 7 Statistical analysis [Andrea, Spyros, Dan]

874 The interpretation of the final results is performed in terms of a profile likelihood fit of a model to the data.
 875 Apart from the simulated backgrounds, $Z' + 2HDM$ and $2HDM + a$ signals are part of the model.

876 7.1 Likelihood definition

877 The statistical analysis of the data uses a binned likelihood function constructed as the product of Poisson
 878 probability terms,

$$\text{Pois}(n|\mu S + B) \left[\prod_{b \in \text{bins}}^n \frac{\mu v_b^{\text{sig}} + v_b^{\text{bkg}}}{\mu S + B} \right], \quad (9)$$

879 where μ , a signal strength parameter, multiplies the expected signal yield v_b^{sig} in each histogram bin b ,
 880 and v_b^{bkg} represents the background content for bin b . The dependence of the signal and background
 881 predictions on the systematic uncertainties is described by a set of nuisance parameters (NP) θ , which
 882 are parameterized by Gaussian or log-normal priors; the latter are used for normalization uncertainties in
 883 order to maintain a positive likelihood.

884 The nominal fit result in terms of μ and σ_μ is obtained by maximising the likelihood function with
 885 respect to all parameters. This is referred to as the maximised log-likelihood value, MLL. The test
 886 statistic q_μ is then constructed according to the profile likelihood: $q_\mu = -2 \ln(\mathcal{L}(\mu, \hat{\theta}_\mu)/\mathcal{L}(\hat{\mu}, \hat{\theta}))$, where
 887 $\hat{\mu}$ and $\hat{\theta}$ are the parameters that maximise the likelihood, and $\hat{\theta}_\mu$ are the nuisance parameter values that
 888 maximise the likelihood for a given μ . This test statistic is used to measure the compatibility of the
 889 background-only model with the observed data and to derive exclusion intervals derived with the CL_s
 890 method [**Cowan:2010js**].

891 7.2 Fit variables and event categories

892 Signal contributions accumulate in the zero-lepton region. Here the fit of the model to the data is based
 893 on the mass of the two signal jets in the resolved channels and on the mass of the single large- R jet in
 894 the merged channels. This Higgs boson candidate mass serves as a variable discriminating signals from
 895 the backgrounds, denoted by m_H . Furthermore, several channels distinguished by the b -tag multiplicity,
 896 the range of E_T^{miss} values and potentially also the number of central small- R jets are used in order to gain
 897 sensitivity to the different signals.

898 For $Z' + 2HDM$ and gluon-gluon induced $2HDM + a$ processes the sensitivity comes almost exclusively
 899 through the channels with 2 b -tags, while for $b\bar{b}$ induced $2HDM + a$ events also the channels with $\geq 3b$ -
 900 tags (resolved) or with 2 b -tags in the leading large- R jet and $\geq 1b$ -jet outside the large- R jet (denoted by
 901 $2_{\text{in}} 1_{\text{out}}$) give a large contribution to the sensitivity. For all three signal models the 1 b -tag category has
 902 barely any effect on the sensitivity. However, the 1 b -tag channel could be helpful to constrain the light
 903 flavor components of the $W/Z+\text{jets}$ backgrounds, so that it might be included in the final fit.

904 Depending on the signal model and its mediators, the momentum of the resulting Dark Matter particles can
 905 largely differ. To be sensitive to signals in different kinematic phase spaces, a binning in E_T^{miss} is applied
 906 in the fit. The resolved region ranges from 150 GeV to 500 GeV and it is divided into three E_T^{miss} bins.
 907 The merged region contains events with > 150 GeV. In the previous iterations of the analysis, a single

	0 lepton	1 muon	2 leptons
Discriminant	m_H distribution	muon charge	single bin
Aim	signal region	$t\bar{t}$ and $W+$ jets control region	$Z+$ jets control region
b -tag multiplicities		resolved: (1), 2, (3) merged: $2_{in}0_{out}, (2_{in}1+_{out})$	
small-R jet multiplicities		to be decided whether it will be used in the resolved region	
E_T^{miss} proxy	E_T^{miss}	$E_{T, \text{lep. invis.}}^{\text{miss}}$	$E_{T, \text{lep. invis.}}^{\text{miss}}$
	resolved: [150,200], [200,350] and [350,500] GeV		
	merged: more than 500 GeV; possibly extra bin at higher values		

Table 12: Event categories entering the combined fit of the model to the data. The discriminant m_H denotes the mass of the light Higgs boson candidate.

908 E_T^{miss} bin was used in the merged region. Including additional high- E_T^{miss} bins can significantly enhance
 909 the sensitivity in the very boosted regime (see Appendix C and Appendix A). However, it still needs to be
 910 studied in the fit which E_T^{miss} binning is feasible within the statistical and systematic uncertainties in the
 911 merged region.

912 As explained in Appendix A, in the resolved region the number of central small-R jets differs between the
 913 signal models: Events coming from a $Z' + 2HDM$ signal have on average a relatively low jet multiplicity,
 914 while the $2HDM + a$ signals predict a higher number of jets in the final state. Therefore, introducing an
 915 additional binning in the number of jets the sensitivity allows to achieve a sensible sensitivity for all signal
 916 models. However, it still needs to be studied which binning choice would be possible within the statistical
 917 and systematic uncertainties.

918 The three dominant backgrounds in the signal regions are $t\bar{t}$, $W+$ jets and $Z+$ jets. They are estimated in
 919 dedicated control regions, which are defined based on the number of leptons in the events. The single-
 920 muon channels are used to constrain the $W+$ jets and the $t\bar{t}$ backgrounds. In order to distinguish between
 921 $W+$ jets and $t\bar{t}$ contributions, a division based on the charge of the muon charge is used. Generically
 922 the muon charge will be equally distributed in $t\bar{t}$ events, whereas the production of $W+$ jets is biased
 923 to produce positively charged W -bosons on account of the LHC being a pp collider. In the one-muon
 924 channels a binning in $E_{T, \text{lep. invis.}}^{\text{miss}}$ is used instead of the E_T^{miss} binning of the zero-lepton channels. The
 925 $Z+$ jets background is constrained by the two-lepton channels. Also here, E_T^{miss} is replaced by $E_{T, \text{lep. invis.}}^{\text{miss}}$
 926 to mimic the kinematics of $Z \rightarrow \nu\nu + \text{jets}$ events.

927 For all channels the fit is restricted on events in which m_H ranges from 50 to 280 GeV, if the event lies in
 928 the resolved region, and to 270 GeV if it is in the merged region. The fit is tested and validated in the side
 929 bands of the m_H distribution with $m_H < 70$ GeV and $m_H > 140$ GeV, while the region within the Higgs
 930 boson mass is blinded until the validation of the fit setup is completed. In the zero-lepton channels an
 931 equidistant binning for m_H is applied, where the number of bins is chosen to be relatively large in case of
 932 high statistics and relatively small in case of low statistics channels. In the previous iteration of the analysis
 933 46 bins were used for the channels in the two lowest E_T^{miss} bins, 23 bins for $350 \text{ GeV} < E_T^{\text{miss}} < 500 \text{ GeV}$
 934 and 11 bins in the merged region. However, the binning as well as the m_H range are going to be optimised
 935 due to the various changes in the analysis, such as the new event selection and the increase in statistics.

936 As the muon charge is used as fit variable in the single-muon channels, only two bins are present for
 937 each E_T^{miss} range, b -tag and jet multiplicity. In the two lepton channels, only a single bin is used for each
 938 $E_{T, \text{lept. invis.}}^{\text{miss}}$ range, b -tag and multiplicity. These choices of coarse binning increase the robustness of the
 939 fitting procedure without compromising the sensitivity, since with this binning the control channels still
 940 serve the purpose of constraining the background normalisation.

941 Table 12 summarizes the variables and event categories which enter the fitting procedure.

942 7.3 Treatment of nuisance parameters [Andrea, Spyros]

943 Two different types of nuisance parameters are used to model systematic uncertainties: floating parameters
 944 and parameters with priors.

945 A floating normalisation is generally associated with the cross-section and acceptance, where absolute
 946 ignorance of the rate is assumed and completely determined from data. Apart from the signal strength,
 947 the model contains some freely-floating background normalisation parameters which are constrained by
 948 the signal and control channels. Given the large amount of available data, the dominant processes can be
 949 constrained without using any prior knowledge about their overall normalisation. These freely floating
 950 parameters are:

- 951 • Signal strength (μ)
- 952 • $W + \text{Heavy Flavour}$ ($W + bb, bc, bl, cc$)
- 953 • $Z + \text{Heavy Flavour}$ ($Z + bb, bc, bl, cc$)
- 954 • $t\bar{t}$ production

955 In addition to the freely floating normalisation parameters, the fit contains nuisance parameters which
 956 are constrained by prior knowledge. They are related to further background normalisations, instrumental
 957 uncertainties and modeling uncertainties (see Section 5).

958 TODO: More information on theory uncertainties, correlation of systematics, pruning, smoothing ...

959 7.4 Results [Andrea, Dan, Spyros]

960 7.5 Discussion [Dan, Spyros]

961 References

- [1] *Search for Dark Matter Produced in Association with a Higgs Boson decaying to $b\bar{b}$ at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS Detector using 79.8 fb^{-1} of proton-proton collision data*, tech. rep. ATLAS-CONF-2018-039, CERN, 2018,
URL: <http://cds.cern.ch/record/2632344>.
- [2] J. Butler et al., *Search for Dark Matter Produced in Association with a Higgs Boson Decaying to $b\bar{b}$ at $\sqrt{s} = 13 \text{ TeV}$ using 79.8 fb^{-1} of pp collision data with the ATLAS Detector*, tech. rep. ATL-COM-PHYS-2018-042, CERN, 2018,
URL: <https://cds.cern.ch/record/2301321>.
- [3] D. Clowe, A. Gonzalez and M. Markevitch, *Weak lensing mass reconstruction of the interacting cluster 1E0657-558: Direct evidence for the existence of dark matter*, *Astrophys. J.* **604** (2004) 596, arXiv: [astro-ph/0312273 \[astro-ph\]](#).
- [4] P. A. R. Ade et al., *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13, arXiv: [1502.01589 \[astro-ph.CO\]](#).
- [5] L. Carpenter et al., *Mono-Higgs-boson: A new collider probe of dark matter*, *Phys. Rev.* **D89** (2014) 075017, arXiv: [1312.2592 \[hep-ph\]](#).
- [6] G. Bertone, D. Hooper and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279, arXiv: [hep-ph/0404175 \[hep-ph\]](#).
- [7] C. P. Burgess, M. Pospelov and T. ter Veldhuis, *The Minimal model of nonbaryonic dark matter: A Singlet scalar*, *Nucl. Phys.* **B619** (2001) 709, arXiv: [hep-ph/0011335 \[hep-ph\]](#).
- [8] J. March-Russell, S. M. West, D. Cumberbatch and D. Hooper, *Heavy Dark Matter Through the Higgs Portal*, *JHEP* **07** (2008) 058, arXiv: [0801.3440 \[hep-ph\]](#).
- [9] M. Aaboud et al., *Constraints on mediator-based dark matter and scalar dark energy models using $\sqrt{s} = 13 \text{ TeV}$ pp collision data collected by the ATLAS detector*, *JHEP* **05** (2019) 142, arXiv: [1903.01400 \[hep-ex\]](#).
- [10] M. Aaboud et al., *Combination of searches for invisible Higgs boson decays with the ATLAS experiment*, *Phys. Rev. Lett.* **122** (2019) 231801, arXiv: [1904.05105 \[hep-ex\]](#).
- [11] M. Aaboud et al., *Search for Dark Matter Produced in Association with a Higgs Boson Decaying to $b\bar{b}$ using 36 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS Detector*, *Phys. Rev. Lett.* **119** (2017) 181804, arXiv: [1707.01302 \[hep-ex\]](#).
- [12] A. M. Sirunyan et al., *Search for associated production of dark matter with a Higgs boson decaying to $b\bar{b}$ or $\gamma\gamma$ at $\sqrt{s} = 13 \text{ TeV}$* , *JHEP* **10** (2017) 180, arXiv: [1703.05236 \[hep-ex\]](#).
- [13] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, *Phys. Rept.* **516** (2012) 1, arXiv: [1106.0034 \[hep-ph\]](#).
- [14] A. Berlin, T. Lin and L.-T. Wang, *Mono-Higgs Detection of Dark Matter at the LHC*, *JHEP* **06** (2014) 078, arXiv: [1402.7074 \[hep-ph\]](#).
- [15] T. Abe et al., *LHC Dark Matter Working Group: Next-generation spin-0 dark matter models*, *Phys. Dark Univ.* (2018) 100351, arXiv: [1810.09420 \[hep-ex\]](#).

Not reviewed, for internal circulation only

- 1002 [16] M. Bauer, U. Haisch and F. Kahlhoefer,
 1003 *Simplified dark matter models with two Higgs doublets: I. Pseudoscalar mediators*,
 1004 *JHEP* **05** (2017) 138, arXiv: [1701.07427 \[hep-ph\]](#).
- 1005 [17] ATLAS Collaboration, TWiki,
 1006 URL: <https://twiki.cern.ch/twiki/bin/viewauth/Atlas/LowestUnprescaled>.
- 1007 [18] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential*
 1008 *cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079,
 1009 arXiv: [1405.0301 \[hep-ph\]](#).
- 1010 [19] R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040,
 1011 arXiv: [1410.8849 \[hep-ph\]](#).
- 1012 [20] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159,
 1013 arXiv: [1410.3012 \[hep-ph\]](#).
- 1014 [21] *ATLAS Run 1 Pythia8 tunes*, tech. rep. ATL-PHYS-PUB-2014-021, CERN, 2014,
 1015 URL: <https://cds.cern.ch/record/1966419>.
- 1016 [22] R. D. Ball et al.,
 1017 *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*,
 1018 *Nucl. Phys.* **B849** (2011) 296, arXiv: [1101.1300 \[hep-ph\]](#).
- 1019 [23] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys.* **B867** (2013) 244,
 1020 arXiv: [1207.1303 \[hep-ph\]](#).
- 1021 [24] T. Sjostrand, S. Mrenna and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*,
 1022 *Comput. Phys. Commun.* **178** (2008) 852, arXiv: [0710.3820 \[hep-ph\]](#).
- 1023 [25] E. Bothmann et al., *Event Generation with Sherpa 2.2*, (2019), arXiv: [1905.09127 \[hep-ph\]](#).
- 1024 [26] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, *JHEP* **12** (2008) 039,
 1025 arXiv: [0808.3674 \[hep-ph\]](#).
- 1026 [27] F. Cascioli, P. Maierhofer and S. Pozzorini, *Scattering Amplitudes with Open Loops*,
 1027 *Phys. Rev. Lett.* **108** (2012) 111601, arXiv: [1111.5206 \[hep-ph\]](#).
- 1028 [28] A. Denner, S. Dittmaier and L. Hofer,
 1029 *Collier: A fortran-based complex one-loop library in extended regularizations*,
 1030 *Comput. Phys. Commun.* **212** (2017) 220, arXiv: [1604.06792 \[hep-ph\]](#).
- 1031 [29] S. Schumann and F. Krauss,
 1032 *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **03** (2008) 038,
 1033 arXiv: [0709.1027 \[hep-ph\]](#).
- 1034 [30] S. Höche, F. Krauss, M. Schönher and F. Siegert,
 1035 *A critical appraisal of NLO+PS matching methods*, *JHEP* **09** (2012) 049,
 1036 arXiv: [1111.1220 \[hep-ph\]](#).
- 1037 [31] S. Höche, F. Krauss, M. Schönher and F. Siegert,
 1038 *QCD matrix elements + parton showers. The NLO case*, *JHEP* **04** (2013) 027,
 1039 arXiv: [1207.5030 \[hep-ph\]](#).
- 1040 [32] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, *QCD Matrix Elements + Parton Showers*,
 1041 *JHEP* **11** (2001) 063, arXiv: [hep-ph/0109231](#).
- 1042 [33] S. Höche, F. Krauss, S. Schumann and F. Siegert, *QCD matrix elements and truncated showers*,
 1043 *JHEP* **05** (2009) 053, arXiv: [0903.1219 \[hep-ph\]](#).

- 1044 [34] S. Frixione, P. Nason and G. Ridolfi,
 1045 *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*,
 1046 *JHEP* **09** (2007) 126, arXiv: [0707.3088 \[hep-ph\]](#).
- 1047 [35] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*,
 1048 *JHEP* **11** (2004) 040, arXiv: [hep-ph/0409146](#).
- 1049 [36] S. Frixione, P. Nason and C. Oleari,
 1050 *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*,
 1051 *JHEP* **11** (2007) 070, arXiv: [0709.2092 \[hep-ph\]](#).
- 1052 [37] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO*
 1053 *calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043,
 1054 arXiv: [1002.2581 \[hep-ph\]](#).
- 1055 [38] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*,
 1056 ATL-PHYS-PUB-2016-020, 2016, URL: <https://cds.cern.ch/record/2216168>.
- 1057 [39] D. J. Lange, *The EvtGen particle decay simulation package*,
 1058 *Nucl. Instrum. Meth. A* **462** (2001) 152.
- 1059 [40] M. Beneke, P. Falgari, S. Klein and C. Schwinn,
 1060 *Hadronic top-quark pair production with NNLL threshold resummation*,
 1061 *Nucl. Phys. B* **855** (2012) 695, arXiv: [1109.1536 \[hep-ph\]](#).
- 1062 [41] M. Cacciari, M. Czakon, M. Mangano, A. Mitov and P. Nason, *Top-pair production at hadron*
 1063 *colliders with next-to-next-to-leading logarithmic soft-gluon resummation*,
 1064 *Phys. Lett. B* **710** (2012) 612, arXiv: [1111.5869 \[hep-ph\]](#).
- 1065 [42] P. Bärnreuther, M. Czakon and A. Mitov, *Percent-Level-Precision Physics at the Tevatron:*
 1066 *Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \rightarrow t\bar{t} + X$* ,
 1067 *Phys. Rev. Lett.* **109** (2012) 132001, arXiv: [1204.5201 \[hep-ph\]](#).
- 1068 [43] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the*
 1069 *all-fermionic scattering channels*, *JHEP* **12** (2012) 054, arXiv: [1207.0236 \[hep-ph\]](#).
- 1070 [44] M. Czakon and A. Mitov,
 1071 *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*,
 1072 *JHEP* **01** (2013) 080, arXiv: [1210.6832 \[hep-ph\]](#).
- 1073 [45] M. Czakon, P. Fiedler and A. Mitov,
 1074 *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$* ,
 1075 *Phys. Rev. Lett.* **110** (2013) 252004, arXiv: [1303.6254 \[hep-ph\]](#).
- 1076 [46] M. Czakon and A. Mitov,
 1077 *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*,
 1078 *Comput. Phys. Commun.* **185** (2014) 2930, arXiv: [1112.5675 \[hep-ph\]](#).
- 1079 [47] M. Botje et al., *The PDF4LHC Working Group Interim Recommendations*, (2011),
 1080 arXiv: [1101.0538 \[hep-ph\]](#).
- 1081 [48] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Parton distributions for the LHC*,
 1082 *Eur. Phys. J. C* **63** (2009) 189, arXiv: [0901.0002 \[hep-ph\]](#).
- 1083 [49] A. D. Martin, W. Stirling, R. Thorne and G. Watt, *Uncertainties on α_S in global PDF analyses*
 1084 *and implications for predicted hadronic cross sections*, *Eur. Phys. J. C* **64** (2009) 653,
 1085 arXiv: [0905.3531 \[hep-ph\]](#).

Not reviewed, for internal circulation only

- 1086 [50] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky et al.,
 1087 *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024,
 1088 arXiv: [1007.2241 \[hep-ph\]](#).
- 1089 [51] J. Gao, M. Guzzi, J. Huston, H.-L. Lai, Z. Li et al.,
 1090 *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. D **89** (2014) 033009,
 1091 arXiv: [1302.6246 \[hep-ph\]](#).
- 1092 [52] E. Re,
 1093 *Single-top Wt-channel production matched with parton showers using the POWHEG method*,
 1094 Eur. Phys. J. C **71** (2011) 1547, arXiv: [1009.2450 \[hep-ph\]](#).
- 1095 [53] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber and C. D. White,
 1096 *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029,
 1097 arXiv: [0805.3067 \[hep-ph\]](#).
- 1098 [54] M. Aliev et al., *HATHOR – HAdronic Top and Heavy quarks crOss section calculatoR*,
 1099 Comput. Phys. Commun. **182** (2011) 1034, arXiv: [1007.1327 \[hep-ph\]](#).
- 1100 [55] P. Kant et al., *HatHor for single top-quark production: Updated predictions and uncertainty*
 1101 *estimates for single top-quark production in hadronic collisions*,
 1102 Comput. Phys. Commun. **191** (2015) 74, arXiv: [1406.4403 \[hep-ph\]](#).
- 1103 [56] R. Frederix, E. Re and P. Torrielli, *Single-top t-channel hadroproduction in the four-flavour*
 1104 *scheme with POWHEG and aMC@NLO*, JHEP **09** (2012) 130, arXiv: [1207.5391 \[hep-ph\]](#).
- 1105 [57] S. Frixione, E. Laenen, P. Motylinski and B. R. Webber, *Angular correlations of lepton pairs*
 1106 *from vector boson and top quark decays in Monte Carlo simulations*, JHEP **04** (2007) 081,
 1107 arXiv: [hep-ph/0702198](#).
- 1108 [58] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk,
 1109 *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*,
 1110 JHEP **03** (2013) 015, arXiv: [1212.3460 \[hep-ph\]](#).
- 1111 [59] S. Alioli, P. Nason, C. Oleari and E. Re,
 1112 *NLO single-top production matched with shower in POWHEG: s- and t-channel contributions*,
 1113 JHEP **09** (2009) 111, [Erratum: JHEP02,011(2010)], arXiv: [0907.4076 \[hep-ph\]](#).
- 1114 [60] K. Hamilton, P. Nason and G. Zanderighi, *MINLO: Multi-Scale Improved NLO*,
 1115 JHEP **10** (2012) 155, arXiv: [1206.3572 \[hep-ph\]](#).
- 1116 [61] G. Luisoni, P. Nason, C. Oleari and F. Tramontano, *$HW^\pm/HZ + 0$ and 1 jet at NLO with the*
 1117 *POWHEG BOX interfaced to GoSam and their merging within MiNLO*, JHEP **10** (2013) 083,
 1118 arXiv: [1306.2542 \[hep-ph\]](#).
- 1119 [62] G. Aad et al., *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions*
 1120 *at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09** (2014) 145, arXiv: [1406.3660 \[hep-ex\]](#).
- 1121 [63] J. Pumplin et al.,
 1122 *New Generation of Parton Distributions with Uncertainties From Global QCD Analysis*,
 1123 JHEP **07** (2002) 012, arXiv: [hep-ph/0201195](#).
- 1124 [64] M. L. Ciccolini, S. Dittmaier and M. Kramer,
 1125 *Electroweak radiative corrections to associated WH and ZH production at hadron colliders*,
 1126 Phys. Rev. **D68** (2003) 073003, arXiv: [hep-ph/0306234 \[hep-ph\]](#).

- Not reviewed, for internal circulation only
- [65] O. Brein, A. Djouadi and R. Harlander,
NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders,
Phys. Lett. **B579** (2004) 149, arXiv: [hep-ph/0307206 \[hep-ph\]](#).
 - [66] G. Ferrera, M. Grazzini and F. Tramontano,
Associated WH production at hadron colliders: a fully exclusive QCD calculation at NNLO,
Phys. Rev. Lett. **107** (2011) 152003, arXiv: [1107.1164 \[hep-ph\]](#).
 - [67] O. Brein, R. Harlander, M. Wiesemann and T. Zirke,
Top-Quark Mediated Effects in Hadronic Higgs-Strahlung, *Eur. Phys. J.* **C72** (2012) 1868,
arXiv: [1111.0761 \[hep-ph\]](#).
 - [68] G. Ferrera, M. Grazzini and F. Tramontano,
Higher-order QCD effects for associated WH production and decay at the LHC,
JHEP **04** (2014) 039, arXiv: [1312.1669 \[hep-ph\]](#).
 - [69] G. Ferrera, M. Grazzini and F. Tramontano,
Associated ZH production at hadron colliders: the fully differential NNLO QCD calculation,
Phys. Lett. **B740** (2015) 51, arXiv: [1407.4747 \[hep-ph\]](#).
 - [70] J. M. Campbell, R. K. Ellis and C. Williams, *Associated production of a Higgs boson at NNLO*,
JHEP **06** (2016) 179, arXiv: [1601.00658 \[hep-ph\]](#).
 - [71] L. Altenkamp, S. Dittmaier, R. V. Harlander, H. Rzehak and T. J. E. Zirke,
Gluon-induced Higgs-strahlung at next-to-leading order QCD, *JHEP* **02** (2013) 078,
arXiv: [1211.5015 \[hep-ph\]](#).
 - [72] R. V. Harlander, A. Kulesza, V. Theeuwes and T. Zirke,
Soft gluon resummation for gluon-induced Higgs Strahlung, *JHEP* **11** (2014) 082,
arXiv: [1410.0217 \[hep-ph\]](#).
 - [73] O. Brein, R. V. Harlander and T. J. E. Zirke, *vh@nnlo - Higgs Strahlung at hadron colliders*,
Comput. Phys. Commun. **184** (2013) 998, arXiv: [1210.5347 \[hep-ph\]](#).
 - [74] H. B. Hartanto, B. Jäger, L. Reina and D. Wackerloth,
Higgs boson production in association with top quarks in the POWHEG BOX,
Phys. Rev. D **91** (2015) 094003, arXiv: [1501.04498 \[hep-ph\]](#).
 - [75] D. de Florian et al.,
Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector, (2016),
arXiv: [1610.07922 \[hep-ph\]](#).
 - [76] ATLASPMG, *ttV NLO cross section*, (2017), URL:
<https://twiki.cern.ch/twiki/bin/view/AtlasProtected/CrossSectionNL0ttV>.
 - [77] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J.* (2019),
arXiv: [1902.04655 \[hep-ex\]](#).
 - [78] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, *JINST* **14** (2019) P03017,
arXiv: [1812.03848 \[hep-ex\]](#).
 - [79] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J. C* **76** (2016) 292,
arXiv: [1603.05598 \[hep-ex\]](#).

- Not reviewed, for internal circulation only
- [80] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2017-029, 2017, URL: <https://cds.cern.ch/record/2261772>.
 - [81] M. Cacciari, G. P. Salam and G. Soyez, *The anti- k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: [0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189).
 - [82] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: [1603.02934 \[hep-ex\]](https://arxiv.org/abs/1603.02934).
 - [83] ATLAS Collaboration, *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702>.
 - [84] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-015, 2015, URL: <https://cds.cern.ch/record/2037613>.
 - [85] ATLAS Collaboration, *Forward Jet Vertex Tagging: A new technique for the identification and rejection of forward pileup jets*, ATL-PHYS-PUB-2015-034, 2015, URL: <https://cds.cern.ch/record/2042098>.
 - [86] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084, arXiv: [0912.1342 \[hep-ph\]](https://arxiv.org/abs/0912.1342).
 - [87] ATLAS Collaboration, *In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector*, Eur. Phys. J. C **79** (2019) 135, arXiv: [1807.09477 \[hep-ex\]](https://arxiv.org/abs/1807.09477).
 - [88] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017, URL: <https://cds.cern.ch/record/2268678>.
 - [89] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*, JHEP **08** (2018) 089, arXiv: [1805.01845 \[hep-ex\]](https://arxiv.org/abs/1805.01845).
 - [90] G. Aad et al., *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, 2019, arXiv: [1907.05120 \[hep-ex\]](https://arxiv.org/abs/1907.05120).
 - [91] D. Krohn, J. Thaler and L.-T. Wang, *Jets with Variable R*, JHEP **06** (2009) 059, arXiv: [0903.0392 \[hep-ph\]](https://arxiv.org/abs/0903.0392).
 - [92] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005, arXiv: [0802.1188 \[hep-ph\]](https://arxiv.org/abs/0802.1188).
 - [93] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, Phys. Lett. **B659** (2008) 119, arXiv: [0707.1378 \[hep-ph\]](https://arxiv.org/abs/0707.1378).
 - [94] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: <https://cds.cern.ch/record/2160731>.
 - [95] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **78** (2018) 903, arXiv: [1802.08168 \[hep-ex\]](https://arxiv.org/abs/1802.08168).

- 1211 [96] ATLAS Collaboration,
 1212 *Object-based missing transverse momentum significance in the ATLAS Detector*,
 1213 ATLAS-CONF-2018-038, 2018, URL: <https://cds.cern.ch/record/2630948>.
- 1214 [97] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, *J. Phys.* **G43** (2016) 023001,
 1215 arXiv: [1510.03865 \[hep-ph\]](https://arxiv.org/abs/1510.03865).
- 1216 [98] S. Hoeche and M. Schonherr, *Uncertainties in next-to-leading order plus parton shower matched*
 1217 *simulations of inclusive jet and dijet production*, *Phys. Rev.* **D86** (2012) 094042,
 1218 arXiv: [1208.2815 \[hep-ph\]](https://arxiv.org/abs/1208.2815).
- 1219 [99] ATLAS Collaboration, *Non-collision backgrounds as measured by the ATLAS detector during*
 1220 *the 2010 proton–proton run*, ATLAS-CONF-2011-137, 2011,
 1221 URL: <https://cds.cern.ch/record/1383840>.
- 1222 [100] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’, *Proceedings of the 22nd*
 1223 *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16,
 1224 ACM, 2016 785, ISBN: 978-1-4503-4232-2,
 1225 URL: <http://doi.acm.org/10.1145/2939672.2939785>.
- 1226 [101] T. A. collaboration,
 1227 *Search for Dark Matter Produced in Association with a Higgs Boson decaying to $b\bar{b}$ at*
 1228 *$\sqrt{s} = 13 \text{ TeV}$ with the ATLAS Detector using 79.8 fb^{-1} of proton-proton collision data*, (2018).
- 1229 [102] S. Argyropoulos et al., *Search for resonances decaying to a W or Z boson and a Higgs boson in*
 1230 *the $vvbb$, ℓvbb and $\ell\ell bb$ final states with $\sqrt{s} = 13 \text{ TeV}$ ATLAS data in the context of models with*
 1231 *2 Higgs doublets or additional heavy vector triplets*, tech. rep. ATL-COM-PHYS-2016-479,
 1232 CERN, 2016, URL: <https://cds.cern.ch/record/2151842>.
- 1233 [103] M. Tanabashi et al., *Review of Particle Physics*, *Phys. Rev.* **D98** (2018) 030001.
- 1234 [104] A. S. Bell and F. Lo Sterzo,
 1235 *Signal and Background Modelling Studies for the Standard Model VH, $H \rightarrow b\bar{b}$ Analysis*,
 1236 tech. rep. ATL-COM-PHYS-2018-505, CERN, 2018,
 1237 URL: <https://cds.cern.ch/record/2316951>.

