

1 An MVA Theme and Variations on
2 $ZH \rightarrow llbb$ with the ATLAS Detector at
3 $\sqrt{s} = 13 \text{ TeV}$

4 A DISSERTATION PRESENTED
5 BY
6 STEPHEN K. CHAN
7 TO
8 THE DEPARTMENT OF PHYSICS

9 IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
10 FOR THE DEGREE OF
11 DOCTOR OF PHILOSOPHY
12 IN THE SUBJECT OF
13 PHYSICS

14 HARVARD UNIVERSITY
15 CAMBRIDGE, MASSACHUSETTS
16 2018

¹⁷ ©2017 – STEPHEN K. CHAN

¹⁸ ALL RIGHTS RESERVED.

19 **An MVA Theme and Variations on $ZH \rightarrow llb\bar{b}$ with the**
20 **ATLAS Detector at $\sqrt{s} = 13$ TeV**

21 **ABSTRACT**

22 *This is a work in progress: All the material is here, but some connective tissue (for the Combina-*
23 *tion chapter in particular) and needed revision is missing.*

24 This thesis describes variations on the two lepton channel of the Run-2 search for the SM Higgs
25 boson produced in association with a vector boson using different variable sets for MVA training.
26 The three variable sets in question are the set of variables from the fiducial analysis, a set based on
27 the Lorentz Invariants (LI) concept, and a set based on a combination of masses and decay angles
28 derived using the RestFrames (RF) package. Aside from the variable sets used for MVA training and
29 discriminant distributions, the analysis is otherwise identical to the fiducial analysis. Both the LI
30 and RF sets perform competitively on the basis of significances, with the RF set showing a $\sim 3.5\%$
31 improvement in expected fits to Asimov and data, though neither set boosts observed significance.
32 Both sets also reduce the observed error on $\hat{\mu}$, with the LI set reducing the error due to systematics
33 by 7.5% and the RF set doing so by 16%.

Contents

34

35	o	INTRODUCTION	I
36	I	THE LARGE HADRON COLLIDER AND THE ATLAS DETECTOR	3
37	1.1	The CERN Accelerator Complex	4
38	1.2	The Large Hadron Collider	4
39	1.3	ATLAS at a Glance	10
40	1.4	The Inner Detector	15
41	1.5	The ATLAS Calorimeters	19
42	1.6	The Muon Spectrometer	27
43	2	THE STANDARD MODEL HIGGS AND COLIDER EVENT VARIABLES	33
44	2.1	The Standard Model Higgs Boson	34
45	2.2	Higgs Boson Production and Decay at the Large Hadron Collider	37
46	2.3	Collider Events and Event Level Variables	40
47	2.4	Characterization with Event-Level Variables	42
48	2.5	Lorentz Invariants	44
49	2.6	RestFrames Variables	46
50	2.7	Extensions to the τ and μ Lepton Channels	48
51	3	DATA AND SIMULATED SAMPLES	50
52	4	SIGNAL AND BACGKROUND MODELING	52
53	4.1	Event Generation In a Nutshell	53
54	4.2	Description of Modeling Uncertainty Categories	55
55	4.3	Process Specific Systematic Summaries	63
56	5	OBJECT AND EVENT RECONSTRUCTION AND SELECTION	73
57	5.1	Triggers	74
58	5.2	Electrons	77
59	5.3	Muons	79
60	5.4	Missing Transverse Energy	81
61	5.5	Jets	81
62	5.6	Miscellania and Sytematics Summary	100
63	5.7	Event Selection and Analysis Regions	100

64	6 MULTIVARIATE ANALYSIS CONFIGURATION	103
65	6.1 Training Samples and Variable Selection	104
66	6.2 MVA Training	115
67	6.3 Statistics Only BDT Performance	120
68	7 STATISTICAL FIT MODEL AND VALIDATION	126
69	7.1 The Fit Model	128
70	7.2 Fit Inputs	130
71	7.3 Systematic Uncertainties Review	131
72	7.4 The VZ Validation Fit	134
73	7.5 Nuisance Parameter Pulls	138
74	7.6 Postfit Distributions	145
75	7.7 VH Fit Model Validation	145
76	8 FIT RESULTS	166
77	9 MEASUREMENT COMBINATIONS	172
78	9.1 The Combined Fit Model	173
79	9.2 Combined Fit Results	188
80	10 CLOSING THOUGHTS	196
81	APPENDIX A MICROMEGAS TRIGGER PROCESSOR SIMULATION	199
82	A.1 Algorithm Overview	200
83	A.2 Monte Carlo Samples	205
84	A.3 Nominal Performance	205
85	A.4 Fit Quantities	206
86	A.5 Efficiencies	212
87	A.6 Incoherent Background	215
88	A.7 BCID	220
89	A.8 Charge Threshold	220
90	A.9 Misalignments and Corrections	224
91	A.10 Individual Cases	229
92	A.11 $ds \neq 0$	230
93	A.12 $dz \neq 0$	230
94	A.13 $dt \neq 0$	230
95	A.14 $\alpha \neq 0$	231
96	A.15 $\beta \neq 0$	231
97	A.16 $\gamma \neq 0$	231
98	A.17 Simulation Correction of the Algorithm Under Nominal Conditions	234
99	A.18 Translation Misalignments Along the Horizontal Strip Direction (Δs)	237

100	A.19 Translation Misalignments Orthogonal to the Beamline and Horizontal Strip Direction (Δz)	239
101		
102	A.20 Translation Misalignments Parallel to the Beamline (Δt)	241
103		
104	A.21 Chamber Tilts Towards and Away from the IP (γ_s Rotation)	242
105		
106	A.22 Rotation Misalignments Around the Wedge Vertical Axis (β_z)	244
105	A.23 Rotation Misalignments Around the Axis Parallel to the Beamline (α_t)	245
106		
	A.24 Conclusion	247
107	APPENDIX B TELESCOPING JETS	249
108		
109	B.1 Monte Carlo Simulation	250
110		
111	B.2 Jet Reconstruction and Calibration	251
112		
113	B.3 Event Reconstruction and Selection	252
114		
115	B.4 Validation of Jet Calibration	253
116		
117	B.5 Truth-Level Analysis	258
118		
119	B.6 Errors on Telescoping Significances	261
120		
	B.7 Counting	261
115	B.8 Multiple Event Interpretations	262
116		
117	B.9 $t(z) = z$	263
118		
119	B.10 $t(z) = \rho_S(z) / \rho_B(z)$	264
120		
	B.11 Reconstructed-Level Analysis	269
	B.12 Conclusions and Prospects	275
120	REFERENCES	281

Acknowledgments

¹²³ THIS THESIS WOULD NOT HAVE BEEN POSSIBLE without large amounts of espresso.

Your life has a limit but knowledge has none you understand this and still strive for knowledge, you will be in danger for certain!

Zhuangzi

0

124

125

Introduction

126 SINCE THE DISCOVERY of a Standard Model (SM) like Higgs boson at the LHC in 2012²¹, one of
127 the main outstanding physics goals of the LHC has been to observe the primary SM Higgs decay
128 mode, $H \rightarrow b\bar{b}$, with efforts primarily targeted at searching for Higgs bosons produced in associa-
129 tion with a leptonically decaying vector (W or Z , denoted generically as V) boson. As the integrated

130 luminosity of data collected at the LHC increases, $H \rightarrow b\bar{b}$ searches will increasingly become limited
131 by the ability to constrain systematic uncertainties, with the latest result from ATLAS at $\sqrt{s} = 13$
132 TeV using 36.1 fb^{-1} of pp collision data already approaching this regime, having a $VH(b\bar{b})$ signal
133 strength of $1.20^{+0.24}_{-0.23}(\text{stat.})^{+0.34}_{-0.28}(\text{syst.})$ at $m_H = 125 \text{ GeV}$ ³⁴.

134 While this effort will likely require a combination of several different methods at various different
135 stages in the analysis chain, one possible avenue forward is to revise the multivariate analysis (MVA)
136 discriminant input variables used, as various schemes offer the promise of reducing systematic uncer-
137 tainties through more efficient use of both actual and simulated collision data. This thesis discusses
138 two such alternate MVA schemes, the RestFrames (RF) and Lorentz Invariants (LI) variables, in the
139 context of the 2-lepton channel of the Run 2 analysis in³⁴ and⁵⁶, henceforth referred to as the “fidu-
140 cial analysis,” before a brief discussion of combinations across channels and datasets.

141 Data and simulation samples used are described in Section ??, and event reconstruction defini-
142 tions and event selection requirements are outlined in Section ??.. The multivariate analysis, includ-
143 ing a description of the LI and RF variable sets and a summary of performance in the absence of
144 systematic uncertainties, is described in Section ??.. The statistical fit model and systematic uncertain-
145 ties are described in Section ??, and the fit results may be found in Section ??.. Combining channels
146 and datasets at different \sqrt{s} values is discussed in the context of the Run 1 + Run 2 SM $VH(b\bar{b})$
147 combination in Chapter 9. Finally, conclusions and closing thoughts are presented in Section A.24.

148 Editorial notes:

- 149 1. pdf will be *probability* distribution function
150 2. PDF will be *parton* distribution function

Noli turbare circulos meos

Archimedes

1

¹⁵¹

¹⁵² The Large Hadron Collider and the ATLAS

Detector

¹⁵³

¹⁵⁴ THE CERN ACCELERATOR COMPLEX AND ITS EXPERIMENTS stand as a testament to human in-

¹⁵⁵ genuity and its commitment to the pursuit of fundamental knowledge. In this chapter, we give a

¹⁵⁶ cursory overview of the CERN accelerator complex, including the Large Hadron Collider (LHC),
¹⁵⁷ before moving on to a more detailed review of the ATLAS detector.

¹⁵⁸ **I.I THE CERN ACCELERATOR COMPLEX**

¹⁵⁹ The journey of protons from hydrogen canister to high energy collisions through the CERN acceler-
¹⁶⁰ ator complex, illustrated in Figure I.I, is also one through the history of CERN’s accelerator program.
¹⁶¹ After being ionized in an electric field, protons are first accelerated in a linear accelerator, LINAC 2*,
¹⁶² to a kinetic energy of 50 MeV. From there, they are fed into the Proton Synchotron Booster†, which
¹⁶³ further accelerates them to 1.4 GeV and, as its name implies, feeds them to the 628 m Proton Syn-
¹⁶⁴ chotron (PS, 1959³) and up to 25 GeV. The penultimate stage is the 7 km Super Proton Synchotron
¹⁶⁵ (SPS, 1976; responsible for the discovery of the W and Z bosons and the 1983 Nobel Prize⁵), which
¹⁶⁶ accelerates the protons to a kinetic energy of 450 GeV. Finally, these 450 GeV protons are injected
¹⁶⁷ into the LHC⁴², a proton-proton collider housed in the 27 km circumference tunnel that housed the
¹⁶⁸ Large Electron Positron Collider (LEP) before its operations ceased in 2000.

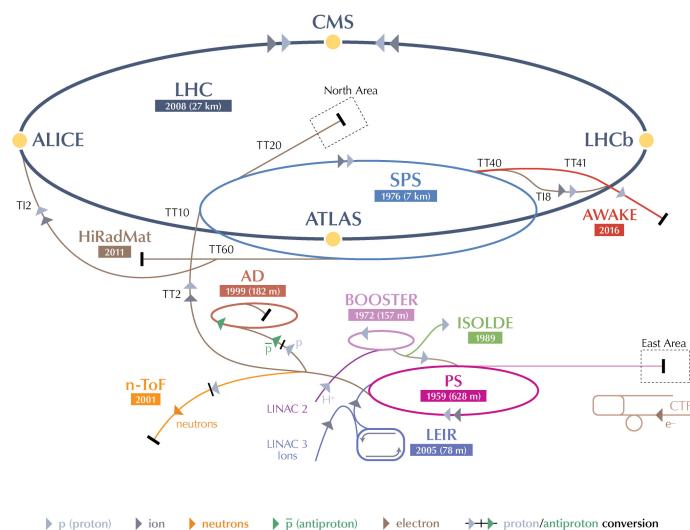
¹⁶⁹ **I.2 THE LARGE HADRON COLLIDER**

¹⁷⁰ The LHC was designed to function primarily as a proton-proton collider with a center of mass en-
¹⁷¹ ergy $\sqrt{s} = 14$ TeV and an instantaneous luminosity of $1 \times 10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$, though it is also capable of
¹⁷² producing heavy ion (Pb-Pb) collisions, which it does for approximately one month in a typical year

* 1978’s LINAC 2 is the successor to 1959’s LINAC 1; it will be replaced in 2020 by LINAC 4; LINAC 3 is responsible for ion production.

† Protons can be directly from a LINAC into the PS, but the higher injection energy allows for approximately 100 times more protons to be used at once⁴, 1972.

CERN's Accelerator Complex



LHC Large Hadron Collider SPS Super Proton Synchrotron PS Proton Synchrotron

AD Antiproton Decelerator CTF3 Clic Test Facility AWAKE Advanced WAKefield Experiment ISOLDE Isotope Separator OnLine Dvice
LEIR Low Energy Ion Ring LINAC LINear ACcelerator n-ToF Neutrons Time Of Flight HiRadMat High-Radiation to Materials

©CERN 2013

Figure 1.1: The CERN Accelerator Complex⁵⁵

¹⁷³ of physics collisions. Owing to an accident at the beginning of the LHC's initial run, the accelerator
¹⁷⁴ has operated at center of mass energies of 7, 8, and now 13 TeV.

¹⁷⁵ The limited size of the LEP tunnel (~ 3.6 m) means that it is impractical to have separate rings
¹⁷⁶ and magnet systems for each proton beam (proton-antiproton colliders like the Tevatron do not face
¹⁷⁷ this complication and can have both beams circulating in the same beam pipe), so the LHC magnets
¹⁷⁸ are coupled in a "twin bore" design. The LHC magnets make use of superconducting NbTi cables
¹⁷⁹ and are cooled using superfluid helium to a temperature of 2 K, which allows for operational field
¹⁸⁰ strengths in excess of 8 T. The layout of an LHC dipole magnet is shown in Figure 1.2. These dipole
¹⁸¹ magnets are responsible for bending the LHC's proton beams, and their strength is the principal
¹⁸² limiting factor in the center of mass energy achievable at a circular collider.

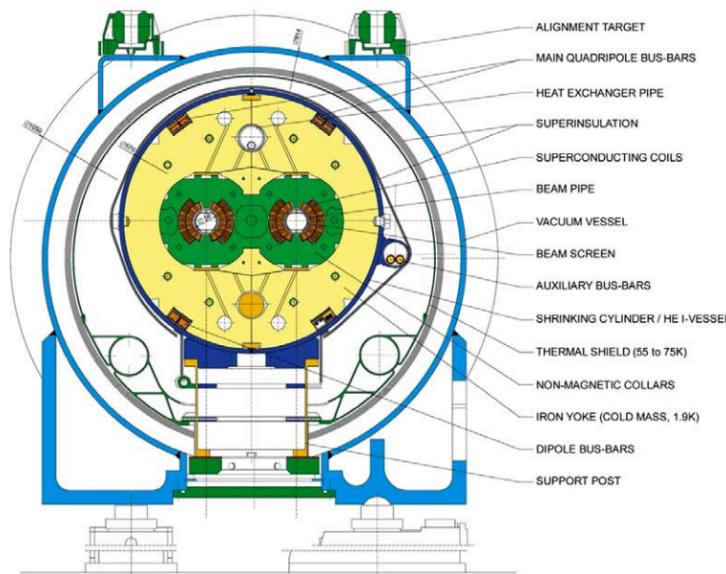


Figure 1.2: Schematic drawing of an LHC dipole magnet and cryogenics system.

