

Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

A DISSERTATION PRESENTED
BY
BAOJIA TONG
TO
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
PHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
MAY 2017

©2014 – BAOJIA TONG
ALL RIGHTS RESERVED.

Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

ABSTRACT

We present a search for Higgs boson pair production, with two $b\bar{b}$ pairs in the final state. This analysis uses the full 2015 and 2016 data collected by the ATLAS Collaboration at $\sqrt{s} = 13$ TeV, corresponding to $X \pm y \text{ fb}^{-1}$ of 2015 and $A \pm B \text{ fb}^{-1}$ of 2016 pp collision data. The data are interpreted in the context of the bulk Randall-Sundrum warped extra dimension model with a Kaluza-Klein graviton decaying to bb . Relative to the 2015 analysis, this analysis focuses on improvements in the boosted analysis in the highest resonance mass range (between 2000 GeV and 3000 GeV). The data is found to be compatible with the Standard model, and no signs of new physics have been observed.

Contents

o INTRODUCTION	I
1 MOTIVATION FOR SEARCHES BEYOND THE STANDARD MODEL	2
1.1 the Standard Model	3
1.2 Problems with the Standard Model	3
1.3 Paths Beyond the Standard Model	3
1.4 Higgs Self Coupling	3
2 MACHINE OF DISCOVERY—THE LARGE HADRON COLLIDER	4
2.1 Design	4
2.2 Performance	5
3 EYES OF GIANT—THE ATLAS DETECTOR	6
3.1 Inner Detector	6
3.2 Trigger and Data Aquasition	6
4 RECONSTRUCTION OF OBJECTS	8
4.1 Leptons	8
4.2 Jets	9
4.3 Flavor Tagging	10
4.4 Boosted Object Tagging	12
5 CENSOR OF WORK—DATA QUALITY	13
5.1 Data Flow in ATLAS	13
5.2 Online Monitoring	13
5.3 Offline Monitoring	13
6 EVENT SELECTION	14
6.1 Trigger Selection	24
6.2 Optimization Strategy	28

6.3	<i>b</i> tagging	28
6.4	Signal Efficiency	28
7	BACKGROUND ESTIMATION	30
7.1	Background Estimation	30
8	SYSTEMATICS	II2
9	RESULT	I42
10	INTEPRETATION	143
II	CONCLUSION	145
	APPENDIX A SOME EXTRA STUFF	146
A.1	Signal Sample Containers	146
A.2	Choice of SB/CR defintions	149
A.3	μ_{multijet} Method Validation	153
A.4	Description of selection optimization method	159
A.5	Optimization Studies	160
A.6	Correlated Error Propagation	167
A.7	Basic selection validation with signal truth MC	169
A.8	Reweighting Studies and Comparisons	172
A.9	Possible Future Additional Cuts	184
A.10	Signal Kinematic Distributions	188
A.11	Reweighting Factor Distributions	192
A.12	Boosted Debug Stream Check	200
	REFERENCES	203

Listing of figures

6.1	Trigger efficiencies as a function of the signal resonance mass with respect to all events with no selection (left) and with respect to events passing the two large- R jets requirement and leading/subleading jet p_T requirement (right). For 1 TeV signal, the trigger efficiency is about 95 %.	16
6.2	Trigger efficiency defined as the fraction of events fired trigger with a given leading jet p_T , measured in 2015 Data (HLT_j360_a10_lcw, left) and 2016 Data (HLT_j420_a10_lcw, right) and MC15c.	16
6.3	Values of the X_{bb} and R_{bb} variables, which are shapes in the two-dimensional plane of the large- R jet masses used to defined signal, control, and sideband regions. For both variables, a smaller value indicates the jets are closer to the Higgs mass.	21
6.4	Absolute (left) and relative (right) signal efficiency as a function of RSG c=1.0 signal resonance mass hypothesis for selection cuts. The relative efficiency is defined from the previous cut, where the order of cuts is given by the legend. PassTrig means the event passes the trigger selection; PassDijetPt means the event passes the leading and sub-leading jet p_T cuts; PassDijetEta means the event passes the leading and sub-leading jet η cuts; PassDeltaH means the events passes the $ \Delta\eta < 1.7$ cut; PassBJetSkim means the event contains at least two b -tagged track jets, inclusive of 2 b , 2 bs , 3 b and 4 b configurations; PassSignal means the event passes the signal region cut $X_{bb} < 1.6$	23
6.5	Signal efficiency in three search b -tag categories (left) and detailed signal efficiency in different track jet and b -tag categories (right) as a function of signal resonance mass hypothesis for selection cuts. The right plot efficiencies are relative to the total number of events in the preselection, whereas the left plot efficiencies are relative to the total number of events in the signal mass region. The green curve in the left plot also corresponds to the PassSignal curve on in Figure 6.4.	24
6.6	Detailed signal efficiency in different signal/control/sideband regions as in 2 bs (top left, 3 b (top right), 4 b (bottom left) and inclusive b-tagged regions, which include 2 b , 1 b and 0 b as well, (bottom right) as a function of signal resonance mass hypothesis for selection cuts. The efficiencies are relative to the total number of events in the preselection.	25

7.1	m_j^{lead} vs. m_j^{subl} in data in the $1b$ -tag (top) and $2b$ -tag (bottom) selection, the plots show the boundary between the Sideband (left) and Control (right) regions.	34
7.2	comparison between the $2b$, $3b$, and $4b$ shapes for the di-large- R -jet mass distributions (the final discriminant) in the SR.	38
7.3	Simultaneous fit of μ_{multijet} and $\alpha_{t\bar{t}}$ in $4b$ (top) and $3b$ (middle) and $2b$ (bottom) sideband region using leading $large - R$ calorimeter jet mass spectrum.	40
7.4	Comparison of different trackjet p_T distributions. Top row is for leading p_T Higgs candidate, and bottom row is for subleading p_T Higgs candidate. Left column is for the leading p_T trackjet of the Higgs candidate, and right column is for the subleading p_T trackjet of the Higgs candidate. Shown in the plot are just data distributions, inclusive of SB, CR, and SR regions for ob and $1b$, while for $2bs$ only the SB region is shown. $1b$ sample is further split into four subcategories, depending on which trackjet gets b tagged. OneTag lead on lead means the b tagged trackjet is the leading trackjet of the leading Higgs candidate, OneTag lead on subl means the b tagged trackjet is the subleading trackjet of the leading Higgs candidate, OneTag subl on lead means the b tagged trackjet is the leading trackjet of the subleading Higgs candidate, and OneTag subl on subl means the b tagged trackjet is the subleading trackjet of the subleading Higgs candidate. At the bottom ratio plot, all the ratio are taken with respect to the ob tagged distribution.	42
7.5	For $2bs$ background estimate: the fits to the ratio of the data in the $1b$ category, of the subleading Higgs candidate $1b$ -tagged events's subleading Higgs candidate distributions(black point), over the leading Higgs candidate $1b$ -tagged events's subleading Higgs candidate distributions(yellow). Distributions and fits to the estimated QCD background for large- R jet p_T (left), the large- R jet's leading trackjet p_T (middle), and large- R jet's subleading trackjet p_T (right) are shown. Figure are shown before reweighting (top row), after the first iteration(second row), after the fourth iteration(third row), and after the last iteration (bottom row). The green line is the spline extrapolation; and the red line is a polynomial fit.	46

7.15	Reweighted $3b$ Control region predictions comaprison. Top row is the dijet Mass, second row is leading large- R jet p_T , third row is the leading large- R jet's leading trackjet p_T and the last row subleading large- R jet's leading trackjet p_T . On the left are the distributions before reweighting, and on the right are the distributions after reweighting.	57
7.16	Reweighted $4b$ Control region predictions comaprison. Top row is the dijet Mass, second row is leading large- R jet p_T , third row is the leading large- R jet's leading trackjet p_T and the last row subleading large- R jet's leading trackjet p_T . On the left are the distributions before reweighting, and on the right are the distributions after reweighting.	58
7.17	Kinematics of the lead large- R jet in data and prediction in the sideband region after requiring $4 b$ -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.	60
7.18	Kinematics of the sub-lead large- R jet in data and prediction in the sideband region after requiring $4 b$ -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.	61
7.19	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the sideband region after requiring $4 b$ -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. The normalization agrees by construction, and the shapes are a feature of the prediction.	62
7.20	Kinematics of the large- R jet system in data and prediction in the sideband region after requiring $4 b$ -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.	63
7.21	Kinematics of the lead large- R jet in data and prediction in the sideband region after requiring $3 b$ -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.	64
7.22	Kinematics of the sub-lead large- R jet in data and prediction in the sideband region after requiring $3 b$ -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.	65

7.23	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the sideband region after requiring 3 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. The normalization agrees by construction, and the shapes are a feature of the prediction.	66
7.24	Kinematics of the large- R jet system in data and prediction in the sideband region after requiring 3 b -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.	67
7.25	Kinematics of the lead large- R jet in data and prediction in the sideband region after requiring 2 b -tags split. The normalization agrees by construction, and the shapes are a feature of the prediction.	68
7.26	Kinematics of the sub-lead large- R jet in data and prediction in the sideband region after requiring 2 b -tags split. The normalization agrees by construction, and the shapes are a feature of the prediction.	69
7.27	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the sideband region after requiring 2 b -tags split. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. The normalization agrees by construction, and the shapes are a feature of the prediction.	70
7.28	Kinematics of the large- R jet system in data and prediction in the sideband region after requiring 2 b -tags split. The normalization agrees by construction, and the shapes are a feature of the prediction.	71
7.29	Kinematics of the lead large- R jet in data and prediction in the control region after requiring 4 b -tags.	73
7.30	Kinematics of the sub-lead large- R jet in data and prediction in the control region after requiring 4 b -tags.	74
7.31	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the control region after requiring 4 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet.	75

7.32	Kinematics of the large- R jet system in data and prediction in the control region after requiring 4 b -tags.	76
7.33	Kinematics of the lead large- R jet in data and prediction in the control region after requiring 3 b -tags.	77
7.34	Kinematics of the sub-lead large- R jet in data and prediction in the control region after requiring 3 b -tags.	78
7.35	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the control region after requiring 3 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet.	79
7.36	Kinematics of the large- R jet system in data and prediction in the control region after requiring 3 b -tags.	80
7.37	Kinematics of the lead large- R jet in data and prediction in the control region after requiring 2 b -tags split.	81
7.38	Kinematics of the sub-lead large- R jet in data and prediction in the control region after requiring 2 b -tags split.	82
7.39	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the control region after requiring 2 b -tags split. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet.	83
7.40	Kinematics of the large- R jet system in data and prediction in the control region after requiring 2 b -tags split.	84
7.41	Kinematics of the lead large- R jet in data and prediction in the signal region after requiring 4 b -tags. Data is blinded, and will be added after unblinding.	86
7.42	Kinematics of the sub-lead large- R jet in data and prediction in the signal region after requiring 4 b -tags. Data is blinded, and will be added after unblinding.	87
7.43	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the signal region after requiring 4 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. Data is blinded, and will be added after unblinding.	88

7.44	Kinematics of the large- R jet system in data and prediction in the signal region after requiring 4 b -tags. Data is blinded, and will be added after unblinding.	89
7.45	Kinematics of the lead large- R jet in data and prediction in the signal region after requiring 3 b -tags. Data is blinded, and will be added after unblinding.	90
7.46	Kinematics of the sub-lead large- R jet in data and prediction in the signal region after requiring 3 b -tags. Data is blinded, and will be added after unblinding.	91
7.47	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the signal region after requiring 3 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. Data is blinded, and will be added after unblinding.	92
7.48	Kinematics of the large- R jet system in data and prediction in the signal region after requiring 3 b -tags. Data is blinded, and will be added after unblinding.	93
7.49	Kinematics of the lead large- R jet in data and prediction in the signal region after requiring 2 b -tags split. Data is blinded, and will be added after unblinding.	94
7.50	Kinematics of the sub-lead large- R jet in data and prediction in the signal region after requiring 2 b -tags split. Data is blinded, and will be added after unblinding.	95
7.51	First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the signal region after requiring 2 b -tags split. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. Data is blinded, and will be added after unblinding.	96
7.52	Kinematics of the large- R jet system in data and prediction in the signal region after requiring 2 b -tags split. Data is blinded, and will be added after unblinding.	97
7.53	Comparison of the 4 b , 3 b and 2 bs signal region $t\bar{t}$ dijet mass shape. On the left is the linear scale, and on the right is the log scale. Both distributions are normalized to 1 for comparison.	99
7.54	Fits for background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the 4 b signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.	100

7.55	Fits for background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $3b$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.	101
7.56	Fits for background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $2bs$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.	102
7.57	Smoothed background estimations the $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal regions. Only smoothing statistical uncertainties are shown here.	103
7.58	Normalized Scaled dijet mass distributions for the $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal regions. For comparison, the unscaled distributions are shown on the same plot.	105
7.59	Fits for scaled background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $4b$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.	106
7.60	Fits for scaled background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $3b$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.	107

7.61	Fits for scaled background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $2bs$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.	108
7.62	Smoothed MJJ (left) and scaled MJJ (right) background estimations the $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal regions. Smoothing statistical and systematic uncertainties from smoothing parameter variations are shown here.	109
7.63	Background prediction for $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal region using scaled di-jet mass before smoothing. The uncertainty band includes only statistical uncertainties.	110
7.64	Background prediction for $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal region using scaled di-jet mass after smoothing. The uncertainty band includes only statistical uncertainties.	III
8.1	Illustration of ZZ (left) and TT (right) signal region as shown in the orange shaded region. Control region shown in green, and Sideband region in blue. The white circle in the midde is the real Signal region, and it is blinded.	122
8.2	ZZ signal region distribution of di-jet mass (left column) and leading large-R jet mass (right column) in low mass signal, for $4b$ (top row), $3b$ (middle row) and $2b$ split (bottom row). The plots are with only statistical uncertainty.	125
8.3	TT signal region distribution of di-jet mass (left column) and leading large-R jet mass (right column) in low mass signal, for $4b$ (top row), $3b$ (middle row) and $2b$ split (bottom row). The plots are with only statistical uncertainty.	127
8.4	(left) Shape of the $t\bar{t}$ di-large- R -jet mass in the sideband region, comparing the $3b$ shape with that of the $2b$, in order to asses the systematic effect of additional b -tags changing the dijet mass distribution. The m_{JJ} distributions is shown on the left, and the ratio of $3b$ to $2b$ distributions on the right.	128
8.5	Dijet mass distribution in the CR along with the prediction (left) and the ratio of the prediction to the CR distribution (right) for the $2bs$ (top) $3b$ (middle) and $4b$ (bottom) samples. Ratios are from the smoothed distributions, the data uncertainty band contains the smoothing parameter variations, and the prediction uncertainty band also contains smoothing parameter variations.	130

8.6	Dijet mass distribution SR prediction fit with several fit ranges (left) and the ratio of nominal to fits with different fit ranges (right) for the $2b$ (top) $3b$ (middle) and $4b$ (bottom) samples.	132
8.7	Dijet mass distribution SR prediction fit with several fit functions (left) and the ratio of nominal to fits with different fit functions (right) for the $2b$ (top) $3b$ (middle) and $4b$ (bottom) samples. The additional fit functions are from Table 8.12.	134
8.8	Impact of each systematic on the signal prediction as a function of the signal mass, in the $4b$ (left) and $3b$ (middle) and $2bs$ signal regions.	137
8.9	The total background estimation in $4b$ signal region, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.	138
8.10	The total background estimation in $3b$ signal region, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.	139
8.11	The total background estimation in $2bs$ signal region, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.	139
8.12	The total background estimation in $4b$ signal region, scaled mJJ, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.	140
8.13	The total background estimation in $3b$ signal region, scaled mJJ, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.	140
8.14	The total background estimation in $2bs$ signal region, scaled mJJ, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.	141
A.1	Fit parameters for μ_{qcd} (left), $\alpha_{\bar{t}t}$ (middle), and ratio of fit uncertainty and stat uncertainty as a functioin of different Sideband values (right) , for $4b$ (top) $3b$ (middle) and $2bs$ (bottom). The x-axis indicates different values of SB R_{bb}^{high} cuts.	150
A.2	Fit parameters for μ_{qcd} (left), $\alpha_{\bar{t}t}$ (middle), and ratio of fit uncertainty and stat uncertainty as a functioin of different Sideband shift values (right) , for $4b$ (top) $3b$ (middle) and $2bs$ (bottom). The Sideband region is chose to be $R_{bb}^{high} < 58$ with different shifts, and Control region is fixed at $33 < R_{bb}$. The x-axis indicates different values of SB center shifts.	151

A.3	Fit parameters for μ_{qcd} (left), $\alpha_{\bar{t}\bar{t}}$ (middle), and ratio of fit uncertainty and stat uncertainty as a function of different Control values (right), for $4b$ (top) $3b$ (middle) and $2bs$ (bottom). The Sideband is chosen to be $R_{bh}^{\text{high}} < 58$ with 10 GeV shifts. The x-axis indicates different values of CR R_{bh} cuts.	152
A.4	$1b$ over $0b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.	154
A.5	$2b$ over $1b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.	154
A.6	$2bs$ over $1b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.	155
A.7	$3b$ over $2b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.	155
A.8	$4b$ over $2b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.	156
A.9	$1b$ over $0b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.	156
A.10	$2b$ over $1b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.	157
A.11	$2bs$ over $1b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot. Poor Statistics of the dijet MC affect the pull distributions.	157

A.12	$3b$ over $2b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot. Poor Statistics of the dijet MC affect the pull distributions.	158
A.13	$4b$ over $2b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot. Poor Statistics of the dijet MC affect the pull distributions.	158
A.14	An example of significance comparison plot. On the top pad, the x-axis is for different RSG masses, while the y-axis is plotting the estimated significance. The black dot is for the $2bs$ signal region with the specified selection, the red square is for the $3b$ signal region, and the green star is for the $4b$ signal region. The red cross is the quadratic sum of the three signal regions above. And the black circle is the default selection total combined significance in this note. At the bottom pad, the x-axis is for different RSG masses, while the y-axis is plotting the combined estimated significance ratio of the specified selection to the reference selection. In this case, 77% working point study is compared to the reference, which has the same b -tagging working point. Hence all the ratios are 1.	159
A.15	Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to the blue combined mass limit. Comparison shown here are analysis repeated with calorimeter mass(green).	160
A.16	Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to the blue 70% b -tagging working point limit. Comparison shown here are analysis repeated with 70% working point but with resolved veto (green), 77 % working point (purple), 77% with resolved veto (red).	162
A.17	Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to scaled MJJ limit. Comparison shown here are analysis repeated with unscaled MJJ (green).	163
A.18	Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to the standard trackjet based analysis limit. Comparison shown here are analysis repeated with variable radius track jets.	164

A.19	Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to the standard signal region analysis limit. Comparison shown here are analysis repeated with the signal region defined above (new, green), with the asymmetric width only (purple), and the p_T -dependent X_{hh} cut only (red).	166
A.20	Percentages of Higgs to Jet matching possibilities as a function of the mass of the Randall-Sundrum Graviton in Monte Carlo data. Note that the case where each Higgs gets matched to the same jet essentially never happens, so it is not included in this plot.	169
A.21	ΔR between Higgs (leading by p_T on left, subleading on right) and the child b quark. Lines are drawn at $\Delta R = 0.4$ (size of large R jets for the resolved $HH \rightarrow 4b$ analysis) and $\Delta R = 1.0$ (size for boosted).	170
A.22	Percentage of the time that different combinations of matching b's to trackjets occur relative to all possible combinations of matching the $b\bar{b}$ pair to 2 trackjets (b's inside leading Higgs on left, inside subleading on right). The situations listed in the legend are all independent, but they are not comprehensive. The situations not listed on the legend (including those cases where a b is not in the calorimeter jet) happen in total at most 1.6% of the time for a given mass point.	170
A.23	Percentage of the time that trackjets matched to b's are b-tagged (b's inside leading Higgs on left, inside subleading on right). b-tagging is based on MV2c2o at a cut of -0.6134 (a working point of 77%). Only those cases where both b's are matched to a trackjet are considered. In the case where one b is matched to two jets, the closer of the two jets is checked against. The b's are allowed to match to the same jet.	171
A.24	Reweighted 2bs inclusive region predictions comparison, for leading Higgs Candidate leading trackjet p_T (top left), leading Higgs Candidate subleading trackjet p_T (top right), subleading Higgs Candidate leading trackjet p_T (bottom left), subleading Higgs Candidate subleading trackjet p_T (bottom right). 1b and reweighted 1b distributions are normalized to 2bs.	174
A.25	Reweighted 2bs Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is not blinded, since it is MC distribution. 1b and reweighted 1b distributions are normalized to 2bs.	175
A.26	Reweighted 2bs Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.	176

A.27 Reweighted $3b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.	177
A.28 Reweighted $4b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.	178
A.29 Reweighted $2bs$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for leading Higgs Candidate leading trackjet p_T (first column), leading Higgs Candidate subleading trackjet p_T (second column), subleading Higgs Candidate leading trackjet p_T (third column), subleading Higgs Candidate subleading trackjet p_T (fourth column). The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.	179
A.30 Reweighted $3b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.	180
A.31 Reweighted $3b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for leading Higgs Candidate leading trackjet p_T (first column), leading Higgs Candidate subleading trackjet p_T (second column), subleading Higgs Candidate leading trackjet p_T (third column), subleading Higgs Candidate subleading trackjet p_T (fourth column). The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.	181
A.32 Reweighted $4b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.	182
A.33 Reweighted $4b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for leading Higgs Candidate leading trackjet p_T (first column), leading Higgs Candidate subleading trackjet p_T (second column), subleading Higgs Candidate leading trackjet p_T (third column), subleading Higgs Candidate subleading trackjet p_T (fourth column). The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.	183
A.34 Monte Carlo results on the ΔR between the truth b's as a result of a Higgs decay (b_0 and b_1). Lines are drawn at the diameter of large-R (Higgs) jets for the resolved ($\Delta R = 0.8$) and boosted ($\Delta R = 2.0$) analyses.	185

A.35 ROC plot for values of $m \in \{112.5, 162.5\}$ and $\delta \in \{0, 0.5\}$ for track jets in the sub-leading and leading large R jets. Signal generated from MC on RSG for masses between 1 and 2 TeV.	185
A.36 Significance of the analysis with the cut v. without.	185
A.37 Significance of the analysis with the cut and switching to the 85% working point v. without.	186
A.38 Distribution of dijet mass in each of the control regions after reweighting.	186
A.39 Comparison of $\Delta\eta$ for largeR jets between signal (RSG) and data. Line is drawn at $\Delta\eta = 1.7$, the current value of the cut in the $HH \rightarrow 4b$ analysis.	186
A.40 Ratio of large-R jet pT contained in leading trackjet vs. large-R jet pT . Monte Carlo plots are generated by a series of different RSG masses between 1 and 3 TeV and contain events only from the jet mass signal region. Data covers $6.69 fb^{-1}$ from 2015 and 2016 and contains events from all jet mass regions. Both plots have events for all b-tag regions.	187
A.41 Ratio of large-R jet pT contained in leading trackjet vs. large-R jet pT . Monte Carlo plots are generated by a series of different RSG masses between 1 and 3 TeV and contain events only from the jet mass signal region. Data covers $6.69 fb^{-1}$ from 2015 and 2016 and contains events from all jet mass regions. Both plots have events only from the 2 tag split b-tag region.	187
A.42 For RSG $c = 1.0$ samples, number of events as a function of leading Higgs candidate mass and subleading Higgs candidate mass, for 1.2 TeV(left) signal and 2.0 TeV(right) signal samples.	188
A.43 For RSG $c = 1.0$ samples, the dijet mass distributions and ΔR between the two large-R jets. The left column is $4b$, the middle column is $3b$, and the right column is $2bs$	189
A.44 For RSG $c = 1.0$ samples, the leading higgs candidate's p_T , mass, leading track jet p_T and subleading trackjet p_T . The left column is $4b$, the middle column is $3b$, and the right column is $2bs$	190
A.45 For RSG $c = 1.0$ samples, the subleading higgs candidate's p_T , mass, leading track jet p_T and subleading trackjet p_T . The left column is $4b$, the middle column is $3b$, and the right column is $2bs$	191
A.46 Sideband Region, left column is $2b$ weights, middle column is $3b$ weights, right column is $4b$ weights.	192

A.47 For $2b$ s Sideband Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T . . .	193
A.48 For $3b$ Sideband Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T . . .	194
A.49 For $4b$ Sideband Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T . . .	195
A.50 Signal Region, left column is $2b$ weights, middle column is $3b$ weights, right column is $4b$ weights.	196
A.51 For $2b$ s Signal Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T . . .	197
A.52 For $3b$ Signal Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T	198
A.53 For $4b$ Signal Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T	199

Listing of tables

6.1	Physics objects and their technical names in the boosted analysis.	17
6.2	List of kinematic pre-selection used in the boosted analysis. These cuts are generally efficient for signal. In particular, waiving the requirement on the number of track jets increases the signal efficiency for resonance masses > 2 TeV relative to previous iterations of this analysis.	18
6.3	Cutflow of data, signal samples of two particular mass points, $t\bar{t}$ and $Z + jets$. Uncertainties are the data/MC stat uncertainty.	26
6.4	The selection efficiency for $G_{KK}^* \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events ($c = 1.0$) at each stage of the event selection. Uncertainties are the MC stat uncertainty only.	26
6.5	The selection efficiency for $G_{KK}^* \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events ($c = 2.0$) at each stage of the event selection. Uncertainties are the MC stat uncertainty only.	27
6.6	The selection efficiency for $H \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events at each stage of the event selection.	27
7.1	Definitions of the Signal (SR), Sideband (SB) and Control (CR) regions.	33
7.2	Background scaling parameters (μ_{multijet} and $\alpha_{t\bar{t}}$) estimated from fits to the leading jet mass distributions in $4b/3b/2bs$ sideband regions. $\xi(\mu_{qcd}, \alpha_{t\bar{t}}) = \frac{\text{Cov}(\mu_{qcd}, \alpha_{t\bar{t}})}{\mu_{qcd} \alpha_{t\bar{t}}} . . .$	39
7.3	Smoothing parameters in $4b$ and $3b$ and $2bs$ signal regions, the correlation between parameters is almost always 0.99.	100
7.4	Smoothing parameters in $4b$ and $3b$ and $2bs$ signal regions for scaled mass distributions, the correlation between parameters is almost always 0.99.	106
8.1	Agreement between data and prediction in $4b$ tag CR. Showing stat uncertainty only.	119
8.2	Agreement between data and prediction in $3b$ tag CR. Showing stat uncertainty only.	120
8.3	Agreement between data and prediction in $2bs$ tag CR. Showing stat uncertainty only.	120
8.4	Background prediction in SR/CR/SB for ZZ SR in $4b$ -tag region. Uncertainties are stat only.	123
8.5	Background prediction in SR/CR/SB for ZZ SR in $3b$ -tag region. Uncertainties are stat only.	123

8.6	Background prediction in SR/CR/SB for ZZ SR in $2bs$ -tag region. Uncertainties are stat only.	123
8.7	Agreement between data and prediction in ZZ SR in $4b$, $3b$ and $2bs$ regions.	124
8.8	Background prediction in SR/CR/SB for TT SR in $4b$ -tag region. Uncertainties are stat only.	124
8.9	Background prediction in SR/CR/SB for TT SR in $3b$ -tag region. Uncertainties are stat only.	124
8.10	Background prediction in SR/CR/SB for TT SR in $2bs$ -tag region. Uncertainties are stat only.	126
8.11	Agreement between data and prediction in TT SR in $4b$, $3b$ and $2bs$ regions.	126
8.12	Functions used to fit the QCD dijet mass distributions, where $x = m_{jj}/\sqrt{s}$	133
8.13	Percent impact of the dominant systematics on the background acceptance and on the signal acceptance of RS $c = 1.0$ graviton predictions in the $4b$ signal region.	136
8.14	Percent impact of the dominant systematics on the background acceptance and on the signal acceptance of RS $c = 1.0$ graviton predictions in the $3b$ signal region.	136
8.15	Percent impact of the dominant systematics on the background acceptance and on the signal acceptance of RS $c = 1.0$ graviton predictions in the $2bs$ signal region.	137
A.1	Cutflow of debug stream data in 2015. No events found in signal regions.	201
A.2	Cutflow of debug stream data in 2016. No events found in signal regions.	202

EVERYTHING IS MEANINGLESS. BUDDHA.

Acknowledgments

THANKS TO EVERYONE IN THE ATLAS COLLABORATION, who have supported this remarkable program and have contributed to every bit of the result in my thesis. Sir Issac Newton said he was standing on the gaint's showlders, and I am standing on the ATLAS(member)'s showlders. It's eight stories high so I hope it doesn't shrug. Without their work on detector design, commisioning, operational works, reconstruction, data processing, performance studies and recommendations, software support, computing support, and analysis discussions and guidance, I could not have completed this work. I truely and sincerely thank everyone deeply for their contributions.

0

Introduction

Why do we look for $hh \rightarrow 4b$?

There are two types of analysis in particle physics. The first one is measurement, which yeilds a observable with an uncertainty. This could either imrpove our knowledge of the Standard Model, or show some inconsistency with the Standard Model. The other type is search, which generally assumes some new physics model and try to justfy in data wether the new model is justified in some observables. A successful search turns the subject into a measurement, yet a null result will set a new limit for a given physics model.

Knowledge knows no bounds.

Creator

1

Motivation for searches beyond the Standard Model

THERE'S SOMETHING OUT THERE THAT WE DON'T KNOW.

I.1 THE STANDARD MODEL

I.2 PROBLEMS WITH THE STANDARD MODEL

I.3 PATHS BEYOND THE STANDARD MODEL

I.4 HIGGS SELF COUPLING

There has been many literatures about modifications of Higgs self coupling. Using standard model measurements and their precisions, we can constrain the self coupling parameter to an order of magnitude, see [note](#).

Run.

Random Person

2

Machine of discovery – The Large Hadron Collider

2.1 DESIGN

The deflection angle $d\vartheta$, with radius $\rho(s)$ of a particle with charge Ze and momentum p in a magnetic field $B(s)$ is (see this [talk](#)):

$$d\vartheta = \frac{ds}{\rho(s)} = \frac{ZeB(s)ds}{p} \quad (2.1)$$

Ingerate this over the circumference:

$$\oint_C d\vartheta = 2\pi = \frac{Ze}{p} \oint_C B(s) ds \quad (2.2)$$

Thus the momemtum is:

$$p = \frac{Ze}{2\pi} \oint_C B(s) ds = Z \times 44.7 \left[\frac{MeV}{cTm} \right] \oint_C B(s) ds \quad (2.3)$$

Given LHC has 1232 14.2m long dipoles with B field 8.33 T, we have p is 7.0 TeV/c.

2.2 PERFORMANCE

Above injection energy the relative beam energy uncertainty is 0.1%, fully correlated between the 2 beams. No correction must be applied to the online energy values. Energy and uncertainty are determined by the magnetic model, see [talk](#).

DAMN.

God

3

Eyes of giant–The ATLAS Detector

We love ATLAS. It even has a writing style <https://cds.cern.ch/record/1110290>.

3.1 INNER DETECTOR

Material of ATLAS Inner Detector for Run 2 of the LHC. [note](#).

3.2 TRIGGER AND DATA ACQUISITION

To avoid too high accept rates for certain triggers, the triggers are often prescaled, which means the accepted events get rejected at the prescale. For example, a prescale of two means only every second event passing all trigger conditions gets accepted.

For trigger information, see [2015 note](#) and [2016 updates](#).

I hate my life.

Tony

4

Reconstruction of Objects

4.1 LEPTONS

Leptons are rare in proton-proton collisions. Less than 1% of the total tracks are made from leptons, and these are often related to very interesting physics processes, especially involving the electroweak forces.

Electrons are maximal ionizing in the tracking system. They get absorbed completely in ECAL and leaves no signature in the HCAL. The amplified, shaped and sampled raw signal from each calorimeter cell is converted into the energy deposited. 3×7 cells in the barrel middle layer or 5×5 in the endcap middle layer are clustered together. If the cluster has no associated tracks, it is an "unconverted photon"; if it is matched to a pair of oppositely charged tracks, it is a "converted photon".

A cluster matched to one track could be a converted photon, and the discrimination from electrons comes from the number of hits from the innermost layers of pixel detector. The calibration uses a multivariate regression algorithm, and validated in data using $Z \rightarrow ee$ events, see [note](#).

Muons are special for leaving minimally ionizing signatures in the detector. They usually penetrate the calorimeter and form tracks in the Muon Spectrometer. This is useful for reducing the generate rate of events, hence for triggering. They are also very clean in reconstruction and have excellent resolutions up to 1 TeV. Material and station information, see [note](#).

For Tau ID, see this [note](#)

4.2 JETS

Jets are reconstructed using local topocluster weighting ([LCW](#)) algorithm.

As a sidenote, there can be quark initiated jets and gluon initiated jets. Gluons carry both a color and anti-color charge, and quarks only carry a single color charge. This causes the splitting functions for gluon radiation off a gluon and a gluon radiation off a quark differ by a factor of 9/4. As a result, quark jets have less constituents than gluon jets. Therefore, the number of charged tracks within the jet could help to distinguish a quark jet from a gluon jet, see [note](#).

Jet mass, see [note](#). Jet mass is one of the best tools for distinguishing massive particle decays from QCD background. Jets are trimmed by re-clustering the constituents of the jet into subjets. k_t algorithm with $R_{sub} = 0.2$ is used, and if the subjets has p_T less than 5% of the original jet p_T , the constituent is removed. The calorimeter-based jet mass (m^{calo}), with calorimeter-cell cluster constituents i

with energy E_i , momentum \vec{p}_i ($|\vec{p}_i| = E_i$) is defined as:

$$m^{calo} = \sqrt{\left(\sum_i E_i\right)^2 - \left(\sum_i p_i\right)^2} \quad (4.1)$$

For a boosted massive particle, the angular spread in the decay products scales as $\frac{1}{p_T}$. For highly boosted cases, the spread could be comparable with the o.i.i calorimeter granularity. Tracking information can be used to maintain performance beyond this. The track-assisted jet mass (m^{TA}) is defined as:

$$m^{TA} = \frac{p_T^{calo}}{p_T^{track}} \times m^{track} \quad (4.2)$$

where p_T^{calo} is the calorimeter measurement, p_T^{track} is the four-vector sum of tracks associated to the large-radius calorimeter jet, and m^{track} is the invariant mass of this four-vector sum (the track mass is set to m_{π}). This ratio corrects for charged-to-neutral fluctuations, thus improve the resolution with respect to track-only jet mass.

NEED TO INSERT A FIGURE HERE.

4.3 FLAVOR TAGGING

Tracks reconstruction, see [this great note](#). The first step is creating clusters based on pixel and SCT measured energy deposits, which are space-points. Then three space-points form a seed, and they are combined to build track candidates using a Kalman filter. After ambiguity solving, an artificial neural network is trained and used to identify merged clusters. The last step is a high resolution fit, which is CPU intensive. The min p_T is 400 MeV, and $|\eta| < 2.5$, and at least seven hits in the pixel or SCT. Total number of holes has to be less than two per track, and no more than one in the pixels. $Z_o^{BL} \sin \vartheta$ is required to be less than 3 mm. Interestingly, the performance of track reconstruction is highly dependent on the momentum of the particle. With higher boost, the decay tracks have

smaller separations in the inner detector, hindering the resolving cluster process, and thus degrading the track identification efficiency. For a 1 TeV B_0 , the reconstruction track efficiency is 83%, compared to 95% for a 200 GeV B_0 .

b-tagging. see [2016 run2 note](#). Recurrent Neural Network(RNN), which could explore more the correlation between different input parameters, especially the fragmentation of jets and the impact parameters, has shown improvements in *b*-tagging efficiency and is under investigation for future Multivariate Taggers, see [RNN note](#). *b*-jet calibration is done using $t\bar{t}t$ events, as can be seen in [Likelihood talk](#) and [Matrix method and Likelihood note](#) or a Tag-and-Probe using semi-leptonic $t\bar{t}t$ events.

There are 3 taggers in MV2 family, MV2coo, MV2c10, MV2c20 depending on their charm composition in training: $MV2coo - > 0\%$ c-fraction in the training, $MV2c10 - > 7\%$ c-fraction in the training and $MV2c20 - > 15\%$ c-fraction in the training. See [twiki](#).

Vertex reconstruction and resolution, can be seen in this [note](#). Track resolution can be seen in this [note](#). In Run2, the d_0 resolution is about 10 μm and z_0 is about 50 μm , both decreases as a function of track momentum.

The increase of tracks from fragmentation in the high jet p_T region is the main reason for the performance degradation. As the jet p_T increases, the number of fake vertices is increasing, while the secondary vertex reconstruction efficiency for b and c jets slightly decreases with jet p_T .

See the reference [here](#). Operating points are defined by a single cut value on the discriminant output distribution and are chosen to provide a specific b-jet efficiency on an inclusive ttbar sample. The 77% working point has a rejection factor of 6 and of 134 on charm and light-jets, respectively.

(More information on the working points can be extracted from Table 2 and the related section in ATL-PHYS-PUB-2016-012).

Correction factors are applied to the simulated event samples to compensate for differences between data and simulation in the b-tagging efficiency for b, c and light-jets. The correction for b-jets is derived from $t\bar{t}$ events with final states containing two leptons, and the corrections are consistent with unity with uncertainties at the level of a few percent over most of the jet pT range.

Uncertainties on the correction factors for the b-tagging identification response are applied to the simulated event samples by looking at dedicated flavour-enriched samples in data. An additional term is included to extrapolate the measured uncertainties to the high-pT region of interest. This term is calculated from simulated events by considering variations on the quantities affecting the b-tagging performance such as the impact parameter resolution, percentage of poorly measured tracks, description of the detector material, and track multiplicity per jet. The dominant effect on the uncertainty when extrapolating to high-pT is related to the different tagging efficiency when smearing the track impact parameters based on the resolution measured in data and simulation.

4.4 BOOSTED OBJECT TAGGING

Although this work doesn't depend on specific boosted W/Z/H/Top taggers, many analyses adopt them and improve search sensitivities. For machine learning techniques applied, see this [W/Top tagger using BDT/DNN](#)

Maybe. Just maybe.

Who

5

Censor of work–Data Quality

5.1 DATA FLOW IN ATLAS

5.2 ONLINE MONITORING

5.3 OFFLINE MONITORING

Pick the berry.

Bear

6

Event Selection

6.0.1 DATA CLEANING

Lumiblocks in the 2015 and 2016 data that fail the following GRL are rejected in order to ensure that all detector components are operating correctly:

```
data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02_PHYS_StandardGRL_All_Good_25ns.xml  
data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04_PHYS_StandardGRL_All_Good_25ns.xml
```

In addition the following data cleaning requirements are made in accordance with the Physics-AnalysisWorkBookRel2oCPRec Twiki⁷:

- Bad events due to problems in TileCal are removed

- Bad events due to problems in LAr are removed
- Events affected by the recovery procedure for single event upsets in the SCT are removed
- Incomplete events are removed.

In both data and MC, an event is vetoed if it contains jets with $p_T > 60$ GeV that fail the "Loose-BAD" jet cleaning cuts[?], which are designed to exclude jets caused by detector noise, non-collision backgrounds and cosmic rays.

The analysis also ran over the debug stream, no events passing the full event selection were found. See Appendix A.12.

6.0.2 TRIGGER REQUIREMENTS

Events in data and simulation are required to pass the lowest unprescaled large- R jet trigger: `HLT_j360_a10_lcw` in 2015 and `HLT_j420_a10_lcw` in 2016. These are seeded by the lowest unprescaled L1 jet trigger, `L1_J100`. Both triggers are found to have $> 98\%$ efficiency for all signal masses above 1200 GeV, with the requirement that the event has two large- R jets that satisfy the loose p_T requirements (leading jet $p_T > 400$ GeV, subleading jet leading jet $p_T > 250$ GeV). The results are shown in Figure 6.1. The trigger turn-on curve in 2015 and 2016 data, as a function of leading jet p_T , is shown in Figure 6.2. For more detailed trigger studies, see Appendix ??.

6.0.3 PRESELECTION

The specific physics objects used in the boosted analysis are described in previous sections and reiterated in Table 6.1.

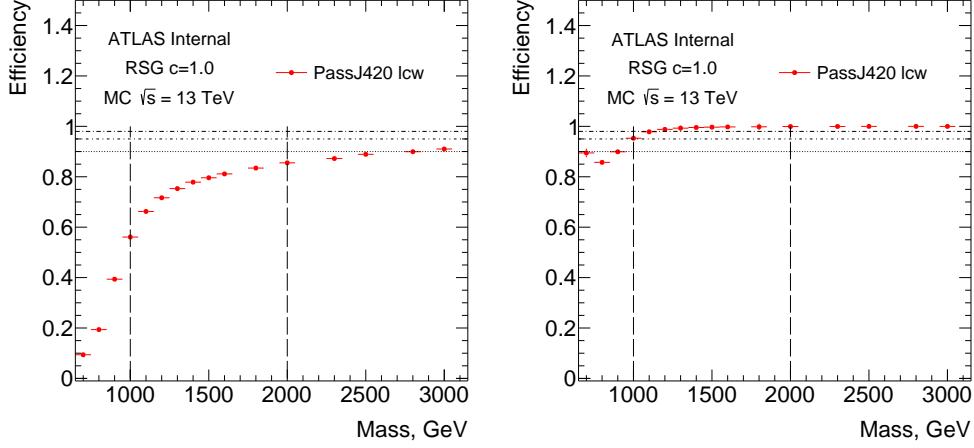


Figure 6.1: Trigger efficiencies as a function of the signal resonance mass with respect to all events with no selection (left) and with respect to events passing the two large- R jets requirement and leading/subleading jet p_T requirement (right). For 1 TeV signal, the trigger efficiency is about 95 %.

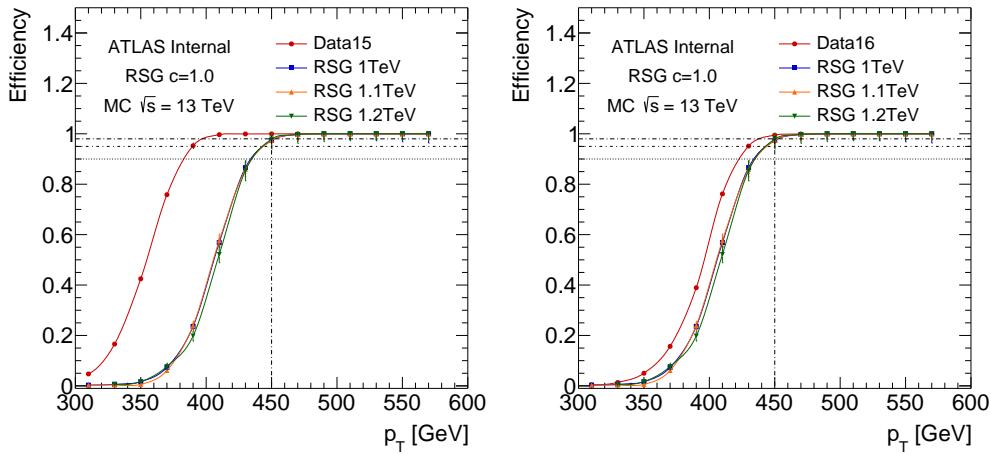


Figure 6.2: Trigger efficiency defined as the fraction of events fired trigger with a given leading jet p_T , measured in 2015 Data (HLT_j360_a10_lcw, left) and 2016 Data (HLT_j420_a10_lcw, right) and MC15c.

In addition to cleaning and trigger requirements, events are pre-selected with simple kinematic criteria shown in Table 6.2 to be consistent with the signal hypotheses. Each event must have at least two high momentum large- R jets, and each of these can have an arbitrary number of ghost associated track jets for b -tagging. The large- R jets are required to have $p_T > 250 \text{ GeV}$, $|\eta| < 2$, and $m > 50 \text{ GeV}$, while the leading large- R jet is required to have $p_T > 450 \text{ GeV}$. The track jets are required to have $p_T > 10 \text{ GeV}$, $|\eta| < 2.5$ and at least two tracks associated with it. The $|\eta|$ and mass thresholds are as recommended by the JetEtmiss performance group.

6.0.4 RESOLVED VETO

In order to have statistical combination with the resolved analysis, events that pass through the resolved signal region selections are vetoed. Specifically, for any event that has at least four AntiKt4EM jets passing $p_T > 40 \text{ GeV}$, $|\eta| < 2.5$, $\text{MV2c10} > 0.8244$ (AntiKt4EM 70% b tagging working point), if the event can make a higgs candidate through the Dhh minimization and passing through the resolved signal region Xhh cut, the event is not considered in the boosted selection. These variables are defined in the resolved supporting note³. For more detail, please see the resolved signal region definition. For the impact on the boosted analysis, see Appendix A.5.3.

6.0.5 SIGNAL REGION SELECTION

After basic object selection, the signal region selection is defined by requiring multiple b -tags and jet masses which are consistent with the Higgs at 125 GeV. The presence of two $h \rightarrow b\bar{b}$ decays in the

object	technical name
large- R calorimeter jets	AntiKt10LCTopoTrimmedPtFrac5SmallR20Jets
small- R track jets	AntiKt2PV0TrackJets
b -tagging	BTagging_AntiKt2Track, MV2c10, 70% working point

Table 6.1: Physics objects and their technical names in the boosted analysis.

final state naturally suggests requiring 4 track jets passing b -tagging requirements, and this is defined as the “ $4b$ ” selection. However, since this requirement has an overall efficiency of roughly ϵ^4 , where ϵ is the efficiency of the tagger, a “ $3b$ ” selection is also introduced to improve the signal efficiency, especially at high masses. In $4b$ and $3b$, one Higgs candidate can have at most two b -tagged trackjets, hence $\geq 3b$ -tagged trackjets cannot be in the same $large - R$ jet. At the highest resonance massed, the Lorentz boost of the Higgs boson can be large enough to collimate the daughter b-quarks below the distance scale resolvable by the track jets ($R = 0.2$). Motivated by this, we introduce a third signal region — which we denote by ”two-tag-split” or simply ” $2bs$ ” — in which exactly one b -tagged track jet is found in each Higgs candidate (plus an arbitrary number of track jets that do not pass the b -tag).

One major change since the 2015 analysis is the requirement on the number of track jets found in each $large - R$ jet. In the past, each $large - R$ jet was required to contain at least two track jets in order to reduce and understand backgrounds. However, above 2 TeV, this cut is inefficient for the signal, when the b-track-jets merge. To recover this efficiency, we do not make any explicit requirement on the number of track jets in each $large - R$ jet. Thus, an event with 3 b -tags must have at least 3 b -tagged track jets, but can have any number of additional un-tagged track jets. Similarly, an event with $2bs$ must have at least one b -tagged track jet in each of the $large - R$ jets. Again, the 70% b -jet

kinematic pre-selection

- at least two large- R jets with $p_T > 250 \text{ GeV}$, $|\eta| < 2$, $m > 50 \text{ GeV}$
- leading large- R jet must have $p_T > 450 \text{ GeV}$, to be above the trigger turn-on plateau
- the leading and sub-leading large- R jets are considered the Higgs candidates
- the Higgs candidates (large- R jets) must have $|\Delta\eta| = |\eta_{\text{leadjet}} - \eta_{\text{subjet}}| < 1.7$

Table 6.2: List of kinematic pre-selection used in the boosted analysis. These cuts are generally efficient for signal. In particular, waiving the requirement on the number of track jets increases the signal efficiency for resonance masses $> 2 \text{ TeV}$ relative to previous iterations of this analysis.

efficiency working point is used in the analysis.

To separate di-Higgs-boson decays from other multi-jet productions like QCD multi-jets and top, requirements on the leading and subleading large- R jet masses are imposed. The signal region is defined using the expression:

$$X_{hh} = \sqrt{\left(\frac{m_j^{\text{lead}} - 124 \text{ GeV}}{0.1(m_j^{\text{lead}})}\right)^2 + \left(\frac{m_j^{\text{subl}} - 115 \text{ GeV}}{0.1(m_j^{\text{subl}})}\right)^2} \quad (6.1)$$

The denominator of each term in the definition can be interpreted as a resolution on the reconstructed mass of 10% for the leading and subleading jets, hence X_{hh} can be interpreted as a χ^2 compatibility with the hh hypothesis. Similarly to the resolved analysis, these $\sigma(m_j)$ are only a rough approximation to the true resolution, but the X_{hh} requirement gives nearly optimal performance. The subleading jet mass value of 115 GeV is chosen after investigating the signal jet masses in MC and noticing that the subleading large- R jet typically has a reconstructed mass which is biased downward. This is due to the ordering of the large- R jets in p_T , which biases the sub-lead jet towards lower energy. The energy losses result from neutrinos in leptonic b decays, cracks in the calorimeter, and other effects. The signal region requires $X_{hh} < 1.6$, same as the resolved analysis.

A similar circular variable can be defined in the two-dimensional mass plane, R_{hh} . The circular region R_{hh} has the same central values as X_{hh} , but without resolution terms in the denominators and is defined as:

$$R_{hh} = \sqrt{(m_j^{\text{lead}} - 124 \text{ GeV})^2 + (m_j^{\text{subl}} - 115 \text{ GeV})^2} \quad (6.2)$$

The region defined by $1.6 < X_{hh}$ and $R_{hh} < 33$ will be used later in the definition of control regions, as discussed in Section 7.1.3. The cut value was optimized to allow for a reasonable sized sample (twice the statistics as the signal region) in the control region with kinematics similar to the

signal region, whilst avoiding the large contributions of the $t\bar{t}$ sample when the large- R jets have a mass near the top quark mass (i.e. with $m_J > 160$ GeV).

Similarly, R_{hh}^{high} , the circular region that has the shifted central values up by 10 GeV is defined using the variable:

$$R_{hh}^{\text{high}} = \sqrt{(m_J^{\text{lead}} - 134 \text{ GeV})^2 + (m_J^{\text{subl}} - 125 \text{ GeV})^2} \quad (6.3)$$

In Run-1, and in the the 2015 analysis, the sideband was defined to be all events not in the signal or control regions. However, the kinematics of events with very large and very small large- R jet masses may not be the same as those within the signal region. Thus, to avoid biasing effects from extremely low mass or extremely high mass large- R jets, the sideband region is also redesigned to be like the control region, but at $R_{hh} > 33$ and $R_{hh}^{\text{high}} < 58$. The shift upwards helps to capture enough $t\bar{t}$ events in the normalization estimates, as described in Section 7.1.5. A similar sideband definition with an upper limit has been made for the resolved analysis.

The values of the X_{hh} and R_{hh} variables can be seen graphically in Figure 6.3.

6.0.6 MUON-IN-JET CORRECTIONS

A further correction to account for energy loss due to leptonic b -hadron decays with a muon in the final state is applied to the large- R jets. Combined muons with $p_T > 4$ GeV, $|\eta| < 2.5$, and passing at least the *medium* MCP quality requirement are ΔR matched to the track jets within the large- R jets. The muons are required to be matched to the b -tagged track jets in order to focus only on semi-leptonic b -hadron decays. In case more than one muon is found within a track jet, only the muon closest in ΔR is considered. If more than one b -tagged track jet is found to have a muon, the muon identified in each track jet are considered (i.e. if two b -tagged track jets are found to have muons, than both muons are considered). The four-momenta of the muons is then added

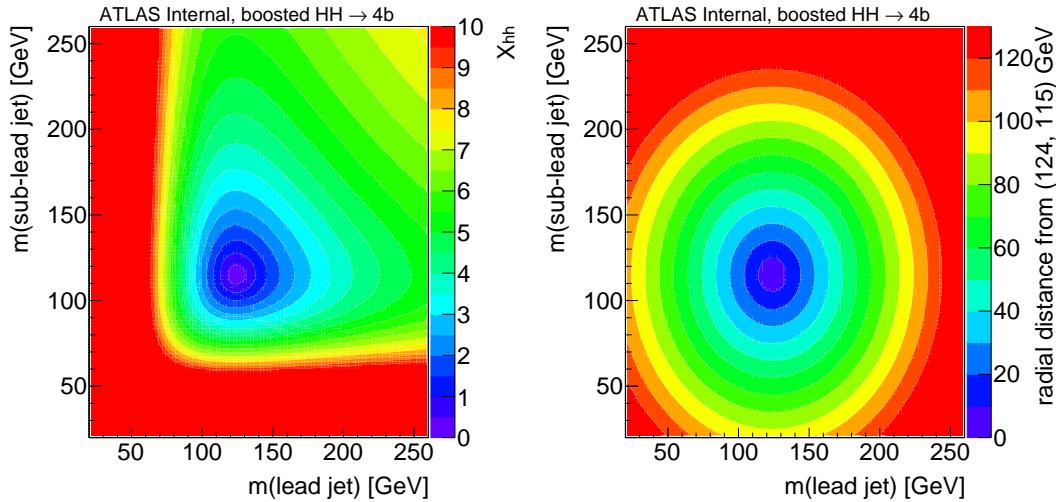


Figure 6.3: Values of the X_{hh} and R_{hh} variables, which are shapes in the two-dimensional plane of the large-R jet masses used to define signal, control, and sideband regions. For both variables, a smaller value indicates the jets are closer to the Higgs mass.

to the large- R jet four-momentum, with the muon calorimeter energy deposite subtracted (after trimming), to better reconstruct the Higgs boson candidate. This correction is only applied to the calorimeter mass portion of the combined mass. The muon-in-jet correction improves the large- R jet mass resolution by approximately 5%.

The muon-in-jet corrections are applied only after preselection, i.e., the fiducial large- R jet requirements on p_T and η are made before the corrections. This is because the uncorrected jet four-momentum is used in the derivation and ntuple-making steps of the analysis. The muon-in-jet corrections will affect the large- R jet mass plane cuts to define the signal region and the final discriminating variable, m_{jj} .

6.0.7 SIGNAL EFFICIENCY

The boosted analysis covers multiple TeV of signal resonance masses, and the backgrounds fall steeply as a function of the reconstructed di-Higgs mass. Understanding how the signal efficiency trends with resonance mass can be helpful in gauging the natural limitations of very boosted objects. It can also indicate ways to improve sensitivity, as the resolved analysis has implemented with mass-dependent cuts, though the boosted analysis is currently not exploring this option.

The signal efficiency as a function of resonance mass is shown in Figure 6.4, both for the absolute signal efficiency and for the efficiency relative to the previous cut in the selection. Above a mass of ~ 1 TeV, the reconstruction of high momentum large- R jets with small $\Delta\gamma$ is efficient. Across the mass range considered, the signal jet masses requirement (X_{hh}) and b -tagging requirements are $\mathcal{O}(20\%)$ efficient relative to the previous cuts.

Above a mass of ~ 2 TeV, the requirement of two track jets per large- R jet becomes increasingly inefficient due to the merging of the track jets. This motivates introducing a selection without a requirement on the number of track jets, and to sort the signals based on the corresponding number of b -tagged track jets, as shown in Figure 6.5. At high masses above 2.5 TeV, the $2b$ s region (where each large- R jet has exactly one b -tagged track jet) significantly improves the sensitivity. In this figure, $2b$ region is defined as a mutually exclusive region (where one large- R jet has two b -tagged track jets, and the other one has no b -tagged track jet). The number of events in the control region and sideband region as a function of Resonance mass is shown in Figure 6.6. For $2b$ s, $3b$ and $4b$, each region has the number of events decrease from signal region to control region to sideband regions.

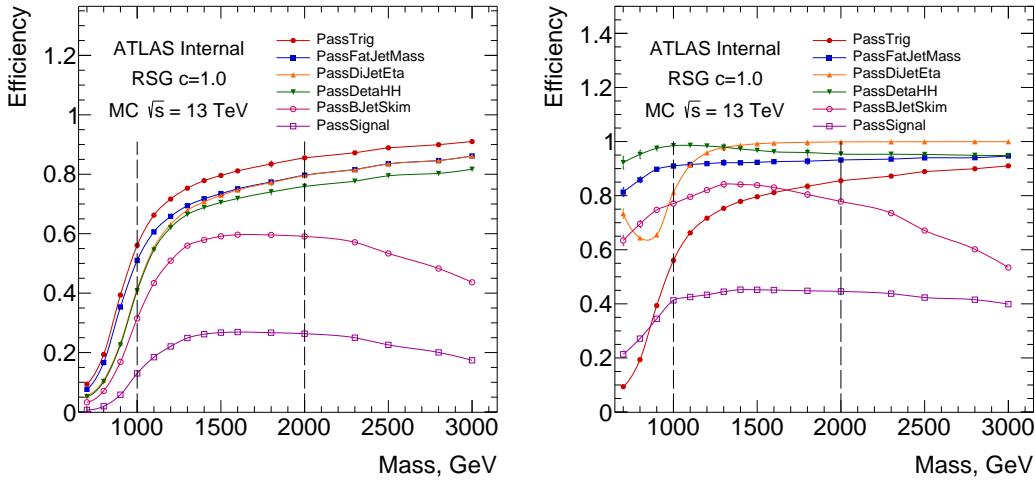


Figure 6.4: Absolute (left) and relative (right) signal efficiency as a function of RSG $c=1.0$ signal resonance mass hypothesis for selection cuts. The relative efficiency is defined from the previous cut, where the order of cuts is given by the legend. PassTrig means the event passes the trigger selection; PassDiJetPt means the event passes the leading and sub-leading jet p_T cuts; PassDiJetEta means the event passes the leading and sub-leading jet η cuts; PassDataH means the events passes the $|\Delta\eta| < 1.7$ cut; PassBJetSkim means the event contains at least two b -tagged track jets, inclusive of $2b$, $2bs$, $3b$ and $4b$ configurations; PassSignal means the event passes the signal region cut $X_{hh} < 1.6$.

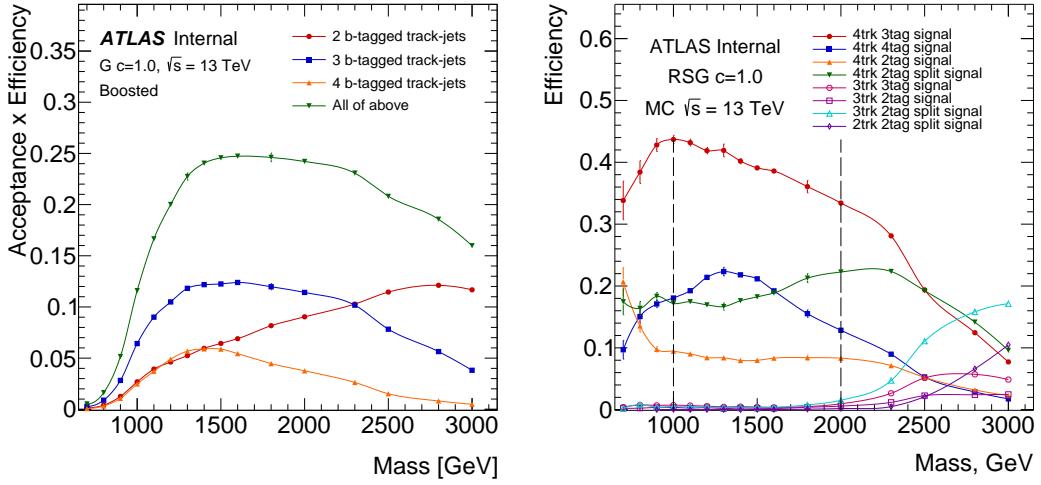


Figure 6.5: Signal efficiency in three search b -tag categories (left) and detailed signal efficiency in different track jet and b -tag categories (right) as a function of signal resonance mass hypothesis for selection cuts. The right plot efficiencies are relative to the total number of events in the preselection, whereas the left plot efficiencies are relative to the total number of events in the signal mass region. The green curve in the left plot also corresponds to the PassSignal curve on in Figure 6.4.

6.0.8 CUTFLOW

Table 6.3 shows the cutflow numbers in data, two different graviton signal MC mass points, $t\bar{t}$ MC samples and $Z+jets$ MC samples. The selection efficiency at various stages for RS graviton (RSG) $c = 1.0$ (narrow width scalar resonance) signal samples of all mass points can be found in Table 6.4, 6.5 and 6.6. Here Sideband indicates the Sideband region as defined before in Equation 6.3, which is not the whole region outside Control and Signal region.

6.1 TRIGGER SELECTION

See 6.4.3 for the 2015 trigger. In data and simulation, there is a small offset in the efficiency curves for the threshold. The trigger `HLT_j360_a10_lcw_sub`, where topo-cluster jets with local calibration weights and pile-up subtraction, is used.

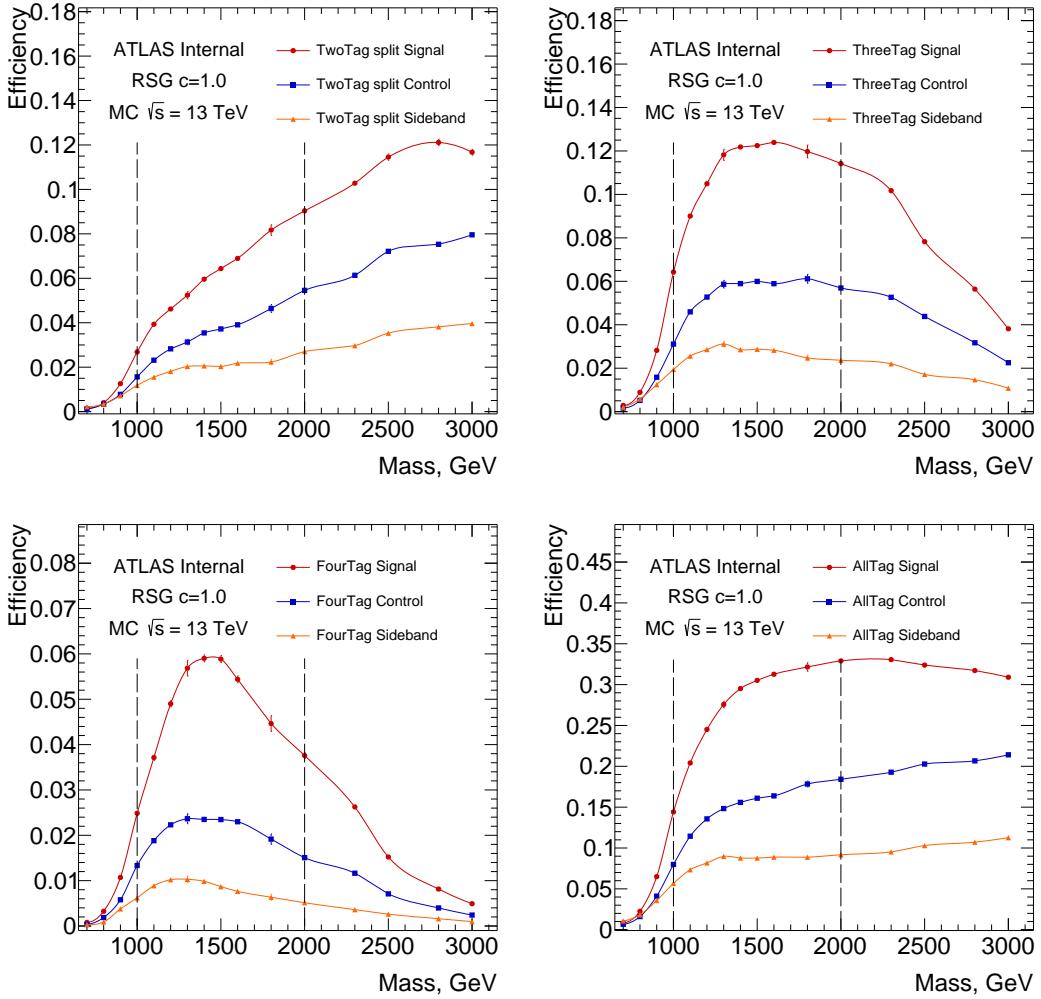


Figure 6.6: Detailed signal efficiency in different signal/control/sideband regions as in $2b\bar{s}$ (top left, $3b$ (top right), $4b$ (bottom left) and inclusive b-tagged regions, which include $2b$, $1b$ and $0b$ as well, (bottom right) as a function of signal resonance mass hypothesis for selection cuts. The efficiencies are relative to the total number of events in the preselection.