1238 The supporting notes for the analysis should also contain a list of contributors. This information should
1239 usually be included in `mydocument-metadata.tex`. The list should be printed either here or before the
1240 Table of Contents.

[Not reviewed, for internal circulation only]

1241 List of contributions

Not reviewed, for internal circulation only

Abreu Juliao Ochoa De Castro, Maria Ines Argyropoulos, Spyridon	research on Xbb tagging and sensitivity study of signal region
Brooijmans, Gustaaf Burr, Jonathan Thomas	Analysis contact, development of framework for theory systematics and statistical interpretation, 2HDM+a signal requests research on Xbb tagging and sensitivity study of signal region Derivations and derivation framework, region boundary studies, Xbb tagger
Chan, Chen-hsun Chen, Jue Gadow, Paul Philipp	optimization with BDT, statistical treatment, signal MC generation research on Xbb tagging and sensitivity study of signal region analysis framework maintenance, derivations, MET significance cut optimisation, tight-jet cleaning study, RECAST
Fabiani, Veronica Filthaut, Frank Frost, James	implementation of truth-level selection, modelling systematics for 0-lepton SR, calibration of VR track jets supervisor for Veronica Fabiani, b-tagging expert analysis optimisation and strategy - channel categorisation, ttbar estimation/CR
Ghosh, Anindya	ntuple production & XAMPP framework maintenance, data/MC comparisons for all channels, EMTopo vs EMPflow comparisons, V+jets flavour composition studies, merging of ttbar, diboson and Znunu samples
Guest, Daniel Junggeburth, Johannes Josef Kortner, Sandra Lorenz, Jeanette	contact, derivation and derivation framework, Xbb tagging analysis framework support supervision of Philipp Gadow Supervisor of Andrea Matic, Theory uncertainties, Advice in SR optimisation
Mallik, Usha Matic, Andrea	supervision of Anindya Ghosh and Spyridon Argyropoulos ntuple production, XAMPP framework maintenance, signal and control region optimisation, truth signal studies, MET trigger calibration, Fitting
Peters, Krisztian Pham, Tuan Minh Pottgen, Ruth Rieck, Patrick Skorda, Eleni	Supervision of Janik von Ahnen optimization with BDT Supervision of Eleni Skorda advice Implementation and validation of 2L selection in truth framework, Studies of a Photon CR (postponed for now), Debugging of Vhbb truth framework and implementing missing objects/variables, Further debugging and implementations of small parts of code in all frameworks, Responsibility for the truth Ntuple production, Validation of 0L, 1L and 2L new selection in the truth framework, Estimation of uncertainties for 0L, 1L, and 2L SR for the new selection
Von Ahnen, Janik	Cut optimisation, b-tagging wp/algorithm studies, muon in jet correction
Wu, Sau Lan Zhou, Chen	supervision of Jay Chan, Tuan Pham, Chen Zhou optimization with BDT, statistical treatment, signal MC generation

1243

1244 Appendices

1245 A Optimisation of cut-based analysis [Andrea]

1246 To be filled properly

1247 The most recent presentations on the subject can be found in the following links:

- 1248 • [1 and 2-lepton selection](#)
1249 • [0-lepton selection](#)

1250 B Removal of non-collision background with tight jet cleaning [Philipp]

1251 In this section, we discuss the contribution of non-collision background (NCB) to the analysis event
 1252 selection.

1253 NCB can arise from beam background (e.g. beam muons) or from upstream proton losses by interactions
 1254 of the beam with residual gas or beam collimators [99]. These processes can produce fake jets with high
 1255 transverse momentum mostly peaking at $\varphi(E_T^{\text{miss}}) = 0, \pm\pi$.

1256 The event selection of the analysis presented in Ref. [ATLAS-CONF-2018-032] contains requirements on
 1257 the $E_T^{\text{miss, track}}$, the missing transverse momentum computed using only tracking information. The analysis
 1258 presented in this note does not contain these requirements due to discontinued support for use of $E_T^{\text{miss, track}}$
 1259 in analyses.

1260 Without the requirements on $E_T^{\text{miss, track}}$ but otherwise similar event selection as specified in Ref. [ATLAS-CONF-2018-033]
 1261 mismodeling in $\varphi(E_T^{\text{miss}}) = 0, \pm\pi$ due to NCB is introduced, as shown in Figure 22.

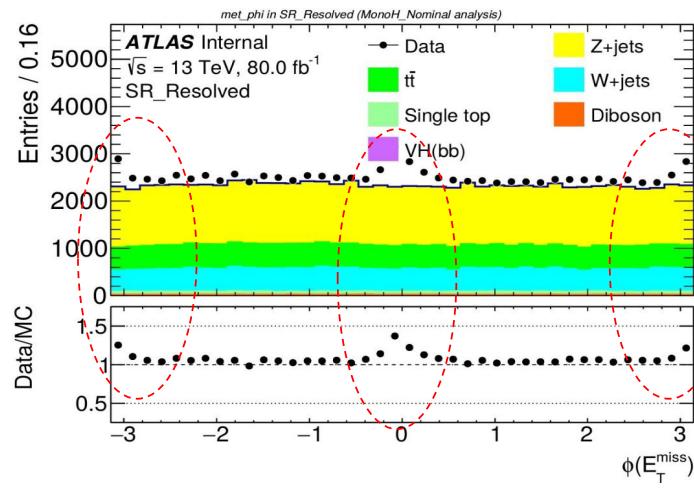


Figure 22: Distribution of the azimuthal E_T^{miss} component for Higgs candidate events in the resolved signal region (blinded), inclusive in b -tags, for Higgs candidate events with the simulated background estimate overlayed. Since the non-collision background is not accounted for in the simulated background estimate, mismodeling can be observed at $\varphi(E_T^{\text{miss}}) = 0, \pm\pi$.

1262 Figure 23 shows distributions of the azimuthal E_T^{miss} component $\varphi(E_T^{\text{miss}})$ for Higgs candidates in
 1263 blinded data with the preselection skimming cuts applied. The distribution on the left includes cuts on
 1264 $E_T^{\text{miss, track}}$, the right figure does not contain those and thus shows peaks at $\varphi(E_T^{\text{miss}}) = 0, \pm\pi$ due to NCB.

1265 In order to remove the mismodeling due to NCB, established cuts on so-called jet cleaning variables are
 1266 introduced, corresponding to the “tight” jet cleaning working point.

1267 Figure 24 shows that using the tight jet cleaning working point is able to eliminate the peaks at $\varphi(E_T^{\text{miss}}) = 0, \pm\pi$ due to NCB.

1269 The effect of moving from the “loose” to the “tight” tight jet cleaning working point on the acceptance \times
 1270 efficiency $A \times \varepsilon$ is small. For the Z' -2HDM model, for most points the yield loss is $A \times \varepsilon < 1 - 2\%$, at
 1271 most $A \times \varepsilon < 5\%$, as shown in Figure 25.

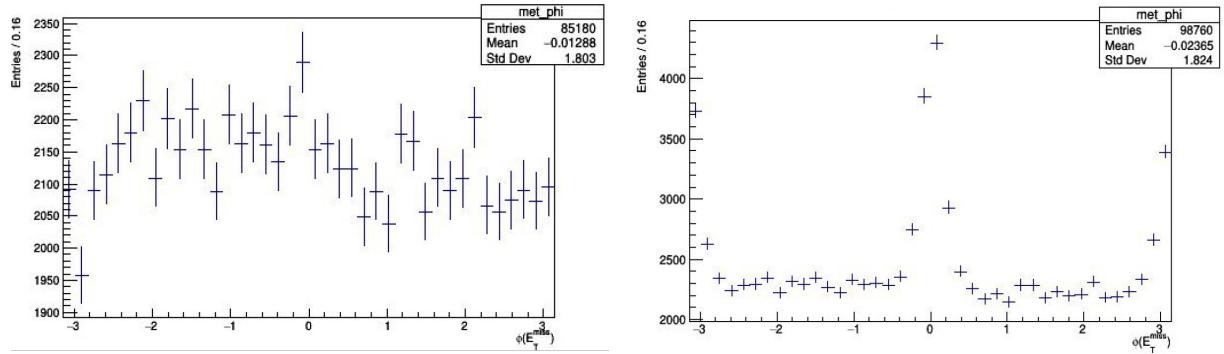


Figure 23: Distribution of the azimuthal E_T^{miss} component $\varphi(E_T^{\text{miss}})$ for Higgs candidate events in blinded data after pre-selection skimming cuts. The left figure includes cuts on $E_T^{\text{miss, track}}$, the right figure does not contain those and thus peaks at $\varphi(E_T^{\text{miss}}) = 0, \pm\pi$ are visible.

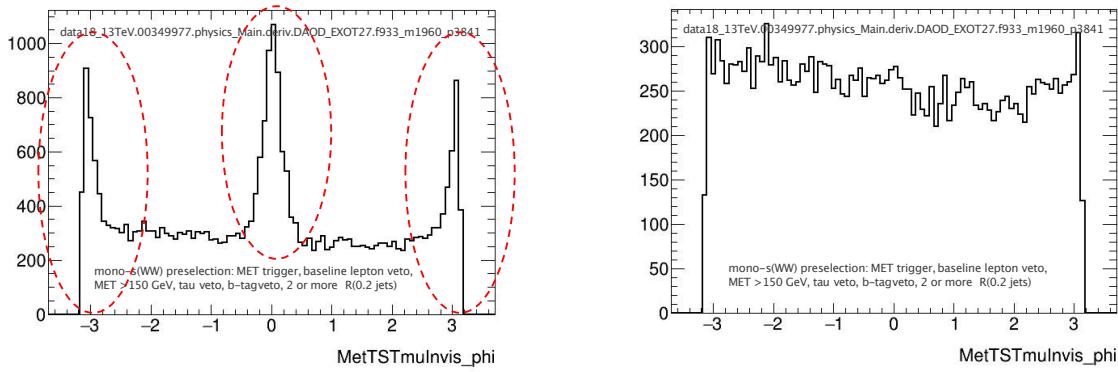


Figure 24: Distribution of the azimuthal E_T^{miss} component $\varphi(E_T^{\text{miss}})$ for Higgs candidate events in data of run 349977 after pre-selection cuts, including E_T^{miss} trigger requirement, lepton veto, $E_T^{\text{miss}} > 150$ GeV, b -tag veto, τ lepton veto, and requiring 2 or more small- R jets. The left (right) figure shows the “loose” (“tight”) jet cleaning working point.

1272 For the 2HDM + a model, that is also considered in the interpretation, the impact of this change is even
 1273 smaller than for the Z' -2HDM model. For the bb -induced production the yield loss is $A \times \varepsilon < 1.1\%$, for
 1274 the gluon-fusion (ggF) induced production the yield loss is $A \times \varepsilon < 1.3\%$, as shown in Figure 26.

1275 In summary, the move from the “loose” towards the “tight” jet cleaning working point does not impact the
 1276 signal $A \times \varepsilon$ but does effectively remove the NCB.

1277 C Optimizaiton using Boosted Decision Tree [Jay]

1278 In this section, we describe the optimization of the categories using boosted decision trees (BDT) with
 1279 the XGBoost package [100]. A general BDT strategy consists of three steps: preselection, BDT training
 1280 and categorisation optimisation. In the baseline BDT strategy, events which pass the preselections will
 1281 first be divided to three MET bins, including one resolved region and two merged regions. Two BDTs
 1282 are trained and used to sort events into multiple categories. The following subsections detail the baseline
 1283 BDT strategy as well as its performance.

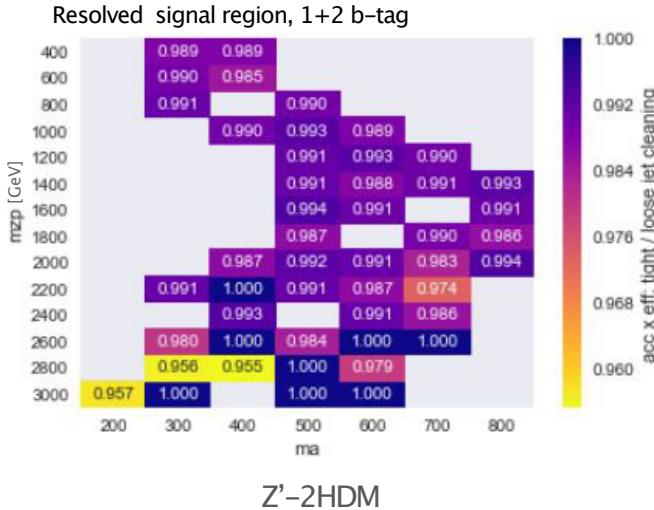


Figure 25: Acceptance \times efficiency for signals of the Z' -2HDM model for different $m_{Z'}$ and m_A in the resolved signal region for events with 1 or 2 b -tags.

1284 In addition, we evaluate the impact on the performance of several scenarios with different training variables
1285 and categorization strategies, which is discussed in the end of this section.

1286 C.1 Preslections

1287 All events are first divided to the resolved region ($150 \text{ GeV} \leq E_T^{\text{miss}} < 500 \text{ GeV}$) and the merged region
1288 ($E_T^{\text{miss}} \geq 500 \text{ GeV}$). In order to enhance the sensitivity for the high mass signals, the merged region
1289 is further split to two MET bins: 1st merged ($E_T^{\text{miss}} \geq 900 \text{ GeV}$) and 2nd merged ($500 \text{ GeV} \leq E_T^{\text{miss}} <$
1290 900 GeV). The preselections are the same as the selections for the baseline cut-based analysis, except that
1291 the H_T ratio is removed and used as a training variable. Table 13 summarizes the preselections used in
1292 the BDT-based categorization.

1293 C.2 BDT training

1294 Following the preselections, a BDT is trained in each of the three MET bins except for the 1st merged due
1295 to the lack of statistics, with Z +jets, W +jets and $t\bar{t}$ as training background. The training signal sample for
1296 each BDT is a mixture of multiple benchmark training signal points samples, which are chosen from the
1297 points close to the expected limit contour. Table 14 summarizes the benchmark training signal points as
1298 well as the training background used for each BDT.

1299 As listed in Table 14, both BDTs are trained using the number of jets n_j , H_T ratio, and the missing
1300 transverse energy E_T^{miss} . In addition, the BDT is trained using the transverse momentum p_T and the
1301 pseudorapidity η of the dijet system, the relative azimuthal angle of the dijet system $j j$ with respect to
1302 E_T^{miss} , and the pseudocontinious b-tagging scores of the two leading jets $\text{PCS}_{j_{1(2)}}$ in the resolved region,
1303 and the p_T and η of the large-R jet J , the relative azimuthal angle of the large-R jet with respect to E_T^{miss} ,
1304 and the pseudocontinious b-tagging scores of the two associated track-jets $\text{PCS}_{tj_{1(2)}}$ in the 2nd merged

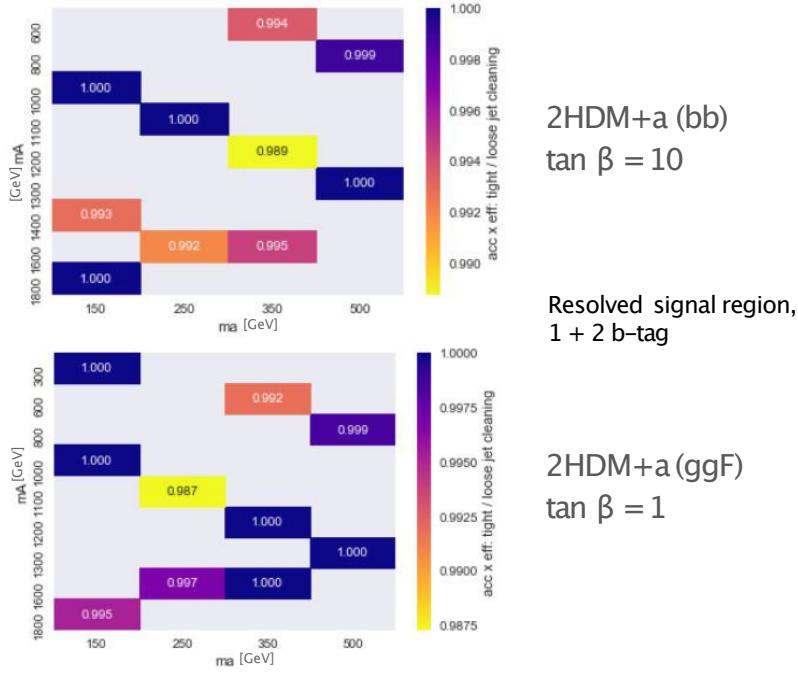


Figure 26: Acceptance \times efficiency for signals of the 2HDM + a model for different m_a and m_A in the resolved signal region for events with 1 or 2 b -tags.

1305 region. Figure 27 and 28 show the distributions of training variables in signals and background. The set of
 1306 variables has been minimized to reduce unnecessary systematics while keeping a similar performance.

1307 In order to avoid bias from tuning on the test samples, we employ a two-fold training method. The training
 1308 samples are split pseudo-randomly and equally into two sets A and B. In the first round of training, we
 1309 train the BDT on set A then apply it on set B, and switch their order in the second round: B for training
 1310 and A for testing. Eventually, these sets are combined for final evaluation. ROC curves of the BDTs for
 1311 the training sets and test sets are shown in Figure 29. Figure 30 shows an importance plot of the training
 1312 variables in each BDT and Figure 31 shows the BDT output distribution in all regions for signals and
 1313 background.

1314 C.3 Categorization optimization

1315 Events are sorted into five categories in the resolved region and three categories in the 2nd merged region
 1316 based on their dedicated merged BDT outputs. The 1st merged region does not have any further splitting
 1317 due to the lack of statistics.

1318 The BDT boundaries that define each category are optimized using an "N-step categorization proced-
 1319 ure". The first boundary is scanned to maximize the number counting significance $\sigma = \frac{s}{\sqrt{b}}$ of the first
 1320 categorization benchmark signal, where s (b) is the signal (background) yield inside the invariant mass
 1321 window, $100 \leq m_{jj(J)} \leq 140$ GeV for resolved (merged) region. A requirement of background statistical
 1322 error $< 20\%$ in each BDT category is enforced when scanning the boundary. After the optimal boundary
 1323 is found, the first BDT category is fixed. The process is repeated; the next boundaries are then scanned
 1324 to maximize the $\frac{s}{\sqrt{b}}$ of the next categorization benchmark signals based on the previously found optimal

Resolved	1^{st} merged	2^{nd} merged
lowest unprescaled $E_{\text{T}}^{\text{miss}}$ trigger		
veto on loose leptons		
$\min \Delta\Phi(E_{\text{T}}^{\text{miss}}, \text{Central + Forward jets}_{1,2,3}) > 20^{\circ}$		
$S > 12$		
$150 \text{ GeV} < E_{\text{T}}^{\text{miss}} < 500 \text{ GeV}$	$500 \text{ GeV} < E_{\text{T}}^{\text{miss}} < 900 \text{ GeV}$	$E_{\text{T}}^{\text{miss}} > 900 \text{ GeV}$
$N(\text{central small-R jets}) \geq 2$	$N(\text{central large-R jets}) \geq 1$	
$p_{j_1}^{\text{T}} > 45 \parallel p_{j_2}^{\text{T}} > 45$	-	
$\sum_{i=1}^{2(3)} p_{\text{T}}^{j_i} > 120(150) \text{ GeV}$	-	
$\Delta\Phi_{jj} < 140^{\circ}$	-	
$\Delta\Phi(p_{\text{T}}, \text{jet}_1 + \text{jet}_2) > 120^{\circ}$	-	
τ -veto	τ -veto	
2 b -tagged	2 b -tagged	
$\Delta R_{jj} < 1.8$	-	
$50 \text{ GeV} < m_{jj} < 280 \text{ GeV}$	$50 \text{ GeV} < m_{jj} < 270 \text{ GeV}$	
b-tag requirement on		
small-R jets	track-jets	

Table 13: Summary of preselections.

1325 BDT categories. The rest of the events that are not selected by the previously found BDT categories will
 1326 be dumped to the last BDT category. Similar to training benchmark signals, the categorization benchmark
 1327 signals are chosen from the signal points near the expected limit contour. The categorization benchmark
 1328 signals for the resolved region and 2^{nd} merged region are summarized in Table 15.

1329 The optimal BDT boundaries for all the categories are summarized in Table 16. The number counting
 1330 significance and the improvement over the baseline cut-based approach (ratio) for each signal point are
 1331 shown in Figure 32 and 33.

1332 C.4 Statistical analysis results

1333 We perform a simple statistical fitting on $m_{jj(J)}$ to better evaluate the performance of the BDT-based
 1334 categorization. For simplicity, no control region is used and only the MC statistical systematic is con-
 1335 sidered so far. The expected upper limit of cross-section for each signal is calculated by performing an
 1336 unconditional maximal likelihood fit. The pre-fit mass spectrum in each category is shown in Figure
 1337 34. For a fair comparison, the upper limit with the baseline cut-based approach is calculated under the
 1338 same setting, namely, no control region and only the MC statistical systematic considered. The ratio

MET bin	Training signal	Training background	training variables
Resolved $150 \leq E_T^{\text{miss}} < 500$	Z'2HDM ($m_{Z'} = 1400, m_A = 600$), Z'2HDM ($m_{Z'} = 1000, m_A = 600$), Z'2HDM ($m_{Z'} = 600, m_A = 300$)	Z+jets, W+jets, $t\bar{t}$	n_j, H_T ratio, E_T^{miss} , $p_{jj}^T, \eta_{jj}, \Delta\phi(jj, E_T^{\text{miss}})$, PCS _{j1} , PCS _{j2}
1 st merged $E_T^{\text{miss}} \geq 900$		No BDT	
2 nd merged $500 \leq E_T^{\text{miss}} < 900$	Z'2HDM ($m_{Z'} = 2000, m_A = 500$), Z'2HDM ($m_{Z'} = 1800, m_A = 600$), Z'2HDM ($m_{Z'} = 1600, m_A = 600$)	Z+jets, W+jets, $t\bar{t}$	n_j, H_T ratio, E_T^{miss} , $p_J^T, \eta_J, \Delta\phi(J, E_T^{\text{miss}})$, PCS _{tj1} , PCS _{tj2}

Table 14: Summary of training samples and variables used in each MET bin.

	Resolved region	2 nd merged region
1 st benchmark signal	Z'2HDM ($m_{Z'} = 1400, m_A = 600$)	2HDM+a ($m_A = 1800, m_a = 150$)
2 nd benchmark signal	2HDM+a ($m_A = 1200, m_a = 350$)	Z'2HDM ($m_{Z'} = 1600, m_A = 600$)
3 rd benchmark signal	Z'2HDM ($m_{Z'} = 1000, m_A = 600$)	
4 th benchmark signal	Z'2HDM ($m_{Z'} = 600, m_A = 300$)	

Table 15: Summary of categorization benchmark signals in the resolved region and 2nd merged region.

	Resolved region	2 nd merged region
1 st category	$O_{\text{BDT}} \geq 0.97$	$O_{\text{BDT}} \geq 0.69$
2 nd category	$0.88 \leq O_{\text{BDT}} < 0.97$	$0.42 \leq O_{\text{BDT}} < 0.69$
3 rd category	$0.75 \leq O_{\text{BDT}} < 0.88$	$O_{\text{BDT}} < 0.42$
4 th category	$0.44 \leq O_{\text{BDT}} < 0.75$	
5 th category	$O_{\text{BDT}} < 0.44$	

Table 16: Summary of categorization benchmark signals in the resolved region and 2nd merged region.

of the expected upper limit of the cross-section between the baseline cut-based approach and the BDT approach for each Z'2HDM signal point, which represents the improvement of the BDT approach over the cut-based, is shown in Figure 35. Furthermore, as shown in Figure 36, the expected exclusion limit contour of the Z'2HDM model is plotted for the baseline cut-based approach, the BDT approach and the re-optimized cut-based approach.

Not reviewed, for internal circulation only

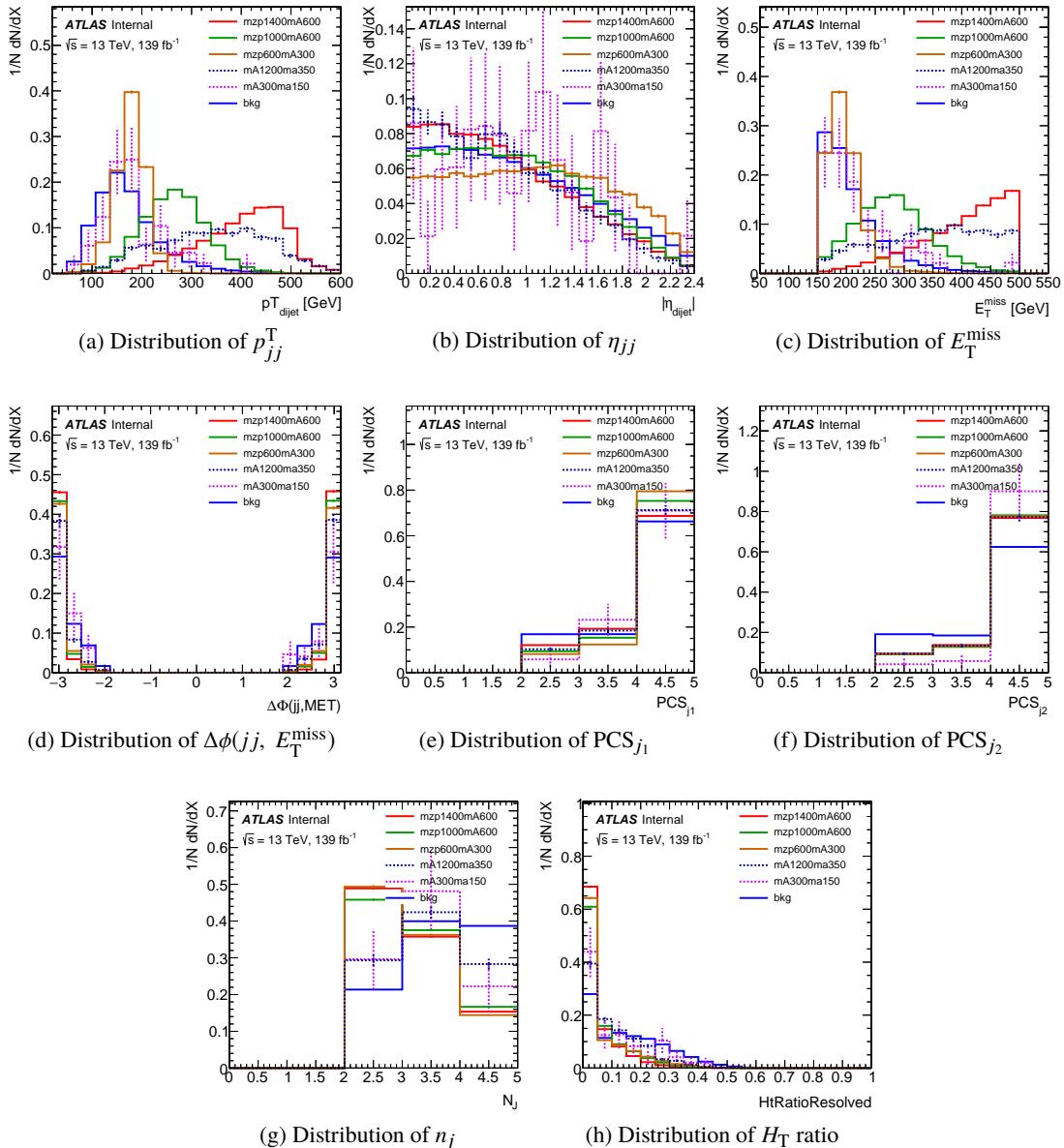


Figure 27: The distributions of training variables for resolved region in signals and background. Events are normalized.

1344
1345

C.5 Performance without the pseudo continuous b-tagging scores of the first two leading (track) jets in the training

Not reviewed, for internal circulation only

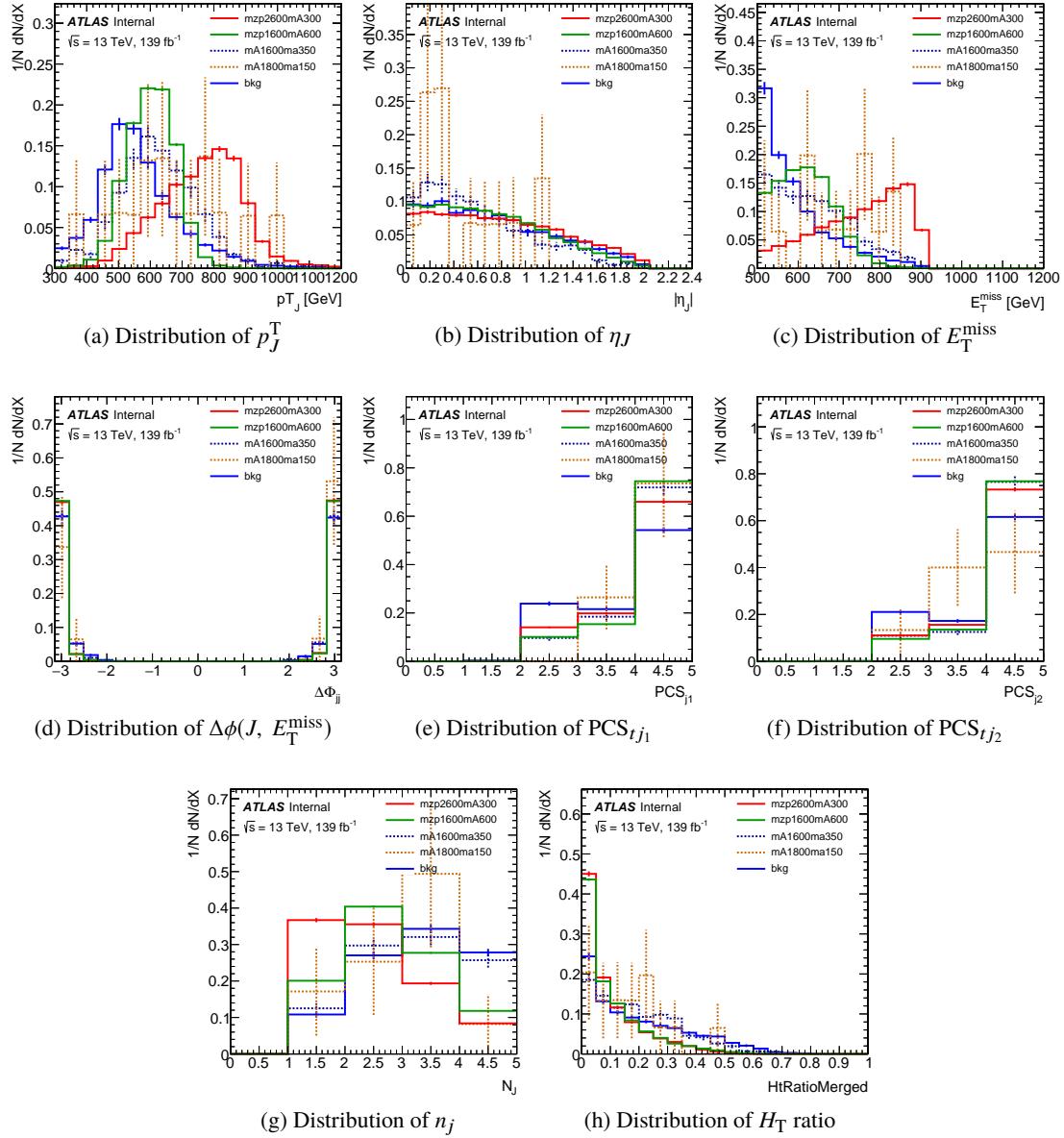


Figure 28: The distributions of training variables for 2nd merged region in signals and background. Events are normalized.