¹⁸³ In addition to the dipole magnets, there are quadrupole magnet assemblies in the short straight
¹⁸⁴ sections (for beam focusing), as well as quadrupole, octupole, and sextupole magnets interspersed
¹⁸⁵ throughout the length of the LHC ring for beam stabilization and other higher order corrections.

¹⁸⁶ The interior of the LHC beam pipe operates at a nominal pressure of $\sim 10^{-7}$ Pa, famously more
¹⁸⁷ rarefied than outer space.

¹⁸⁸ The LHC ring itself is between 45 m and 170 m below ground and has a 1.4% incline towards
¹⁸⁹ Lac Léman with eight arcs and eight straight sections. In the middle of each of the eight straight
¹⁹⁰ sections, there are potential interaction points (each colloquially referred to by its number as “Point
¹⁹¹ N”), with each point housing either accelerator infrastructure or an experiment. A schematic of the
¹⁹² contents of each component, as well as a more detailed view of the infrastructure in the LHC ring,
¹⁹³ can be found in Figure 1.3.

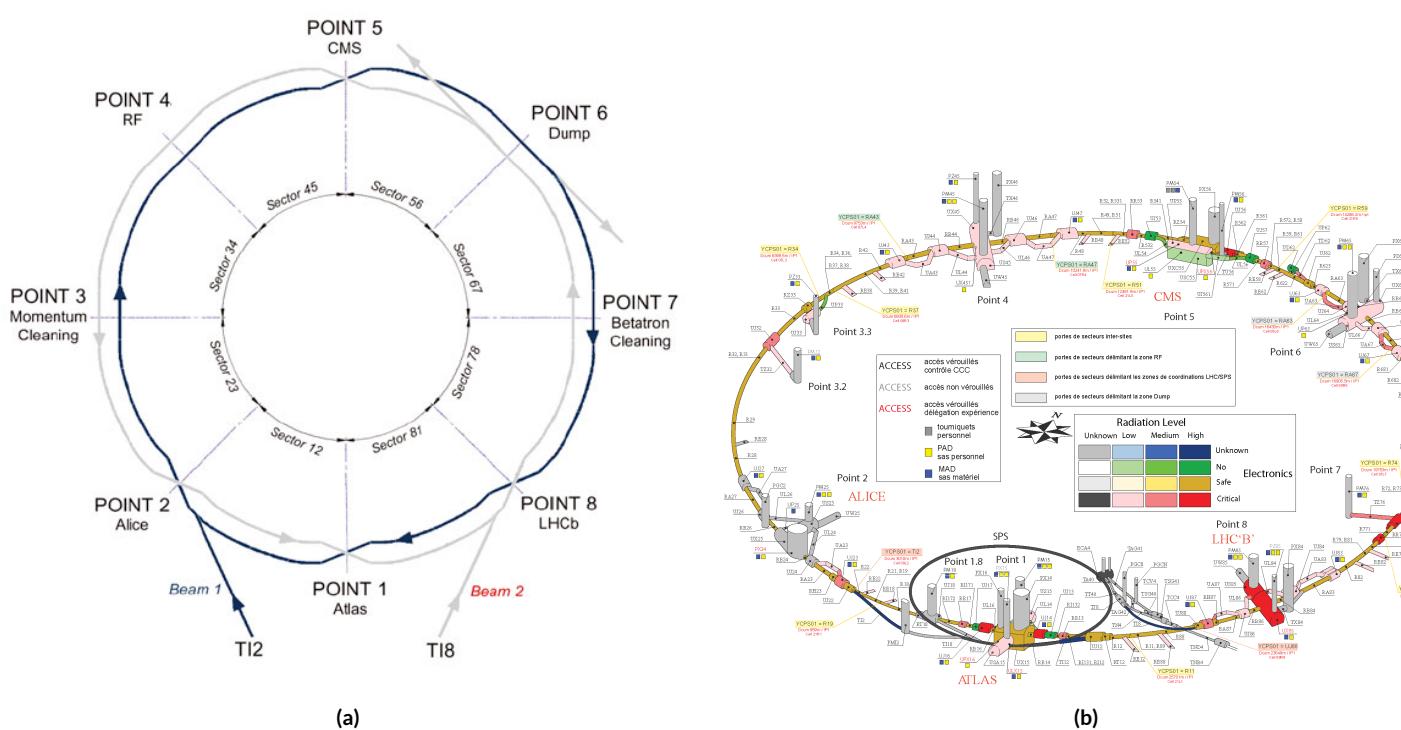


Figure 1.3: Schematic and detailed views of the LHC ring. IC.^{27 63},

¹⁹⁴ Points 1, 2, 5, and 8 house the LHC’s experiments, ATLAS (one of the two general purpose de-
¹⁹⁵ tectors, discussed in detail below), ALICE (A Large Ion Collider Experiment, a dedicated heavy
¹⁹⁶ ion experiment), CMS (Compact Muon Solenoid, the other general purpose detector), and LHCb
¹⁹⁷ (LHC beauty, a B physics experiment), respectively. Point 3 houses a series of collimators that scat-
¹⁹⁸ ter an absorbs particles in the beam with a large momentum deviation from other particles in the
¹⁹⁹ beam (“momentum cleaning”), while Point 7 has a similar setup to remove particles with large beta-
²⁰⁰ tron amplitudes (“betatron cleaning”). Betatron amplitudes are related to how well focused beams
²⁰¹ are and can be thought of as giving a characteristic size for a beam; just as one wants to screen out
²⁰² particles deviating in physical space, one also wants protons in the beam to have nearly identical mo-
²⁰³ mentum. Well focused beams in both position and momentum space are crucial to high quality
²⁰⁴ collisions. Point 4 contains the LHC’s RF (radio frequency; 400 MHz) acceleration system, respon-
²⁰⁵ sible for taking protons from their injection energy of 450 GeV to their collision energy of 3.5, 4, 6.5,
²⁰⁶ or 7 TeV. Point 6 is where the energetic ionizing radiation of circulating beams can be safely taken
²⁰⁷ out of the collider into a block of abosrbing material, either at the end of a data-taking run or in the
²⁰⁸ event of an emergency (in the event of irregular behavior, it is essential to do this as quickly as possi-
²⁰⁹ ble to minimize damage to the accelerator and to experiments); this is known as a “beam dump.”

210 1.3 ATLAS AT A GLANCE

211 1.3.1 COORDINATES AND DISTANCES IN THE ATLAS DETECTOR

212 *A Toroidal LHC ApparatuS* is one of the two general purpose, high luminosity detectors at the
213 LHC, located at Interaction Point 1, as described above. With a length of 44 m and a height of 25 m,
214 it is the detector with largest physical dimensions at the LHC.[‡]. While primarily a high luminosity
215 proton-proton collision detector, ATLAS does collect heavy ion collision data, typically for one
216 month during a year of typical operation.

217 The ATLAS coordinate system is shown in Figure 1.4. It is a right-handed coordinate system
218 centered at the nominal collision point, with the x axis pointing towards the center of the LHC ring,
219 the z axis pointing up, and the y axis completing the right-handed coordinate system.

220 While the Cartesian coordinates are useful for specifying the locations of things like detector
221 components and activated calorimeter cells, cylindrical polar coordinates with the same origin, z
222 axis, and handedness are often more suitable, with a point in 3-space expressed as (r, ϕ, η) . r is the
223 perpendicular distance from the beam axis. This differs from the usual spherical ρ , the distance of a
224 point from the origin, because the ATLAS detector is cylindrical[§], and so detector components are
225 more easily located using r instead of ρ . In some contexts, the latter is used, though this is (or should
226 be) made clear. ϕ is the usual (right-handed) azimuthal angle around the beam axis, with 0 at the $+x$
227 axis.

[‡]This is the only reason CMS can call itself “compact.”

[§]“toroidal;” the hole is the beam pipe

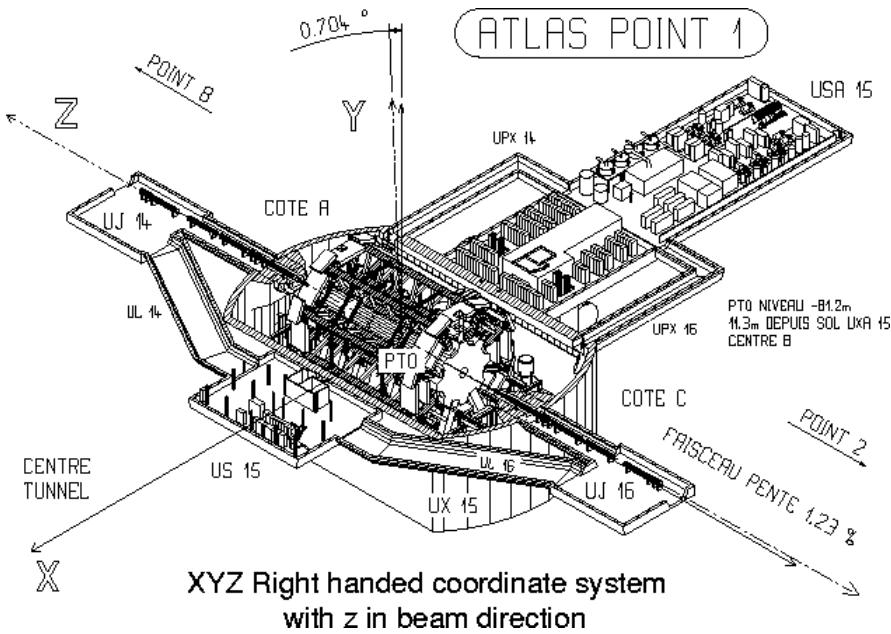


Figure 1.4: The ATLAS coordinate system. "A" side is the airport, and "C" side is "Charlie's," a pub in Saint-Genis, France.

²²⁸ In a lepton collider where total momentum is conserved, a useful coordinate is the relativistic

²²⁹ rapidity of a particle:

$$y = \frac{1}{2} \ln \left[\frac{E + p_z}{E - p_z} \right] \quad (1.1)$$

²³⁰ with E and p_z as the energy and longitudinal momentum of the particle, respectively. The rapidity

²³¹ is the relativistic analog of a rotation angle; boosts can be added in a manner similar to rotations[¶],

²³² and differences in rapidity are invariant under boosts. In a hadronic collider, where the participants

²³³ in the hard scatter are partons inside of the proton of unknown momentum fraction, longitudinal

²³⁴ momentum is not conserved. Nevertheless, since the incident momentum is entirely longitudinal,

[¶]Generally, one need only insert the appropriate factor of i , the square root of -1 ; this introduces differences in sign and changes all of the trigonometric functions associated with rotations into hyperbolic trigonometric functions.

235 momentum is still conserved in the transverse plane, so quantities like trasnverse momentum \vec{p}_T
 236 or energy (E_T)^{||} are often very useful in analysis. However, in the massless limit^{**}, we can take $E =$
 237 $\sqrt{p_T^2 + p_z^2}$. Hence, with θ taken as the zenith angle and o corresponding to the $+z$ direction, for a
 238 massless particle, $p_z = E \cos \theta$. Using the usual half angle formula $\cos \theta = (1 - \tan^2 \theta) / (1 + \tan^2 \theta)$

239

240

$$= 1 - \frac{1}{2} \ln \left[\frac{1}{\tan^2(\theta/2)} \right] = -\ln \left(\tan \frac{\theta}{2} \right) \quad (1.2)$$

241 This last expression, denoted η , is known as the pseudorapidity.

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (1.3)$$

242 Lower values of $|\eta|$ (1.3) correspond to more central areas of the detector known as the “barrel,”
 243 with the typical layout here being concentric, cylindrical layers. Larger values of $|\eta|$ (to ~ 2.5 for
 244 some systems and up to as much as $\sim 4.5 - 5$ for others) are known as the “end caps,” where ma-
 245 terial is typically arranged as disks of equal radius centered on the beam pipe stacked to ever greater

||Energy is not a vectoral quantity, but one can take the scalar or vectoral sum of vectors formed from energy deposits with their location as the direction and energy value as magnitude. In practice, primitives are almost always assumed to be massless, so transverse energy and momentum may loosely be thought of as equivalent, with $E_T = |\vec{p}_T| = p_T$

** not a terrible one for most particles depositing energy in the calorimeter; pions have masses of ~ 130 MeV, and typical energies of calorimeter objects are $\sim 10^3$ GeV, making for a boost of roughly 100.

246 values of $|z|$. This terminology will be useful when discussing the various subsystems of the ATLAS
 247 detector. Since decay products from a collision propagate radially (in the calorimeter portions of
 248 the detector with no magnetic field), the radial coordinate is not so important for composite physics
 249 objects like electrons or jets, which are typically expressed as momentum 4-vectors. Hence, only η
 250 and ϕ are the only useful spatial coordinates. Distances between objects are often expressed not as a
 251 difference in solid angle, but as a distance, ΔR , in the plane, where

$$\Delta R_{12} = (\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2 \quad (1.4)$$

252 In addition to the physical distance traveled by particles, also important in talking about particle
 253 detectors are radiation lengths and (nuclear) interaction lengths, both of which characterize typical
 254 lengths for the energy loss of energetic particles traveling through materials. In general, the energy
 255 loss is modeled as an exponential

$$E = E_0 e^{-l/L} \quad (1.5)$$

256 where E_0 is the initial energy, and L is the characteristic length. These lengths are characteristics of
 257 the materials through which particles pass. In the case of uniform, composite materials, the length
 258 may be found by calculating the reciprocal of the sum of mass fraction weighted reciprocal char-
 259 acteristic lengths of the components. This formula works quite well for modeling the very regular
 260 behavior of electromagnetic showers (energetic photons convert into electron/positron pairs, which
 261 emit photons...). In this case, L is denoted X_0 ; this is the radiation length. Hadronic showers are far
 262 more complicated, with shower multiplicity and makeup dependent on the particle in question and

263 generally being highly variable (and things like pions converting to photons, which then shower elec-
264 tromagnetically). Nevertheless, a characteristic length can be tabulated for a standard particle type,
265 typically pions, and is called the nuclear interaction length.

266 **I.3.2 GENERAL LAYOUT OF ATLAS**

267 The ATLAS detector and its main components are shown in Figure 1.5. ATLAS is designed as a
268 largely hermetic detector, offering full coverage in ϕ and coverage in $|\eta|$ up to 4.7. The multiple sub-
269 systems allow for good characterization of the decay products from collisions in the LHC. The in-
270 nermost system is the inner detector (ID); composed primarily of silicon pixels and strips immersed
271 in a magnetic field, it is designed to construct the curved trajectories of charged particles produced in
272 collisions while taking up as little material as possible.

273 Surrounding the ID is the liquid argon based electromagnetic calorimeter (ECAL), which is de-
274 signed to capture all of the energy of the electromagnetic showers produced by electrons and pho-
275 tons coming from particle collisions. The ECAL is in turn encapsulated by a scintillating tile based
276 and liquid argon based hadronic calorimeter (HCAL) that captures any remaining energy from the
277 jets produced by hadronizing quarks and gluons coming from collisions.