Cut	Data	$m_G = 1\text{TeV}$	$m_G = 2\text{TeV}$	$m_G = 3\text{TeV}$	$t\bar{t}$	$Z + jets$
Initial	23889570.0 ± 4887.7	238.2 ± 0.98	5.72 ± 0.021	0.3 ± 0.0011	210298.2 ± 415.91	35770.71 ± 365.93
Pass GRL	22416007.0 ± 4734.55	238.2 ± 0.98	5.72 ± 0.021	0.3 ± 0.0011	210298.2 ± 415.91	35770.71 ± 365.93
Pass Trigger	21360954.0 ± 4621.79	226.99 ± 0.96	5.71 ± 0.021	0.3 ± 0.0011	190758.37 ± 392.63	30708.09 ± 329.84
Pass Jet Cleaning	21358219.0 ± 4621.5	226.98 ± 0.96	5.71 ± 0.021	0.3 ± 0.0011	190741.16 ± 392.62	30701.46 ± 329.8
N(fiducial large-R jets) ≥ 2	21358219.0 ± 4621.5	226.98 ± 0.96	5.71 ± 0.021	0.3 ± 0.0011	190741.16 ± 392.62	30701.46 ± 329.8
Pass Large-R jet Selection	9852994.0 ± 3138.95	167.76 ± 0.82	5.32 ± 0.02	0.29 ± 0.0011	122171.9 ± 299.65	18466.6 ± 241.77
$ \Delta\eta(J/\psi) < 1.7$	7671253.0 ± 2769.7	165.2 ± 0.82	5.07 ± 0.019	0.27 ± 0.0011	104353.21 ± 280.17	16879.76 ± 231.61
0 b-tags, Sideband	990693.0 ± 995.34	6.63 ± 0.054	0.029 ± 0.0016	0.0041 ± 0.00014	7562.7 ± 74.79	2634.39 ± 91.9
0 b-tags, Control	401071.0 ± 633.3	0.51 ± 0.048	0.035 ± 0.0018	0.0066 ± 0.00019	1557.48 ± 33.42	1432.14 ± 67.2
0 b-tags, Signal	196205.0 ± 442.95	0.5 ± 0.048	0.044 ± 0.0021	0.0082 ± 0.00021	713.8 ± 22.77	263.61 ± 29.93
1 b-tags, Sideband	269076.0 ± 518.73	3.77 ± 0.13	0.15 ± 0.0036	0.015 ± 0.00027	15677.95 ± 105.56	885.1 ± 62.2
1 b-tags, Control	104862.0 ± 323.82	3.72 ± 0.13	0.23 ± 0.0045	0.026 ± 0.00036	3323.38 ± 47.64	431.92 ± 44.1
1 b-tags, Signal	51791.0 ± 227.58	4.98 ± 0.15	0.33 ± 0.0055	0.036 ± 0.00042	1726.68 ± 34.02	71.72 ± 18.42
2 b-tags, Sideband	28146.0 ± 167.77	3.35 ± 0.12	0.066 ± 0.0023	0.0017 ± 8.8e-05	1249.73 ± 36.31	333.49 ± 33.25
2 b-tags, Control	11204.0 ± 105.85	3.76 ± 0.13	0.12 ± 0.0032	0.0035 ± 0.00013	261.07 ± 17.22	212.37 ± 32.03
2 b-tags, Signal	5495.0 ± 74.13	5.92 ± 0.16	0.21 ± 0.0042	0.0052 ± 0.00015	136.35 ± 11.5	23.94 ± 10.09
2 b-tags, split, Sideband	25137.0 ± 158.55	4.79 ± 0.14	0.18 ± 0.0039	0.013 ± 0.00025	7960.88 ± 71.27	67.74 ± 16.82
2 b-tags, split, Control	8486.0 ± 92.12	6.33 ± 0.16	0.36 ± 0.0056	0.027 ± 0.00034	1505.09 ± 29.65	26.44 ± 10.08
2 b-tags, split, Signal	blinded	10.87 ± 0.22	0.6 ± 0.0072	0.039 ± 0.00041	870.13 ± 22.53	0.13 ± 0.091
3 b-tags, Sideband	4403.0 ± 66.36	7.86 ± 0.18	0.16 ± 0.0035	0.0036 ± 0.00013	1066.24 ± 32.15	32.8 ± 11.34
3 b-tags, Control	1553.0 ± 39.41	12.58 ± 0.23	0.38 ± 0.0054	0.0075 ± 0.00018	202.91 ± 13.94	11.21 ± 5.65
3 b-tags, Signal	blinded	26.0 ± 0.33	0.76 ± 0.0076	0.013 ± 0.00023	99.18 ± 10.73	0.49 ± 0.49
4 b-tags, Sideband	204.0 ± 14.28	2.52 ± 0.1	0.034 ± 0.0015	0.00032 ± 3.7e-05	31.25 ± 5.79	0 ± o
4 b-tags, Control	81.0 ± 9.0	5.4 ± 0.15	0.1 ± 0.0026	0.0008 ± 5.6e-05	7.16 ± 3.31	6.18 ± 5.12
4 b-tags, Signal	blinded	10.07 ± 0.2	0.25 ± 0.0041	0.0016 ± 8e-05	1.89 ± 1.37	o ± o

Table 6.3: Cutflow of data, signal samples of two particular mass points, $t\bar{t}$ and $Z + jets$. Uncertainties are the data/MC stat uncertainty.

Resonance Mass [GeV]	Mini-ntuple Skimming	2 large-R jets	$\Delta\gamma$	Xhh < 1.6	2bs SR	3b SR	4b SR
500	317.31 ± 6.0	295.75 ± 5.79	164.5 ± 4.32	8.45 ± 0.99	1.08 ± 0.37	2.14 ± 0.52	o ± o
600	269.07 ± 3.64	247.94 ± 3.5	136.31 ± 2.59	11.31 ± 0.76	2.57 ± 0.37	3.84 ± 0.45	0.66 ± 0.19
700	253.68 ± 3.35	226.93 ± 3.16	124.83 ± 2.35	16.79 ± 0.86	3.74 ± 0.42	6.99 ± 0.56	1.91 ± 0.29
800	286.26 ± 2.28	245.36 ± 2.11	129.2 ± 1.53	24.41 ± 0.67	5.11 ± 0.31	11.27 ± 0.46	4.13 ± 0.27
900	306.51 ± 1.61	275.57 ± 1.52	158.03 ± 1.15	40.72 ± 0.59	8.81 ± 0.28	19.76 ± 0.41	7.5 ± 0.25
1000	238.2 ± 0.98	226.98 ± 0.96	165.2 ± 0.82	52.86 ± 0.47	10.87 ± 0.22	26.0 ± 0.33	10.07 ± 0.2
1100	164.5 ± 0.63	160.94 ± 0.63	132.53 ± 0.57	45.26 ± 0.34	9.55 ± 0.16	21.88 ± 0.23	9.03 ± 0.14
1200	109.24 ± 0.41	107.92 ± 0.4	93.45 ± 0.38	33.53 ± 0.23	6.96 ± 0.11	15.8 ± 0.16	7.38 ± 0.1
1300	72.72 ± 0.59	72.2 ± 0.59	63.74 ± 0.56	24.19 ± 0.35	5.02 ± 0.17	11.33 ± 0.24	5.45 ± 0.16
1400	48.83 ± 0.17	48.61 ± 0.17	42.96 ± 0.16	16.62 ± 0.1	3.72 ± 0.052	7.61 ± 0.07	3.68 ± 0.046
1500	33.13 ± 0.12	33.02 ± 0.12	29.25 ± 0.11	11.31 ± 0.07	2.67 ± 0.036	5.08 ± 0.047	2.44 ± 0.031
1600	22.81 ± 0.08	22.75 ± 0.08	20.16 ± 0.075	7.74 ± 0.048	1.93 ± 0.025	3.48 ± 0.032	1.53 ± 0.02
1800	11.2 ± 0.1	11.18 ± 0.1	9.93 ± 0.094	3.71 ± 0.059	1.1 ± 0.034	1.6 ± 0.038	0.6 ± 0.022
2000	5.72 ± 0.021	5.71 ± 0.021	5.07 ± 0.019	1.83 ± 0.012	0.6 ± 0.0072	0.76 ± 0.0076	0.25 ± 0.0041
2250	2.61 ± 0.0088	2.61 ± 0.0088	2.32 ± 0.0083	0.78 ± 0.005	0.31 ± 0.0032	0.3 ± 0.003	0.078 ± 0.0014
2500	1.24 ± 0.0054	1.24 ± 0.0054	1.11 ± 0.0051	0.33 ± 0.0028	0.16 ± 0.002	0.11 ± 0.0016	0.021 ± 0.00066
2750	0.6 ± 0.0026	0.6 ± 0.0026	0.54 ± 0.0025	0.14 ± 0.0013	0.081 ± 0.00099	0.038 ± 0.00065	0.0055 ± 0.00024
3000	0.3 ± 0.0011	0.3 ± 0.0011	0.27 ± 0.0011	0.058 ± 0.00051	0.039 ± 0.00041	0.013 ± 0.00023	0.0016 ± 8e-05

Table 6.4: The selection efficiency for $G_{KK}^* \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events ($c = 1.0$) at each stage of the event selection. Uncertainties are the MC stat uncertainty only.

Resonance Mass [GeV]	Mini-ntuple Skimming	2 large-R jets	$\Delta\eta$	Xhh < 1.6	2bs SR	3b SR	4b SR
500	3705.15 ± 40.86	3479.44 ± 39.59	2325.18 ± 32.37	568.04 ± 16.17	122.56 ± 7.76	253.49 ± 10.78	100.7 ± 6.53
600	2549.14 ± 22.55	2374.01 ± 21.76	1591.92 ± 17.82	396.96 ± 9.03	89.01 ± 4.46	178.63 ± 6.05	74.31 ± 3.71
700	1928.4 ± 13.57	1782.83 ± 13.04	1183.86 ± 10.63	320.53 ± 5.62	71.38 ± 2.76	148.41 ± 3.81	59.31 ± 2.31
800	1595.14 ± 8.82	1457.71 ± 8.43	958.89 ± 6.84	268.75 ± 3.67	64.14 ± 1.86	123.48 ± 2.47	49.43 ± 1.51
900	1264.78 ± 5.77	1179.88 ± 5.58	819.75 ± 4.65	251.29 ± 2.61	55.63 ± 1.27	119.44 ± 1.79	48.72 ± 1.11
1000	891.0 ± 3.66	856.95 ± 3.59	662.54 ± 3.15	219.04 ± 1.84	49.45 ± 0.91	104.31 ± 1.26	42.97 ± 0.78
1100	595.58 ± 2.98	581.72 ± 2.95	481.67 ± 2.68	167.96 ± 1.61	37.64 ± 0.79	78.28 ± 1.09	34.59 ± 0.7
1200	390.84 ± 1.69	385.41 ± 1.68	330.23 ± 1.55	118.0 ± 0.94	26.23 ± 0.46	54.18 ± 0.64	25.34 ± 0.42
1300	257.66 ± 0.94	255.35 ± 0.94	222.37 ± 0.88	82.11 ± 0.54	19.04 ± 0.27	37.8 ± 0.37	16.99 ± 0.23
1400	172.09 ± 0.72	171.02 ± 0.71	150.22 ± 0.67	56.23 ± 0.42	13.6 ± 0.22	25.36 ± 0.28	11.78 ± 0.18
1500	116.25 ± 0.41	115.72 ± 0.41	101.92 ± 0.39	38.5 ± 0.24	9.94 ± 0.13	17.04 ± 0.16	7.64 ± 0.1
1600	80.09 ± 0.28	79.82 ± 0.28	70.48 ± 0.26	26.24 ± 0.16	7.01 ± 0.09	11.62 ± 0.11	4.92 ± 0.067
1800	38.99 ± 0.14	38.9 ± 0.14	34.39 ± 0.13	12.65 ± 0.081	3.82 ± 0.047	5.46 ± 0.053	2.02 ± 0.03
2000	19.94 ± 0.088	19.91 ± 0.088	17.68 ± 0.083	6.17 ± 0.05	2.15 ± 0.031	2.52 ± 0.032	0.85 ± 0.017
2250	9.02 ± 0.031	9.01 ± 0.031	7.99 ± 0.029	2.62 ± 0.017	1.03 ± 0.011	1.02 ± 0.01	0.28 ± 0.0051
2500	4.28 ± 0.016	4.28 ± 0.016	3.8 ± 0.015	1.13 ± 0.0083	0.52 ± 0.0058	0.4 ± 0.0048	0.098 ± 0.0022
3000	1.07 ± 0.004	1.07 ± 0.004	0.96 ± 0.0038	0.23 ± 0.0019	0.13 ± 0.0015	0.062 ± 0.00097	0.013 ± 0.00043

Table 6.5: The selection efficiency for $G_{KK}^* \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events ($c = 2.0$) at each stage of the event selection. Uncertainties are the MC stat uncertainty only.

Resonance Mass [GeV]	Mini-ntuple Skimming	2 large-R jets	$\Delta\eta$	Xhh < 1.6	2bs SR	3b SR	4b SR
500	1557.94 ± 136.12	1022.77 ± 110.29	95.14 ± 33.64	11.69 ± 11.69	0 ± 0	11.69 ± 11.69	0 ± 0
600	3289.78 ± 123.99	2542.11 ± 108.99	485.99 ± 47.66	54.55 ± 15.77	18.73 ± 9.39	9.17 ± 6.49	0 ± 0
700	4655.21 ± 94.59	3855.64 ± 86.09	1237.8 ± 48.78	142.42 ± 17.03	28.55 ± 7.74	52.6 ± 10.75	7.69 ± 3.85
800	7506.31 ± 81.79	6020.56 ± 73.25	2150.64 ± 43.78	320.63 ± 17.02	67.63 ± 7.83	139.97 ± 11.23	47.57 ± 6.75
900	9732.89 ± 61.17	8400.91 ± 56.83	3574.63 ± 37.07	806.13 ± 17.76	188.71 ± 8.71	377.92 ± 12.12	127.7 ± 6.99
1000	7516.07 ± 37.72	7033.18 ± 36.49	4496.85 ± 29.18	1351.1 ± 16.2	303.71 ± 7.88	650.89 ± 11.19	234.2 ± 6.57
1100	4731.39 ± 21.54	4563.4 ± 21.15	3485.58 ± 18.49	1135.18 ± 10.7	251.77 ± 5.18	539.51 ± 7.36	215.39 ± 4.5
1200	2853.51 ± 12.23	2782.42 ± 12.07	2253.95 ± 10.87	786.92 ± 6.53	175.53 ± 3.21	366.01 ± 4.44	158.29 ± 2.8
1300	1700.83 ± 7.01	1668.05 ± 6.94	1362.91 ± 6.27	494.36 ± 3.84	107.0 ± 1.86	224.19 ± 2.58	107.55 ± 1.7
1400	1016.14 ± 4.03	999.98 ± 4.0	802.44 ± 3.59	296.46 ± 2.22	65.49 ± 1.1	133.86 ± 1.49	65.58 ± 0.99
1500	621.34 ± 2.41	613.46 ± 2.39	484.92 ± 2.13	179.29 ± 1.32	42.75 ± 0.68	79.32 ± 0.88	36.77 ± 0.56
1600	386.3 ± 1.46	382.44 ± 1.46	297.63 ± 1.28	109.99 ± 0.8	27.68 ± 0.42	49.3 ± 0.53	20.92 ± 0.32
1800	154.5 ± 0.58	153.5 ± 0.57	116.52 ± 0.5	42.24 ± 0.31	12.41 ± 0.18	18.2 ± 0.2	6.66 ± 0.11
2000	65.4 ± 0.24	65.02 ± 0.24	48.57 ± 0.2	16.89 ± 0.12	5.64 ± 0.076	7.01 ± 0.079	2.18 ± 0.041
2250	23.84 ± 0.085	23.73 ± 0.085	17.44 ± 0.073	5.57 ± 0.042	2.25 ± 0.028	2.11 ± 0.025	0.52 ± 0.012
2500	9.2 ± 0.032	9.17 ± 0.032	6.72 ± 0.028	1.9 ± 0.015	0.92 ± 0.011	0.64 ± 0.0086	0.11 ± 0.0034
2750	3.73 ± 0.013	3.73 ± 0.013	2.71 ± 0.011	0.63 ± 0.0054	0.37 ± 0.0042	0.17 ± 0.0027	0.021 ± 0.00093
3000	1.59 ± 0.0054	1.59 ± 0.0054	1.15 ± 0.0046	0.22 ± 0.0021	0.15 ± 0.0017	0.044 ± 0.0009	0.0038 ± 0.00025

Table 6.6: The selection efficiency for $H \rightarrow hh \rightarrow b\bar{b}b\bar{b}$ events at each stage of the event selection.

6.2 OPTIMIZATION STRATEGY

To improve the analysis, a quantity which defines the sensitivity of analysis is maximized. Technically, the optimal sensitivity is described as $\sqrt{2((S + B) \ln i + \frac{S}{B} - S)}$, see [note](#). This is usually considered at the $S \ll B$ limit and simplified as $\frac{S}{\sqrt{B}}$, when no knowledge of the model cross section is available, or $\frac{S}{\sqrt{S+B}}$, if the signal cross section is known. These parametrizations have limitations, particularly when the signal yield and the number of estimated background are both small. A better parametrization for low signal strength is $\frac{S}{\sqrt{i+B}}$, where the extra $\sqrt{i+B}$ accounts for poisson fluctuations. For a discussion of p-values, please see this [note](#).

For this analysis, two methods are used: one is calculate the number of signal and backgrounds within 68% of the signal m_{hh} mass window, the other is to implement the full signal and background predictions after smoothing and compare the asymptotic expected exclusion limits. Both methods yield comparable results.

To avoid bias during the optimization process, data's signal regions are blinded.

6.3 b TAGGING

b -tagging, which is the identification of the b hadron, ⁱ is the core and main limiting factor of this analysis. Because of the relatively long lifetime, it is possible to tag the b hadron using the inner detector informations. A higher b -tagging efficiency will increase the signal selection efficiency, while a lower b -tagging fake rate will reduce the background like $gg \rightarrow \bar{c}c$ in the signal regions.

6.4 SIGNAL EFFICIENCY

Acceptance refers to purely geometric fiducial volume of the detector. Efficiency refers to purely detector effectiveness in finding objects. The final value in the study is Acceptance \times Efficiency,

where both effects are considered.

Tough part.

Revolutionaries.

7

Background Estimation

7.1 BACKGROUND ESTIMATION

7.1.1 OVERVIEW

The primary backgrounds to this analysis in the four b -jet signal region, in order of size, are QCD multi-jet production ($\sim 95\%$), $t\bar{t}$ ($\sim 5\%$), and $Z+jets$ ($< 1\%$), where the percentages are the expected fraction of the background coming from each source. In the three b -jet signal region, the fractions are QCD ($\sim 90\%$), $t\bar{t}$ ($\sim 10\%$), and $Z+jets$ ($< 1\%$). In the two b -jet split signal region, the fractions are QCD ($\sim 80\%$), $t\bar{t}$ ($\sim 20\%$), and $Z+jets$ ($< 1\%$).

QCD is by far the dominant background. However, there is no reliable, high-statistics Monte Carlo simulation sample in this region of phase space (i.e with three or four b -jets collected into two high- p_T large radius jets) and thus a data-driven background estimation is needed. (See Appendix A.3.) For the $t\bar{t}$ background, Monte Carlo simulation samples of reasonable size are available, and thus can be used to guide an estimation of this background. The $Z+jets$ background is small enough that we will rely on the Monte Carlo simulation of $Z+heavy flavor jets$. $ZZ \rightarrow b\bar{b}b\bar{b}$ has been estimated to be completely negligible using a particle-level analysis, with less than one event expected after three b -tags are required, which will be further heavily suppressed by the X_{hh} requirement.

The QCD background prediction relies on finding a region which is similar enough in event properties that it can be used to estimate the shapes of the expected QCD background. This region is identical to the signal region defined by the full selection, with the exception that the events must have less b -tagged track jets:

- For the $2bs$ category, the $1b$ sample is used for modeling.
- For the $3b$ and $4b$ categories, the $2b$ sample - where the two b -tagged trackjet are in the same large- R jet - is used for modeling.

To prevent differences in the number of track jets from biasing the dijet mass distribution, the $1b$ -tagged region requires that each large- R jet has at least one track jet (to model $2bs$, ie. $2b$ tag split). Similarly, the $2b$ -tagged region requires that one large- R jet has at least one track jet and the other one has at least two track jets (to model $3b$), and each large- R jet has at least two track jets (to model $4b$).

However, this less b -tagged region only supplies the shapes of the expected background and not the total yield, and a second control sample, which we denote the *Sideband* region, is used to estimate the yield. The Sideband is obtained by doing the full analysis selection, except instead of the

X_{bb} cut an alternative criteria on the large radius jet masses is used, that $33 < R_{bb}$ and $R_{bb}^{\text{high}} < 58$ GeV. To validate this approach, a third region, which we denote the *Control* region, is centered on the signal region in the plane of the two large radius jet masses but does not include the signal region, such that $R_{bb} < 33$ GeV. The control region is used to validate the background estimations before unblinding. The control and sideband regions are optimized, as shown in the following sections, to accurately estimate the rate of the QCD background (and thus allow for an extrapolation from the $1b/2b$ estimate to a prediction in the $4b/3b/2bs$ signal regions), whilst giving a control region which has kinematic properties similar to that of the signal region.

The $t\bar{t}$ background shape is taken from MC. A data-based estimation of the $t\bar{t}$ background yield is performed simultaneously with the QCD background yield estimation, by means of a binned likelihood fit. In the plane of the two leading large radius jet masses, the main contribution of the $t\bar{t}$ background lies in the Sideband region. The data distribution in the Sideband region of the leading- p_T large radius jet mass is fit simultaneously with the QCD shape estimate (from the less b -tagged sample) and with the $t\bar{t}$ Monte Carlo shape. This fit is done separately in the $4b$, $3b$, and $2bs$ Sideband regions. From this fit, two terms are determined simultaneously: μ_{QCD} and α_{tt} . μ_{QCD} is the ratio of the QCD event yield in the $2bs/3b/4b$ regions to the amount in each corresponding less b -tagged region. α_{tt} is the ratio of the fitted ttbar event yield to the yield predicted from ttbar MC. These two numbers are then used as multiplicative constants in other regions of the mass plane (i.e. the Control or Signal regions) to extrapolate from the rates of the less b -tagged regions to predictions of rates in the $2bs/3/4$ b -tagged regions, for estimating the amount of QCD, and to correct the rate of $t\bar{t}$ production wrt. MC. Hence, the underlying assumption is that these scale factors are roughly constant over the 2D large-R jet mass plane for Sideband/Control/Signal regions, which has been verified by performing these fits in small bins across the 2D mass plane. This is shown in Appendix A.3. The correction factors are derived separately for the $4b$, $3b$, and $2bs$ regions.

In this section, we describe this approach in more detail and show its validation in data.

7.1.2 DEFINITION OF THE SIDEBAND AND CONTROL REGIONS (SB, CR)

The definitions of the SB, CR, and SR in the leading (m_J^{lead}) and sub-leading (m_J^{subl}) large- R jet mass plane are found in Table 7.1. These regions can be seen in the leading and sub-leading large- R jet mass plane in Figure 7.1. As a reminder, the definition of X_{hh} , R_{hh} and R_{hh}^{high} are:

$$X_{hh} = \sqrt{\left(\frac{m_J^{\text{lead}} - 124 \text{ GeV}}{\sigma(m_J^{\text{lead}})}\right)^2 + \left(\frac{m_J^{\text{subl}} - 115 \text{ GeV}}{\sigma(m_J^{\text{subl}})}\right)^2}, R_{hh} = \sqrt{(m_J^{\text{lead}} - 124)^2 + (m_J^{\text{subl}} - 115)^2}, R_{hh}^{\text{high}} = \sqrt{(m_J^{\text{lead}} - 132)^2 + (m_J^{\text{subl}} - 125)^2} \quad (7.1)$$

Region	Definition
Signal Region (SR)	$X_{hh} < 1.6$
Control Region (CR)	$R_{hh} < 33 \text{ GeV}$ and $X_{hh} > 1.6$
Sideband Region (SB)	$33 \text{ GeV} < R_{hh} \text{ and } R_{hh}^{\text{high}} < 58 \text{ GeV}$

Table 7.1: Definitions of the Signal (SR), Sideband (SB) and Control (CR) regions.

The CR is chosen to be as close as possible to the signal region, thus allows a good test for the background predictions, avoids the top mass peak around 175 GeV, and still gives reasonably statistics. The SB definition is optimized so as to also be a reasonable proxy for the events contained in the CR and SR. Being farther from the SR means that the exact kinematics will not be the same, but one can avoid very large and very small mass jets not present in the SR by appropriate choice of SB. The optimization of the CR and SB definitions can be found in Appendix A.2. The choice of SB's impact on the predicted QCD background normalization, which is derived from the SB, can be found in Appendix A.3.

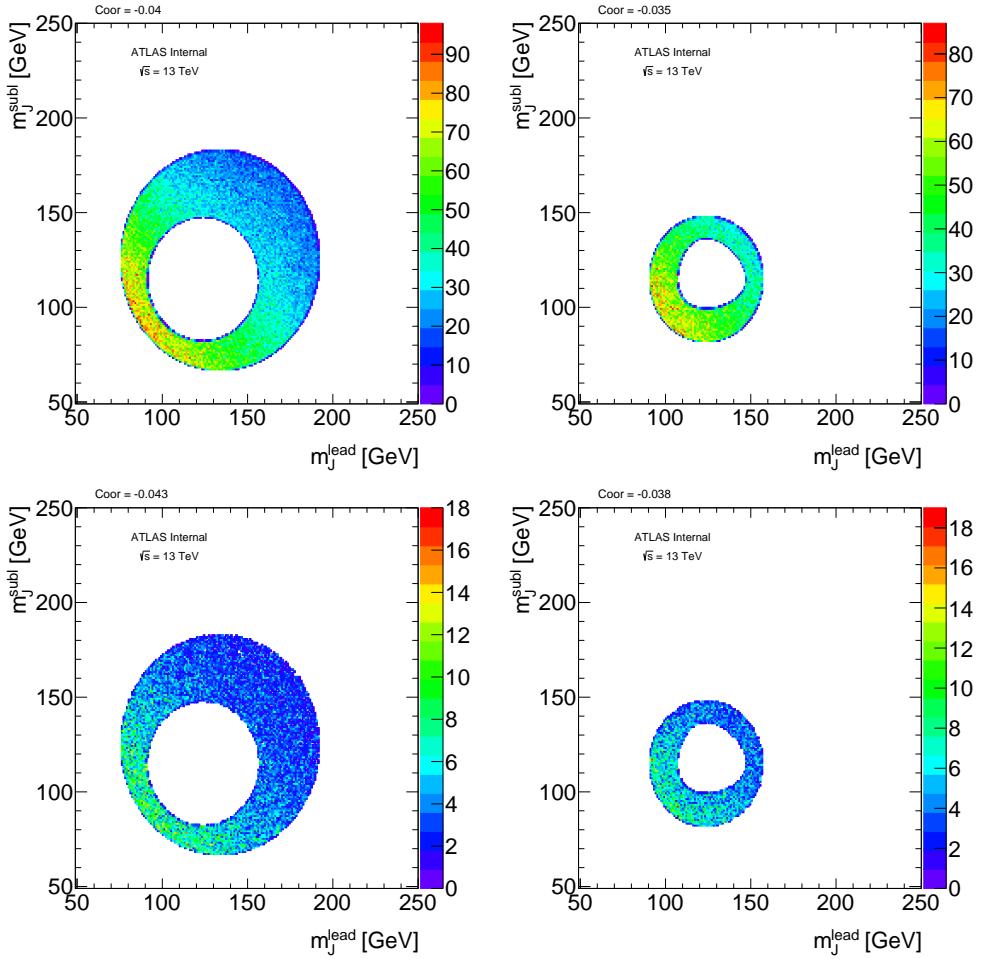


Figure 7.1: m_j^{lead} vs. m_j^{subl} in data in the $1b$ -tag (top) and $2b$ -tag (bottom) selection, the plots show the boundary between the Sideband (left) and Control (right) regions.

7.1.3 QCD MULTI-JETS

The QCD multi-jets prediction relies on finding a region which is similar enough in event properties so that it can be used to estimate the shapes of the expected background. This region is defined to be identical to the signal region except requiring both of the large- R jets to pass the ≥ 2 (like in $4b$) or $\geq 1/2$ (like in $3b$) or ≥ 1 (like in $2b$ split) track jet requirement, but have two or one associated b -tagged track jets only on one of the large- R jets. However, this $1b/2b$ region only provides shapes of the expected background and not the total yield.

It should be noted that the $1b/2b$ region is orthogonal to the $4b/3b/2bs$ signal regions. In addition, the MC predicted $t\bar{t}$ events in the $1b/2b$ regions are subtracted from the data to produce the $1b/2b$ QCD estimation. This procedure follows closely the method used in Run 1 and also used in the resolved analysis, but requiring $1b$ -tag for the $2bs$ background estimation.

It should also be noted that the $2b$ sample is further split into 80% – 20% parts, where each is used separately for $3b$ and $4b$ background estimations. This ensures the shape estimations of $3b$ and $4b$ QCD estimates are uncorrelated.

It should also be noted that the resolved veto will impact the $4b$ background estimation. Specifically, $2b$ events are excluded when they have at least two resolved jets that are b tagged (passing resolved 70% working point) and passing the resolved $Xbb < 1.6$ cut if using two other non b -tagged resolved jets to make the Higgs candidates. This ensures that a similar sculpting effect is reflected in the background estimation, and a check for this can be found in Appendix ??.

Given the $1b/2b$ samples, which predict shapes for the QCD background, the normalization of the QCD background is determined in the sideband by fitting the leading jet mass distribution simultaneously with QCD and $t\bar{t}$ background templates, as described in section 7.1.5. This fit gives a scaling factor for QCD, called μ_{QCD} (and for $t\bar{t}$, called $\alpha_{t\bar{t}}$) which can be applied to scale the $1b/2b$

predictions in the CR or SR to the predicted normalizations in those regions.

It should be noted that there can be kinematic differences between the $1b/2b$ samples and the $4b/3b/2bs$ regions. Thus a kinematic reweighting is applied to correct for such differences, as described in Section 7.1.6.

7.1.4 $t\bar{t}$ BACKGROUND

The number of $t\bar{t}$ events in the signal region coming mainly from the all-hadronic decay mode (with a smaller contribution from the leptonic + jets decay mode) comprises of around 5-20% of the inclusive total background in the $4b/3b/2bs$ regions due to the high p_T threshold imposed on the leading *large - R* calorimeter jet. In addition, the normalization and the shape of $t\bar{t}$ events in the sideband region can affect the QCD estimate described in the previous section.

For the normalization of the $t\bar{t}$ background, we start with the MC prediction estimated by scaling the MC sample by the cross section and luminosity, and then applying the boosted event selection. To account for possible differences between data and MC, a normalization scaling factor is derived from a fit to data in the sideband region. Although estimated in the SB, this normalization scaling factor will be applied also in the CR and SR. Separate fits are done in the $4b/3b/2bs$ SB, thus deriving separate normalization scaling factor for the $4b/3b/2bs$ samples. For the shape of the $t\bar{t}$ background, no data driven methods were identified, and thus the MC shape is used.

However, it should be noted that in the $4b$ and $3b$ signal region, there were not sufficient MC statistics to get a reasonable shape estimate. As a result, in the $4b$ and $3b$ signal region, the $2bs$ shapes will be used (but the normalization will still be that estimated for the $4b/3b$ sample). Since the same shape is used for the $4b/3b/2bs$ SR predictions of $t\bar{t}$, the shape systematics are considered correlated in the final results and limit setting. A comparison between the $4b/3b/2bs$ shapes for the di-large- R -jet mass distributions (the final discriminant) in the SR can be found in Figure 7.2. As we can see, the shapes are compatible, with the $4b$ having much larger statistical uncertainties. Differences between these distributions will be used as a systematic, as described in Section 8.

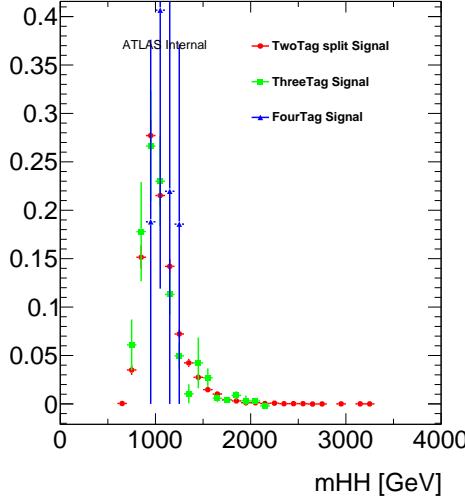


Figure 7.2: comparison between the $2b$, $3b$, and $4b$ shapes for the di-large- R -jet mass distributions (the final discriminant) in the SR.

7.1.5 FITTING PROCEDURE FOR QCD AND $t\bar{t}$ NORMALIZATION

The number of $4b/3b/2b$ s events in data observed in a given region (SB / CR / SR) can be described as:

$$N_{\text{data}}^b = \mu_{\text{qcd}}^b N_{\text{qcd}}^{xb} + \alpha_{t\bar{t}}^b N_{t\bar{t}}^b + N_{Z+jets}^b \quad (7.2)$$

where ν_b is the number of b -tagged track jets required, x is 1 for $2b$ s and 2 for $3b$ and $4b$. μ_{multijet} is essentially an estimate of the ratio of the number of QCD events with ν_b b -tagged track jets, to the number of $1b/2b$ QCD events, while the $t\bar{t}$ normalization parameter $\alpha_{t\bar{t}}$ applied after the $t\bar{t}$ is scaled to the total integrated luminosity, is a correction to the MC prediction in this phase space. The same equation can be applied to the $4b/3b/2b$ s region (replacing ν_b by $4b/3b/2b$ s b in Equation 7.2).

In order to constrain the QCD and $t\bar{t}$ background normalizations using data, a simultaneous fit is applied to extract both the $t\bar{t}$ normalization with respect to the yields from simulation and the number of $1b/2b$ data events for the QCD background. These scaling parameters are determined independently for the $4b/3b/2s$ signal regions. But as the procedure is the same for those three signal regions, we denote these scaling factors simply μ_{multijet} and $\alpha_{t\bar{t}}$ in the following text.

A binned maximum likelihood fit is employed to find the values of μ_{multijet} and $\alpha_{t\bar{t}}$, as well as the correlation between the two parameters. The fit is performed on the leading- p_T jet mass spectrum in the sideband region, as it has the best separation between QCD and $t\bar{t}$ shapes. Due to the $p_T > 450$ GeV cut imposed on the leading *large* – R_{jet} , the hadronic top quark is likely to be fully reconstructed inside of the *large* – R_{jet} and the leading jet mass in the $t\bar{t}$ sample has a clean peak around $M = 170$ GeV in the sideband region.

The values of μ_{multijet} and $\alpha_{t\bar{t}}$ as estimated by the fits in the $4b/3b/2bs$ sideband regions can be found in Table 7.2, along with the correlation of the fitted parameters $\rho(\mu_{qcd}, \alpha_{t\bar{t}}) = \frac{\text{Cov}(\mu_{qcd}, \alpha_{t\bar{t}})}{\sqrt{\text{Cov}(\mu_{qcd})} \sqrt{\text{Cov}(\alpha_{t\bar{t}})}}$. μ_{multijet} and $\alpha_{t\bar{t}}$ are approximately 70% negatively correlated, which is not surprising as they are the only two components fit to the data distribution and their sum needs to predict the SB total event count.

Sample	μ_{qcd}	$\alpha_{t\bar{t}}$	$\rho(\mu_{qcd}, \alpha_{t\bar{t}})$
FourTag	0.0332 ± 0.00428	0.891 ± 0.599	-0.785
ThreeTag	0.163 ± 0.00434	0.8 ± 0.0733	-0.72
TwoTag split	0.0627 ± 0.000573	0.986 ± 0.0186	-0.47

Table 7.2: Background scaling parameters (μ_{multijet} and $\alpha_{t\bar{t}}$) estimated from fits to the leading jet mass distributions in $4b/3b/2bs$ sideband regions. $\rho(\mu_{qcd}, \alpha_{t\bar{t}}) = \frac{\text{Cov}(\mu_{qcd}, \alpha_{t\bar{t}})}{\sqrt{\text{Cov}(\mu_{qcd})} \sqrt{\text{Cov}(\alpha_{t\bar{t}})}}$

Figure 7.3 shows the post-fit spectrum of the leading *large-R* calorimeter jet mass in the $4b/3b/2b$ sideband regions. The normalization of $t\bar{t}$ is constrained by the top quark mass peak around 170 GeV. The shapes of the data is also well modeled by the predicted background. The fitting errors on μ_{multijet} and $\alpha_{t\bar{t}}$ are applied as systematic uncertainties taking into account their correlation. This will be explained in more detail in the systematics section.

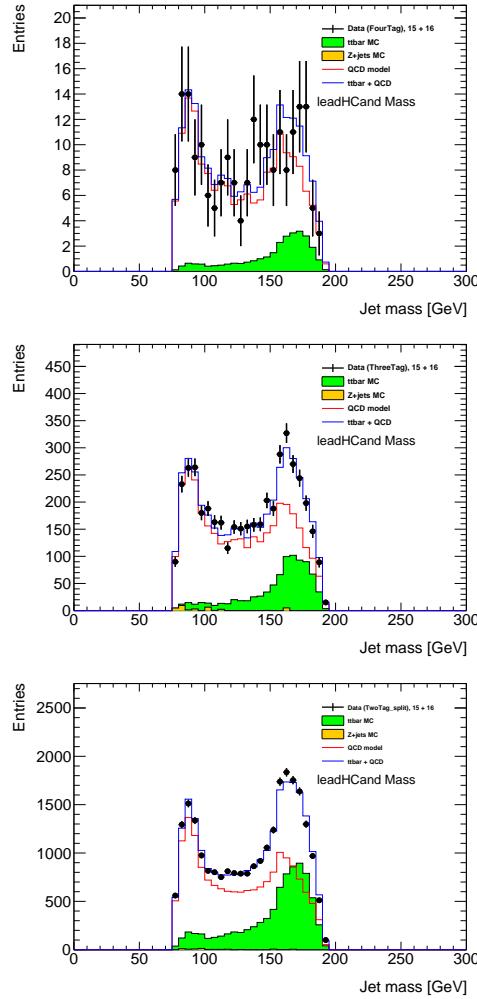


Figure 7.3: Simultaneous fit of μ_{multijet} and $\alpha_{t\bar{t}}$ in $4b$ (top) and $3b$ (middle) and $2b$ (bottom) sideband region using leading *large-R* calorimeter jet mass spectrum.

7.1.6 KINEMATIC REWEIGHTING

Due to the large contribution from the completely data-driven QCD background, it is important to model this background as good as possible in all regions of the analysis. Using the $1/2b$ region to model the $2bs$, $3b$, and $4b$ regions can introduce discrepancies in the modeling of the estimated QCD background versus the real $n b$ data. These discrepancies arise possibly from the non-trivial effect that b -tagging has on jet kinematics. The natural choice of reweighting variable is the p_T of the track jets in the event, since these are the objects that we apply the b -tagging to. Also, large- R jet p_T is also reweighted to account for the effect from light and charm quark composition difference at different energy scales. The three chosen variables are the leading large- R jet p_T , leading large- R jet leading trackjet p_T and subleading large- R jet leading trackjet p_T .

In order to account for the b -tagging effect, a reweighting on the $1/2b$ data is adopted. The basic idea is to reweight the non b -tagged Higgs candidate to have kinematic distributions just like a b -tagged Higgs candidate. The idea is demonstrated in Figure 7.4. It shows that $2bs$ has very similar kinematic distributions on the trackjet p_T as the $1b$ sample, when the variable is the b tagged trackjet.

For $2bs$, the $1b$ non-tagged Higgs candidate is reweighted to be like a $1b$ tagged Higgs candidate; for $3b$, the $2b$ non-tagged Higgs candidate is reweighted to be like a $1b$ tagged Higgs candidate; for $4b$, the $2b$ non-tagged Higgs candidate is reweighted to be like a $2b$ tagged Higgs candidate. For each category, the events are split into two orthogonal subgroups, depending on whether leading/subleading Higgs candidate is b -tagged, the event is then reweighted such that the untagged Higgs candidate's distribution behaves like the corresponding b -tagged Higgs candidate's.

To avoid potential biases in the final distributions used for the analysis, a reweighting technique is applied to the $1/2b$ data only. Since each signal region is modeled by a different $1/2b$ tag category: $2bs$ by $1b$ tag events with at least 1 track jets on both large- R jets, $3b$ by $2b$ tag events with at least one

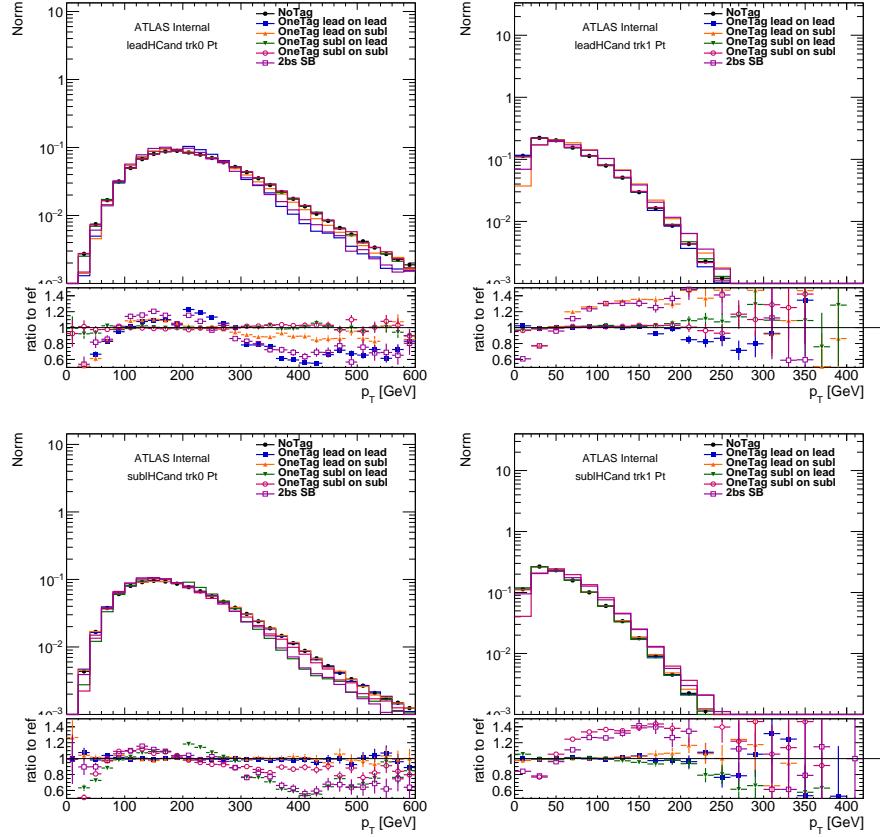


Figure 7.4: Comparison of different trackjet p_T distributions. Top row is for leading p_T Higgs candidate, and bottom row is for subleading p_T Higgs candidate. Left column is for the leading p_T trackjet of the Higgs candidate, and right column is for the subleading p_T trackjet of the Higgs candidate. Shown in the plot are just data distributions, inclusive of SB, CR, and SR regions for $0b$ and $1b$, while for $2bs$ only the SB region is shown. $1b$ sample is further split into four subcategories, depending on which trackjet gets b tagged. OneTag lead on lead means the b tagged trackjet is the leading trackjet of the leading Higgs candidate, OneTag lead on subl means the b tagged trackjet is the subleading trackjet of the leading Higgs candidate, OneTag subl on lead means the b tagged trackjet is the leading trackjet of the subleading Higgs candidate, and OneTag subl on subl means the b tagged trackjet is the subleading trackjet of the subleading Higgs candidate. At the bottom ratio plot, all the ratio are taken with respect to the $0b$ tagged distribution.

track jets on one large- R jet and at least two track jets on the other large- R jet, and $4b$ by $2b$ tag events with at least two track jets on both large- R jets, the reweighting procedure is the same but orthogonal for the three different channels. Note the $2b$ sample is already split into separate parts, as described in paragraph 7.1.3.

The detailed procedure is listed as follows:

- Subtracting $1/2b$ tag $t\bar{t}$ and $Z+jets$ samples in the sideband from the $1/2b$ tag data in the Sideband + Control + Signal regions to get the $1/2b$ QCD inclusive estimate.
- Separate the $1/2b$ tag sample further to sample A. that has the b -tagged Higgs is the leading p_T Higgs candidate, and B. that the b -tagged Higgs is the subleading p_T Higgs candidate.
- For each variable, i.e. the large- R jet p_T , normalize sample A to sample B total number of events, take the ratio of sample A distribution over sample B distribution, and fit the ratio with a spline function. (TSpline3)
- Use this functional form to extract reweighting values for each variable that is considered. The reweighting value for each variable is also constrained to be within a -30% to $+40\%$ range compared to one, to avoid over corrections and failed fit situations.
- For each event, all the weights are multiplied together to change the $1/2b$ tag data event weight. Another constraint is applied, such that each total reweighting value is constrained to be within a 10% to $+1000\%$ range compared to one, again to avoid over corrections.
- The reweighting is done on the three variables: large- R jet p_T and the two track jet $p_{T\gamma}$ s, which is counted as one iteration of reweighting.
- A total of ten iterations are used to stabilize the reweighting. The reweighting is roughly converging after three iterations.

For reweighting method comparisons and validations in data and Dijet MC, see Appendix A.8.

REWEIGHTING FITS

The first iteration, second iteration, and last iteration of fits for $2b$ s, where in $1b$ data, the non-tagged Higgs candidate are reweighted to be like a $1b$ tagged Higgs candidate, can be seen in Figure 7.5 and 7.6. Similar distributions for $3b$, where in $2b$ data, the non-tagged Higgs candidate are

reweighted to be like a $1b$ tagged Higgs candidate, are shown in Figure 7.7 and 7.8. Similar distributions for $4b$, where in $2b$ data, the non-tagged Higgs candidate are reweighted to be like a $2b$ tagged Higgs candidate, are shown in Figure 7.9 and 7.10. The before reweighting distribution (first row), the reweighting result after the first iteration (second row), and the final distribution after reweighting (last row) are presented.

It should be noted that in some plots, like Figure 7.9 and 7.10, the last ratio bin sometimes still doesn't converge to unity. This is a feature from the limited statistics from the last bin, especially in the $4b$ case, where only 20% number of events in $2b$ is used for background prediction and therefore reweighted. One could choose a different binning and use more iterations to help this converge to one, yet the last bin's few event will also likely to end up with a large unphysical weight and therefore harm the background prediction later.

For the distribution of weights and the weight as a function of different kinematic ranges, see Appendix A.II.

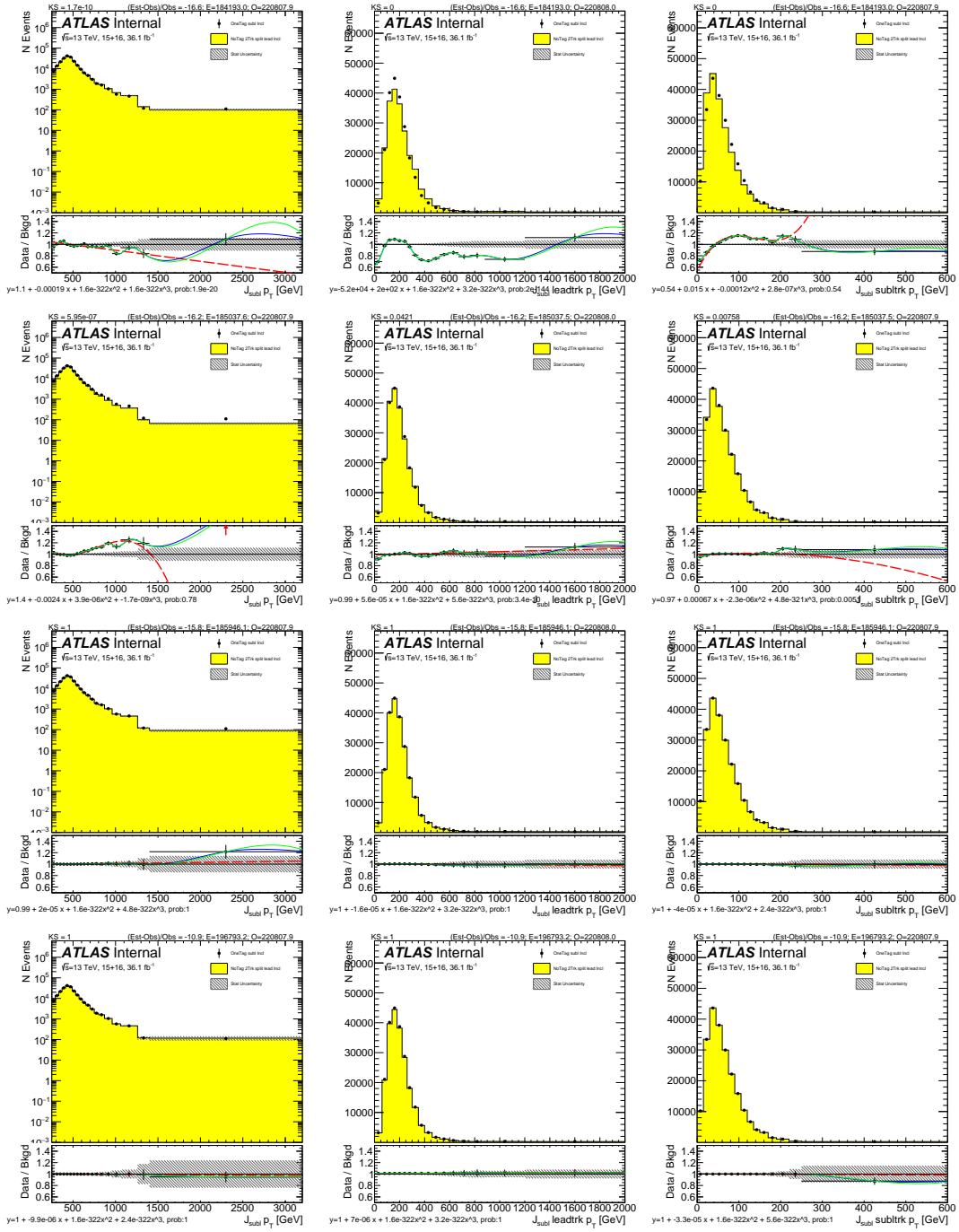


Figure 7.5: For zbs background estimate: the fits to the ratio of the data in the $1b$ category, of the subleading Higgs candidate $1b$ -tagged events's subleading Higgs candidate distributions(black point), over the leading Higgs candidate $1b$ -tagged events's subleading Higgs candidate distributions(yellow). Distributions and fits to the estimated QCD background for large- R jet p_T (left), the large- R jet's leading trackjet p_T (middle), and large- R jet's subleading trackjet p_T (right) are shown. Figure are shown before reweighting (top row), after the first iteration(second row), after the fourth iteration(third row), and after the last iteration (bottom row). The green line is the spline extrapolation; and the red line is a polynomial fit.

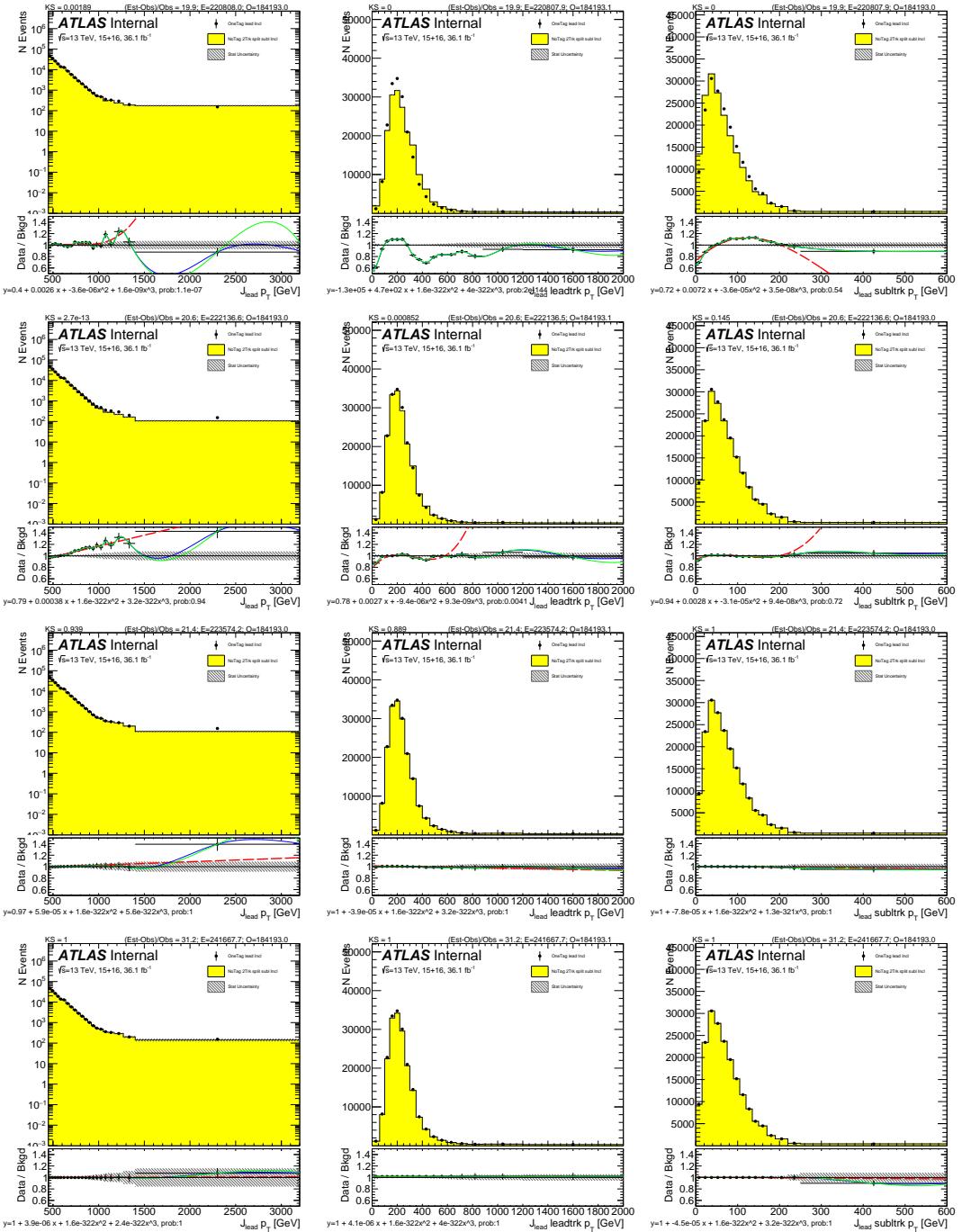


Figure 7.6: For $z\bar{b}s$ background estimate: the fits to the ratio of the data in the $1b$ category, of the leading Higgs candidate $1b$ -tagged events's leading Higgs candidate distributions(black point), over the subleading Higgs candidate $1b$ -tagged events's leading Higgs candidate distributions(yellow). Distributions and fits to the estimated QCD background for large- R jet p_T (left), the large- R jet's leading trackjet p_T (middle), and large- R jet's subleading trackjet p_T (right) are shown. Figure are shown before reweighting (top row), after the first iteration(second row), after the fourth iteration(third row), and after the last iteration (bottom row). The green line is the spline extrapolation; and the red line is a polynomial fit.

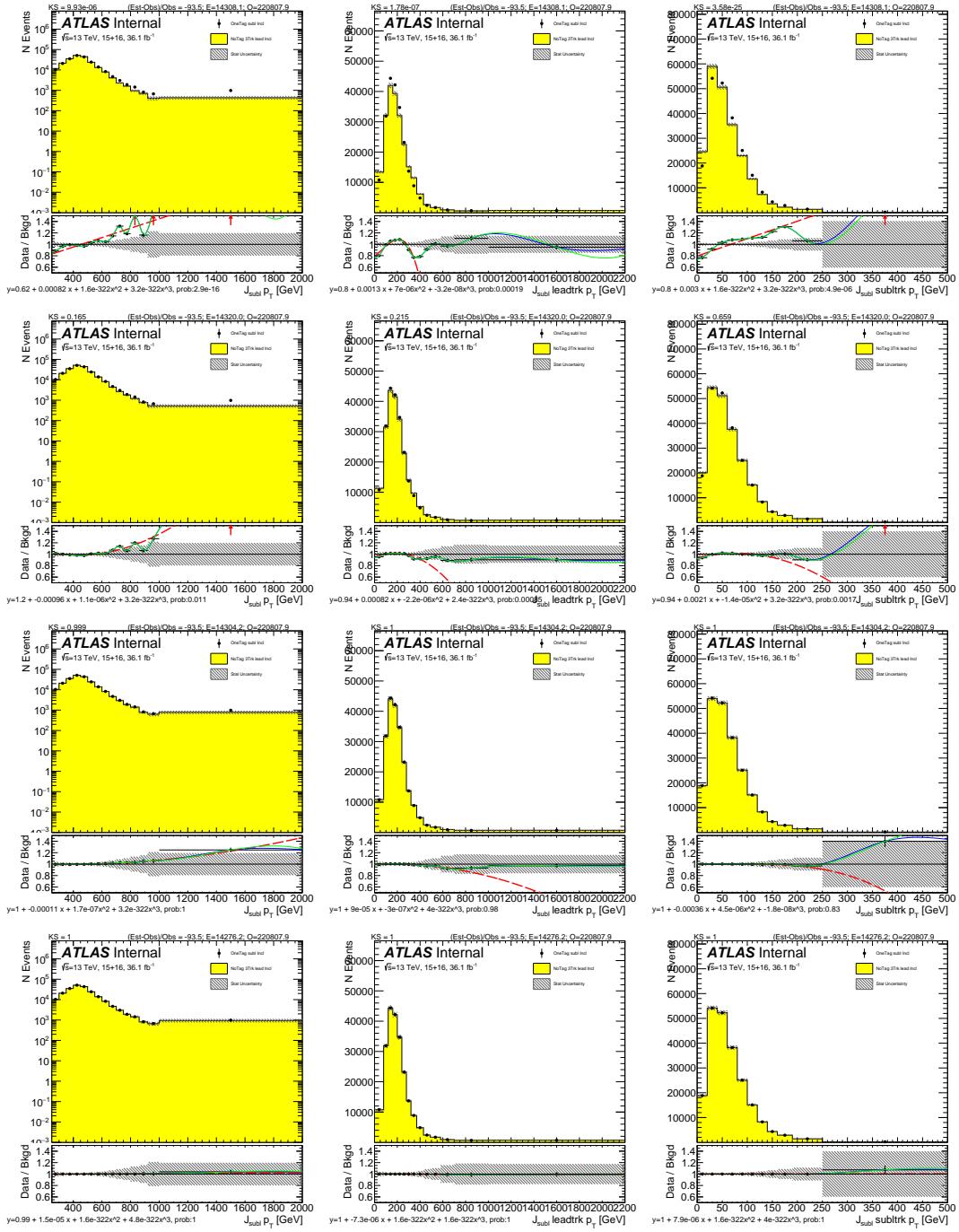


Figure 7.7: For $3b$ background estimate: the fits to the ratio of the data in the $2b$ category, of the subleading Higgs candidate $2b$ -tagged events's subleading Higgs candidate distributions(black point), over the leading Higgs candidate $1b$ -tagged events's subleading Higgs candidate distributions(yellow). Distributions and fits to the estimated QCD background for large- R jet p_T (left), the large- R jet's leading trackjet p_T (middle), and large- R jet's subleading trackjet p_T (right) are shown. Figure are shown before reweighting (top row), after the first iteration(second row), after the fourth iteration(third row), and after the last iteration (bottom row). The green line is the spline extrapolation; and the red line is a polynomial fit.

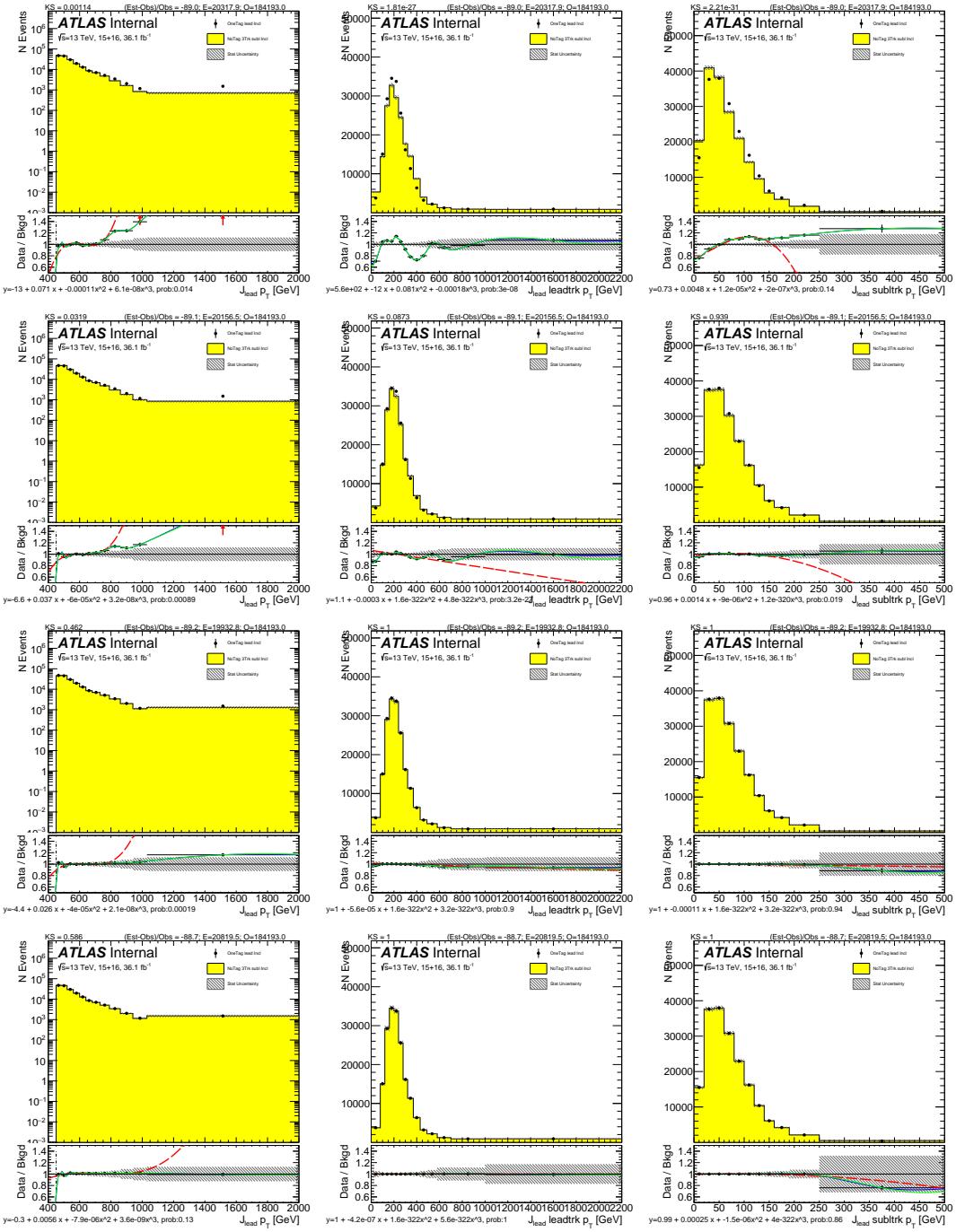


Figure 7.8: For $3b$ background estimate: the fits to the ratio of the data in the $2b$ category, of the leading Higgs candidate $2b$ -tagged events's leading Higgs candidate distributions(black point), over the subleading Higgs candidate $1b$ -tagged events's leading Higgs candidate distributions(yellow). Distributions and fits to the estimated QCD background for large- R jet p_T (left), the large- R jet's leading trackjet p_T (middle), and large- R jet's subleading trackjet p_T (right) are shown. Figure are shown before reweighting (top row), after the first iteration(second row), after the fourth iteration(third row), and after the last iteration (bottom row). The green line is the spline extrapolation; and the red line is a polynomial fit.