Not reviewed, for internal circulation only

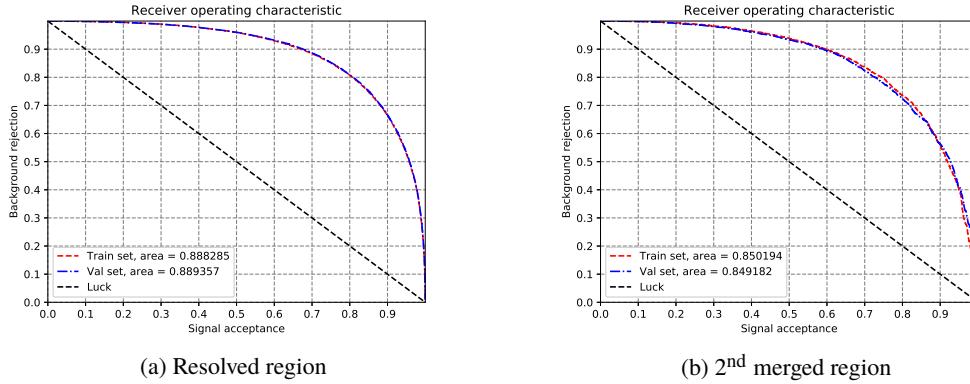


Figure 29: ROC curves of the BDTs in the training sets and test sets.

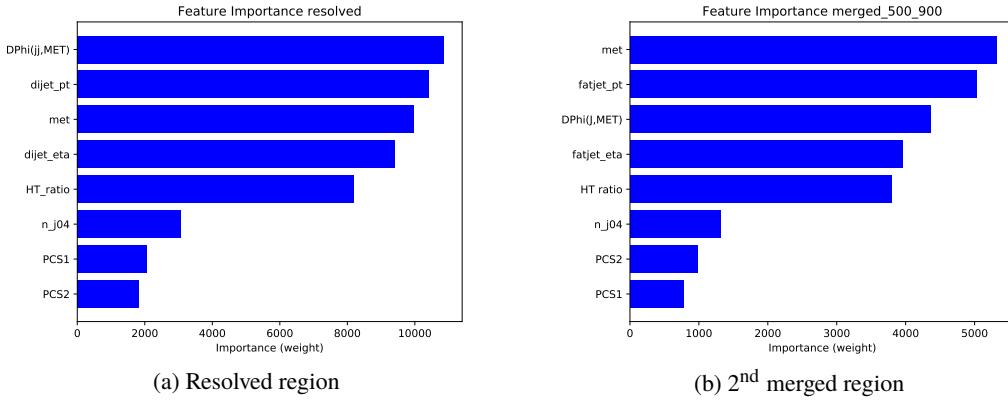


Figure 30: Feature importance of each variable for each BDT.

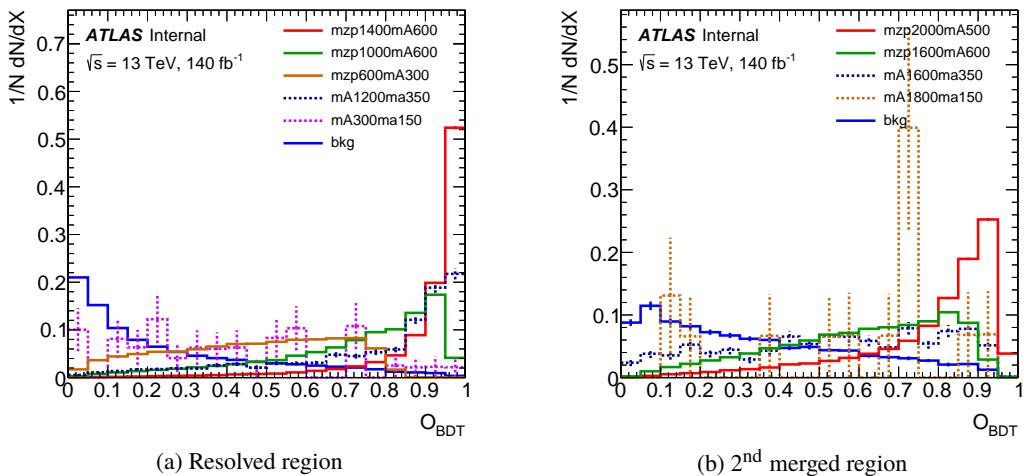


Figure 31: BDT output distribution for several signal points and background. Note that the 2HDM+a model signals shown here are all from the ggF production. Histograms are normalized.

Not reviewed, for internal circulation only

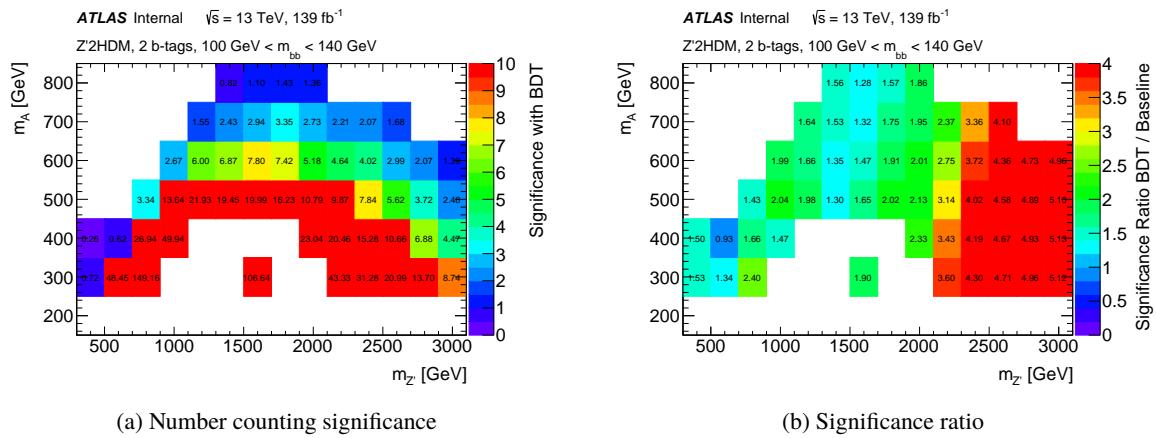


Figure 32: The number counting significance $\frac{s}{\sqrt{b}}$ and the improvement over the baseline cut-based approach (ratio) for each Z'2HDM signal point.

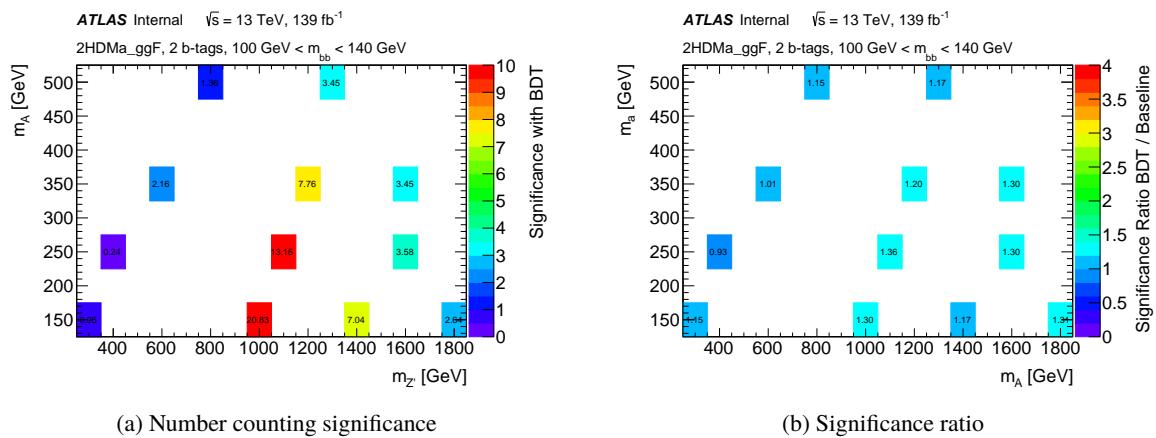


Figure 33: The number counting significance $\frac{s}{\sqrt{b}}$ and the improvement over the baseline cut-based approach (ratio) for each 2HDM+a (ggF) signal point.

Not reviewed, for internal circulation only

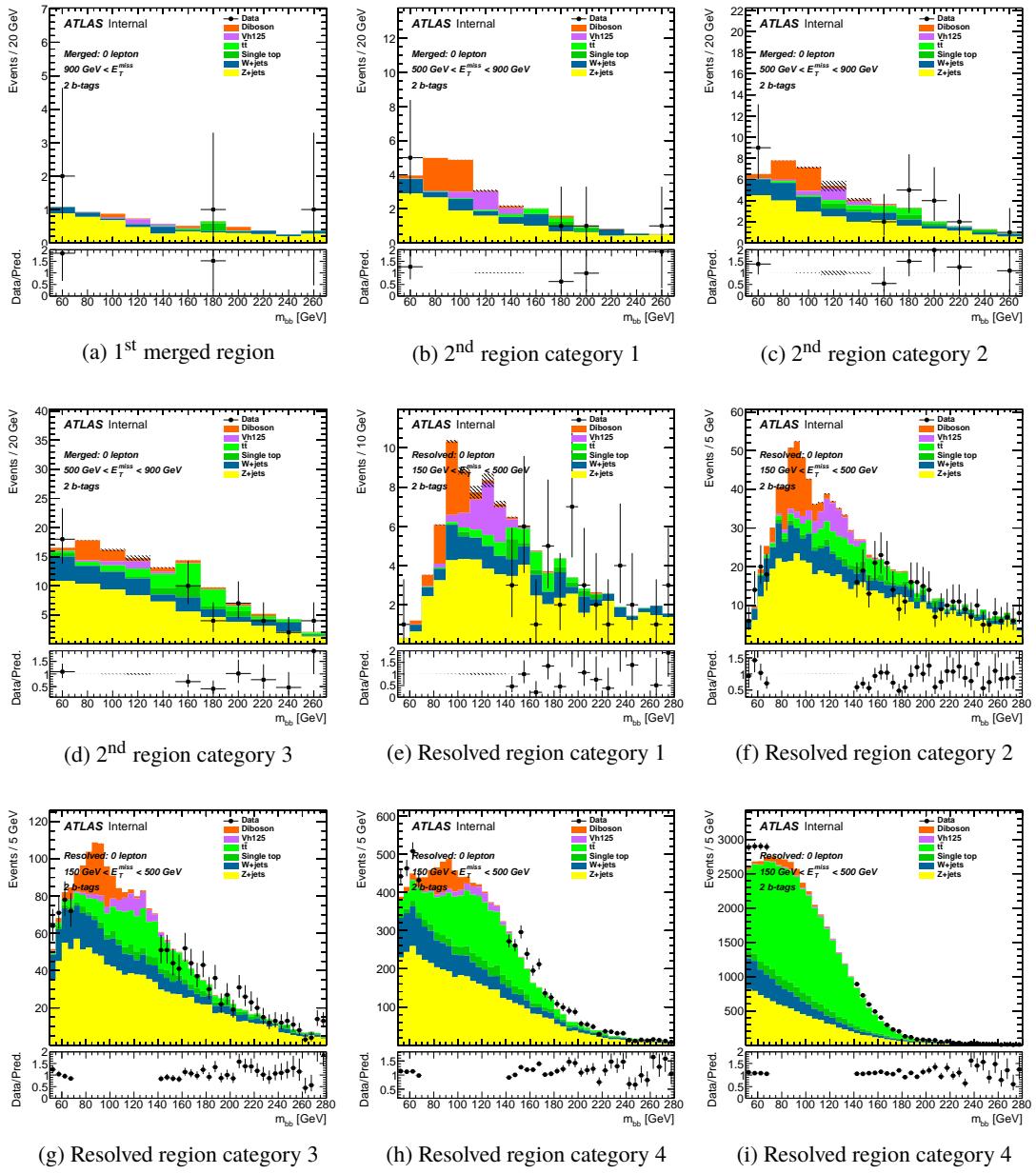


Figure 34: The pre-fit mass spectrum in each category.

Not reviewed, for internal circulation only

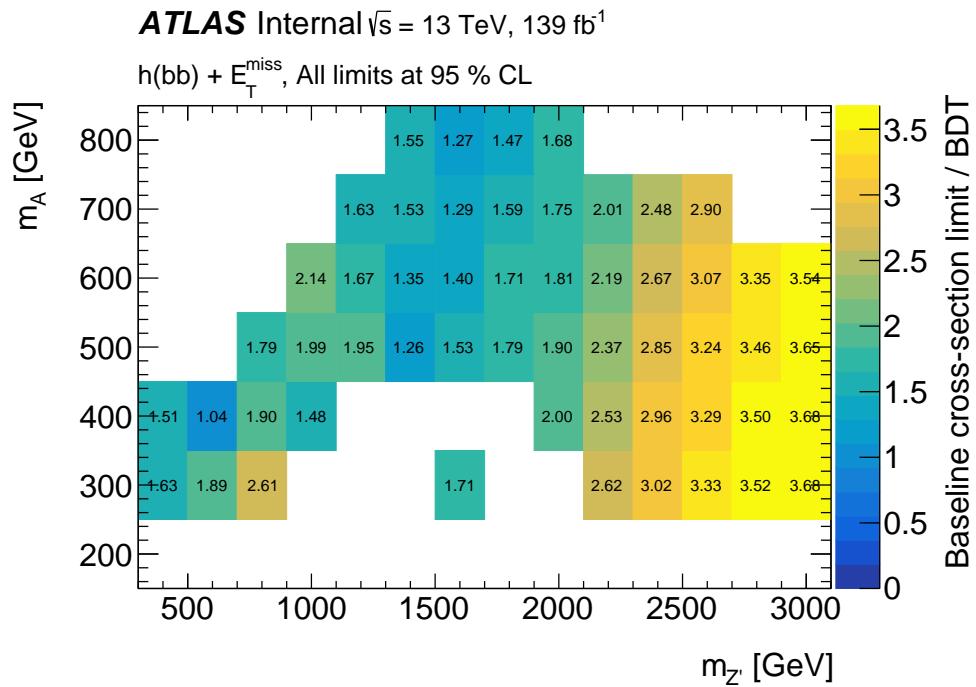


Figure 35: The ratio of the expected upper limit of the cross-section between the baseline cut-based approach and the BDT approach for each Z'2HDM signal point, which represents the improvement of the BDT approach over the cut-based.

Not reviewed, for internal circulation only

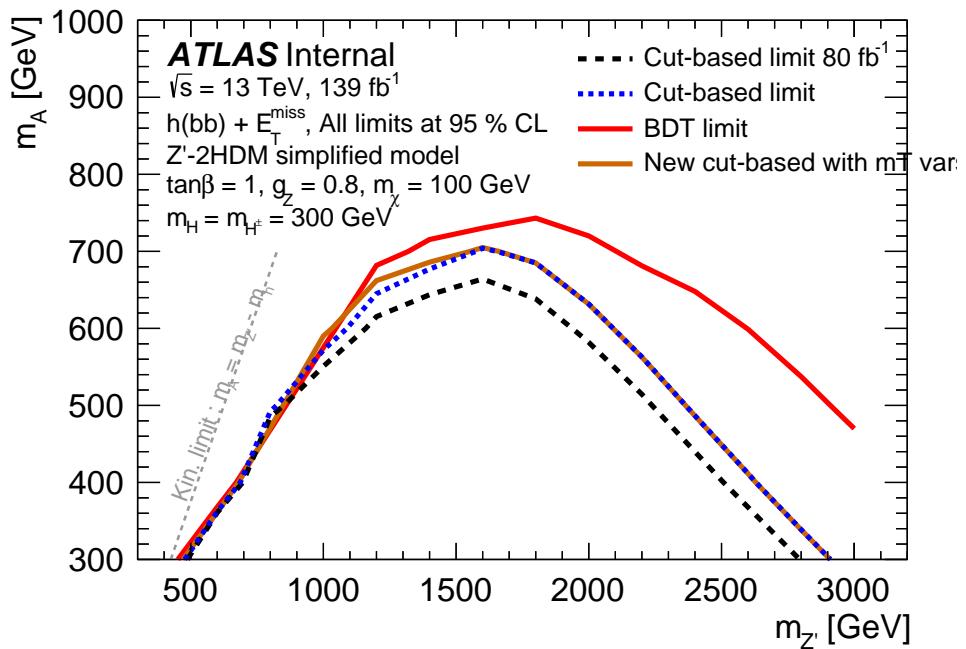


Figure 36: The expected exclusion limit contour of the Z'2HDM model with the baseline cut-based approach, the BDT approach and the new optimized cut-based approach. Note that the expected limit contour with the BDT approach has exceeded the current signal grid, resulting a straight falling line from $m_{Z'} = 700 \text{ GeV}$ to $m_{Z'} = 1200 \text{ GeV}$.

1346 D Higgs tagging [Jue, Ines, Dan]

1347 To be filled properly

1348 The latest update can be found at: [link](#).

1349 E MV2 vs DL1r [Veronica, Janik]

1350 To be filled properly

1351 The latest updates can be found at:

- 1352 • [link 1](#)

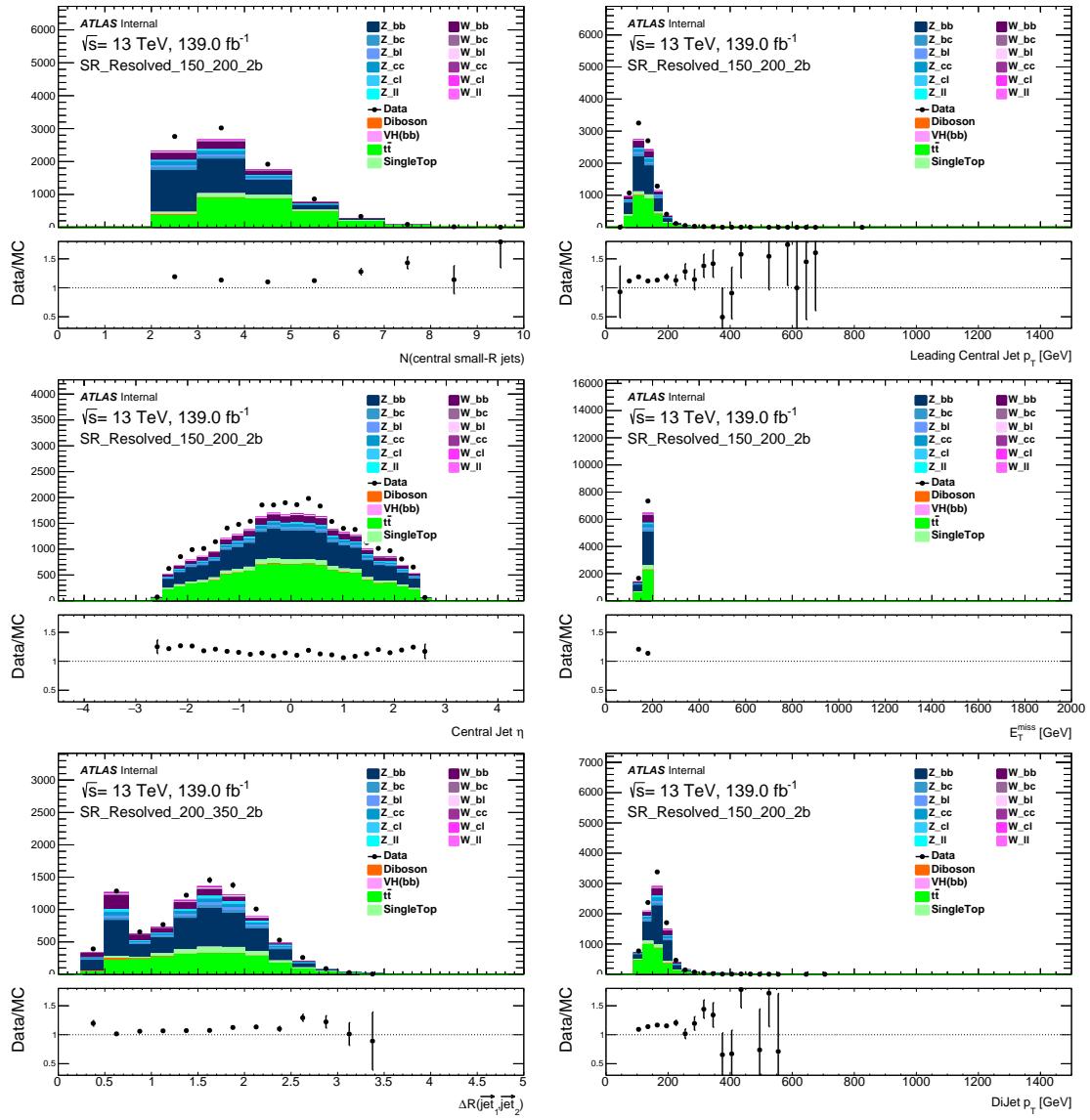
- 1353 • [link 2](#)

Not reviewed, for internal circulation only

1354 F Pre-fit results [Anindya]

1355 F.1 0-lepton [Anindya]

Not reviewed, for internal circulation only

Figure 37: Kinematic distributions in 0-lepton channel with 2 b-tagged jets and E_T^{miss} between 150 and 200 GeV.

Not reviewed, for internal circulation only

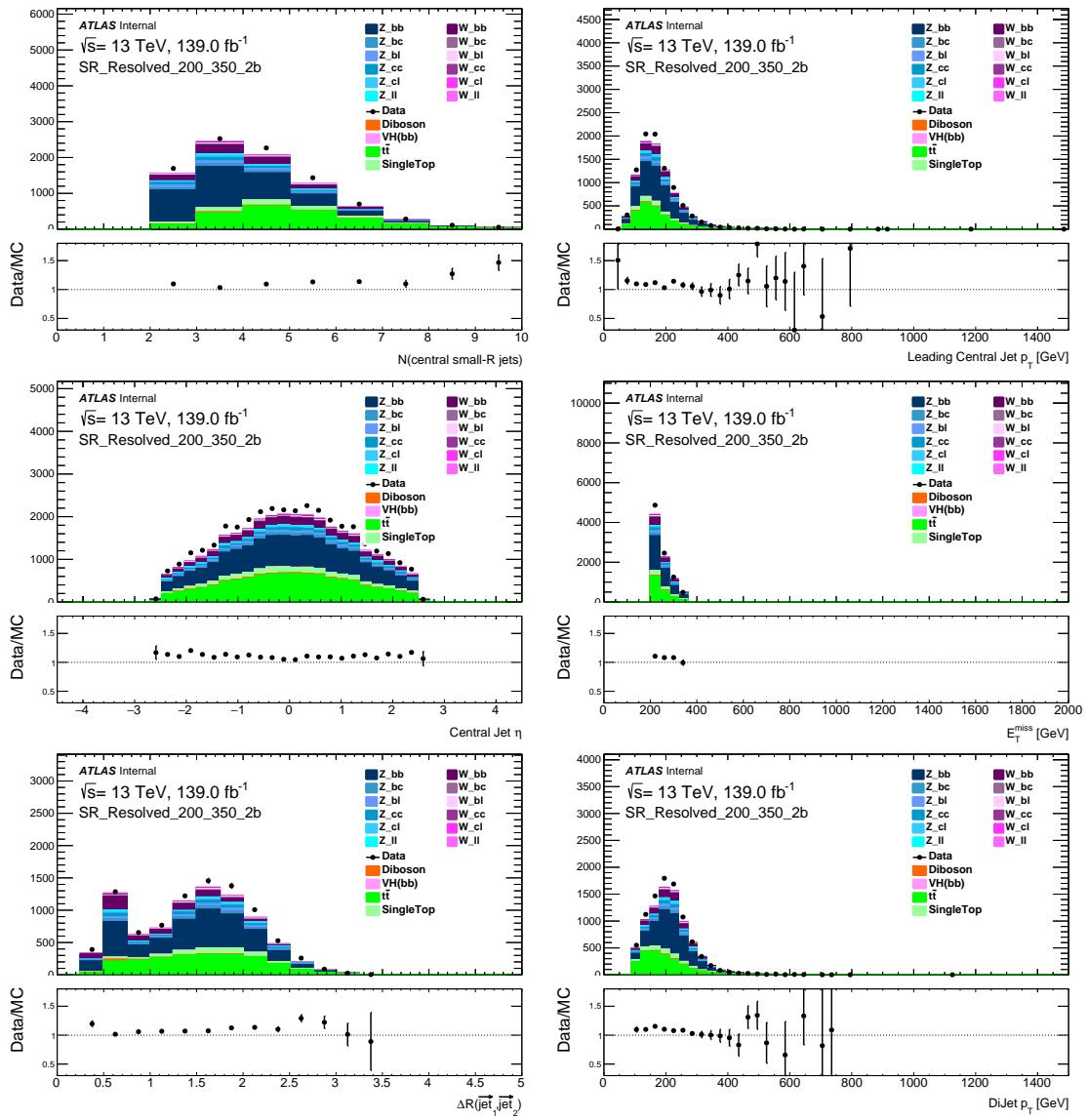


Figure 38: Kinematic distributions in 0-lepton channel with 2 b-tagged jets and E_T^{miss} between 200 and 350 GeV.

Not reviewed, for internal circulation only

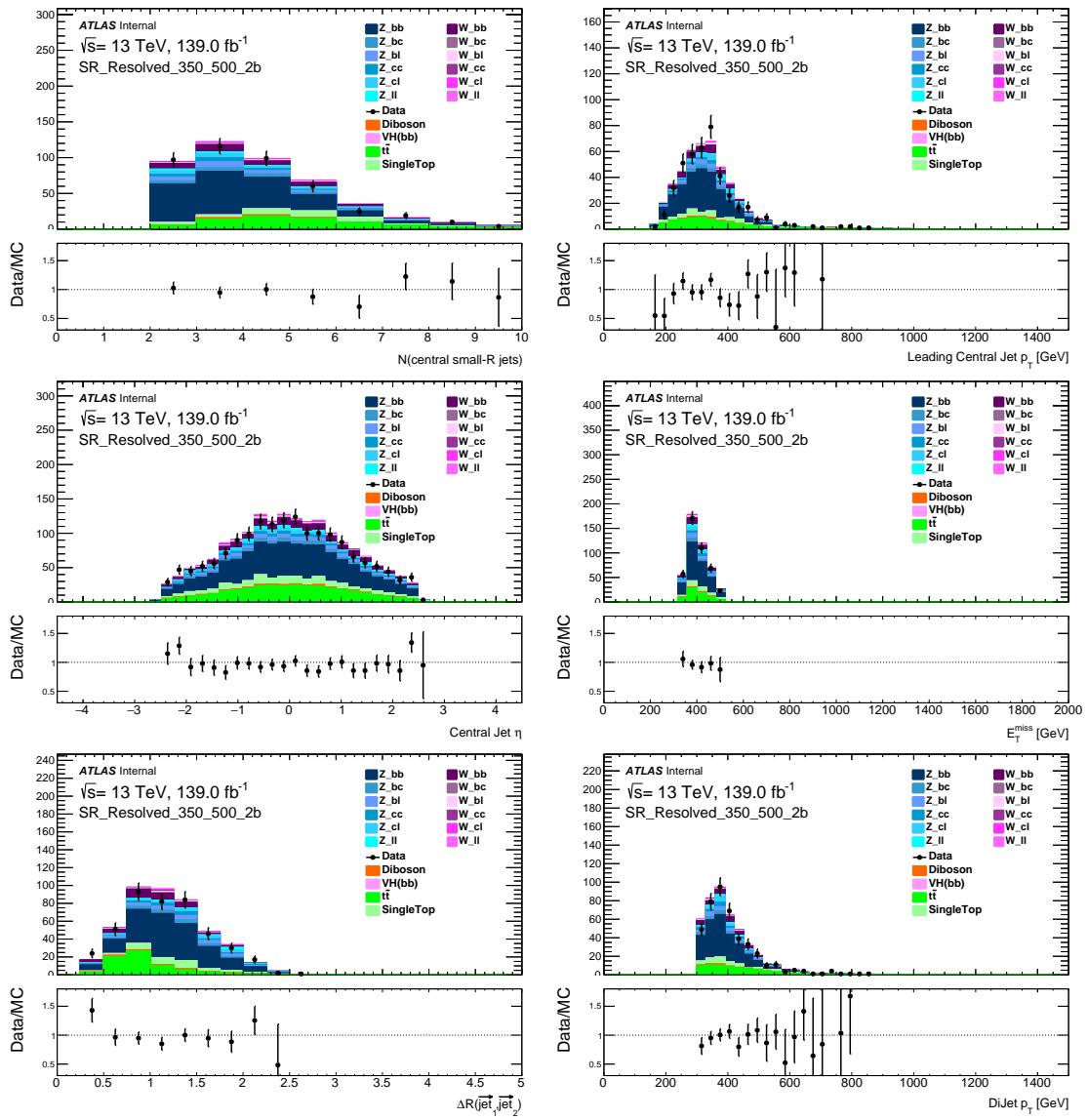


Figure 39: Kinematic distributions in 0-lepton channel with 2 b-tagged jets and E_T^{miss} between 350 and 500 GeV.