278 The outermost layer of ATLAS is the muon spectrometer (MS) with its own magnetic field pro-
279 duced by toroidal magnets. Muons are highly penetrating particles that escape the calorimeters with
280 most of their initial momentum, so the MS and its magnets are designed to curve these charged
281 particles and measure their trajectories to measure their outgoing momenta. Each of these detector
282 systems has several principal subsystems and performance characteristics, which will be described in

²⁸³ turn below.

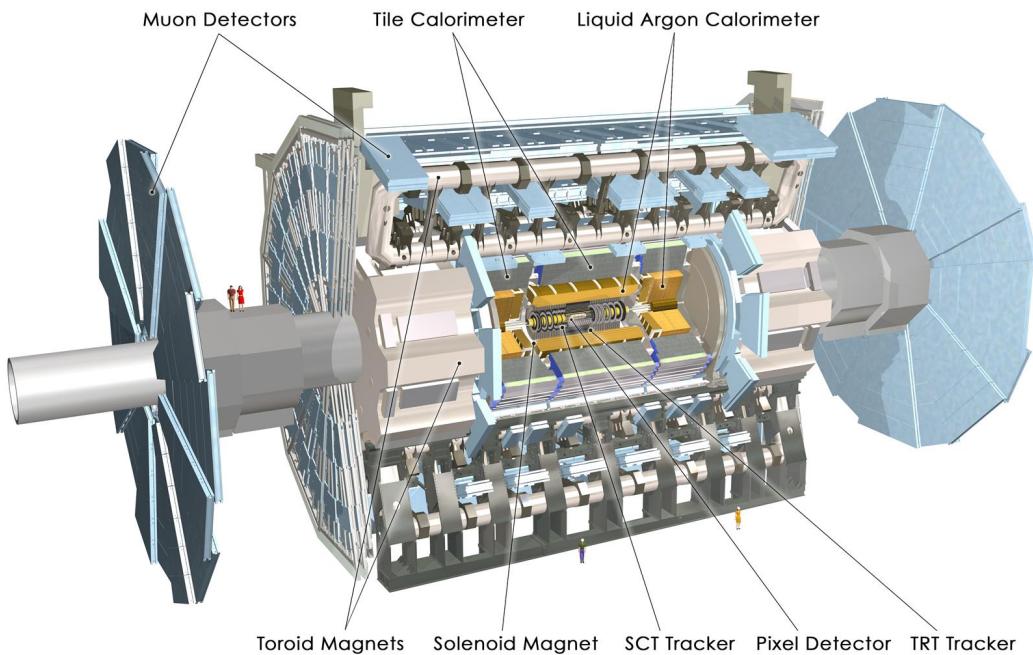


Figure 1.5: The ATLAS detector with principal subsystems shown.

²⁸⁴ I.4 THE INNER DETECTOR

²⁸⁵ ATLAS's inner detector (ID) is surrounded by a superconducting solenoid that is cryogenically
²⁸⁶ cooled to a temperature of 4.5 K, which immerses it in a 2 T magnetic field. The ID uses two sili-
²⁸⁷ con detector subsystems (the Pixel and SemiConductor (strip) Tracker (SCT)) to track the curved
²⁸⁸ trajectories of charged particles emanating from particle collisions and one subsystem composed of
²⁸⁹ gas straw detectors with filaments for e/π discrimination, as shown in Figure 1.9. The ID offers full
²⁹⁰ coverage in ϕ and extends to an $|\eta|$ of 2.5.

²⁹¹ Since the components of the ID do not provide an energy measurement, it is desirable for a track-

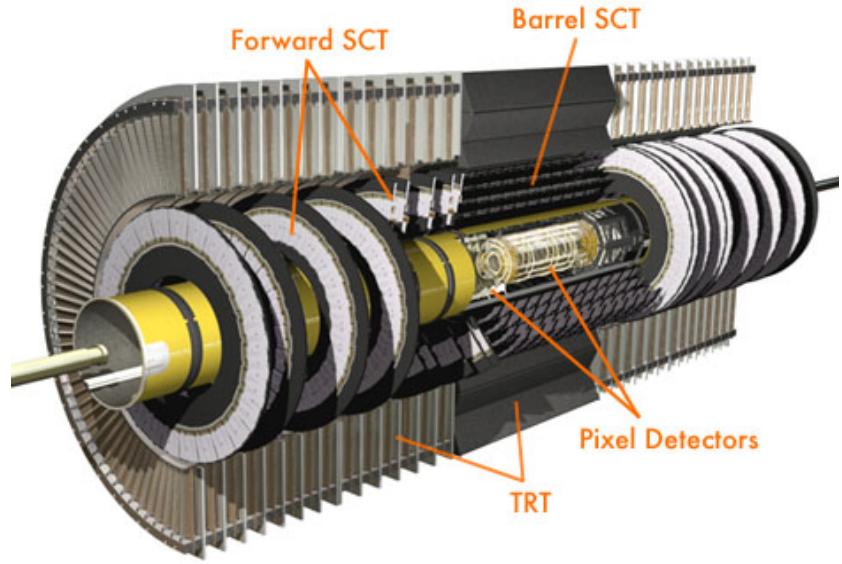


Figure 1.6: The ATLAS inner detector. IC:³⁶

ing system to have as small a material budget as possible so that more accurate energy measurements
 may be done in the calorimeters. Generally, there are two radiation lengths in the inner detector (it
 varies with η); the full material budget, with the layout of the individual layers in each subsystem,
 can be seen in 1.8.

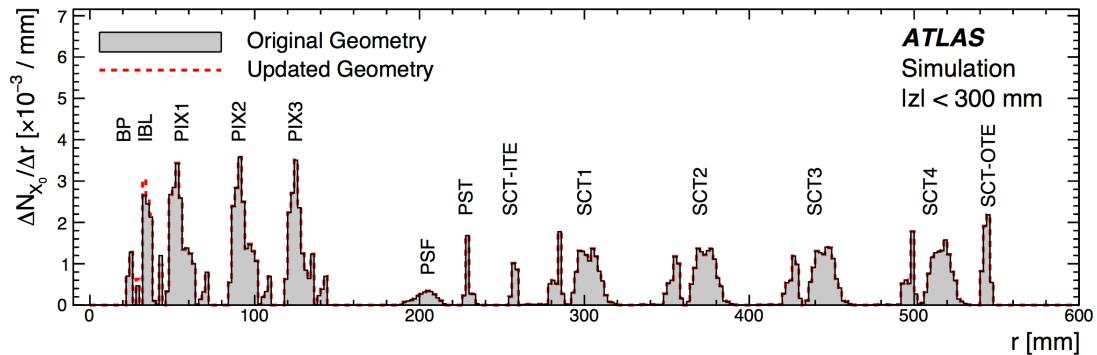


Figure 1.7: The ID material budget. IC:³⁵

²⁹⁶ **I.4.1 THE PIXEL DETECTOR**

²⁹⁷ The innermost part of ATLAS is the pixel detector, which, as the name suggests, is comprised of
²⁹⁸ four layers of pixels in the barrel at 32, 51, 89, and 123 mm from the beam pipe, and three years in the
²⁹⁹ end caps at 495, 580, and 650 mm from the beam pipe, with over 80 million channels total. The in-
³⁰⁰ nermost layer of pixels, the insertable *B* layer (IBL) was installed during the 2013–14 LHC shutdown.

³⁰¹ The pixels are cooled to a temperature of ~ -5 with N_2 gas and operate at 150–600 V. The pixels
³⁰² themselves come in two sizes $50 \times 400(600) \times 250 \mu\text{m}$, with the larger pixels being in the outer
³⁰³ layers. They provide nominal resolution of $10(115) \mu\text{m}$ resolution in $r - \phi(z)$ direction.

³⁰⁴ In order to improve total coverage in the detector and prevent any gaps, pixels are not installed
³⁰⁵ flush with each other. Pixels in the barrel are tilted at about 20 with an overlap in $r - \phi$, as shown in
³⁰⁶ Figure I.8. The disks of the ID end caps are rotated with respect to each other by 3.75

³⁰⁷ **I.4.2 THE SILICON MICROSTRIP DETECTOR (SCT)**

³⁰⁸ The layout of the SCT is similar to that of the Pixel detector, except, for cost considerations, the
³⁰⁹ SCT uses silicon strips that are also cooled to ~ -5 with N_2 gas and operate from 150–350 V. Strip
³¹⁰ dimensions are $80 \times 6000 \times 285 \mu\text{m}$, and provide nominal $17(580) \mu\text{m}$ resolution in $r - \phi(z)$. Barrel
³¹¹ strips feature an 11tilt and come in four layers at 299, 371, 443, and 514 mm. There are nine end cap
³¹² disks on each side at z values varying from 934–2720 mm.

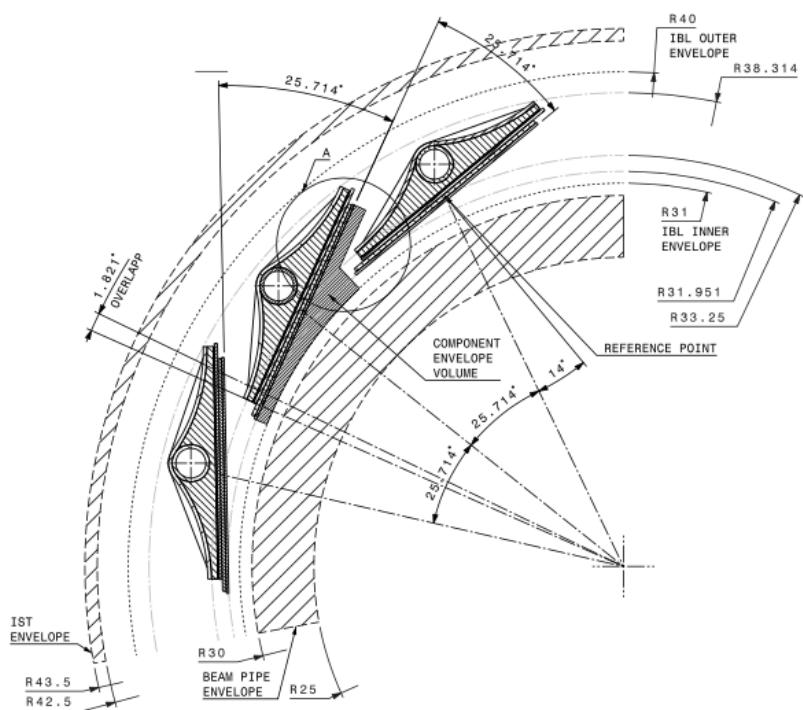


Figure 1.8: Arrangement of pixels in the barrel. IC:²⁶

313 1.4.3 TRANSITION RADIATION TRACKER (TRT)

314 The final and outermost subsystem in the ID is the transition radiation tracker (TRT). It provides
315 coverage for $|\eta|$ up to 2.0 and is composed of straw detectors with a 4 mm diameter and run the
316 length of the detector module, which provide $130 \mu\text{m}$ resolution and filled with a Xe-CO₂-O₂ (70-
317 27-3) gas combination and operate at -1500 V. The filaments and foil lining inside the straws induce
318 X-ray emission in electrons and pions passing through the TRT as they move from a dielectric to
319 a gas; this “transition radiation” is the source of the TRT’s name. Since the energy deposited due
320 to transition radiation is proportional to the relativistic boost γ , for constant momentum, this is
321 inversely proportional to mass, so electrons will have $\sim 130/0.5 = 260\times$ more transition radiation
322 than pions, in principle enabling excellent electron/pion discrimination. The TRT will be replaced
323 by silicon strips in the Phase II upgrade.

324 1.5 THE ATLAS CALORIMETERS

325 ATLAS has four main calorimeter systems: the liquid argon based electromagnetic (ECAL), hadronic
326 end cap (HEC), and forward (FCAL) calorimeters, and the scintillating tile based hardronic Tile
327 calorimeter in the barrel. Their layout and material budget in interaction lengths can be seen in Fig-
328 ure 1.10.

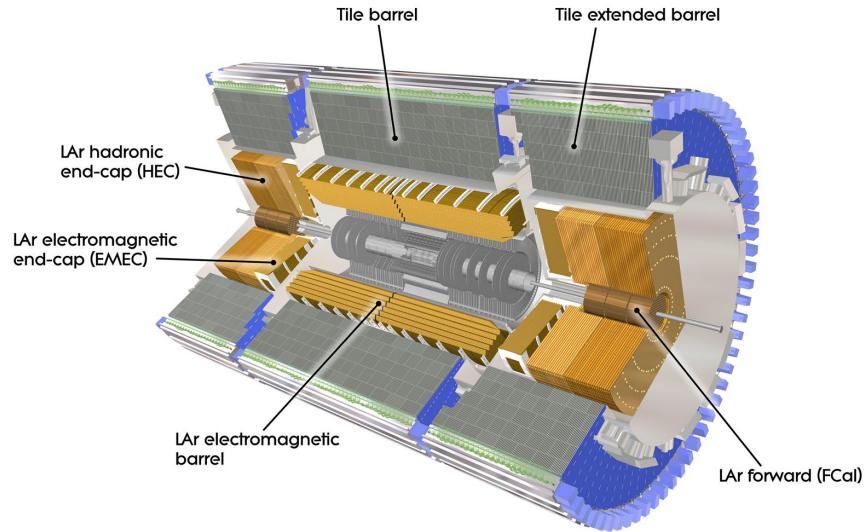


Figure 1.9: The ATLAS calorimeters.

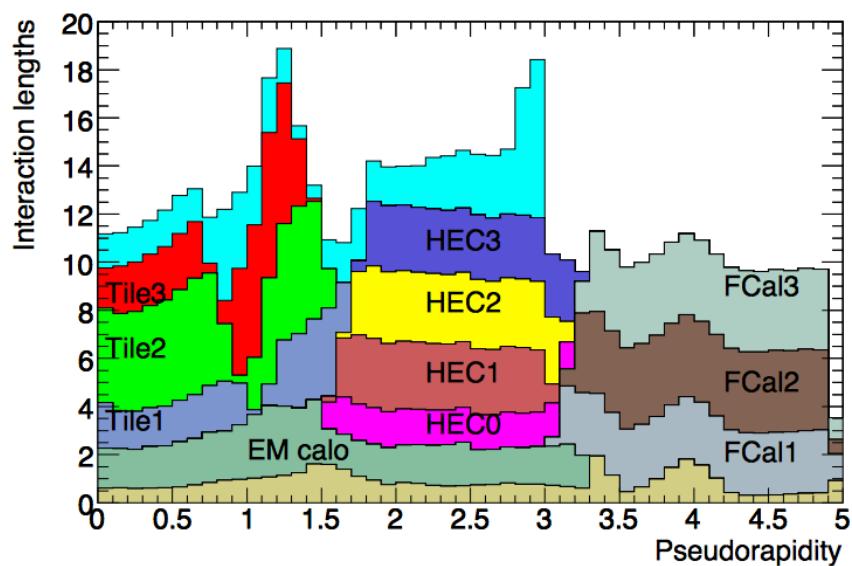


Figure 1.10: Material depth of the ATLAS calorimeters. IC;³⁶

329 1.5.1 CALORIMETER RESOLUTION

330 Before diving into the specifics of each of the ATLAS calorimeters, we review how calorimeter en-
331 ergy resolution performance is typically quoted. A calorimeters relative energy resolution (a ratio)
332 can be broken up into three orthogonal components, as shown in Equation 1.6.