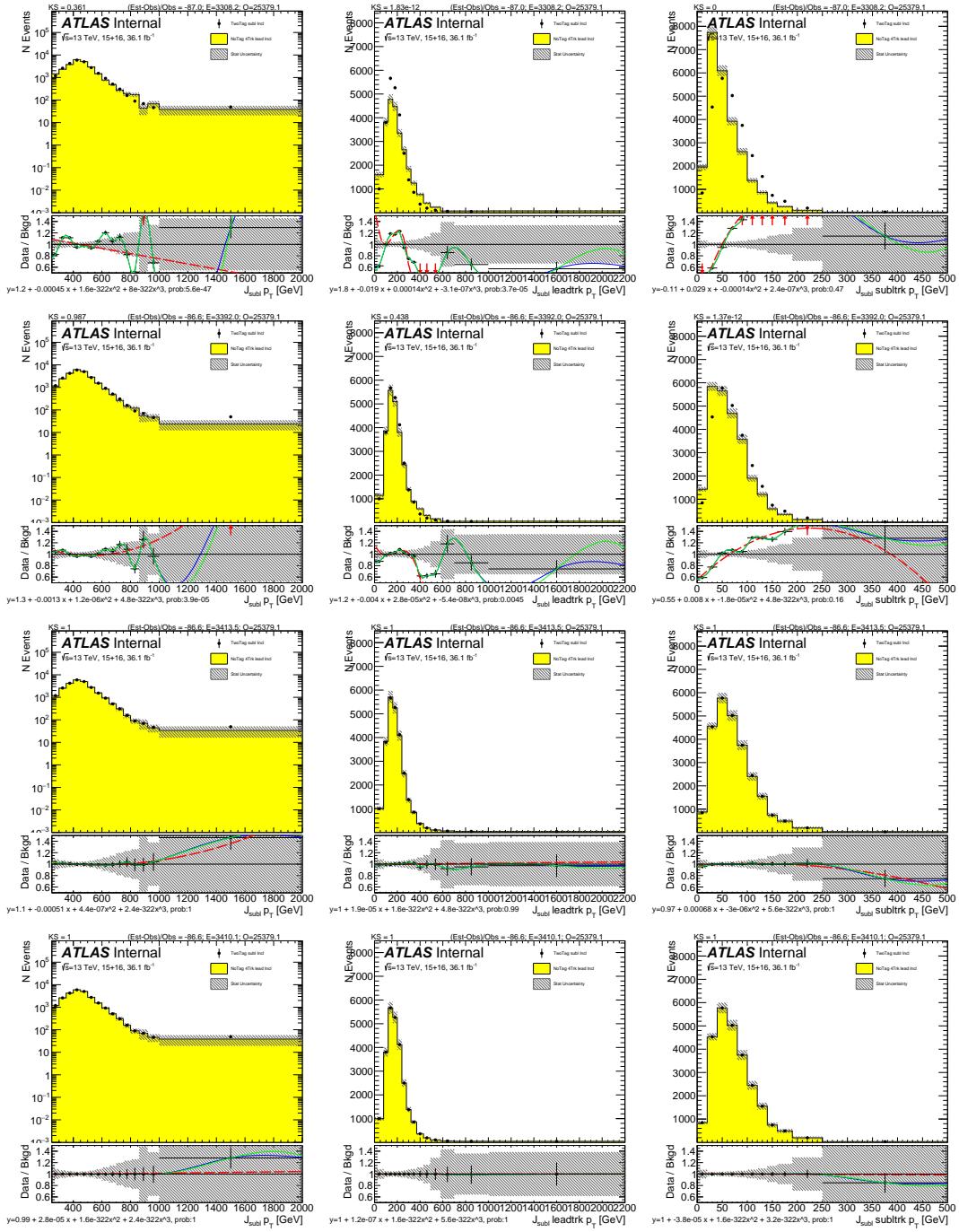


Figure 7.9: For $4b$ background estimate: the fits to the ratio of the data in the $2b$ category, of the subleading Higgs candidate $2b$ -tagged events's subleading Higgs candidate distributions(black point), over the leading Higgs candidate $2b$ -tagged events's subleading Higgs candidate distributions(yellow). Distributions and fits to the estimated QCD background for large- R jet p_T (left), the large- R jet's leading trackjet p_T (middle), and large- R jet's subleading trackjet p_T (right) are shown. Figure are shown before reweighting (top row), after the first iteration(second row), after the fourth iteration(third row), and after the last iteration (bottom row). The green line is the spline extrapolation; and the red line is a polynomial fit.

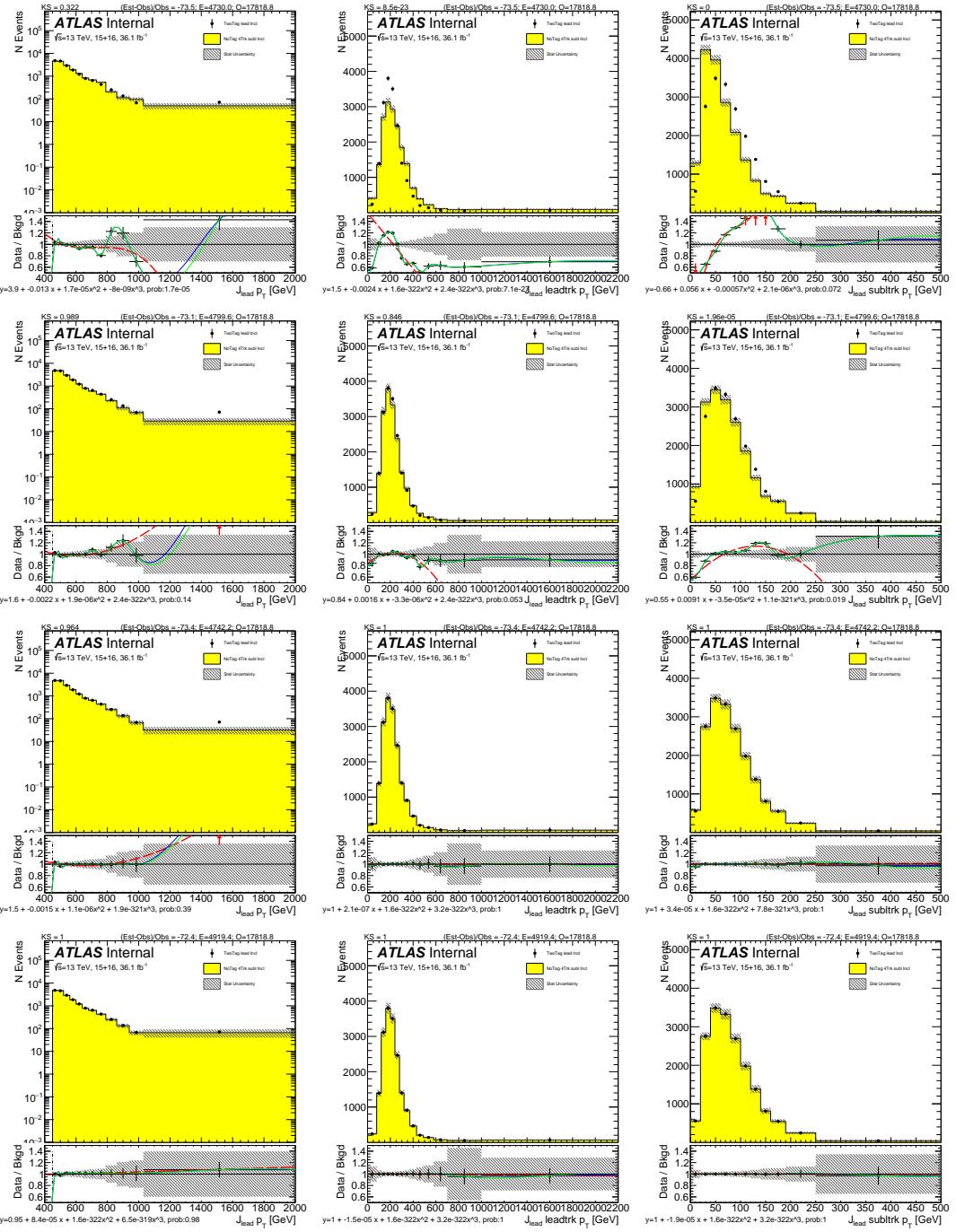


Figure 7.10: For $4b$ background estimate: the fits to the ratio of the data in the $2b$ category, of the leading Higgs candidate $2b$ -tagged events's leading Higgs candidate distributions(black point), over the subleading Higgs candidate $2b$ -tagged events's leading Higgs candidate distributions(yellow). Distributions and fits to the estimated QCD background for large- R jet p_T (left), the large- R jet's leading trackjet p_T (middle), and large- R jet's subleading trackjet p_T (right) are shown. Figure are shown before reweighting (top row), after the first iteration(second row), after the fourth iteration(third row), and after the last iteration (bottom row). The green line is the spline extrapolation; and the red line is a polynomial fit.

REWEIGHTING RESULTS COMPARISON IN SIDEBAND AND CONTROL REGION

A comparison of the Sideband shapes before and after reweighting for $2bs$, $3b$ and $4b$ can be seen in Figures 7.11, 7.12, and 7.13. Also, a comparison of the Control Region shapes before and after reweighting for $2bs$, $3b$ and $4b$ can be seen in Figures 7.14, 7.15, and 7.16. In almost all cases, both the reweighted/non-reweighted prediction agrees fairly well with the data, and the reweighted plots' KS score improved from non-reweighted distributions.

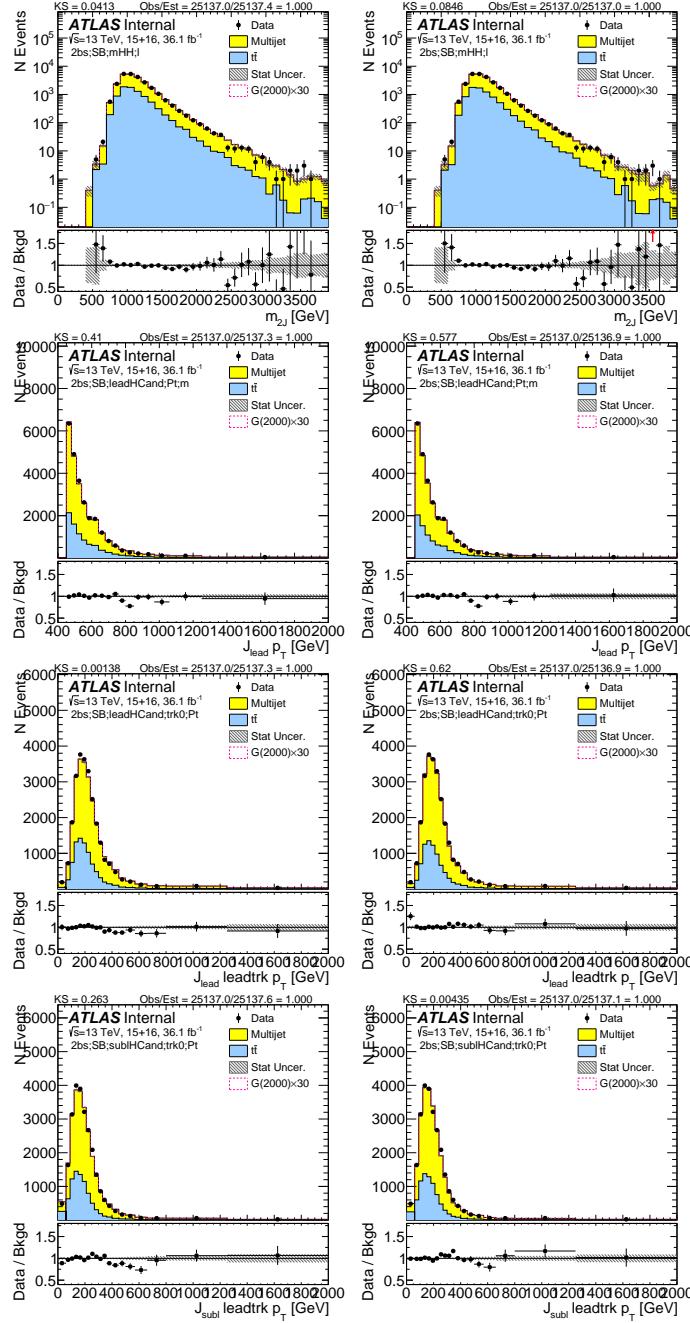


Figure 7.11: Reweighted $2bs$ Sideband region predictions comparison. Top row is the dijet Mass, second row is leading large- R jet p_T , third row is the leading large- R jet's leading trackjet p_T and the last row subleading large- R jet's leading trackjet p_T . On the left are the distributions before reweighting, and on the right are the distributions after reweighting.

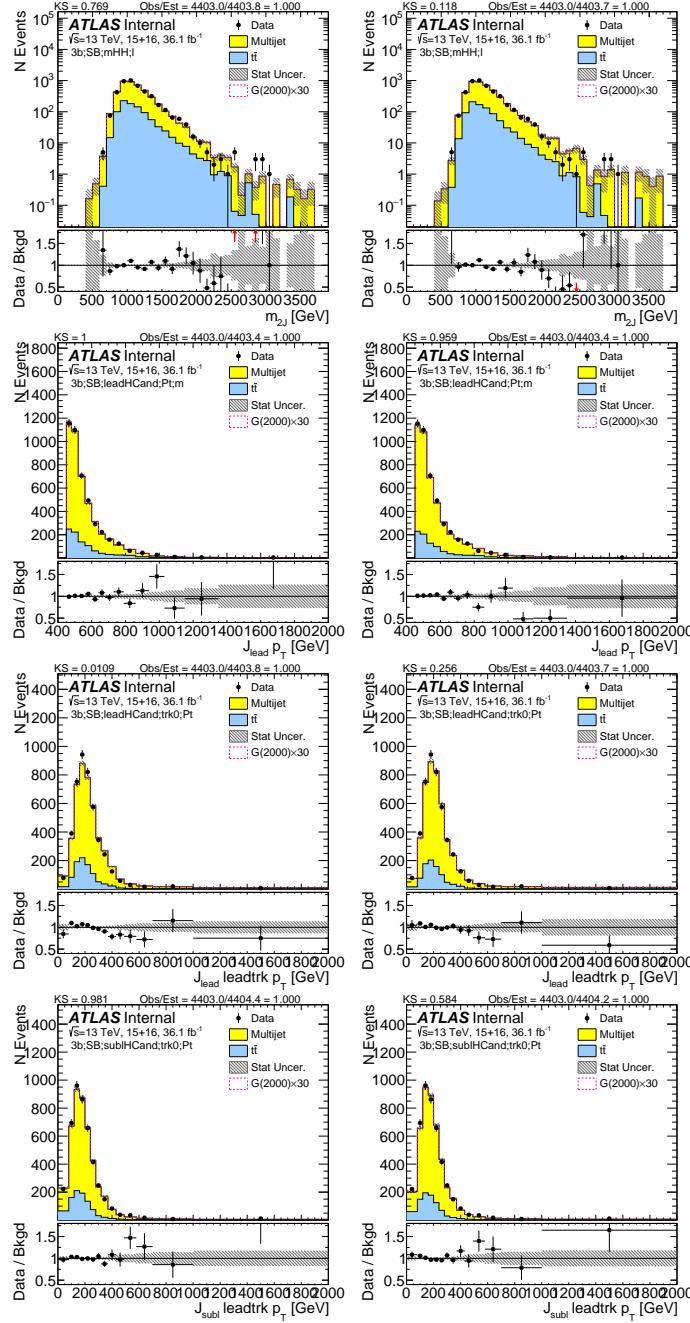


Figure 7.12: Reweighted 3 b Sideband region predictions comparison. Top row is the dijet Mass, second row is leading large- R jet p_T , third row is the leading large- R jet's leading trackjet p_T and the last row subleading large- R jet's leading trackjet p_T . On the left are the distributions before reweighting, and on the right are the distributions after reweighting.

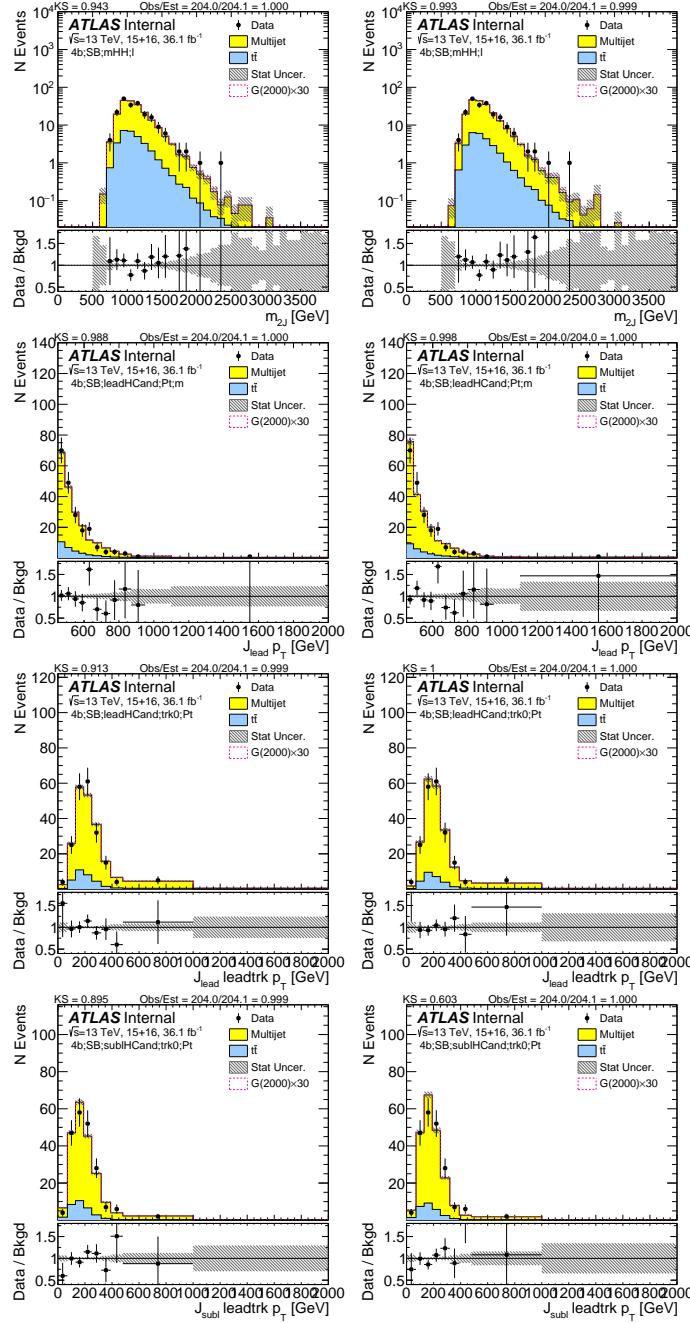


Figure 7.13: Reweighted $4b$ Sideband region predictions comparison. Top row is the dijet Mass, second row is leading large- R jet p_T , third row is the leading large- R jet's leading trackjet p_T and the last row subleading large- R jet's leading trackjet p_T . On the left are the distributions before reweighting, and on the right are the distributions after reweighting.

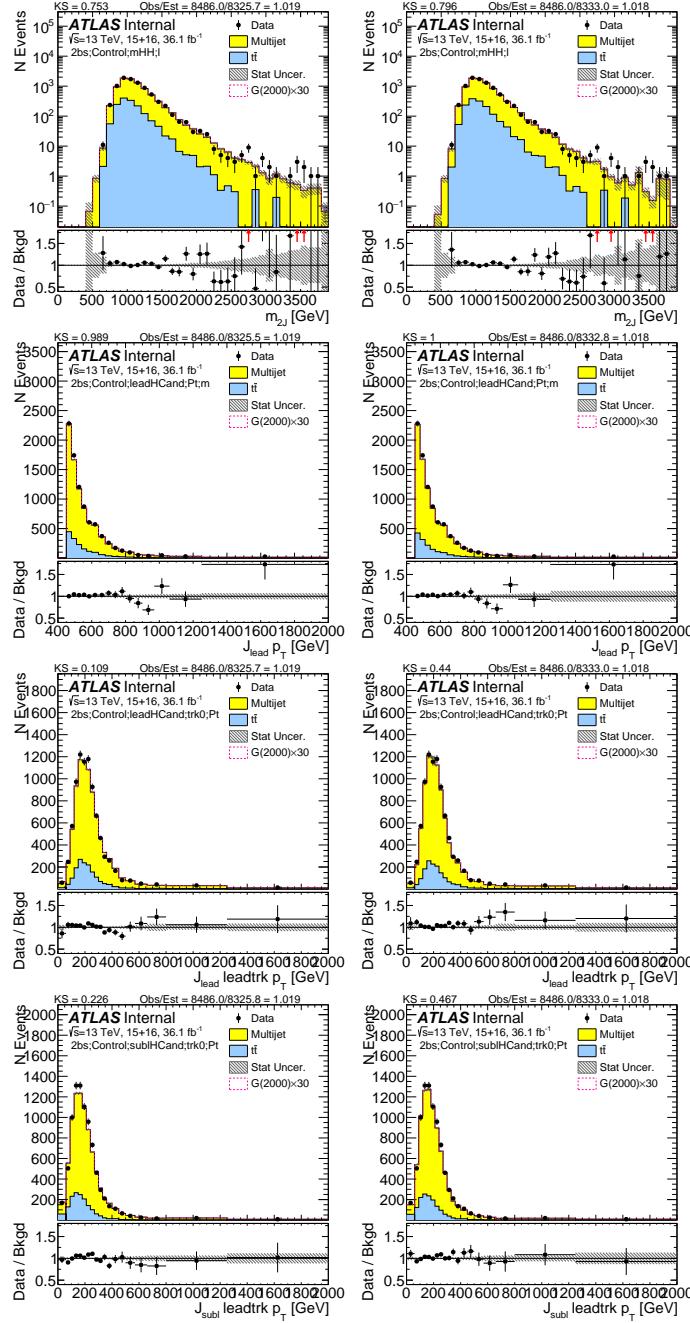


Figure 7.14: Reweighted 2bs Control region predictions comparison. Top row is the dijet Mass, second row is leading large- R jet p_T , third row is the leading large- R jet's leading trackjet p_T and the last row subleading large- R jet's leading trackjet p_T . On the left are the distributions before reweighting, and on the right are the distributions after reweighting.

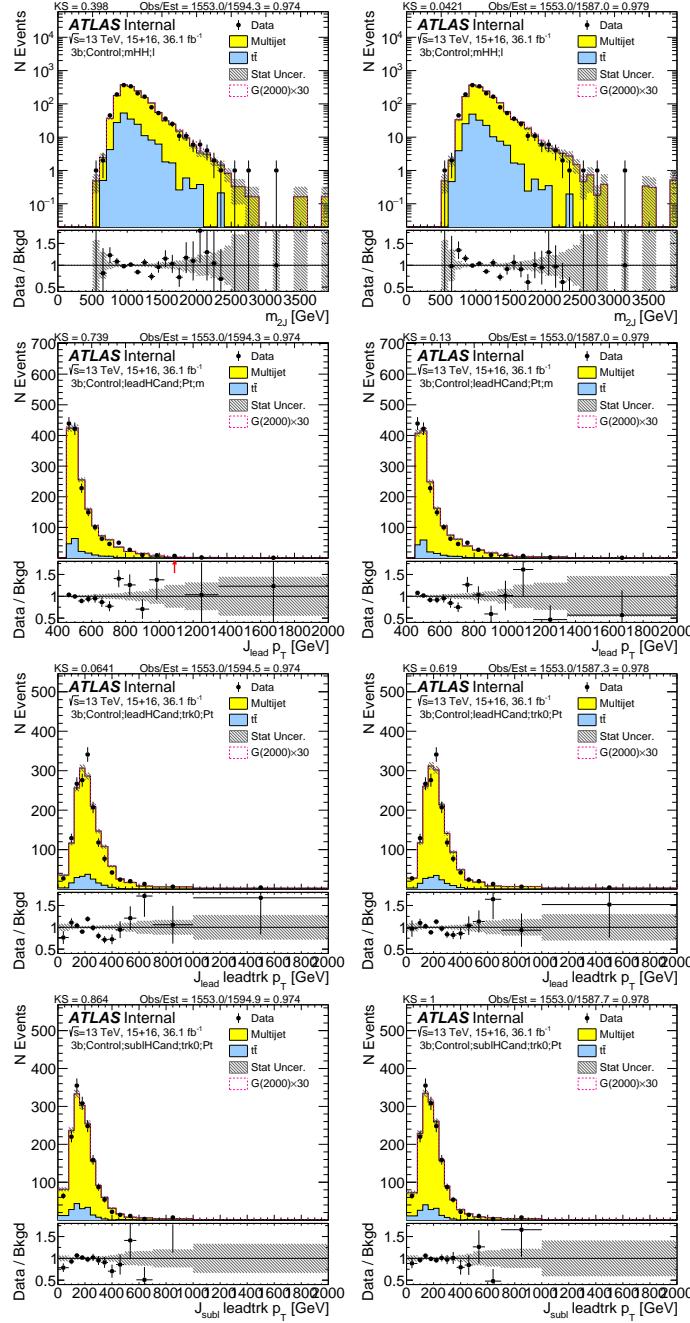


Figure 7.15: Reweighted 3 b Control region predictions comparison. Top row is the dijet Mass, second row is leading large- R jet p_T , third row is the leading large- R jet's leading trackjet p_T and the last row subleading large- R jet's leading trackjet p_T . On the left are the distributions before reweighting, and on the right are the distributions after reweighting.

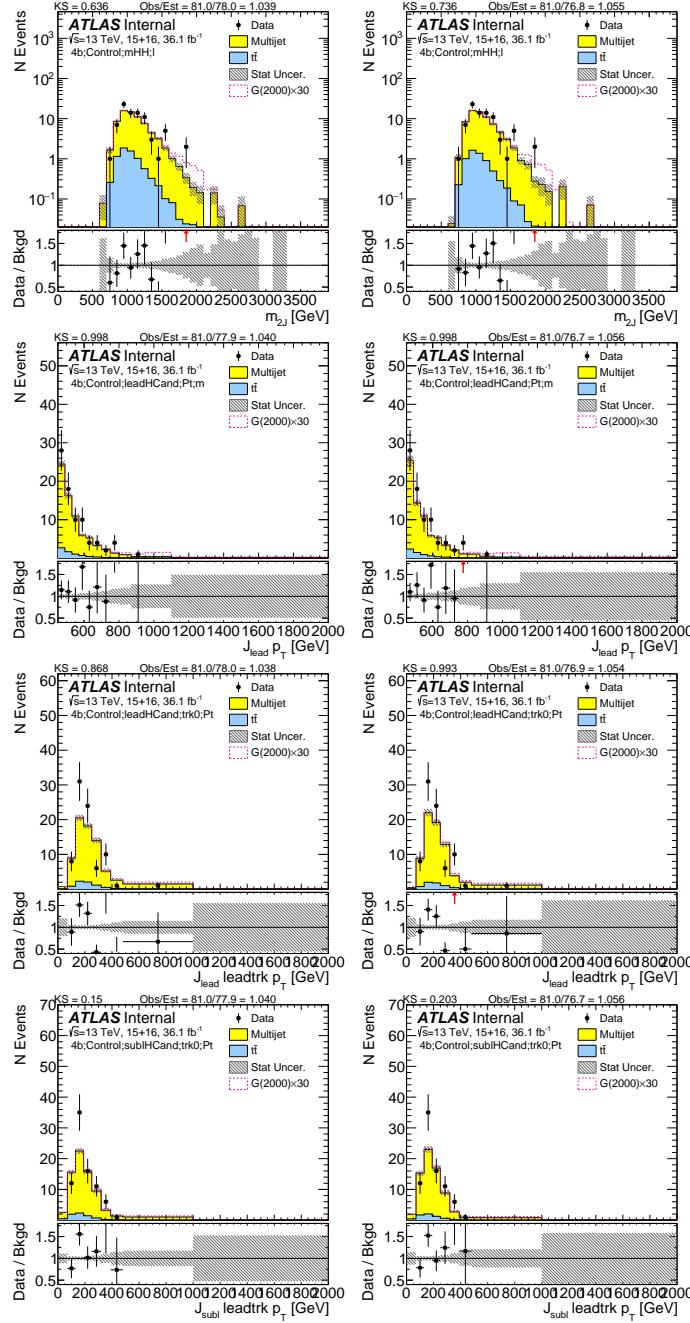


Figure 7.16: Reweighted 4*b* Control region predictions comparison. Top row is the dijet Mass, second row is leading large- R jet p_T , third row is the leading large- R jet's leading trackjet p_T and the last row subleading large- R jet's leading trackjet p_T . On the left are the distributions before reweighting, and on the right are the distributions after reweighting.

7.1.7 PREDICTIONS IN THE SIDEBAND REGION (SB)

This section shows comparisons of data with the prediction of QCD multi-jets and $t\bar{t}$ in the sideband region (SB), which is identical to the signal region (SR) except the large- R jets are required to have masses far from the Higgs mass. The definition of the sideband and control regions is discussed in Section 7.1.2. The predicted and observed event yields are summarized in Tables ?? and ??.

Figures 7.17, 7.18, 7.19, and 7.20 show predictions of various kinematics of the large- R jets and their associated track jets in the $4b$ selection. The predicted normalization agrees perfectly by construction, but the shapes are a feature of the prediction. The quality of the prediction is generally good, and no clear systematic biases are observed.

Figures 7.21, 7.22, 7.23, and 7.24 show predictions of various kinematics of the large- R jets and their associated track jets in the $3b$ selection. The predicted normalization agrees perfectly by construction, but the shapes are a feature of the prediction. The quality of the prediction is generally good, and no clear systematic biases are observed.

Figures 7.25, 7.26, 7.27, and 7.28 show predictions of various kinematics of the large- R jets and their associated track jets in the $2bs$ selection. The predicted normalization agrees perfectly by construction, but the shapes are a feature of the prediction. The quality of the prediction is generally good, and no clear systematic biases are observed.

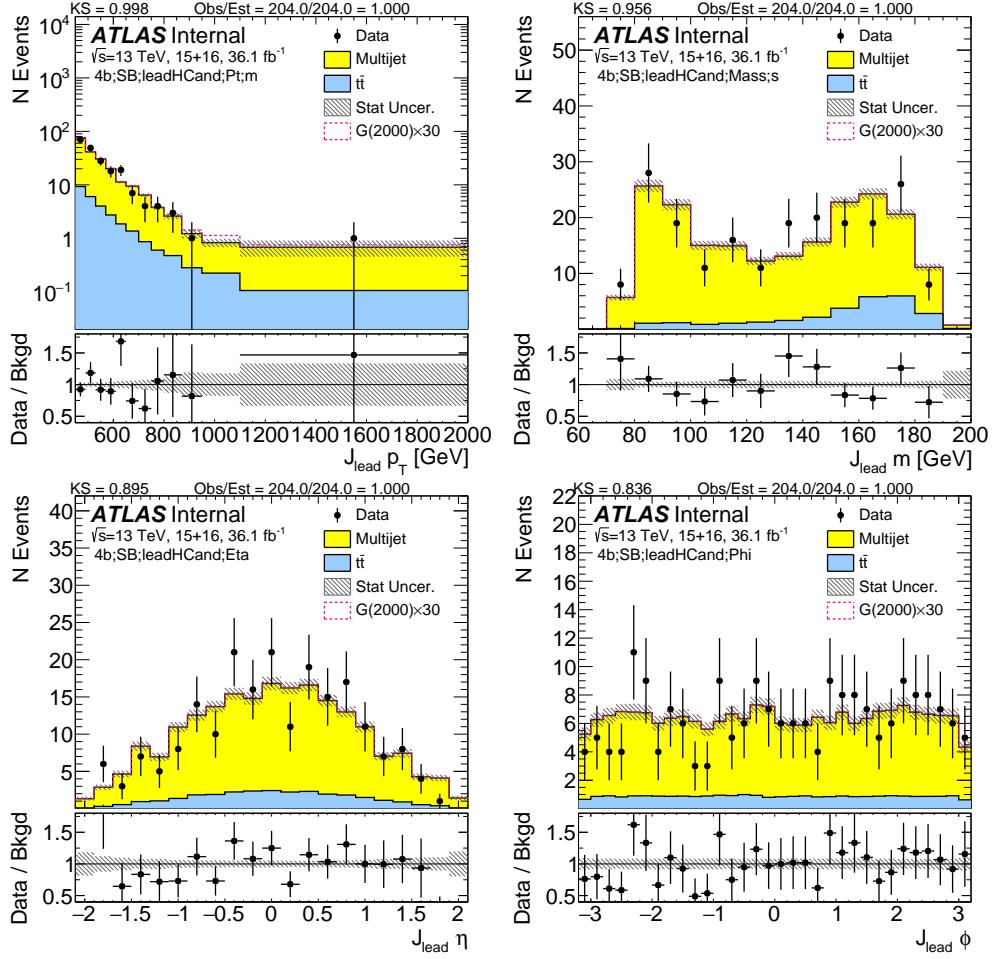


Figure 7.17: Kinematics of the lead large- R jet in data and prediction in the sideband region after requiring 4 b -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.

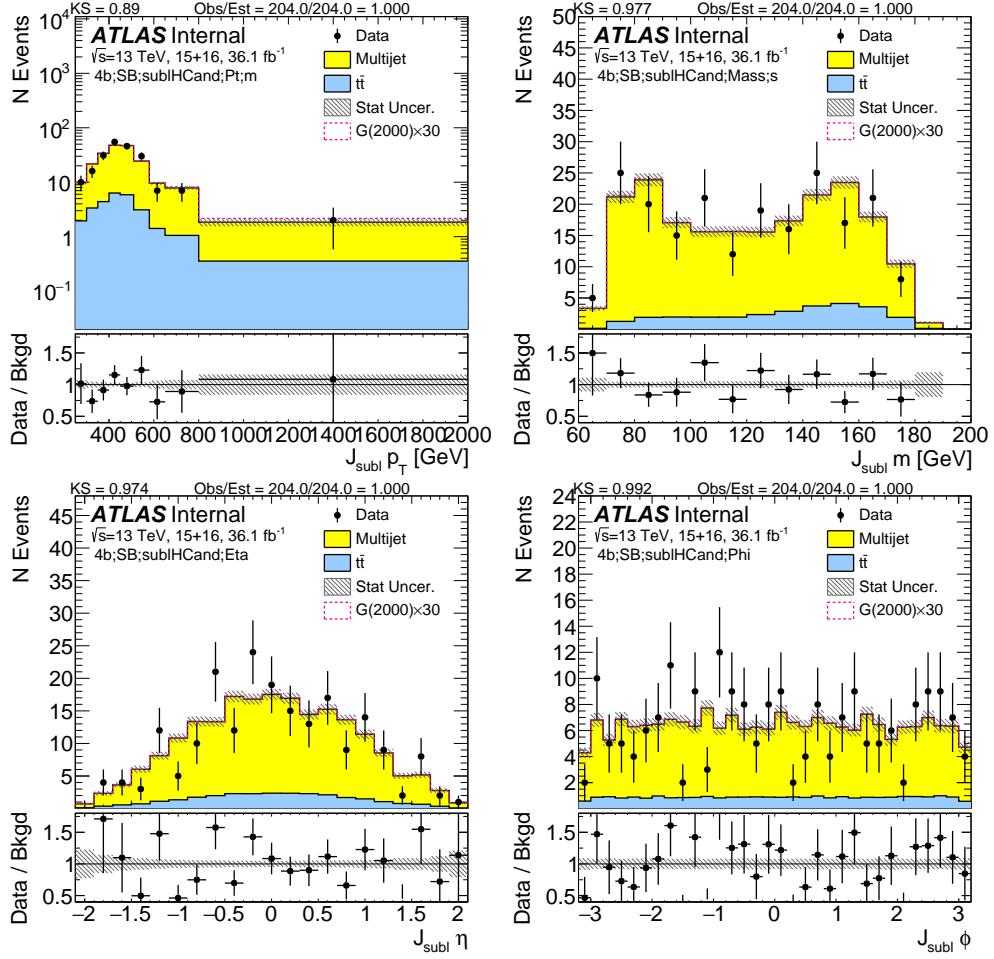


Figure 7.18: Kinematics of the sub-lead large- R jet in data and prediction in the sideband region after requiring 4 b -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.

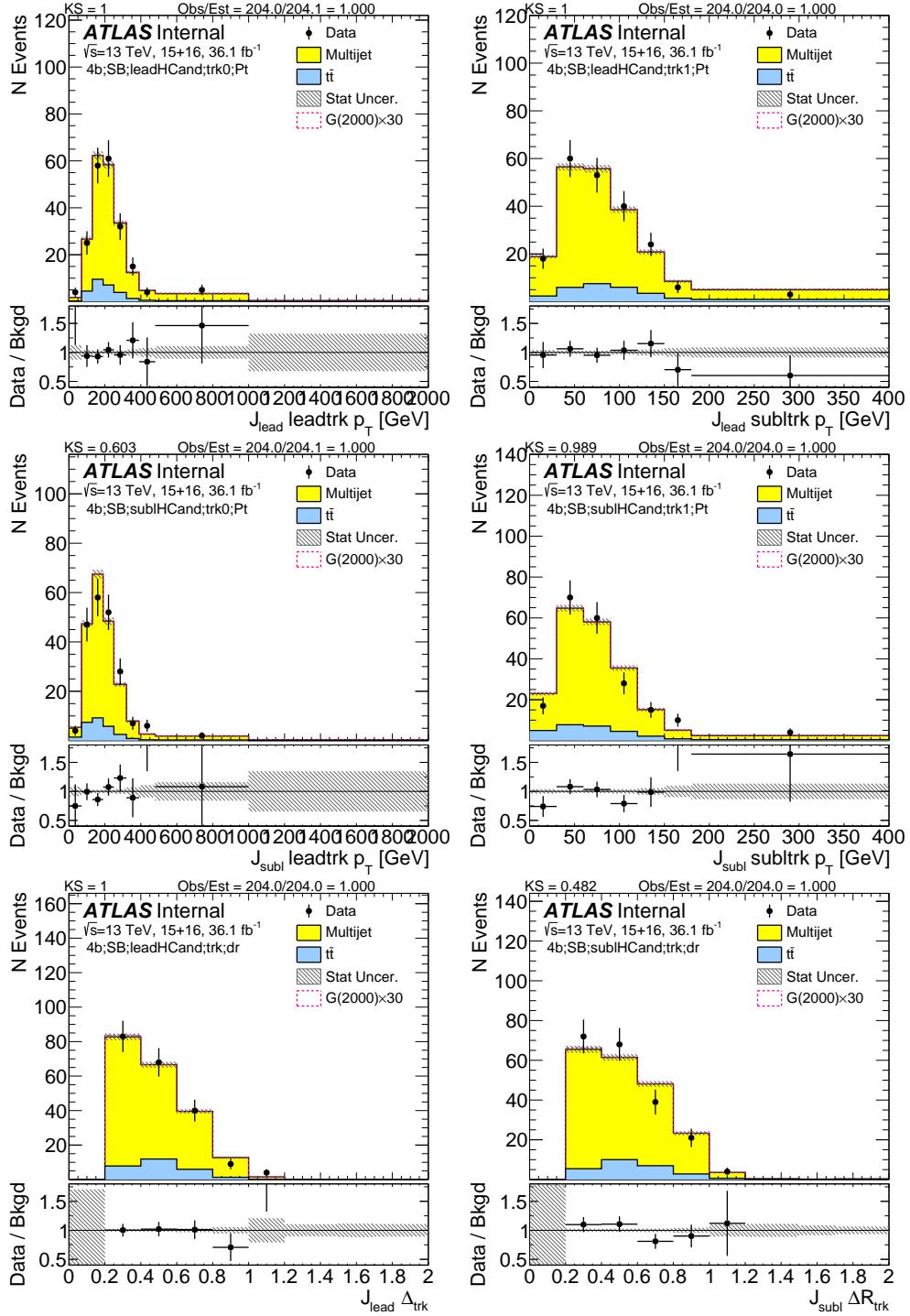


Figure 7.19: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the sideband region after requiring 4 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. The normalization agrees by construction, and the shapes are a feature of the prediction.

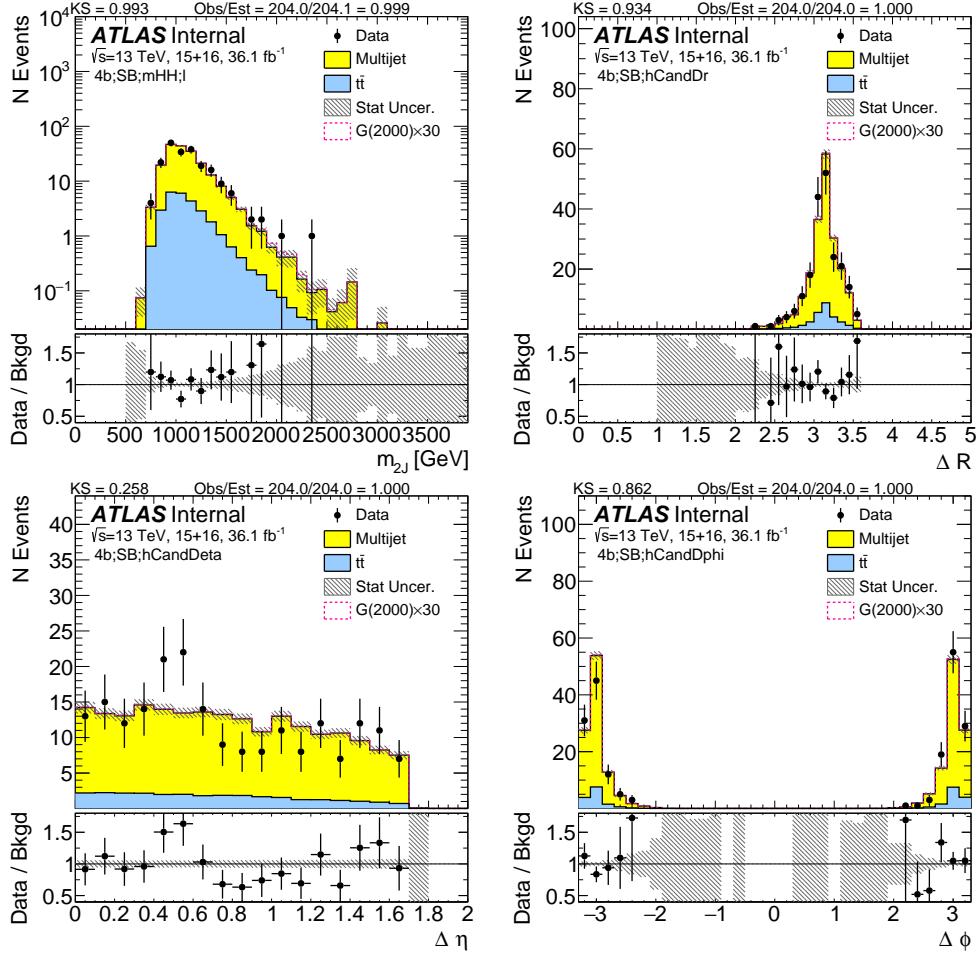


Figure 7.20: Kinematics of the large- R jet system in data and prediction in the sideband region after requiring 4 b -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.

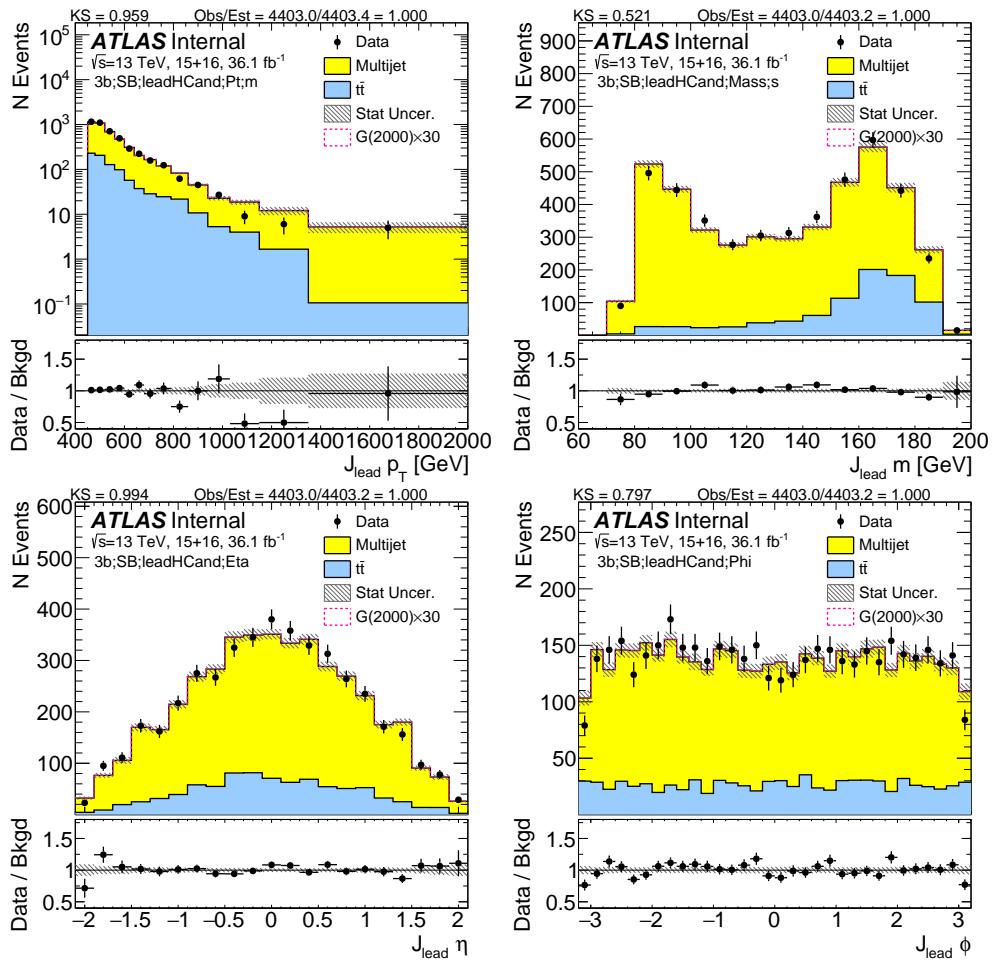


Figure 7.21: Kinematics of the lead large- R jet in data and prediction in the sideband region after requiring 3 b -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.

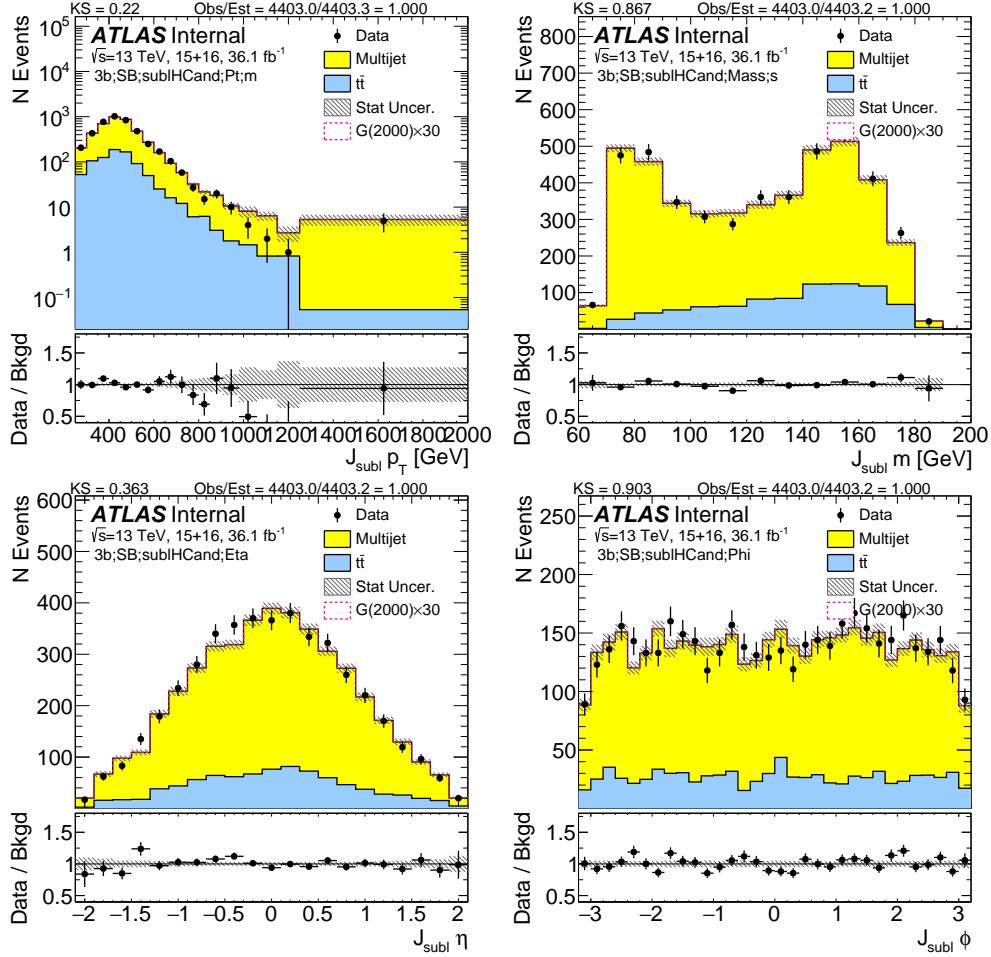


Figure 7.22: Kinematics of the sub-lead large- R jet in data and prediction in the sideband region after requiring 3 b -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.

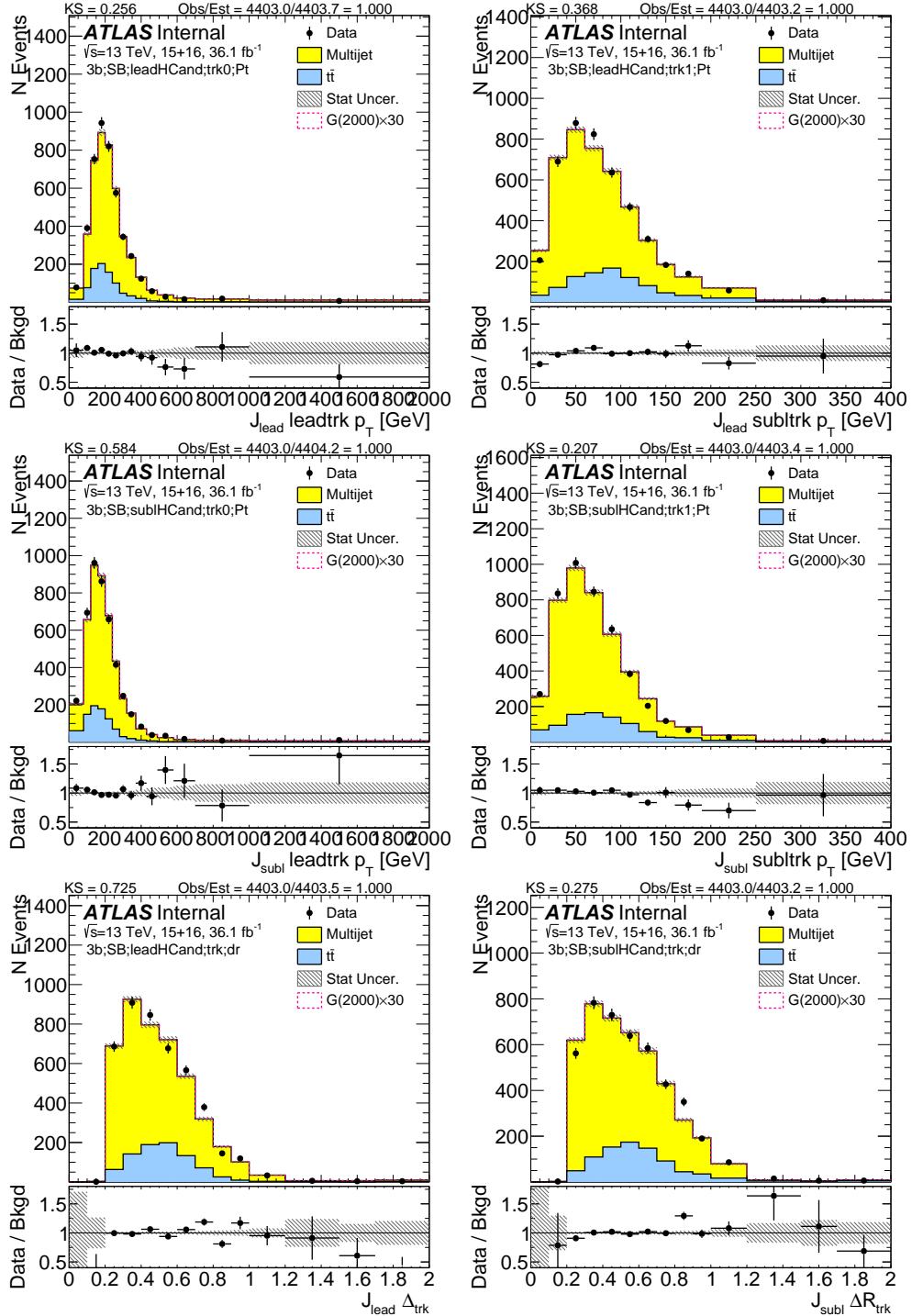


Figure 7.23: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the sideband region after requiring 3 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. The normalization agrees by construction, and the shapes are a feature of the prediction.

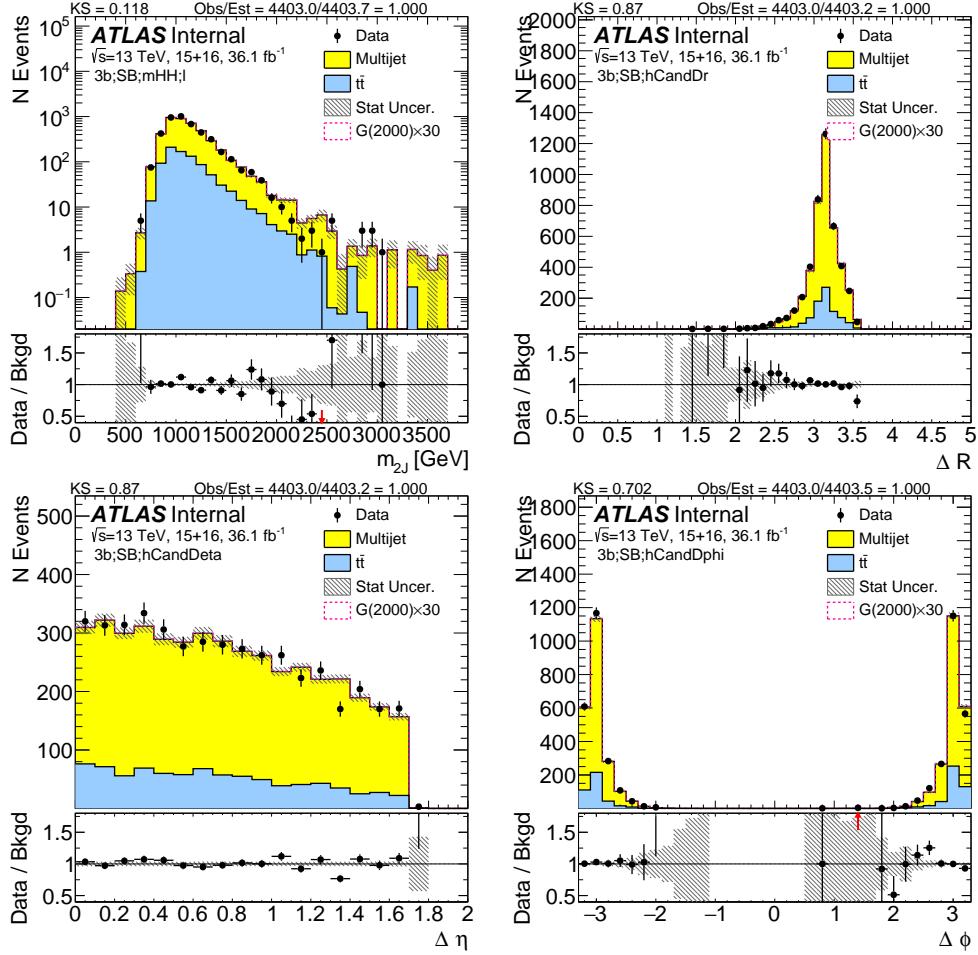


Figure 7.24: Kinematics of the large- R jet system in data and prediction in the sideband region after requiring 3 b -tags. The normalization agrees by construction, and the shapes are a feature of the prediction.

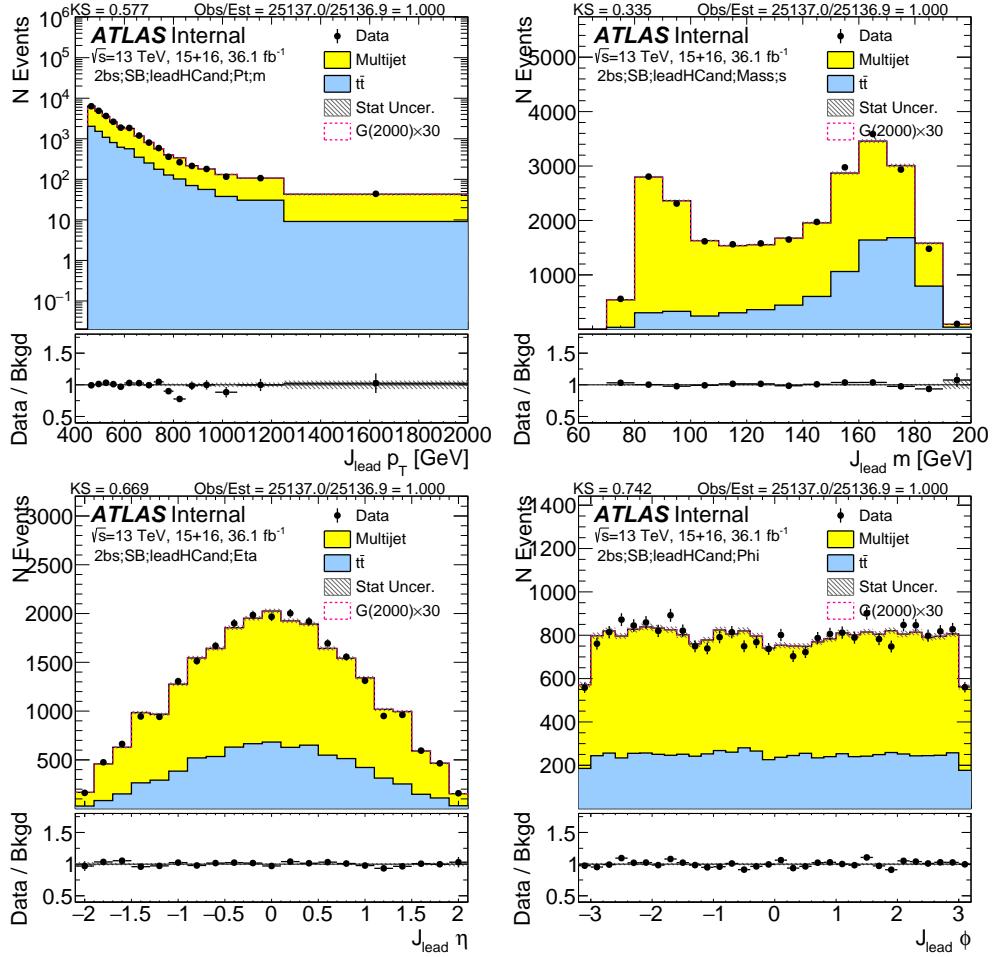


Figure 7.25: Kinematics of the lead large- R jet in data and prediction in the sideband region after requiring 2 b -tags split. The normalization agrees by construction, and the shapes are a feature of the prediction.

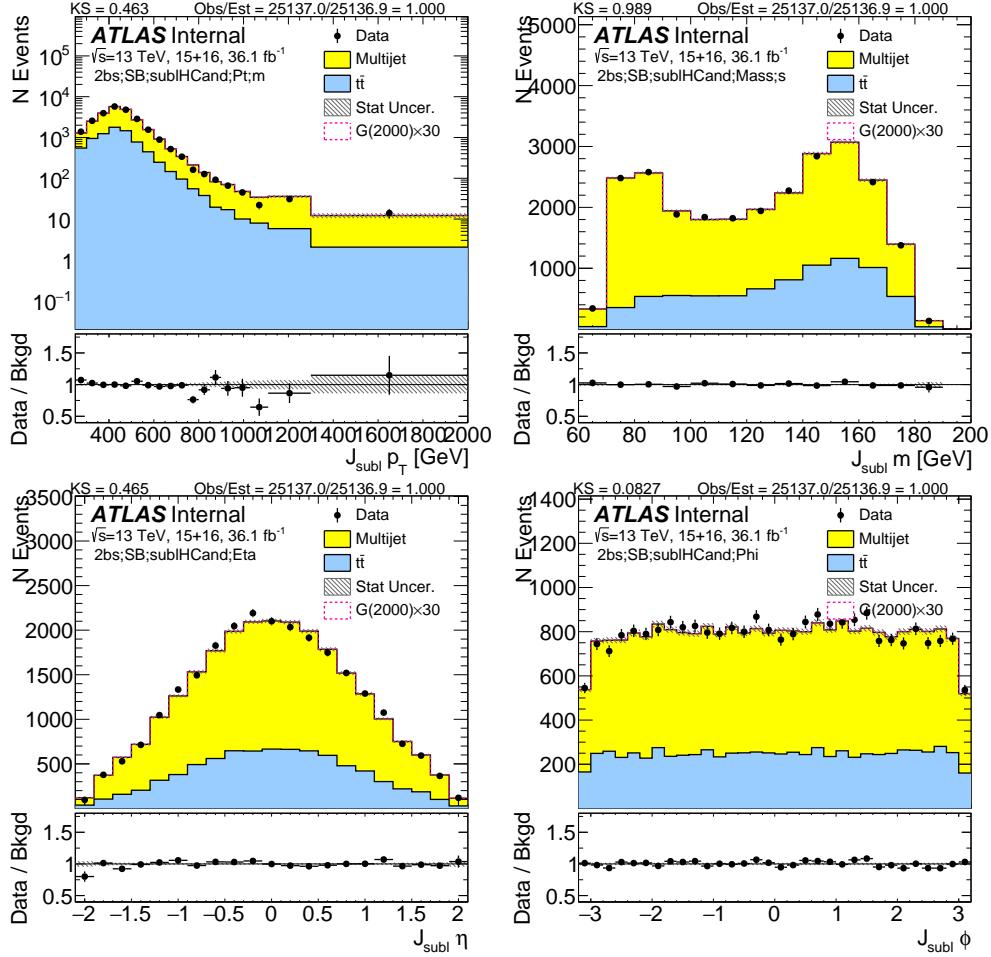


Figure 7.26: Kinematics of the sub-lead large- R jet in data and prediction in the sideband region after requiring 2 b -tags split. The normalization agrees by construction, and the shapes are a feature of the prediction.