Not reviewed, for internal circulation only

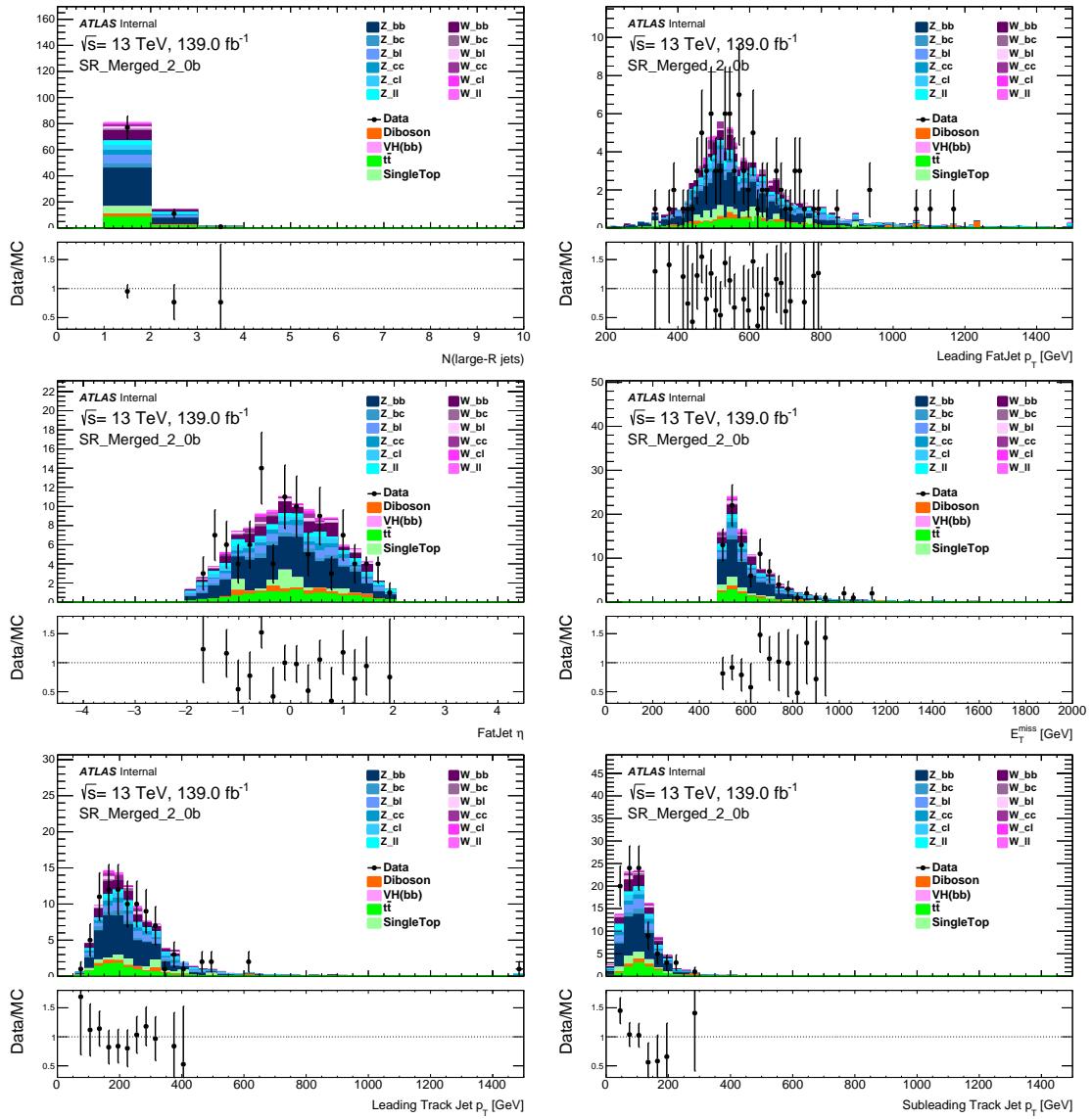


Figure 40: Kinematic distributions in 0-lepton channel with 2 b-tagged jets and E_T^{miss} above 500 GeV.

Not reviewed, for internal circulation only

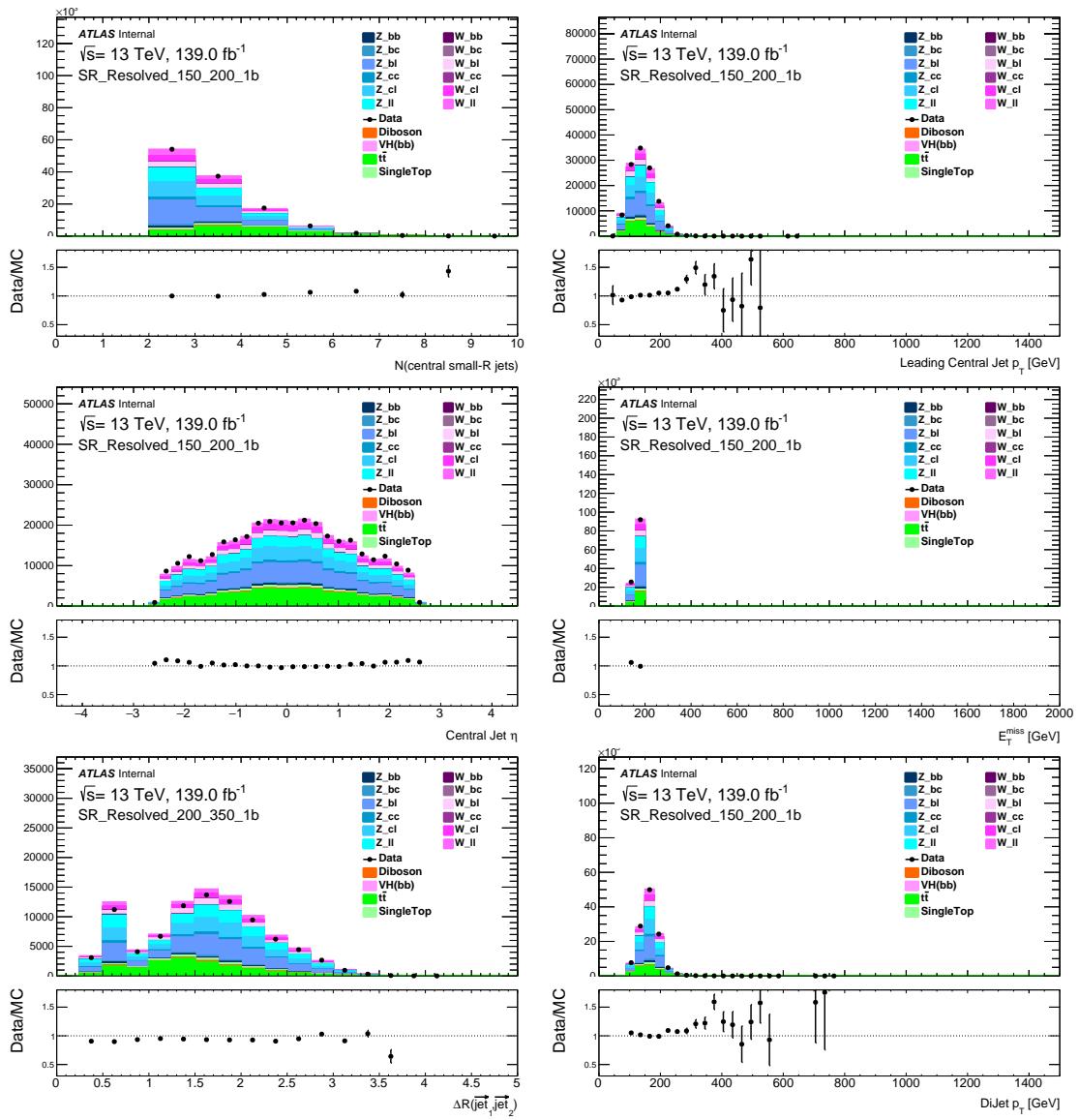


Figure 41: Kinematic distributions in 0-lepton channel with 1 b-tagged jets and E_T^{miss} between 150 and 200 GeV.

Not reviewed, for internal circulation only

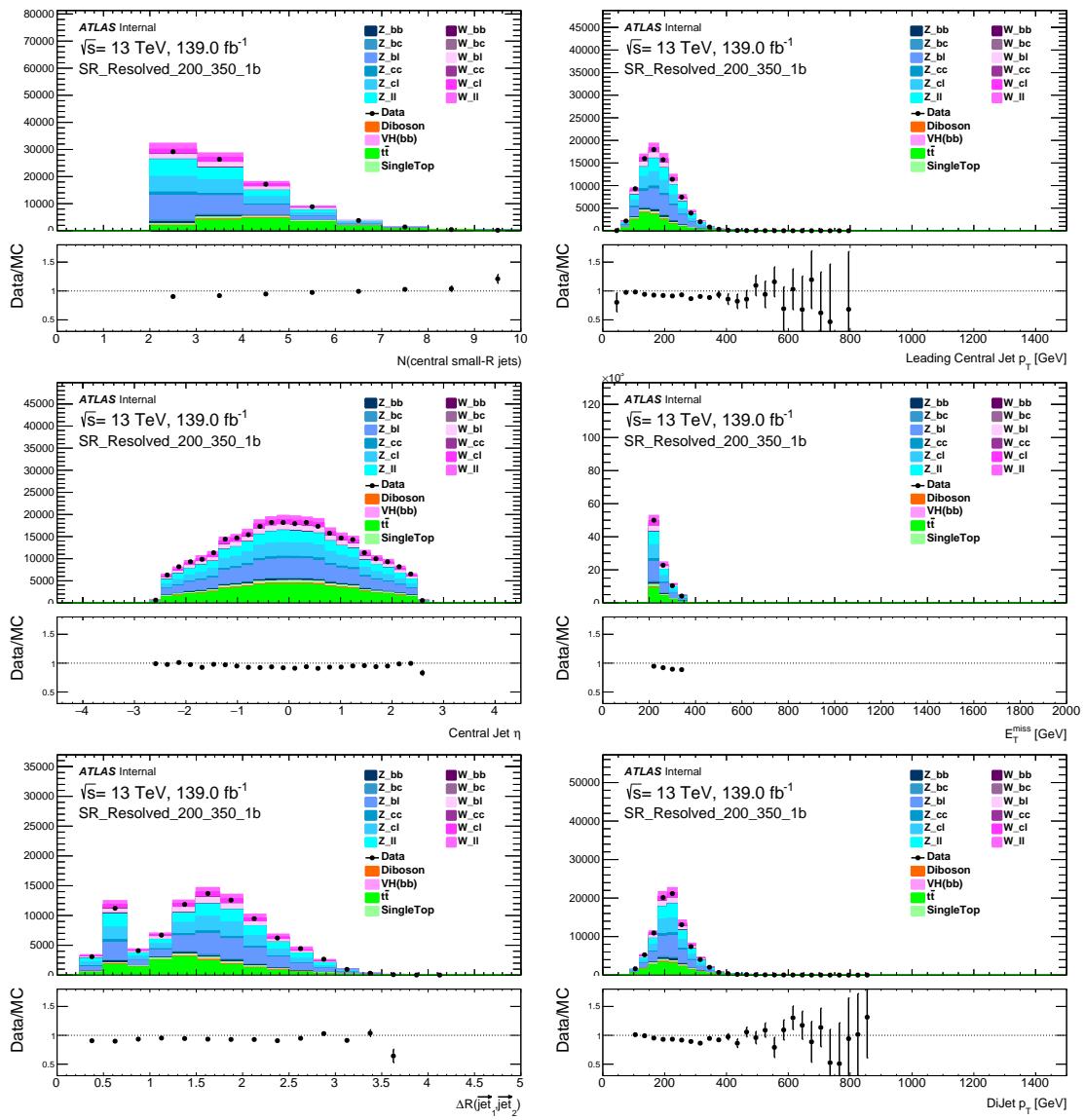


Figure 42: Kinematic distributions in 0-lepton channel with 1 b-tagged jets and E_T^{miss} between 200 and 350 GeV.

Not reviewed, for internal circulation only

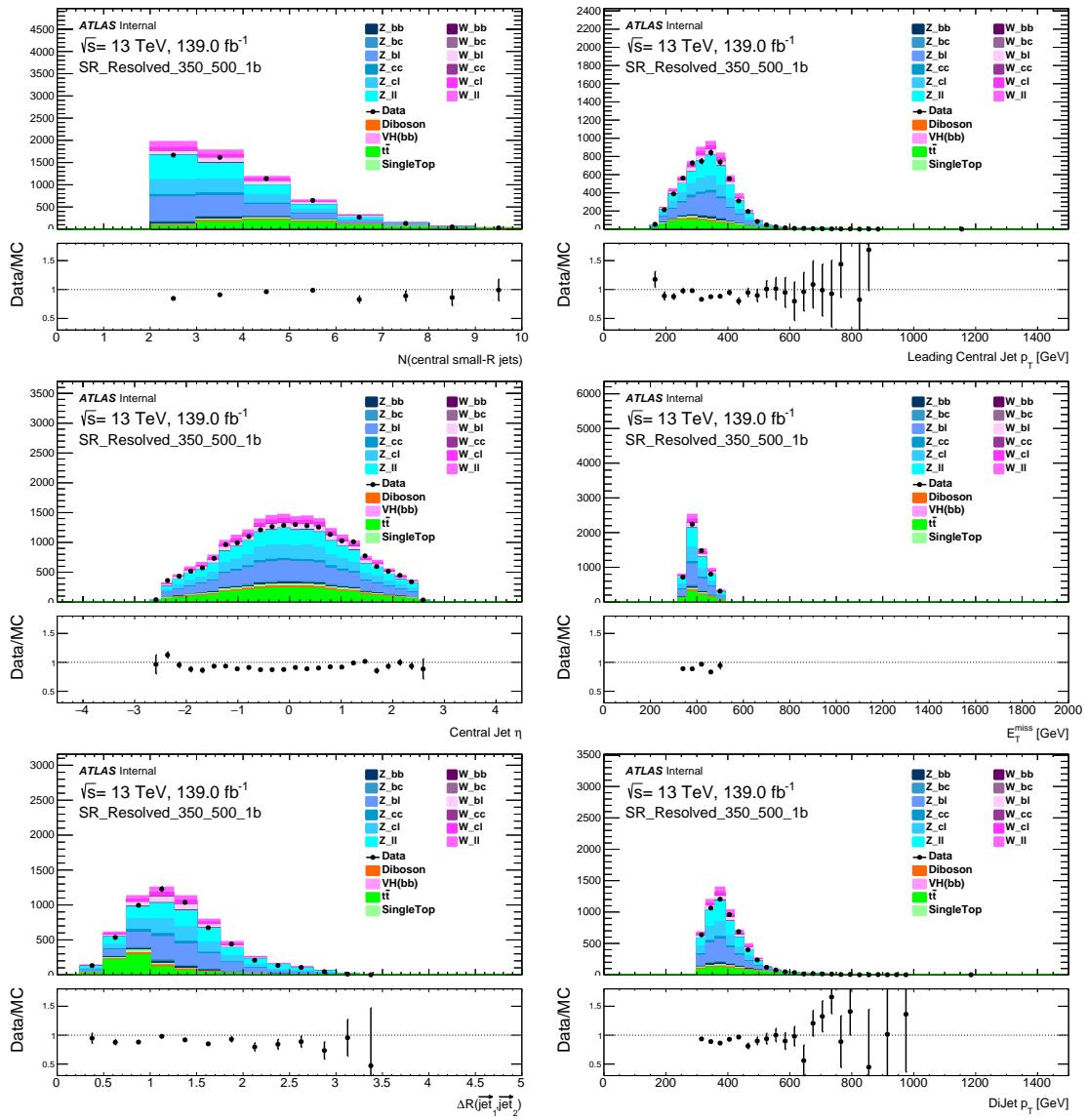


Figure 43: Kinematic distributions in 0-lepton channel with 1 b-tagged jets and E_T^{miss} between 350 and 500 GeV.

Not reviewed, for internal circulation only

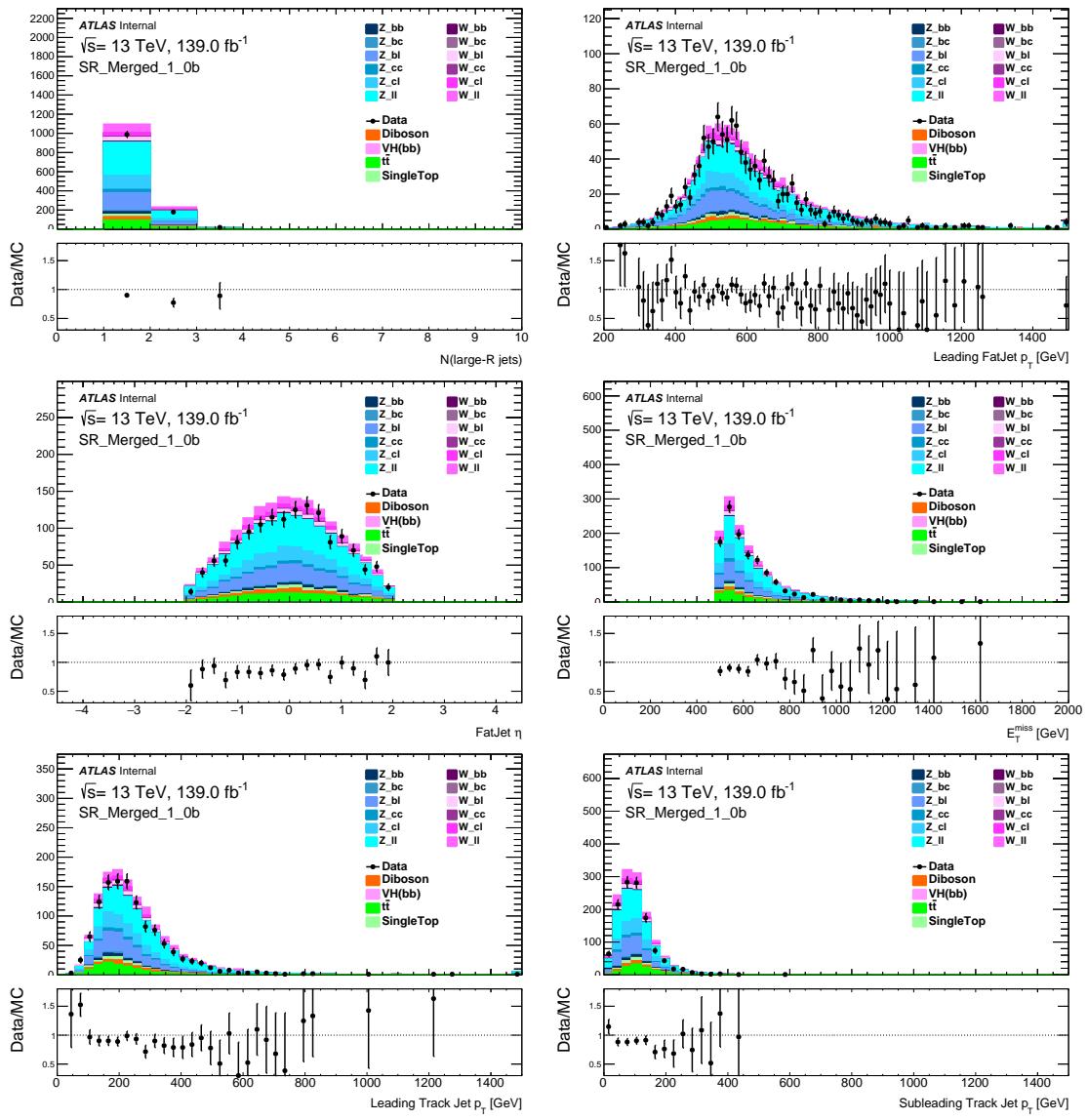
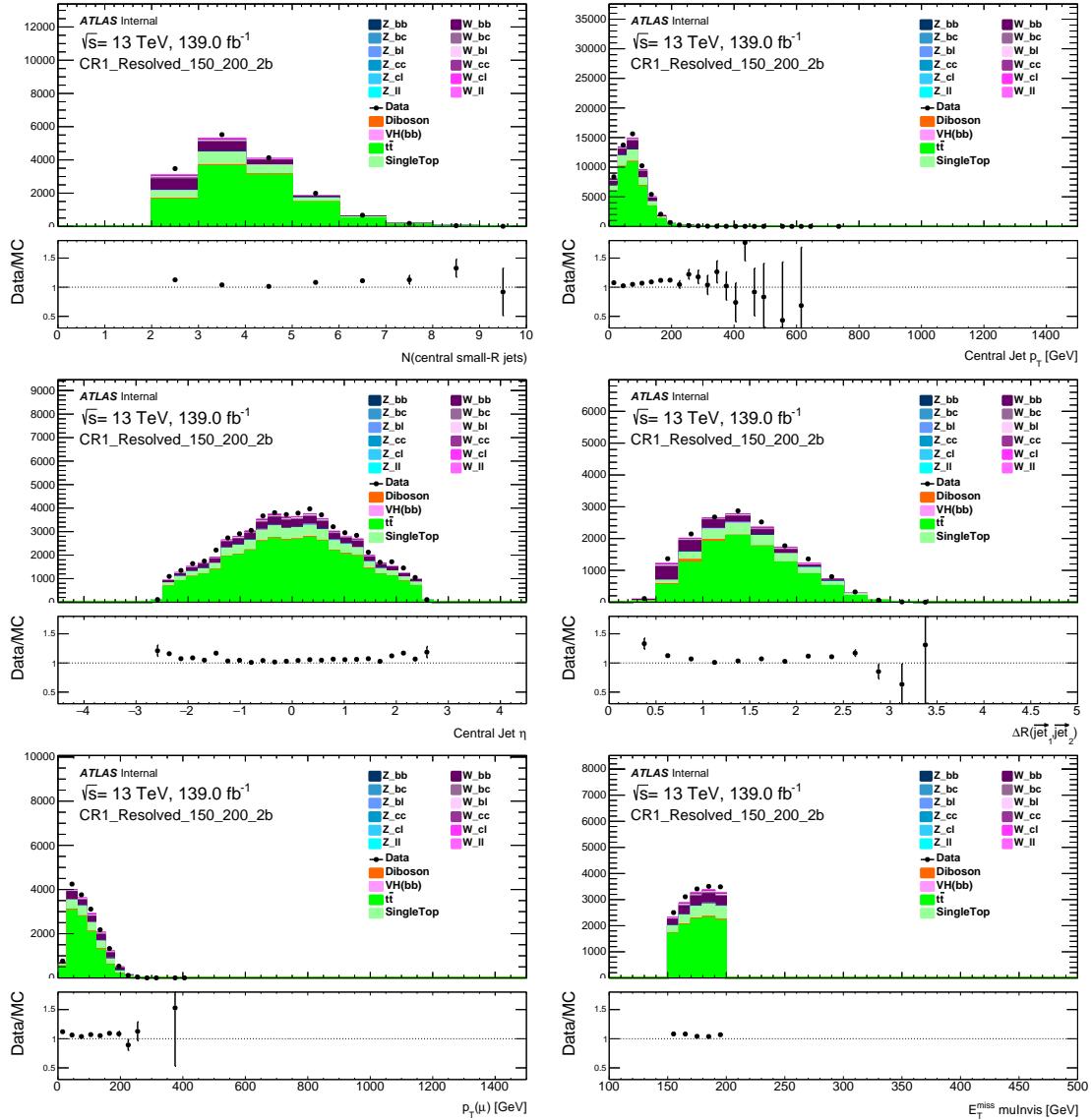


Figure 44: Kinematic distributions in 0-lepton channel with 1 b-tagged jets and E_T^{miss} above 500 GeV.

1356 F.2 1-lepton [Anindya]

Not reviewed, for internal circulation only

Figure 45: Kinematic distributions in 1-muon CR with 2 b-tagged jets and E_T^{miss} between 150 and 200 GeV.

Not reviewed, for internal circulation only

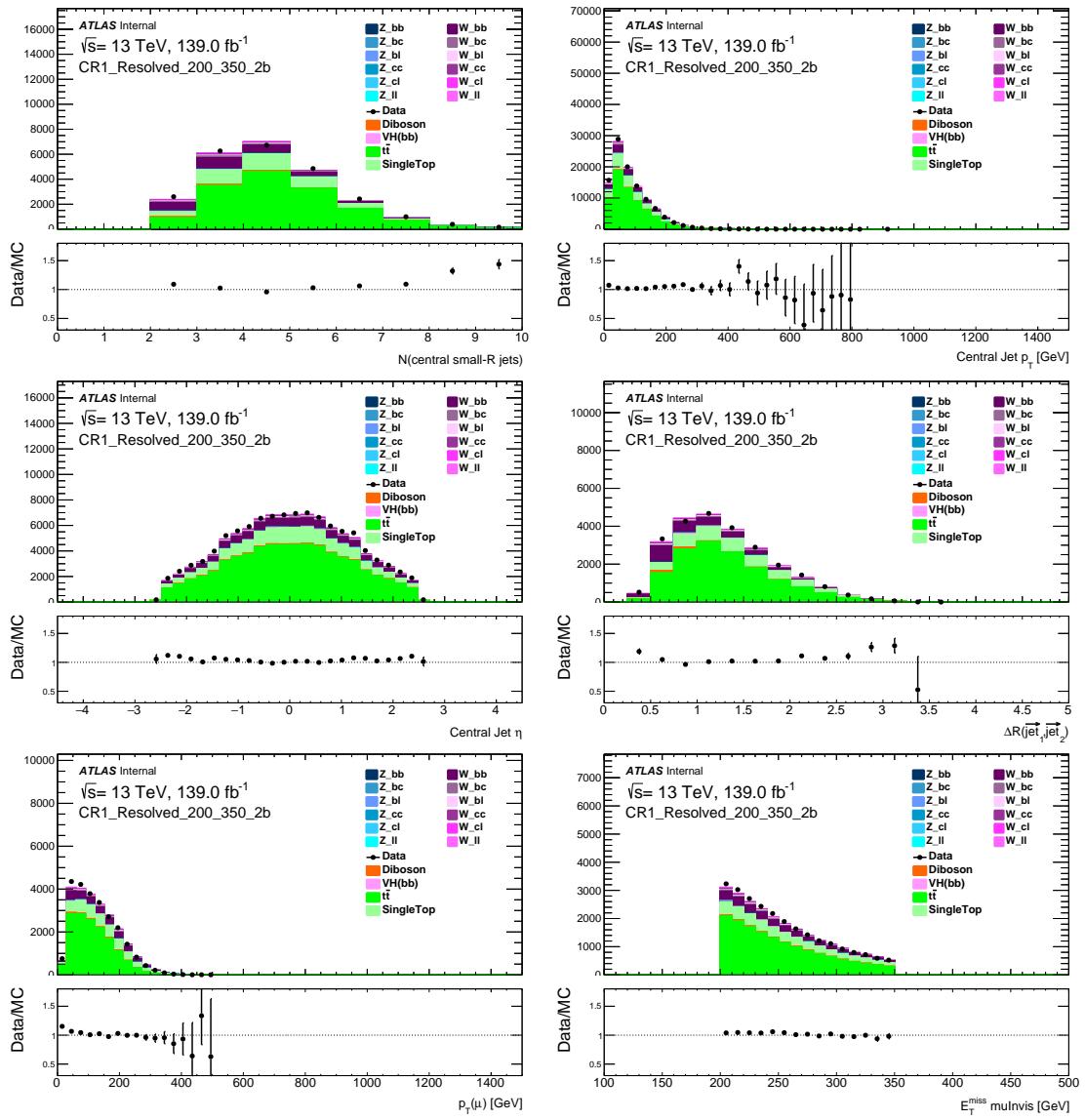


Figure 46: Kinematic distributions in 1-muon CR with 2 b-tagged jets and E_T^{miss} between 200 and 350 GeV.

Not reviewed, for internal circulation only

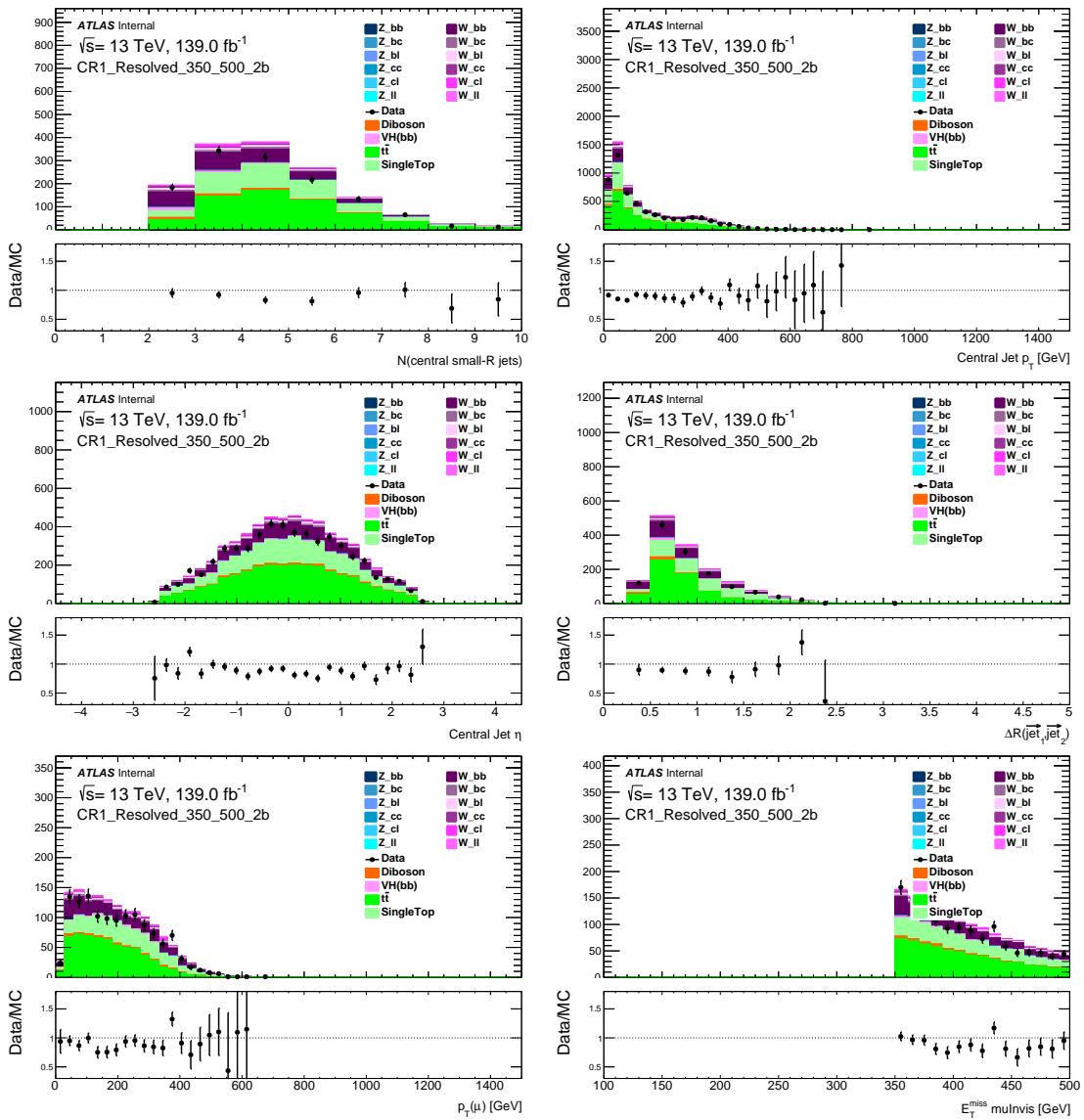


Figure 47: Kinematic distributions in 1-muon CR with 2 b-tagged jets and E_T^{miss} between 350 and 500 GeV.