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \quad (1.6)$$

333 S is the photoelectron statistics or stochastic term represents the coefficient to the usual counting
334 term (assuming Gaussian statistics); N is a noise term, which is constant per channel (and hence
335 comes in as $1/E$ in the relative energy resolution); and C is a constant “calibration” term, which re-
336 flects how well one intrinsically understands a detector (i.e. mismodelling introduces an irreducible
337 component to the energy resolution). If any detector were perfectly modeled/understood, it’s C
338 term would be zero. $N \sim 0.1 - 0.5$ GeV for the typical calorimeter regardless of type, so S and C are
339 typically quoted.

340 A typical stochastic term scales as $S \sim \text{few\%} \sqrt{d_{active} [\text{mm}] / f_{samp}}$, where f_{samp} is the sampling
341 fraction or the ratio of a calorimeter by mass is composed of an active volume capable (i.e. one
342 that registers energy deposits). The tile calorimeter, for example, has a sampling fraction of about
343 $1/36$. There are several reasons that this fraction is so low. First, many active volumes have insuffi-
344 cient stopping power; one wants to capture as much energy as possible from electromagnetic and
345 hadronic showers inside the calorimeter, and this simply is not possible for most active media (one

346 notable exception to this is the CMS crystal-based calorimeter; ATLAS is a more conservative de-
347 sign), so well-behaved absorbers like lead or iron are necessary to ensure all the energy is contained
348 within a calorimeter. Another factor is cost; things like liquid argon are expensive. Finally, most
349 active media are unsuitable for structural support, so sturdy absorbing materials help relieve engi-
350 neering constraints.

351 1.5.2 THE ELECTROMAGNETIC CALORIMETER (ECAL)

352 The ECAL has liquid argon (LAr) as an active material with lead as an absorber. The ECAL barrel
353 extends to $|\eta|$ of 1.475 with three layers at 1150, 1250, and 2050 mm, and its end cap, comprised of
354 two wheels extends covers $1.375 < |\eta| < 2.5(3.2)$ for the inner (outer) wheel, with 3 (2) layers out to
355 3100 mm. There is a 1.1 (0.5) cm thick layer of LAr pre-sampler up to $|\eta|$ of 1.8 in the barrel (end cap)
356 of the ECAL, which is designed to aid in correcting for electron and photon energy loss in the ID.

357 The LAr and lead absorber are arranged in alternating beveled sawtooth layers in what is known
358 as an “accordion” geometry, shown in Figure 1.11, which shows the layout of a barrel module in the
359 ECAL. The absorber thickness is 1.53 (1.13) mm for $|\eta|$ less (more) than 0.8 to ensure a constant sam-
360 pling fraction. This arrangement helps provide greater coverage in ϕ .

361 The ECAL overall typically covers 2–4 interaction lengths or about 20–40 radiation lengths. Its
362 performance corresponds to resolution coefficients $S = 0.1 \text{ GeV}^{-1/2}$ and $C = 0.002$ with a 450
363 ns drift time. In order to optimize the material budget and overall detector construction, the ECAL
364 barrel infrastructure is integrated with that of the ID’s solenoid. The granularity of the ECAL barrel
365 middle layer, $\Delta\eta \times \Delta\phi$ cells of size 0.025×0.025 , are used to define the granularity of calorimeter

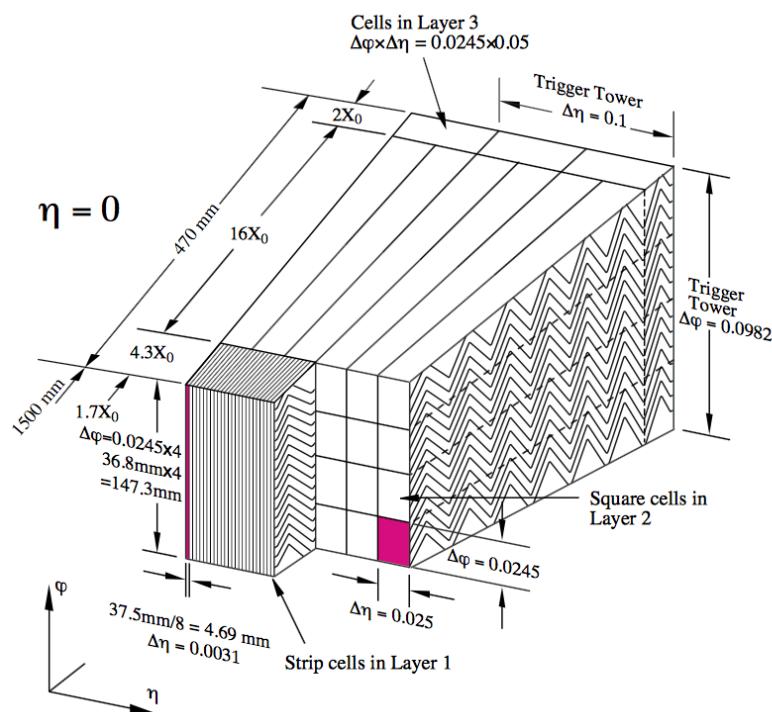


Figure 1.11: The accordion geometry of the LAr electromagnetic calorimeters is prominently shown in this illustration of an ECAL barrel module. IC:³⁶

³⁶⁶ cluster reconstruction in ATLAS.

³⁶⁷ **I.5.3 HADRONIC END CAPS (HEC)**

³⁶⁸ The HEC covers an $|\eta|$ range of 1.5 to 3.2. Like the ECAL end caps, the HEC consists of two identi-
³⁶⁹ cal wheels out to a distance from the beam axis of 2030 mm; its layout is shown in Figure I.12. The
³⁷⁰ HEC also has LAr as the active material, but instead has flat copper plates as absorbers for sampling
³⁷¹ fraction of 4.4% and 2.2% in the first and second wheels, respectively. Its granularity in ϕ is 0.1×0.1
³⁷² for $|\eta|$ up to 2.5 and 0.2×0.2 in the more forward regions.

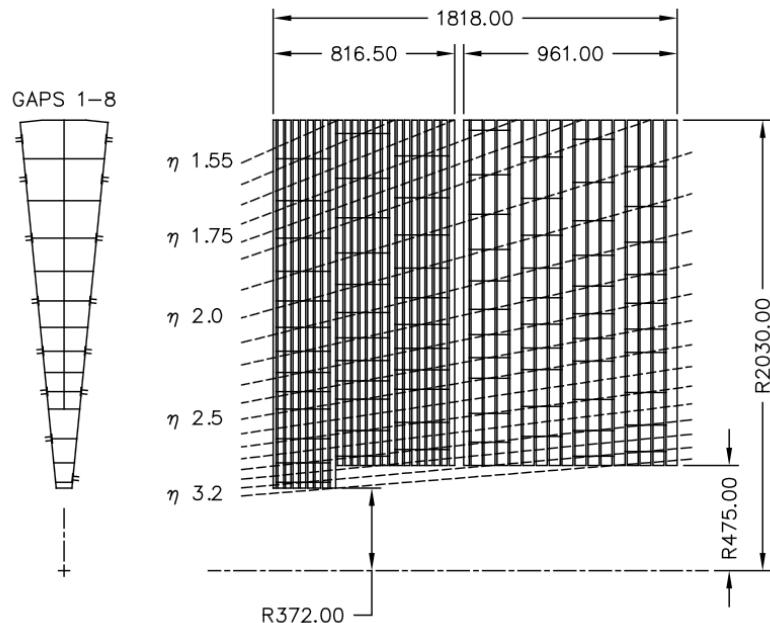


Figure 1.12: The layout of the HEC in $r - \phi$ and $r - z$ dimensions are in millimeters. IC:³⁶

³⁷³ 1.5.4 THE FORWARD CALORIMETER (FCAL)

³⁷⁴ The FCAL covers an $|\eta|$ range from 3.1 to 4.9, again using LAr as the active material in gaps between
³⁷⁵ rods and tubes in a copper-tungsten matrix, as shown in Figure 1.13. These system has characteris-
³⁷⁶ tic performance corresponding to stochastic term of $S \approx 1 \text{ GeV}^{-1/2}$. There are three modules in
³⁷⁷ the FCAL: one electromagnetic and two hadronic, with the latter two featuring a higher tungsten
³⁷⁸ content for a larger absorption length.

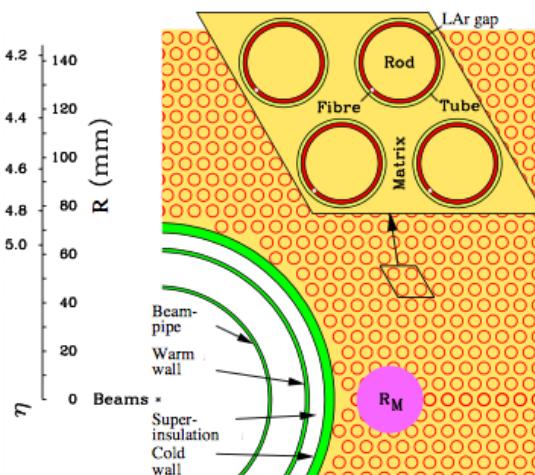


Figure 1.13: The material layout for a typical section of the FCAL in the transverse plane. IC:³⁶

³⁷⁹ 1.5.5 THE HADRONIC TILE CALORIMETER

³⁸⁰ The tile calorimeter, covering at $|\eta|$ of up to 1.7 is made up of 64 modules in the barrel (each cover-
³⁸¹ ing $\Delta\phi$ of $360/64 = 5.625$, each with a layout as in Figure 1.14. It is designed to be self-supporting
³⁸² for structural reasons, and so is the only calorimeter without LAr as a an active medium, with a stag-
³⁸³ gered matrix of active scintillating polystyrene and supporting steel. It operates at 1800 V with a 400

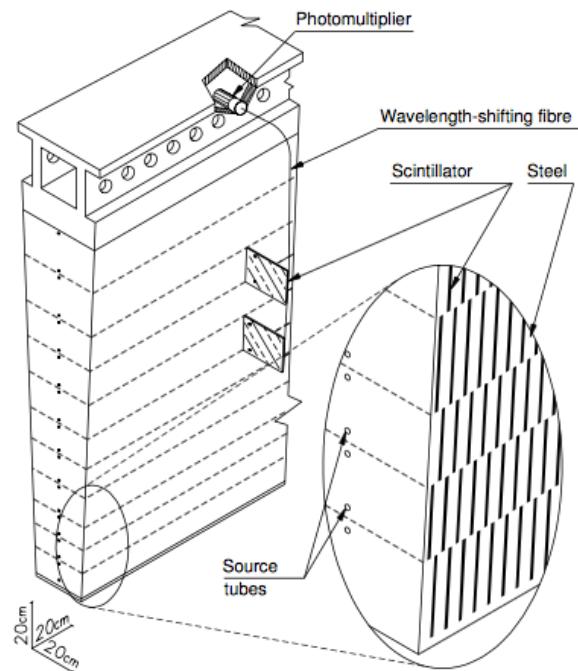


Figure 1.14: The material layout for a typical section of the hadronic tile calorimeter. IC.³⁶

384 ns dead time and has a thickness corresponding to 10–20 interaction lengths (2.28–4.25 m). Its cell
 385 have a $\Delta\eta \times \Delta\phi$ granularity of 0.1×0.1 in the first two layers and 0.2×0.1 in the last layer. Its
 386 performance corresponds to $S = 0.5 \text{ GeV}^{-1/2}$ and $C = 0.05$ (0.03 after calibration).

387 1.6 THE MUON SPECTROMETER

388 Since the energy of muons is not captured within the calorimeters, the stations of the ATLAS MS
 389 surrounds the entire detectors and provides tracks of outgoing muons that can be matched to tracks
 390 in the ID. The ATLAS toroids, which provide field strengths of up to 2.5 (3.5) T in the barrel (end
 391 cap) with typical strengths of 0.5–1.0 T, bends the muons, which allows for a muon momentum
 392 measurement since the mass of the muon is known. The relative momentum resolution of a tracker
 393 (assuming, as in ATLAS, that bending primarily happens in the ϕ direction) may be expressed as

$$\frac{\sigma_{p_T}}{p_T} = c_o \oplus c_i \cdot p_T \quad (1.7)$$

394 The c_o term represents a degradation in resolution due to multiple scattering, and is typically 0.5–
 395 2%². The c_i term describes the effect that, for a constant magnetic field, higher momentum muons
 396 will curve less. This term has typical values of $10^{-3} - 10^{-4} \text{ GeV}^{-1}$. At very high p_T values, this is of
 397 particular concern since a very small curvature can result in charge misidentification.

398 A cross-sectional view (in $r-z$) of the muon spectrometer with station names, detector types, and
 399 layouts is shown in Figure 1.15. There are three layers of muon detectors in both the barrel (at 5 000,
 400 7 500, and 10 000 mm) and end cap (at 7 000 (11 000), 13 500, and 21 000 mm), with the innermost

401 end cap layer split in two due to the end cap toroid. This corresponds to an $|\eta|$ range up to 2.4 for
 both precision and trigger coverage, and up to 2.7 for precision detection only.

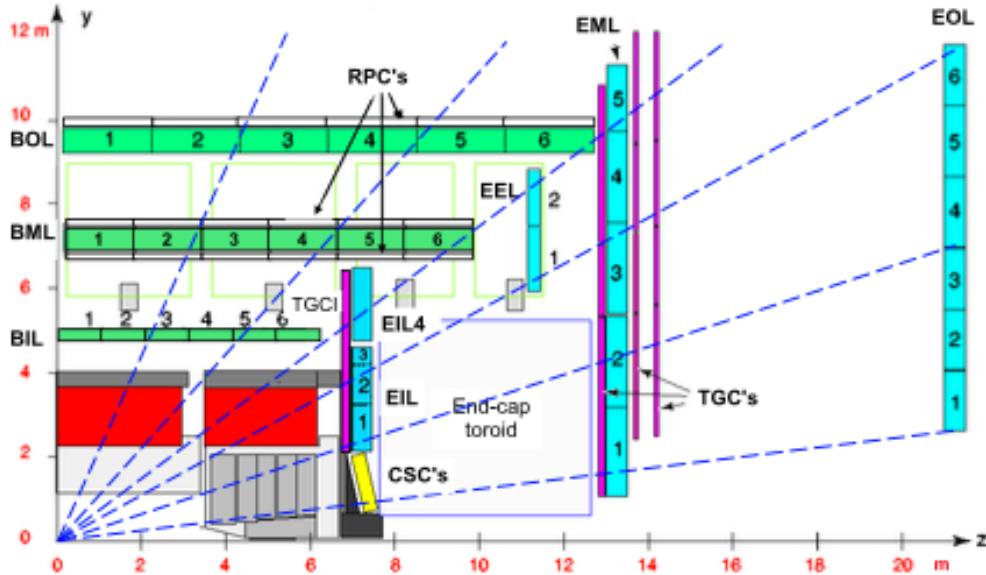


Figure 1.15: The ATLAS muon spectrometer. Naming of the MDT stations obeys the following convention [BE] (barrel or end cap) [IEMO] (inner, inner extended (end cap only), middle, or outer layer) [1-6] (increasing in z (r) for the barrel (end cap)), so EI1 is the station in the inner most end cap layer closest to the beam pipe. IC:³⁶

402
 403 The MS can reconstruct muons with transverse momenta from 5 GeV up to 3 TeV (with 10%
 404 resolution at 1 TeV (3% at 100 GeV)). Detectors in the MS fall into two broad headings, precision
 405 detectors and trigger detectors, both described below. Nominal performance of the current detec-
 406 tor types in the MS is summarized in Figure 1.16, a table taken from³⁶. It should be noted that $|\eta|$
 407 ranges quoted below, where applicable, do not include the range 0-0.1, where this a gap in the MS to
 408 allow for cabling and other services to the ATLAS detector; for a discussion of compensatory mea-
 409 sures in muon reconstruction, see Chapter 5.