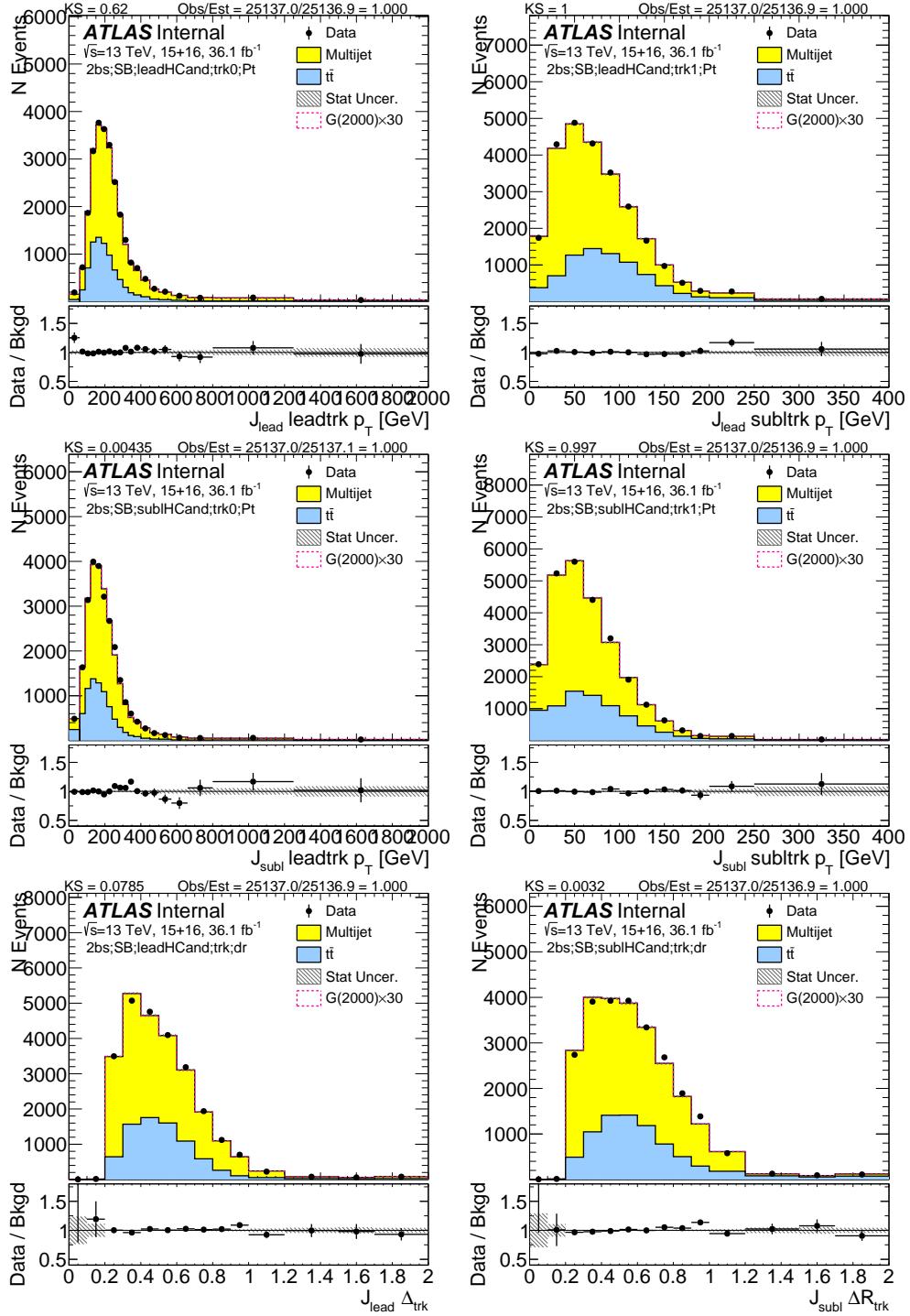


Figure 7.27: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the sideband region after requiring 2 b -tags split. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. The normalization agrees by construction, and the shapes are a feature of the prediction.

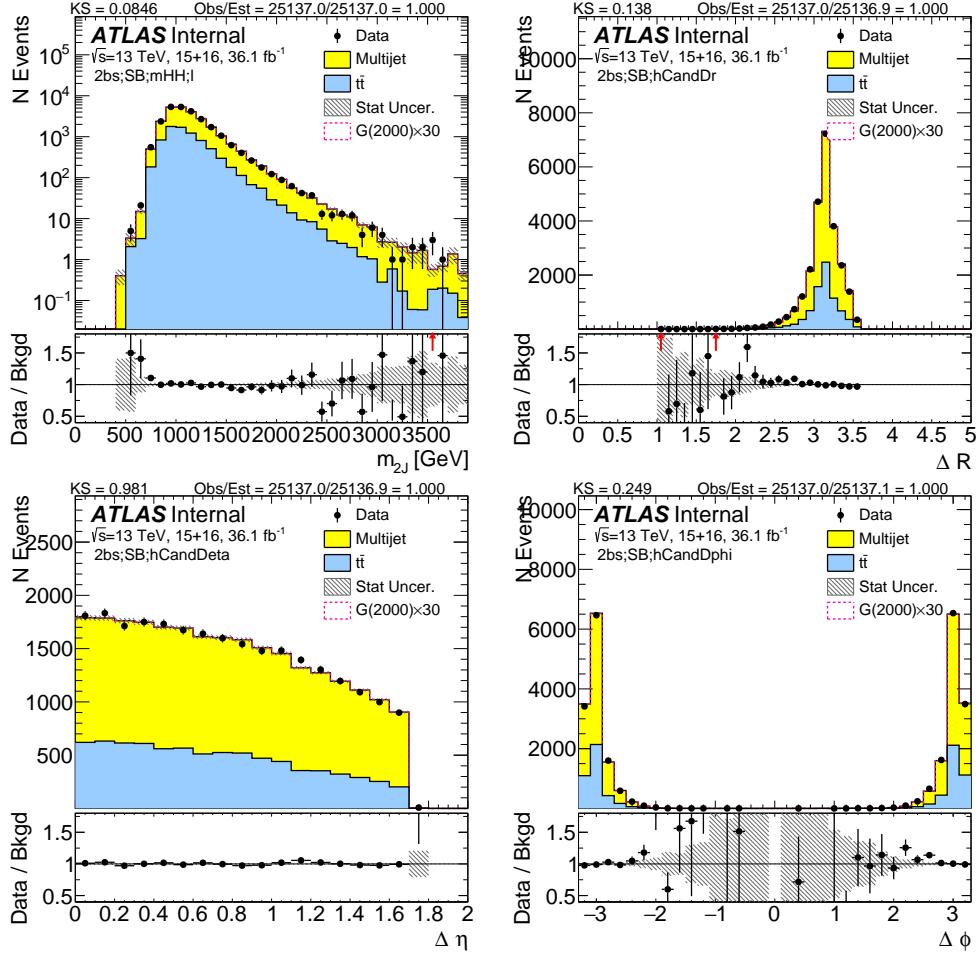


Figure 7.28: Kinematics of the large- R jet system in data and prediction in the sideband region after requiring 2 b -tags split. The normalization agrees by construction, and the shapes are a feature of the prediction.

7.1.8 PREDICTIONS IN THE CONTROL REGION (CR)

This section shows comparisons of data with the prediction of QCD multi-jets and $t\bar{t}$ in the control region (CR), which is identical to the signal region (SR) except the large- R jets are required to have masses close but not too close to the Higgs mass. The definition can be seen in Section 7.1.2. The predicted and observed event yields are summarized in Tables ?? and ??.

Figures 7.29, 7.30, 7.31, and 7.32 show predictions of various kinematics of the large- R jets and their associated track jets in the $4b$ selection. The shapes and normalization are a feature of the prediction, where the normalization is derived in the SB. The quality of the prediction is generally good, and no clear systematic biases are observed.

Figures 7.33, 7.34, 7.35, and 7.36 show predictions of various kinematics of the large- R jets and their associated track jets in the $3b$ selection. The shapes and normalization are a feature of the prediction, where the normalization is derived in the SB. The quality of the prediction is generally good, and no clear systematic biases are observed.

Figures 7.37, 7.38, 7.39, and 7.40 show predictions of various kinematics of the large- R jets and their associated track jets in the $2bs$ selection. The shapes and normalization are a feature of the prediction, where the normalization is derived in the SB. The quality of the prediction is generally good, and no clear systematic biases are observed.

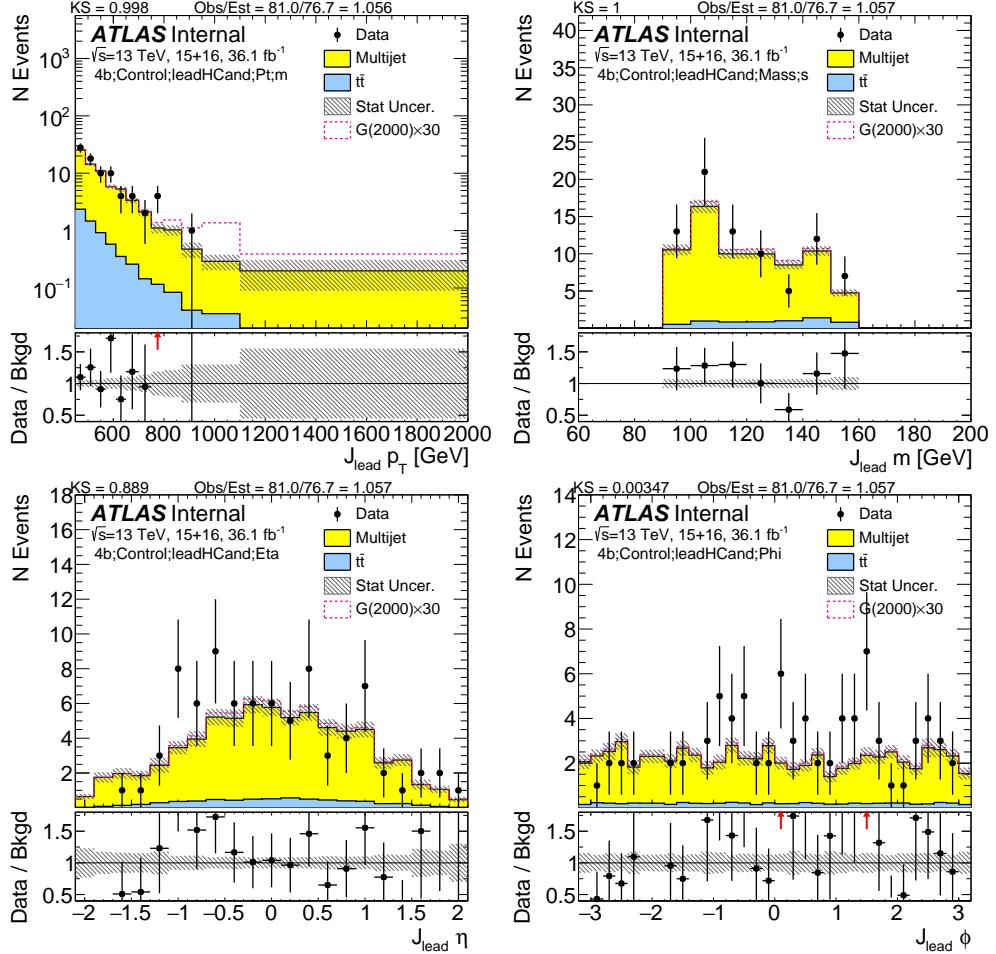


Figure 7.29: Kinematics of the lead large- R jet in data and prediction in the control region after requiring 4 b -tags.

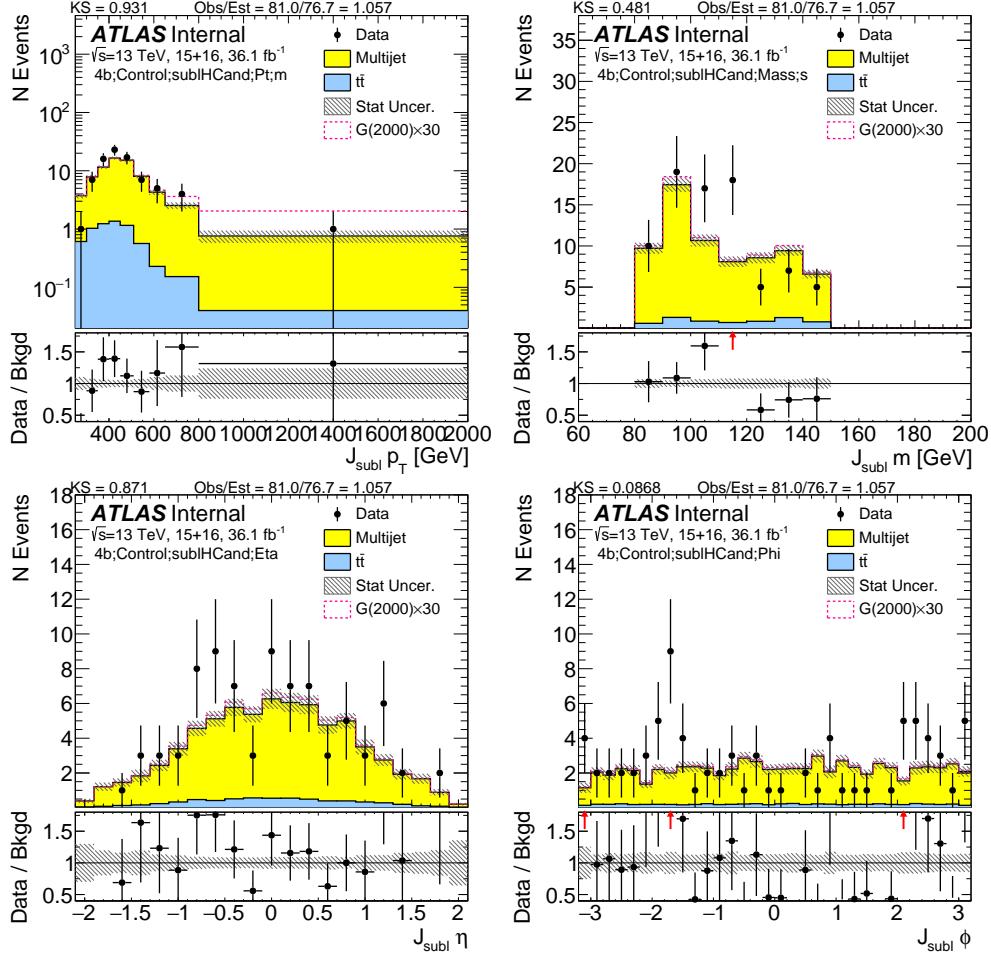


Figure 7.30: Kinematics of the sub-lead large- R jet in data and prediction in the control region after requiring 4 b -tags.

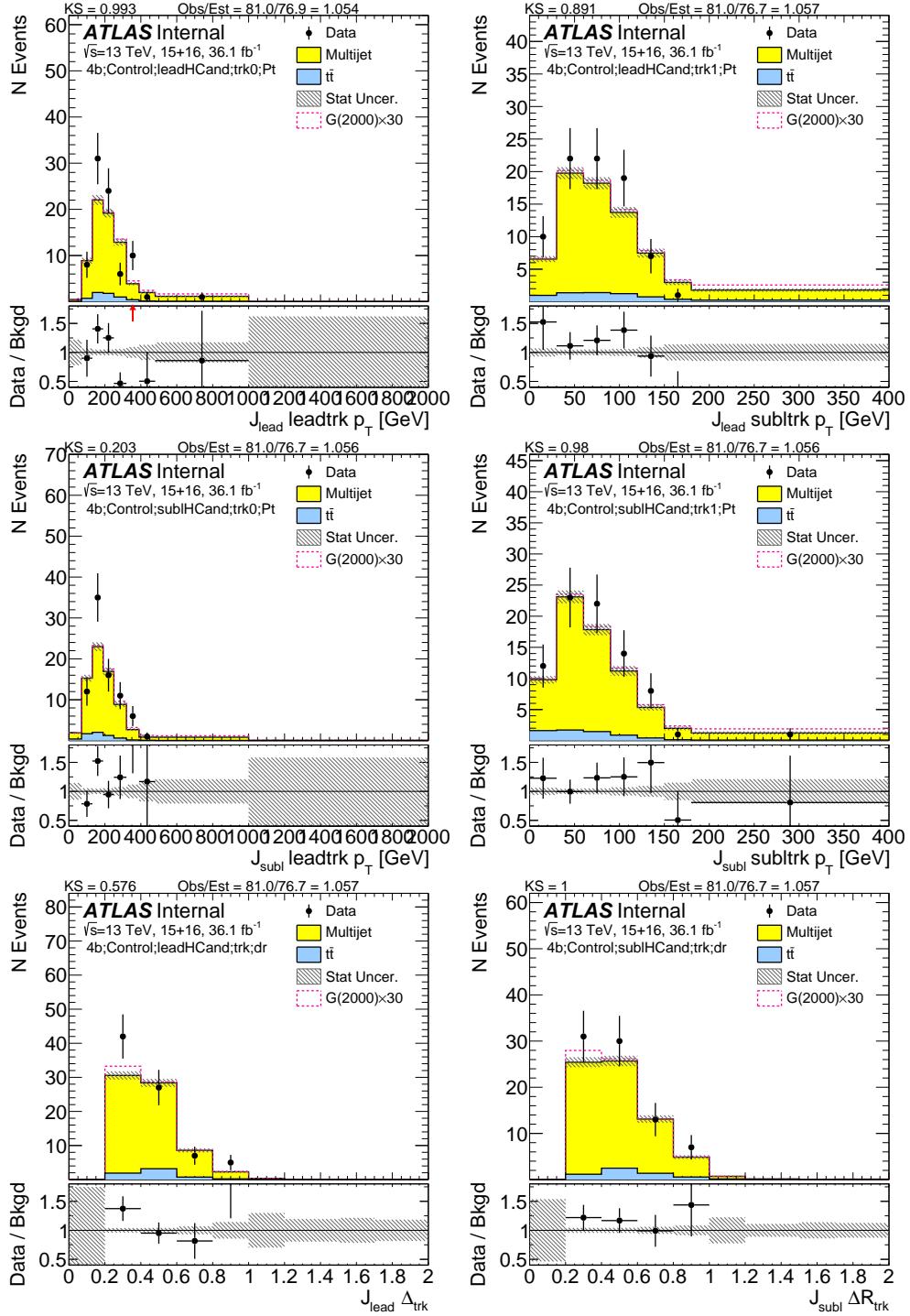


Figure 7.31: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the control region after requiring 4 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet.

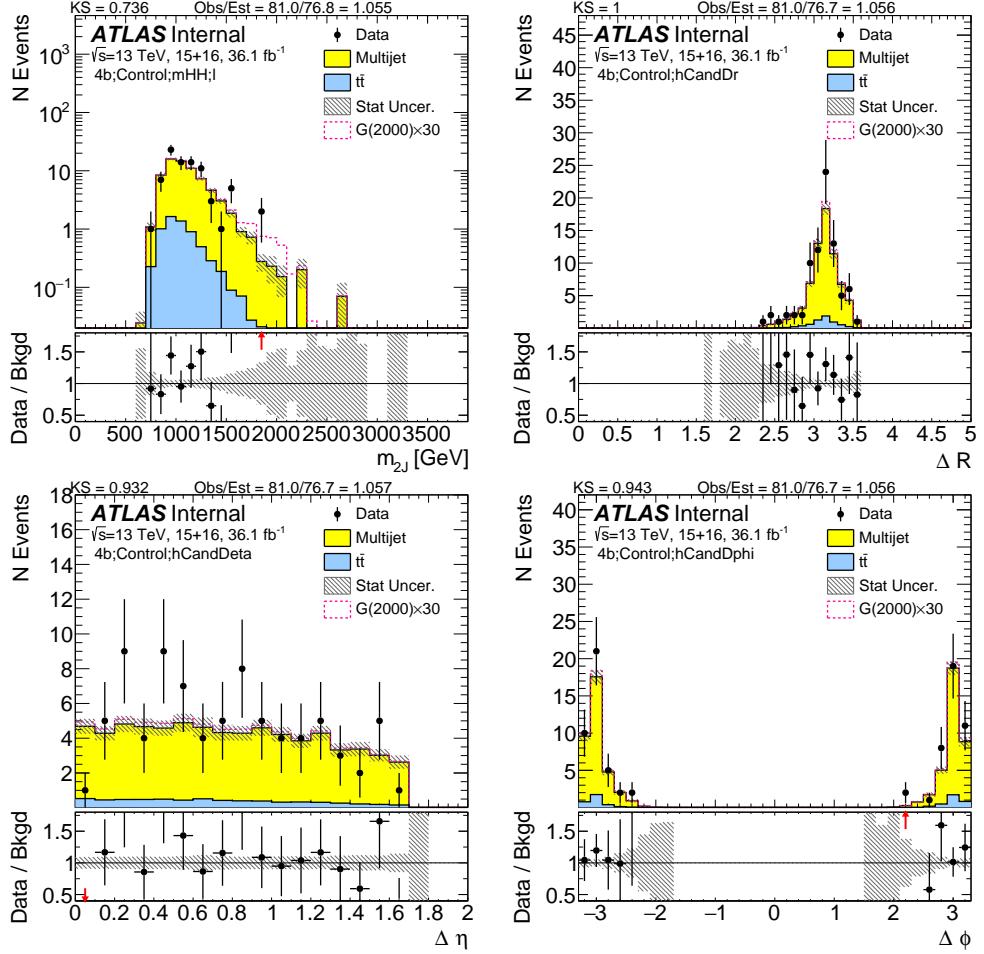


Figure 7.32: Kinematics of the large- R jet system in data and prediction in the control region after requiring 4 b -tags.

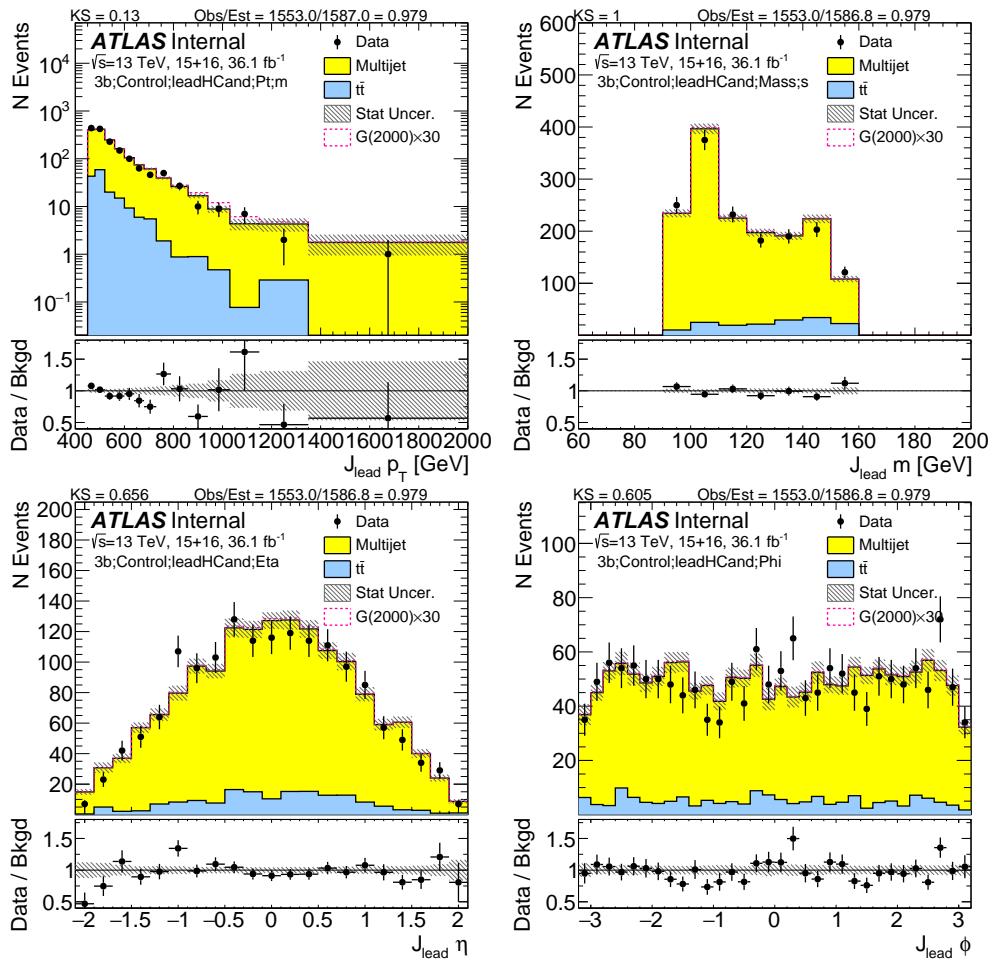


Figure 7.33: Kinematics of the lead large- R jet in data and prediction in the control region after requiring 3 b -tags.

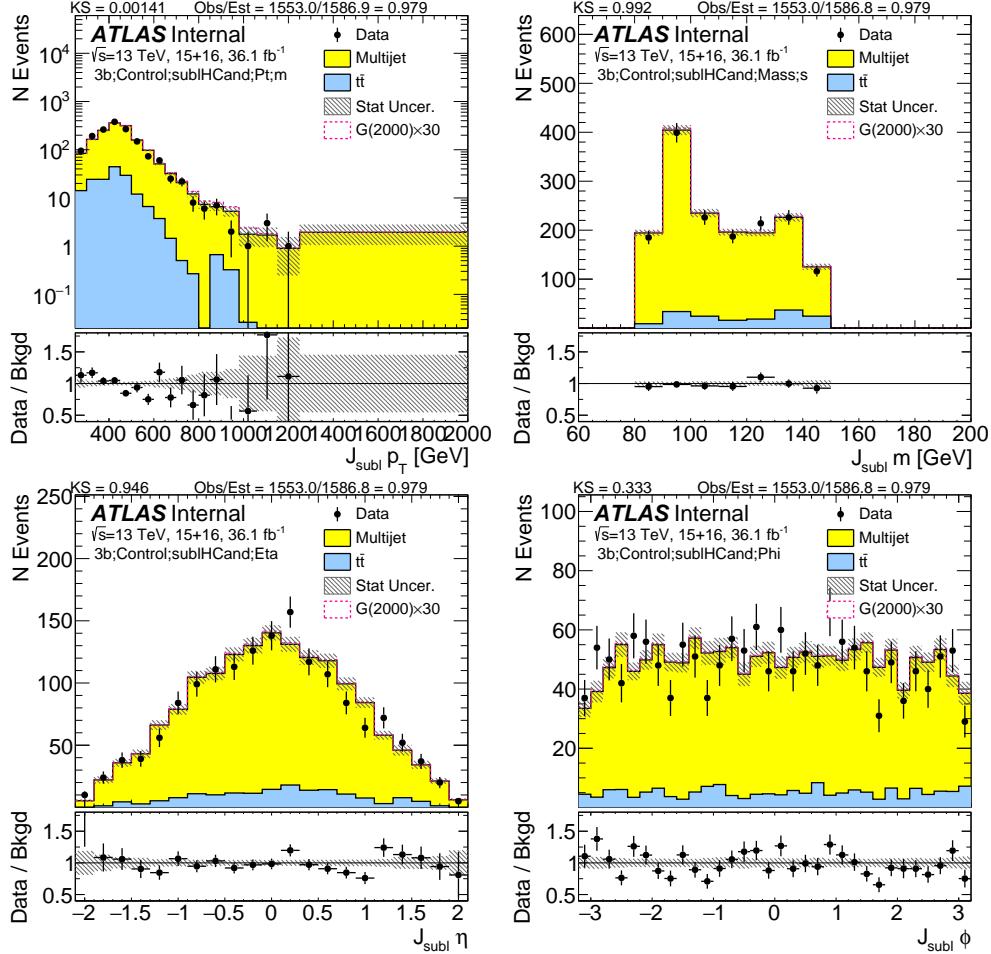


Figure 7.34: Kinematics of the sub-lead large- R jet in data and prediction in the control region after requiring 3 b -tags.

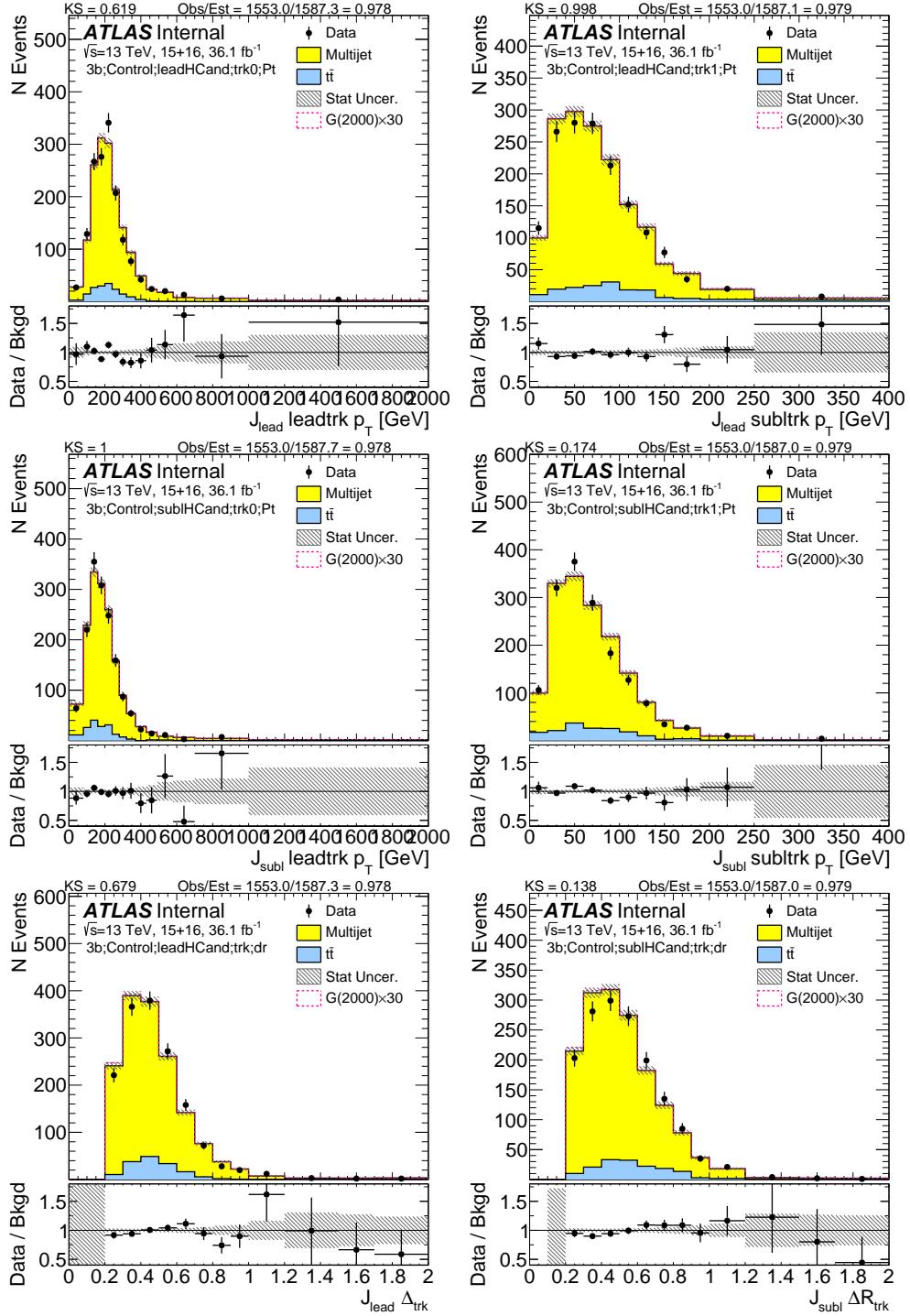


Figure 7.35: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the control region after requiring 3 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet.

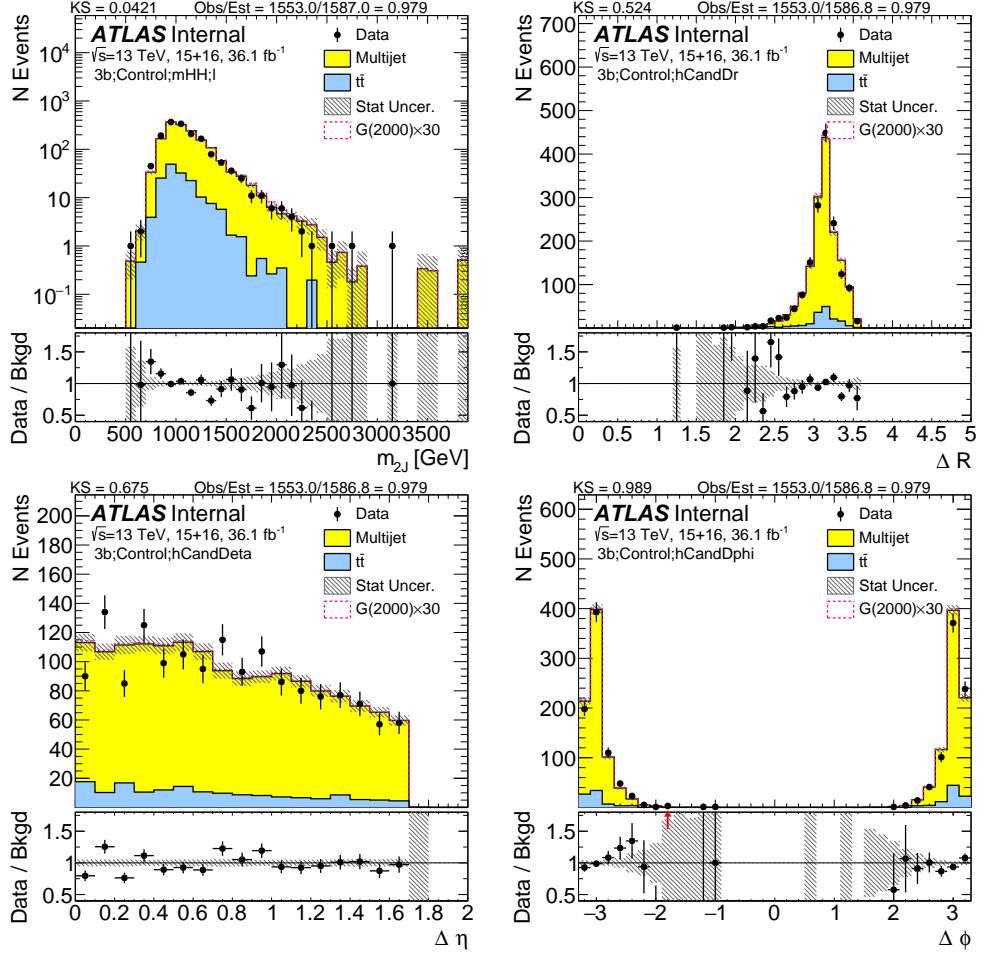


Figure 7.36: Kinematics of the large- R jet system in data and prediction in the control region after requiring 3 b -tags.

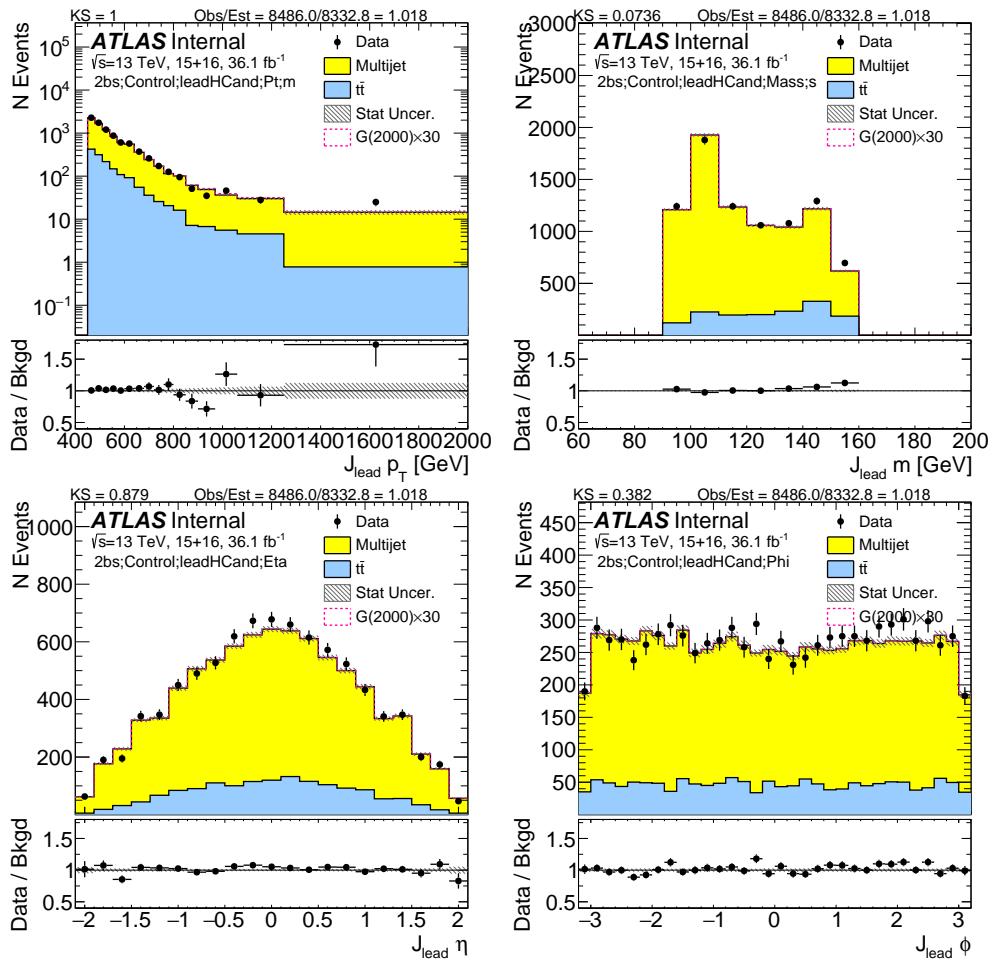


Figure 7.37: Kinematics of the lead large- R jet in data and prediction in the control region after requiring 2 b -tags split.

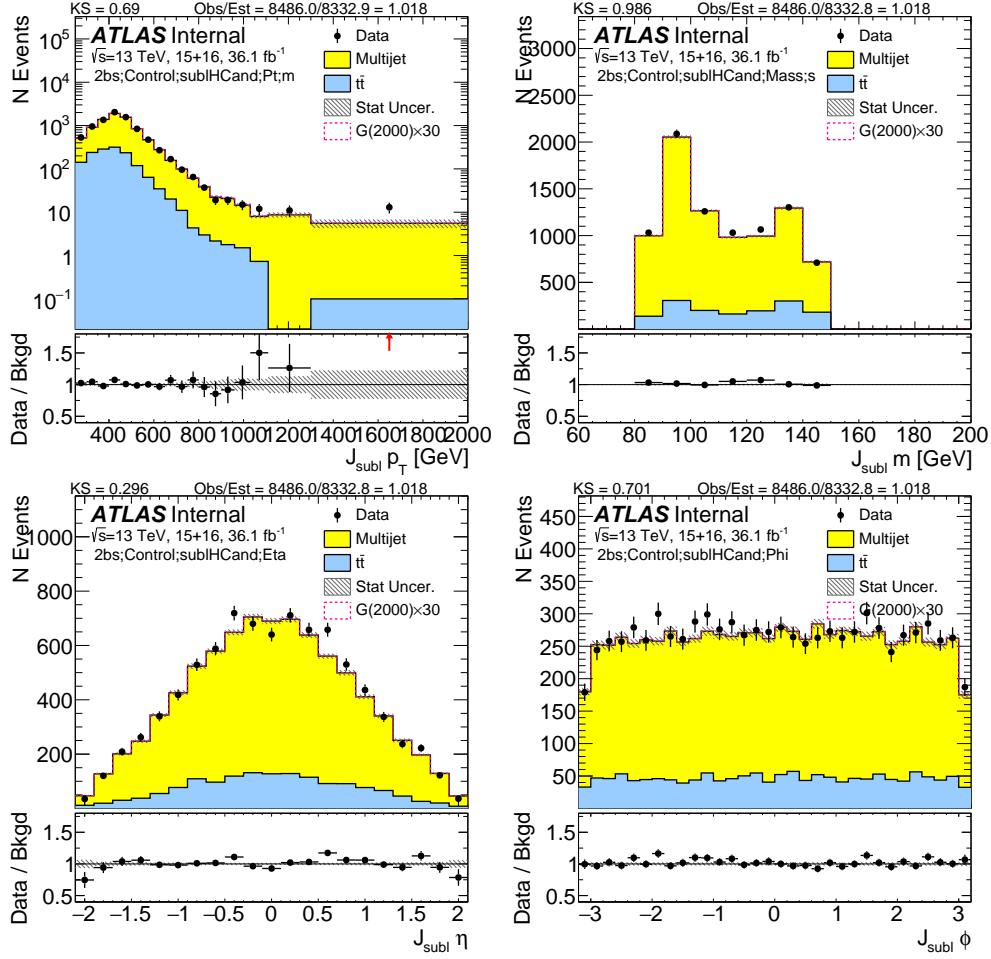


Figure 7.38: Kinematics of the sub-lead large- R jet in data and prediction in the control region after requiring 2 b -tags split.

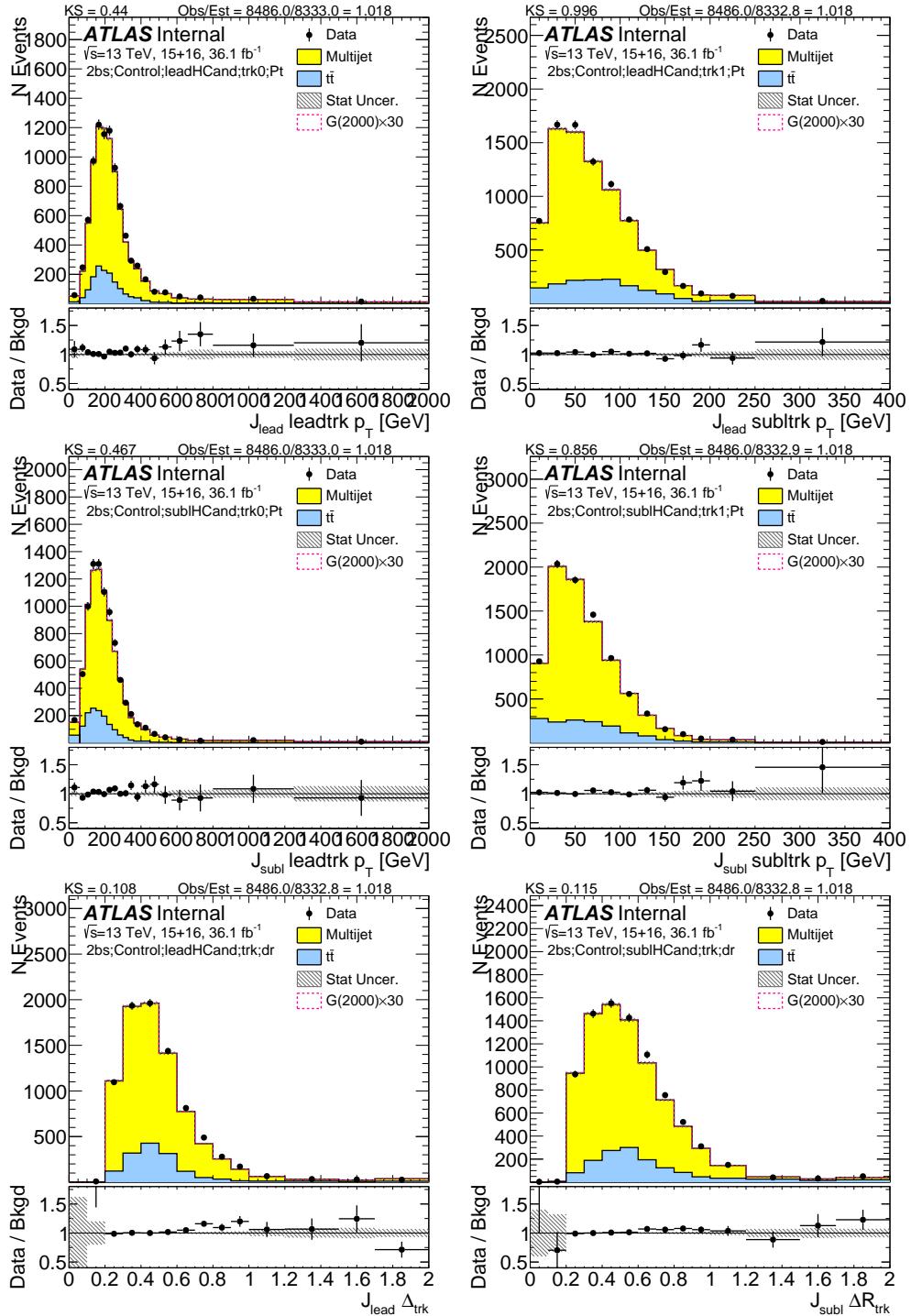


Figure 7.39: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the control region after requiring 2 b -tags split. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet.

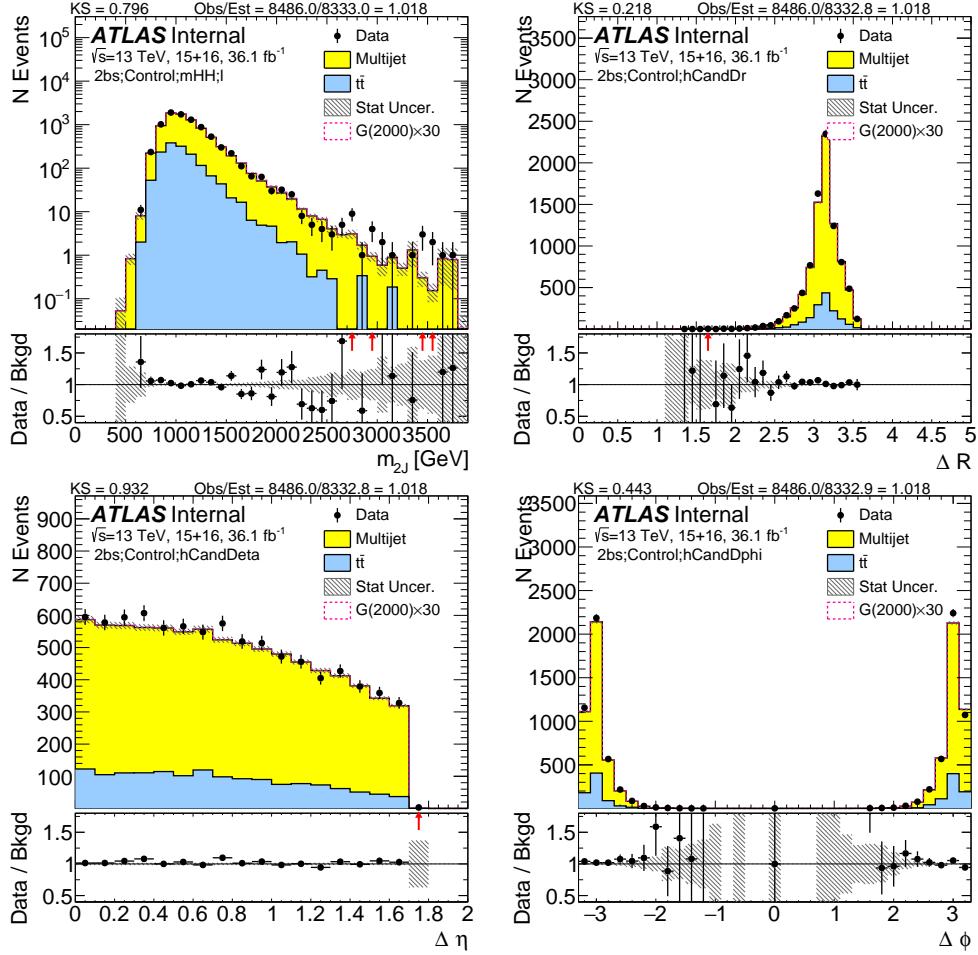


Figure 7.40: Kinematics of the large- R jet system in data and prediction in the control region after requiring 2 b -tags split.

7.1.9 SIGNAL REGION PREDICTIONS

This section shows comparisons of data with the prediction of QCD multi-jets and $t\bar{t}$ in the signal region (SR). Plots shown are blinded. The unblinded data is shown in the Result Section, section ??.

Figures 7.41, 7.42, 7.43, and 7.44 show predictions of various kinematics of the large- R jets and their associated track jets in the $4b$ selection. The shapes and normalization are a feature of the prediction, where the normalization is derived in the SB.

Figures 7.45, 7.46, 7.47, and 7.48 show predictions of various kinematics of the large- R jets and their associated track jets in the $3b$ selection. The shapes and normalization are a feature of the prediction, where the normalization is derived in the SB.

Figures 7.49, 7.50, 7.51, and 7.52 show predictions of various kinematics of the large- R jets and their associated track jets in the $2b$ selection. The shapes and normalization are a feature of the prediction, where the normalization is derived in the SB.

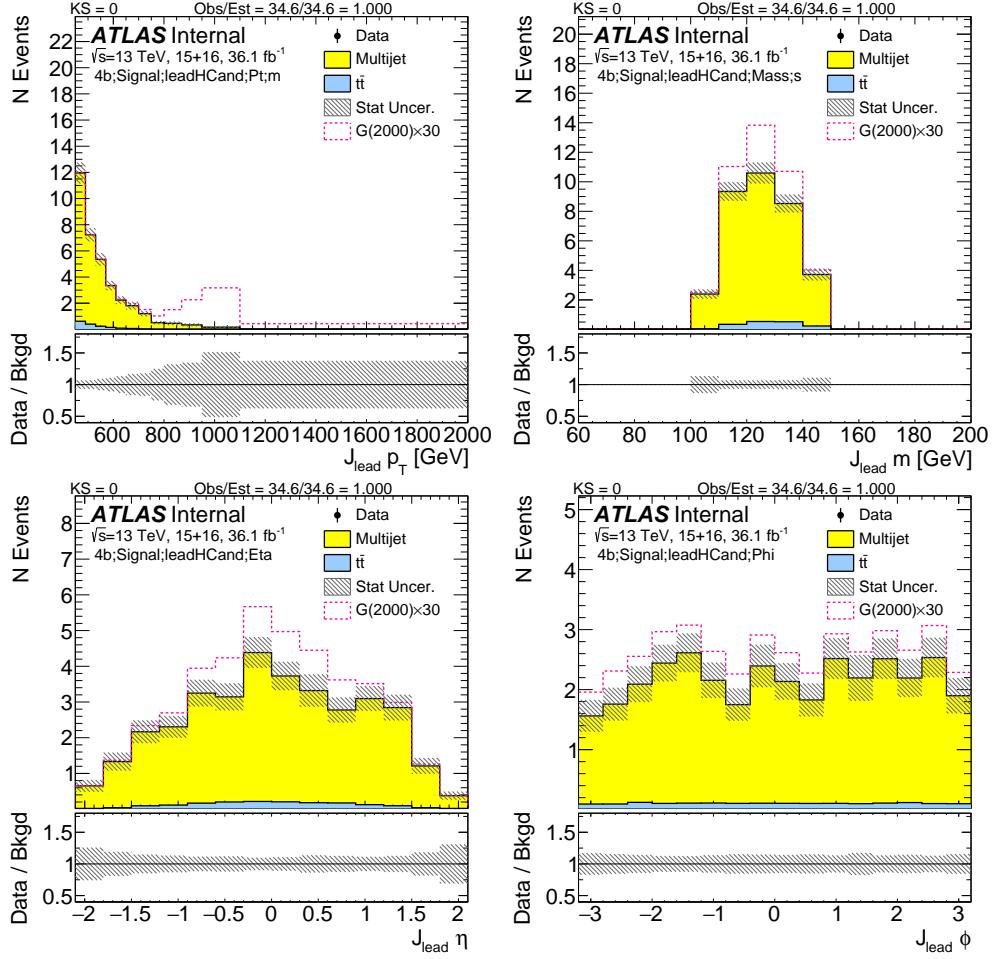


Figure 7.41: Kinematics of the lead large- R jet in data and prediction in the signal region after requiring 4 b -tags. Data is blinded, and will be added after unblinding.

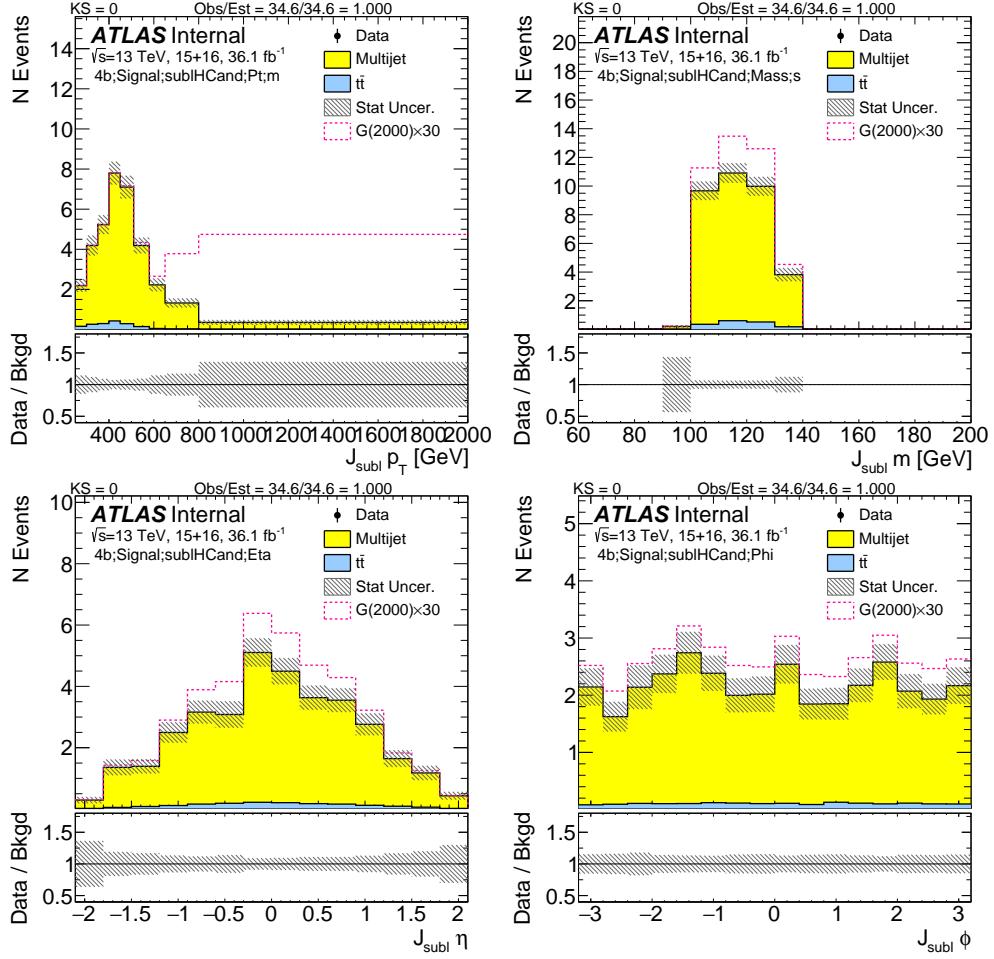


Figure 7.42: Kinematics of the sub-lead large- R jet in data and prediction in the signal region after requiring 4 b -tags. Data is blinded, and will be added after unblinding.

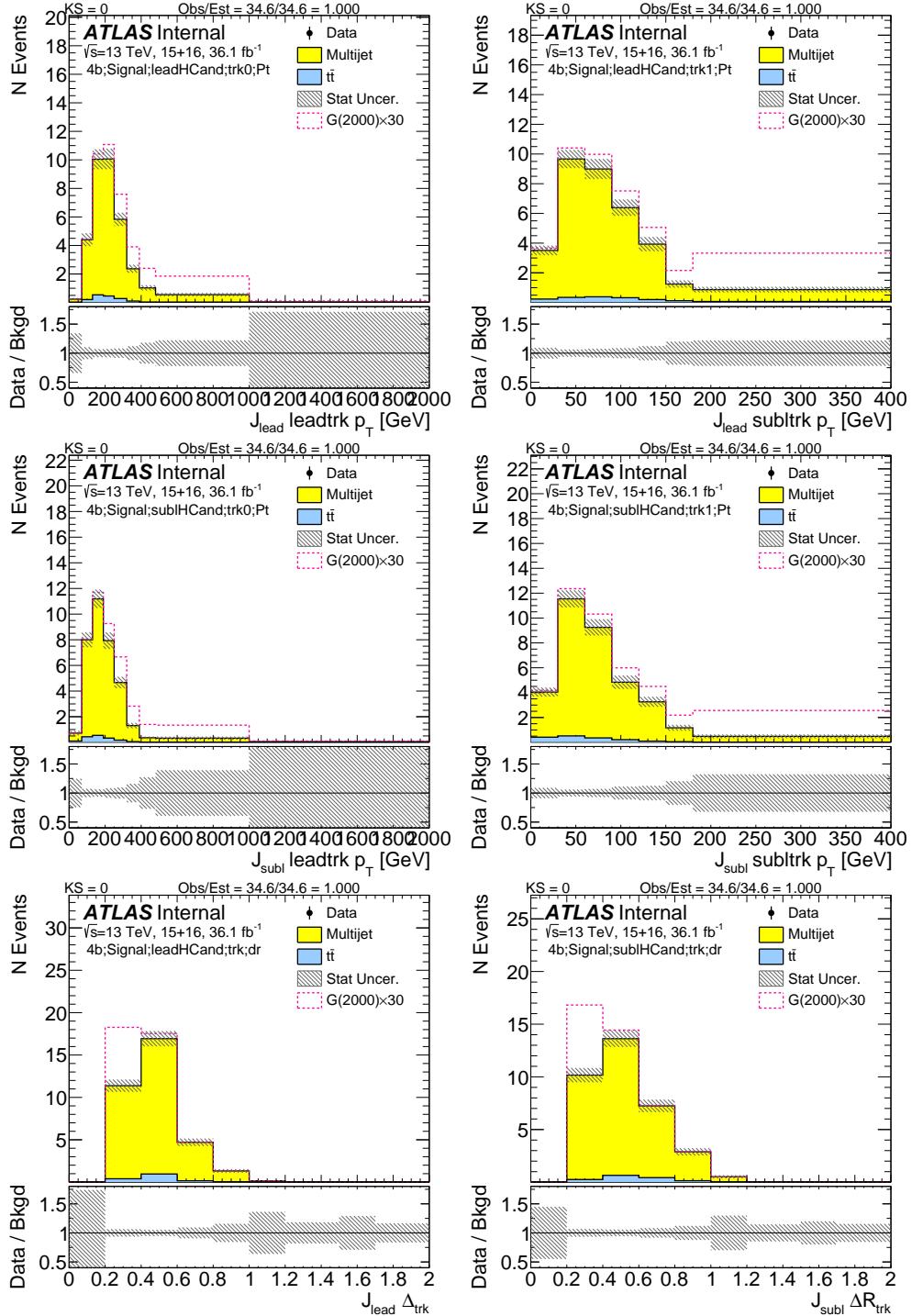


Figure 7.43: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the signal region after requiring 4 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. Data is blinded, and will be added after unblinding.

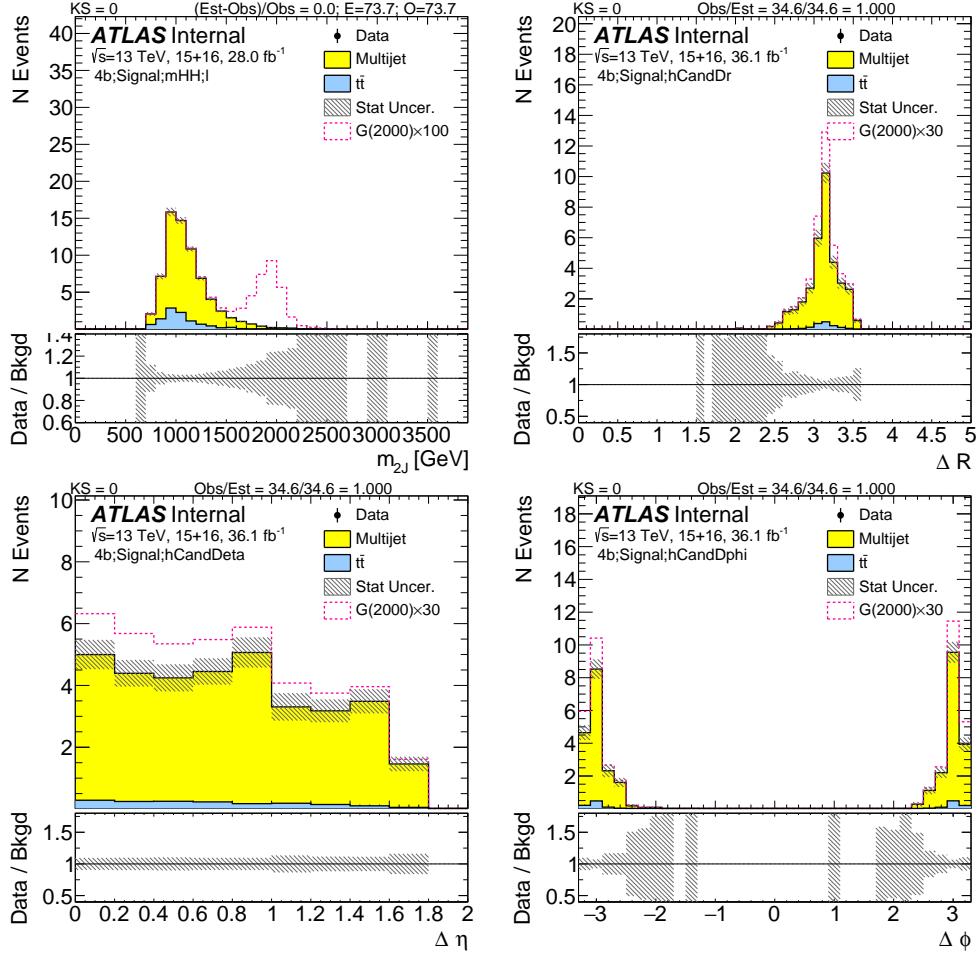


Figure 7.44: Kinematics of the large- R jet system in data and prediction in the signal region after requiring 4 b -tags. Data is blinded, and will be added after unblinding.