Not reviewed, for internal circulation only

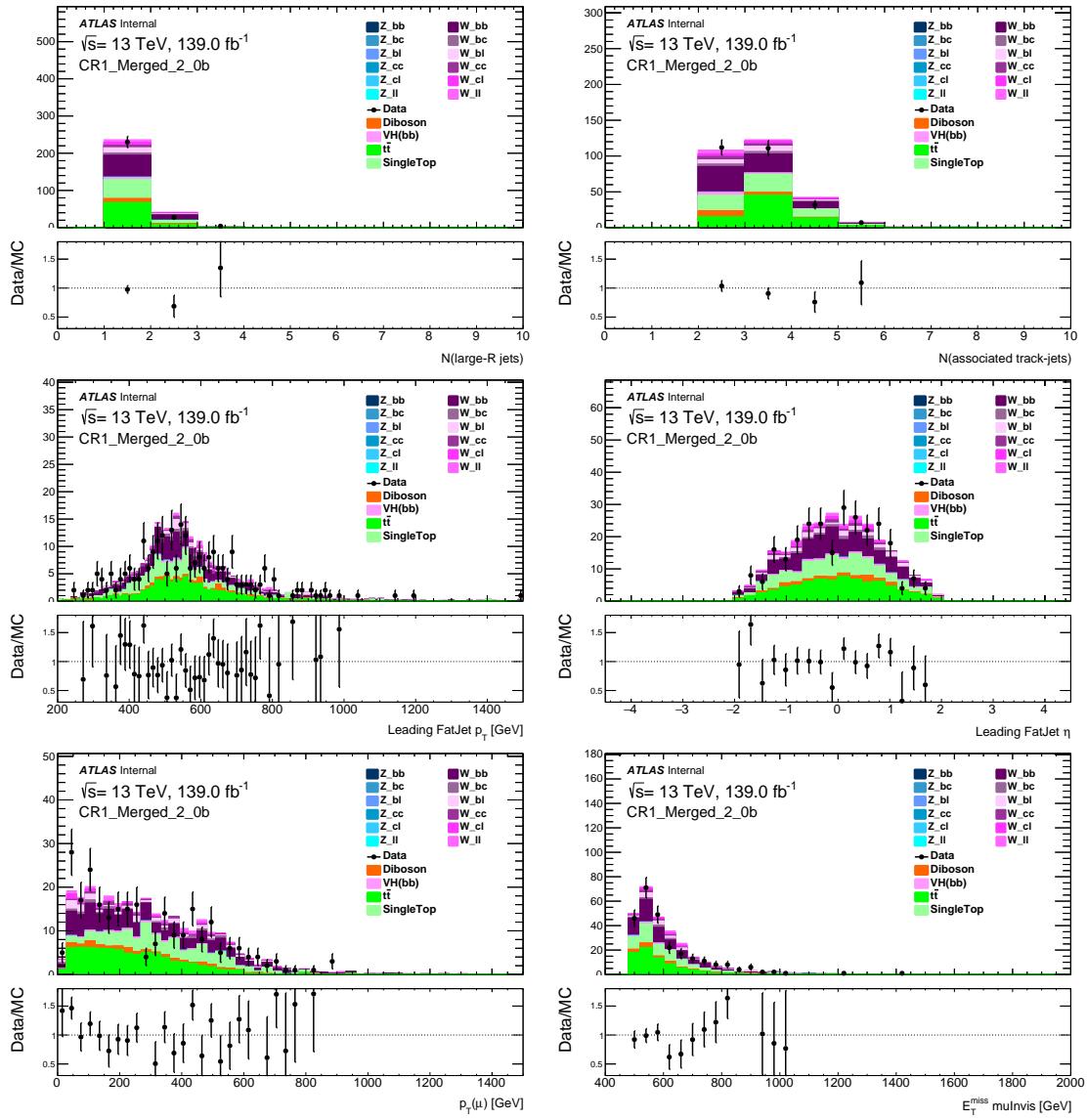


Figure 48: Kinematic distributions in 1-muon CR with a 2 b-tagged fatjet and E_T^{miss} above 500 GeV.

Not reviewed, for internal circulation only

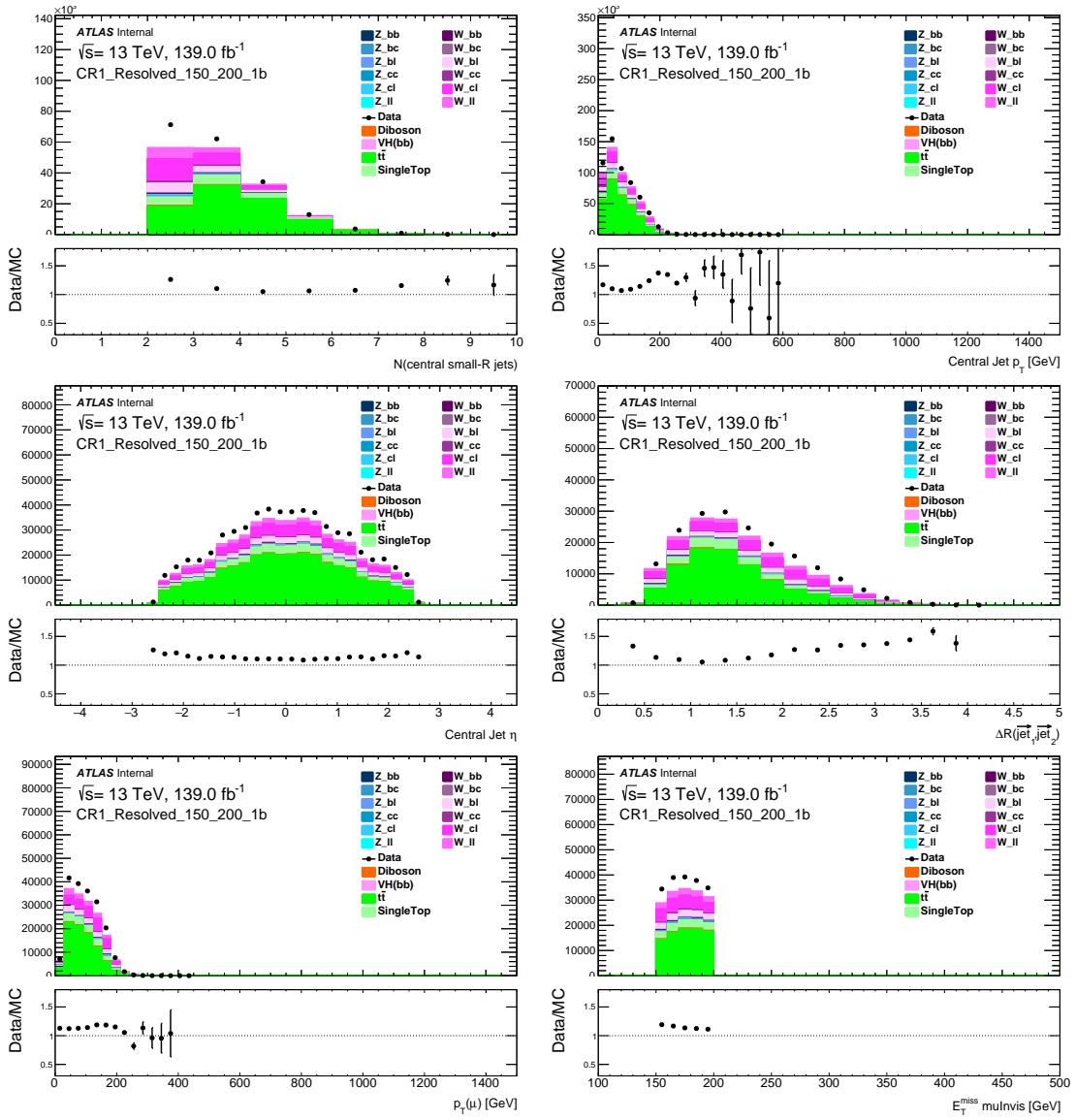


Figure 49: Kinematic distributions in 1-muon CR with 1 b-tagged jet and E_T^{miss} between 150 and 200 GeV.

Not reviewed, for internal circulation only

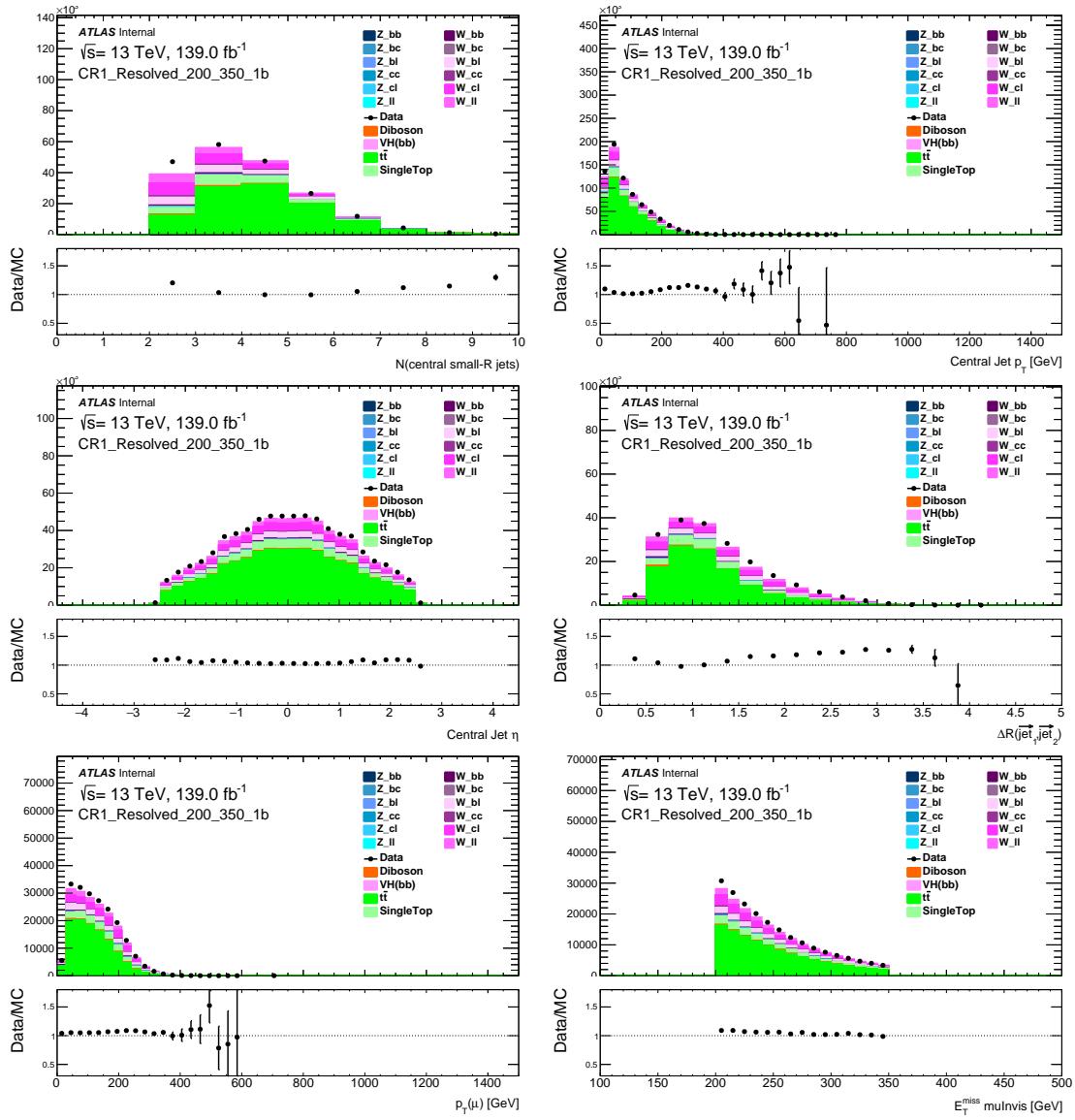


Figure 50: Kinematic distributions in 1-muon CR with 1 b-tagged jet and E_T^{miss} between 200 and 350 GeV.

Not reviewed, for internal circulation only

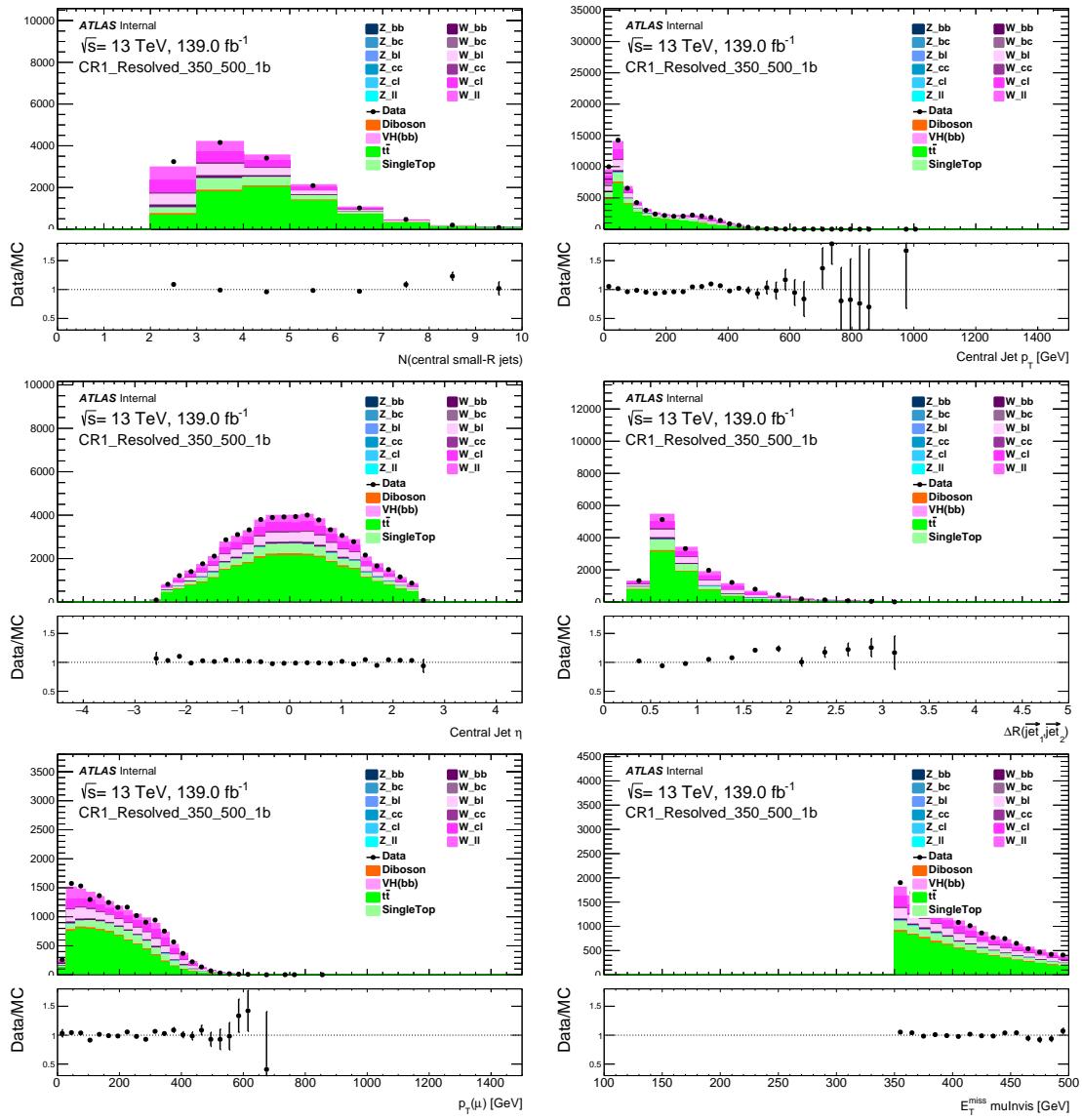


Figure 51: Kinematic distributions in 1-muon CR with 1 b-tagged jet and E_T^{miss} between 350 and 500 GeV.

Not reviewed, for internal circulation only

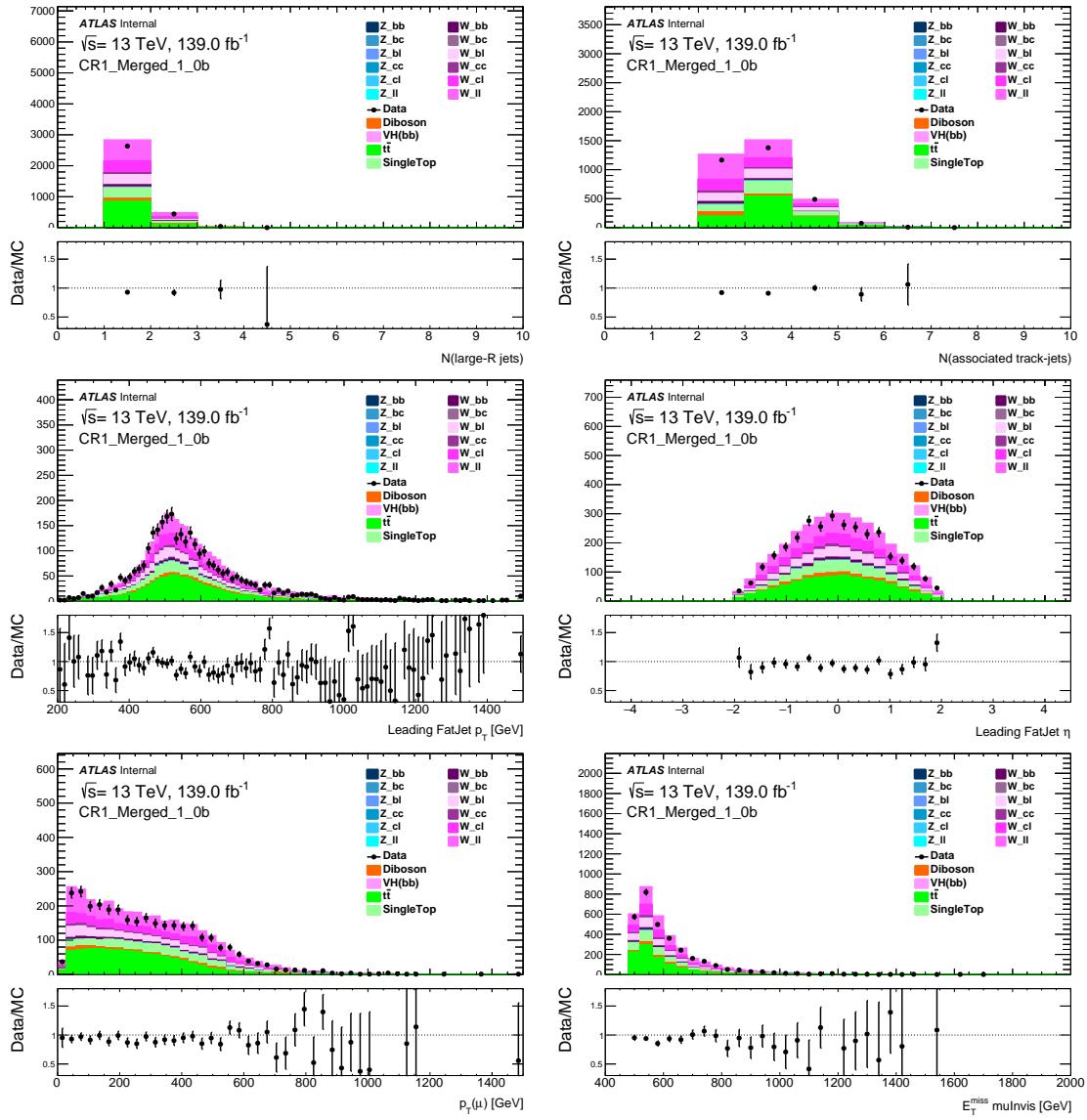
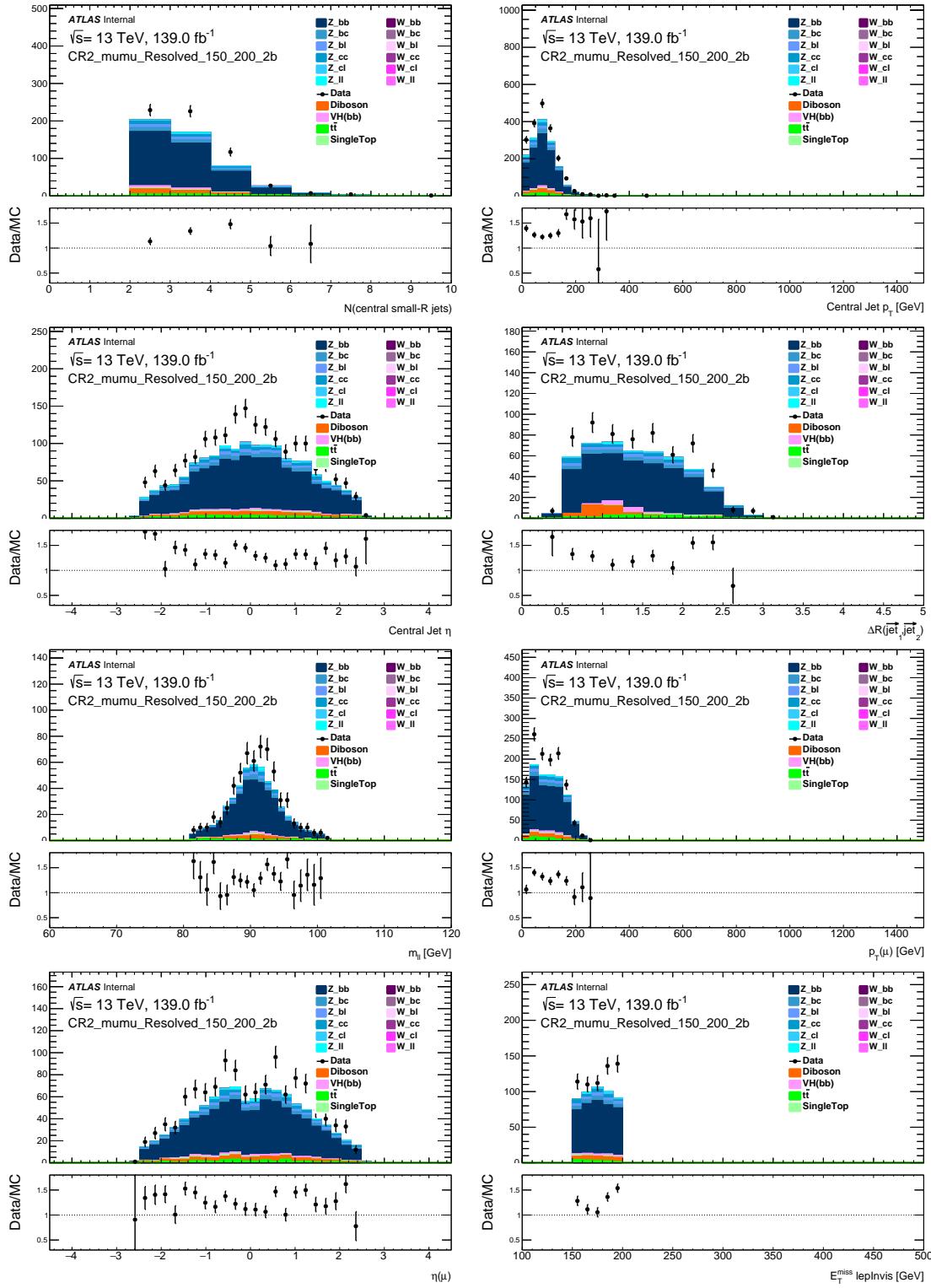


Figure 52: Kinematic distributions in 1-muon CR with a 1 b-tagged fatjet and E_T^{miss} above 500 GeV.

1357 F.3 2-lepton [Anindya]

Not reviewed, for internal circulation only

Figure 53: Kinematic distributions in 2-muon CR with 2 b-tagged jets and E_T^{miss} between 150 and 200 GeV.

Not reviewed, for internal circulation only

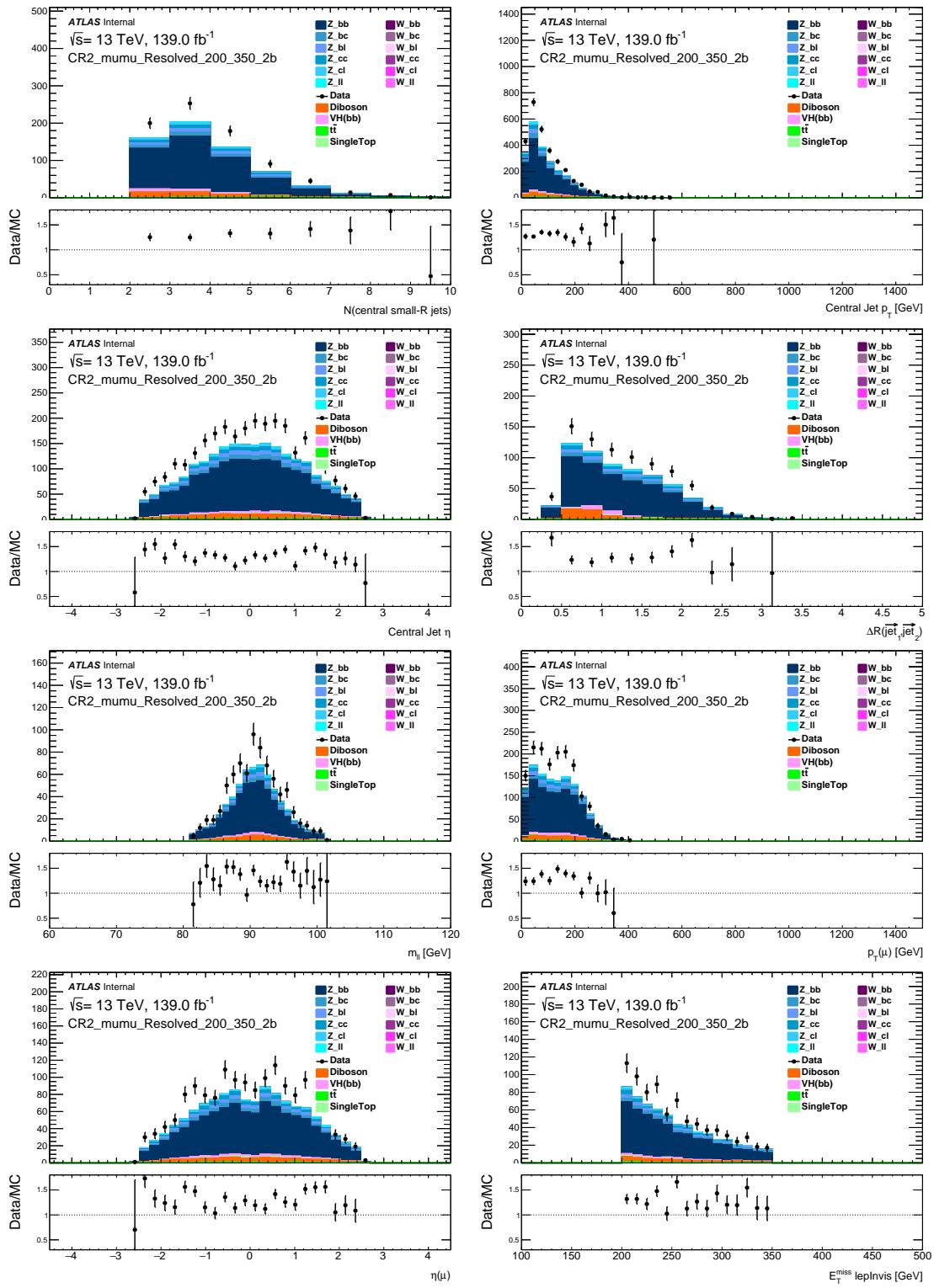


Figure 54: Kinematic distributions in 2-muon CR with 2 b-tagged jets and E_T^{miss} between 200 and 350 GeV.

Not reviewed, for internal circulation only

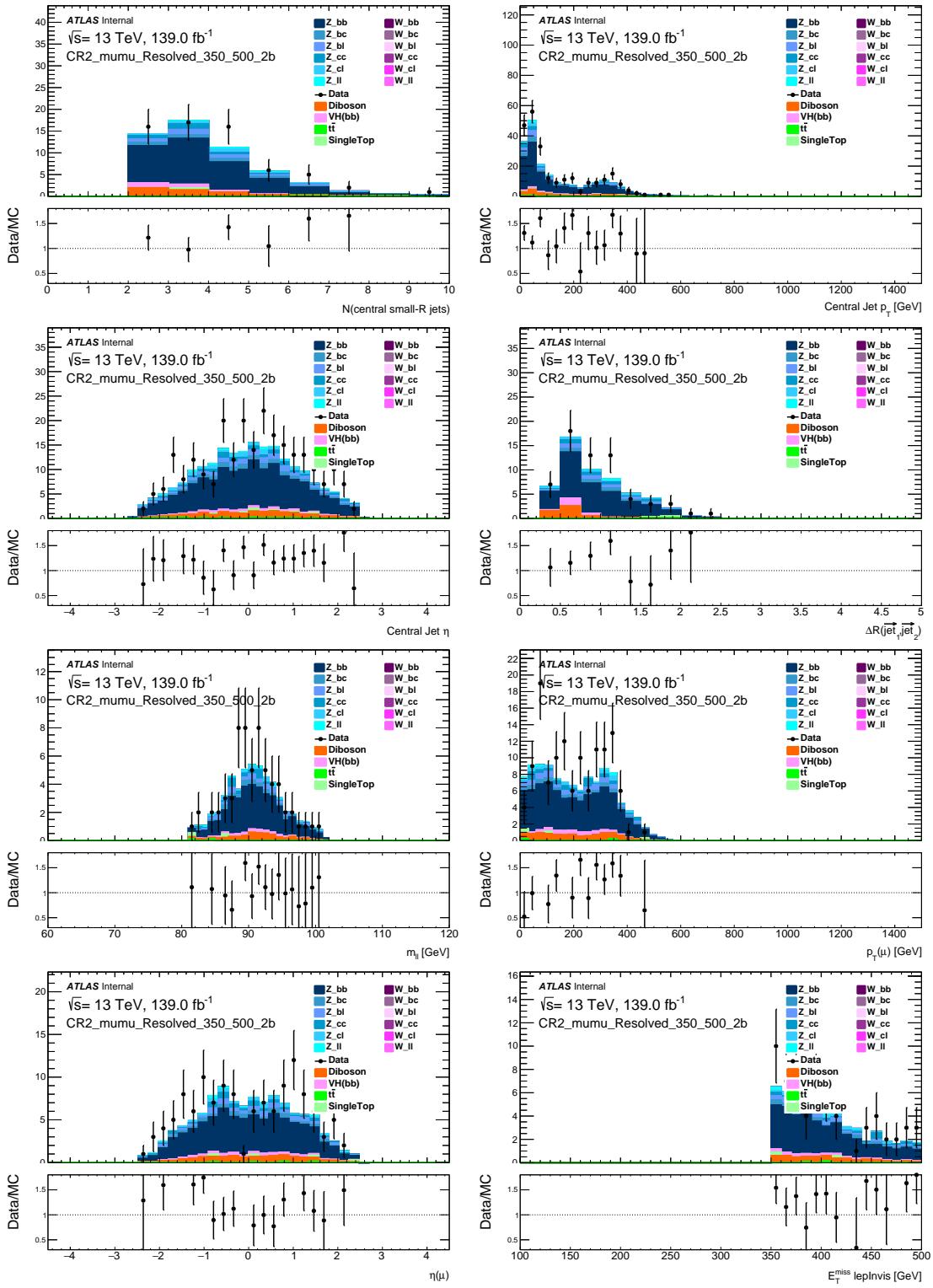


Figure 55: Kinematic distributions in 2-muon CR with 2 b-tagged jets and E_T^{miss} between 350 and 500 GeV.