Type	Function	Chamber resolution (RMS) in			Measurements/track		Number of	
		z/R	ϕ	time	barrel	end-cap	chambers	channels
MDT	tracking	35 μm (z)	—	—	20	20	1088 (1150)	339k (354k)
CSC	tracking	40 μm (R)	5 mm	7 ns	—	4	32	30.7k
RPC	trigger	10 mm (z)	10 mm	1.5 ns	6	—	544 (606)	359k (373k)
TGC	trigger	2–6 mm (R)	3–7 mm	4 ns	—	9	3588	318k

Figure 1.16: ATLAS MS detector performance. IC:³⁶

⁴¹⁰ 1.6.1 PRECISION DETECTORS

⁴¹¹ The ATLAS MS has two types of precision detectors: monitored drift tubes (MDT's) and cathode
⁴¹² strip chambers (CSC's). An MDT is a tube with a 3 cm diameter with length depending on the sta-
⁴¹³ tion in which the tube is located. The tube is filled with an Ar/CO₂ gas mixture and has a tungsten-
⁴¹⁴ rhenium wire at its center that is kept at 3 000 V when operational. The MDT's provide 35 μm res-
⁴¹⁵ olution (per chamber) in their cross-sectional dimension (there is no sensitivity along the axis of the
⁴¹⁶ wire). Resolution of this magnitude requires very precise knowledge of the location of the wires
⁴¹⁷ within the MDT's; this is generally true for detectors in the MS (trigger as well as precision); to this
⁴¹⁸ end, stations of the MS are aligned using an optical laser system. For a detailed description of how
⁴¹⁹ misalignment can affect performance, see Appendix A for a detailed discussion of misalignment's
⁴²⁰ simulated effets on the performance of the proposed micromegas trigger processor in the New Small
⁴²¹ Wheel (NSW) of the Phase I upgrade. Their 700 ns dead time, however, percludes their use as trig-
⁴²² ger detectors and also in the region of the small wheel (innermost endcap) closest to the beam pipe
⁴²³ ($|\eta|$ from 2.0 to 2.7), where rates are highest.
⁴²⁴ In this region, the precision detectors are the CSC's, which have a much lower dead time of ~ 40

⁴²⁵ ns. These are multiwire proportional chambers with cathode planes that have orthogonal sets of
⁴²⁶ strips, allowing for a measurement in both the principal directions. CSC detector sizes also vary
⁴²⁷ by station, coming in both small and large chambers. The CSC strip pitch is 5.31 (5.56) mm for the
⁴²⁸ large (small) chambers, with position determined from the induced charge distribution in the strips.

⁴²⁹ This corresponds to a nominal resolution of 60 (5 000) μm per plane in the bending (non-bending)
⁴³⁰ direction. These are slated to be replaced by micromegas detectors in the NSW.

⁴³¹ 1.6.2 TRIGGER DETECTORS

⁴³² Trigger detectors have a fundamentally different role than the precision detectors, instead needing
⁴³³ to deliver “good enough” approximate values of muon track positions and p_T values. The MS has
⁴³⁴ two types of trigger detectors: resistive plate chambers (RPC’s) in the barrel and thin gap chambers
⁴³⁵ (TGC’s) in the end caps. They collectively cover an $|\eta|$ range to 2.4, and their arrangement is shown
⁴³⁶ in Figure 1.17.

⁴³⁷ The RPC’s are parallel plate detectors with a dead time of 5 ns and a thickness of 2 mm, kept at
⁴³⁸ a potential of 9 800 V; they are deployed in three layers. RPC’s, too, feature strips with orthogonal
⁴³⁹ arrangements on the top and bottom planes, with a strip pitch of 23–35 mm.

⁴⁴⁰ The TGC’s are multiwire proportional chambers with a dead time of 25 ns. Also, featuring or-
⁴⁴¹ thogonal strips, the TGC’s also provide a ϕ measurement to compensate for the lack of MDT sensi-
⁴⁴² tivity in this direction. There are four layers of TGC’s in the end cap. TGC’s will be supplanted by
⁴⁴³ sTGC’s (small thin gap chambers) in the NSW.

⁴⁴⁴ For more details on how detector level trigger objects work in the ATLAS MS, see Appendix A

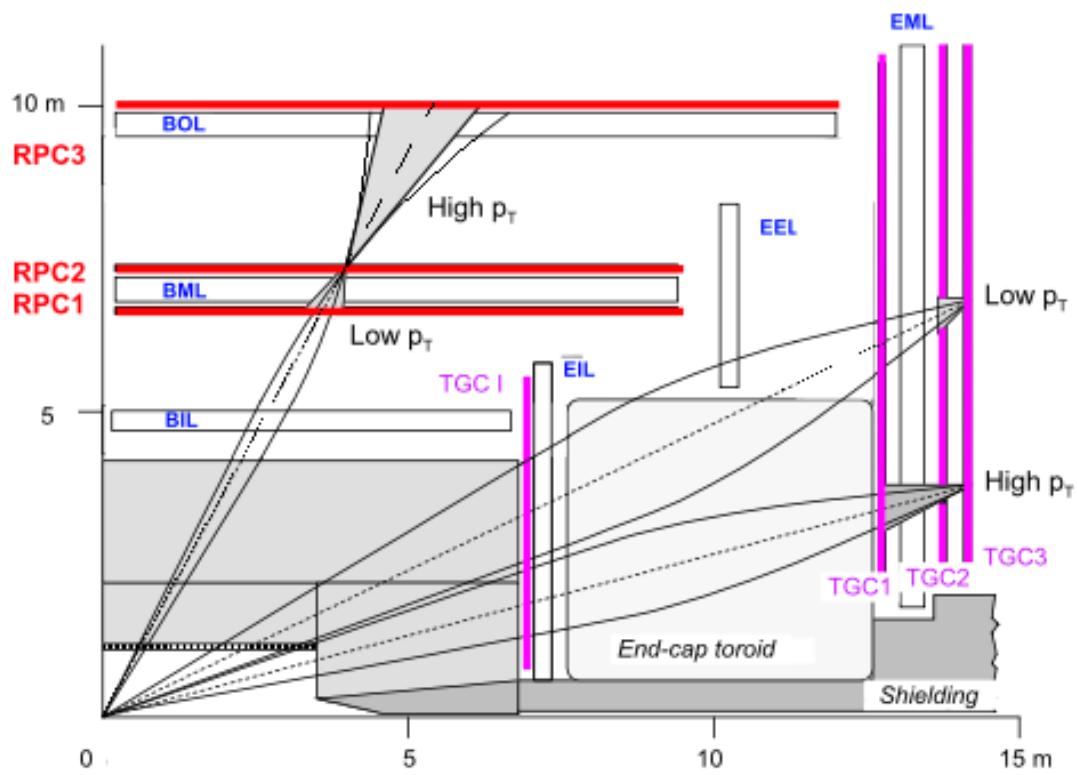


Figure 1.17: ATLAS MS trigger detector arrangement. IC:³⁶

⁴⁴⁵ for details on the micromegas trigger processor algorithm.

*The relationship between theorists and experimentalists
is like that between a truffle farmer and his pig*

Howard Georgi

2

446

447

The Standard Model Higgs and Colider

448

Event Variables

449 MUCH HAS BEEN SAID about the so-called Standard Model (SM) of particle physics, so only the
450 bare essentials of electroweak symmetry breaking and Higgs production relevant to SM $VH(b\bar{b})$ will

⁴⁵¹ be addressed here, following the discussion and notation of⁵⁷ Chapter 11. We then move onto the
⁴⁵² treatment of kinematic variables in collider events, including the two novel schemes considered in
⁴⁵³ this thesis, the Lorentz Invariants (LI) and RestFrames (RF) variables.

⁴⁵⁴ 2.1 THE STANDARD MODEL HIGGS BOSON

⁴⁵⁵ The generic scalar Lagrangian potential (the kinetic term will be addressed later) for a scalar in the
⁴⁵⁶ SM is:

$$V(\Phi) = m^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \quad (2.1)$$

⁴⁵⁷ where Φ is the Higgs field, a complex scalar doublet under $SU(2)$. Its four degrees of freedom are
⁴⁵⁸ typically decomposed as follows:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}\phi^+ \\ \phi^0 + i\alpha^0 \end{pmatrix} \quad (2.2)$$

⁴⁵⁹ ϕ^+ is the complex charged component of the Higgs doublet, and ϕ^0 and α^0 are the CP-even and
⁴⁶⁰ CP-odd neutral components, respectively.

⁴⁶¹ If the sign of $m^2 \Phi^\dagger \Phi$ is negative, Φ acquires a *vacuum expectation value* or VEV:

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \sqrt{\frac{2m^2}{\lambda}} \end{pmatrix} \quad (2.3)$$

⁴⁶² with this value typically denoted $v = \sqrt{2m^2/\lambda} = \sqrt{\sqrt{2}G_F}246$ GeV (with the coupling of the

⁴⁶³ 4-Fermi effective theory of weak interactions measured through experiments involving muon decay),

⁴⁶⁴ and ϕ^0 is rewritten as $\phi^0 = H + v$.

⁴⁶⁵ It is this non-zero VEV that induces spontaneous symmetry breaking in the SM's gauge (local)

⁴⁶⁶ symmetry group of $SU(3)_C \times SU(2)_L \times U(1)_Y$ since the VEV does not respect the $SU(2)_L \times U(1)_Y$

⁴⁶⁷ symmetry of the Lagrangian (i.e. $\langle \Phi \rangle$ is not invariant under a gauge transformation of this group).

⁴⁶⁸ Three of the four generators of this subgroup are spontaneously broken, which implies the existence

⁴⁶⁹ of three massless Goldstone bosons, which are in turn “eaten” by linear combinations of the W^a

⁴⁷⁰ and B bosons to form the longitudinal components of the familiar W^\pm and Z bosons, with the last

⁴⁷¹ generator giving rise to the usual, unbroken $U(1)_{EM}$ symmetry and its massless photon, A , as well

⁴⁷² as the scalar Higgs boson H . To see this one starts with the full Higgs SM Lagrangian (kinetic minus

⁴⁷³ potential only)

$$\mathcal{L}_{Higgs} = (D_\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi), \quad \mu \Phi = (\partial_\mu + ig\sigma^a W_\mu^a + ig' Y B_\mu/2) \Phi \quad (2.4)$$

⁴⁷⁴ One simply plugs in the reparametrized Φ with $\phi^0 = H + v$, collects the terms involving v together

⁴⁷⁵ with the appropriate W and B kinetic terms to extract:

$$M_W^2 = \frac{g^2 v^2}{4}, \quad M_Z^2 = \frac{(g'^2 + g^2) v^2}{4} \quad (2.5)$$

⁴⁷⁶ This is left as an exercise for the reader; this exercise also makes it manifest that the Higgs couplings

⁴⁷⁷ with the W^\pm and Z scale quadratically with their masses. Since the Higgs field also respects the

⁴⁷⁸ $SU(3)_C$ color symmetry, the eight gluons are also left massless, and the H is left interacting with
⁴⁷⁹ photons and gluons primarily through heavy quark loops.

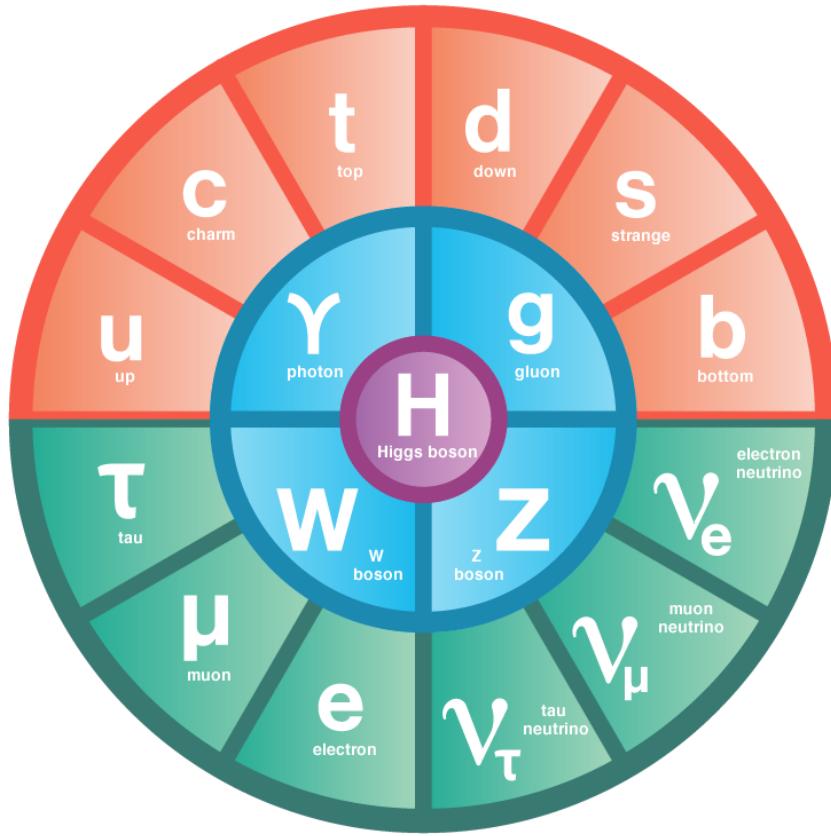


Figure 2.1: The fundamental particles of the Standard Model. IC:⁶⁵

⁴⁸⁰ The Higgs is often introduced to the public at large as the mechanism through which fundamen-
⁴⁸¹ tal fermions (enumerated in Figure 2.1) acquire mass—this is through the Yukawa interactions of the
⁴⁸² Higgs:

$$\mathcal{L}_{Yukawa} = -\hat{b}_{d_{ij}} \bar{q}_{L_i} \tilde{\Phi} d_{R_j} - \hat{b}_{u_{ij}} \bar{q}_{L_i} \tilde{\Phi} u_{R_j} - \hat{b}_{l_{ij}} \bar{l}_{L_i} \tilde{\Phi} e_{R_j} + h.c. \quad (2.6)$$

⁴⁸³ where $\tilde{\Phi} = i\sigma_2 \Phi^*$, q_L (l_L) and u_R , d_R (e_R) are the quark (lepton) left-handed doublets and right

484 handed singlets of the weak $SU(2)_L$ group, with each term parametrized by a 3×3 matrix in family
 485 space (also known as the fermion generations). The neutrinos have been purposely omitted since
 486 the mechanism that generates their mass is as of yet unknown, though these Yukawa interactions
 487 could have a non-zero contribution. Once the Higgs VEV value is known and the Yukawa inter-
 488 action matrices \hat{b}_{f_i} ($i, j \in 1, 2, 3$) are diagonalized, the fermion masses can simply be written as
 489 $m_{f_i} = b_{f_i} v / \sqrt{2}$. The SM has no motivation for any of these masses.

490 Note that from \mathcal{L}_{Yukawa} , it is easy to see that the Higgs couplings with fermions scale linearly
 491 with fermion mass. Higgs self-couplings and beyond the standard model (BSM) Higgs scenarios are
 492 beyond the scope of this thesis.