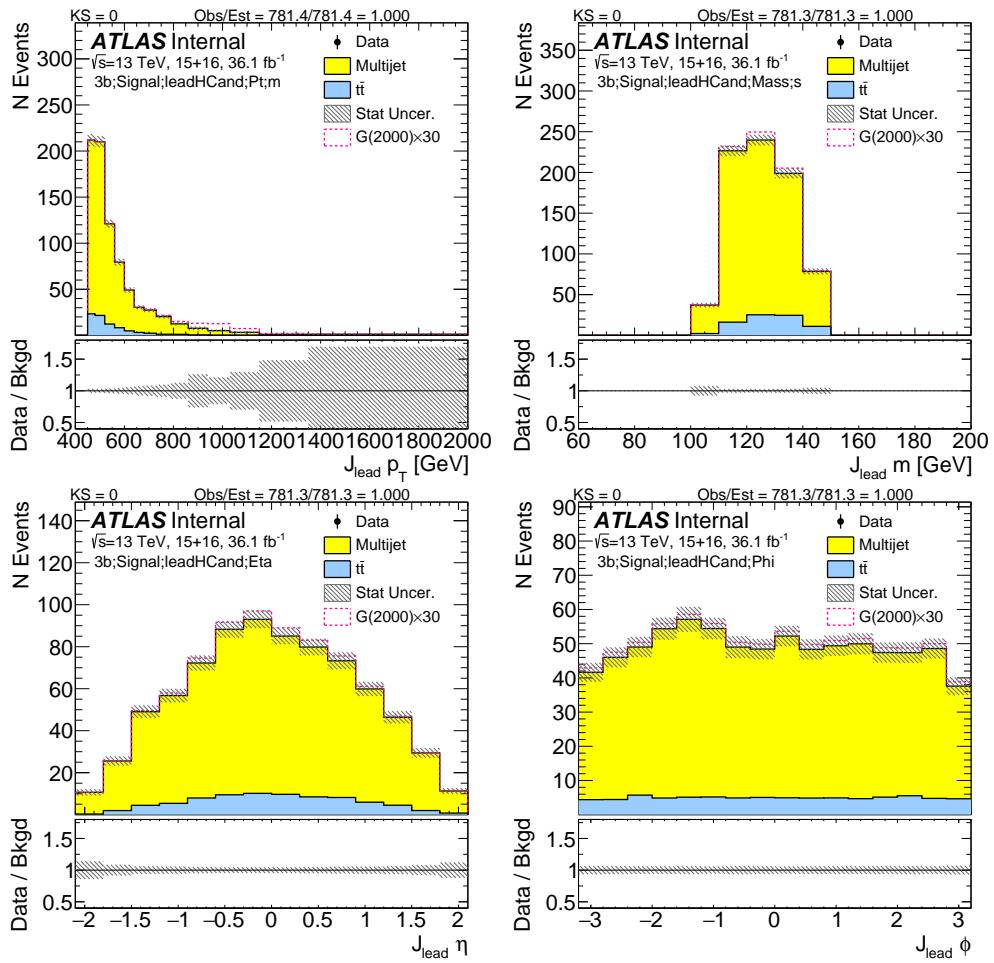


Figure 7.45: Kinematics of the lead large- R jet in data and prediction in the signal region after requiring 3 b -tags. Data is blinded, and will be added after unblinding.

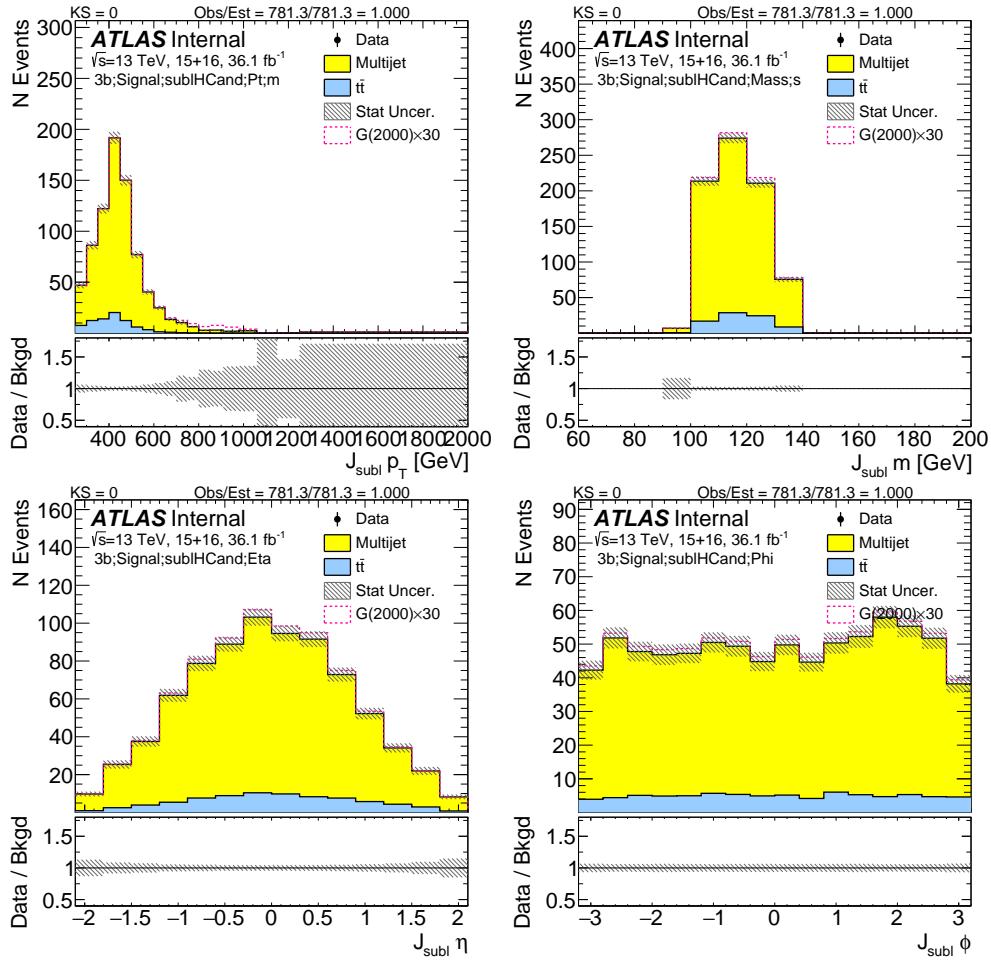


Figure 7.46: Kinematics of the sub-leading large- R jet in data and prediction in the signal region after requiring 3 b -tags. Data is blinded, and will be added after unblinding.

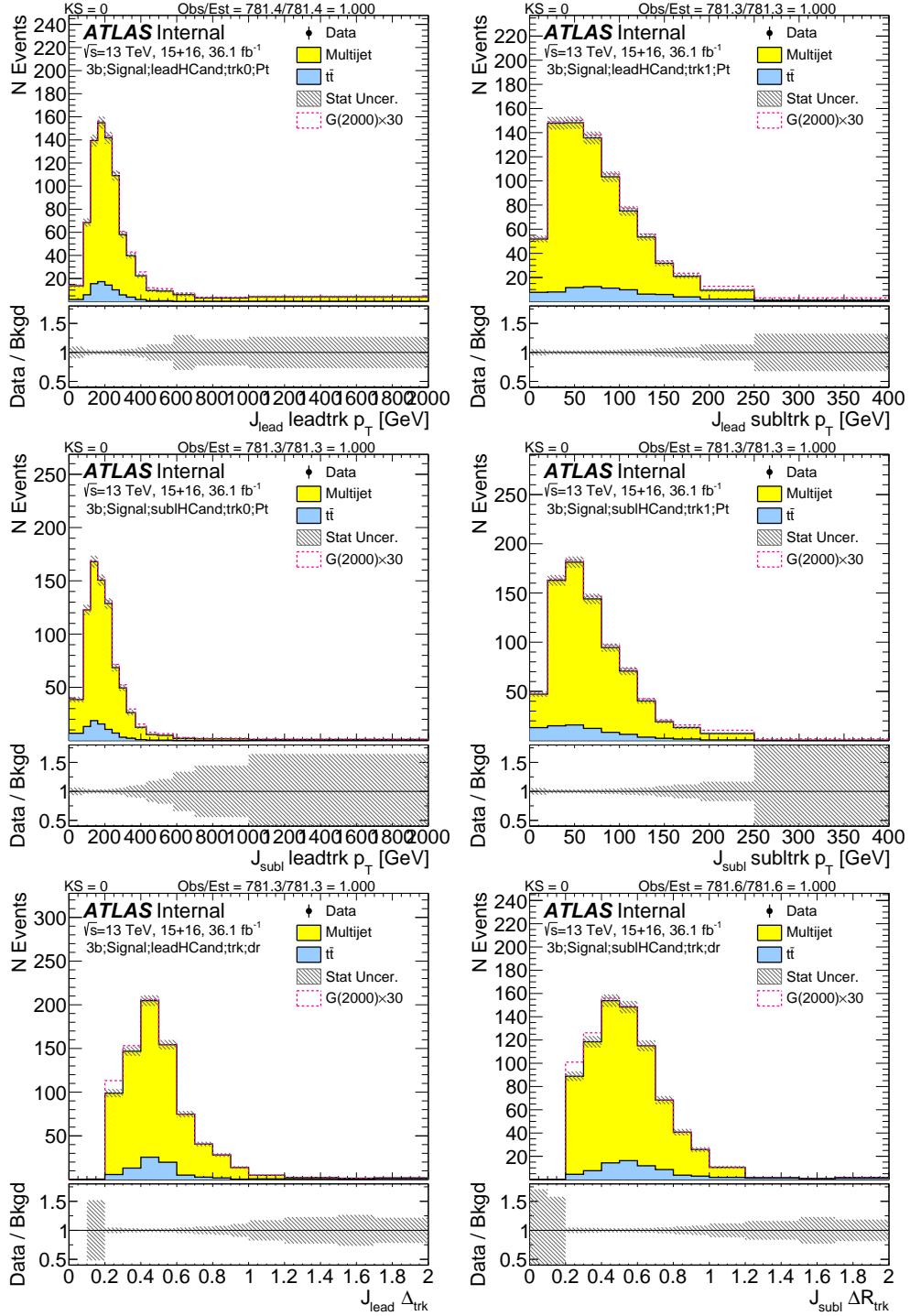


Figure 7.47: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the signal region after requiring 3 b -tags. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. Data is blinded, and will be added after unblinding.

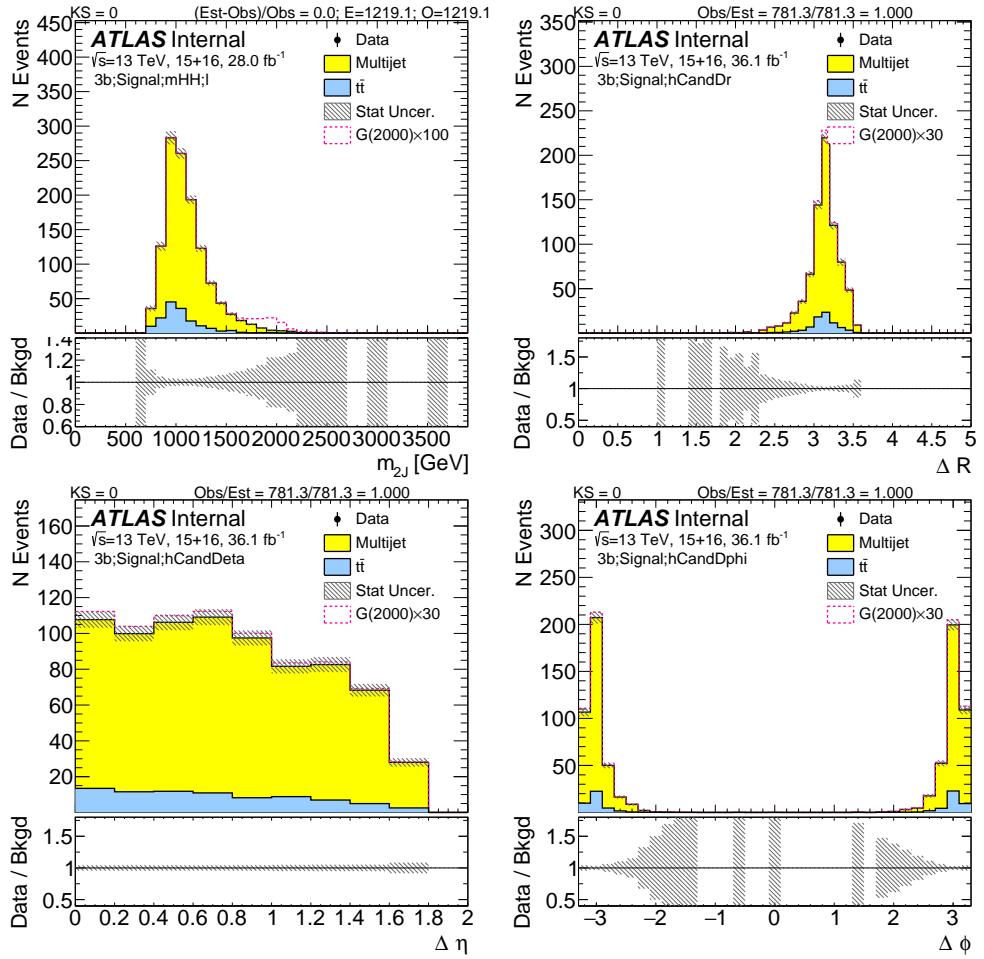


Figure 7.48: Kinematics of the large- R jet system in data and prediction in the signal region after requiring 3 b -tags. Data is blinded, and will be added after unblinding.

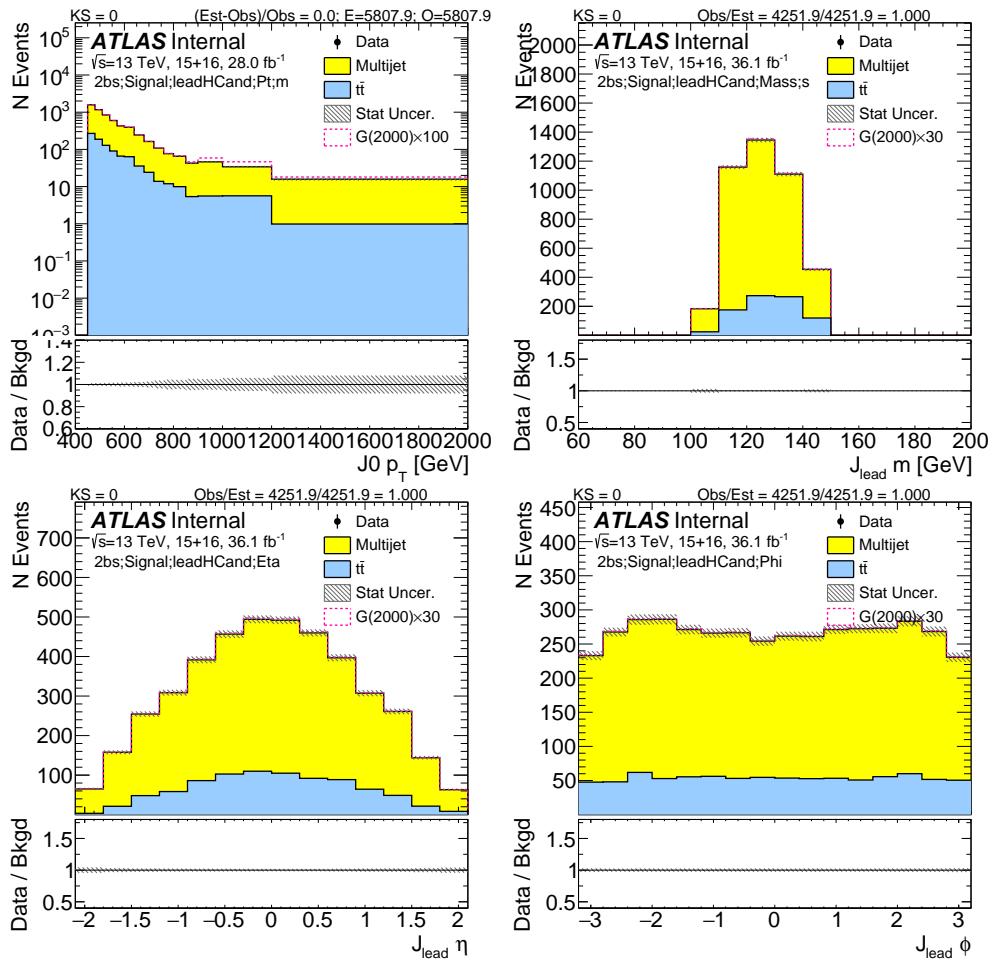


Figure 7.49: Kinematics of the lead large- R jet in data and prediction in the signal region after requiring 2 b -tags split. Data is blinded, and will be added after unblinding.

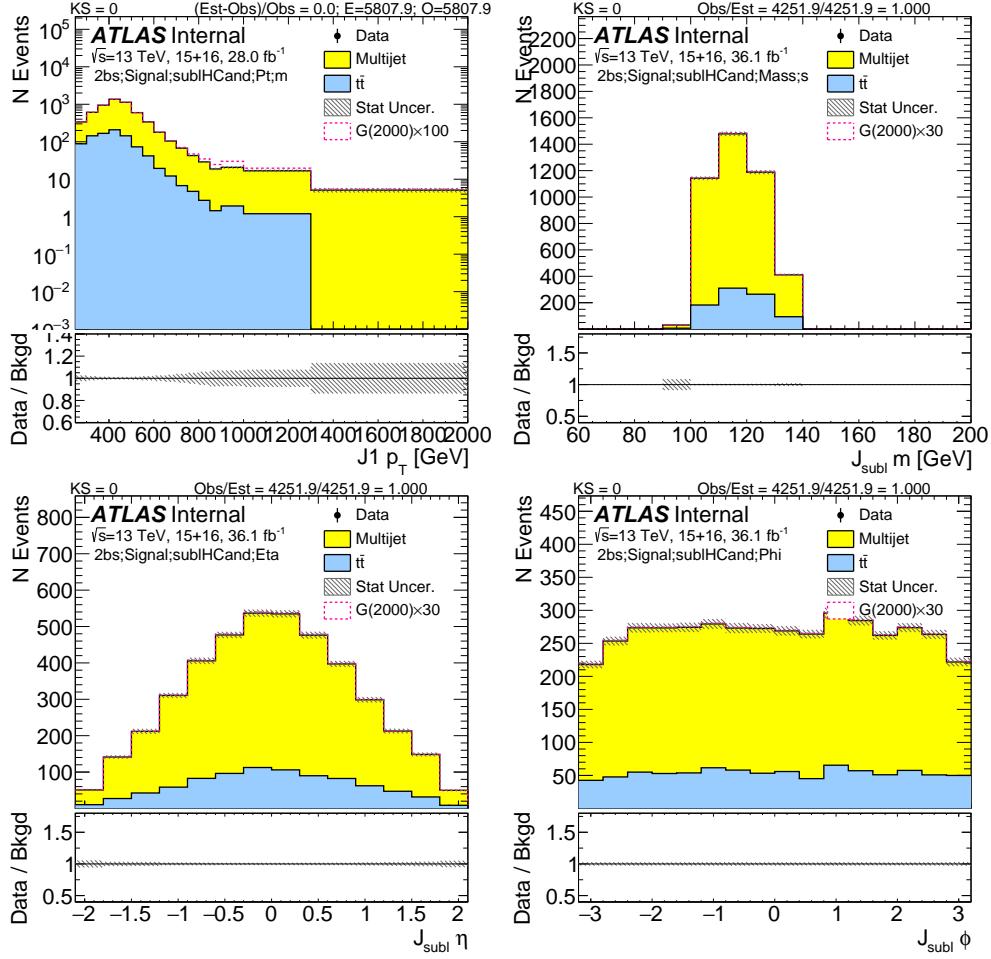


Figure 7.50: Kinematics of the sub-leading large- R jet in data and prediction in the signal region after requiring 2 b -tags split. Data is blinded, and will be added after unblinding.

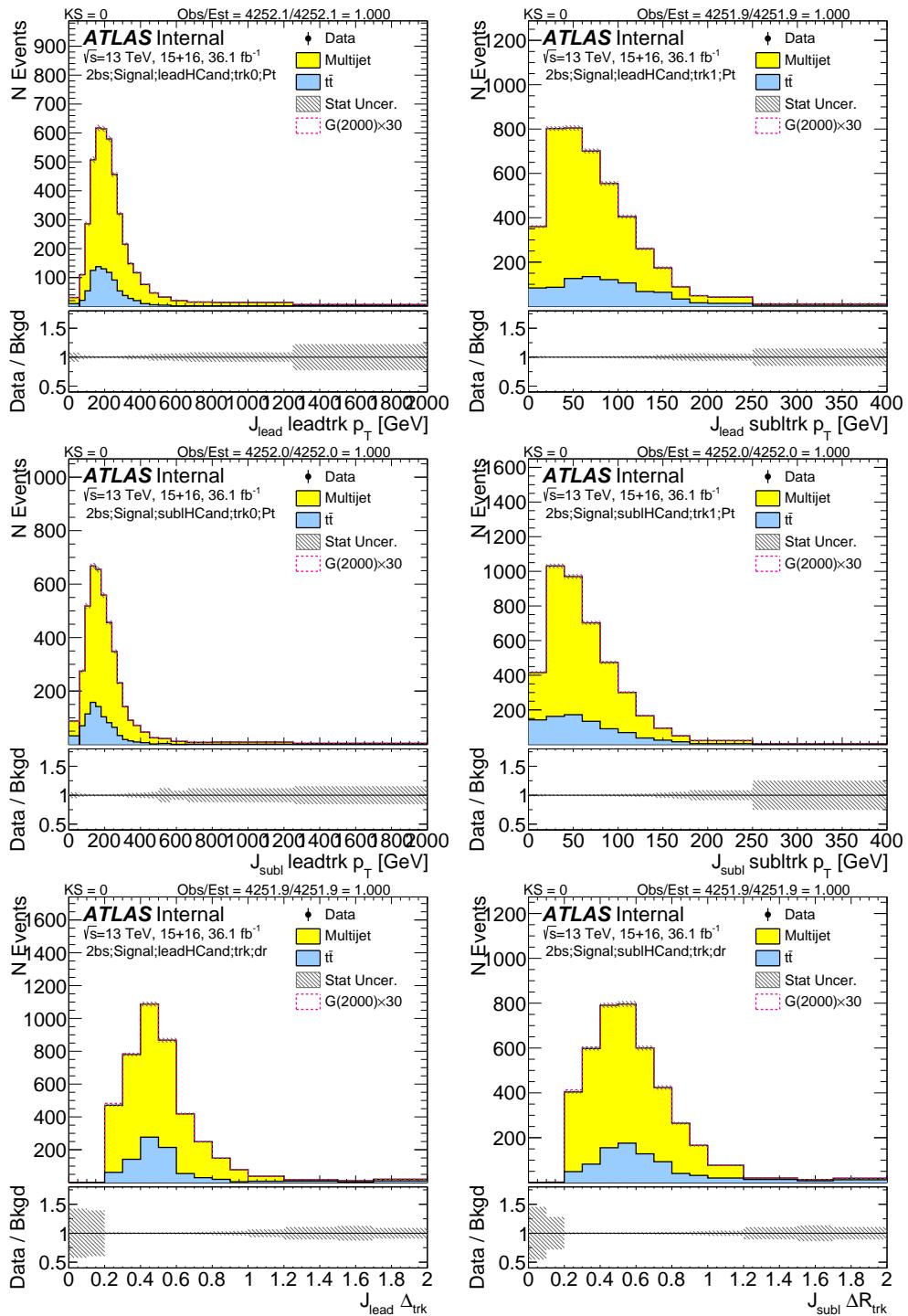


Figure 7.51: First two rows show the kinematics of the lead (left) and sub-lead (right) small- R track jets associated to the lead (first-row) and sub-lead (second-row) large- R jet in data and prediction in the signal region after requiring 2 b -tags split. Third row shows the ΔR between two leading small- R track-jets associated to the leading (left) and sub-leading (right) large- R jet. Data is blinded, and will be added after unblinding.

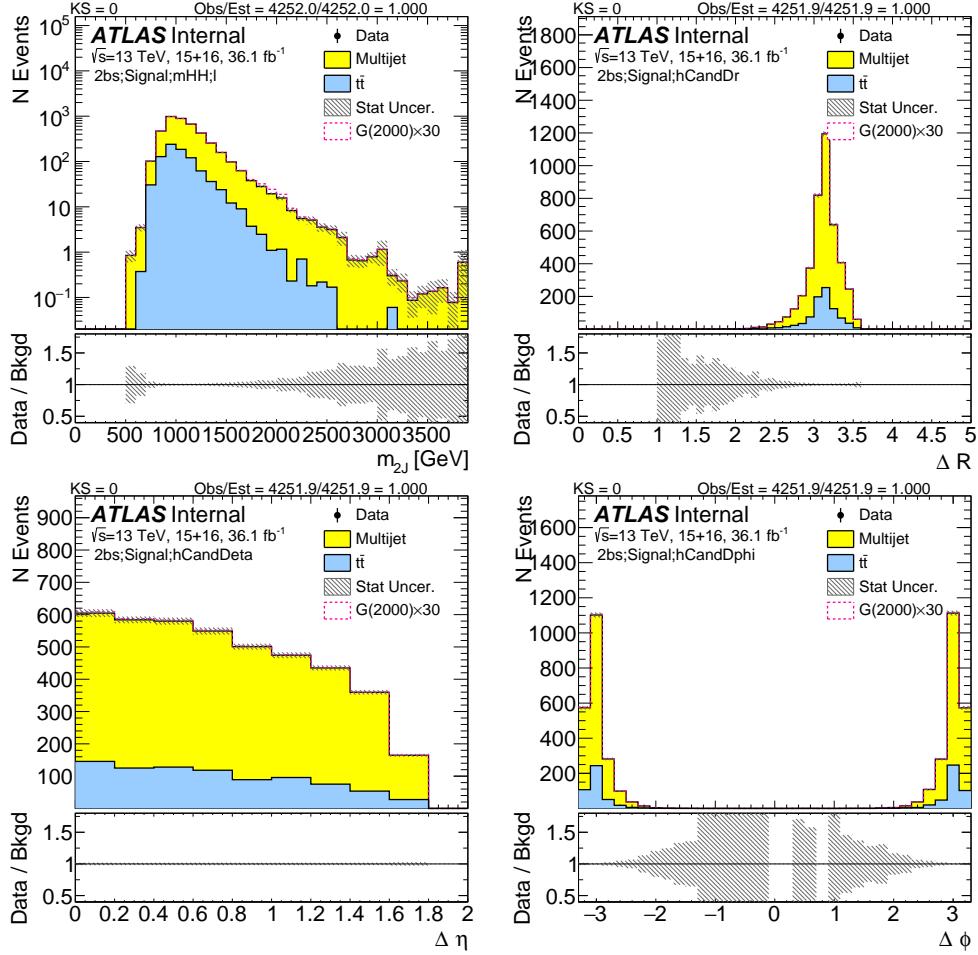


Figure 7.52: Kinematics of the large- R jet system in data and prediction in the signal region after requiring 2 b -tags split. Data is blinded, and will be added after unblinding.

SIGNAL REGION SMOOTHING

Due to the improved $1/2b$ statistics at high di-large – R jet invariant mass above 1500 GeV and the limited $t\bar{t}$ statistics above 1100 GeV, different fits are performed to smooth the di-large – R jet mass distribution in the signal region. The $1/2b$ QCD background is fit with the MJ8 functional form:

$$y = \frac{\alpha}{\frac{x}{\sqrt{s}}^2} \left(1 - \frac{x}{\sqrt{s}}\right)^{b-c} \log\left(\frac{x}{\sqrt{s}}\right) \quad (7.3)$$

where $\sqrt{s} = 13000$ GeV, in the range $1200 < M_{JJ} < 3000$ GeV, and the three free parameters are a , b and c . This form is used in fitting, as it was seen to be easier for the fits to converge. The signal region $t\bar{t}$ distribution is fitted also with the dijet functional form, also in the range $1200 < M_{JJ} < 3000$ GeV, without parameter constraints. The values of the estimated fit parameters in the $4b$ and $3b$ and $2bs$ signal regions can be found in Table 7.3.

Given that the similar $1/2b$ sample is used for deriving the QCD shape for the $4/3/2bs$ signal regions, it is not surprising that the slope parameter (a) is similar in the $4/3/2bs$ signal regions for each the QCD backgrounds.

Due to the very limited statistics of the $4b$ $t\bar{t}$ sample, the $4b$ $t\bar{t}$ dijet mass shape is used from the $3b$ $t\bar{t}$ dijet mass shape normalized to the number of $4b$ $t\bar{t}$ events. A comparison of the shape is shown in Figure 7.53. Good agreement between the $4b$ and $3b$ signal region plot is shown.

Figure 7.54 shows the smoothing fits for the QCD background and the $t\bar{t}$ background in the $4b$ signal region. Figure 7.55 shows the same for the $3b$ signal region. Figure 7.56 shows the same for the $2bs$ signal region. The smoothing statistical uncertainties are also shown on these two plots. More additional uncertainties, such as uncertainty from choice of smoothing function, will be discussed in the Section 8.0.3.

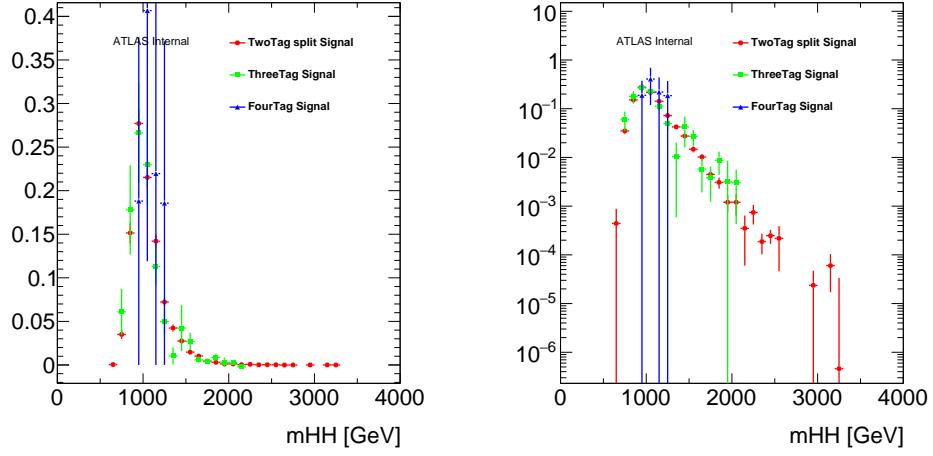


Figure 7.53: Comparison of the $4b$, $3b$ and $2bs$ signal region $t\bar{t}$ dijet mass shape. On the left is the linear scale, and on the right is the log scale. Both distributions are normalized to 1 for comparison.

The final smoothed background predictions for the $4b$ and $3b$ and $2bs$ signal regions can be found in Figure 7.57. This includes smoothing statistical uncertainties only. More details on other systematics, including smoothing systematics, shape uncertainties and other sources of uncertainties would be discussed in Section 8.0.3.

Uncertainties on the fit parameters are propagated as systematic uncertainties, though they are essentially replacing the bin-by-bin statistical uncertainties of the background estimates (which are not used once smoothing is applied). Correlations in the fit parameters of the backgrounds are taken into account when propagating the uncertainties, as described in Appendix A.6.

Region	$a_{t\bar{t}}$	$b_{t\bar{t}}$	$c_{t\bar{t}}$	a_{qcd}	b_{qcd}	c_{qcd}
FourTag	-1.02 ± 1.22	35.62 ± 10.83	9.05 ± 9.59	7.75 ± 1.8	-18.0 ± 16.75	54.36 ± 14.28
ThreeTag	2.83 ± 1.22	35.62 ± 10.83	9.05 ± 9.59	10.42 ± 1.73	-27.1 ± 15.78	56.91 ± 13.89
TwoTag split	5.21 ± 1.22	35.62 ± 10.83	9.05 ± 9.59	7.74 ± 0.36	7.22 ± 3.11	24.54 ± 2.78

Table 7.3: Smoothing parameters in $4b$ and $3b$ and $2bs$ signal regions, the correlation between parameters is almost always 0.99.

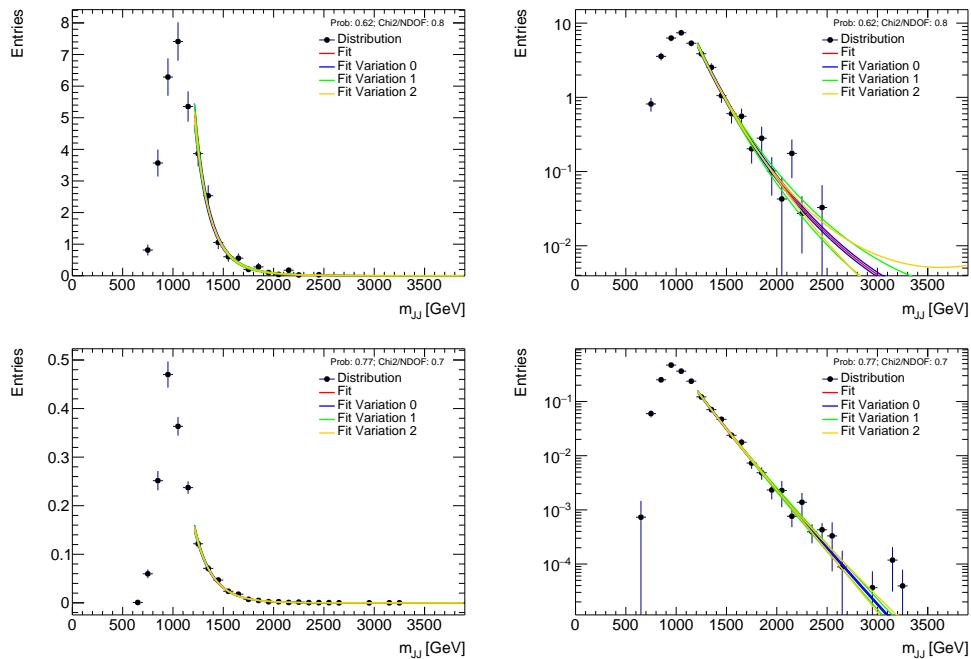


Figure 7.54: Fits for background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $4b$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.

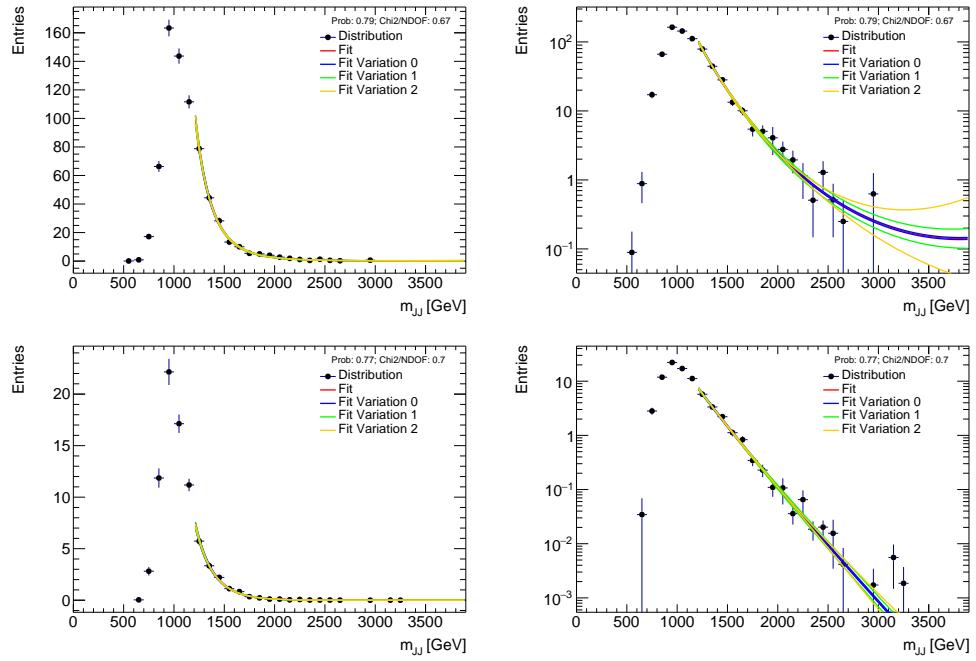


Figure 7.55: Fits for background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $3b$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.

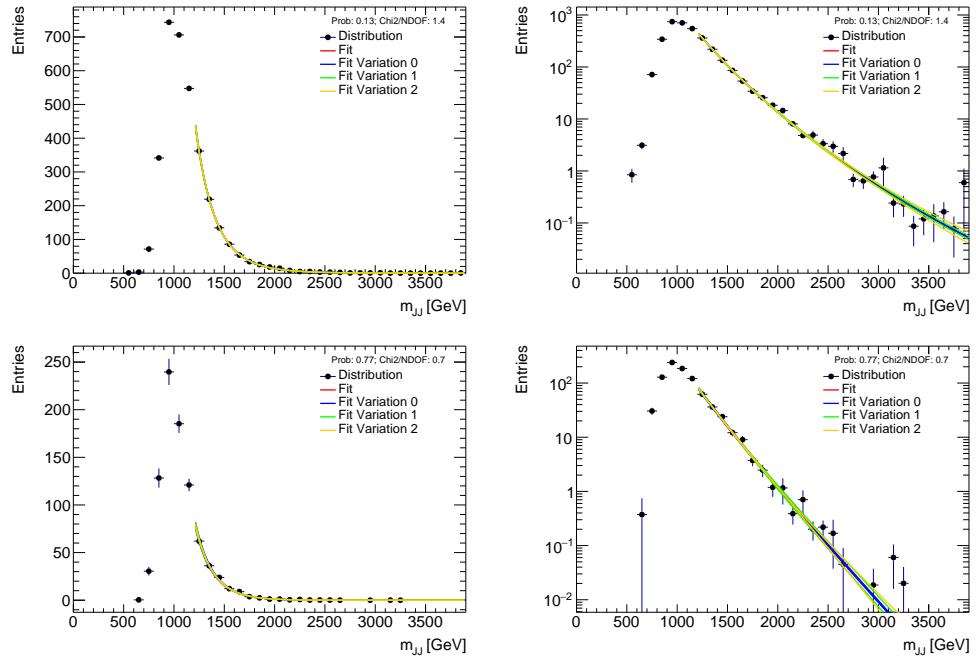


Figure 7.56: Fits for background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $2bs$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.

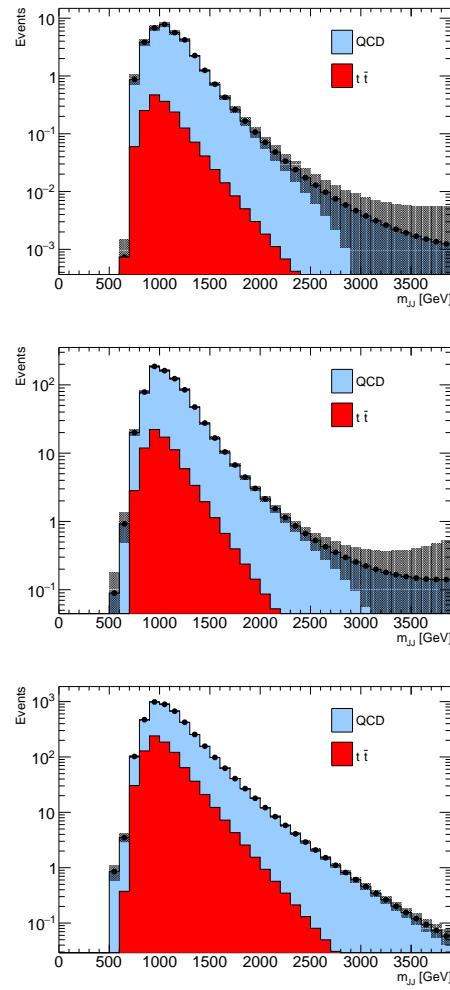


Figure 7.57: Smoothed background estimations the $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal regions. Only smoothing statistical uncertainties are shown here.

SCALED DIJET MASS DISTRIBUTION IN SIGNAL REGION

As is done in the resolved analysis, we also consider the scaled M_{JJ} distribution. In this case, the two higgs candidate 4-vectors are scaled by m_h/m_J , where $m_h = 125$ GeV, and m_J is the *large* – R jet mass of the Higgs candidate. While this distribution is expected to have less impact on the boosted analysis, because the mass correction is small relative to the signal masses being considered, we investigate this variable for possible improvements and for consistency with the resolved analysis. The scaled dijet mass distribution can be found in Figure 7.58. Its impact of the boosted analysis limit can be found in Appendix A.5.4.

For determining the choice of the final discriminant, both the expected limits on the nominal and scaled dijet mass distribution have been computed. Since the scaled dijet mass distribution based limits are consistent (slightly better at low mass and slightly worse at high mass, with differences of the order of 10%) than the nominal dijet mass limits, we proceed to use the scaled dijet mass distribution for consistency with the resolved analysis.

As is done for the dijet mass distribution, the scaled dijet mass distribution is smoothed. The smoothing is performed between 1200 GeV and 3000 GeV for the QCD and $t\bar{t}$. The smoothed distributions can be seen in Figures 7.59, 7.60, and 7.61. The values of the estimated fit parameters in the $4b$ and $3b$ and $2bs$ signal regions can be found in Table 7.4.

The final signal region prediction, using scaled di-jet mass distribution, with only statistical uncertainties, are shown in Figure 7.63(Figure 7.64) as before(after) smoothing.

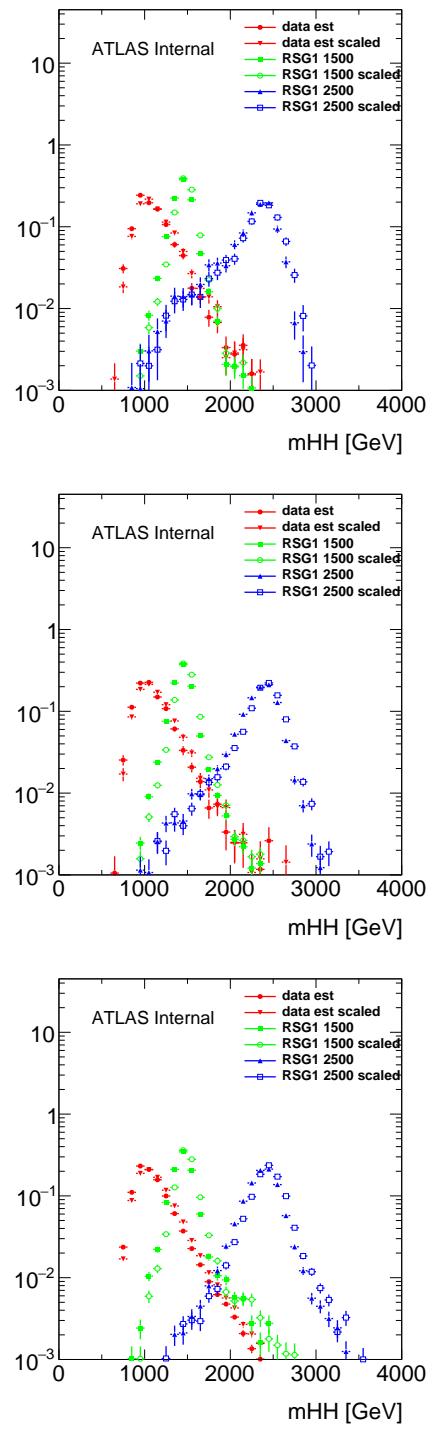


Figure 7.58: Normalized Scaled dijet mass distributions for the $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal regions. For comparison, the unscaled distributions are shown on the same plot.

Region	$a_{t\bar{t}}$	$b_{t\bar{t}}$	$c_{t\bar{t}}$	a_{qcd}	b_{qcd}	c_{qcd}
FourTag	-2.02 ± 1.17	42.46 ± 9.87	1.31 ± 8.98	-0.49 ± 1.59	53.06 ± 15.2	-11.1 ± 12.8
ThreeTag	1.84 ± 1.17	42.45 ± 9.88	1.32 ± 8.98	8.51 ± 0.98	-13.8 ± 9.14	42.58 ± 7.87
TwoTag split	4.22 ± 1.17	42.45 ± 9.88	1.32 ± 8.98	7.06 ± 0.32	11.54 ± 2.77	19.05 ± 2.48

Table 7.4: Smoothing parameters in $4b$ and $3b$ and $2bs$ signal regions for scaled mass distributions, the correlation between parameters is almost always 0.99.

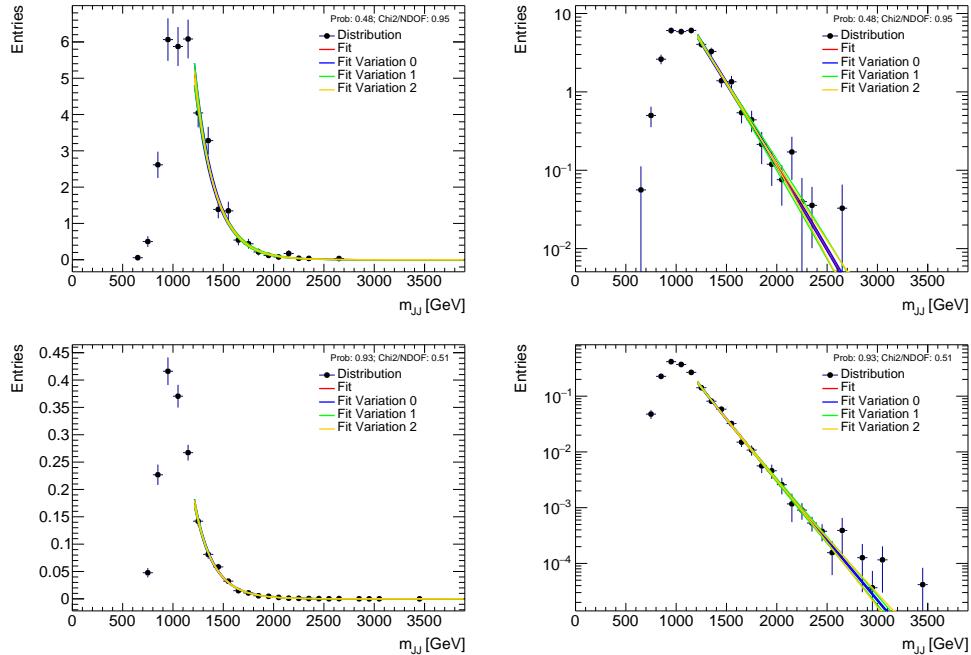


Figure 7.59: Fits for scaled background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $4b$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.

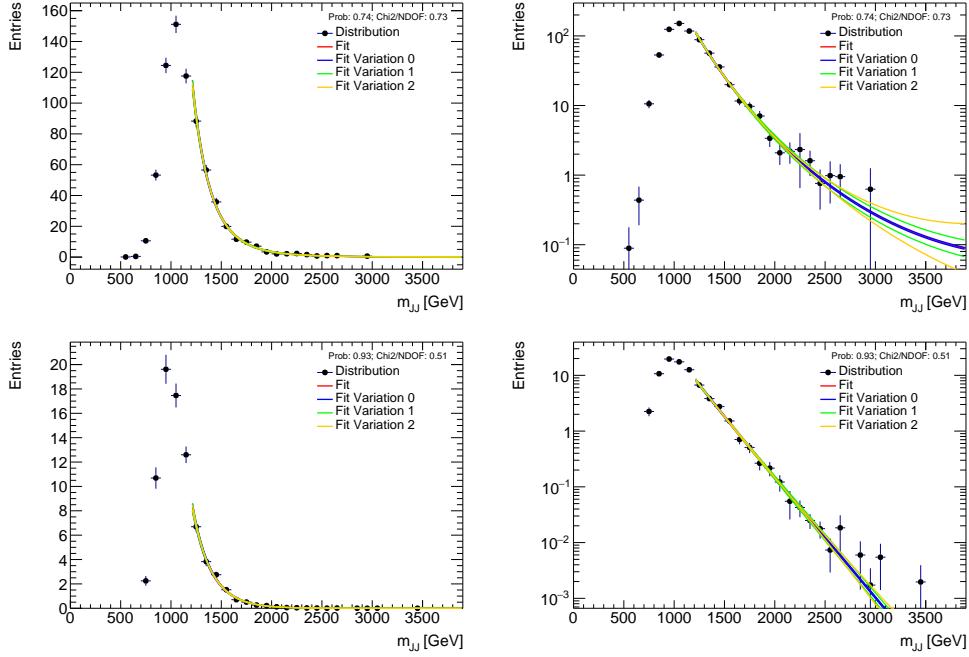


Figure 7.60: Fits for scaled background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $3b$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying fit parameters within uncertainties whilst taking into account parameter correlations.

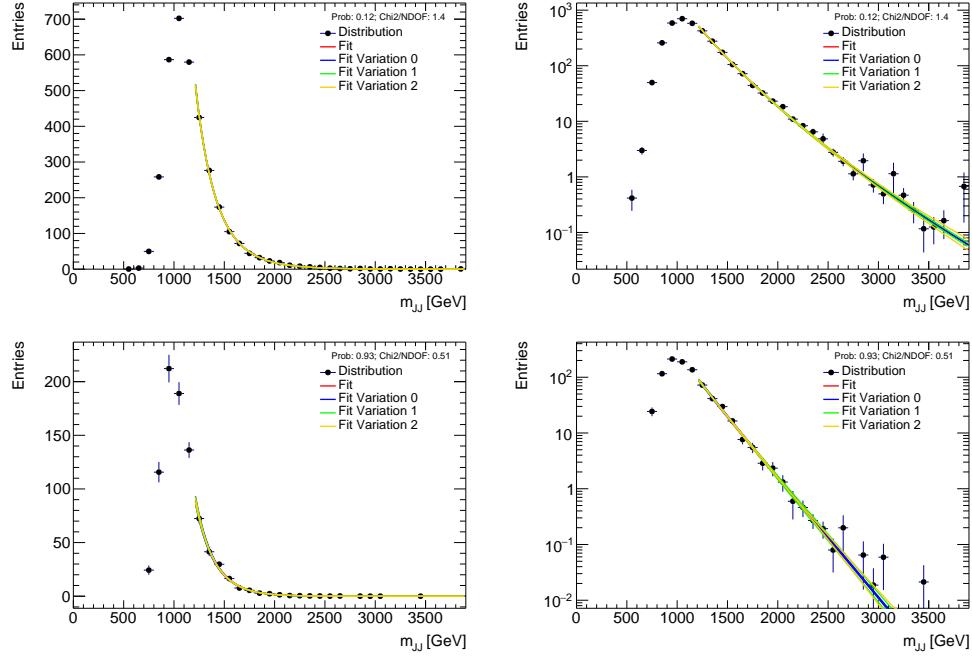


Figure 7.61: Fits for scaled background smoothing are shown for QCD (top row) and $t\bar{t}$ (bottom row) in the $2bs$ signal region. The left figures show the distributions with linear y -axis scale along with the fit central value and variations. The right figures show the distributions with log y -axis scale along with the fit central value and the fit variations as determined by the varying the fit parameters within uncertainties whilst taking into account parameter correlations.

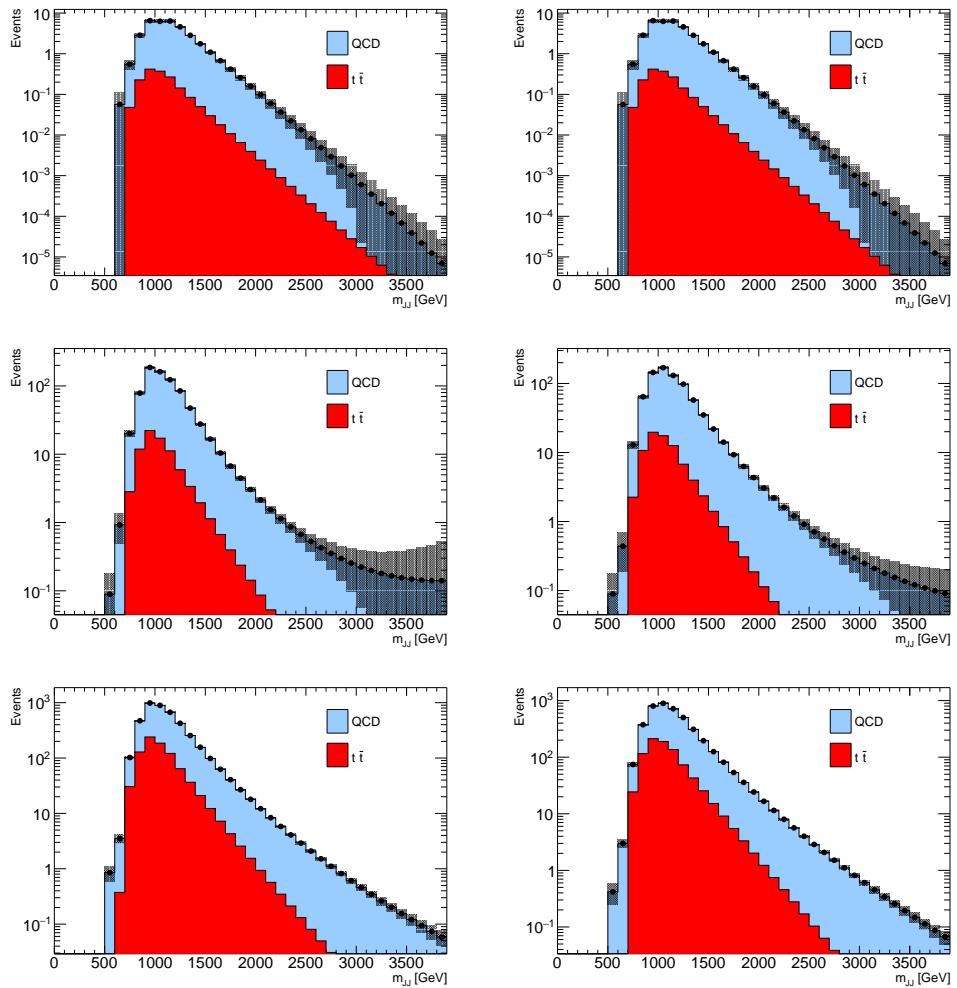


Figure 7.62: Smoothed MJJ (left) and scaled MJJ (right) background estimations the $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal regions. Smoothing statistical and systematic uncertainties from smoothing parameter variations are shown here.

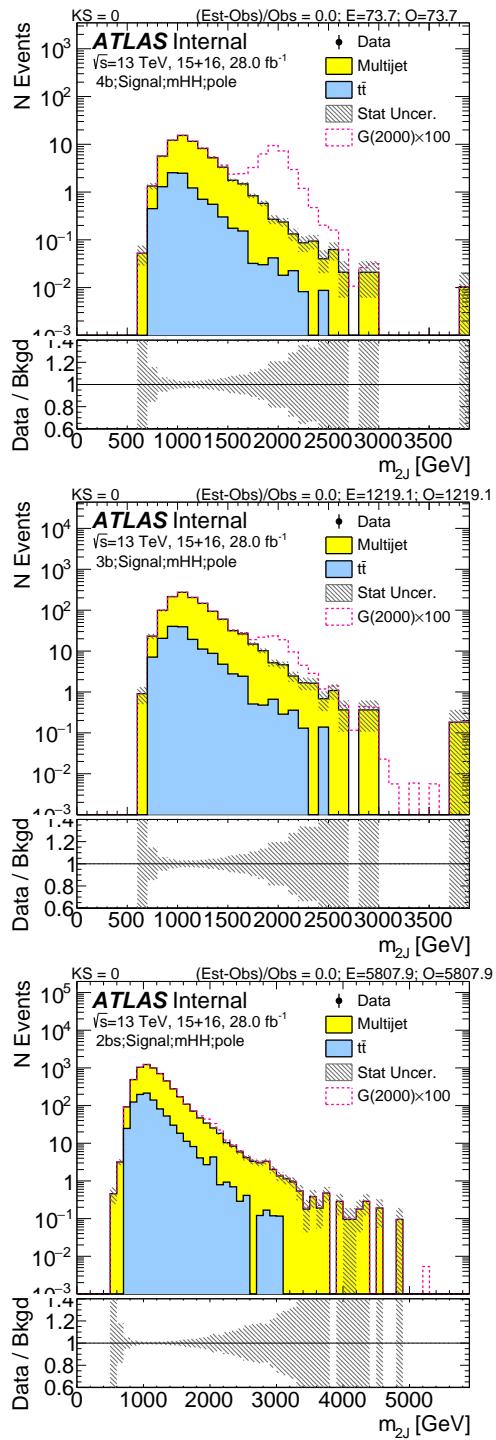


Figure 7.63: Background prediction for $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal region using scaled di-jet mass before smoothing. The uncertainty band includes only statistical uncertainties.

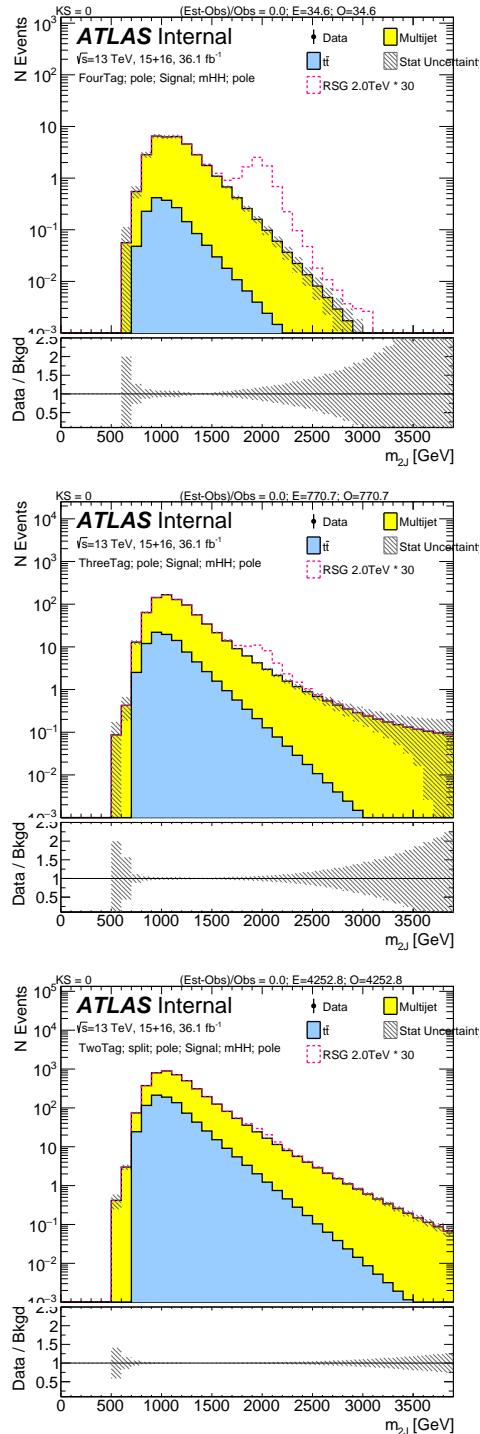


Figure 7.64: Background prediction for $4b$ (top), $3b$ (middle), and $2bs$ (bottom) signal region using scaled di-jet mass after smoothing. The uncertainty band includes only statistical uncertainties.

Since we do physics.

Tony Tong

8

Systematics

All backgrounds are data-driven, with the exception of Z +jets, which is very small. The $t\bar{t}$ simulation is only used to determine the shape of expected events, where the $t\bar{t}$ normalization is derived in data. MC uncertainties are applied to all simulated samples and propagated anywhere a MC based shape is used.

8.0.1 MC UNCERTAINTIES

MC based uncertainties are propagated in the analysis using standard CP group recommendations. These uncertainties can change both the shape and normalization of the signal and of the MC-based background prediction (Z +jets). The multijet and $t\bar{t}$ backgrounds normalisations are estimated with a data driven method (the likelihood fit). Since the shape for the $t\bar{t}$ component is taken

from MC, the fit is redone for each MC variations.

LUMINOSITY UNCERTAINTY : The preliminary uncertainty on the combined 2015+2016 integrated luminosity is 3.2%, assuming uncorrelated uncertainties between years. It is derived, following a methodology similar to that detailed in Refs.[?] and[?], from a preliminary calibration of the luminosity scale using x-y beam-separation scans performed in August 2015 and May 2016. This uncertainty is applicable to the backgrounds with normalizations determined from simulation, and further propagated to the multijet prediction through data-driven background estimation procedure. It is expected to have a small impact on this analysis. This uncertainty is also applied to the signal normalization prediction

LARGE-R JET RESOLUTION AND SCALE UNCERTAINTIES : The uncertainties on the jet energy and mass (JES, JMS scale) are evaluated by the combined performance groups using track-to-calorimeter double ratios between data and MC, measured in dijet data^{??}. Discrepancies observed between data and MC are assigned as uncertainties on the energy/mass scales of the jet. Different correlation scenarios are supported by the Jet Substructure group, where all uncertainties can be decorrelated, fully correlated, or only correlated between energy and mass. Currently the latter is being used, known as the “medium” configuration. The uncertainties on the jet energy, mass resolutions are estimated by applying a Gaussian smearing which degrades the nominal resolution by an absolute 2% for p_T , relative 20% for mass.

The uncertainties in our kinematic regime for signal yield predictions are below 7% for JES/JMS. For $t\bar{t}$ yield predictions the uncertainties are \sim 7-24% for JES/JMS. Details are listed in section 8.0.4. The uncertainties in our kinematic regime for signal yield predictions are below 7% / 15% for JER/JMR. For $t\bar{t}$ yield predictions the uncertainties are \sim 4-27% for JER/JMR. Details are listed in section 8.0.4.

B-TAGGED TRACK JET SCALE FACTOR UNCERTAINTIES The uncertainties related to the b -tagging efficiency calibrations as measured in $t\bar{t}$ events for track-jets are considered, using the official prescriptions. The procedure to define these calibrations is similar to that described in reference[?].

The effect of the different experimental uncertainties on the signal yield is shown in section 8.0.4. The signal yield uncertainty due to b -tagging is less than 30% for the signal, and less than 12% for the $t\bar{t}$ background yield. The main difference with respect to the previous result is on the b -tagging uncertainties, which have been reduced by approximately 50%. The total effect of the b -tagging uncertainty on the expected limits is shown in Sec. ??, along with other uncertainties.

$t\bar{t}$ MC UNCERTAINTY

In addition to the $t\bar{t}$ fit uncertainties, following the recommendations, extra $t\bar{t}$ MC samples are used with different variations: Hadronization, Fragmentation, Matrix Element and Additional Radiation. The top quark mass variations are also considered. The MC samples used are:

```

mc15_13TeV.410001.PowhegPythiaEvtGen_P2012radHi_ttbar_hdamp345_down_nonallhad.merge.DAOD_EXOT8.e3783_s2608_r7725_r7676_p2949
mc15_13TeV.410002.PowhegPythiaEvtGen_P2012radLo_ttbar_hdamp172_up_nonallhad.merge.DAOD_EXOT8.e3783_s2608_r7725_r7676_p2949
mc15_13TeV.410003.aMcAtNloHerwigppEvtGen_ttbar_nonallhad.merge.DAOD_EXOT8.e4441_s2726_r7772_r7676_p2949
mc15_13TeV.410004.PowhegHerwigppEvtGen_UEEE5_ttbar_hdamp172p5_nonallhad.merge.DAOD_EXOT8.e3836_a766_a821_r7676_p2949
mc15_13TeV.410008.aMcAtNloHerwigppEvtGen_ttbar_allhad.merge.DAOD_EXOT8.e3964_s2726_r7772_r7676_p2949
mc15_13TeV.410022.Sherpa_CT10_ttbar_SingleLeptonP_MEPS_NLO.merge.DAOD_EXOT8.e3957_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410022.Sherpa_CT10_ttbar_SingleLeptonP_MEPS_NLO.merge.DAOD_EXOT8.e3959_a766_a818_r7676_p2949
mc15_13TeV.410023.Sherpa_CT10_ttbar_SingleLeptonM_MEPS_NLO.merge.DAOD_EXOT8.e3957_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410023.Sherpa_CT10_ttbar_SingleLeptonM_MEPS_NLO.merge.DAOD_EXOT8.e3959_a766_a818_r7676_p2949
mc15_13TeV.410024.Sherpa_CT10_ttbar_AllHadron_MEPS_NLO.merge.DAOD_EXOT8.e3957_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410024.Sherpa_CT10_ttbar_AllHadron_MEPS_NLO.merge.DAOD_EXOT8.e3959_a766_a818_r7676_p2949
mc15_13TeV.410037.PowhegPythiaEvtGen_P2012_ttbar_hdamp170_nonallhad.merge.DAOD_EXOT8.e4529_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410038.PowhegPythiaEvtGen_P2012_ttbar_hdamp171p5_nonallhad.merge.DAOD_EXOT8.e4529_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410039.PowhegPythiaEvtGen_P2012_ttbar_hdamp173p5_nonallhad.merge.DAOD_EXOT8.e4529_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410040.PowhegPythiaEvtGen_P2012_ttbar_hdamp175_nonallhad.merge.DAOD_EXOT8.e4529_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410041.PowhegPythiaEvtGen_P2012_ttbar_hdamp177p5_nonallhad.merge.DAOD_EXOT8.e4529_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410042.PowhegPythiaEvtGen_P2012_ttbar_hdamp170_allhad.merge.DAOD_EXOT8.e4510_s2608_s2183_r7725_r7676_p2949

```

```

mc15_13TeV.410043.PowhegPythiaEvtGen_P2012_ttbar_hdamp171p5_allhad.merge.DAOD_EXOT8.e4510_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410044.PowhegPythiaEvtGen_P2012_ttbar_hdamp173p5_allhad.merge.DAOD_EXOT8.e4510_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410045.PowhegPythiaEvtGen_P2012_ttbar_hdamp175_allhad.merge.DAOD_EXOT8.e4510_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410046.PowhegPythiaEvtGen_P2012_ttbar_hdamp177p5_allhad.merge.DAOD_EXOT8.e4510_s2608_s2183_r7725_r7676_p2949
mc15_13TeV.410161.PowhegPythiaEvtGen_P2012radHi_ttbar_hdamp345_down_allhad.merge.DAOD_EXOT8.e4837_s2726_r7772_r7676_p2949
mc15_13TeV.410162.PowhegPythiaEvtGen_P2012radLo_ttbar_hdamp172p5_up_allhad.merge.DAOD_EXOT8.e4837_s2726_r7772_r7676_p2949
mc15_13TeV.410163.PowhegHerwigppEvtGen_UEEE5_ttbar_hdamp172p5_allhad.merge.DAOD_EXOT8.e4836_s2726_r7772_r7676_p2949

```

These $t\bar{t}$ samples are used to replace the normal had and nonhad MCs, stiched with Mtt slices samples, and the variation in the $t\bar{t}$ yield and background predictions are considered. The total varation on $t\bar{t}$ yield in signal region is between 12-58%. These are shown in section 8.0.4. UPDATE.