Not reviewed, for internal circulation only

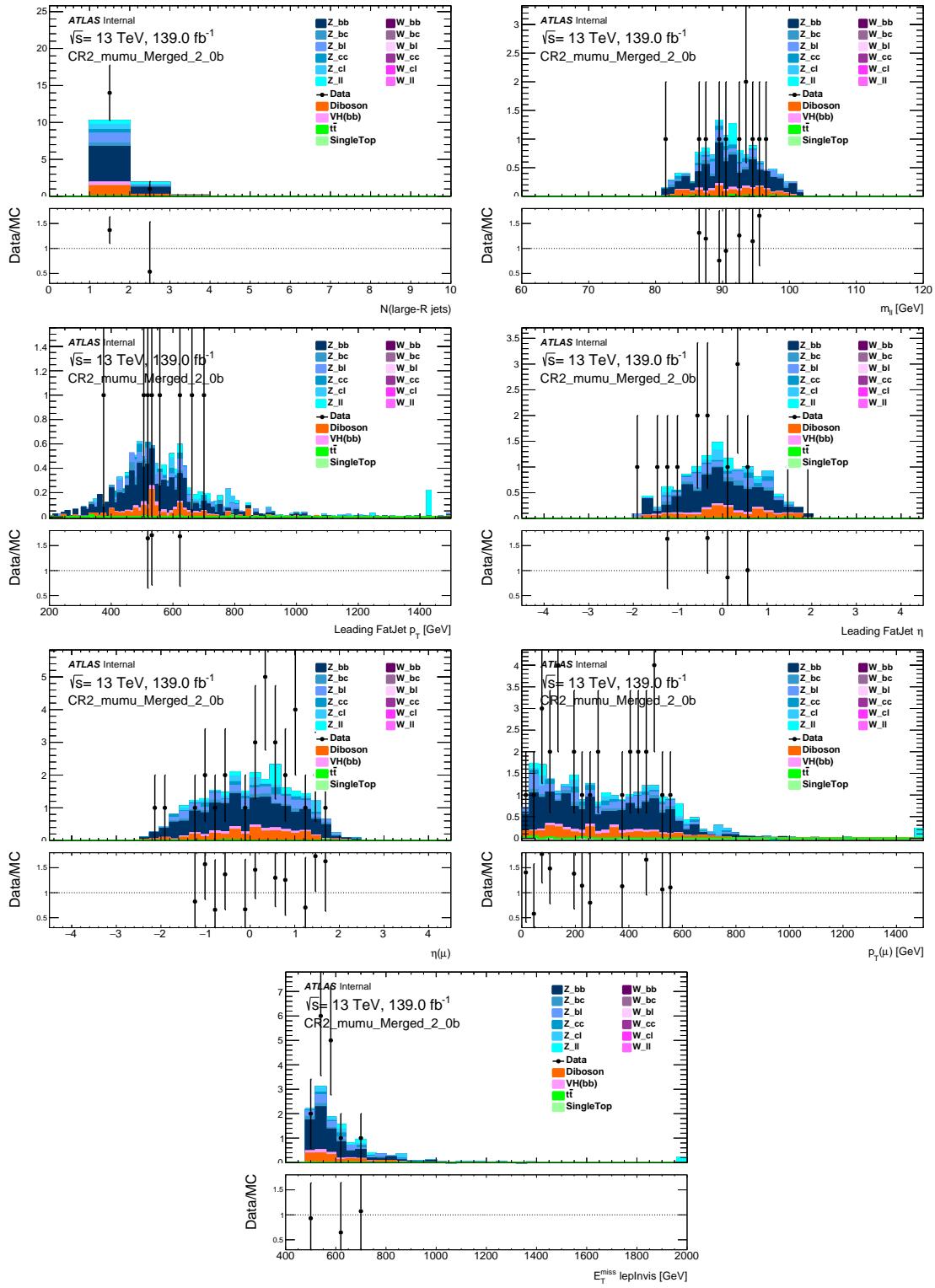


Figure 56: Kinematic distributions in 2-muon CR with a 2 b-tagged fatjet and E_T^{miss} above 500 GeV.

Not reviewed, for internal circulation only

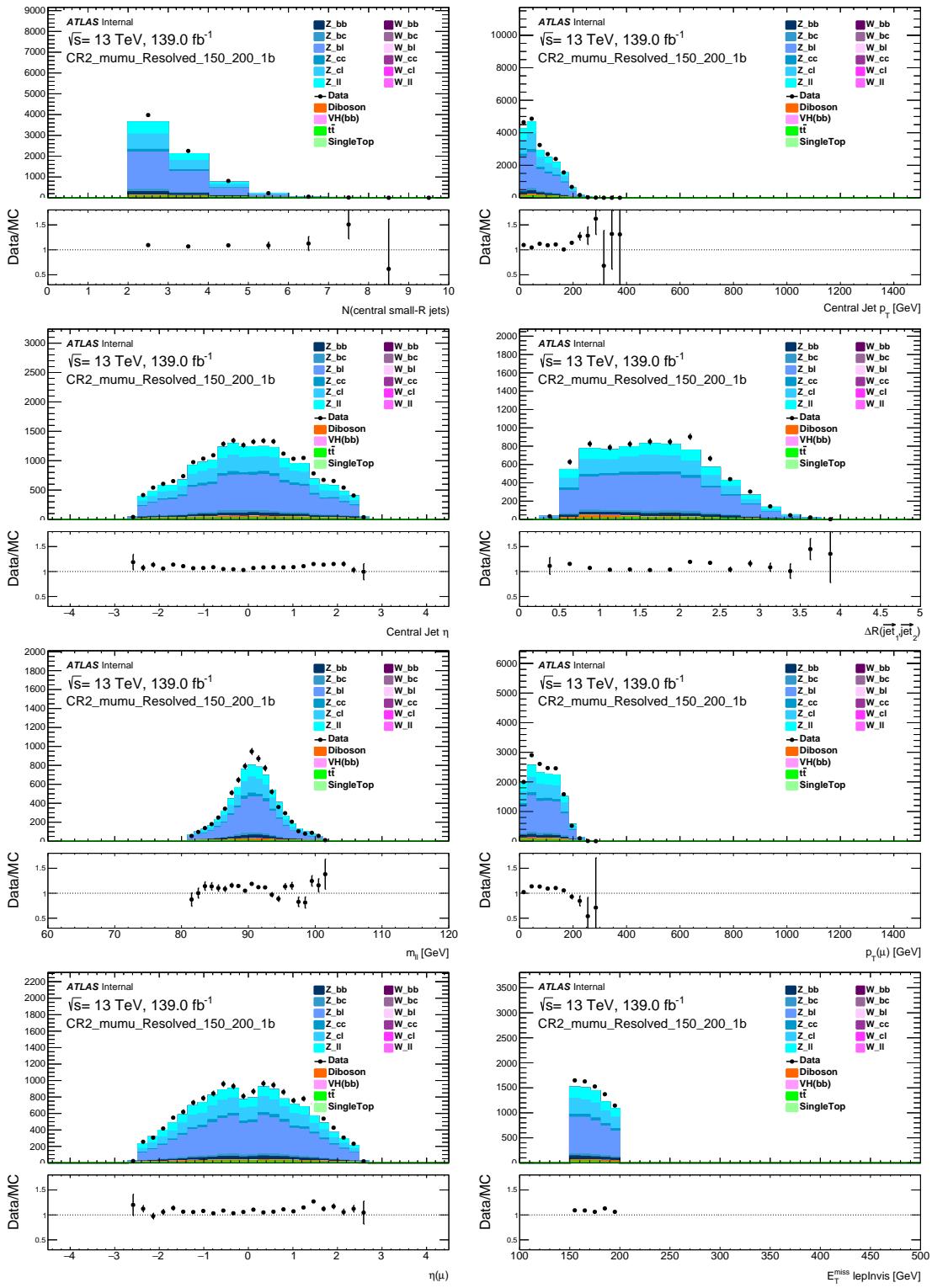


Figure 57: Kinematic distributions in 2-muon CR with 1 b-tagged jet and E_T^{miss} between 150 and 200 GeV

Not reviewed, for internal circulation only

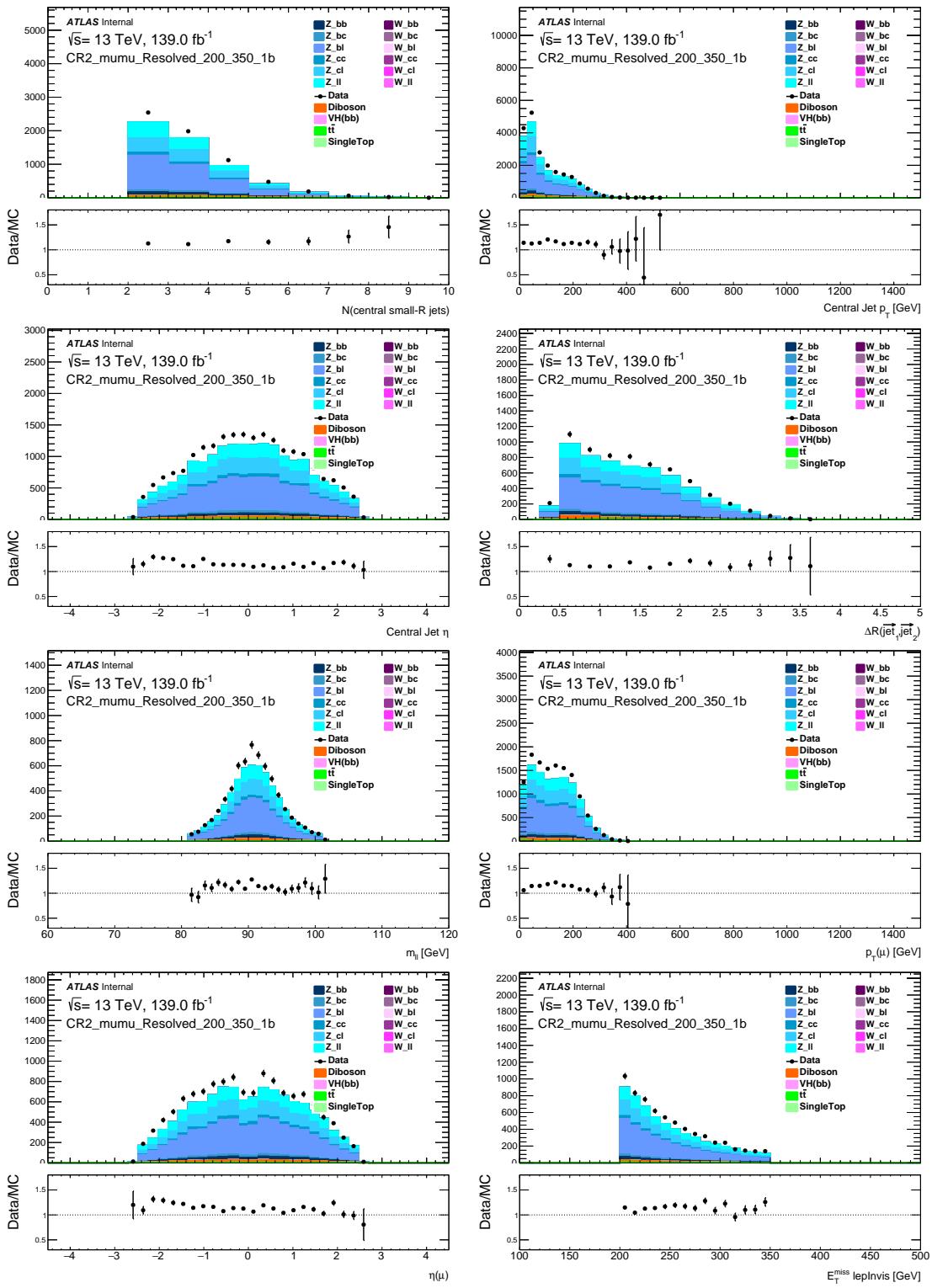


Figure 58: Kinematic distributions in 2-muon CR with 1 b-tagged jet and E_T^{miss} between 200 and 350 GeV

Not reviewed, for internal circulation only

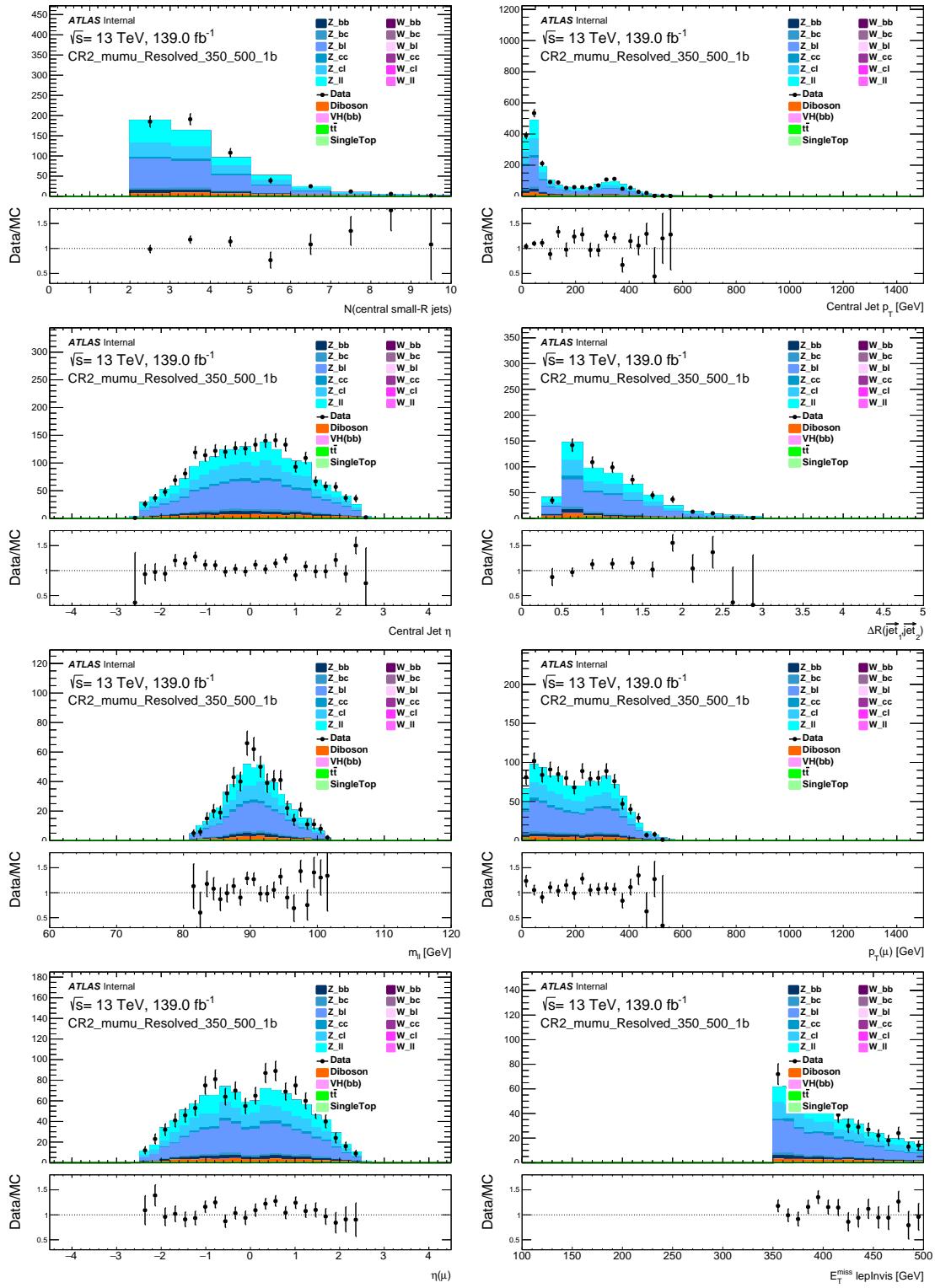


Figure 59: Kinematic distributions in 2-muon CR with 1 b-tagged jet and E_T^{miss} between 350 and 500 GeV

Not reviewed, for internal circulation only

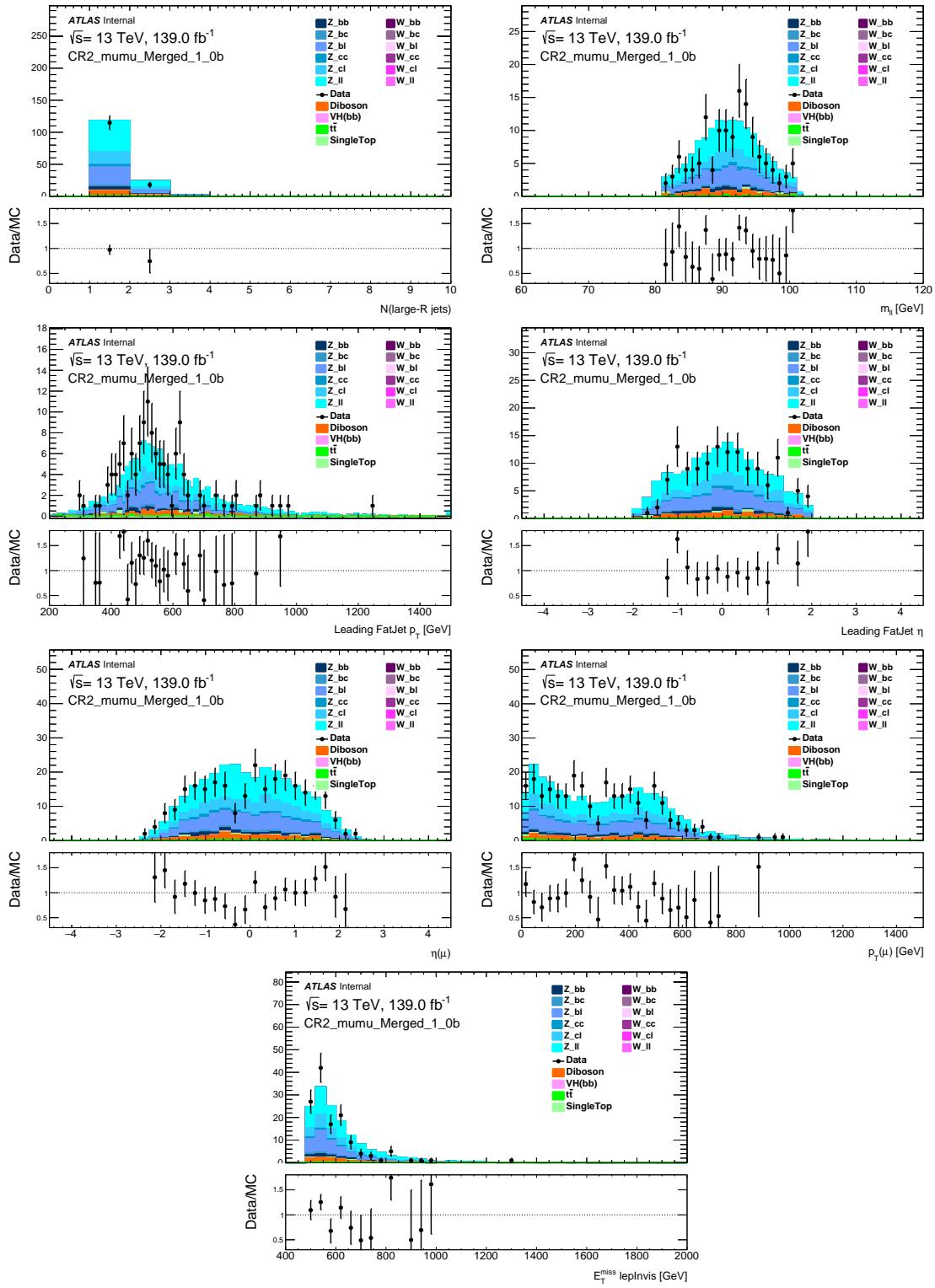


Figure 60: Kinematic distributions in 2-muon CR with a 1 b-tagged fatjet and E_T^{miss} above 500 GeV

Not reviewed, for internal circulation only

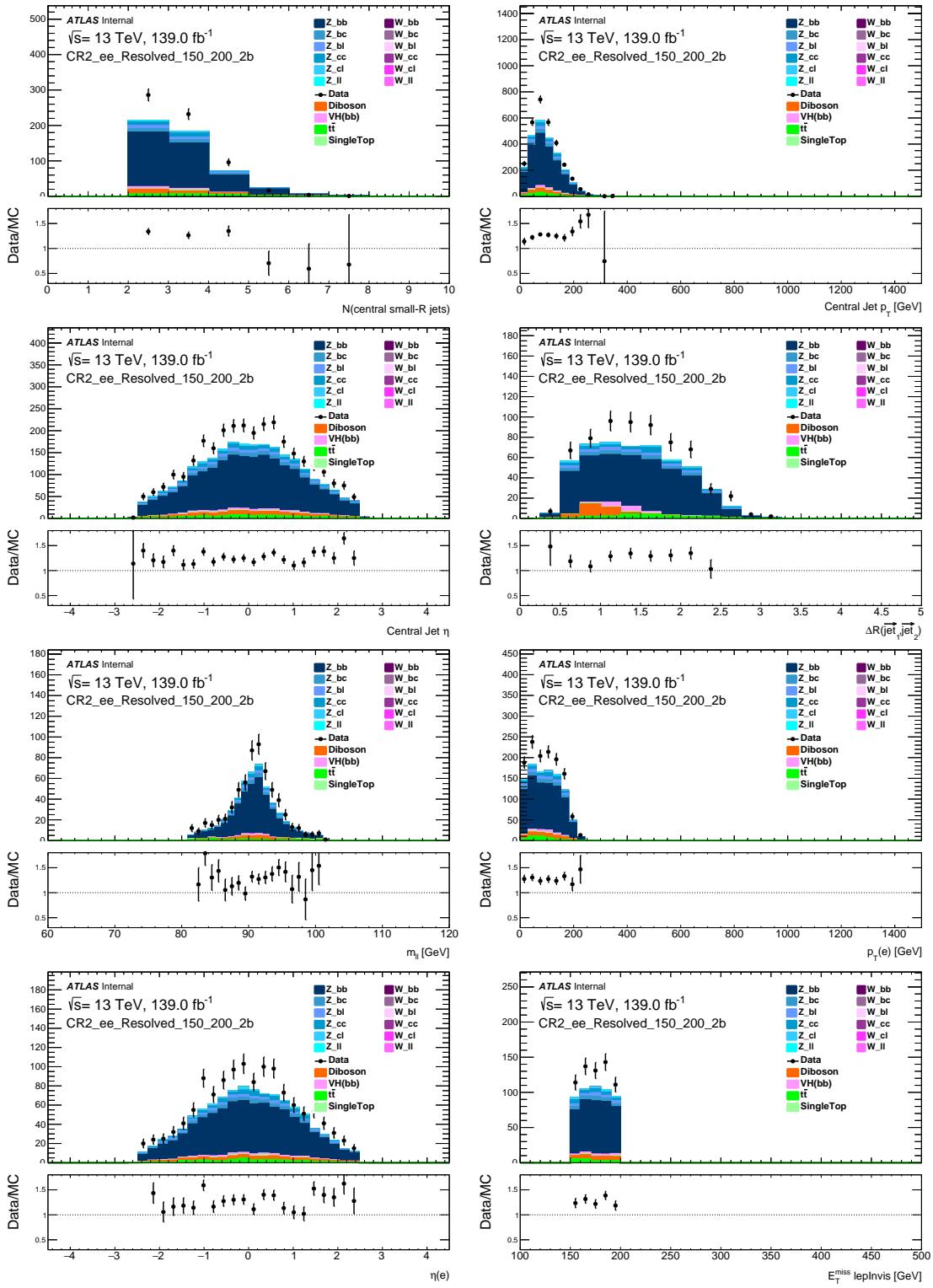


Figure 61: Kinematic distributions in 2-electron CR with 2 b-tagged jets and E_T^{miss} between 150 and 200 GeV

Not reviewed, for internal circulation only

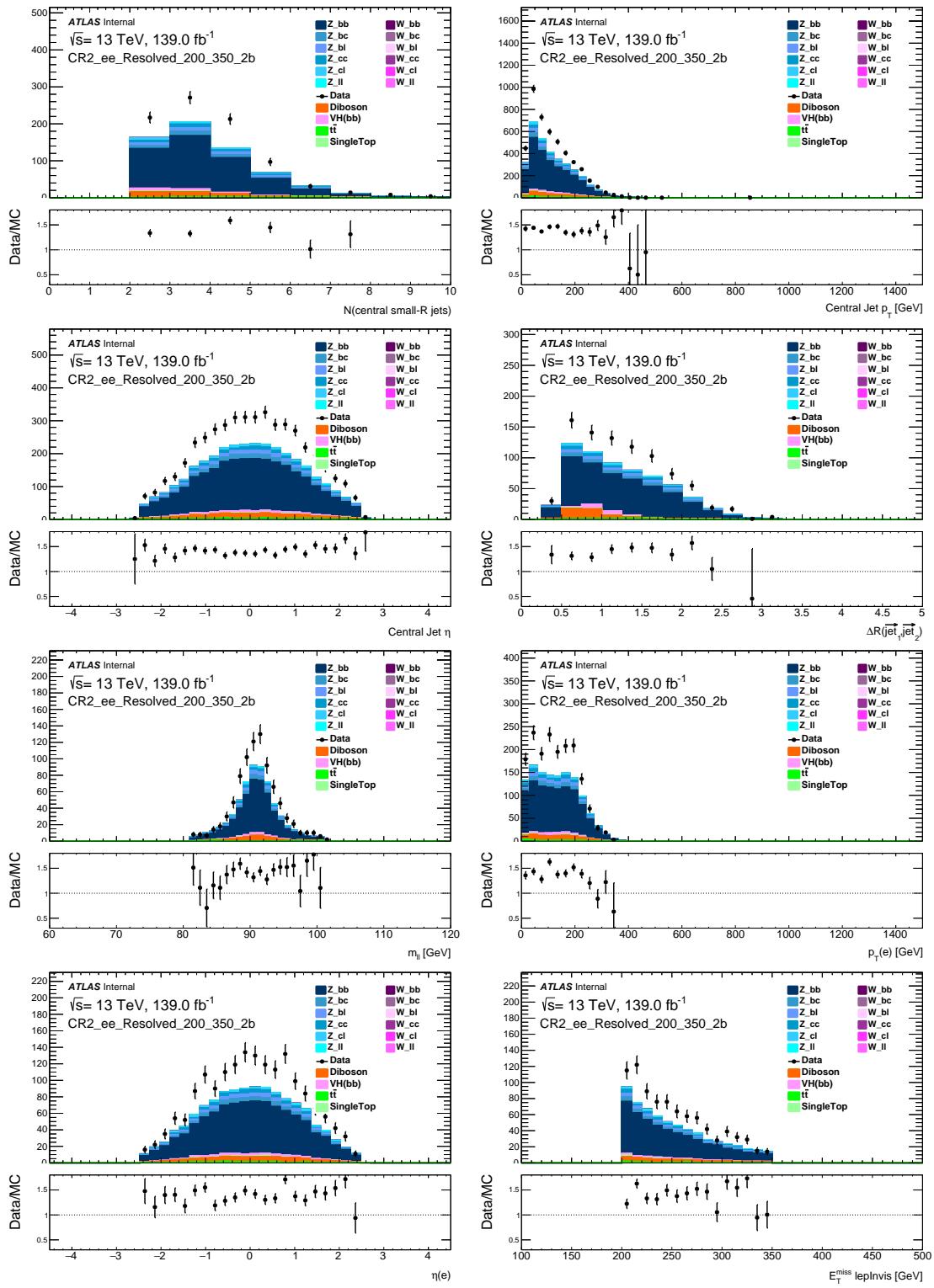


Figure 62: Kinematic distributions in 2-electron CR with 2 b-tagged jets and E_T^{miss} between 200 and 350 GeV.