493 2.2 HIGGS BOSON PRODUCTION AND DECAY AT THE LARGE HADRON COLLIDER

494 The leading order Feynmann diagrams for the four dominant modes of Higgs production at the
 LHC are shown in Figure 2.2, each described briefly in turn. The dominant process, accounting for

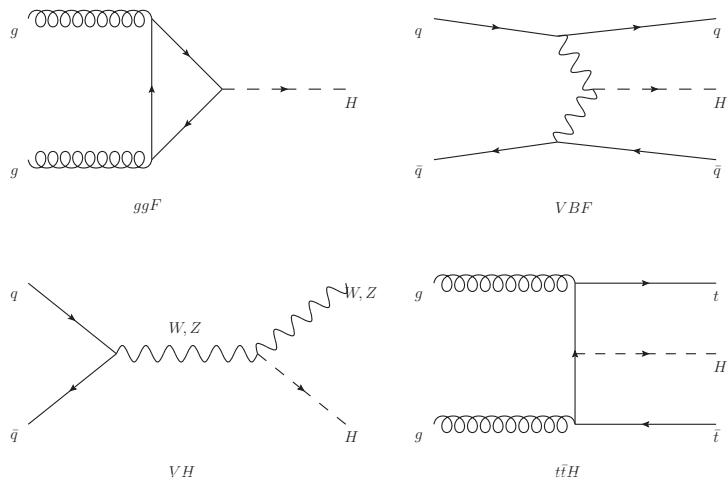


Figure 2.2: Dominant Higgs production modes.

⁴⁹⁵ some 87% of Higgs production at the nominal LHC center of mass energy of 14 TeV, is gluon-gluon
⁴⁹⁶ fusion (ggF), shown in Figure 2.2 (a). At high center of mass energies, most of a proton's momen-
⁴⁹⁷ tum is predominantly carried by sea gluons (as opposed to the constituent valence quarks associated
⁴⁹⁸ with the hadron's identity). This, along with the difficulties associated with high luminosity antipro-
⁴⁹⁹ ton beam production, is why the LHC was designed as a proton-proton collider instead of a proton-
⁵⁰⁰ antiproton collider (like the Tevatron). As mentioned above, the Higgs does not couple directly to
⁵⁰¹ gluons but must instead be produced through the fermion loop shown in the figure. The heaviest
⁵⁰² fundamental fermion by far is the top quark, with $m_t = 173$ GeV, so top loops dominate this pro-
⁵⁰³ cess. While not particularly relevant for this thesis, about 14% of events in the 2-lepton channel of
⁵⁰⁴ the $H \rightarrow b\bar{b}$ analysis are ggF initiated.

⁵⁰⁵ The next most prevalent process is vector boson fusion (VBF), where vector bosons (W or Z ,
⁵⁰⁶ denoted generically as V) from quarks in the colliding quarks “fuse” to form a Higgs. These quarks
⁵⁰⁷ typically form jets in the forward region, which provides a unique signature for this process. This
⁵⁰⁸ process is not relevant for this thesis.

⁵⁰⁹ The third leading process is “Higgsstrahlung” or Higgs production in association with a vector
⁵¹⁰ boson, often simply VH production. In this process, a quark-antiquark pair in the colliding protons
⁵¹¹ forms an energetic vector boson, which then radiates a Higgs (this is similar to photon emission of
⁵¹² accelerating electrons, called “bremsstrahlung,” hence the name). Some fraction of the time (about
⁵¹³ 21% of the time for WH and 6.7% of the time for ZH), the energetic V will decay leptonically (i.e.
⁵¹⁴ into a decay involving an electron or a muon), which provides a unique and triggerable signature for
⁵¹⁵ this process. Another 20% of the time for ZH production, the Z will decay to neutrinos, which are

516 not absorbed by detectors and show up as missing transverse energy (\vec{E}_T^{miss}), another triggerable sin-
 517 gature. This ability to trigger and require that this leptonic signature be consistent with a V allows
 518 one to significantly reduce the impact of multijet background (a very common generic processes at
 519 the LHC) on analysis. Hence, this is the process of primary importance to this thesis.

520 The final important Higgs production process is $t\bar{t}H$ production, the box diagram in Figure 2.2
 521 (d). Again, the top pair provides a useful signature for analysis. This, like VBF, is also not considered
 522 in this thesis.

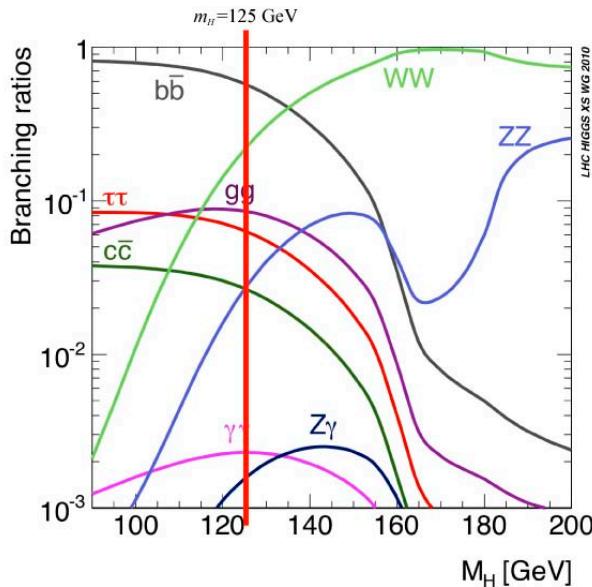


Figure 2.3: Higgs decay modes as a function of its mass; a line has been drawn at the observed Higgs mass of 125 GeV.

523 Once the Higgs has been produced, it can decay in a number of ways, as shown in Figure ???. By
 524 far the most dominant decay mode of the Higgs is to $b\bar{b}$ at 58% of all decays. This b -quark pair then
 525 hadronizes into two b -jets (for a more thorough discussion of jets and b -jets in particular, see Sec-
 526 tion 5.5). However, many processes at the LHC create pairs of b -jets with invariant masses consistent

527 with the Higgs have much higher production rates ($t\bar{t}$ production at the LHC is in the neighbor-
 528 hood of hundreds of pb), so a clear process signature is necessary to study $H \rightarrow b\bar{b}$ production at the
 529 LHC. This is why the bulk of search efforts have focused on VH production. A summary of Higgs
 530 production cross sections and simple extrapolations to raw numbers of Higgs bosons produced for
 531 VH for leptonically decaying V is shown in Table 2.1

\sqrt{s} (TeV)	ZH	WH	ggF	total σ	$N_{V \rightarrow \ell^+ \nu} H$
7	$0.34^{+4\%}_{-4\%}$	$0.58^{+3\%}_{-3\%}$	$15.3^{+10\%}_{-10\%}$	17.5	$4.7 \text{ fb}^{-1} \rightarrow 589$
8	$0.42^{+5\%}_{-5\%}$	$0.70^{+3\%}_{-3\%}$	$19.5^{+10\%}_{-10\%}$	22.3	$20.3 \text{ fb}^{-1} \rightarrow 3100$
13	$0.88^{+5\%}_{-5\%}$	$1.37^{+2\%}_{-2\%}$	$44.1^{+11\%}_{-11\%}$	50.6	$36.1 \text{ fb}^{-1} \rightarrow 11100$
14	$0.99^{+5\%}_{-5\%}$	$1.51^{+2\%}_{-2\%}$	$49.7^{+11\%}_{-11\%}$	57.1	$1000 \text{ fb}^{-1} \rightarrow 343000$

Table 2.1: Cross sections for processes important to the SMVHbb analysis and the total Higgs cross section as a function of center of mass energy. Also given are the total number of Higgs bosons produced for given luminosities through both WH and ZH processes.

532 2.3 COLLIDER EVENTS AND EVENT LEVEL VARIABLES

533 Collision data in experiments like ATLAS is structured using what is known as the *event data model*.
 534 In this model, one collision corresponds to one event. The raw data, the various tracks, energy de-
 535 posits, and hits in the detector, undergo reconstruction (described at length in Chapter 5) both
 536 through automated experiment-wide standards and through analysis-specific level selections, cor-
 537 rections, and calibrations. The result of this considerable effort is a collection of labeled 4-vectors,
 538 representing the final state objects. This is shown in Figure 2.4.

539 In the process that is the focus of this thesis, every event ultimately is condensed into a lepton

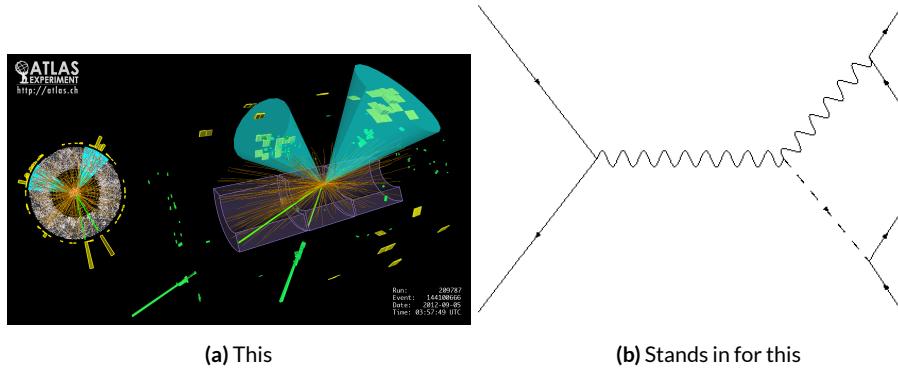


Figure 2.4: Reconstruction in a nutshell

540 pair (two electrons or two muons), two or three jets*, all 4-vectors and a \vec{E}_T^{miss} vector in the trans-
 541 verse plane. Further selection then takes place to winnow down events into interesting regions of
 542 phase space hopefully more rich in signal-like events. Once events are selected in a search like the
 543 one in this thesis, one then analyzes the data to test its consistency with some background only hy-
 544 pothesis to produce the usual significances quoted. This can be done in various ways, with main
 545 approaches being: a simple counting experiment (often referred to as the “cut and count” approach),
 546 a functional fit for excesses over a falling background spectrum (the so-called “bump hunt” used in
 547 analyses like the $H \rightarrow \gamma$ discovery channel), or the use of discriminant distributions as PDF’s in a
 548 likelihood fit (the approach of this analysis). These distributions can be simple counts (i.e. single bin
 549 distributions) in analysis regions, quantities of interest (the distribution of the invariant mass of the
 550 two b -jets in selected events with the greatest transverse momenta, m_{bb} , is used as a validation), or
 551 something more complicated like a multivariate analysis (MVA) discriminant.

*Sometimes more, though this is a small fraction of events, and the wisdom of this choice may be questioned

552 2.4 CHARACTERIZATION WITH EVENT-LEVEL VARIABLES

553 This is where our story truly begins. Traditionally, particle physicists have favored the approach of
554 using distributions of physical variables since it is easier to develop “physical intuition” for what
555 these distributions should “look like” during validation, so it is no surprise that as many LHC anal-
556 yses have transitioned to using MVA techniques that these variables form the basis of many very
557 robust physics results. These variables do quite well summarize many of the main physics features
558 of an event for the signal topology. In $ZH \rightarrow \ell\ell b\bar{b}$ events, for example, one wishes to characterize
559 the ZH system by using the lepton pair as a stand-in for the Z and the b -jet pair as a stand-in for the
560 H , and composite variables like m_{bb} and $m_{\ell\ell}$ can be used to check whether events are consistent with
561 these objects. There are also variables like p_T^V that characterize the momentum scale of the event,
562 angles like $\Delta R(b_1, b_2)$ and $\Delta\phi(V, H)$ that can be further used to characterize the overall “shape” of
563 these events, and variables like \vec{E}_T^{miss} that can discriminate against backgrounds like $t\bar{t}$ that do not
564 have a closed topology.

565 Nevertheless, the intuition based approach, with incremental addition of variables as they prove
566 useful in the lifetime of an analysis’s iterations, does beg the question of whether there is a more sys-
567 tematic way to treat this information. There are clearly patterns to which variables are useful: these
568 correspond to important information about the hypothesized physics objects and their relation-
569 ships, so could there be some gain in finding a way to systematize the way these are found? Such
570 systematic, top-down approaches often promise to increase performance in two ways. The first is
571 by having higher descriptive power, often through some sophisticated treatment of the missing

572 transverse energy in an event, \vec{E}_T^{miss} . \vec{E}_T^{miss} is just a single number, and if there is just one invisible
573 object in a desired event topology, using \vec{E}_T^{miss} on its own often provides sufficient sensitivity. In
574 more complicated topologies with multiple invisible particles in the final state, for example in many
575 supersymmetry searches, a more careful treatment of the missing energy is often necessary.

576 The second means of improvement is through using a more orthogonal basis of description,
577 which allows one to more efficiently use data and simulation samples. A more orthogonal basis
578 implies that variables contain less overlapping information with each other and so allow for a more
579 efficient exploration of parameter space. This means one can gain higher sensitivity from equiva-
580 lent datasets using a more orthogonal basis. To see why this might be the case, take an MVA dis-
581 criminant for $ZH \rightarrow \ell\ell b\bar{b}$ formed using only the classic variables $\Delta R(b_1, b_2)$ and p_T^V . In the
582 $ZH \rightarrow \ell\ell b\bar{b}$ topology, transverse mass of the Z and H (and hence the lepton pair and jet pair) are
583 equivalent. This means that at higher p_T^V the p_T of b -jets will also be higher, which in turn implies
584 that they will have a smaller angle of separation and hence a smaller $\Delta R(b_1, b_2)$. This correlation
585 is not unity—each variable still does have information the other does not—but it is still very high.
586 Hence, when training an MVA, which in principle knows nothing about these variables other than
587 some set limits, an undue number of training events will be wasted converging upon relations that
588 could be known *a priori*, and while this might be easy to hard code in for a two variable toy example,
589 the dimensionality of any real discriminant makes this prohibitive. An MVA that uses data (both
590 actual and simulated) more efficiently will also tend to be more robust to variations, offering a po-
591 tential avenue for reduction in the error on quantities of interest due to systematic uncertainties.
592 Details of how this plays out in a likelihood fit will be deferred to the discussion of the fit model

593 used in the VHbb search in Chapter 7.

594 Many of these novel schemes are designed to explicitly address the first issue in channels where
595 it is of paramount importance while having the second issue as something of a fringe benefit. How-
596 ever, as the amount of data taken at the LHC grows, analyses will increasingly become systematics
597 limited, so an exploration to the veracity of the second claim has great potential for the high lumi-
598 nosity era of the LHC. The $ZH \rightarrow \ell\ell b\bar{b}$ process offers a great setting for investigating this issue on
599 its own since its closed topology largely mitigates any improvement from more sophisticated treat-
600 ments of \vec{E}_T^{miss} . We introduce two of these more top-down approaches to event-level variables be-
601 low: the “Lorentz Invariant” (LI)⁴⁵ and “RestFrames inspired” (RF)⁴⁷ variable schemes. A broad
602 overview of the concepts behind these schemes will be given here, with a more in-depth discussion
603 of their implementation deferred until Chapter 6.