8.0.2 THEORETICAL UNCERTAINTIES

THIS SECTION IS STILL TO BE UPDATED

The effect of the following theoretical uncertainties on the signal acceptance and efficiency have been evaluated: uncertainties in the parton density functions; uncertainties due to missing higher order terms in the matrix element; uncertainties in the modelling of the underlying event (including multi-parton interactions), of hadronic showers and of initial and final state radiation. All of the uncertainties are evaluated for the Bulk RS $c = 1.0$ benchmark model. These have been evaluated in exactly the same way as for the resolved analysis.

To evaluate the potential effect of missing higher order terms in the matrix element, the renormalisation and factorisation scales used in the signal generation were varied coherently by factors of $0.5\times$ and $2\times$. The shift induced by doubling the scales is consistent with zero across the full mass range, however halving the scales results in a constant -1% shift.

Uncertainties due to modelling of the parton shower and the underlying event (including multi-parton interactions) are evaluated by varying the parameters of the A14 Pythia 8 tune used in the simulation of the graviton signal samples. Ten “eigentune” variations are produced for each mass point. These variations, along the principal directions of the covariance matrix at the A14 tune minimum as described in Ref.⁷, provide good coverage of the experimental and modelling uncertainties implicit in the tuning. Variation 1 is influenced mainly by underlying event effects; variation 2 by jet structure; and variations 3a-3c are mainly influenced by extra jet production.

A14 tune variations 1 does not have a significant effect on the signal acceptance. Variation 2 has an asymmetric impact on the acceptance that is quadratically dependent on the resonance mass, reaching +4% and -8% at $m_{G_{KK}^*} = 3$ TeV. Variation 3a is consistent with zero shift for all masses when the downward variation is applied, while the upward variation leads to a signal acceptance uncertainty which is quadratic in $m_{G_{KK}^*}$ reaching -8% for $m_{G_{KK}^*} = 3$ TeV. Variation 3b is again asymmetric with quadratic behaviour, with the upward variation reaching +2% and the downward variation -6% at $m_{G_{KK}^*} = 3$ TeV. Variation 3c is symmetric, with the corresponding acceptance uncertainty $\pm 2\%$ for all masses. These uncertainties will be considered in the statistical analysis.

The PDF uncertainty is evaluated using the 100 NNPDF 2.3 LO replicas, as in the resolved analysis. In this case, the uncertainty in acceptance due to PDF uncertainties is grows to $\pm 3\%$ at the very highest masses considered for the boosted analysis.

8.0.3 BACKGROUND PREDICTION UNCERTAINTY

w

A statistical uncertainty on the value of μ_{multijet} for the 4b (3b, 2bs) Signal Region was determined from the fitting procedure described in Section 7.1.5.

The statistical uncertainty of the $t\bar{t}$ normalization is accounted for through the uncertainties on $\alpha_{t\bar{t}}$ from the fit to data, as described in Section 7.1.5. The statistical uncertainty of 69% on $\alpha_{t\bar{t}}$ are 79% anti-correlated to the value of μ_{multijet} found in the fitting procedure in the 4 b -tag region, the uncertainties of 9% on $\alpha_{t\bar{t}}$ are 76% anti-correlated to the value of μ_{multijet} found in the fitting procedure in the 3 b -tag region, and the uncertainties of 2.6% on $\alpha_{t\bar{t}}$ are 75% anti-correlated to the value of μ_{multijet} found in the fitting procedure in the 2 bs -tag region.

The background systematic uncertainties in the signal region are divided into the following components:

- Non-closure uncertainty on μ_{multijet} found by comparing the value derived from the sideband to the control region normalization.
- Effects on the QCD prediction from variations of the SideBand and Control Region Definitions
- The impact of the shape uncertainty of the $t\bar{t}$ distribution in the 4 b signal region.
- The impact of the shape uncertainty of the 1/2 b QCD distribution derived in the control region.
- The impact of the smoothing function fit range and function choice on the QCD prediction

The $t\bar{t}$ normalization uncertainty on $\alpha_{t\bar{t}}$ is derived in the fit to data as described in Section 7.1.5. This has a negligible impact on the signal sensitivity, but is still propagated as the uncertainty in the $t\bar{t}$ normalization in the signal region.

NON-CLOSURE UNCERTAINTY ON μ_{multijet} DETERMINED IN THE CONTROL REGION

A further uncertainty is derived by comparing the value of μ_{multijet} to the overall difference between predicted to observed events in the control region. While the total predicted background

(showing stat error only) of $4b$: 76.7 ± 5.4 vs obs 81.0 , $3b$: 1565.6 ± 18.1 vs obs 1553.0 , $2bs$: 8332.4 ± 38.8 vs obs 8486.0 , the number events agrees with the total data in the control region within statistical error, we consider an added systematic on the background prediction normalization, taken as the maximum between either the difference between the central value of the prediction to the observed number of events (4.3 events, or 5% , for $4b$; 13 events, or 1% , for $3b$; 154 events, or 2% , for $2bs$) or the statistical uncertainty of the observed $4b$ ($3b$) data in the CR (11.1% for $4b$; 2.5% for $3b$; and 1.1% for $2bs$). For the detailed numbers, please refer to section ??.

Although we have derived our non-closure uncertainty on μ_{QCD} from comparison between data and prediction in the control region, we need to test how this number is sensitive to our choice of control region (CR) and sideband region (SB). In addition, we also want to check how our background prediction in signal region is sensitive to the choice of control region and sideband region. All these tests are done on the control/sideband regions after the full reweighting procedure as described in 7.1.6, while applying the nominal reweighting values.

Besides the nominal control region as described above, we design three additional control regions, as illustrated in Figure ??, ??, ??, ??, ??, ??, ??:

- Low-mass CR: the center position of the circle that defines nominal CR is moved down by 3 GeV, in both leading and sub-leading large-jet mass.
- High-mass CR: the center position of the circle that defines nominal CR is moved up by 3 GeV, in both leading and sub-leading large-R jet mass.
- Signal-depletion (Small) CR: the X_{hh} cut that defines signal region is increased to 2.0 from 1.6 . This variation only affect CR, while SR remains unchanged (i.e. signal region is still defined by $X_{hh} < 1.6$, while CR is defined as $X_{hh} > 2.0$ and $R_{hh} < 33$).
- High-mass SB: The signal region and control region remain unchanged, the center position of the circle that defines nominal SB is moved up by 3 GeV in both leading and sub-leading large-R jet mass.

- Low-mass SB: The signal region and control region remain unchanged, the center position of the circle that defines nominal SB is moved up by 3 GeV in both leading and sub-leading large-R jet mass.
- Large SB: The signal region and control region remain unchanged, while the SB is $33 < R_{hh}$ and $R_{hh}^{\text{high}} < 61$. μ_{QCD} will change.
- Small SB: The signal region and control region remain unchanged, while the SB is $33 < R_{hh}$ and $R_{hh}^{\text{high}} < 55$. μ_{QCD} will change.

The results are summarized in Table 8.1, 8.2 and 8.3, while the details are presented in the Appendix ???. Based on all the variations, a 2.8% normalization uncertainty is assigned to $2b$ s region, 4.2% to $3b$ region (which is the statistical uncertainty), and a 12.2% normalization uncertainty is assigned to $4b$ region.

CR Variations Four Tag	Data	Prediction	(Predict - Data)/Data
Nominal	81.0 ± 9.0	76.77 ± 5.43	-5.22 % ± 17.23 %
CR High	76.0 ± 8.72	71.12 ± 5.41	-6.43 % ± 17.85 %
CR Low	91.0 ± 9.54	79.87 ± 5.45	-12.2 % ± 15.19 %
CR Small	58.0 ± 7.62	55.96 ± 5.35	-3.52 % ± 21.89 %
SB Large	81.0 ± 9.0	74.71 ± 5.4	-7.76 % ± 16.91 %
SB Small	81.0 ± 9.0	74.15 ± 5.38	-8.45 % ± 16.81 %
SB High	81.0 ± 9.0	78.72 ± 5.46	-2.82 % ± 17.54 %
SB Low	81.0 ± 9.0	76.51 ± 5.38	-5.54 % ± 17.14 %

Table 8.1: Agreement between data and prediction in $4b$ tag CR. Showing stat uncertainty only.

CR Variations ThreeTag	Data	Prediction	(Predict - Data)/Data
Nominal	1553.0 ± 39.41	1587.04 ± 21.4	$2.19\% \pm 3.97\%$
CR High	1461.0 ± 38.22	1473.89 ± 20.77	$0.88\% \pm 4.06\%$
CR Low	1628.0 ± 40.35	1697.38 ± 21.75	$4.26\% \pm 3.92\%$
CR Small	1134.0 ± 33.67	1127.34 ± 17.66	$-0.59\% \pm 4.51\%$
SB Large	1553.0 ± 39.41	1574.23 ± 21.47	$1.37\% \pm 3.95\%$
SB Small	1553.0 ± 39.41	1601.44 ± 21.64	$3.12\% \pm 4.01\%$
SB High	1553.0 ± 39.41	1602.74 ± 21.48	$3.2\% \pm 4.0\%$
SB Low	1553.0 ± 39.41	1576.56 ± 21.5	$1.52\% \pm 3.96\%$

Table 8.2: Agreement between data and prediction in 3b tag CR. Showing stat uncertainty only.

CR Variations TwoTag split	Data	Prediction	(Predict - Data)/Data
Nominal	8486.0 ± 92.12	8332.97 ± 38.84	$-1.8\% \pm 1.52\%$
CR High	8174.0 ± 90.41	7937.59 ± 39.61	$-2.89\% \pm 1.56\%$
CR Low	8907.0 ± 94.38	8800.86 ± 39.51	$-1.19\% \pm 1.49\%$
CR Small	5999.0 ± 77.45	5873.52 ± 32.31	$-2.09\% \pm 1.8\%$
SB Large	8486.0 ± 92.12	8341.7 ± 38.44	$-1.7\% \pm 1.52\%$
SB Small	8486.0 ± 92.12	8333.25 ± 39.12	$-1.8\% \pm 1.53\%$
SB High	8486.0 ± 92.12	8378.14 ± 38.45	$-1.27\% \pm 1.52\%$
SB Low	8486.0 ± 92.12	8356.86 ± 39.06	$-1.52\% \pm 1.53\%$

Table 8.3: Agreement between data and prediction in 2bs tag CR. Showing stat uncertainty only.

VALIDATION OF BACKGROUND ESTIMATION FROM LOW MASS AND HIGH MASS SIGNAL REGION

Another check is the so-called "low mass signal region rehearsal" (or ZZ region) and "high mass signal region rehearsal" (or TT region). Instead of a signal region around di-Higgs mass region on leading-subleading large-R jet mass 2D plane, we redefine a separate lower mass (ZZ) and higher mass (TT) signal region:

$$X_{ZZ} = \sqrt{\left(\frac{m(J_1) - 103 \text{ GeV}}{\text{o.}1m(J_1)}\right)^2 + \left(\frac{m(J_2) - 96 \text{ GeV}}{\text{o.}1m(J_2)}\right)^2} < 1.6 \quad (8.1)$$

$$X_{TT} = \sqrt{\left(\frac{m(J_1) - 164 \text{ GeV}}{\text{o.}1m(J_1)}\right)^2 + \left(\frac{m(J_2) - 155 \text{ GeV}}{\text{o.}1m(J_2)}\right)^2} < 1.6 \quad (8.2)$$

which is also illustrated in Figure 8.1. The analysis is repeated, using the same definition of Sideband and Control region as nominal (but with events contained in ZZ signal region excluded) for normalization fit. Then the low mass signal region is unblinded. This helps to validate the background estimation strategy, and the stability for other similar analysis.

The summary of background estimation for ZZ signal region can be found in Table 8.4, 8.5 and 8.6. The difference between data and prediction in ZZ signal region is summarized in Table 8.7 for all the regions. The discrepancy between data and prediction is either covered by statistical uncertainty of data or comparable with data statistical uncertainty in 4b, 3b and 2bs ZZ SR respectively. We further check the kinematic distribution between data and prediction in ZZ SR, as shown in Figure 8.2. The data agrees with prediction well in general, though a few bins might not agree perfectly.

The summary of background estimation for TT signal region can be found in Table 8.8, 8.9 and 8.10. The difference between data and prediction in TT signal region is summarized in Table 8.11 for

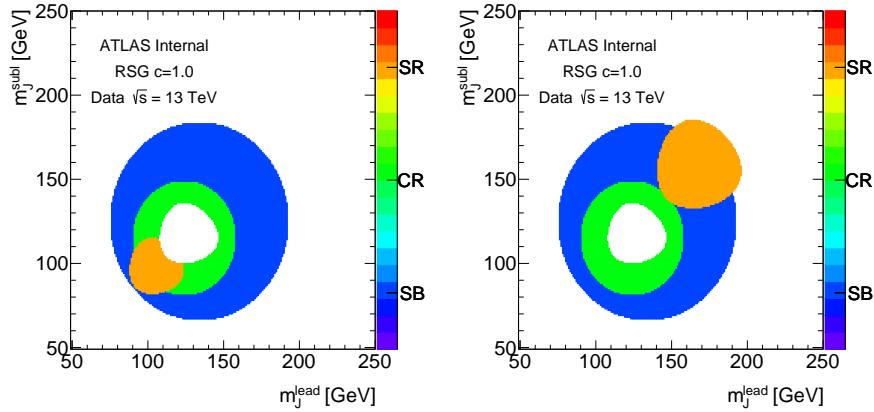


Figure 8.1: Illustration of ZZ (left) and TT (right) signal region as shown in the orange shaded region. Control region shown in green, and Sideband region in blue. The white circle in the midde is the real Signal region, and it is blinded.

all the regions. The discrepancy between data and prediction is either covered by statistical uncertainty of data or comparable with data statistical uncertainty in $4b$, $3b$ and $2bs$ TT SR respectively. We further check the kinematic distribution between data and prediction in TT SR, as shown in Figure 8.3. The data agrees with prediction well in general, though a few bins might not agree perfectly.

Based on all the variation tests done above, we think there is no need to introduce extra uncertainty on non-closure systematics since most of the data/prediction disagreements are well covered by the data statistical uncertainty.

FourTag	Sideband	Control	Signal
QCD Est	166.65 ± 2.88	45.83 ± 1.51	27.37 ± 1.16
$t\bar{t}$ Est.	27.52 ± 0.25	6.31 ± 0.14	0 ± 0
$Z + jets$	0 ± 0	6.18 ± 5.12	0 ± 0
Total Bkg Est	194.17 ± 2.89	58.32 ± 5.34	27.37 ± 1.16
Data	194.0 ± 13.93	54.0 ± 7.35	37.0 ± 6.08
$c = 1.0, m = 1.0 TeV$	2.45 ± 0.098	4.47 ± 0.13	0.99 ± 0.063
$c = 1.0, m = 2.0 TeV$	0.032 ± 0.0015	0.075 ± 0.0022	0.028 ± 0.0014
$c = 1.0, m = 3.0 TeV$	$0.00029 \pm 3.5e-05$	$0.00064 \pm 5e-05$	$0.0002 \pm 2.7e-05$

Table 8.4: Background prediction in SR/CR/SB for ZZ SR in $4b$ -tag region. Uncertainties are stat only.

ThreeTag	Sideband	Control	Signal
QCD Est	3344.46 ± 26.85	998.41 ± 14.63	637.78 ± 11.87
$t\bar{t}$ Est.	826.66 ± 25.11	136.58 ± 10.23	30.07 ± 1.24
$Z + jets$	32.49 ± 11.34	8.22 ± 5.29	3.3 ± 2.0
Total Bkg Est	4203.61 ± 38.47	1143.2 ± 18.62	671.15 ± 12.11
Data	4203.0 ± 64.83	1108.0 ± 33.29	645.0 ± 25.4
$c = 1.0, m = 1.0 TeV$	7.56 ± 0.18	9.84 ± 0.2	3.05 ± 0.11
$c = 1.0, m = 2.0 TeV$	0.15 ± 0.0033	0.27 ± 0.0046	0.12 ± 0.003
$c = 1.0, m = 3.0 TeV$	0.0034 ± 0.00012	0.0056 ± 0.00016	$0.0021 \pm 9.5e-05$

Table 8.5: Background prediction in SR/CR/SB for ZZ SR in $3b$ -tag region. Uncertainties are stat only.

TwoTag split	Sideband	Control	Signal
QCD Est	16387.44 ± 37.6	4827.76 ± 19.86	3026.83 ± 15.61
$t\bar{t}$ Est.	7671.95 ± 69.14	1229.96 ± 26.54	332.29 ± 13.66
$Z + jets$	44.37 ± 13.23	13.34 ± 6.6	36.47 ± 12.88
Total Bkg Est	24103.77 ± 79.8	6071.07 ± 33.8	3395.59 ± 24.42
Data	24104.0 ± 155.25	6261.0 ± 79.13	3258.0 ± 57.08
$c = 1.0, m = 1.0 TeV$	4.57 ± 0.14	4.65 ± 0.14	1.91 ± 0.089
$c = 1.0, m = 2.0 TeV$	0.16 ± 0.0038	0.26 ± 0.0047	0.12 ± 0.0032
$c = 1.0, m = 3.0 TeV$	0.012 ± 0.00024	0.019 ± 0.00029	0.0085 ± 0.00019

Table 8.6: Background prediction in SR/CR/SB for ZZ SR in $2bs$ -tag region. Uncertainties are stat only.

ZZ Signal Region	Data	Prediction	(Predict - Data)/Data
FourTag	37.0 ± 6.08	27.37 ± 1.16	$-26.0\% \pm 15.3\%$
ThreeTag	645.0 ± 25.4	671.15 ± 12.11	$4.05\% \pm 5.97\%$
TwoTag split	3258.0 ± 57.08	3395.59 ± 24.42	$4.22\% \pm 2.58\%$

Table 8.7: Agreement between data and prediction in ZZ SR in $4b$, $3b$ and $2bs$ regions.

FourTag	Sideband	Control	Signal
QCD Est	152.28 ± 2.72	63.47 ± 1.77	28.6 ± 1.21
$t\bar{t}$ Est.	19.86 ± 0.22	7.45 ± 0.15	15.02 ± 0.2
$Z + jets$	0 ± 0	6.18 ± 5.12	0 ± 0
Total Bkg Est	172.14 ± 2.73	77.1 ± 5.42	43.62 ± 1.23
Data	172.0 ± 13.11	81.0 ± 9.0	46.0 ± 6.78
$c = 1.0, m = 1.0 TeV$	2.38 ± 0.097	5.4 ± 0.15	0.15 ± 0.024
$c = 1.0, m = 2.0 TeV$	0.033 ± 0.0015	0.1 ± 0.0026	0.0011 ± 0.00027
$c = 1.0, m = 3.0 TeV$	$0.00031 \pm 3.6e-05$	$0.0008 \pm 5.6e-05$	$1.5e-05 \pm 7.7e-06$

Table 8.8: Background prediction in SR/CR/SB for TT SR in $4b$ -tag region. Uncertainties are stat only.

ThreeTag	Sideband	Control	Signal
QCD Est	3106.11 ± 25.79	1427.41 ± 17.53	570.01 ± 11.6
$t\bar{t}$ Est.	495.21 ± 18.75	148.55 ± 10.21	406.57 ± 5.42
$Z + jets$	32.5 ± 11.34	11.21 ± 5.65	0.3 ± 0.3
Total Bkg Est	3633.82 ± 33.85	1587.17 ± 21.05	976.88 ± 12.81
Data	3633.0 ± 60.27	1553.0 ± 39.41	1017.0 ± 31.89
$c = 1.0, m = 1.0 TeV$	7.57 ± 0.18	12.58 ± 0.23	0.32 ± 0.037
$c = 1.0, m = 2.0 TeV$	0.15 ± 0.0034	0.38 ± 0.0054	0.0047 ± 0.0006
$c = 1.0, m = 3.0 TeV$	0.0034 ± 0.00012	0.0075 ± 0.00018	$0.00023 \pm 3.3e-05$

Table 8.9: Background prediction in SR/CR/SB for TT SR in $3b$ -tag region. Uncertainties are stat only.

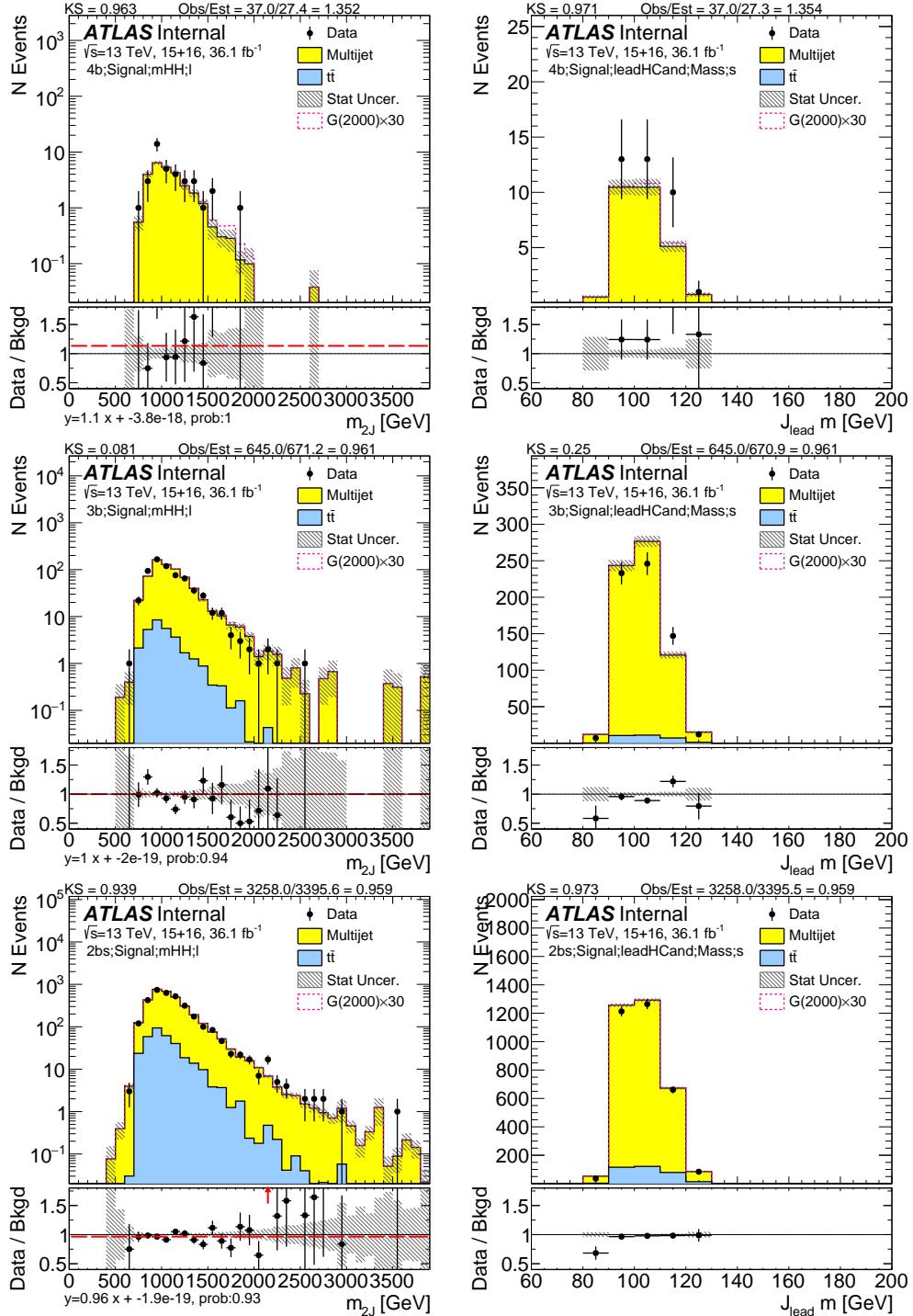


Figure 8.2: ZZ signal region distribution of di-jet mass (left column) and leading large-R jet mass (right column) in low mass signal, for $4b$ (top row), $3b$ (middle row) and $2b$ split (bottom row). The plots are with only statistical uncertainty.

TwoTag split	Sideband	Control	Signal
QCD Est	14980.05 \pm 35.33	6803.06 \pm 23.41	2817.92 \pm 16.54
$t\bar{t}$ Est.	5170.92 \pm 56.22	1468.85 \pm 28.93	3628.91 \pm 48.42
$Z + jets$	61.34 \pm 16.04	26.44 \pm 10.08	6.4 \pm 5.05
Total Bkg Est	20212.31 \pm 68.31	8298.34 \pm 38.56	6453.23 \pm 51.41
Data	20212.0 \pm 142.17	8486.0 \pm 92.12	6446.0 \pm 80.29
$c = 1.0, m = 1.0 \text{ TeV}$	4.59 \pm 0.14	6.33 \pm 0.16	0.24 \pm 0.033
$c = 1.0, m = 2.0 \text{ TeV}$	0.17 \pm 0.0039	0.36 \pm 0.0056	0.0066 \pm 0.00077
$c = 1.0, m = 3.0 \text{ TeV}$	0.012 \pm 0.00024	0.027 \pm 0.00034	0.00089 \pm 6.7e-05

Table 8.10: Background prediction in SR/CR/SB for TT SR in $2bs$ -tag region. Uncertainties are stat only.

TT Signal Region	Data	Prediction	(Predict - Data)/Data
FourTag	46.0 \pm 6.78	43.62 \pm 1.23	-5.18 % \pm 16.66 %
ThreeTag	1017.0 \pm 31.89	976.88 \pm 12.81	-3.95 % \pm 4.27 %
TwoTag split	6446.0 \pm 80.29	6453.23 \pm 51.41	0.11 % \pm 2.04 %

Table 8.11: Agreement between data and prediction in TT SR in $4b$, $3b$ and $2bs$ regions.

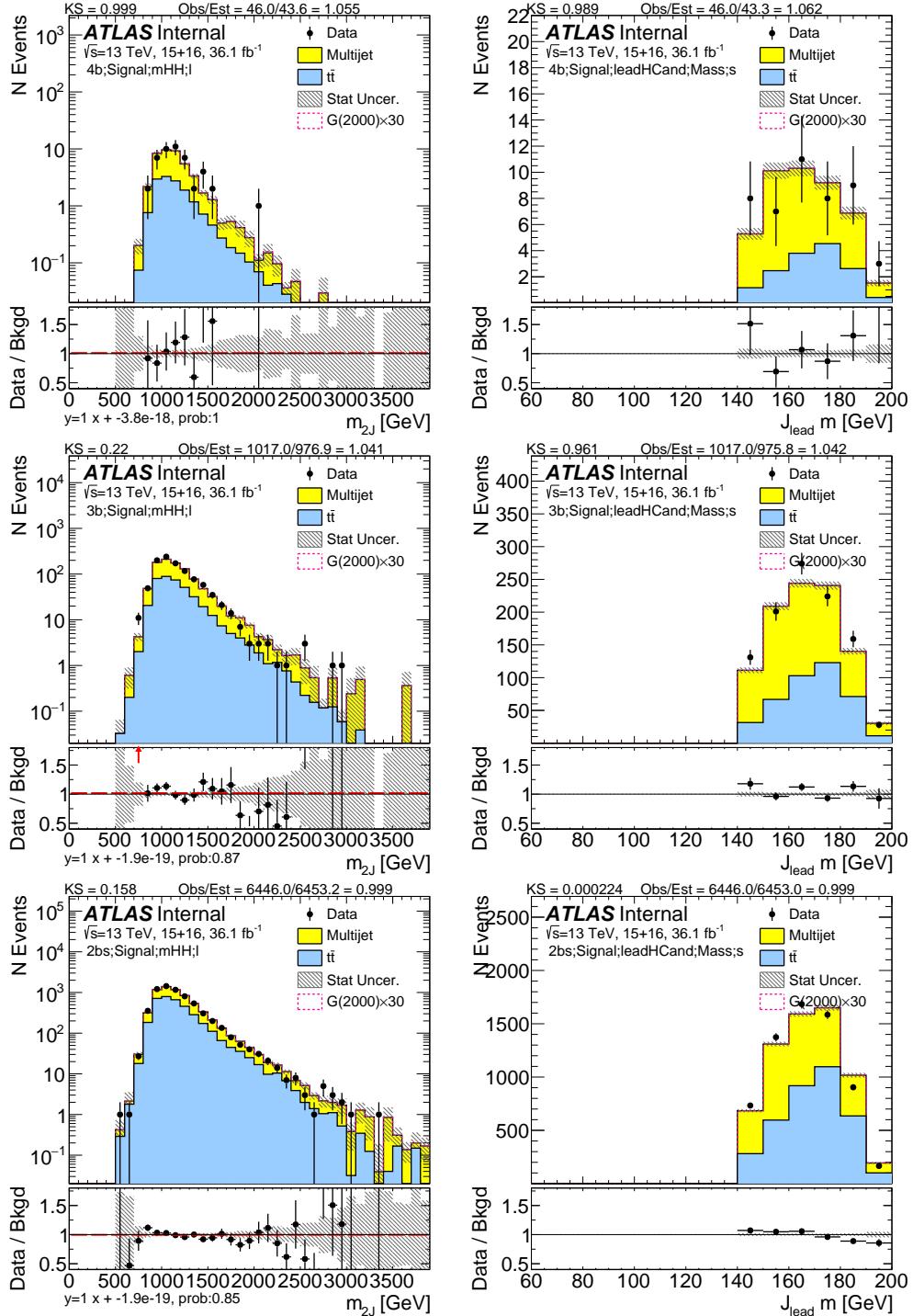


Figure 8.3: TT signal region distribution of di-jet mass (left column) and leading large-R jet mass (right column) in low mass signal, for $4b$ (top row), $3b$ (middle row) and $2b$ split (bottom row). The plots are with only statistical uncertainty.

UNCERTAINTY ON THE SHAPE OF THE $t\bar{t}$ JET MASS IN THE $4/3b$ SIGNAL REGION

Because the $4/3b$ $t\bar{t}$ MJJ distribution is extremely statistically limited, the $4/3b$ shape is used to predict the final $t\bar{t}$ background shape in the $2bs$ signal region. In order to estimate the possible shape uncertainty, the $2bs$ and $3b$ sideband shapes are compared in Figure 8.4 (after being normalized to the same area). The $2bs$ is used as there are not sufficient $4b$ statistics to assess the comparison quality. In order to avoid large statistical uncertainties, the distributions of the $3b$ and $2b$ are smoothed. The ratio of the two smoothed distributions is taken as the shape systematic. We then use this function to apply a bin-by-bin scaling of the $t\bar{t}$ background prediction in the signal region, maintaining the same normalization given by nominal $t\bar{t}$ normalization prediction.

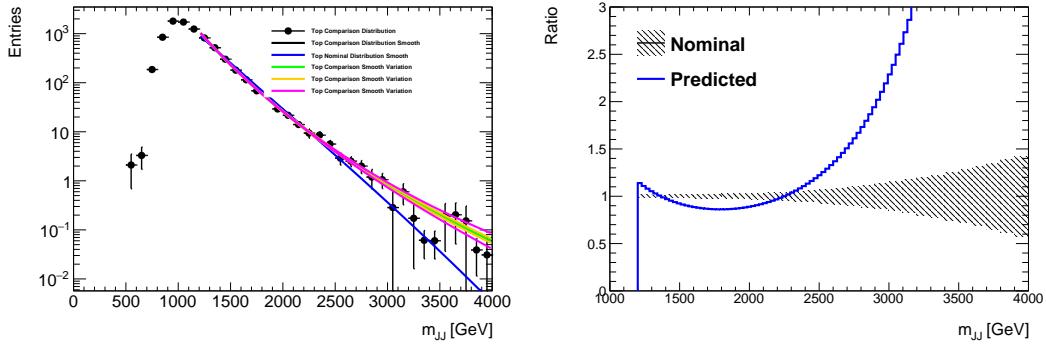


Figure 8.4: (left) Shape of the $t\bar{t}$ di-large – R -jet mass in the sideband region, comparing the $3b$ shape with that of the $2b$, in order to assess the systematic effect of additional b -tags changing the dijet mass distribution. The m_{JJ} distribution is shown on the left, and the ratio of $3b$ to $2b$ distributions on the right.

UNCERTAINTY ON THE SHAPE OF THE $1/2b$ QCD DISTRIBUTION IN THE SIGNAL REGION

As shown in Figures 7.32 and 7.36, the shape distribution of the total predicted background using the scaled $1/2b$ QCD sample was found to be in good agreement with the $4b$, $3b$, and $2bs$ data in the control region. However due to the low statistics in the data in the control region, the comparison is performed by first smoothing the $1/2b$, and the $4/3/2bs$ distributions. The ratio of the smoothed $1/2b$ distributions to that of the smoothed $4/3/2bs$ distributions is taken as the shape systematic. This function is then used to apply a bin-by-bin scaling of the QCD background prediction in the signal region, maintaining the same normalization given by μ_{multijet} . The CR distributions and the smoothing fit ratios can be found in Figure 8.5. This systematics is further split into two parts: one below 2000 GeV and the other above 2000 GeV, to ensure the low and high mass shape variation post-fit pulls can vary independently. It should be noted, that this uncertainty is used for both the dijet mass, and the scaled dijet mass distribution, and the correction to scaling is expected to be small relative to the dijet mass.

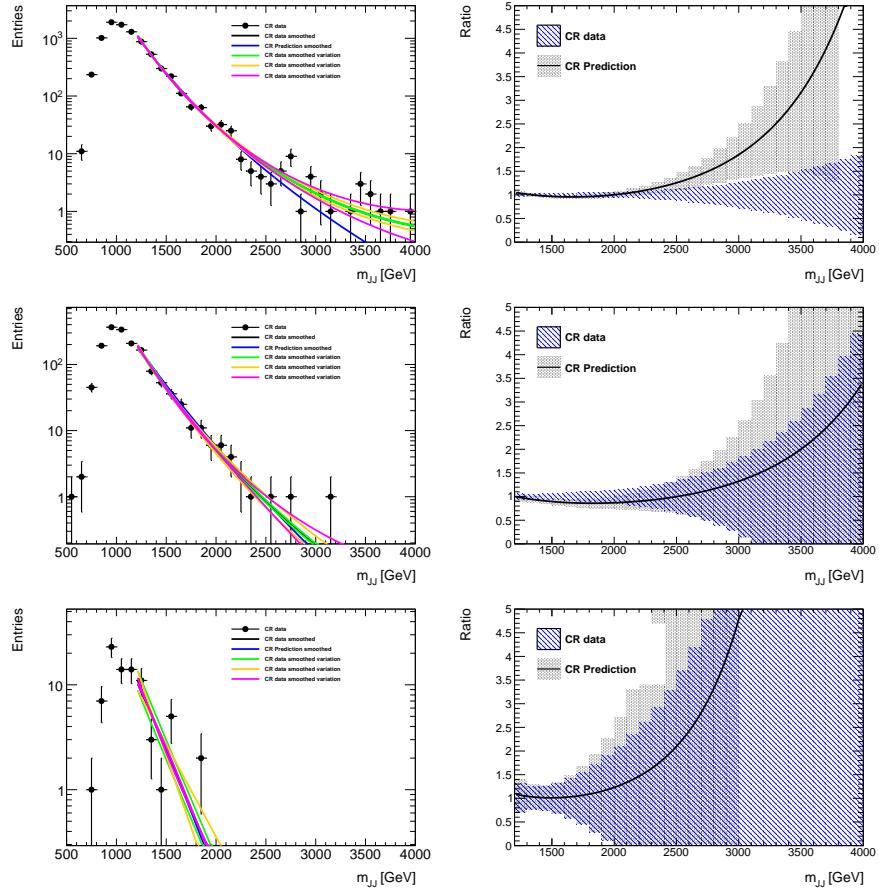


Figure 8.5: Dijet mass distribution in the CR along with the prediction (left) and the ratio of the prediction to the CR distribution (right) for the $2b$ s (top) $3b$ (middle) and $4b$ (bottom) samples. Ratios are from the smoothed distributions, the data uncertainty band contains the smoothing parameter variations, and the prediction uncertainty band also contains smoothing parameter variations.

UNCERTAINTY ON QCD SMOOTHING FUNCTION IN THE SIGNAL REGION

The MJ8 function has been used to fit the QCD background prediction in order to smooth the distribution and provide non-zero background estimates up to dijet masses beyond which we have $1/2b$ statistics. While this distribution is observed to fit the $1/2b$ data well, it does not have a concrete physical motivation, and in principle the high mass tail of the distribution could be larger than predicted by an exponential. Two checks are performed, changing the boundaries where the fit is performed, and changing the fit function.

To test the impact of the region in which the fit is performed, we varied the upper bound on the dijet fit region to be each of the values $\{2800, 3000, 3200\}$ GeV and the starting value between $\{1200, 1300, 1400\}$ GeV. The ratio of the fits for each upper bound, to that of the nominal (1200–3000 GeV) can be found in Figure 8.6, along with a hash band showing the statistical uncertainty of the nominal fit. The maximum deviation from the nominal fit, per bin, is taken as the shape systematic uncertainty. This is estimated separately for $2b$, $3b$, and $4b$ samples.

It should be noted that fits in which the fit χ^2 probability was less than 0.001%, or in which the fit integrals between 1500–2000 GeV, 2000–2500 GeV, or >2500 GeV were not in agreement with the original $1/2b$ distribution within a factor of 2 or 0.5, were not used to estimate the uncertainty. The aforementioned checks ensure that we do not use poor fits of the $1/2b$ distribution to estimate the uncertainty.

As a second test, we fit the $1/2b$ QCD prediction with a variety of other distributions which can show both power law behavior in the bulk of the distribution as well as longer tails. The set of additional functions examined (labelled MJ1–MJ7) can be found in Table 8.12, where $x = m_{JJ}/\sqrt{s}$.

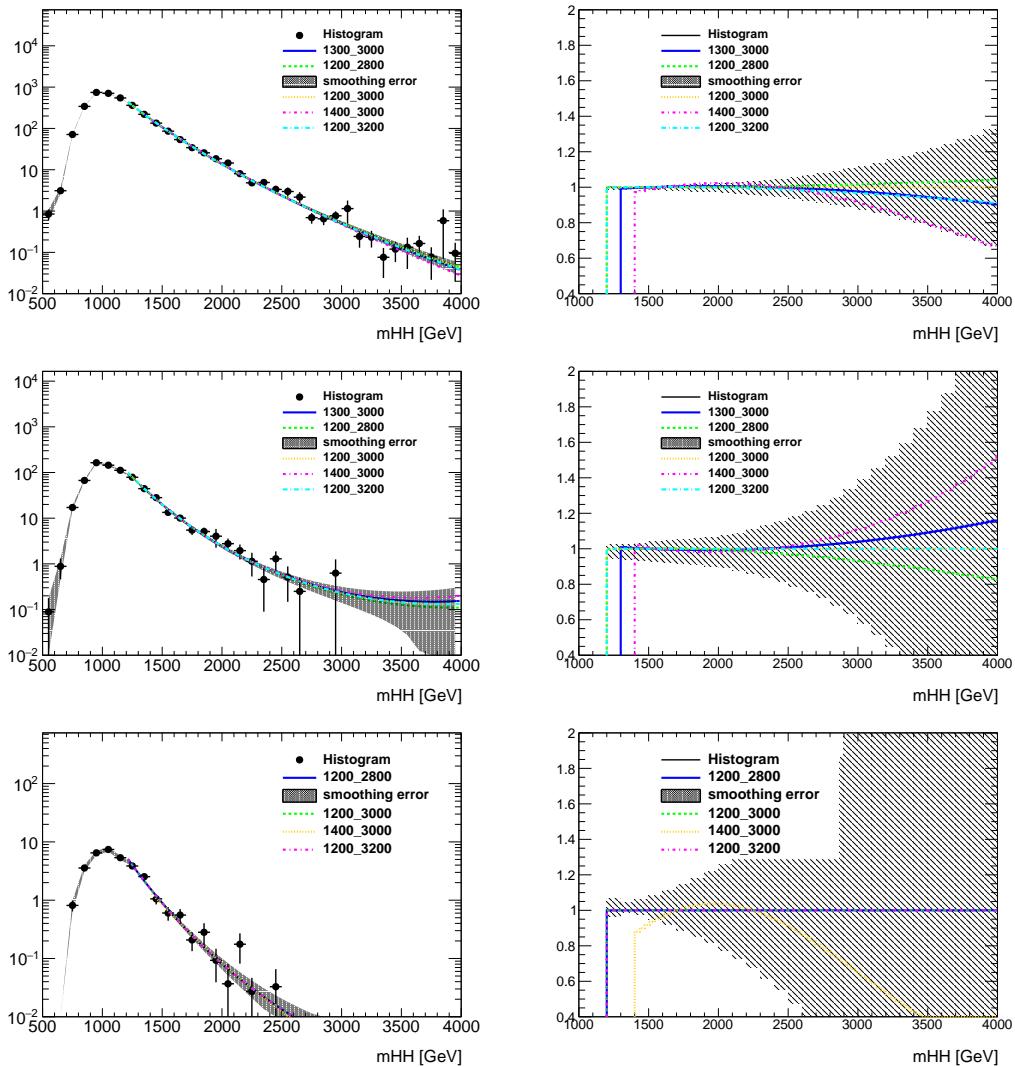


Figure 8.6: Dijet mass distribution SR prediction fit with several fit ranges (left) and the ratio of nominal to fits with different fit ranges (right) for the 2b (top) 3b (middle) and 4b (bottom) samples.

Name	Functional Form
MJ1 (Dijet)	$f_1(x) = p_o(1-x)^{p_1}x^{p_2}$
MJ2	$f_2(x) = p_o(1-x)^{p_1}e^{p_2 x^2}$
MJ3	$f_3(x) = p_o(1-x)^{p_1}x^{p_2 x}$
MJ4	$f_4(x) = p_o(1-x)^{p_1}x^{p_2 \ln x}$
MJ5	$f_5(x) = p_o(1-x)^{p_1}(1+x)^{p_2 x}$
MJ6	$f_6(x) = p_o(1-x)^{p_1}(1+x)^{p_2 \ln x}$
MJ7	$f_7(x) = \frac{p_o}{x}(1-x)^{p_1-p_2 \ln x}$
MJ8	$f_8(x) = \frac{p_o}{x^2}(1-x)^{p_1-p_2 \ln x}$

Table 8.12: Functions used to fit the QCD dijet mass distributions, where $x = m_{jj}/\sqrt{s}$.

Figure 8.7 shows the fits to the QCD prediction in the 4/3/2b signal regions, and the nominal dijet fit, as well as the ratios of the nominal fit to that of the additional functions. The maximum per bin deviation is taken as the shape systematic, separately for the 4/3/2b SRs.

As before, fits in which the fit χ^2 probability was less than 0.1%, or in which the fit integrals between 1500-2000 GeV, 2000-2500 GeV, or >2500 GeV were not in agreement with the original ob distribution within a factor of 2 or 0.5, were not used to estimate the uncertainty. The aforementioned checks ensure that we do not use poor fits of the ob distribution to estimate the uncertainty.

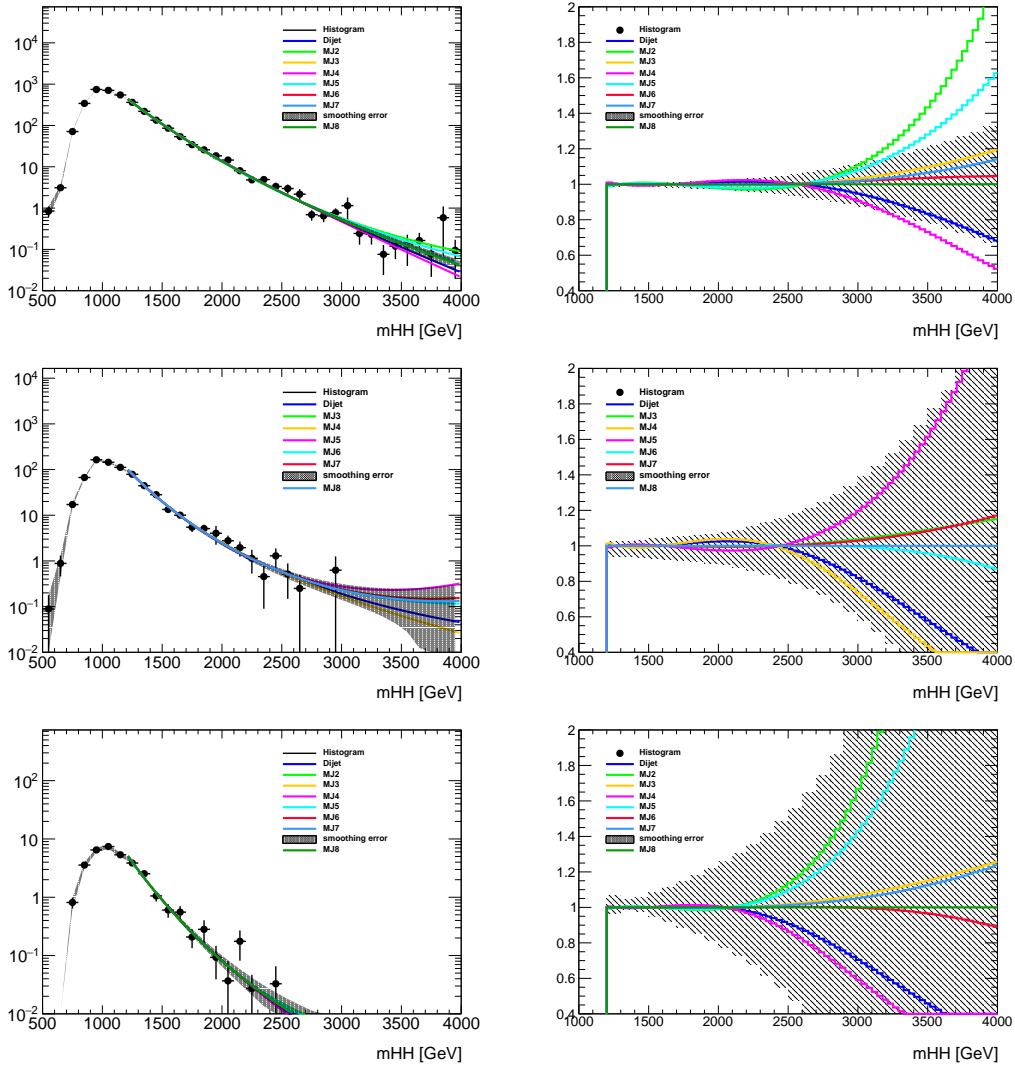


Figure 8.7: Dijet mass distribution SR prediction fit with several fit functions (left) and the ratio of nominal to fits with different fit functions (right) for the 2b (top) 3b (middle) and 4b (bottom) samples. The additional fit functions are from Table 8.12.

8.0.4 SUMMARY OF SYSTEMATICS

Table 8.13 shows the percent impact of systematics used in this analysis on the backgrounds yields and on the expected yields for RSG $c = 1.0$ signals in the $4b$ signal region. The correspondent values are shown for the $3b$ signal region in Table 8.14, and are shown for the $2bs$ signal region in Table 8.15.

A 3.2% luminosity uncertainty is also considered for the Z+jets background and the RSG signal predictions. The JER/JMR/JES/JMS/track jet b-tag scale factor uncertainties are applied to RSG and $t\bar{t}$ samples.

The “ $t\bar{t}$ SB shape” is the background normalization uncertainty due to the mis-modeling of the shape of the $t\bar{t}$ dijet mass distribution in the sideband region, taken from comparing the shape with $3/4$ b -tags to the shape with only a $2bs$ requirement. The “Background Normalization Fit” uncertainty comes from summing in quadrature the independent uncertainty components calculated from the correlated statistical errors of μ_{multijet} and $\alpha_{t\bar{t}}$. The “QCD Non-Closure in CR” systematic is derived as the maximum of (a) difference between the predicted and observed $4b/3b/2bs$ QCD yields in the control region, (b) the fractional change in SR predictions from varying the CR and SB definitions. Both options gave similar sized uncertainties, but the uncertainty from the CR/SB variations was found to be larger. All these uncertainties are summed in quadrature and shown in the “Bkg Est” row in the table.

The remaining systematics not listed in this table, as they do not impact the acceptance, include uncertainties on the shape of the QCD and $t\bar{t}$ backgrounds in the signal region, and the uncertainties from the smoothing / extrapolation procedure.

The size of the Monte Carlo modeling systematics on the signal predictions as a function of the signal mass can be found in Figure 8.8. The largest uncertainty in the $4b$ and $2b$ signal region is from b -tagging, followed by the JMR uncertainty. In the $3b$ signal region, although b -tagging systematics is still one of the largest uncertainty, it has been much reduced compared to $4b$ region, as discussed in Section 8.0.1. Then the jet mass scale and resolution are the largest uncertainties following b -tagging.

It should be noted that as numbers in this version of the note, the $t\bar{t}$ MC systematics are reduced, because of the larger stat $2bs$ $t\bar{t}$ to model $3b$ and $4b$ $t\bar{t}$ shape estimation. This is applied for JER, JMR, and $t\bar{t}MC$ generator variations.

FourTag	totalbkg	qcd	ttbar	RSG1 1000	RSG1 2000	RSG1 3000
JER	0.45	0.27	3.98	2.44	1.07	0.67
JMR	7.9	10.35	39.96	12.33	13.16	15.08
Top	1.65	2.8	26.17	-	-	-
JES/JMS	1.32	1.49	24.36	5.18	3.72	5.62
Bkg Est	15.64	18.15	67.83	-	-	-
b-tag SF	1.11	0.79	18.85	18.34	28.11	27.73
Total Sys	17.69	21.15	88.58	22.83	31.28	32.07
Stat	3.2	3.36	2.48	1.97	1.63	4.9
Estimated Events	34.6	32.91	1.68	10.07	0.25	0.0016

Table 8.13: Percent impact of the dominant systematics on the background acceptance and on the signal acceptance of RS $c = 1.0$ graviton predictions in the $4b$ signal region.

The final background prediction of MJJ along with total systematic uncertainties can be found in Figure 8.9, 8.10, and 8.11. The final background prediction of scaled MJJ along with total uncertainties can be found in Figure 8.12, 8.13, and 8.14.

ThreeTag	totalbkg	qcd	ttbar	RSG ₁ 1000	RSG ₁ 2000	RSG ₁ 3000
JER	1.38	3.52	17.5	1.41	0.93	1.05
JMR	1.35	4.26	24.39	14.3	12.33	15.53
Top	1.91	5.58	49.05	-	-	-
JES/JMS	2.02	1.26	26.22	5.19	1.94	6.3
Bkg Est	5.03	6.32	12.59	-	-	-
b-tag SF	0.48	0.53	8.45	2.45	2.01	9.27
Total Sys	6.08	10.17	65.0	15.47	12.68	19.18
Stat	1.39	1.52	2.48	1.26	1.0	1.83
Estimated Events	781.61	702.23	79.38	26.0	0.76	0.013

Table 8.14: Percent impact of the dominant systematics on the background acceptance and on the signal acceptance of RS $c = 1.0$ graviton predictions in the $3b$ signal region.

TwoTag split	totalbkg	qcd	ttbar	RSG ₁ 1000	RSG ₁ 2000	RSG ₁ 3000
JER	0.25	0.48	3.14	1.18	0.74	0.5
JMR	0.52	1.73	9.43	10.96	12.3	13.05
Top	4.82	6.98	26.63	-	-	-
JES/JMS	0.43	1.67	7.17	6.72	1.69	3.44
Bkg Est	2.76	3.38	2.35	-	-	-
b-tag SF	0.83	1.43	1.82	19.28	27.36	2.7
Total Sys	5.66	8.26	29.46	23.2	30.05	13.77
Stat	0.6	0.42	2.48	2.0	1.2	1.07
Estimated Events	4252.44	3393.74	858.7	10.87	0.6	0.039

Table 8.15: Percent impact of the dominant systematics on the background acceptance and on the signal acceptance of RS $c = 1.0$ graviton predictions in the $2bs$ signal region.

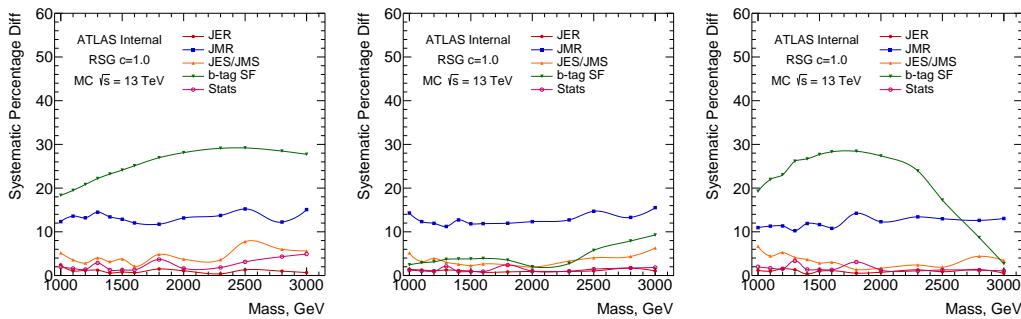


Figure 8.8: Impact of each systematic on the signal prediction as a function of the signal mass, in the $4b$ (left) and $3b$ (middle) and $2bs$ signal regions.

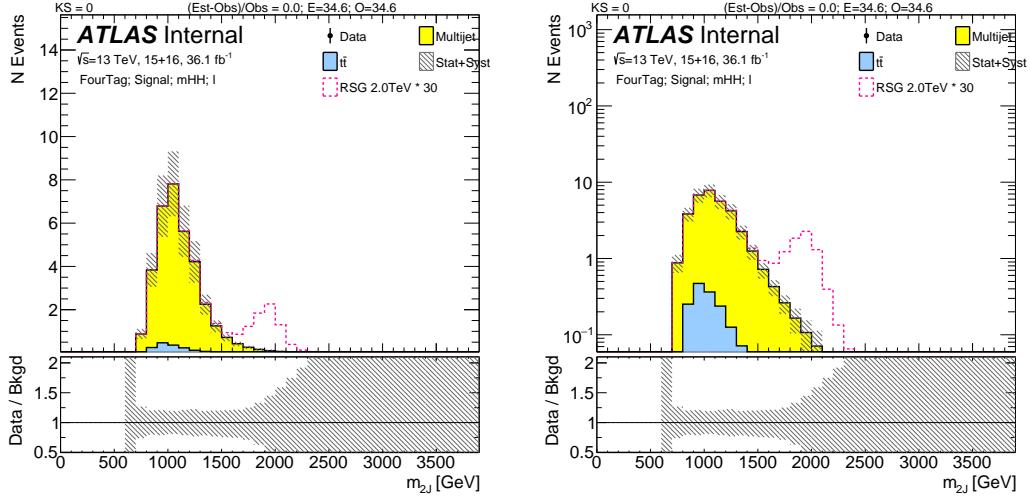


Figure 8.9: The total background estimation in $4b$ signal region, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.

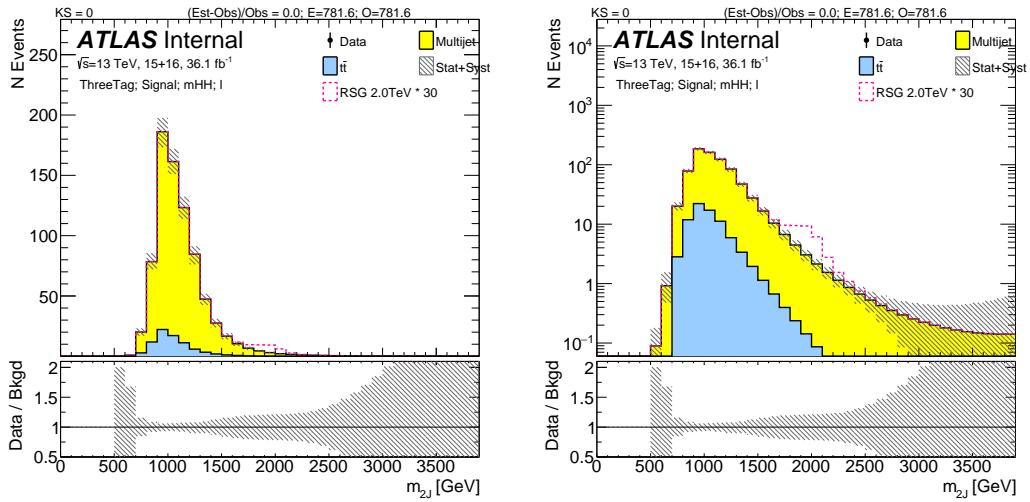


Figure 8.10: The total background estimation in $3b$ signal region, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.

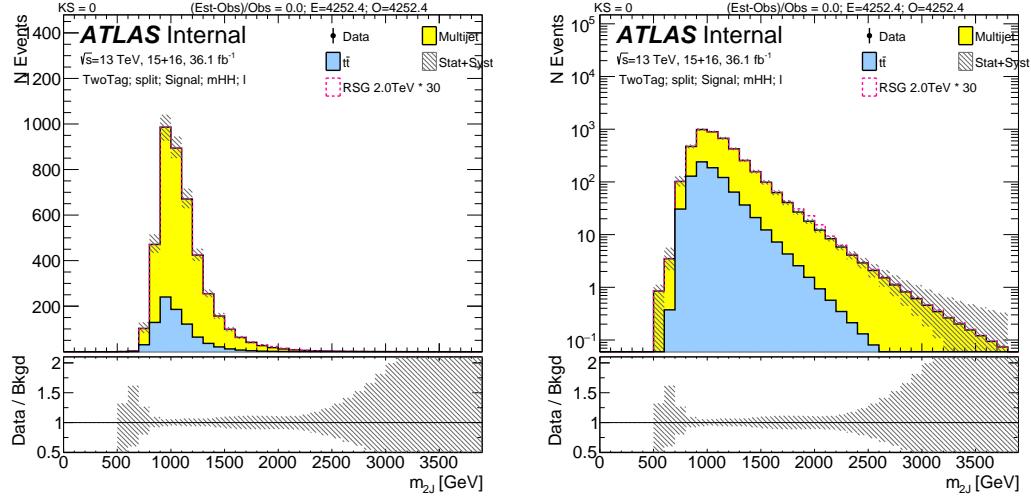


Figure 8.11: The total background estimation in $2b$ signal region, with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.

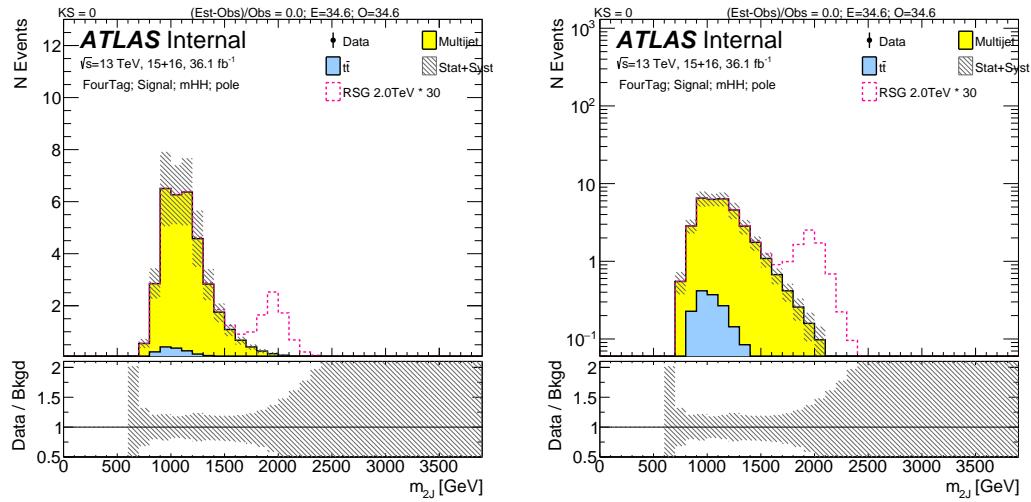


Figure 8.12: The total background estimation in $4b$ signal region, scaled m_{JJ} , with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.

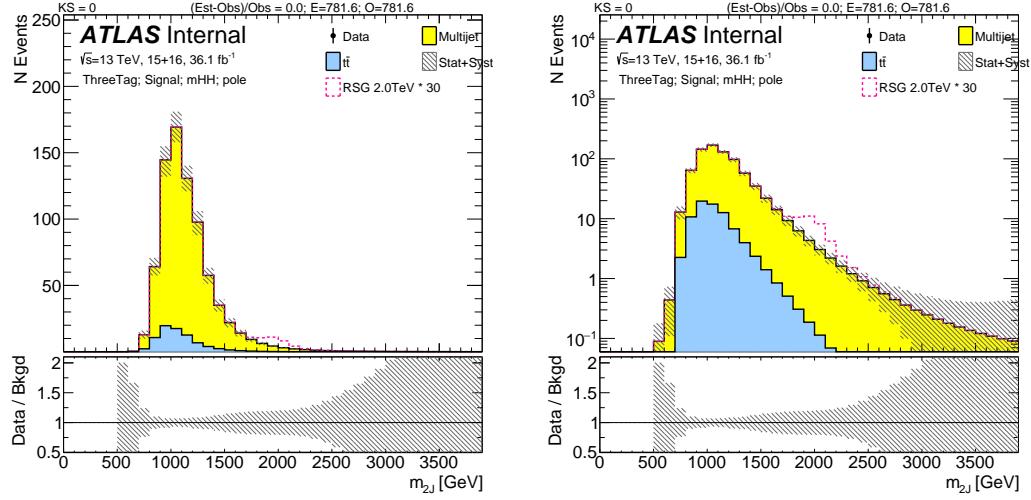


Figure 8.13: The total background estimation in $3l$ signal region, scaled m_{JJ} , with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.