Not reviewed, for internal circulation only

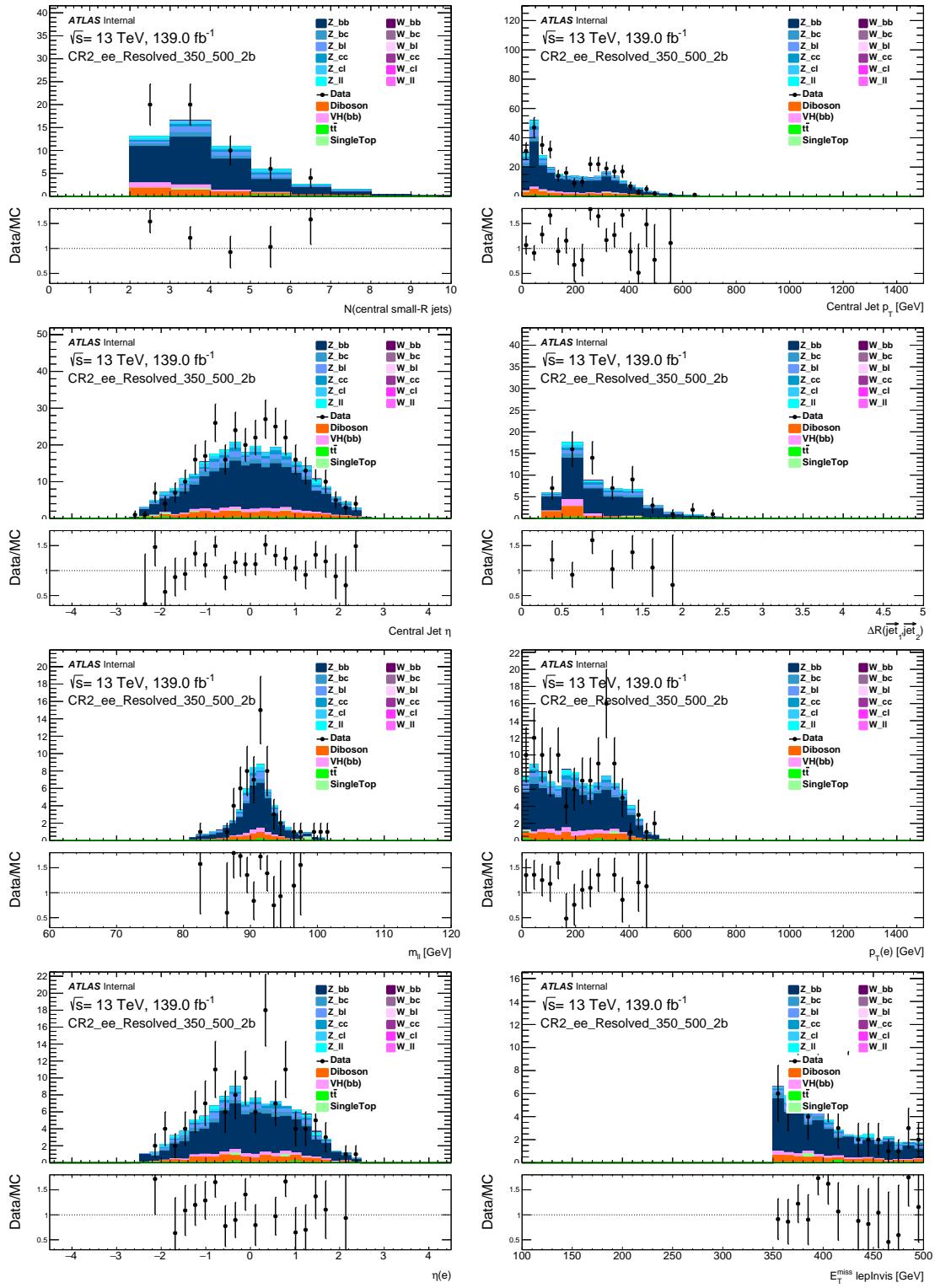


Figure 63: Kinematic distributions in 2-electron CR with 2 b-tagged jets and E_T^{miss} between 350 and 500 GeV.

Not reviewed, for internal circulation only

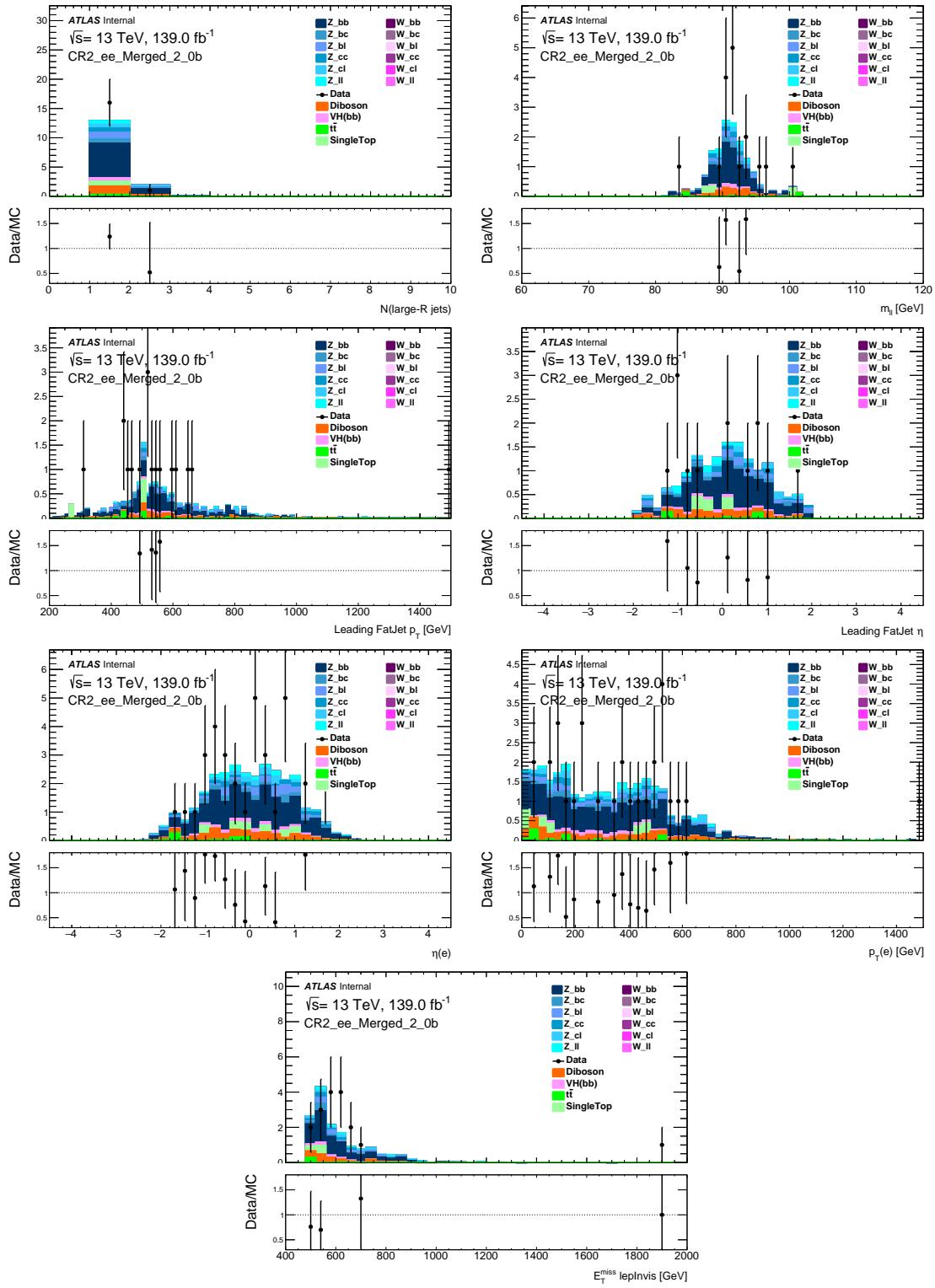


Figure 64: Kinematic distributions in 2-electron CR with a 2 b-tagged fatjet and E_T^{miss} above 500 GeV.

Not reviewed, for internal circulation only

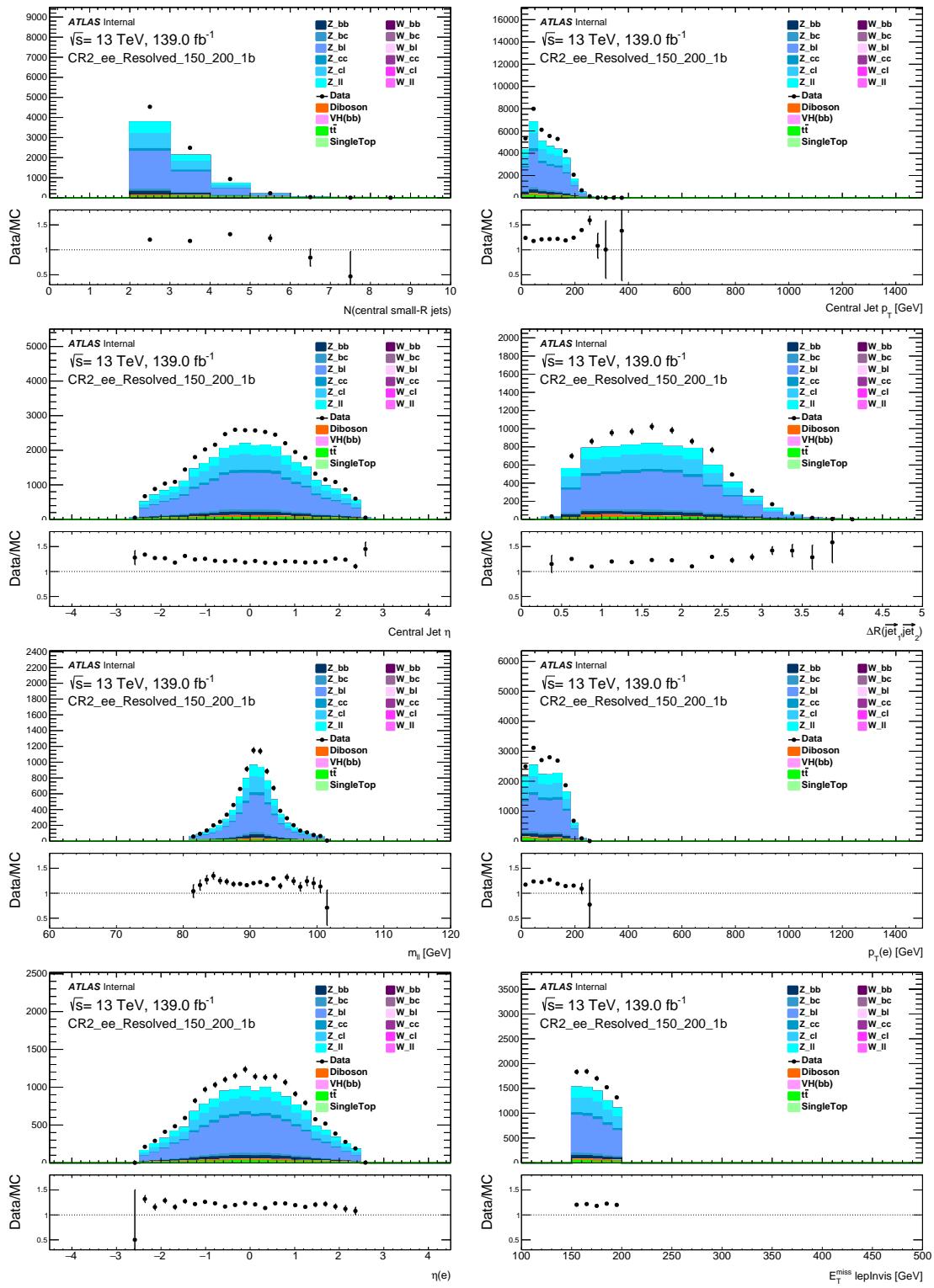


Figure 65: Kinematic distributions in 2-electron CR with 1 b-tagged jet and E_T^{miss} between 150 and 200 GeV

Not reviewed, for internal circulation only

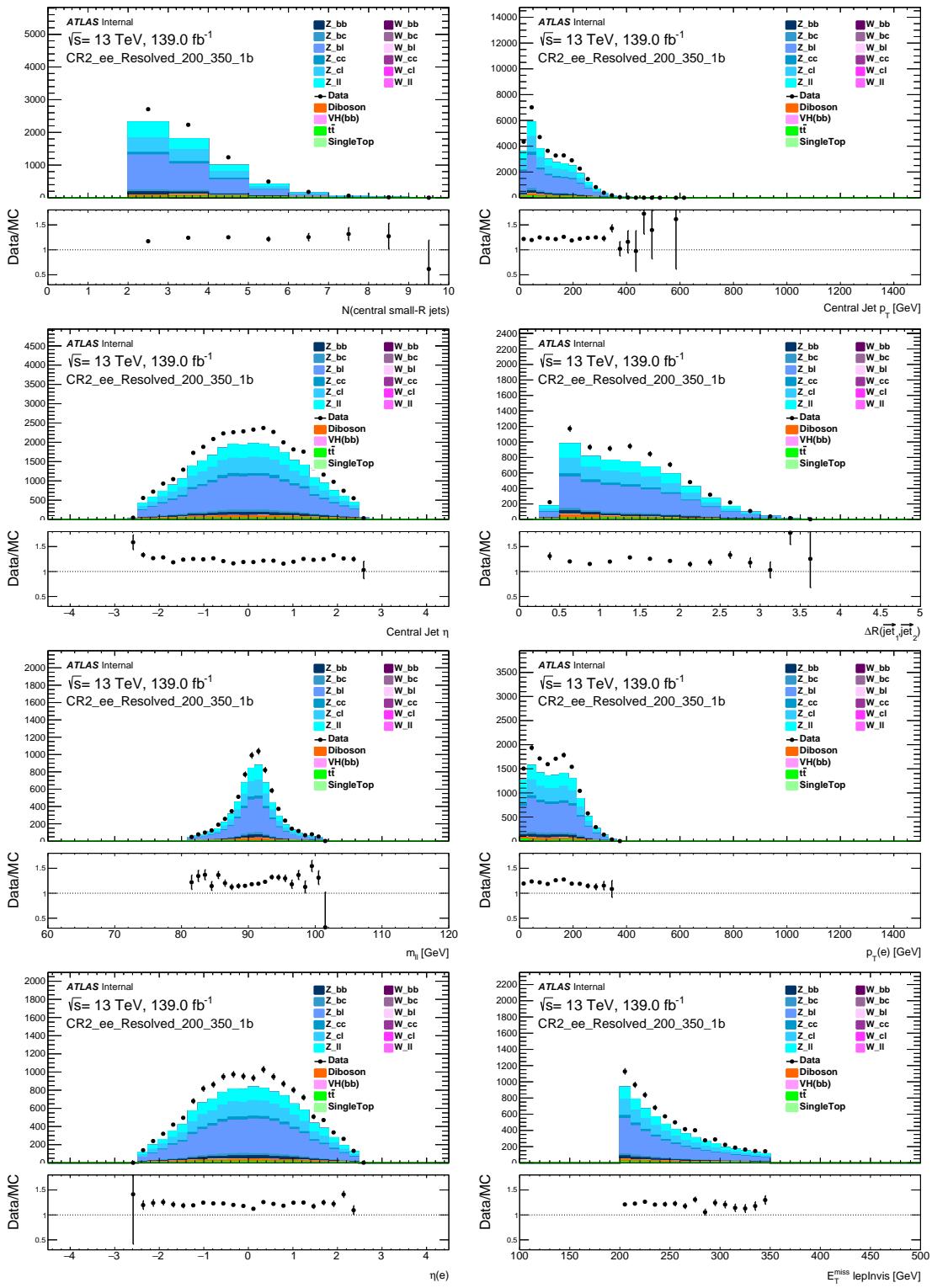


Figure 66: Kinematic distributions in 2-electron CR with 1 b-tagged jet and E_T^{miss} between 200 and 350 GeV

Not reviewed, for internal circulation only

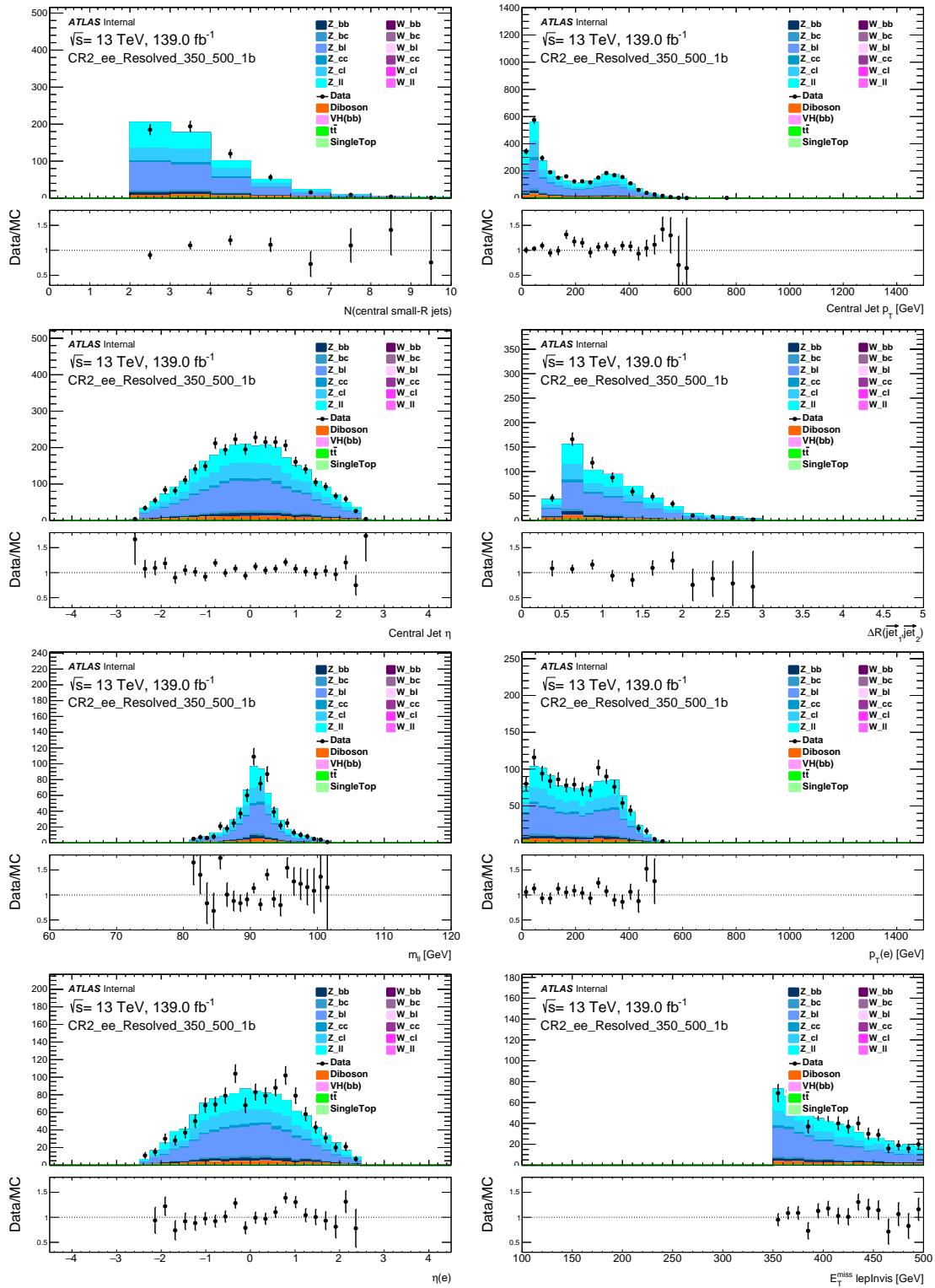


Figure 67: Kinematic distributions in 2-electron CR with 1 b-tagged jet and E_T^{miss} between 350 and 500 GeV

Not reviewed, for internal circulation only

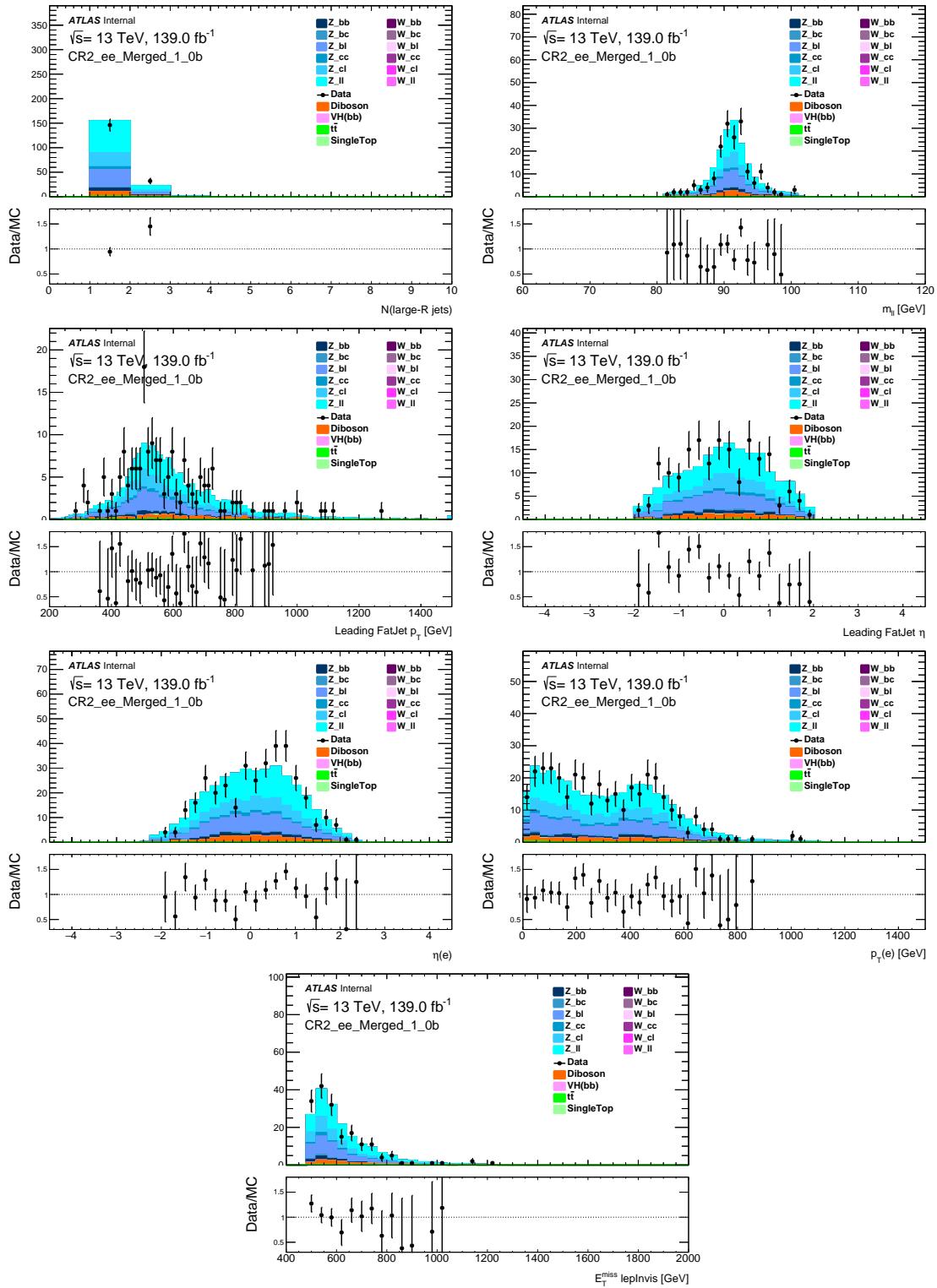


Figure 68: Kinematic distributions in 2-electron CR with a 1 b-tagged fatjet and E_T^{miss} above 500 GeV.

1358 G Studies of V+jets flavour composition [Anindya]

1359 V+jets is the most dominant background for 0 lepton Merged 2 b-tag region and second most dominating
 1360 background for Resolved 2 b-tag region in MonoH(bb) search and plays a crucial role in the fit. This
 1361 section describes the shape studies for different V+jets flavours and how to merge them in an optimum
 1362 way in the fit.

1363 **G.1 V+jets flavour labelling**

1364 A jet is labelled as a b, c or light-jet if it contains respectively a B-hadron, C-hadron or no B or C-hadrons
 1365 within a distance of $\Delta R = 0.3$ from the jet axis. The flavour of the two $R=0.4$ calorimeter jets that define
 1366 the Higgs-boson candidate in the resolved region or the two leading VR track-jets associated to the leading
 1367 $R=1.0$ calorimeter jet in the merged region define the event flavour. V+jet events are decomposed into the
 1368 following components: Vbb , Vbc , Vbl , Vcc , Vcl , Vll .

1369 **G.2 How to merge V+jets flavours**

1370 In previous round of MonoH(bb) analysis [101] the V+jets flavours were handled in the fit as:
 1371 1) **HF**: $Vbb + Vbc + Vbl + Vcc$ 2) Vcl 3) Vll

1372 In the fit *HF* normalizations are floating and Vcl and Vll have a prior of 30% and 10% respectively and
 1373 normalization of $\{bc, bl, cc\}/\{bb\}$ has a prior of 20%. Whether this V+jets flavour merging strategy is
 1374 optimal or not for MonoH(bb) analysis and if a new strategy is required are the questions addressed in
 1375 this study. With infinite MC statistics and systematics that cover mismodelling sufficiently, the V+jets
 1376 flavour merging strategy would have a negligible effect on the fit. However because of limited statistics
 1377 and systematic fluctuations the merging can affect the fit in various ways, i.e smoothing and pruning. For
 1378 example systematics with statistical fluctuations can have multiple minima, maxima which can lead to
 1379 strange pulls and cause convergence issues in the fit (i.e affecting the smoothing). Besides the V+jets
 1380 flavours which have small systematics variation from the nominal could be removed by pruning in the
 1381 fit.

1382 There are different ways to merge the flavours:

1383 1. Theoretically motivated:

- 1384 • Keep HF template separate from light (since light is well modelled).
- 1385 • Merge V+jets flavours which originate from similar processes (e.g. bb, cc from gluon splitting).

1386 2. Shape driven

- 1387 • Maximise shape differences between the merged V+jets flavours .

1388 The goal here is to

- 1389 • See if any prior knowledge can be injected on what to merge based on the V+jets flavours shape.
- 1390 • Find if few easily interpretable V+jets flavours merging strategy can be obtained.

1391 G.2.1 Merging strategy

1392 The strategy considered for merging the different V+jet flavours is :

- 1393 1. Take the ratio of flavour1/flavour2 and fit the ratio with a linear fit as a function of Higgs mass(fit
1394 variable for 0 lepton channel).

1395 2. Take the slope and error on the slope from the fit result and calculate slope/error and merge the two
1396 flavours which gives the smallest slope and slope/error, e.g for the 0 lep Resolved 2b category V_{cl}
1397 and V_{cc} has lowest slope and slope/error, hence merged together as shown in Table 17.

1398 3. Compare the new merged flavour with other flavours and depending on the slope and slope/error
1399 decide whether to merge or not.

1400 4. The iteration continues until a template is found where the shape of the flavours is very distinct.

1401 5. The main criterion to merge or not merge is the slope and slope/error. χ^2 and Kolmogorov-Smirnov
1402 tests are also used as cross-checks.

1403 The background compositions(69) in different regions give a better indication why the merging is neces-
 1404 sary.In Resolved 2b-tag category the most dominant V+jets flavour is Vbb (62%), which helps in better
 1405 estimation of Vbb systematics. However the other V+jets flavour contributions in the Resolved 2b-tag cat-
 1406 egory is very small which limits the precision of systematics estimation for those flavours, hence merging
 1407 of the V+jets flavours is necessary.

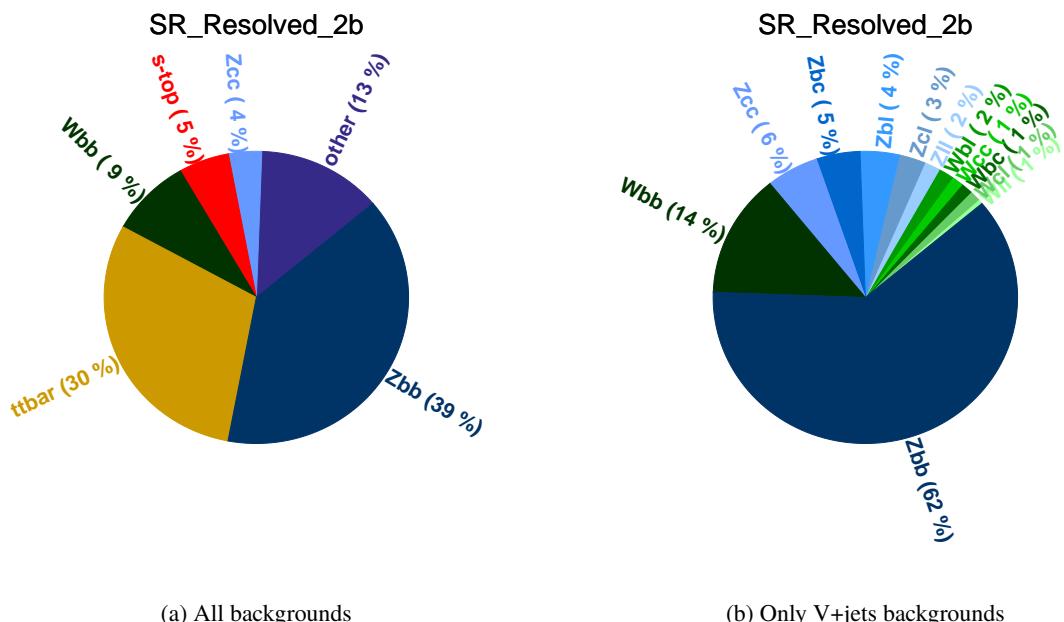


Figure 69: Background composition of (a) all backgrounds included (b) only V+jets included in SR Resolved 2b region. Backgrounds contributing <3% are combined in "**other**" category.

The shape tests are performed with an inclusive E_T^{miss} binning(150-500 GeV) with the Higgs mass distribution. An inclusive E_T^{miss} binning can be used as the flavour decomposition does not change with E_T^{miss} (fig 70). Similar characteristics are observed for the Higgs mass distribution as well.

Not reviewed, for internal circulation only

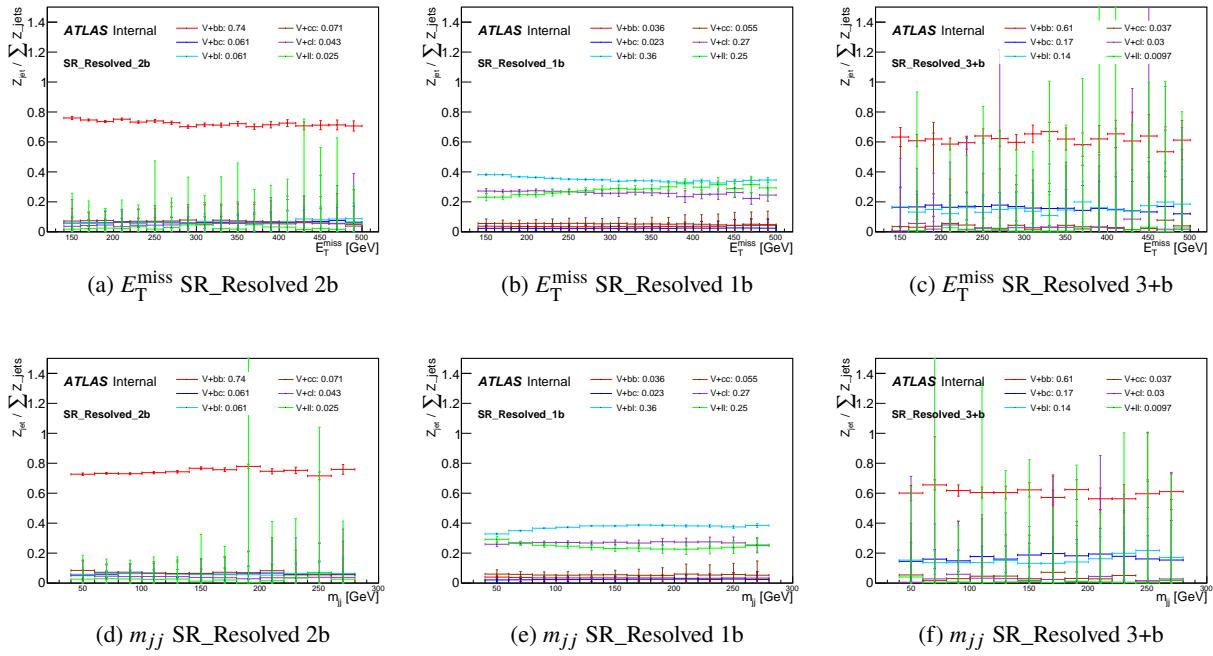


Figure 70: $Z_{\text{jet}} / \sum Z_{\text{jet}}$ plots for E_T^{miss} (top) and Higgs mass m_{jj} (bottom) distributions for different b-tag Resolved categories in 0 lep channel. The flavour decomposition doesn't change with E_T^{miss} or Higgs mass(m_{jj}).