604 2.5 LORENTZ INVARIANTS

605 The LI variables, first put forth by S. Hagebeck and others⁴⁵, are based upon the fact that the four-
606 vectors of an event are determined, all of the information in an event are encoded into their inner
607 products (Lorentz invariant quantities, hence the name) and the angles between them. This makes
608 for 16 quantities in all: the ten inner products of the four vectors, the three Euler angles, and the
609 three parameters specifying the boost of the ZH system. The masses of the four final state objects
610 are not considered very useful and so can be removed to leave six meaningful inner products (the
611 ${}_4C_2$ combinations between distinct final state four vectors). Since these inner products can have an
612 ill-defined physical interpretation and in order to help MVA training, each inner product is scaled

by:

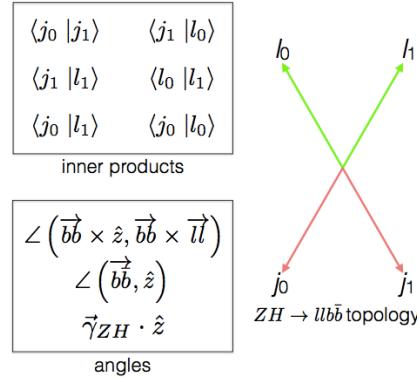


Figure 2.5: Summary of LI variables in the $ZH \rightarrow \ell\ell b\bar{b}$ topology.

613

$$x \rightarrow \frac{x}{x + c} \quad (2.7)$$

614 where c is the mean of the distribution in the signal MC distribution. These inner products are de-
 615 noted $x_i_y_j$, where x and y are either j (for jet) or ℓ (for lepton) and the indices are either o (i) for
 616 the leading (subleading) object by p_T in the event.

617 The number of useful angles can be reduced by recognizing some symmetries inherent in the
 618 final state. The symmetry around the beam axis eliminates one angle. Furthermore, the boost of the
 619 VH system is primarily in the beam direction (z) direction, marginalizing the utility of the transverse
 620 boost angles. This leaves the boost in the z direction, denoted `gamma_ZHz`, and two angles chosen to
 621 be the angle between the $b\bar{b}$ system and the beam (`angle_bb_z`) and the angle between $(b_1 + b_2) \times$
 622 \hat{z} and $(b_1 + b_2) \times (l_1 + l_2)$ (`angle_bbz_bbll`).

623 These variables do contain a lot of information similar to the usual set: there are mass equivalents

624 ($j\theta_{j1\leftrightarrow mBB}$, and $\ell\theta_{\ell1\leftrightarrow mLL}$) and angles. Instead of individual final state object scales, there are
 625 the four jet-lepton inner products, though this correspondence (and indeed any physical interpretation)
 626 is far from clear. An important advantage of the LI variable set is that all of the variables are in
 627 it are orthogonal in the signal case by construction. A drawback of this framework in a completely
 628 closed final state is that there is no way to treat E_T^{miss} in a Lorentz invariant way.

629 There is also no prescription for any additional jets in the event beyond the two b -tagged jets.
 630 They are simply ignored in these variable calculations since the fiducial analysis requirement of ex-
 631 actly two b -tagged jets eliminates any combinatoric ambiguity.

632 2.6 RESTFRAMES VARIABLES

633 The RestFrames variables⁴⁷, calculated using the software package of the same name, is based upon
 634 the idea that the most natural frame in which to analyze objects of the signal decay tree is in their in-
 dividual production (rest) frames. The signal decay tree for $ZH \rightarrow \ell\ell b\bar{b}$ is show in Figure 2.6. Gen-

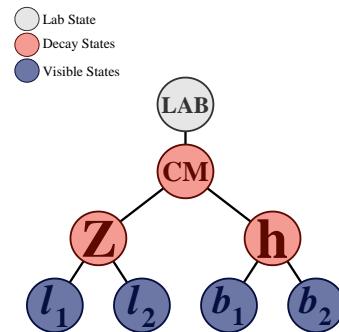


Figure 2.6: The $ZH \rightarrow \ell\ell b\bar{b}$ decay tree.

635
 636 erally, one does not typically have enough information to determine exactly each of the intermediate
 637 rest frames or the boosts between the frames, but in a completely closed final state like $ZH \rightarrow \ell\ell b\bar{b}$,

638 this can be done in the usual way by adding the four-vectors of the final state objects and solving the
639 usual equations from special relativity (`RestFrames` does this automatically for each event).

640 Each frame has associated with it the boost from its immediate parent and a mass scale; that mass
641 (in this case the correspondence between RF mass variables and standard mass variables is exact) and
642 the angles between the Euclidean three vector associated with boost and the axis of the decay prod-
643 ucts provide useful variables. In general, the polar angle (typically given as a cosine) is considered
644 more useful than the azimuthal angle (typically just a $\Delta\phi$), though this is dependent on the candi-
645 date decay tree. The Z frame, for example, has `MZ`, which is just the usual $m_{\ell\ell}$, `cosZ`, the cosine of
646 the polar angle between the lepton momentum axis in their production frame and the boost from
647 the ZH center of mass (CM) frame, and the angle `dphiCMZ`.

648 In addition to the masses and angles attached to individual object rest frames, energy scales associ-
649 ated with the CM frame can be used to contextualize other event level quantities. In particular, one
650 can use the mass of the CM frame as a natural scale to evaluate the momentum of the CM frame,
651 and the p_T of the CM frame as a natural scale for the event's E_T^{miss} , yielding the variables:

$$R_{p_T} = \frac{p_{T,CM}}{p_{T,CM} + M_{CM}}, \quad R_{p_z} = \frac{p_{z,CM}}{p_{z,CM} + M_{CM}}, \quad R_{met} = \frac{E_T^{miss}}{E_T^{miss} + p_{T,CM}} \quad (2.8)$$

652 denoted `Rpt`, `Rpz`, and `Rmet`. These can be thought of as behaving like significance based variables
653 in particle physics, like METHT or impact parameter significances. These are used instead of the
654 final state object scales and standard E_T^{miss} of the standard variable set.

655 Unlike the LI variables, the physical interpretation of RF variables is very clear. Everything has

656 physical units, and these are variables one might have introduced in the usual process of develop-
 657 ing an MVA with the traditional mindset. The solution to the issue of additional jets in an event is
 658 not immediately clear. In order to keep the two non-standard MVA's on as equal footing as possi-
 659 ble, the approach of simply ignoring additional jets is taken in this thesis. Nevertheless, it would be
 660 easy enough to redefine the H intermediate frame to have, for example, the two b -tagged jets and the
 661 highest p_T untagged jet for any subset of events. This flexibility is not a feature of the Lorentz Invari-
 662 ants framework. Of course, `RestFrames` cannot tell you what approach to take, but it is capable of
 663 handling more flexible topologies once optimization studies have been completed.

664 2.7 EXTENSIONS TO THE 1 AND 0 LEPTON CHANNELS

665 Both the LI and RF variable concepts are readily extendible to the 1-lepton channel. In this topology,
 666 one of the leptons in the $ZH \rightarrow \ell\ell b\bar{b}$ diagram is replaced by a neutrino, the lone invisible particle
 667 in the final state. We can assume that the neutrino has zero mass and transverse momentum equal to
 668 the \vec{E}_T^{miss} in the event, leaving one undetermined degree of freedom, the longitudinal momentum
 669 of the neutrino, p_z^ν .

670 The LI concept was in fact initially formulated for sensitivity improvement in the 1-lepton chan-
 671 nel, with the same orthogonality of variables described in the 2-lepton case being the main draw.
 672 The LI approach to estimating the neutrino longitudinal momentum is outlined in ⁴⁵, which we
 673 reproduce here. We first guess the neutrino energy in its rest frame and then boost to the lab frame:

$$\langle E_\nu \rangle = \frac{1}{4} m_{WH} \implies \langle p_z^\nu \rangle = \beta \gamma \langle E_\nu \rangle = \frac{p_z^{WH}}{m_{WH}} \langle E_\nu \rangle = \frac{1}{4} p_z^{WH} \quad (2.9)$$

⁶⁷⁴ Finally, assuming energy and momentum in aggregate are equally shared among final state constituents, we arrive at

$$\langle p_z^\nu \rangle = \frac{1}{4} \times \frac{4}{3} \left(p_z^l + p_z^{j_0} + p_z^{j_1} \right) \quad (2.10)$$

⁶⁷⁶ The RF approach for the 1-lepton case amounts to replacing the $Z \rightarrow \ell\ell$ in 2.6 with $W \rightarrow \nu$.

⁶⁷⁷ As alluded to in the 2-lepton discussion, when there is missing information in the final state from invisible particles and/or combinatoric ambiguities, recursive jigsaw reconstruction (RJR) offers a standard toolkit for deriving estimated boosts between rest frames by analytically minimizing on unknown quantities. While in more exotic final states with multiple invisible particles and combinatoric ambiguities the choice of jigsaw rule can be subjective, the case of W is well-studied and outlined in detail in Section V.A. of⁴⁷. It reproduces the usual transverse mass of the W in place of m_Z in the 2-lepton case.

⁶⁸⁴ The 0-lepton channel would appear to present some difficulty as two neutrinos in the final state introduce extra degrees of freedom, but both concepts may be extended by treating the invisibly decaying Z as a single invisible particle and requiring the Z to be on-shell. Both of these requirements may be folded into the 1-lepton framework to produce similar sets of variables.

⁶⁸⁸ *Maybe do the 0-lep calculation and some cute RF diagrams*

If it's stupid but it works, it isn't stupid.

Conventional Wisdom

3

689

690

Data and Simulated Samples

691 THE DATA AND Monte Carlo simulation (MC) samples are the same as in the fiducial analysis. The
692 data corresponds to 36.1 fb^{-1} of pp collision data collected in 2015+16 at the ATLAS detector at $\sqrt{s} =$
693 13 TeV. Only events recorded with all systems in ATLAS in good working order and passing certain
694 quality requirements, according to a Good Run List (GRL), are analyzed.

695 Details about MC samples may be found in³⁸. The $ZH \rightarrow \ell\ell b\bar{b}$ process is considered for both
696 multivariate analysis (MVA) optimization and the final statistical analysis, while $WH \rightarrow \ell\nu b\bar{b}$
697 and $ZH \rightarrow \nu\nu b\bar{b}$ production are included in the final statistical analysis only. Signal MC samples
698 were generated separately for qq and gg initiated VH processes. $qqVH$ samples were generated with
699 POWHEG MiNLO + PYTHIA8 with the AZNLO tune set and NNPDF3.0 PDF. Nominal $ggZH$
700 samples were generated using POWHEG for the matrix element (ME) and PYTHIA8 for the parton
701 shower (PS), underlying event (UE), and multiple parton interactions (MPI), again applying the
702 AZNLO tune and NNPDF3.0 PDF set.³¹

703 The background processes considered in these studies are $Z+jets$, $t\bar{t}$, and diboson production
704 for both MVA optimization and the final statistical analysis and single top production and $W+jets$
705 only considered in the final statistical analysis. $V+jets$ samples are generated using SHERPA 2.2.1³²
706 for both the ME and PS. These samples are generated in different groups, according to the identity
707 of the V , the max (H_T, p_T^V) of events, and also further subdivided according to the flavor of the
708 two leading jets in an event, b , c , or l , for a total of six categories.. $t\bar{t}$ samples are generated using
709 POWHEG with the NNPDF3.0 PDF set interfaced with PYTHIA8 using the NNPDF2.3 PDF's and
710 the A14 tune³⁸. Single top samples use POWHEG with the CT10 PDF's interfaced with PYTHIA6
711 using the CTEQ6L1 PDF's^{14,48}. Diboson samples are generated with SHERPA 2.2.1 interfaced with
712 the NNPDF3.0 NNLO PDF set normalized to NLO cross sections²⁵.

If it's stupid but it works, it isn't stupid.

Conventional Wisdom

4

713

Signal and Bacground Modeling

714

715 THIS CHAPTER summarizes the modeling of the dominant signal and background processes in this
716 analysis, including corrections and systematic uncertainties (set in **this** font) related to each process.
717 Further details on the specifics of these topics, including in-depth studies for the derivation and def-
718 initions of some of the quantities cited, may be in ⁵⁸. We start with a general discussion of modeling

719 and associated major categories of uncertainties before addressing each of the physics processes in
720 turn.

721 **4.1 EVENT GENERATION IN A NUTSHELL**

722 Before diving into the minutiae of the modeling and systematic uncertainties associated with each
723 major set of physics processes considered in this analysis, we review at a schematic level*, the problem
724 of simulation event generation, namely, once a physics processes of interest has been determined,
725 how does one simulate an ensemble of particle collisions with the process in question. This is il-
726 lustrated in Figure 4.1. Note that the scope of this problem does not include how these generated
727 collision products propagate through one's detector. This problem is left for Chapter 5.

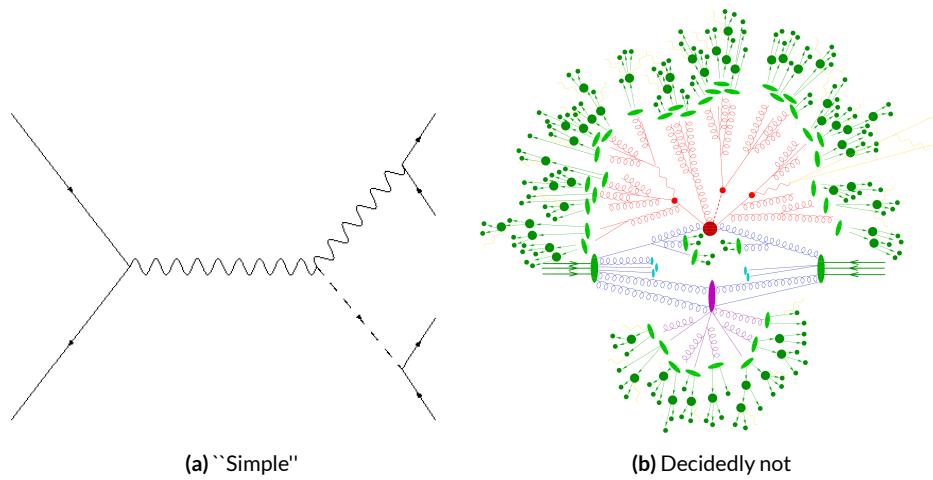


Figure 4.1: The problem here is how to get from (a) to (b).

728 The primary source of complication in event generation comes from dealing with hadronic ob-

*i.e. this will not be a technically rigorous discussion. For a more thorough treatment, the reader is directed to the usual references.

729 jects both in the initial state (the lefthand side of Figure 4.1 (a); the LHC is a hadron collider) and
730 the final state (this analysis searches for Higgs decays to b -jets, the lower righthand side of Figure 4.1).
731 Common to all hadronic objects, by definition, are the many considerations that go into calculations
732 in quantum chromodynamics (QCD). In calculating the hard scatter process itself, one must make
733 a variety of choices, such as the parton distribution function (PDF) set to use and to what order in
734 perturbation theory to do the calculation (common choices are leading order (LO), (next to) next
735 to leading order ((N)NLO), and (next to) next to leading log (NNLL)). Similar considerations of-
736 ten need to be made for the electroweak parts of an event. These considerations and others will be
737 discussed in more detail below.

738 The initial state includes not only the hard scatter partons that generate the physics process of
739 interest but also the rest of partons in the colliding protons, known as the underlying event (UE).
740 Moreover, the hard scatter partons may not be the only interacting partons in an event, further com-
741 plicating matters; this phenomenon is known as multiple parton interactions (MPI). Specific to the
742 final state are the kinematic distributions of the final state objects—what their energies and angular
743 distributions will be in addition to the overall cross section of the process as measurable by the de-
744 tector (acceptance effects). Furthermore, one has to model hadronization, the process by which any
745 free (colored) partons in an event transform into colorless hadrons.