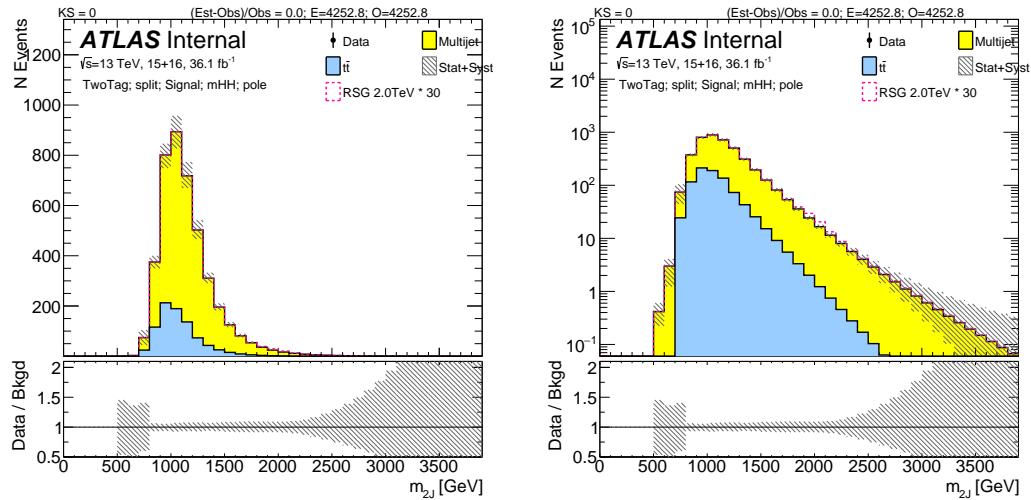


Figure 8.14: The total background estimation in $2bs$ signal region, scaled m_{JJ} , with linear scale on the left and with log scale on the right, along with total uncertainties (stats.+systematic) variation up and down.

Enfin.

French Person

9

Result

Not yet.

Tell the world.

Marathon

10

Intepretation

In order to avoid the **Oops-Leon** cases, commonly accepted standard for announcing the discovery of a particle is that the number of observed events is 5 standard deviations (σ) above the expected level of the background.

With no excess observed, a limit needs to be set on the cross section of the signal ².

One method to parametrise signal is called **singal morphing**. Specifically, in momentum morphing, not only the signal strength is scaled, but also the width is modified by a non-linear transformation.

An estimator should satisfy three criterias:

- Consistency: the value of the estimator should converge to the truth value if the sample size goes to infinity
- Efficiency: theory limits the variance of the true value, given a sample size N (Minimum Variance bound, or MVB). If the variance of the estimator is equal to the MVB the estimator is called efficient.
- Unbiased: the estimator should have no difference from the true value, otherwise it is biased.

Maximum likelihood estimator is a good estimator, based on these criterias.

For measurements, most of the time <https://arxiv.org/abs/1611.01927> is done to compare with generator level distributions. This accounts for detector effects, statistical fluctuations and background mis-identification. Since the analysis is a search, this $bb \rightarrow b\bar{b}b\bar{b}$ is a search, unfolding is less applicable in this case.

11

Conclusion

We will find something new one day, because there is always soemthing new.

A

Some extra stuff

A.I SIGNAL SAMPLE CONTAINERS

RS graviton samples with $c \equiv k/\bar{M}_P = 1.0$:

```
mc15_13TeV.301488.MadGraphPythia8EvtGen_A14NNPDF23LO_RS_G_hh_bbbb_c10_M300.merge.DAOD_EXOT8.e3820_s2608_s2183_r7772_r7676_p2952
mc15_13TeV.301489.MadGraphPythia8EvtGen_A14NNPDF23LO_RS_G_hh_bbbb_c10_M400.merge.DAOD_EXOT8.e3820_s2608_s2183_r7772_r7676_p2952
mc15_13TeV.301490.MadGraphPythia8EvtGen_A14NNPDF23LO_RS_G_hh_bbbb_c10_M500.merge.DAOD_EXOT8.e3820_s2608_s2183_r7772_r7676_p2952
mc15_13TeV.301491.MadGraphPythia8EvtGen_A14NNPDF23LO_RS_G_hh_bbbb_c10_M600.merge.DAOD_EXOT8.e3820_s2608_s2183_r7772_r7676_p2952
mc15_13TeV.301492.MadGraphPythia8EvtGen_A14NNPDF23LO_RS_G_hh_bbbb_c10_M700.merge.DAOD_EXOT8.e3820_s2608_s2183_r7772_r7676_p2952
mc15_13TeV.301493.MadGraphPythia8EvtGen_A14NNPDF23LO_RS_G_hh_bbbb_c10_M800.merge.DAOD_EXOT8.e3820_s2608_s2183_r7772_r7676_p2952
```



```
mc15_13TeV.301525.MadGraphPythia8EvtGen_A14NNPDF23LO_RS_G_hh_bbbb_c20_M2500.merge.DAOD_EXOT8.e3820_s2608_s2183_r7772_r7676_p2952
```

```
mc15_13TeV.301527.MadGraphPythia8EvtGen_A14NNPDF23LO_RS_G_hh_bbbb_c20_M3000.merge.DAOD_EXOT8.e3820_s2608_s2183_r7772_r7676_p2952
```

The 2HDM H samples are not simulated yet, although particle-level events have been generated.

DAOD_EXOT8 samples will shortly be produced from the following set of samples:

```
mc15_13TeV.343394.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m260_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343395.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m300_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343396.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m400_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343397.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m500_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343398.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m600_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343399.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m700_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343400.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m800_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343401.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m900_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343402.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m1000_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343403.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m1100_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343404.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m1200_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343405.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m1300_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343406.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m1400_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343407.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m1500_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343408.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m1600_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343409.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m1800_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343410.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m2000_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343411.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m2250_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343412.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m2500_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
mc15_13TeV.343413.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m2750_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952

mc15_13TeV.343414.MadGraphHerwigppEvtGen_UEEE5_CTEQ6L1_CT10ME_Xhh_m3000_4b.merge.DAOD_EXOT8.e4729_s2726_r7772_r7676_p2952
```

A.2 CHOICE OF SB/CR DEFINITIONS

This section is for detailed description of Sideband and Control region choice, for Section 7.1.2.

Sideband size is limited to be outside the control region. The choice of the upper bound of the Sideband is constrained such that the leading jet mass spectrum will contain the $t\bar{t}$ peak and so that the Sideband has enough statistics and is as close to the signal and control region as possible. One test is to vary the Sideband size and to re-estimate the μ_{qcd} and $\alpha_{t\bar{t}}$ values. This is shown in Figure A.1. The values of fitted μ_{qcd} and $\alpha_{t\bar{t}}$ are fairly stable for different Sideband choices.

To evaluate the statistical effect of smaller Sideband regions, we plot the ratio of the stat uncertainty of the predicted number of events vs. the fit uncertainty of μ_{qcd} and $\alpha_{t\bar{t}}$. This is shown in Figure A.1. In Sideband region, the ratio (right) is close to 1, because the Sideband normalization is fixed in the fit. Based on the choice that for the signal region predictions, the fit uncertainty should be around half of the statistical uncertainty, the Sideband upper limit is chosen to be $R_{bb}^{\text{high}} < 58$ GeV. Note that this will be the way to assign non-cllosure test in the systematics, as described in Section 8.0.3.

A similar test can be done on different Sideband Region shifts. This is shown in Figure A.2. Based on the choice that for the signal region predictions, the fit uncertainty should be around half of the statistical uncertainty, the Sideband center shift is chosen to be 10 GeV.

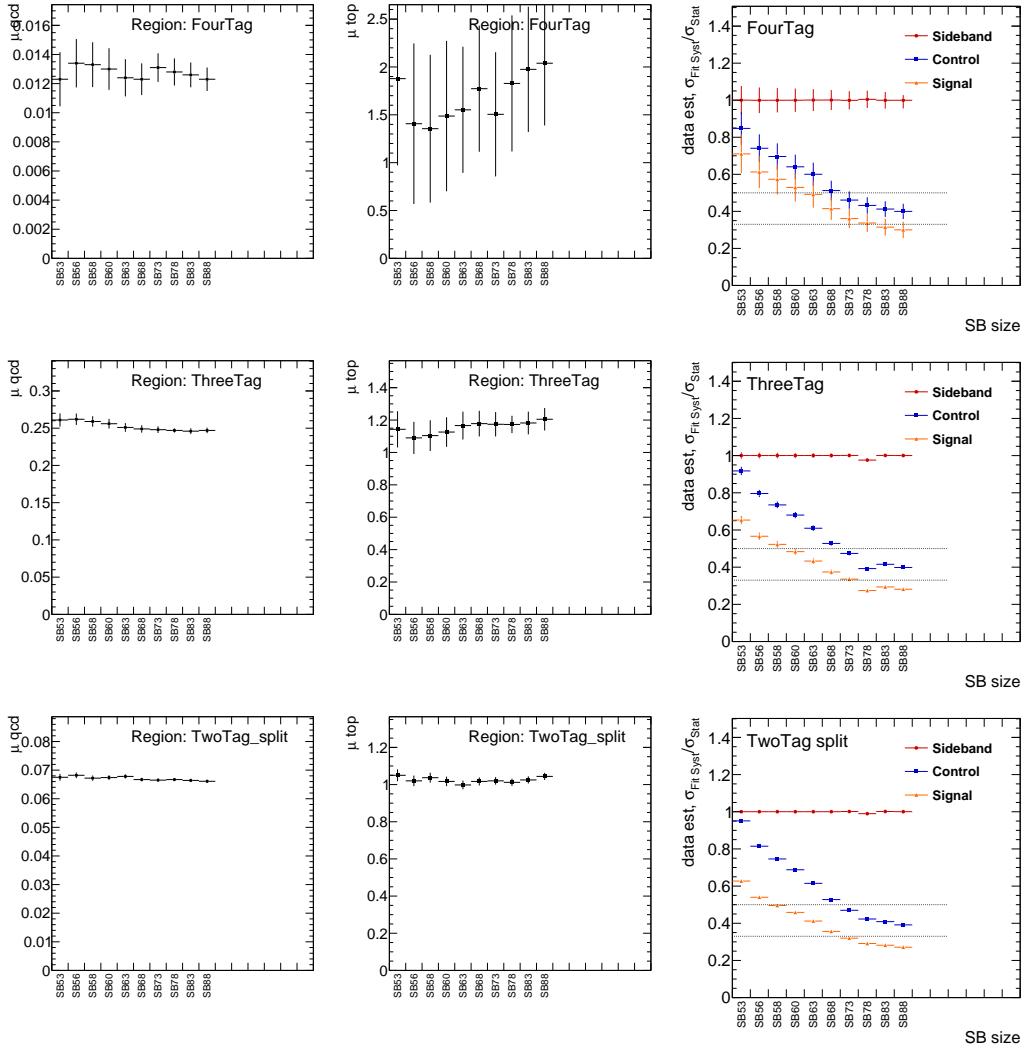


Figure A.1: Fit parameters for μ_{qcd} (left), $\alpha_{\bar{t}t}$ (middle), and ratio of fit uncertainty and stat uncertainty as a function of different Sideband values (right), for $4b$ (top) $3b$ (middle) and $2bs$ (bottom). The x-axis indicates different values of SB R_{hh}^{high} cuts.

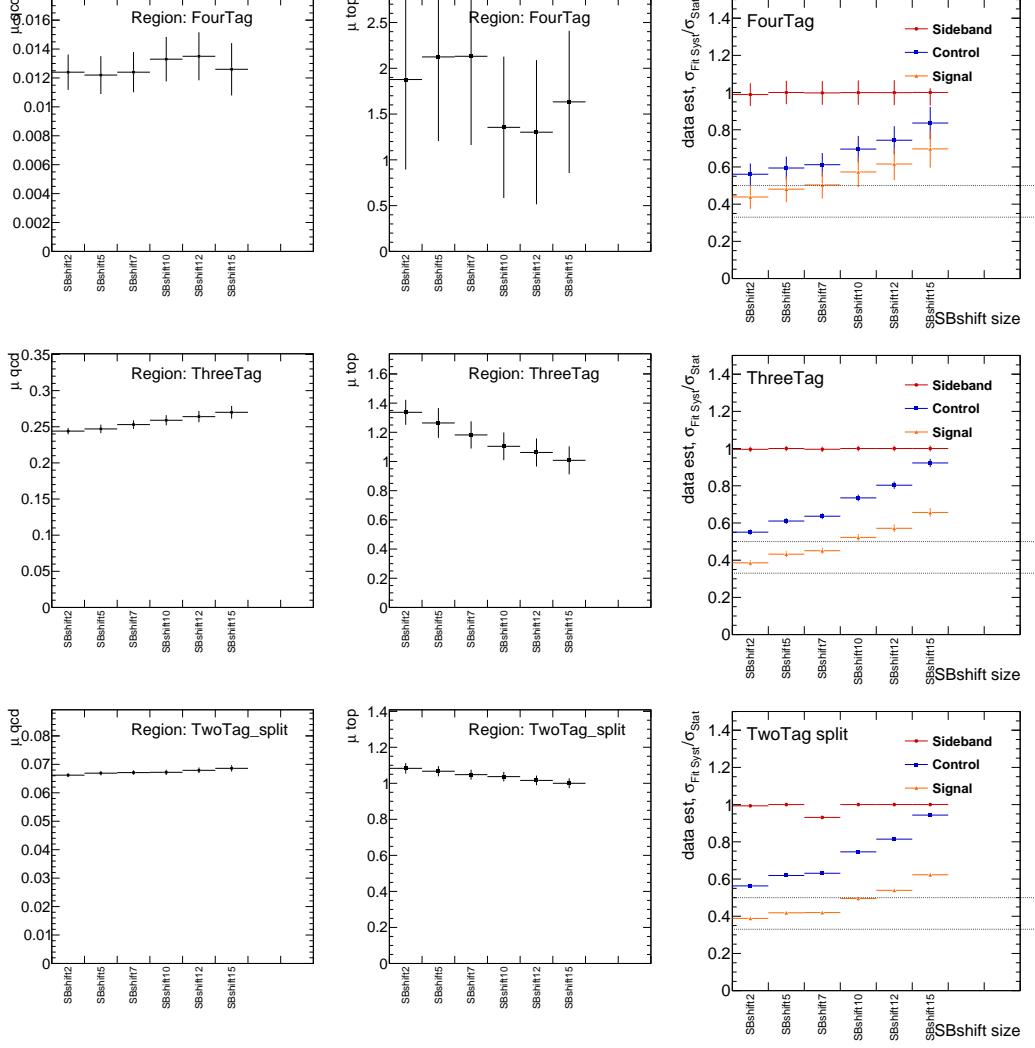


Figure A.2: Fit parameters for μ_{qcd} (left), $\alpha_{\bar{t}\bar{t}}$ (middle), and ratio of fit uncertainty and stat uncertainty as a function of different Sideband shift values (right), for $4b$ (top) $3b$ (middle) and $2bs$ (bottom). The Sideband region is chosen to be $R_{bb}^{high} < 58$ with different shifts, and Control region is fixed at $33 < R_{bb}$. The x-axis indicates different values of SB center shifts.

A similar test can be done on different Control Region sizes. This is shown in Figure A.3. Based on the choice that for the signal region predictions, the fit uncertainty should be around half of the statistical uncertainty, the Control upper limit is chosen to be $R_{hh} < 33$ GeV.

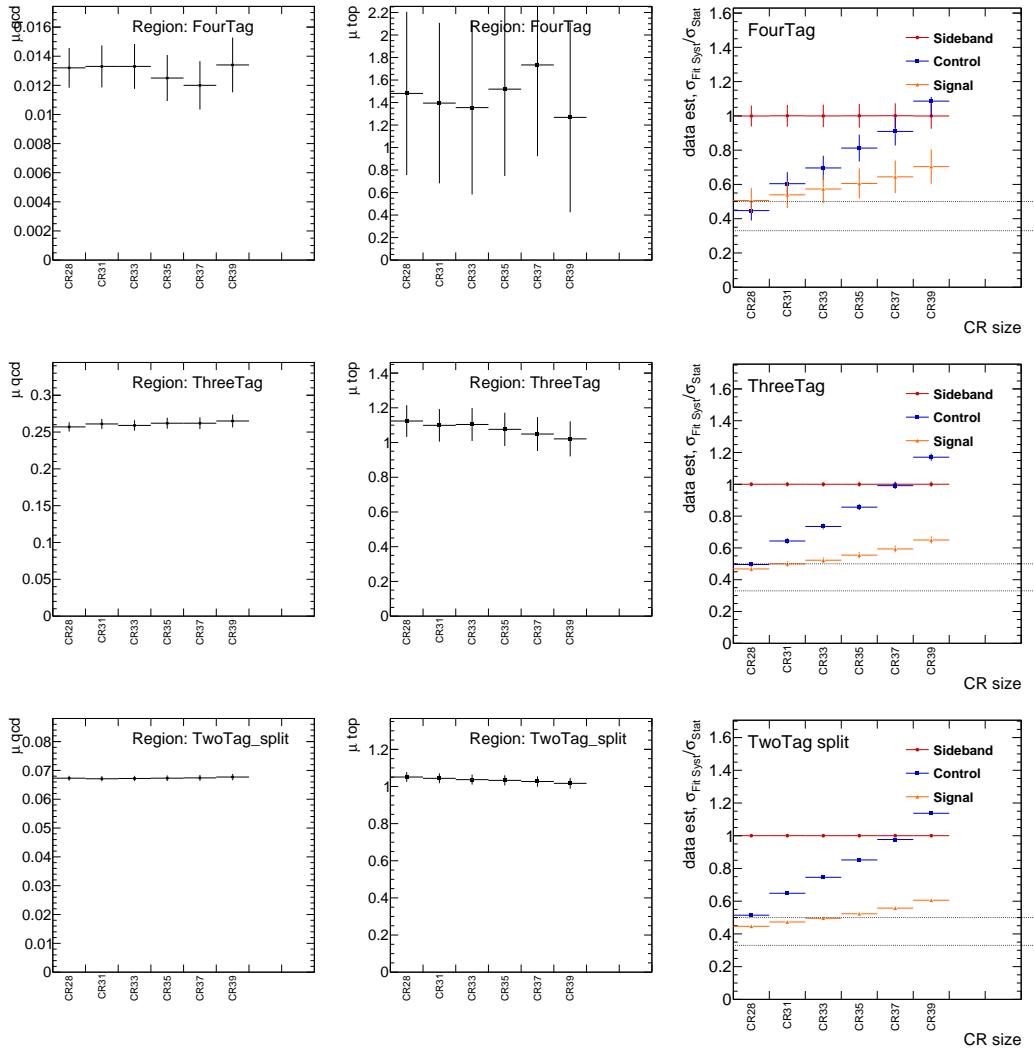


Figure A.3: Fit parameters for μ_{qcd} (left), $\alpha_{\bar{t}\bar{t}}$ (middle), and ratio of fit uncertainty and stat uncertainty as a function of different Control values (right), for $4b$ (top) $3b$ (middle) and $2bs$ (bottom). The Sideband is chosen to be $R_{hh}^{high} < 58$ with 10 GeV shifts. The x-axis indicates different values of CR R_{hh} cuts.

A.3 μ_{multijet} METHOD VALIDATION

One important validation is whether the μ_{multijet} is the same for different number of b tags in multijet event. This can be validated in data, excluding signal regions. For data, the $t\bar{t}$ contribution is estimated directly from MC and subtracted in the distributions. Then two plots are produced: one is the ratio of the number of $N b$ tagged events versus the number of $N-1 b$ tagged events in each leading Higgs candidate/subleading Higgs candidate mass bin, the second is the pull of the bins in Sideband/Control/Signal regions. The two distributions show roughly the consistency of μ_{multijet} in different regions, as seen in Figure A.4 ($1b$ over $0b$), A.5 ($2b$ over $1b$), A.6 ($2bs$ over $1b$), A.7 ($3b$ over $2b$), A.8 ($4b$ over $2b$). For $4/3/2bs$, the μ_{multijet} value can be compared with 7.2. In Figure A.4 and A.5, the agreement of μ_{multijet} weighted mean value in Sideband/Control/Signal (especially) regions validates the constant μ_{multijet} assumption in the analysis.

Also, the Dijet MC can be used for validation. The same distributions, without subtracting $t\bar{t}$, are shown in Figure A.9 ($1b$ over $0b$), A.10 ($2b$ over $1b$), A.11 ($2bs$ over $1b$), A.12 ($3b$ over $2b$), A.13 ($4b$ over $2b$). Poor Statistics of the dijet MC affect the pull distributions, especially for $3b$ and $4b$ ratios, yet the consistency of μ_{multijet} in different regions can still be validated. This also proves that the dijet MC could not be used directly for background estimation.

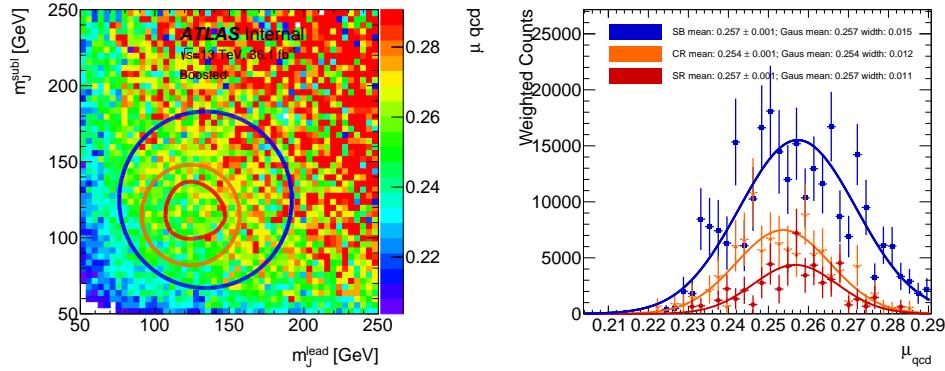


Figure A.4: $1b$ over $0b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.

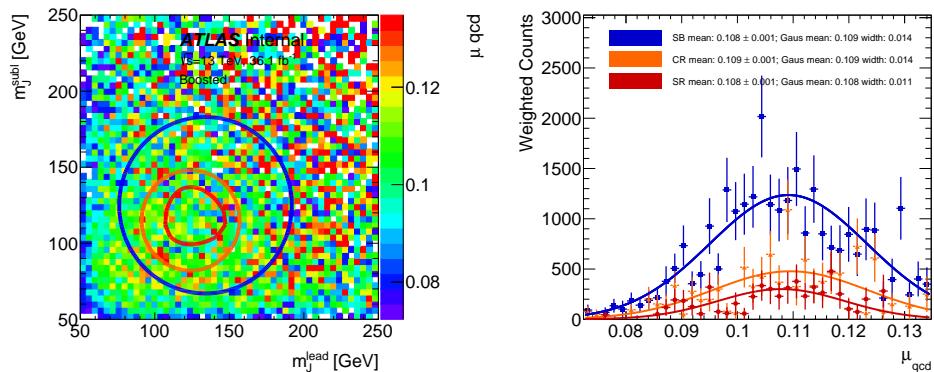


Figure A.5: $2b$ over $1b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.

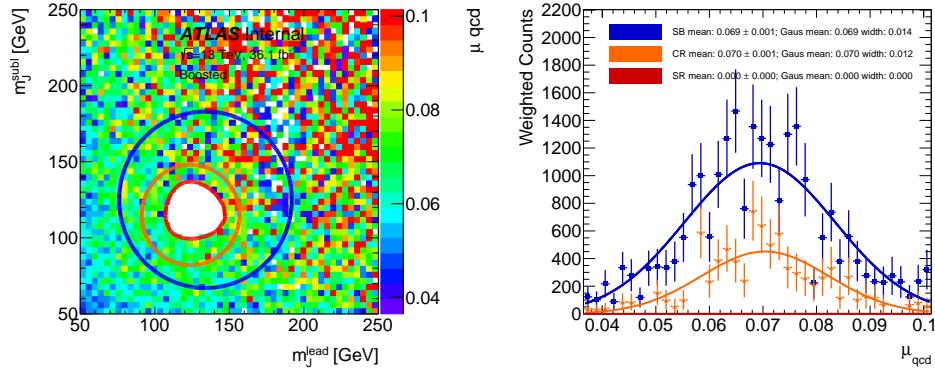


Figure A.6: $2b$ over $1b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.

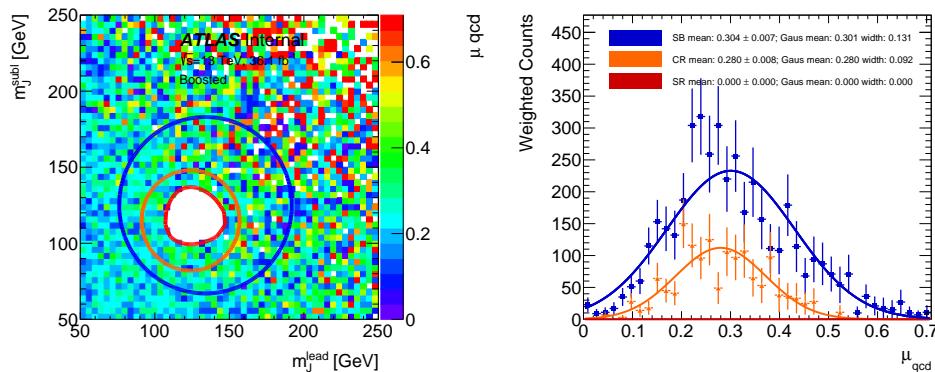


Figure A.7: $3b$ over $2b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.

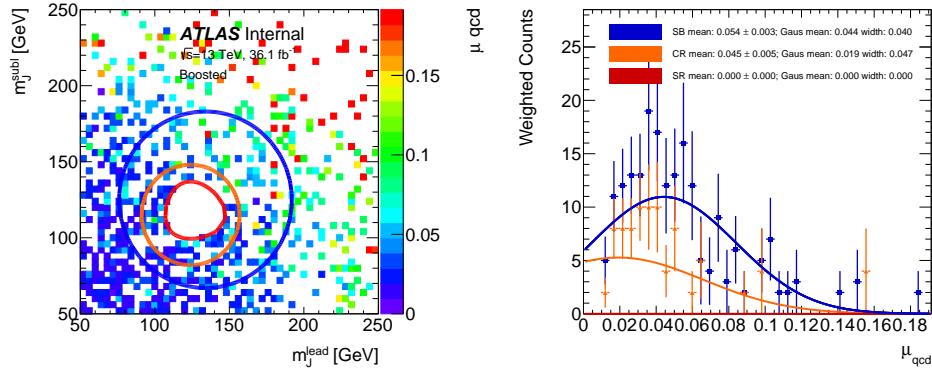


Figure A.8: $4b$ over $2b$ data driven μ_{multijet} values: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.

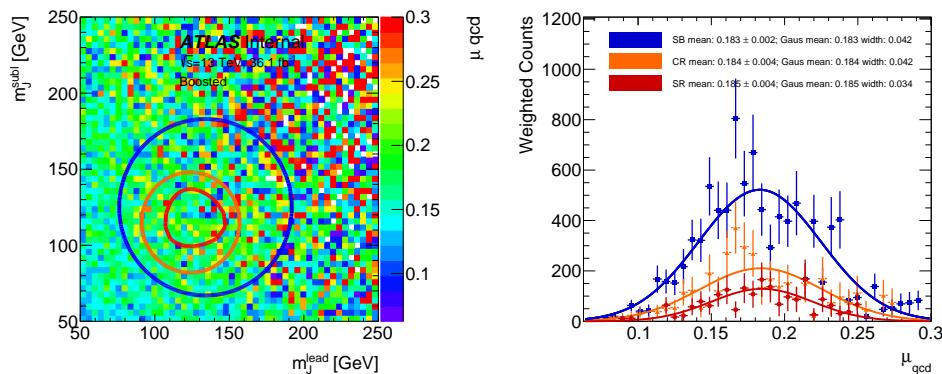


Figure A.9: $1b$ over $0b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Gaussian fit mean value shown on the plot.

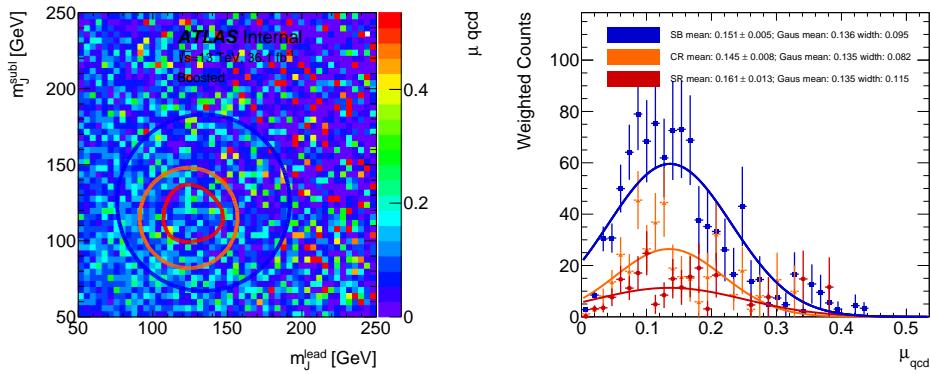


Figure A.10: $2b$ over $1b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Guassian fit mean value shown on the plot.

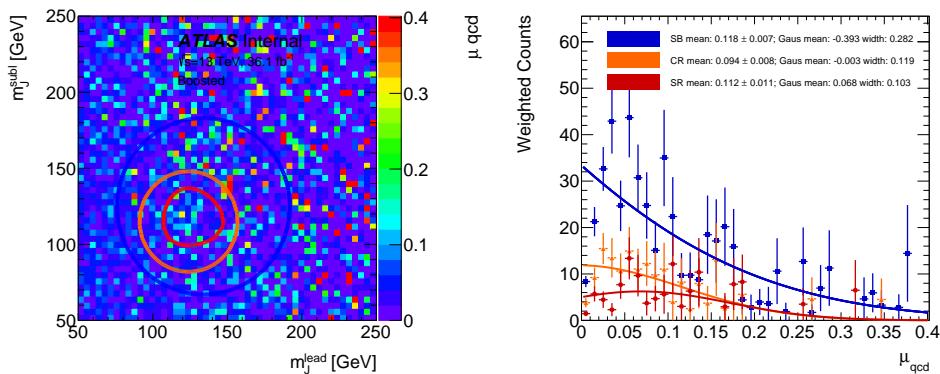


Figure A.11: $2bs$ over $1b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Guassian fit mean value shown on the plot. Poor Statistics of the dijet MC affect the pull distributions.

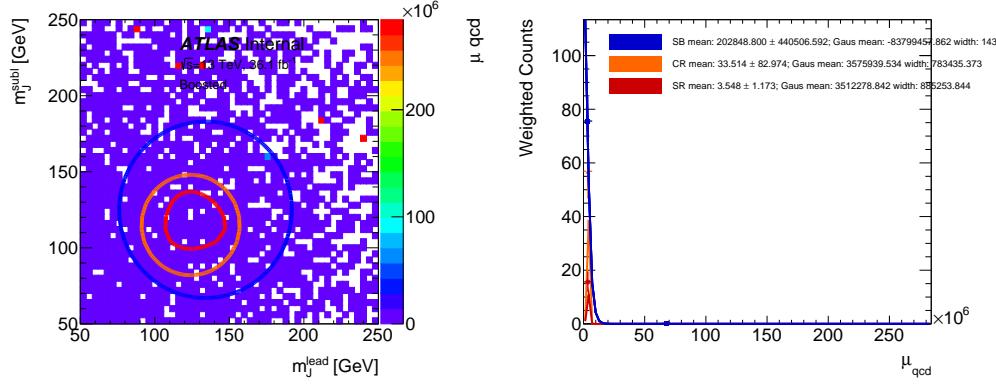


Figure A.12: $3b$ over $2b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Guassian fit mean value shown on the plot. Poor Statistics of the dijet MC affect the pull distributions.

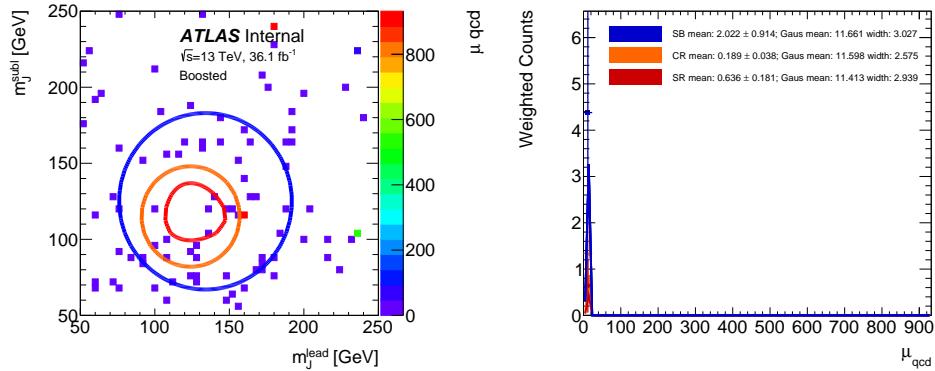


Figure A.13: $4b$ over $2b$ data driven μ_{multijet} values in dijet MC: μ_{multijet} as a function of leading Higgs candidate/subleading Higgs candidate mass(left); and μ_{multijet} value pull in Sideband/Control/Signal regions(right), with the weighted mean value and the Guassian fit mean value shown on the plot. Poor Statistics of the dijet MC affect the pull distributions.

A.4 DESCRIPTION OF SELECTION OPTIMIZATION METHOD

To study the effect of different selections, a signal significance is estimated for RSG $c=1.0$ sample. A full analysis is run with the difference selection, including a separate μ_{qcd} and $\alpha_{t\bar{t}}$ estimate. Then, for each signal sample's signal region, with the background prediction, the significance is estimated by first finding the maximum bin in the signal dijet mass distribution. Then, a window is selected around the maximum bin, where there are 68% of the total events within the mass windows. The total number of events is integrated within that window, for both the signal sample as N_{signal} and the background estimation as N_{bkg} , and then the sensitivity/significance is defined to be $\frac{N_{signal}}{1+\sqrt{N_{bkg}}}$.

To gain a comparison, the result is compared with the default selections as documented in the note. An example is shown in Figure A.14.

Figure A.14: An example of significance comparison plot. On the top pad, the x-axis is for different RSG masses, while the y-axis is plotting the estimated significance. The black dot is for the $2b$ signal region with the specified selection, the red square is for the $3b$ signal region, and the green star is for the $4b$ signal region. The red cross is the quadratic sum of the three signal regions above. And the black circle is the default selection total combined significance in this note. At the bottom pad, the x-axis is for different RSG masses, while the y-axis is plotting the combined estimated significance ratio of the specified selection to the reference selection. In this case, 77% working point study is compared to the reference, which has the same b -tagging working point. Hence all the ratios are 1.

A.5 OPTIMIZATION STUDIES

A.5.1 COMBINED MASS IMPACT

Using combined mass instead of the calorimeter based mass improves the boosted exclusion limit across the mass range, as can be seen in A.15.

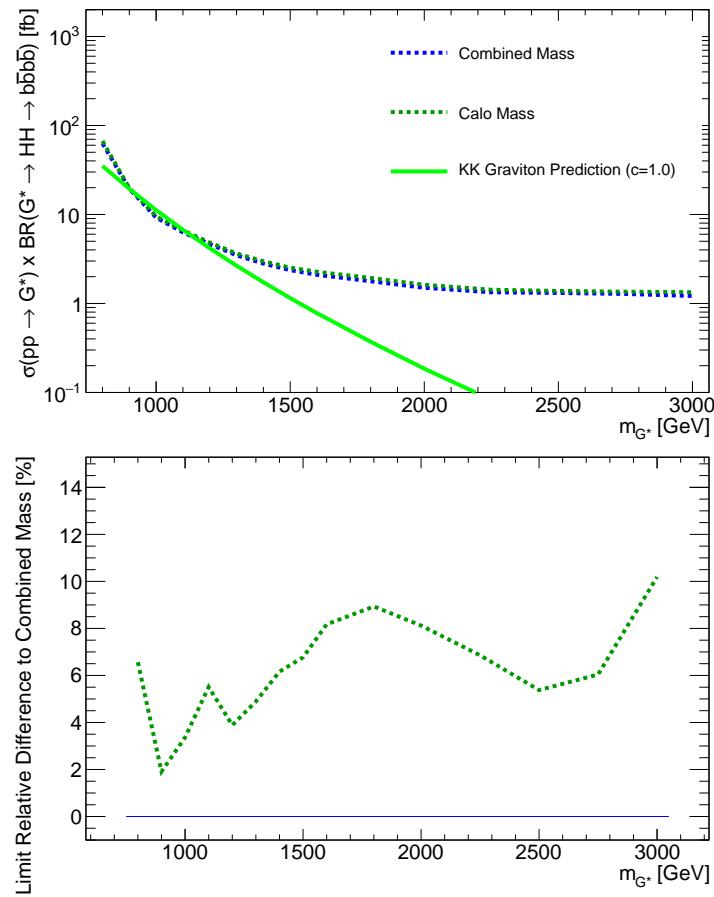


Figure A.15: Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to the blue combined mass limit. Comparison shown here are analysis repeated with calorimeter mass(green).

A.5.2 *b*-TAGGING IMPACT

b-tagging working point on track jets is chosen to be 70% for this analysis, which used to be 77%. Different *b*-tagging working points have been tested. The resulting plot is shown as Figure A.16. As seen, the 70% *b*-tagging working point is better than/comparable to 77% *b*-tagging working point in terms of cross section asymptotic statistical limit, especially for higher mass signals. Hence it is chosen to be used by this analysis.

A.5.3 RESOLVED VETO IMPACT

Impact from veto resolved events passing resolved signal region selection is shown also in Figure A.16. The impact comes mostly from 800 GeV to 1.5 TeV, with largest impact at 900 GeV, where the cross section asymptotic statistical limit differes by a factor of 10. Since those events are in the resolved signal region and will be combined in the final statistical limit, it won't harm the overall resolved boosted combined result.

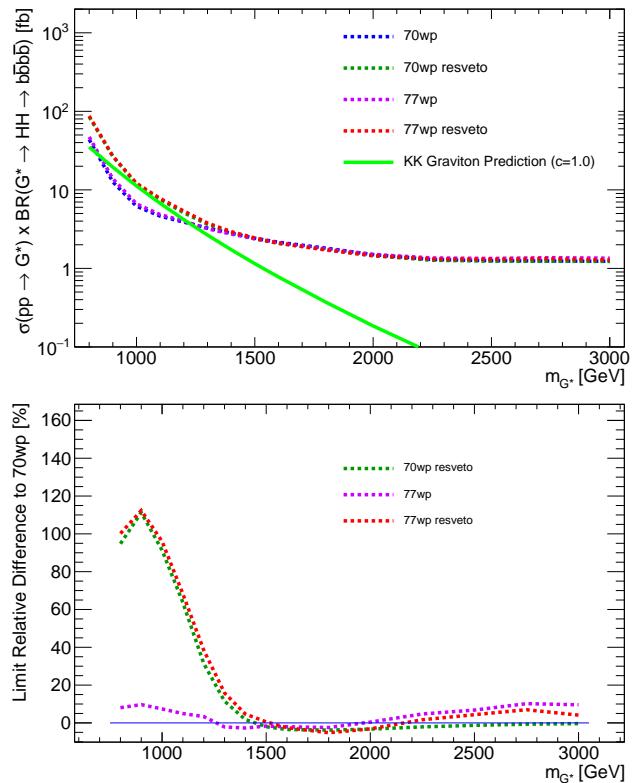


Figure A.16: Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to the blue 70% b -tagging working point limit. Comparison shown here are analysis repeated with 70% working point but with resolved veto (green), 77 % working point (purple), 77% with resolved veto (red).

A.5.4 SCALED MASS IMPACT

Using scaled mass helps to be consistent with the resolved analysis, and gives very similar exclusion limit as using the unscaled MJJ. The difference on the limit is minimal, as shown in Figure A.17.

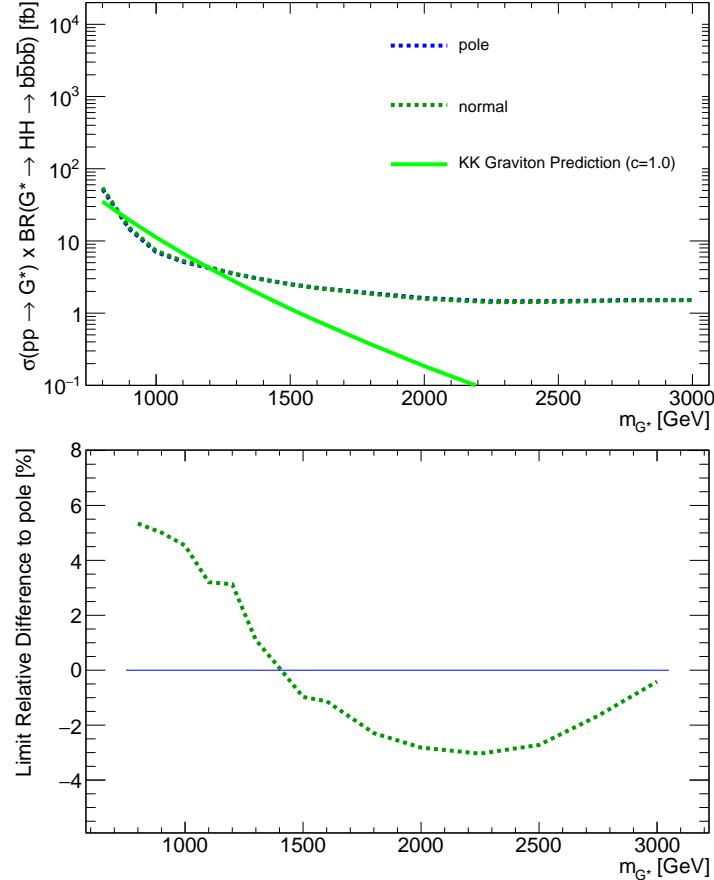


Figure A.17: Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to scaled MJJ limit. Comparison shown here are analysis repeated with unscaled MJJ (green).

A.5.5 VARIABLE RADIUS TRACKJET IMPACT

Using variable radius (VR) trackjet's impact on the analysis is evaluated. At the time of the test, the b -tagging calibration on MCs is missing. The improvements mostly show for high mass signals, yet a relative lower sensitivity on the lower mass signals is found, as shown in Figure A.18.

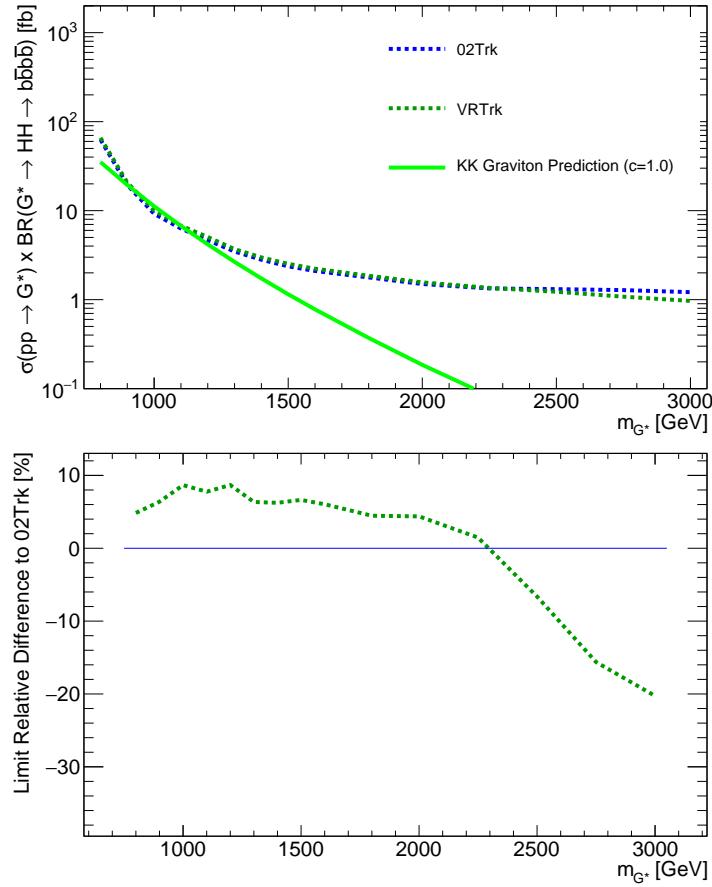


Figure A.18: Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to the standard trackjet based analysis limit. Comparison shown here are analysis repeated with variable radius track jets.

A.5.6 BOOSTED SIGNAL REGION OPTIMIZATION

The signal region is optimized and the optimal expression is:

(A.1)

which is different from the chosen signal region, as shown in Equation 6.1. This optimization has two changes, the different width for leading Higgs Candidate and subleading Higgs Candidate, and the p_T -dependent X_{hh} cut, which effectively increase the signal region for higher resonant masses. The impact on the exclusion limit is shown in Figure A.19. The improvement in exclusion limit is around 2 to 8 %, with a strong dependence on signal mass. In order to keep the analysis simple and consistent with the resolved analysis, this signal region is not adopted.

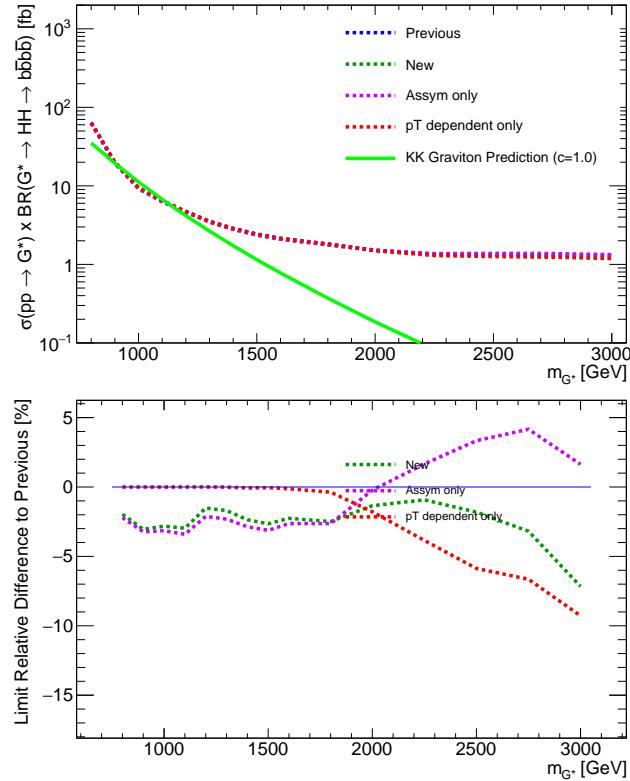


Figure A.19: Expected limit for combined boosted channels. The top plot is the limit for RSC $c = 1.0$, and the bottom plot is the ratio of the limit relative to the standard signal region analysis limit. Comparison shown here are analysis repeated with the signal region defined above (new, green), with the asymmetric width only (purple), and the p_T -dependent Xhh cut only (red).

A.6 CORRELATED ERROR PROPAGATION

In several instances, we fit distributions with multiple parameters, and the resulting parameter errors are found to be correlated. When propagating uncertainties, this implies that the errors of the correlated parameters of the fit can not be varied independently to estimate the uncertainty.

To handle the correlated error propagation, we make use of the fact that the covariance matrix of the fit parameters is a symmetric matrix and can be diagonalized by unitary congruence. Specifically, if the fit has two parameters and we write the fit parameters as a vector, $\alpha = (\alpha_1, \alpha_2)$, and covariance matrix of α is C where $C_{ij} = Cov(\alpha_i, \alpha_j)$, then we can decompose C as:

$$C = U\Lambda U^T \quad (\text{A.2})$$

where U is a unitary matrix whose columns are the eigenvectors of C and $\Lambda = Diag(\lambda_1, \lambda_2)$ is a real diagonal matrix whose entries are the eigenvalues of C .

If we were project the vector α along the direction of the eigenvectors, i.e. $\alpha \rightarrow \alpha' = U^T \alpha$, then the resulting eigenvalues would be uncorrelated because $C \rightarrow C' = U^T C U = \Lambda$. Thus we could vary the values of α'_i independently by $\pm \lambda_i$ to propagate the uncertainty. To return to the original space of α , we simply need to rotate the system back with the U matrix.

This set of operations can be simplified. Let $l_1 = (\lambda_1, \alpha)^T$ and $l_2 = (\alpha, \lambda_2)^T$, then a variation of the correlated parameters takes the form,

$$\alpha \rightarrow U(\alpha' \pm l_i) = U(U^T \alpha \pm l_i) = \alpha \pm Ul_i \quad (\text{A.3})$$

and thus the fit parameters errors can be varied in a correlated way by simply adding $\pm Ul_i$ to the original parameter vector α .

A.7 BASIC SELECTION VALIDATION WITH SIGNAL TRUTH MC

Monte Carlo data on $RSG \rightarrow HH \rightarrow 4b$ is studied to validate the current methods to match Higgs to calorimeter jets and b's to trackjets. Namely, that the two leading (by p_T) calorimeter jets, each with a size of $\Delta R = 1.0$ are treated as Higgs candidates, and that the two leading trackjets (also by p_T) within each calorimeter jets are treated as the b candidates.

Large-R Jets Higgs Matching: First it is examined whether the two leading calorimeter jets match with the Higgs particles, i.e. whether each Higgs has a ΔR less than 1.0 for a matching calorimeter jet. Figure A.20 demonstrates that $\sim 95\%$ of the time, the two leading large-R jets match up with the two leading Higgs. Thus, it is not necessary to include the third leading calorimeter jet to identify the Higgs, as doing so could only improve the matching rate by 5%, and doing so would likely lower background rejection.

Figure A.20: Percentages of Higgs to Jet matching possibilities as a function of the mass of the Randall-Sundrum Graviton in Monte Carlo data. Note that the case where each Higgs gets matched to the same jet essentially never happens, so it is not included in this plot.

In addition, the calorimeter jets almost always contain the $b\bar{b}$ pair that each Higgs decay into. Figure A.21 shows that the Higgs and its leading child b almost always have an angular displacement less than 1.0, so as long as the Higgs is matched closely with a calorimeter jet, each b will also be inside that calorimeter jet. Thus, a size of $\Delta R = 1.0$, is enough to contain both b's.

Track Jets-Matching: Now that we know that the b's are contained inside the large-R jets, we can investigate whether they actually match with the track jets inside each large-R jet. Figure A.22 shows

Figure A.21: ΔR between Higgs (leading by p_T on left, subleading on right) and the child b quark. Lines are drawn at $\Delta R = 0.4$ (size of large R jets for the resolved $HH \rightarrow 4b$ analysis) and $\Delta R = 1.0$ (size for boosted).

that for low RSG mass, b's are matched to track jets about 80% of the time. At high mass, the b-jets begin to merge, so at that point the b's tend to be identified with the same trackjet (this is accounted for in the Boosted analysis by the 2tag split signal region). Including a third trackjet could help only in the case where one of the b's is not matched but the Higgs is matched correctly, which happens only about 5 – 15% of the time. Thus using the two leading trackjets is good enough for matching the b's, since including a third trackjet could improve the matching rate by at most 10%.

Figure A.22: Percentage of the time that different combinations of matching b's to trackjets occur relative to all possible combinations of matching the $b\bar{b}$ pair to 2 trackjets (b's inside leading Higgs on left, inside subleading on right). The situations listed in the legend are all independent, but they are not comprehensive. The situations not listed on the legend (including those cases where a b is not in the calorimeter jet) happen in total at most 1.6% of the time for a given mass point.

b-Tagging: Finally, it is examined how often the track jets that the b's are matched to are b-tagged. B-tagging is based on (the old method) MV2c20 with a cut of -0.6134 (at the 77% working point). B-tagging does not perform particularly well on the trackjets, especially for high RSG mass (Figure A.23). Note that the rise in the percentage of time that both b's are b-tagged at high RSG mass is due to the fact that the b-jets begin to merge (as shown in figure A.22) and because in figure A.23 the b's are allowed to match to the same track jet. Thus, the rise is not due to an effect of the b-tagging algorithm.

Figure A.23: Percentage of the time that trackjets matched to b's are b-tagged (b's inside leading Higgs on left, inside subleading on right). b-tagging is based on MV2c20 at a cut of -0.6134 (a working point of 77%). Only those cases where both b's are matched to a trackjet are considered. In the case where one b is matched to two jets, the closer of the two jets is checked against. The b's are allowed to match to the same jet.

A.8 REWEIGHTING STUDIES AND COMPARISONS

More reweighting methods are tested to further validate our current reweighting method. The original method described in 7.1.6, which tries to reweight the non-Tagged Higgs candidate in the lower b -tagged region to be like a b -tagged Higgs candidate. AddTag reweighting is done on three variables: large- R jet p_T and the two track jet $p_{T\text{s}}$, which is counted as one iteration of reweighting.

Other reweighting methods, that share similar idea with the AddTag method, where $1/2b$ samples are inclusively reweighted, have been tested:

- bkgeta: AddTag reweighting, but on four variables: large- R jet p_T , the two track jet $p_{T\text{s}}$, and large- R jet η .
- bkgdr: AddTag reweighting, but on four variables: large- R jet p_T , the two track jet $p_{T\text{s}}$, and two track jet ΔR .
- bkgtrk: AddTag reweighting, but on two variables: only the two track jet $p_{T\text{s}}$.
- bkgsb: AddTag reweighting, using the same three variables but instead of reweighting the $1b$ non-tagged Higgs candidate to be like the $1b$ tagged Higgs candidate, reweight $1b$ non-tagged Higgs candidate's distributions to be like the $2bs$ Higgs candidate in the sideband. Similarly, for $3b$, reweight $2b$ non-tagged Higgs candidate's distributions to be like the $3b$ Higgs candidate in the sideband, and for $4b$, reweight $2b$ non-tagged Higgs candidate's distributions to be like the $4b$ Higgs candidate in the sideband.

Also, reweighting the $1b$ Sideband directly to be like $2bs$ Sideband, and propagate the weight to control and signal regions, as the method used for ICHEP analysis, has been tested:

- leadtrk: on three variables: leading Higgs candidate large- R jet p_T , the two lead-track jet $p_{T\text{s}}$.

- sublitrk: on three variables: leading Higgs candidate large- R jet p_T , the two sublead-track jet $p_{T\text{s}}$.
- alltrk: on five variables: leading Higgs candidate large- R jet p_T , , the two lead-track jet $p_{T\text{s}}$, and the two sublead-track jet $p_{T\text{s}}$.

For MJJ, the dijet mass distribution, the distribution comparisons are shown in Figure A.26, A.30, A.32. Singal regions are blinded, and the data distribution is replaced with the non-reweighted background estimate. For trackjets p_T distributions, the comparisons are shown in Figure A.29, A.31, A.33. In most cases, the χ^2 is improved for the background estimations, and very similar distributions in signal region is observed. Hence no extra systematics is assigned for the reweighting method.

The reweighting method can be further validated using Dijet MC. Due to the limited statistics, only $2bs$ dijet MC has enough event for testing. The weights derived in data are applied in the MC $1b$ using the AddTag method, and compared with the $1b$ un-reweighted distribution and the $2bs$ distribution. The inclusive region (sideband + control + signal) trackjet p_T distributions are shown in Figure A.24. As can be seen, reweighting helps to model different b tagging sculpting effect. The MJJ distributions can be seen in Figure A.25. Reweighting differs from the non-reweighted distribution only by a small amount, yet improves the overall χ^2 value.

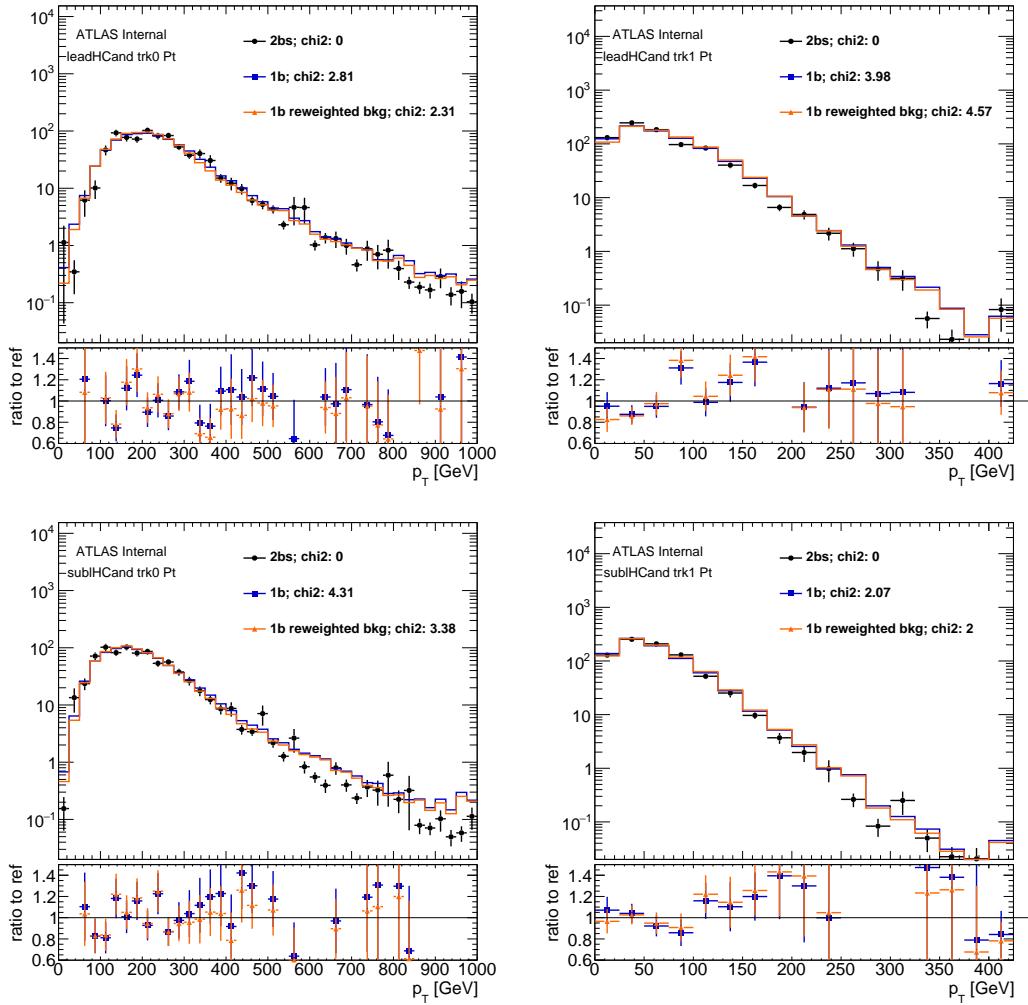


Figure A.24: Reweighted $2bs$ inclusive region predictions comparison, for leading Higgs Candidate leading trackjet p_T (top left), leading Higgs Candidate subleading trackjet p_T (top right), subleading Higgs Candidate leading trackjet p_T (bottom left), subleading Higgs Candidate subleading trackjet p_T (bottom right). $1b$ and reweighted $1b$ distributions are normalized to $2bs$.

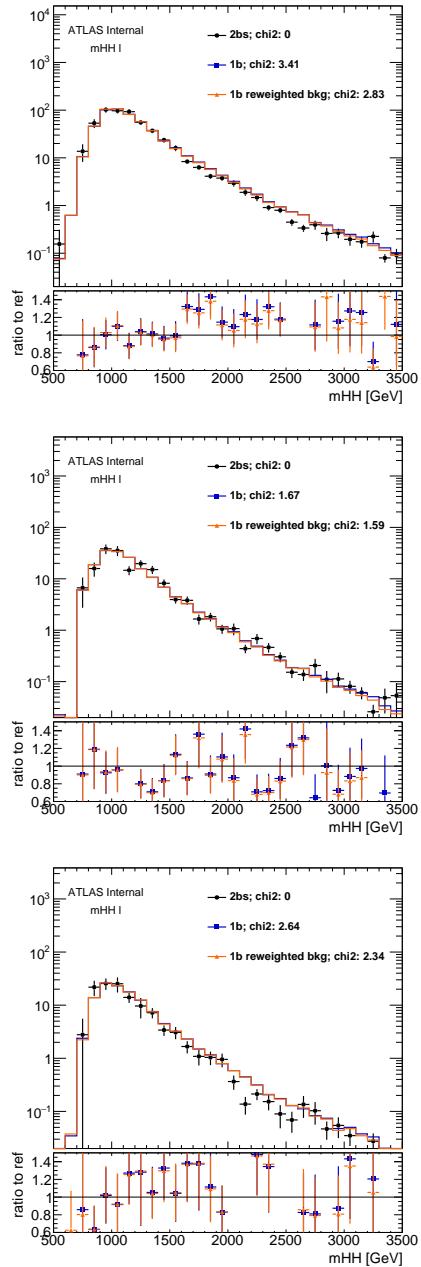


Figure A.25: Reweighted $2bs$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is not blinded, since it is MC distribution. $1b$ and reweighted $1b$ distributions are normalized to $2bs$.

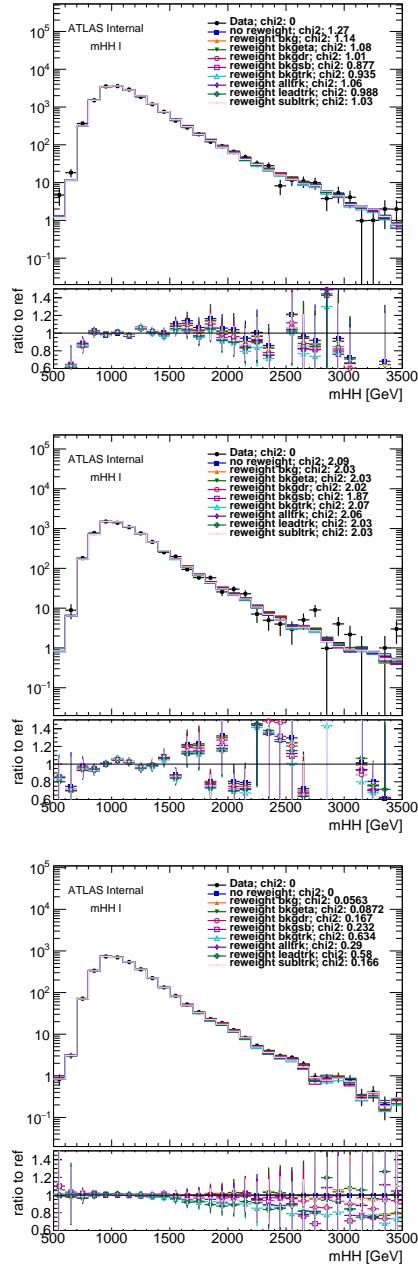


Figure A.26: Reweighted $2bs$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.

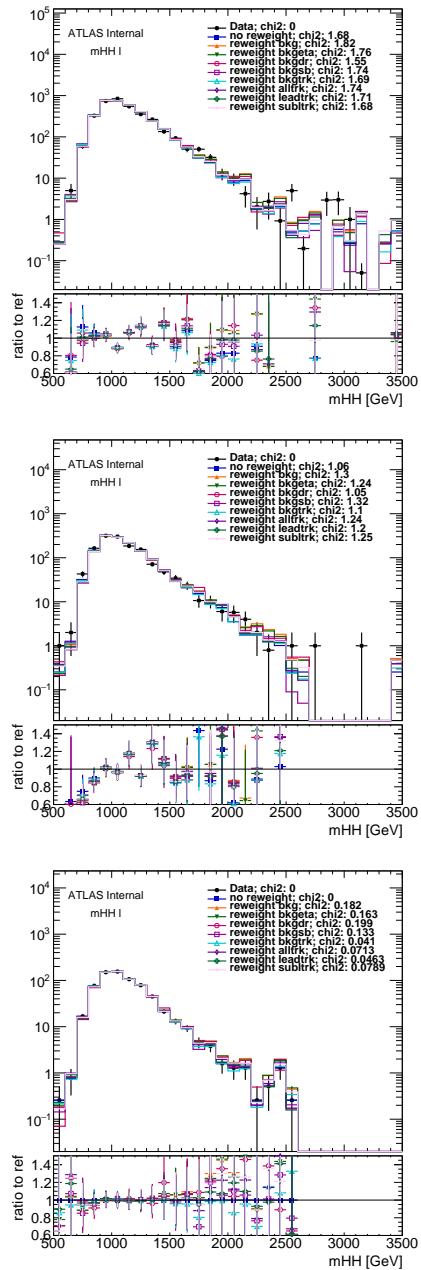


Figure A.27: Reweighted $3b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ.
The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.