Following the procedure described in step 1 and 2 of merging strategy the most compatible flavours found for the resolved 2 b-tag category are Z_{cl} and Z_{cc} as shown in Table 17, hence they are merged together (fig 71).

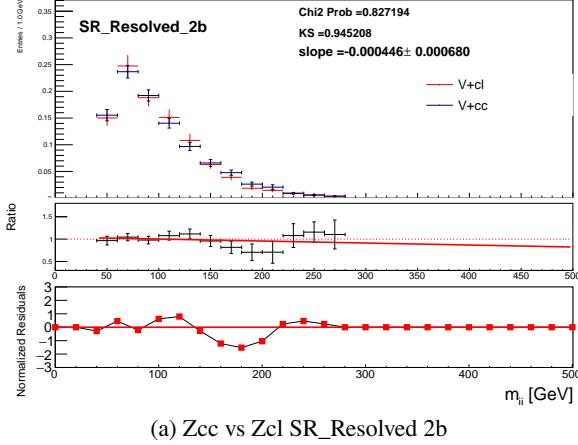


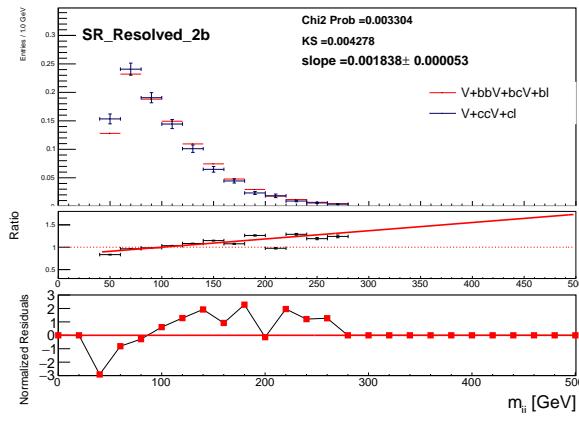
Figure 71: Compatibility of Z_{cc} and Z_{cl} shapes

Not reviewed, for internal circulation only

Flavour	Slope	Error	Slope/Error
Zbb_Zbc	0.000298	0.000057	5.23
Zbb_Zbl	0.000702	0.000057	12.32
Zbb_Zcc	0.001602	0.000057	28.11
Zbb_Zcl	0.002581	0.000057	45.28
Zbb_Zll	0.003228	0.000057	55.26
Zbl_Zbc	0.000858	0.000212	4.05
Zbl_Zcl	0.000858	0.000212	8.20
Zbl_Zcc	0.000858	0.000212	4.34
Zbl_Zll	0.000858	0.000212	9.60
Zcl_Zbc	0.002309	0.00068	3.40
Zcl_Zcc	0.000446	0.00068	0.66
Zcl_Zll	0.00185	0.000669	2.77
Zbc_Zcc	0.00118	0.000214	5.22
Zcc_Zll	0.003846	0.000506	7.60

Table 17: Table for shape comparison results in step 1 of V+jets shape comparison

1414 The combined $Zcl+Zcc$ is then compared with other flavours and combinations and the best scheme where
 1415 the final template has maximum shape difference is $Zbb+Zbc+Zbl$, $Zcl+Zcc$, Zll as shown in figure 72.



(a) Merging scheme for SR_Resolved 2b

Figure 72: Most distinct merging template : Zbb+bc+bl , Zcc+c1 , Zll

1416 No significant differences found between V+jets flavour merging templates for 1) Previous round of
 1417 MonoH(bb)[101] 2) Shape driven scheme and 3) VHbb Resonance scheme[102] as shown in figure 73
 1418 and there is no strong preference for any of the scheme.

Not reviewed, for internal circulation only

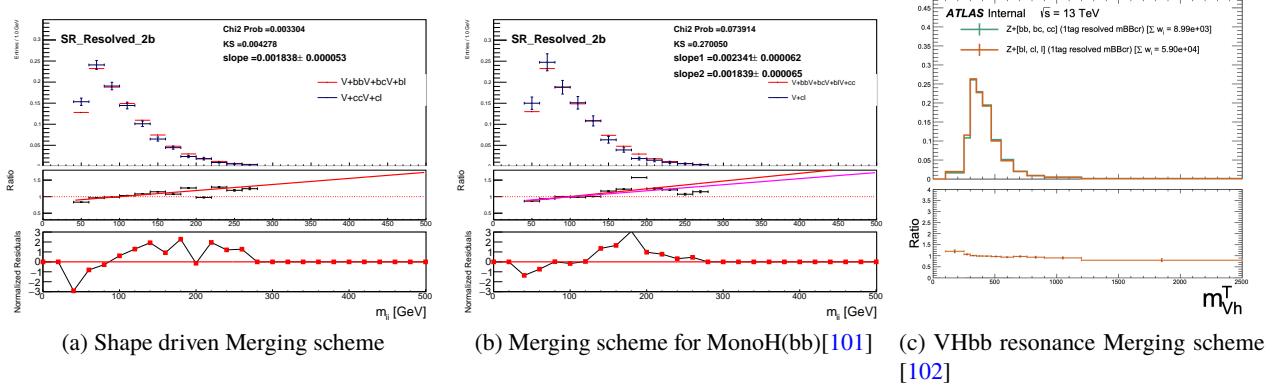


Figure 73: Comparison between Shape driven, previous MonoH(bb) analysis[101] and VHbb Resonance merging scheme. No scheme has a clear preferene.

1419 However the V+jets merging scheme for resolved 2 b-tag is not the optimal for 1b or at least 3 b-tag regions.
 1420 For 1 b-tag region the best scheme found following the same procedure is $Zbb+Zbc+Zbl+Zcc+Zcl$, Zll
 1421 (fig 74).

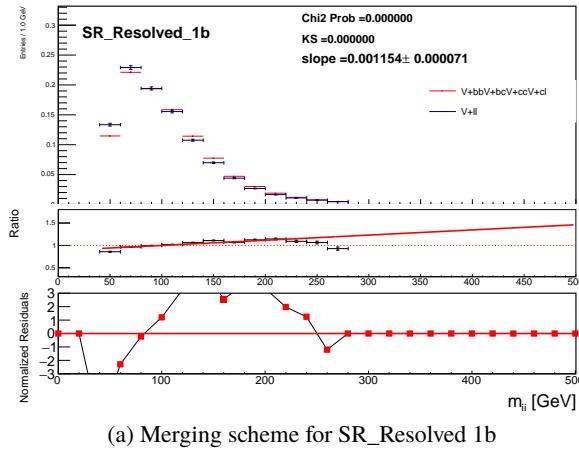


Figure 74: Most distinct merging template : $Zbb+bc+bl+Zcc+ccl$, Zll

1422 From these studies no clear preferred scheme can be identified, a decision can only be made with the fit
 1423 studies including full modelling systematics on flavour composition. However, these studies will help to
 1424 interpret the fit results.

1425 H Implemntation of E_T^{miss} filtered $t\bar{t}$, B-filtered Diboson VZ($Z \rightarrow b\bar{b}$) and 1426 Zvv+jets PTV sliced samples [Anindya]

1427 This section describes the merging of inclusive $t\bar{t} + E_T^{miss}$ samples with E_T^{miss} filtered $t\bar{t}$ samples, inclusive
1428 diboson VZ($Z \rightarrow q\bar{q}$) with B-filtered diboson VZ($Z \rightarrow b\bar{b}$) samples and Zvv+jets MAX(HT,PTV) samples
1429 with Zvv+jets PTV sliced samples to increase the MC statistics of these processes.

1430 H.1 Merging inclusive with E_T^{miss} filtered $t\bar{t}$ samples

1431 In previous iteration of MonoH(bb) analysis [101] a E_T^{miss} inclusive non-all hadronic $t\bar{t}$ sample produced
1432 with Powheg+Pythia8+EvtGen was used for modelling the $t\bar{t}$ background. In 0 lepton channel a $E_T^{miss} >$
1433 150 cut removes significant statistics from the inclusive non-all hadronic $t\bar{t}$ sample. In order to improve
1434 the MC statistical uncertainties for the analysis increase in MC statistics is necessary. This can be achieved
1435 by either :

- 1436 1. Generating more MC for the E_T^{miss} inclusive non-all hadronic $t\bar{t}$ sample which requires significant
1437 CPU time.
- 1438 2. Producing E_T^{miss} filtered $t\bar{t}$ samples targetting different phase spaces of the analysis, which is a much
1439 less CPU consuming approach.

1440 There are different E_T^{miss} filtered $t\bar{t}$ samples available as described in table 18 which are used in the analysis
1441 to gain statistics.

DSID	Generator	Process	Xsec $\sigma [pb]$	K-factor	ϵ_{filter}
410470	Powheg+Pythia8+EvtGen	non-allhad $t\bar{t}$, E_T^{miss} inclusive	729.77	1.14	0.54385
345935	Powheg+Pythia8+EvtGen	non-allhad $t\bar{t}$, $100 < E_T^{miss} < 200$ GeV	729.77	1.0	0.095208
407345	Powheg+Pythia8+EvtGen	non-allhad $t\bar{t}$, $200 < E_T^{miss} < 300$ GeV	729.77	1.14	0.00806
407346	Powheg+Pythia8+EvtGen	non-allhad $t\bar{t}$, $300 < E_T^{miss} < 400$ GeV	729.77	1.14	0.00119
407347	Powheg+Pythia8+EvtGen	non-allhad $t\bar{t}$, $E_T^{miss} > 400$ GeV	729.77	1.14	0.000314

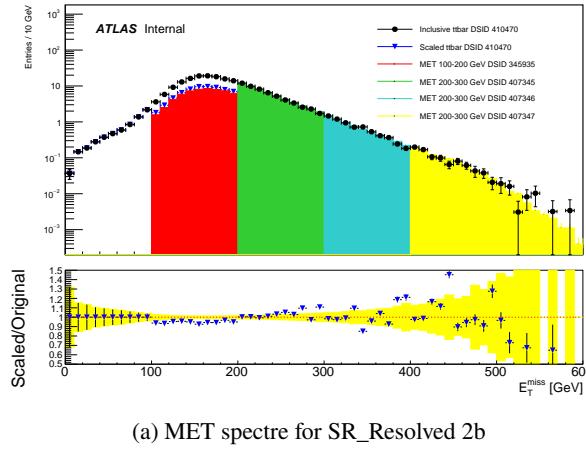
Table 18: List of $t\bar{t} + E_T^{miss}$ samples used

1442 The $t\bar{t}$ with $100 < E_T^{miss} < 200$ GeV filter sample is generated in such a way that there is no overlap with
1443 the inclusive non-all hadronic $t\bar{t}$ sample while the $E_T^{miss} > 200$ GeV filtered $t\bar{t}$ samples can have overlaps.
1444 The following strategy is chosen to merge these samples :

- 1445 1. For events with $E_T^{miss} < 100$ GeV only the E_T^{miss} inclusive non-all hadronic $t\bar{t}$ events are used with
1446 original generator weight.
- 1447 2. For events with $100 < E_T^{miss} < 200$ GeV both the E_T^{miss} inclusive non-all hadronic and E_T^{miss} (345935)
1448 filtered $t\bar{t}$ events are used with half of the original generator weight.
- 1449 3. For events with $E_T^{miss} > 200$ GeV only the E_T^{miss} filtered $t\bar{t}$ events are used with original generator
1450 weight.

1451 The merging strategy is validated using MC16d samples in the 0 lepton Merged and Resolved regions for
 1452 2 b-tag categories.

1453 Figure 75 shows the E_T^{miss} spectrum for the E_T^{miss} inclusive $t\bar{t}$ sample and different E_T^{miss} filtered $t\bar{t}$ slices
 1454 before and after the merging. The ratio plot here corresponds to the ratio of E_T^{miss} inclusive $t\bar{t}$ sample and
 1455 the final $t\bar{t} + E_T^{miss}$ sample used in the analysis. The yellow band corresponds to the statistical uncertainties
 1456 for the E_T^{miss} inclusive $t\bar{t}$ sample.

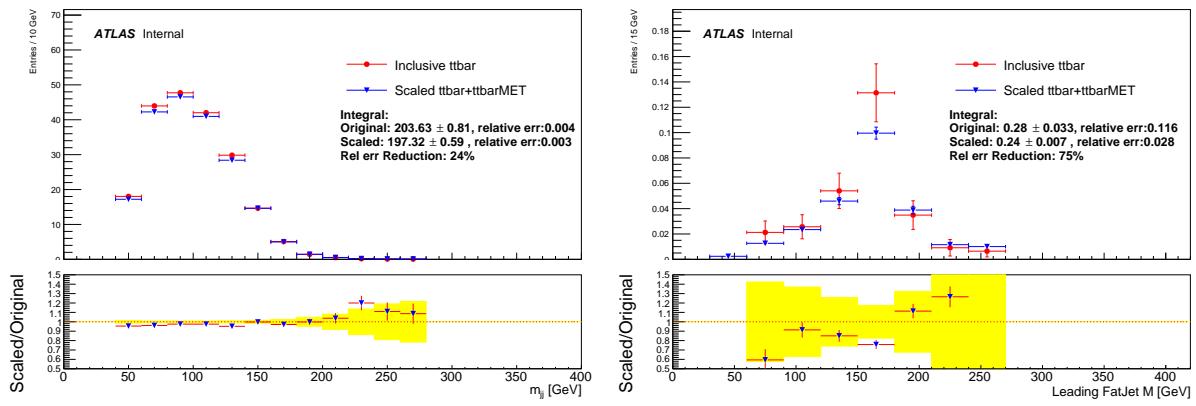


(a) MET spectre for SR_Resolved 2b

Figure 75: MET spectrum for the nominal $t\bar{t}$ and different E_T^{miss} filtered slices after reweighting

1457 The plot 75 shows that the ratio of the E_T^{miss} inclusive $t\bar{t}$ sample and the reweighted $t\bar{t} + E_T^{miss}$ sample used
 1458 in the analysis are within the statistical uncertainties of the E_T^{miss} inclusive $t\bar{t}$ sample.

1459 The increase in MC statistics decreases the MC statistical uncertainties for both the Merged and Resolved
 1460 2 b-tag category in 0 lepton channel as shown with the m_{bb} distribution in figure 76. The ratio plot
 1461 corresponds to the E_T^{miss} inclusive $t\bar{t}$ sample and the final $t\bar{t} + E_T^{miss}$ sample, while the yellow band is the
 1462 statistical uncertainty band for the E_T^{miss} inclusive $t\bar{t}$ sample.



(a) m_{bb} for SR_Resolved 2b

(b) m_{bb} for SR_Merged 2b

Figure 76: m_{bb} for the nominal $t\bar{t}$ and reweighted $m_{bb} + E_T^{miss}$ sample

1463 For the Resolved and Merged 2 b-tag categories the relative error is reduced by 24% and 75% respect-
 1464 ively.

1465 H.2 Merging $Z\nu\nu$ +jets MAX(HT,PTV) with $Z\nu\nu$ +jets $p_T(V)$ sliced samples

1466 $Z\nu\nu$ +jets contributes 33% and 60% of the total backgrounds in the 0 lepton Resolved and Merged 2 b-tag
 1467 region respectively. In the previous round of MonoH(bb) [101] only $Z\nu\nu$ +jets MAX(HT, $p_T(V)$) samples
 1468 were used, while in this iteration $p_T(V)$ sliced $Z\nu\nu$ +jets B-filtered samples are also used to gain MC
 1469 statistics. This section describes how the $p_T(V)$ sliced $Z\nu\nu$ +jets B-filtered samples are merged with the
 1470 $Z\nu\nu$ +jets MAX(HT, $p_T(V)$) samples.

1471 The $Z\nu\nu$ +jets MAX(HT, $p_T(V)$) and $Z\nu\nu$ +jets B-filtered $p_T(V)$ sliced samples used are described in table
 1472 [19](#) and [20](#).

Z $\nu\nu$ +jets MAX(HT, $p_T(V)$) samples					
DSID	Generator	Filter	Xsec $\sigma[pb]$	K-factor	ϵ_{filter}
364144	Sherpa	MAX(HT, $p_T(V)$)0_70	10705.0	0.9728	0.067425
364147	Sherpa	MAX(HT, $p_T(V)$)70_140	607.97	0.9728	0.11955
364150	Sherpa	MAX(HT, $p_T(V)$)140_280	222.38	0.9728	0.15813
364153	Sherpa	MAX(HT, $p_T(V)$)280_500	47.421	0.9728	0.18047
364154	Sherpa	MAX(HT, $p_T(V)$)500_1000	9.9101	0.9728	1.0
364155	Sherpa	MAX(HT, $p_T(V)$)1000_E_CMS	0.818	0.9728	1.0

Table 19: List of $Z\nu\nu$ +jets MAX(HT, $p_T(V)$) samples

Z $\nu\nu$ +jets $p_T(V)$ sliced B-filtered samples					
DSID	Generator	Filter	Xsec $\sigma[pb]$	K-factor	ϵ_{filter}
366010	Sherpa	$p_T(V)70_100$	275.21	1.0	0.076979
366011	Sherpa	$p_T(V)100_140_MJ0_500$	109.1	1.0	0.009147
366012	Sherpa	$p_T(V)100_140_MJ500_1000$	4.5514	1.0	0.12478
366013	Sherpa	$p_T(V)100_140_MJ1000_E_CMS$	1.2029	1.0	0.11405
366014	Sherpa	$p_T(V)140_280_MJ0_500$	51.779	1.0	0.10208
366015	Sherpa	$p_T(V)140_280_MJ500_1000$	4.4678	1.0	0.12876
366016	Sherpa	$p_T(V)140_280_MJ1000_E_CMS$	1.376	1.0	0.11958
366017	Sherpa	$p_T(V)280_500$	4.2467	1.0	0.11401
364222	Sherpa	$p_T(V)500_1000$	0.3044	1.0	1.0
364223	Sherpa	$p_T(V)1000_E_CMS$	0.0058558	1.0	1.0

Table 20: List of $p_T(V)$ sliced $Z\nu\nu$ +jets B-filtered samples used in the analysis

1473 Only B-filtered $Z\nu\nu$ +jets $p_T(V)$ sliced samples are merged to the $Z\nu\nu$ +jets MAX(HT, $p_T(V)$) samples
 1474 because the contributions from C-filtered and light filtered $Z\nu\nu$ +jets $p_T(V)$ sliced samples in 0 lepton
 1475 Resolved and Merged 2 b-tag regions are negligible compared to the B-filtered $Z\nu\nu$ +jets $p_T(V)$ sliced
 1476 samples as shown in table [21](#).

Topology	N_b-tag	Total $Z\nu\nu$ +jets $p_T(V)$ yield	BFilterCveto	CFilterBVeto	BvetoCveto
Merged	2	0.758	97.6%	1.4%	1.0%
Resolved	2	31.67	89.6%	8.2%	2.2%

Table 21: Yields comparison between different filters of $p_T(V)$ sliced $Z\nu\nu$ +jets samples

1477 The two sets of samples are merged with generator weight halved for each event. The increase of MC
 1478 statistics decreases the MC statistical uncertainties in both 0 lepton Merged and Resolved 2 b-tag region
 1479 as shown in 77. The ratio plot corresponds to the original $Z\nu\nu$ +jets MAX(HT, $p_T(V)$) and combined
 1480 $Z\nu\nu$ +jets MAX(HT, $p_T(V)$), $Z\nu\nu$ +jets B-filtered $p_T(V)$ sliced samples. The yellow band is the statistical
 1481 uncertainty band for $Z\nu\nu$ +jets MAX(HT, $p_T(V)$) samples.

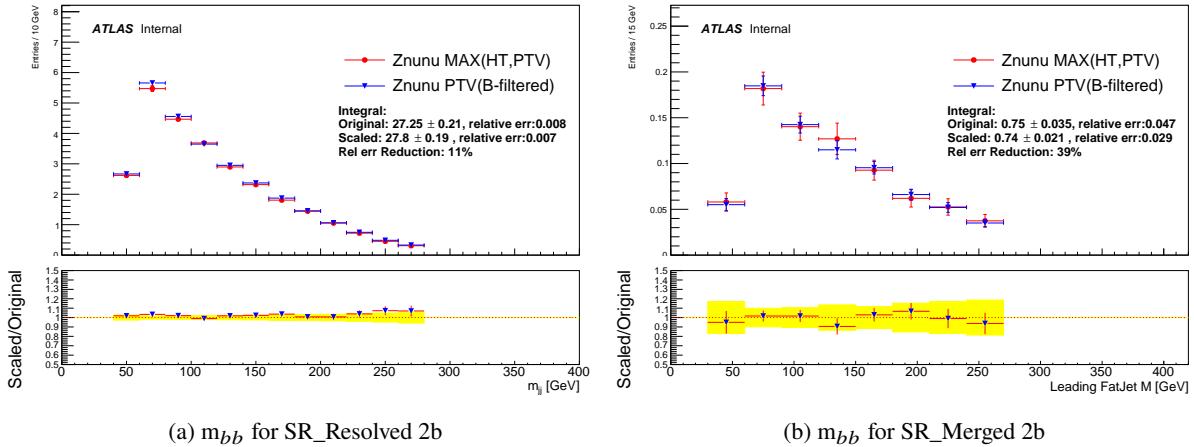


Figure 77: m_{bb} for the $Z\nu\nu$ +jets MAX(HT, $p_T(V)$) and combined $Z\nu\nu$ +jets MAX(HT, $p_T(V)$), $Z\nu\nu$ +jets B-filtered $p_T(V)$ sliced samples

1482 For the 0 lepton Resolved and Merged 2 b-tag categories the relative error is reduced by 11% and 39%
 1483 respectively.

1484 H.3 Merging inclusive diboson VZ($Z \rightarrow q\bar{q}$) with B-filtered diboson VZ($Z \rightarrow b\bar{b}$) samples

1485 This section describes the merging of inclusive diboson samples VZ($Z \rightarrow q\bar{q}$) with B-filtered diboson
 1486 samples VZ($Z \rightarrow b\bar{b}$). The addition of B-filtered diboson samples VZ($Z \rightarrow b\bar{b}$) increases the MC statistics,
 1487 hence the MC statistical uncertainty is decreased. In previous MonoH(bb)[101] analysis only the inclusive
 1488 diboson samples VZ($Z \rightarrow q\bar{q}$) were used, while in this iteration of MonoH(bb) analysis the diboson samples
 1489 used are described in table22.

Inclusive diboson VZ($Z \rightarrow q\bar{q}$) samples					
DSID	Generator	Channel	Xsec $\sigma [pb]$	K-factor	ϵ_{filter}
363355	Sherpa	ZqqZvv	15.56	1.0	0.27686
363356	Sherpa	ZqqZll	15.563	1.0	0.14135
363489	Sherpa	ZqqWlv	11.419	1.0	1.0

B-filtered diboson VZ($Z \rightarrow b\bar{b}$) samples					
DSID	Generator	Channel	Xsec $\sigma [pb]$	K-factor	ϵ_{filter}
345043	Sherpa	ZbbZvv	15.565	1.0	0.06156
345044	Sherpa	ZbbZll	15.569	1.0	0.03127
345045	Sherpa	ZbbWlv	2.5045	1.0	0.12478

Table 22: List of inclusive and B-filtered diboson samples used

1490 The two sets of samples are merged with the following strategy:

- 1491 1. All the events from the B-filtered diboson VZ($Z \rightarrow b\bar{b}$) samples are reweighted.
- 1492 2. All the $Z \rightarrow b\bar{b}$ events from the inclusive diboson VZ($Z \rightarrow q\bar{q}$) sample are reweighted.
- 1493 3. All the non $Z \rightarrow b\bar{b}$ events from the inclusive diboson VZ($Z \rightarrow q\bar{q}$) sample are used as it is .

1494 The reweighting factors for both the samples are calculated such that: 1) MC statistical uncertainty is
 1495 minimized and 2) sum of generator event weights for the pair of events which are reweighted equals to 1.
 1496 The following set of equations satisfies the above-mentioned criterions

$$\frac{W_{bf}^{bb}}{W_{inc}^{bb}} = \frac{N_{bf}^{bb}}{N_{inc}^{bb}} \quad \text{and} \quad W_{bf}^{bb} + W_{inc}^{bb} = 1 \quad (10)$$

1497 where

1498 W_{bf}^{bb} is the generator event weight of B-filter sample, W_{inc}^{bb} is the generator event weight of a Zbb event in
 1499 the inclusive sample

1500 N_{bf}^{bb} is the number of events in the B-filter sample, N_{inc}^{bb} is the number of Zbb events in inclusive sample

1501

1502 Solving equation 10 the weights can be written as

$$W_{bf}^{bb} = \frac{N_{bf}^{bb}}{N_{bf}^{bb} + N_{inc}^{bb}} \quad \text{and} \quad W_{inc}^{bb} = \frac{N_{inc}^{bb}}{N_{bf}^{bb} + N_{inc}^{bb}} \quad (11)$$

1503 while

$$N_{bf}^{bb} = N_{bb}^{yield} \quad \text{and} \quad N_{inc}^{bb} = N_{inc}^{yield} \times \frac{Br(Z \rightarrow bb)}{Br(Z \rightarrow qq)} \quad (12)$$

1504 The branching ratio of $\frac{Br(Z \rightarrow bb)}{Br(Z \rightarrow qq)}$ is 0.2161 [103].

1505 From equation 11 and 12 the new event weights for the Zbb events in the inclusive sample(13a) and all
1506 the events in B-filtered sample(13b) can be written as

$$W_{bf}^{bb_new} = W_{bf}^{bb} \times \frac{W_{bf}^{bb}}{W_{bf}^{bb} + W_{inc} \times \frac{Br(Z \rightarrow bb)}{Br(Z \rightarrow qq)}} \quad (13a)$$

$$W_{inc}^{bb_new} = W_{inc}^{bb} \times \frac{W_{inc} \times \frac{Br(Z \rightarrow bb)}{Br(Z \rightarrow qq)}}{W_{bf}^{bb} + W_{inc} \times \frac{Br(Z \rightarrow bb)}{Br(Z \rightarrow qq)}} \quad (13b)$$

1507 Following the above-mentioned strategy, the samples are merged and are validated with the MC16d
1508 samples in the 0 lepton channel for both Merged and Resolved 2 b-tag category. The results are shown
1509 in figure 78 where the ratio plot corresponds to the original inclusive diboson VZ($Z \rightarrow q\bar{q}$) sample and
1510 merged inclusive VZ($Z \rightarrow q\bar{q}$) and B-filtered VZ($Z \rightarrow b\bar{b}$) diboson samples. The yellow band is the
1511 statistical uncertainty band for inclusive diboson VZ($Z \rightarrow q\bar{q}$) samples.

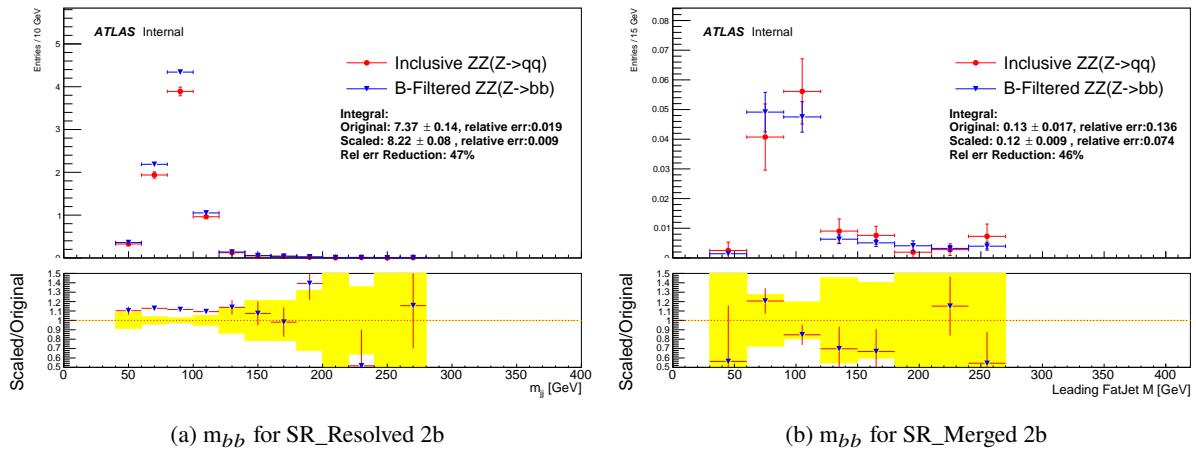


Figure 78: m_{bb} for the inclusive diboson and B-filtered diboson samples

1512 For the Resolved and Merged 2 b-tag categories the relative error is reduced by 47% and 46% respect-
1513 ively.

1514 I Resolved/Merged combination strategy [Jon, James]**1515 J Photon CR [Eleni]**

1516 To be filled properly

1517 The latest update on the subject can be found at the following [link](#)

1518 K Truth-level studies of signals [Anindya, Andrea, Spyros]

1519 To be filled properly

1520 A summary can be found in the following presentation: [link](#)

1521 K.1 Comparison of ggF and b-associated production**1522 K.2 Cross-sections, widths, BR etc as a function of $m_A, m_a, \tan \beta, \sin \theta$**

1523 [Anindya]

1524 L Merging of $V+jets, t\bar{t}$ and diboson samples [Anindya]

1525 This procedure describes the combination of the filtered and nominal $t\bar{t}$, $V+jets$ and diboson samples. It
1526 follows the procedure developed in [\[104\]](#).