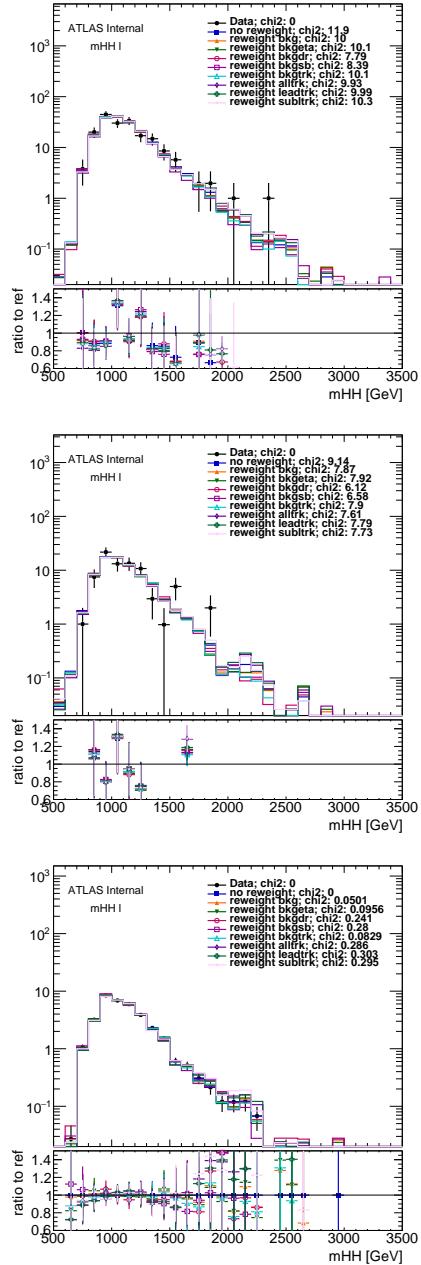


Figure A.28: Reweighted $4b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.

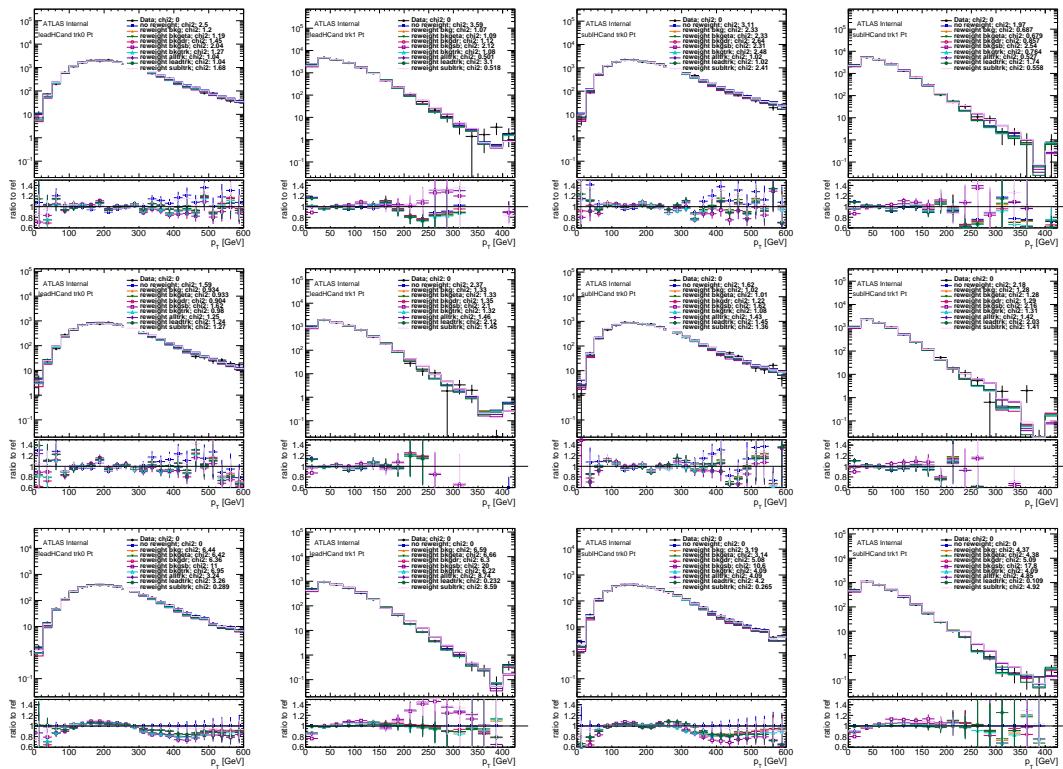


Figure A.29: Reweighted $2b$ s Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for leading Higgs Candidate leading trackjet p_T (first column), leading Higgs Candidate subleading trackjet p_T (second column), subleading Higgs Candidate leading trackjet p_T (third column), subleading Higgs Candidate subleading trackjet p_T (fourth column). The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.

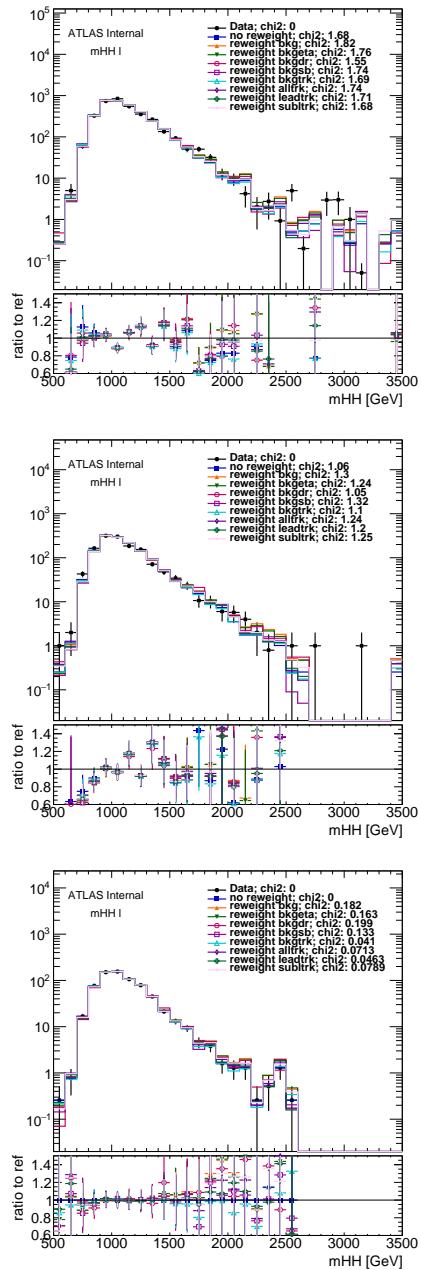


Figure A.30: Reweighted $3b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.

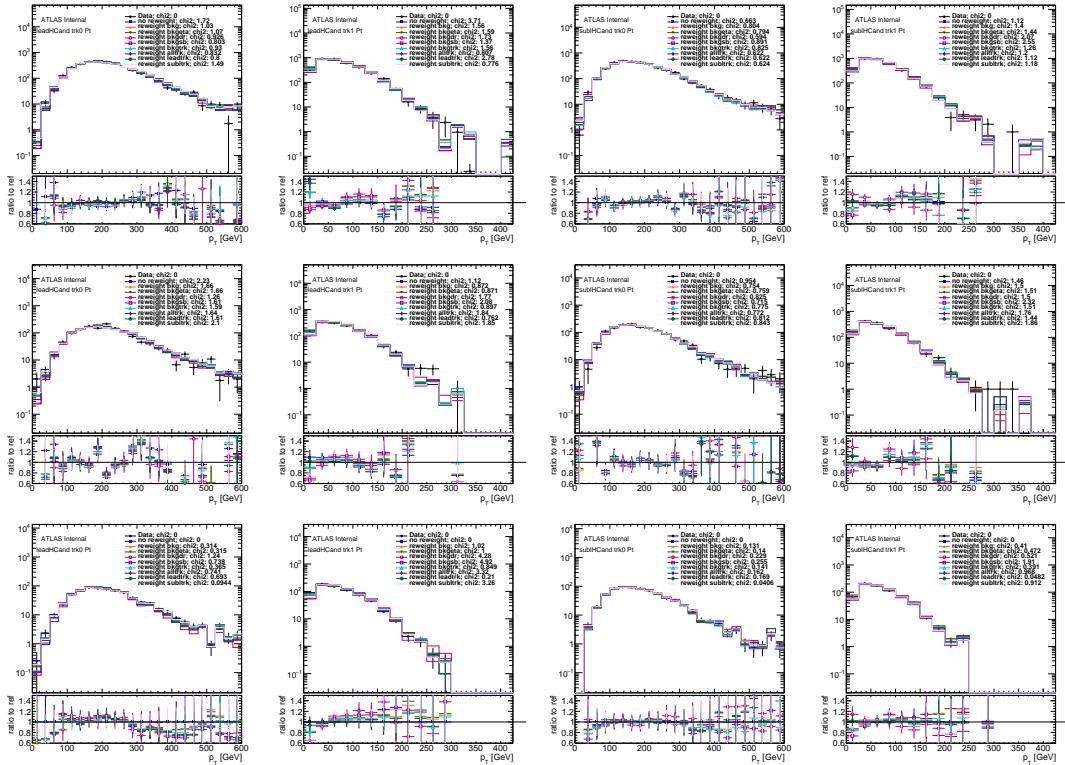


Figure A.31: Reweighted 3 b Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for leading Higgs Candidate leading trackjet p_T (first column), leading Higgs Candidate subleading trackjet p_T (second column), subleading Higgs Candidate leading trackjet p_T (third column), subleading Higgs Candidate subleading trackjet p_T (fourth column). The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.

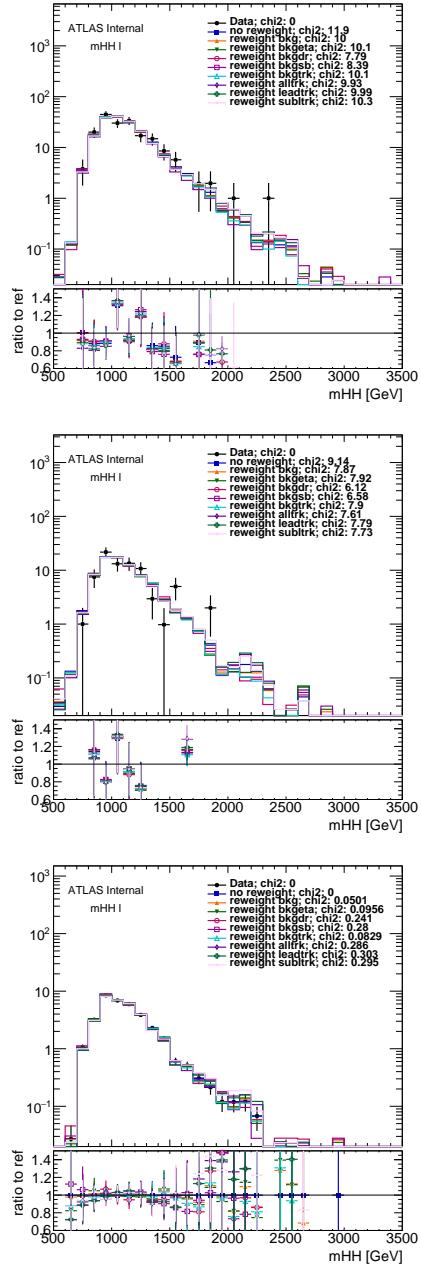


Figure A.32: Reweighted $4b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for MJJ. The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.

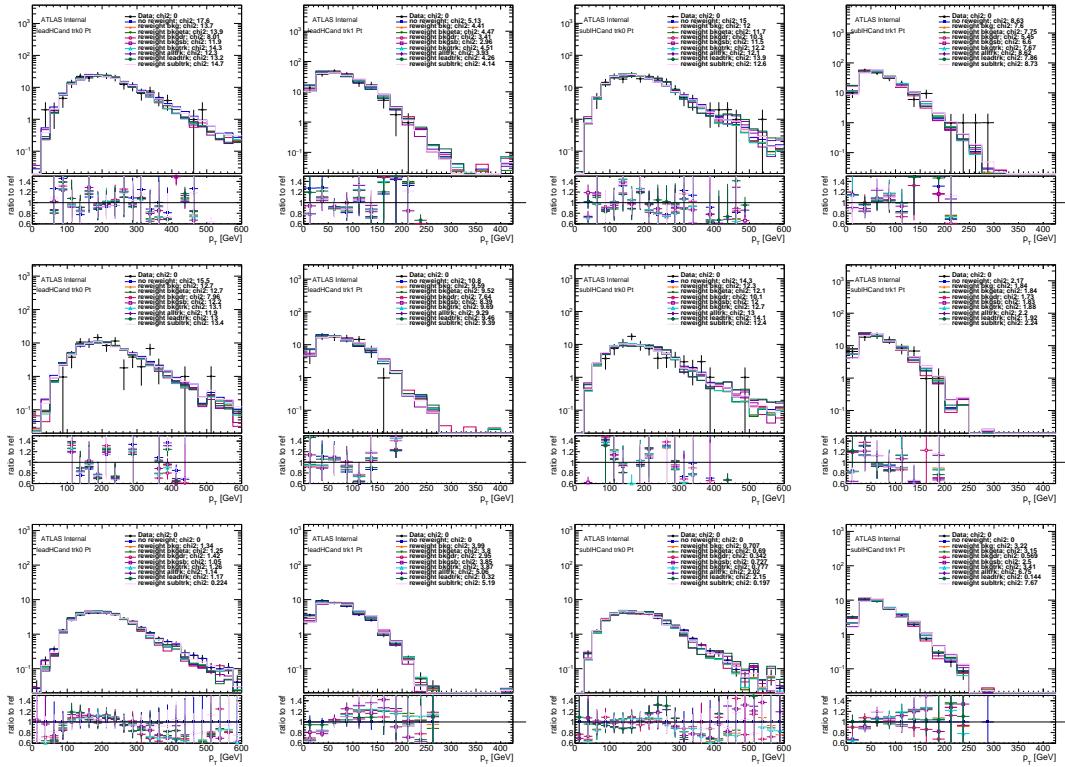


Figure A.33: Reweighted $4b$ Sideband (top)/Control (middle)/Signal(bottom) region predictions comparison, for leading Higgs Candidate leading trackjet p_T (first column), leading Higgs Candidate subleading trackjet p_T (second column), subleading Higgs Candidate leading trackjet p_T (third column), subleading Higgs Candidate subleading trackjet p_T (fourth column). The Signal region is blinded, where the distribution is replaced with the non-reweighted distributions.

A.9 POSSIBLE FUTURE ADDITIONAL CUTS

This section is not going to be applied in the ICHEP results but will be used as a reference for 2017 Moriond Studies.

A.9.1 TRACKJET ΔR CUT

Theoretically, the value of ΔR between the 2 b quarks that decay from a higgs should be roughly $\Delta R = \frac{2m_H}{pT}$ where pT is the transverse momentum of the Higgs and m_H is the mass of the Higgs. Although in practice this formula is not exact, on average the ΔR between the two b's does decrease with increasing pT at a rate of about $1/pT$, as shown in figure A.34. We propose adding a cut that ensures that the ΔR between the two trackjets follows this distribution. This cut effectively cuts on the mass of the parent higgs particle, and is able to do so independently of the energy scale of the trackjets. One complication to this cut is that at high pT , the b-quarks tend to start appearing in the same track jet, and thus doing a cut in this regime on the leading and subleading trackjets would no longer identify b's very well. Thus, we limit this cut to those cases where $pT < 1\text{TeV}$.

We define the cut as: $\text{abs}(\frac{2m}{pT} - \Delta R) < \delta$, where ΔR is between the two trackjets, and m and δ are free parameters to be optimized. Based on figure A.35, we take $m = 142.5$, $\delta = .125$ for the leading large R jet, and $m = 132.5$, $\delta = .125$ for the subleading large R jet. As shown in figure A.36, this results in an improvement in the significance of about 10 – 15% in the boosted regime (RSG Mass $> 1\text{TeV}$).

However, adding this cut on top of the 77% working point for M_{V2C10} leads to very low statis-

tics: there is only one event in the 4tag control region in 6.6 fb^{-1} of 2015 and 2016 data. Therefore, we suggest using this cut in conjunction with switching to the 85% working point. The significance of this combined update is given in A.37.

These two updates together give good stats in the control and sideband regions. However, even after using the reweighting procedure from ICHEP, there is not good agreement between background predictions and data in the various control regions, as shown in figure A.38. To incorporate this cut we would likely have to change the reweighting procedure.

Figure A.34: Monte Carlo results on the ΔR between the truth b's as a result of a Higgs decay (b0 and b1). Lines are drawn at the diameter of large-R (Higgs) jets for the resolved ($\Delta R = 0.8$) and boosted ($\Delta R = 2.0$) analyses.

Figure A.35: ROC plot for values of $m \in \{112.5, 162.5\}$ and $\delta \in \{0, 0.5\}$ for track jets in the subleading and leading large R jets. Signal generated from MC on RSG for masses between 1 and 2 TeV.

Figure A.36: Significance of the analysis with the cut v. without.

A.9.2 LARGE-R JET $\Delta\eta$ CUT

Currently, in the $HH \rightarrow 4b$ analysis, there is a cut on events that have a $\Delta\eta$ between the two jets larger than 1.7. As shown in figure A.39, this cut does a good job excluding background and excluding events with an anomalously high $\Delta\eta$. Additionally, it is clear from figure A.39 that there is little variation of $\Delta\eta$ with the mass of the truth particle, so a mass dependent cut would not be helpful.

Figure A.37: Significance of the analysis with the cut and switching to the 85% working point v. without.

Figure A.38: Distribution of dijet mass in each of the control regions after reweighting.

A.9.3 LEAD TRACKJET pT CUT

Another possible cut to add to the $HH \rightarrow 4b$ analysis would be on the pT of the leading trackjet inside each large R jet. In the regime where the b's start merging into one trackjet, the momentum of both particles from the Higgs decay are concentrated in the lead trackjet, so this trackjet should have a higher percentage of the total momentum from the large R jet. More b's merge where the higgs pT is high and inside the ztag-split region, so we would expect the leading trackjet pT ratio to be higher to be higher in these areas.

As shown in figure A.40, when looking at all signal regions, the distributions of the leading trackjet pT ratio in MC are not different enough from those in data to be able to place a cut. When just looking at the ztag region (in figure A.41), we do see that comparatively more of the MC events have high pT ratios, but still not enough for a cut to be successful. Note that sometimes the ratio of the trackjet pT to the large-R jet pT is bigger than 1 due to mis-measurements in the detector.

Figure A.39: Comparison of $\Delta\eta$ for largeR jets between signal (RSG) and data. Line is drawn at $\Delta\eta = 1.7$, the current value of the cut in the $HH \rightarrow 4b$ analysis.

Figure A.40: Ratio of large-R jet pT contained in leading trackjet vs. large-R jet pT . Monte Carlo plots are generated by a series of different RSG masses between 1 and 3TeV and contain events only from the jet mass signal region. Data covers 6.69 fb^{-1} from 2015 and 2016 and contains events from all jet mass regions. Both plots have events for all b-tag regions.

Figure A.41: Ratio of large-R jet pT contained in leading trackjet vs. large-R jet pT . Monte Carlo plots are generated by a series of different RSG masses between 1 and 3TeV and contain events only from the jet mass signal region. Data covers 6.69 fb^{-1} from 2015 and 2016 and contains events from all jet mass regions. Both plots have events only from the 2 tag split b-tag region.

A.10 SIGNAL KINEMATIC DISTRIBUTIONS

This is a quick check of the signal region kinematic distributions; it is compared with data from ob Tag signal region, as a rough estimate of possible QCD background shape. Figures A.43, A.44, A.45, are the normalized distributions. All plots shown are RSG samples with $c = 1.0$. Figure A.42 shows the signal sample distribution on the mHH plane.

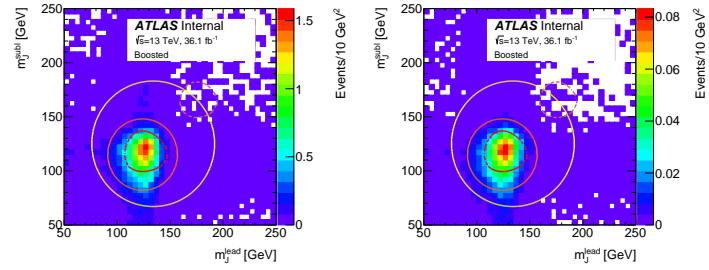


Figure A.42: For RSG $c = 1.0$ samples, number of events as a function of leading Higgs candidate mass and subleading Higgs candidate mass, for 1.2 TeV(left) signal and 2.0 TeV(right) signal samples.

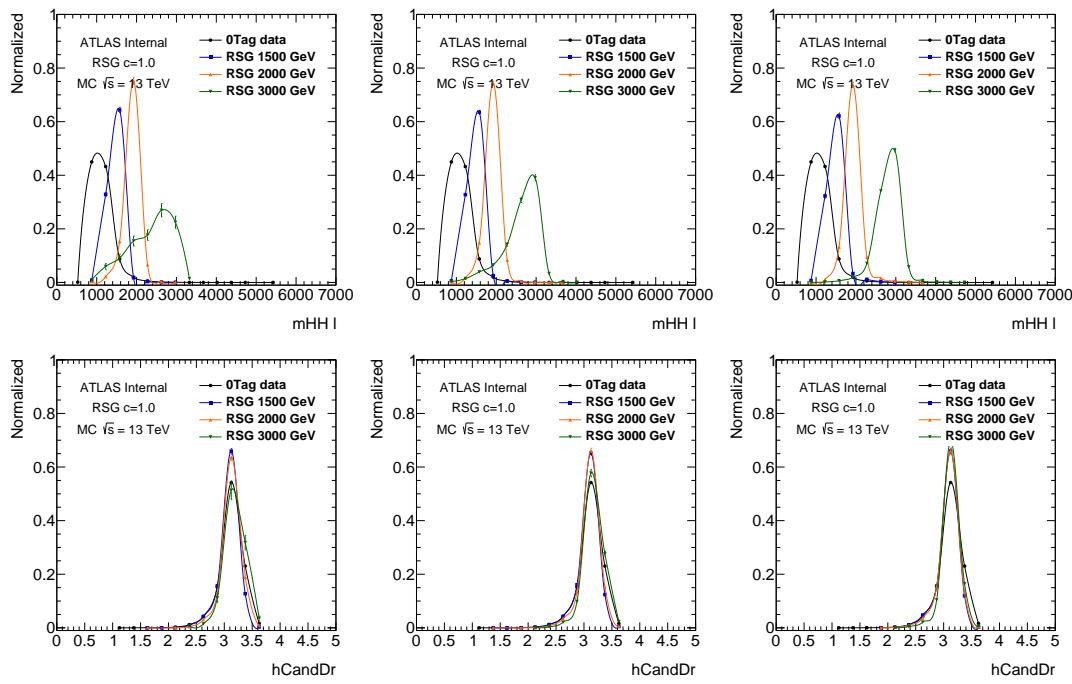


Figure A.43: For RSG $c = 1.0$ samples, the dijet mass distributions and ΔR between the two large- R jets. The left column is $4b$, the middle column is $3b$, and the right column is $2b$.

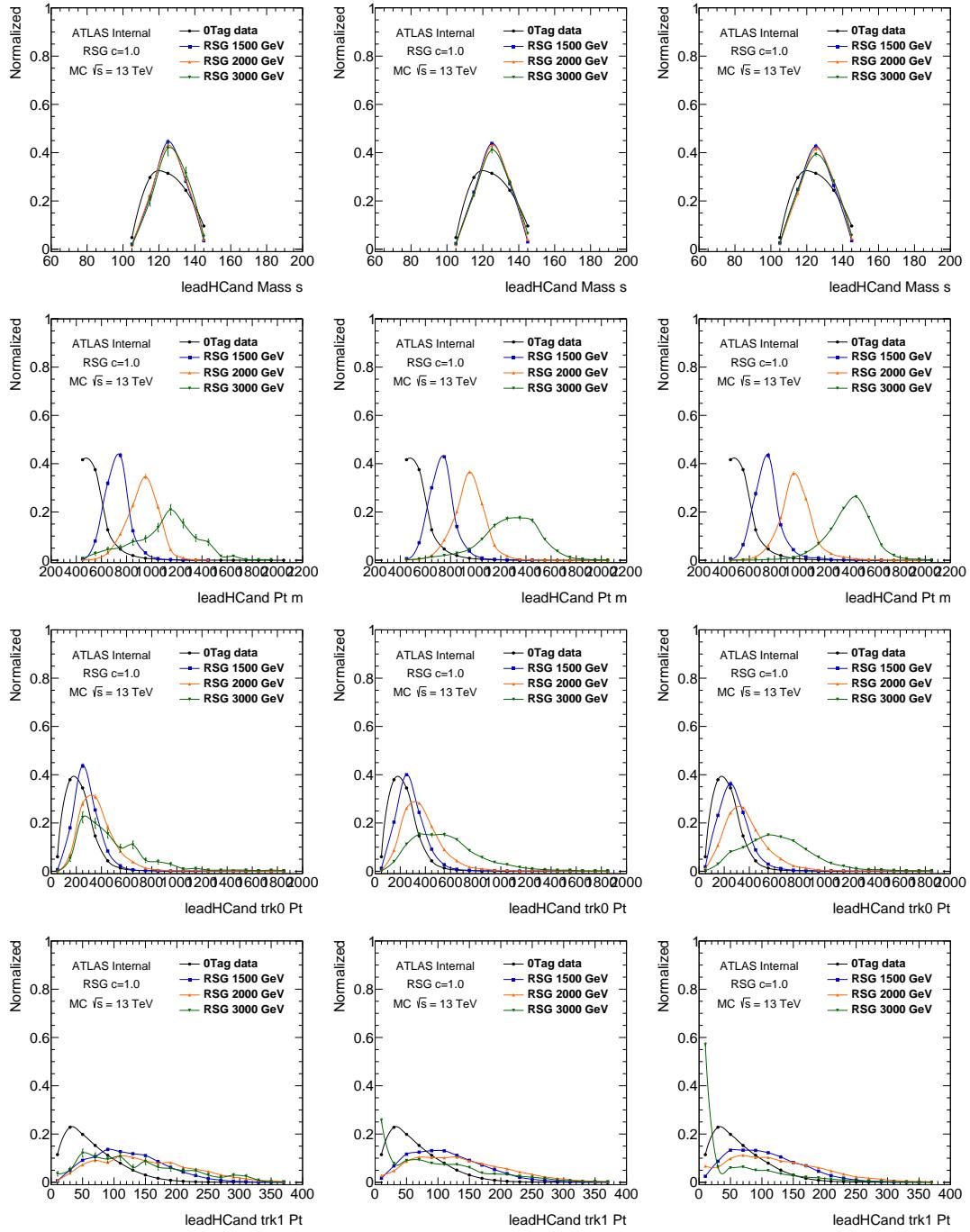


Figure A.44: For RSG $c = 1.0$ samples, the leading higgs candidate's p_T , mass, leading track jet p_T and subleading trackjet p_T . The left column is $4b$, the middle column is $3b$, and the right column is $2b$.

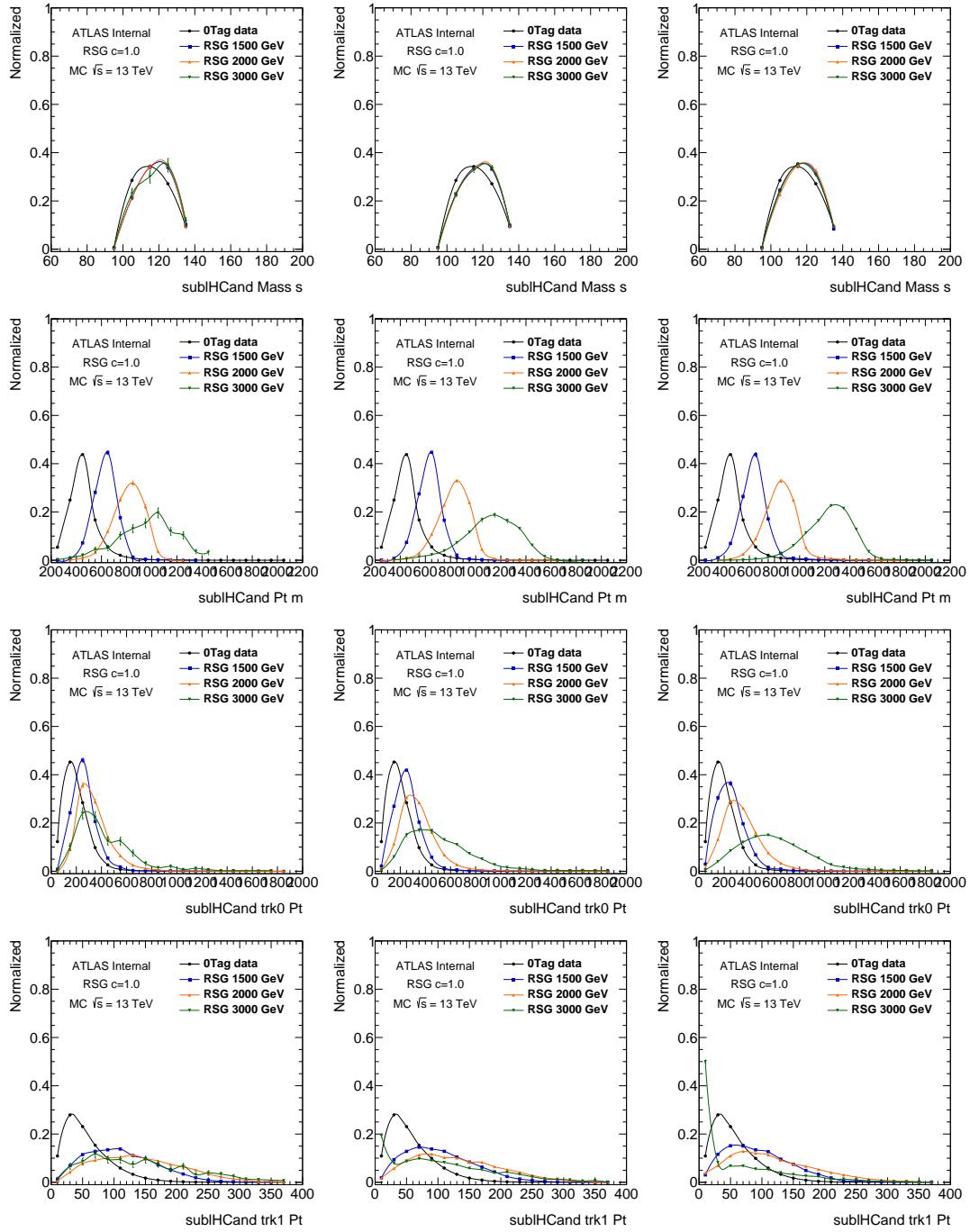


Figure A.45: For RSG $c = 1.0$ samples, the subleading higgs candidate's p_T , mass, leading track jet p_T and subleading trackjet p_T . The left column is $4b$, the middle column is $3b$, and the right column is $2b$ s.

A.II REWEIGHTING FACTOR DISTRIBUTIONS

This section presents the reweighting factors derived. Total reweighting factors after all iterations are shown. $2b$, $3b$ and $4b$ each has its own $1/2b$ Tag region based on the number of minimum track jet requirement, so the three reweights are independent. The reweighting factors for $2b$, $3b$ and $4b$ is shown in Figure A.46. $2b$ and $3b$ show very reasonable distributions with mean value 1. As for $4b$, the stat is limited in the fit so the reweighting factors are distorted. For Sideband regions, detailed 2D distributions, see Figure A.47, A.48 and A.49. For Signal regions, see Figure A.50, A.51, A.52 and A.53.

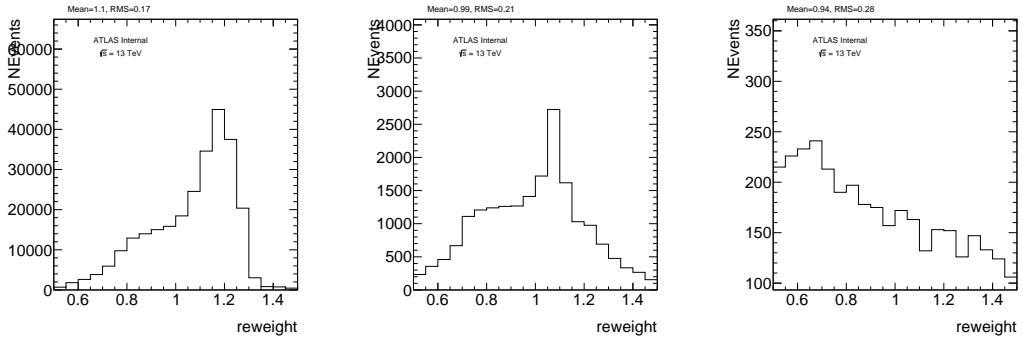


Figure A.46: Sideband Region, left column is $2b$ weights, middle column is $3b$ weights, right column is $4b$ weights.

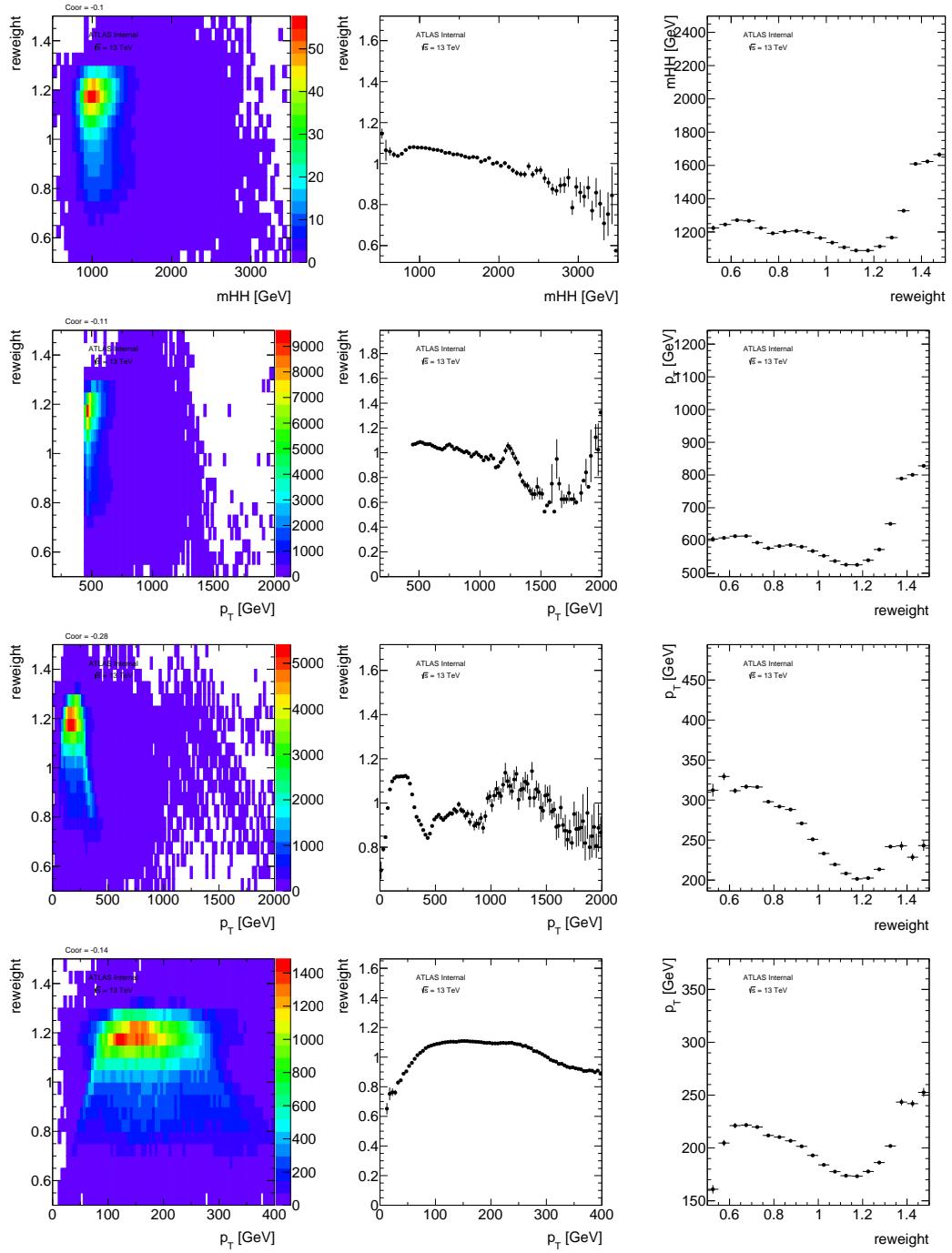


Figure A.47: For $2bs$ Sideband Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is mJJ , the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T .

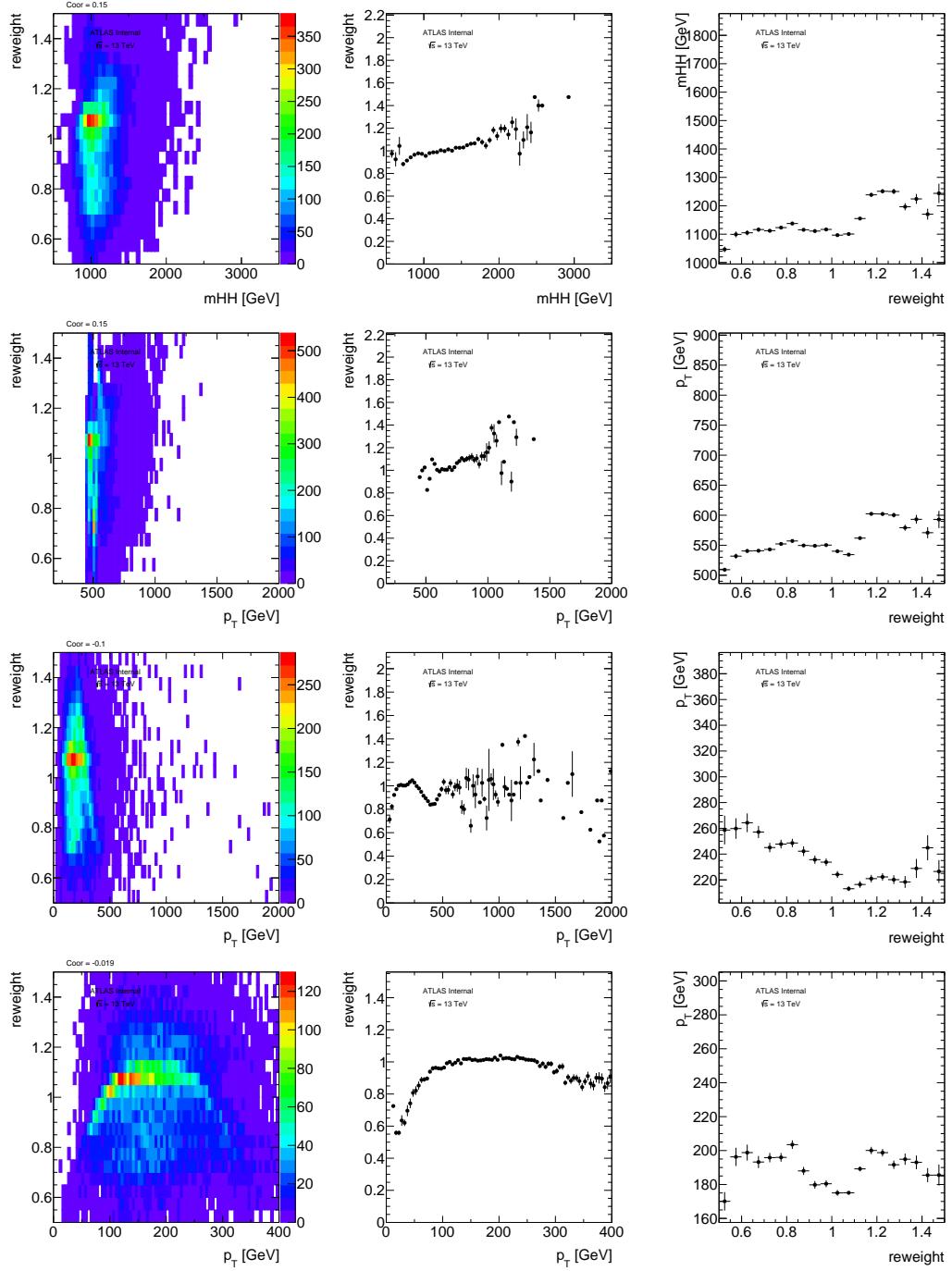


Figure A.48: For 3b Sideband Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T .

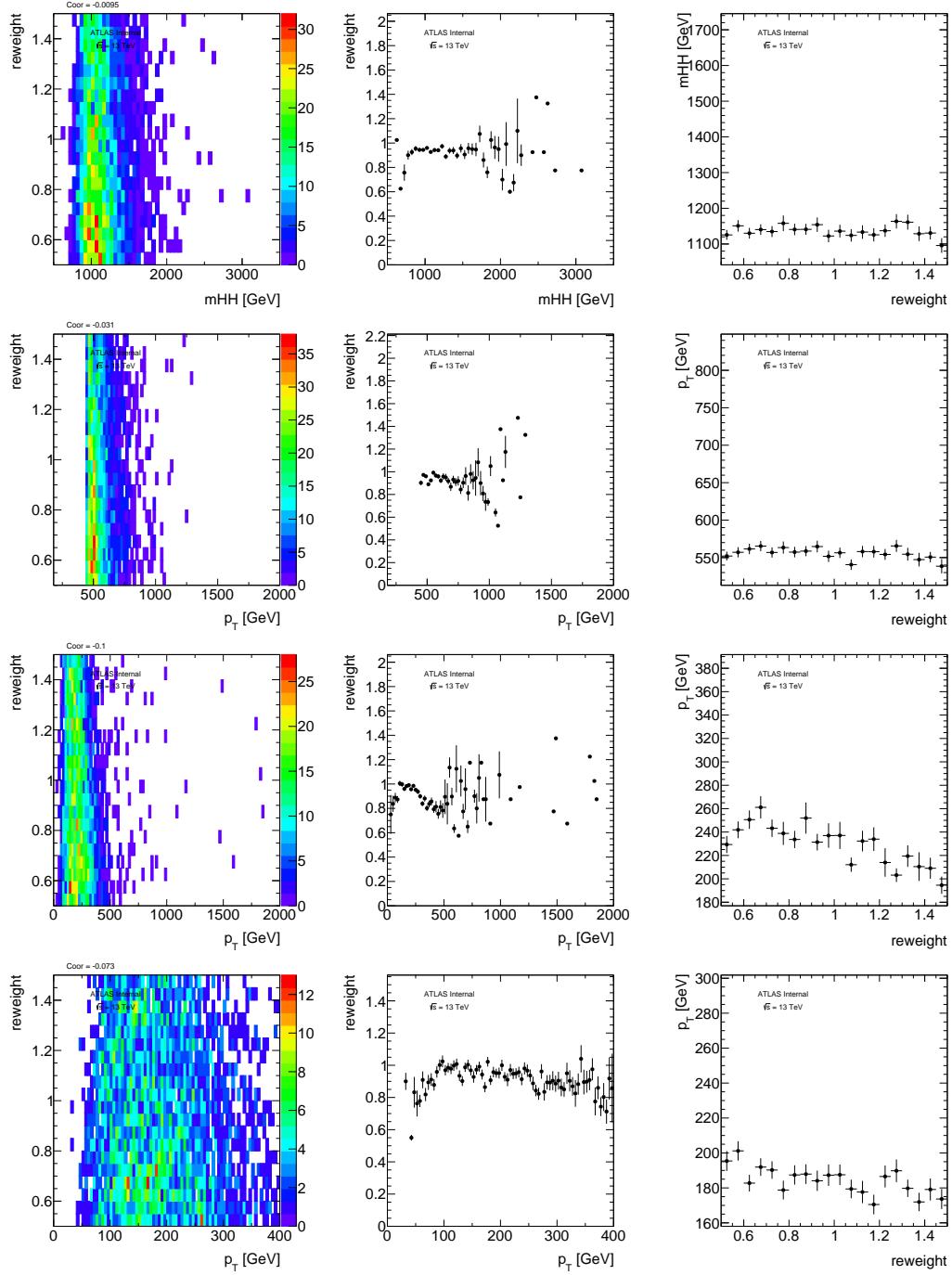


Figure A.49: For 4b Sideband Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is m_{HH} , the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T .

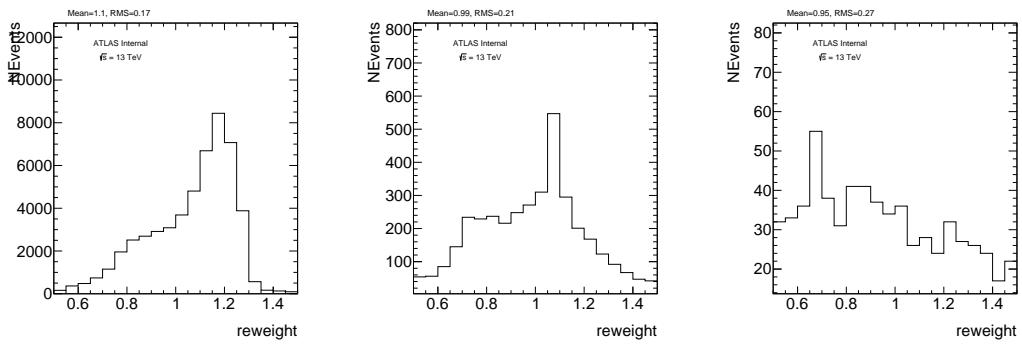


Figure A.50: Signal Region, left column is $2b$ weights, middle column is $3b$ weights, right column is $4b$ weights.

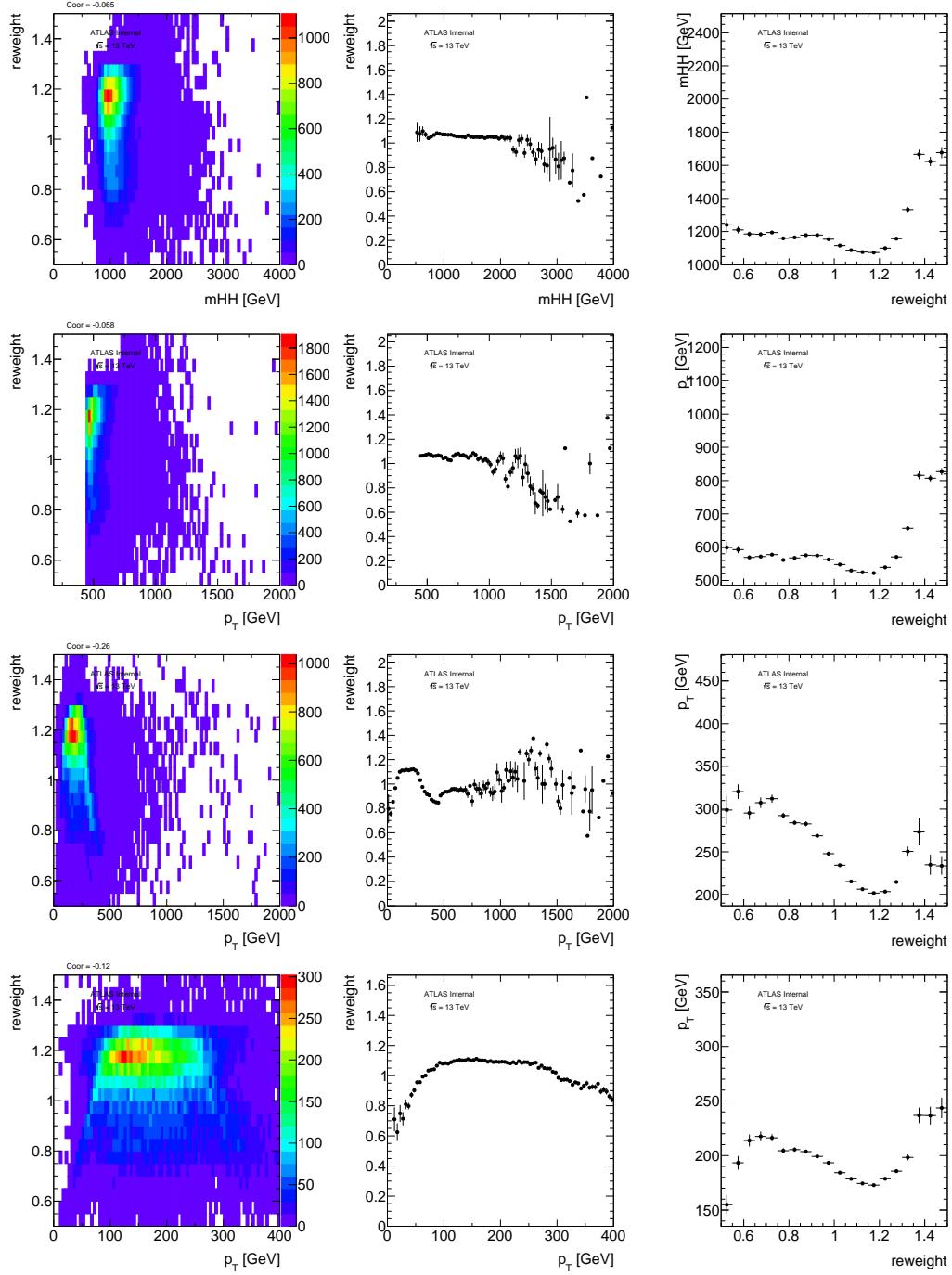


Figure A.51: For $2bs$ Signal Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is mJJ , the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T .

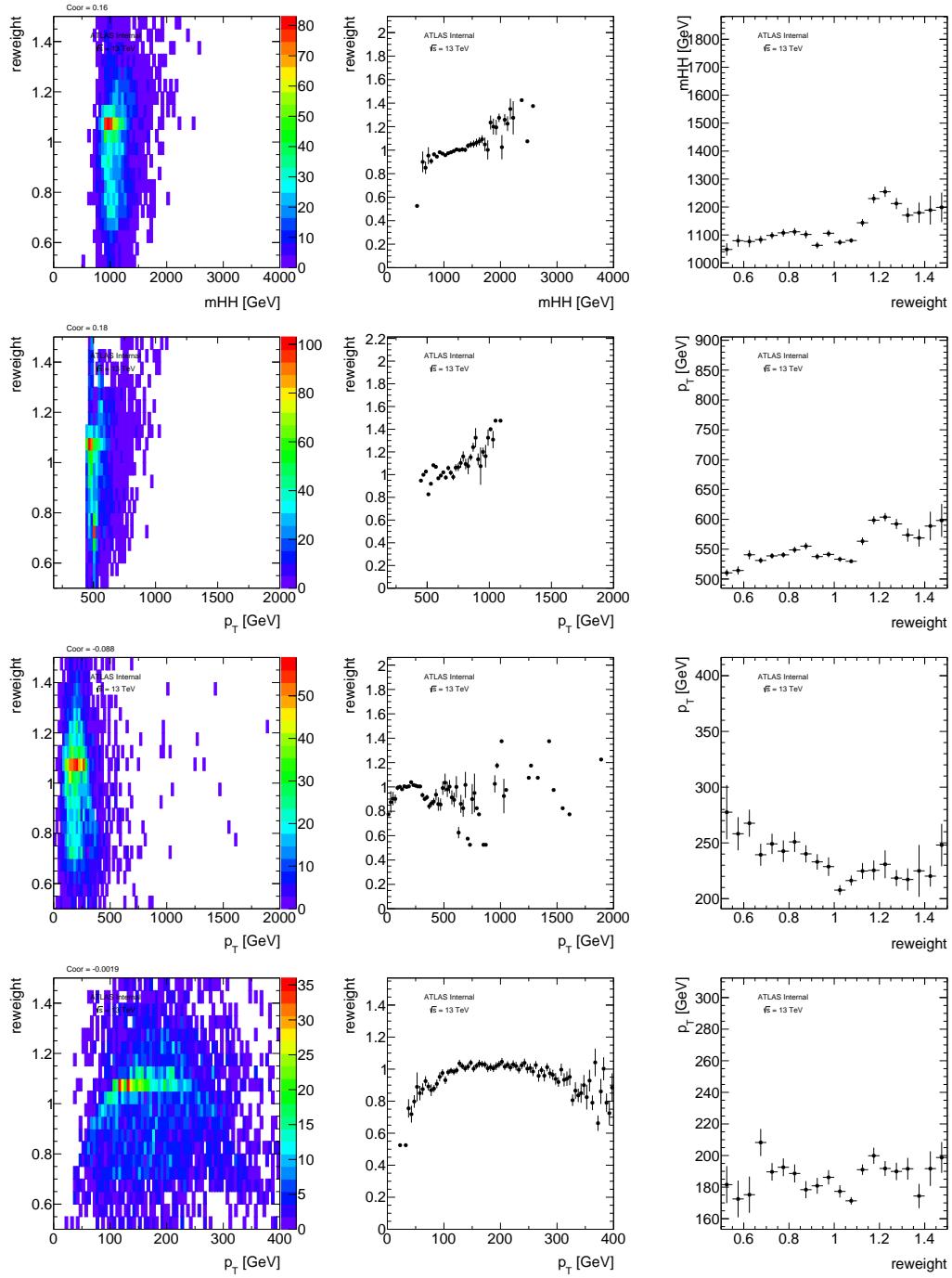


Figure A.52: For 3b Signal Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T .

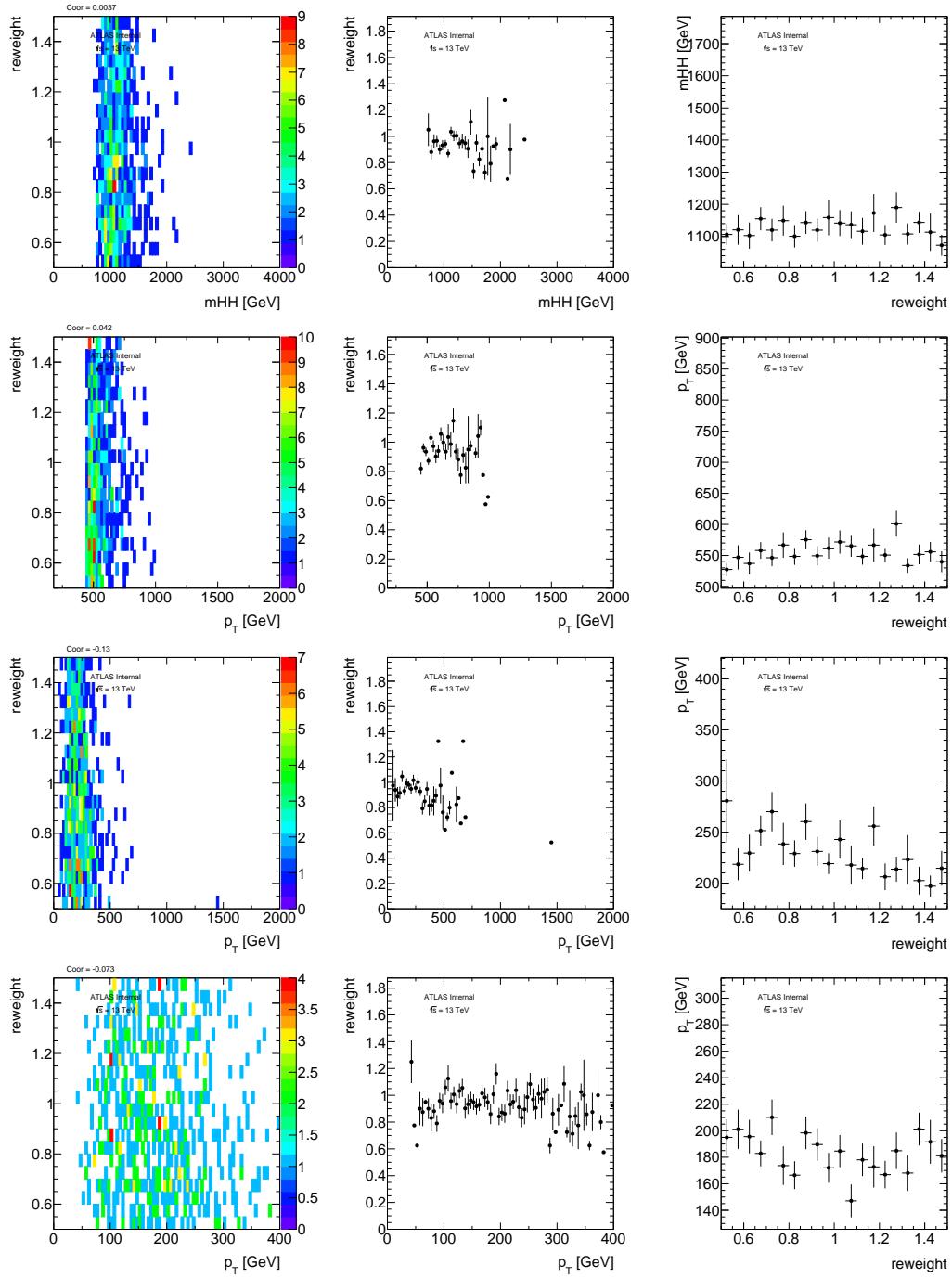


Figure A.53: For 4b Signal Region, left column is the 2D distribution of variable vs. weights, middle column is the profile X plot, right column is the profile Y plot. The top row is MJJ, the second row is leading large- R jet p_T , the third row is the leading large- R jet's leading track jet p_T , and the fourth row is the subleading large- R jet's leading track jet p_T .

A.12 BOOSTED DEBUG STREAM CHECK

To make sure no events are ignored, the 2015 and 2016 debug stream data is also analyzed. The list of 2016 debug stream samples are:

```
data15_13TeV.00279928.debugrec_hlt.merge.DAOD_EXOT8.g49_r7562_ContE_p2982  
data15_13TeV.00267639.debugrec_hlt.merge.DAOD_EXOT8.g17_r7600_ContA_p2982  
data15_13TeV.00281411.debugrec_hlt.merge.DAOD_EXOT8.g49_r7562_ContH_p2982  
data15_13TeV.00280368.debugrec_hlt.merge.DAOD_EXOT8.g49_r7562_ContF_p2982  
data15_13TeV.00281075.debugrec_hlt.merge.DAOD_EXOT8.g49_r7562_ContG_p2982  
data15_13TeV.00284484.debugrec_hlt.merge.DAOD_EXOT8.g49_r7562_ContJ_p2982  
data15_13TeV.00271744.debugrec_hlt.merge.DAOD_EXOT8.g49_r7600_ContC_p2982  
data15_13TeV.00276954.debugrec_hlt.merge.DAOD_EXOT8.g49_r7562_ContD_p2982  
data16_13TeV.00310216.debugrec_hlt.merge.DAOD_EXOT8.g53_f750_m1689_ContH_p2982  
data16_13TeV.00300908.debugrec_hlt.merge.DAOD_EXOT8.g53_f711_m1620_ContB_p2982  
data16_13TeV.00302393.debugrec_hlt.merge.DAOD_EXOT8.g53_f716_m1620_ContC_p2982  
data16_13TeV.00303560.debugrec_hlt.merge.DAOD_EXOT8.g53_f716_m1620_ContD_p2982  
data16_13TeV.00303892.debugrec_hlt.merge.DAOD_EXOT8.g53_f716_m1620_ContE_p2982  
data16_13TeV.00311481.debugrec_hlt.merge.DAOD_EXOT8.g53_f758_m1710_ContL_p2982  
data16_13TeV.00304494.debugrec_hlt.merge.DAOD_EXOT8.g53_f716_m1620_ContF_p2982  
data16_13TeV.00309759.debugrec_hlt.merge.DAOD_EXOT8.g53_f750_m1689_ContK_p2982  
data16_13TeV.00306657.debugrec_hlt.merge.DAOD_EXOT8.g53_f735_m1659_ContG_p2982  
data16_13TeV.00308084.debugrec_hlt.merge.DAOD_EXOT8.g53_f741_m1673_ContI_p2982  
data16_13TeV.00299390.debugrec_hlt.merge.DAOD_EXOT8.g49_f701_m1594_ContA_p2982
```

Trigger requirement is waived for this test. The cutflow table for 2015 and 2016 is shown respectively in Table A.1 and A.2. Comparing this with Table 6.3, debug stream contribution to background estimation and signal yield should be zero. After unblinding, this table is double checked no

events show up in signal region as expected.

Cut	Data
Initial	1639
Pass GRL	890
Pass trigger	890
$N(\text{fiducial large-R jets}) \geq 2$	591
leading large-R jet $p_T > 450 \text{ GeV}$	504
$ \Delta\eta(JJ) < 1.7$	370
Pass at least 2 bjets	18
<hr/>	
2 b oneside, $X_{bb} < 1.6$	1
2 b oneside, $X_{bb} > 1.6$ and $R_{bb} < 33$	1
2 b oneside, $R_{bb} > 33$	11
<hr/>	
2 b split, $X_{bb} < 1.6$	0
2 b split, $X_{bb} > 1.6$ and $R_{bb} < 33$	0
2 b split, $R_{bb} > 33$	5
<hr/>	
3 b-tags, $X_{bb} < 1.6$	0
3 b-tags, $X_{bb} > 1.6$ and $R_{bb} < 33$	0
3 b-tags, $R_{bb} > 33$	0
<hr/>	
4 b-tags, $X_{bb} < 1.6$	0
4 b-tags, $X_{bb} > 1.6$ and $R_{bb} < 33$	0
4 b-tags, $R_{bb} > 33$	0

Table A.1: Cutflow of debug stream data in 2015. No events found in signal regions.

Cut	Data
Initial	68
Pass GRL	63
Pass trigger	63
$N(\text{fiducial large-R jets}) \geq 2$	63
leading large-R jet $p_T > 450 \text{ GeV}$	32
$ \Delta\eta(JJ) < 1.7$	15
Pass at least 2 bjets	0
<hr/>	
2 b oneside, $X_{bb} < 1.6$	0
2 b oneside, $X_{bb} > 1.6$ and $R_{bb} < 33$	0
2 b oneside, $R_{bb} > 33$	0
<hr/>	
2 b split, $X_{bb} < 1.6$	0
2 b split, $X_{bb} > 1.6$ and $R_{bb} < 33$	0
2 b split, $R_{bb} > 33$	0
<hr/>	
3 b-tags, $X_{bb} < 1.6$	0
3 b-tags, $X_{bb} > 1.6$ and $R_{bb} < 33$	0
3 b-tags, $R_{bb} > 35.8$	0
<hr/>	
4 b-tags, $X_{bb} < 1.6$	0
4 b-tags, $X_{bb} > 1.6$ and $R_{bb} < 33$	0
4 b-tags, $R_{bb} > 33$	0

Table A.2: Cutflow of debug stream data in 2016. No events found in signal regions.

References

- [1] Georges Aad et al. Performance of b -Jet Identification in the ATLAS Experiment. *JINST*, 11 (04):Po4008, 2016. doi: 10.1088/1748-0221/11/04/Po4008.
- [2] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. doi: 10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z. [Erratum: Eur. Phys. J.C73,2501(2013)].