746 Typically, it takes several steps and tools to accomplish this. The hard scatter itself is often done
747 with a dedicated event generator like PowHEG⁵⁹ or MadGraph¹⁶, with events generated being in-
748 terfaced with a tool like Pythia⁶² for the PS, UE, and MPI, though there are exceptions (Sherpa⁴³,
749 for example, can do both for some processes).

750 4.2 DESCRIPTION OF MODELING UNCERTAINTY CATEGORIES

751 The general idea behind a systematic uncertainty[†] is to characterize the scale of variation for a given
752 process in such a way that this knowledge can be easily combined with all other sources of uncer-
753 tainty in an analysis to give total uncertainties for categories of systematics, different channels of
754 analysis, and the analysis as a whole (as well as the single systematic itself).

755 The modeling systematics in this analysis are parametrized in a number of ways but are often
756 characterized as Gaussians centered at zero, and so the relevant quantity to derive is the standard
757 deviation for this distribution. In practice, one does this by, in addition to making distributions
758 of discriminants (usually an MVA or m_{bb}) for the nominal case, distributions corresponding to a
759 variation in a given systematic uncertainty by both one standard deviation up and one standard
760 deviation down. Modeling systematics are exclusively weight systematics; that is, they only effect the
761 weight with which an event is added to discriminating distributions.

762 Modeling systematics are derived separately for each physics process (simulation sample). Some-
763 times, all of the variation for a given process is encapsulated in a single systematic, but oftentimes the
764 variations from multiple considerations are distinct enough to be treated separately. Furthermore,
765 each of these separate systematics for a given sample may be treated in a number of ways. We will
766 begin by outlining in more detail the former issue of theoretical considerations before describing the
767 latter problem of different types of systematics.

[†]The terms “systematic uncertainty,” “systematic,” “uncertainty,” and “nuisance parameter” (often simply NP in both writing and speech) are often used interchangeably, though each has its own pedantically distinct definition.

768 The motivation of this section is to give prototypical examples for certain treatments of systemat-
769 ics to prevent verbosity and redundancy when describing each sample in detail.

770 **4.2.1 PHYSICS CONSIDERATIONS**

771 As described in the preceding section, many choices must be made when generating events for anal-
772 ysis studies. In general, evaluating the uncertainties arising from these choices entails generating
773 alternate samples of events, which practically means tuning parameters in the various software pack-
774 ages and/or using alternate packages/libraries to make new samples. Once these samples have been
775 created, they are compared at truth-level (particle level) using a package called Rivet²³ instead of us-
776 ing the full ATLAS detector reconstruction for computational considerations. Given the nature of
777 the problem and the tools, there are generally three main categories of physics issues, each described
778 below.

779 **UNDERLYING EVENT AND PARTON SHOWER**

780 The modeling of the underlying event (UE) and the parton shower (PS) are usually handled by the
781 same package, typically PYTHIA8 and so are usually treated together. One approach to modeling
782 these uncertainties is to simply see what happens when a different model is used and then compare
783 this alternate set of events to the nominal set, taking the difference as the scale of variation. Another
784 approach is to vary some parameter within a given model, for example, using different tunes in the
785 A14 set for PYTHIA8 with their accompanying variations, to characterize the scale of variation.

786 A natural question is how to treat these two approaches on the same footing. When examining

787 a set of potential variations related to the same process or effect, oftentimes the largest single varia-
 788 tion in a set is picked as defining the scale for the systematic uncertainty; another approach is to use
 789 the average over a set of variations. The ATLAS_UEPS_VH_hbb systematic, for example, uses the
 790 Pythia8 + A14 tunes approach to determine the scale of UE variation and compares Pythia8 with
 791 Herwig7 to characterize the PS variation. Each of the A14 tunes comes with an up and down varia-
 792 tion, and the difference between each of these variations and a nominal setup may be expressed as a
 793 ratio, R , of total events.

794 As is often done when a physical argument can be made for combining related, but ultimately
 795 orthogonal categories/measurements/uncertainties/systematics, the combined UE+PS systematic is
 796 taken to be the sum in quadrature of these two effects:

$$\sum_{tunes} \max_{tune} (|R_{up} - R_{down}|) \oplus \sigma_{PS} \quad (4.1)$$

797 QCD SCALE

798 The term “QCD scale” in the context of modeling uncertainties refers to the choice of renormal-
 799 ization (μ_R) and factorization (μ_F) scales used in QCD calculations. These are typically treated to-
 800 gether. Usually, some multiplicative scale factor, f , is chosen, and each scale is varied in concert with
 801 the other scale by $1/f$ and $1/f^2$ (nine total combinations), sometimes with a cap on how large the com-
 802 bined variation can be (so ignoring the (f,f) and $(1/f, 1/f)$ cases). Just as in the UE+PS, the largest
 803 variation is usually taken as the systematic uncertainty.

804 PARTON DISTRIBUTION FUNCTIONS AND α_s

805 Finally, separate uncertainties are often made for the choice of parton distribution function (PDF)
806 set and associated choice of strong coupling for QCD (α_s). Much as in the previous two cases, one
807 can vary the parameter α_s and study what samples of simulation events made using different PDF
808 sets relative some nominal setup look like. Similarly, one can take the maximum, average, or sum in
809 quadrature of different variations to characterize a systematic uncertainty.

810 4.2.2 MODELING SYSTEMATIC TYPES

811 ACCEPTANCE/NORMALIZATION

812 The most basic type of modeling uncertainty is a normalization uncertainty, often called an accep-
813 tance uncertainty. This simply denotes the uncertainty on the number of predicted events for a
814 given process in a given region of phase space (usually delineated by the number of leptons in the
815 final state sometimes by the number of and jets the p_T^V of an event) and is usually expressed as a per-
816 cent.

817 As an example, the uncertainty on the theoretical prediction of the $H \rightarrow b\bar{b}$ branching ratio,
818 denoted ATLAS_BR_bb (it is an ATLAS-wide systematic), is expressed as a normalization systematic
819 with a value of 1.7%, affecting all VH processes. Now imagine we have an event in a VH sample with
820 weight 1.0. The nominal histograms for this region gets filled with this event's relevant information
821 with weight 1.0, while the ATLAS_BR_bb__1up (__1do) histograms get filled with weight 1.017
822 (0.983).

823 **SHAPE SYSTEMATICS**

824 In addition to normalization systematics expressed as single numbers attached to different processes
825 in different regions, there are also the so-called “shape systematics” and “shape corrections.” These
826 have the schematic form

$$w_{event} = \mathcal{A}_{region} \times f_{region}(event)$$

827 where w_{event} is the simulated event’s weight, \mathcal{A}_{region} is the overall normalization (in principle in-
828 cluding any systematics), and $f_{region}(event)$ is some function of event-level variables, usually a single
829 variable, like p_T^V or m_{bb} . The purpose of these systematics is to take into account (in the case of a
830 systematic) or correct (in the case of a correction applied to the event weight) the non-trivial depen-
831 dence of a normalization on one of these quantities. Some of these are taken from histograms while
832 others are parametric functions (in this analysis, usually linear ones).

833 An example of the former case is the quantity δ_{EW} , the difference between the nominal $qqVH$
834 cross section and the differential cross section as a function of p_T^V at next to leading order (NLO). As
835 a correction, this term is simply used as a correction factor $k_{EW}^{NLO} = (1 + \delta_{EW})$.

836 An example of the latter case is the systematic associated with the m_{bb} dependence of the the
837 $t\bar{t}$ normalization for 2 jet, $p_T^V \in [75, 150]$ GeV, 2 lepton events. In this case, a variety of effects are
838 studied (ME, PS, UE), but the treatment of the ME calculation was seen to have the largest effect
839 on normalizaiton, so a linear fit to reasonably envelope the largest variation was done, and this was
840 taken as a systematic variation. [‡]

[‡]If this all seems a little ad-hoc, that’s because it is, but, at least in this instance, the idea was that a single

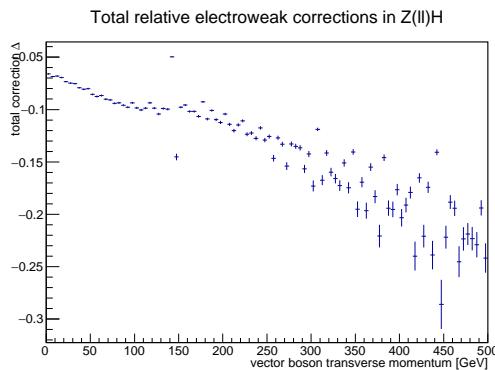


Figure 4.2: The δ_{EW} correction term for 2-lepton $qqZH$.

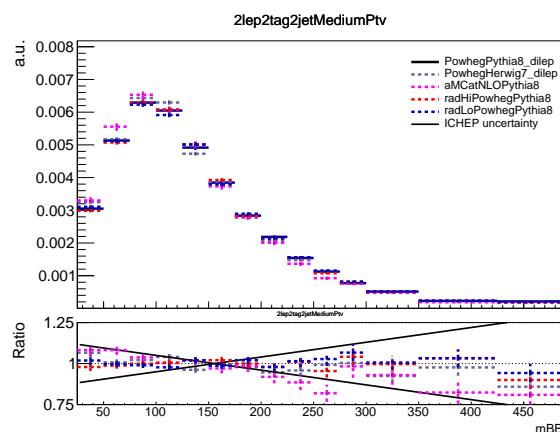


Figure 4.3: The derivation of the 2-lepton $t\bar{t} m_{bb}$ shape systematic.

841 DIVIDING MODELING UNCERTAINTIES: ACCEPTANCE RATIOS

842 In addition to uncertainties on absolute normalizations (both inclusive and region specific), model-
843 ing uncertainties are sometimes introduced for the ratio of normalizations in different regions. The
844 two main categories are ratios between different flavor regions and between regions with different
845 numbers of jets (henceforth n_{jet} regions). The former category is only relevant for $V+jets$ system-
846 atics and will be treated in that process's dedicated section below. In order to discuss the n_{jet} ratio
847 systematics, we must first describe how exclusive n_{jet} cross section calculations are done.[§]

848 THEORETICAL ASIDE: STEWART-TACKMANN A way to calculate uncertainties on processes in re-
849 gions with different numbers of jets was developed by Stewart and Tackmann and is implicitly used
850 for most n_{jet} ratio systematics⁶⁴. The problem is how to calculate the cross section and associated
851 uncertainty for a process with exclusively N jets in the final state. Generically:

$$\sigma_{\geq N} = \sigma_N + \sigma_{\geq N+1} \quad (4.3)$$

852 The physical interpretation of one parton to one jet is an idealized case. In order to demarcate
853 between jets, one has some quantity that is used as a cutoff in an integral that defines the border
854 between jet regions.

$$\sigma_{\geq N} = \int_0^{p_{cut}} \frac{d\sigma_N}{dp} + \int_{p_{cut}} \frac{d\sigma_{\geq N+1}}{dp} \quad (4.4)$$

systematic covered the largest effect, so others were unnecessary.

[§]We don't really need to do this, but everyone seems to mention Stewart-Tackmann, and no one ever explains it. The upshot is fairly simple, but the reasoning isn't necessarily so obvious.

855 Since these cutoffs (not necessarily constant, etc.) can make calculations more complicated, includ-
 856 sive cross sections tend to be easier to calculate. Hence, it is usually much easier to evaluate the two
 857 inclusive cross sections and find the uncertainties on these by varying α_s in the usual way. One then
 858 assumes the inclusive uncertainties are uncorrelated, for a covariance matrix for $\{\sigma_{\geq N}, \sigma_N, \sigma_{\geq N+1}\}$
 859 of:

$$\Sigma = \begin{pmatrix} \Delta_{\geq N}^2 & \Delta_{\geq N}^2 & 0 \\ \Delta_{\geq N}^2 & \Delta_{\geq N}^2 + \Delta_{\geq N+1}^2 & -\Delta_{\geq N+1}^2 \\ 0 & -\Delta_{\geq N+1}^2 & \Delta_{\geq N+1}^2 \end{pmatrix} \quad (4.5)$$

860 The main idea is that you have Sudakov double logs of p/Q , where $Q = m_H$ or whatever scale
 861 your hard process occurs at, and p_{cut} is usually something like a p_T cutoff. Now, the $N + 1$ term
 862 in that matrix is actually some uncertainty associated with your cutoff, but your double logs will
 863 dominate your higher order terms with Stewart and Tackmann giving the following reasoning:

864 “In the limit $\alpha_s L^2 \approx 1$, the fixed-order perturbative expansion breaks down and the logarithmic
 865 terms must be resummed to all orders in α_s to obtain a meaningful result. For typical experimental
 866 values of p_{cut} fixed-order perturbation theory can still be considered, but the logarithms cause large
 867 corrections at each order and dominate the series. This means varying the scale in α_s in Eq. (9) di-
 868 rectly tracks the size of the large logarithms and therefore allows one to get some estimate of the size
 869 of missing higher-order terms caused by p_{cut} , that correspond to Δ_{cut} . Therefore, we can approxi-
 870 mate $\Delta_{cut} = \Delta_{\geq 1}$, where $\Delta_{\geq 1}$ is obtained from the scale variation for $\sigma_{\geq 1}$.”

871 The above considerations are important for this analysis since phase space is separated into got 2
 872 and ≥ 3 jet regions, and the uncertainties for these regions are anti-correlated.

873 Of particular importance to ratio systematics is the so-called “double ratio” that is often take as
874 the scale of variation (plus one). The ATLAS_UEPS_VH_hbb systematic mentioned above, for exam-
875 ple, has associated with it, ATLAS_UEPS_VH_hbb_32JR. This systematic is evaluated by dividing the
876 3 jet to 2 jet ratio in the nominal setup by the same ratio in an alternate setup. Such a ratio generically
877 looks like:

$$\frac{\text{Acceptance}[\text{Category}_A(\text{nominalMC})]}{\text{Acceptance}[\text{Category}_B(\text{nominalMC})]} \Big/ \frac{\text{Acceptance}[\text{Category}_A(\text{alternativeMC})]}{\text{Acceptance}[\text{Category}_B(\text{alternativeMC})]} \quad (4.6)$$

878 4.3 PROCESS SPECIFIC SYSTEMATIC SUMMARIES

879 Brief descriptions of modeling systematics, including recapitulations of nominal sample generation,
880 are given in the following sections. The general approach here is to copy the relevant summary tables
881 and describe any major deviations from the general procedures described in the previous section.
882 The dominant backgrounds for the 2-lepton channel are $Z+hf$ and $t\bar{t}$, accounting for well over 90%
883 of all background events. Diboson samples are the next-leading background and are an important
884 validation sample; others are included for completeness. A summary of all the modeling systematics
885 in this analysis are given in Table 4.1.

886 4.3.1 SIGNAL PROCESSES

887 Nominal signal $qqVH$ samples are generated using PowHEG with the MiNLO (multiscale improved
888 NLO)⁵³ procedure applied interfaced with PyTHIA8 using the AZNLO tune¹² and NNPDF3.0

Process	Systematics
Signal	$H \rightarrow bb$ decay, QCD scale, PDF+ α_s scale, UE+PS (acc., p_T^V , m_{bb} , 3/2 jet ratio)
Z+jets	Acc, flavor composition, $p_T^V+m_{bb}$ shape
$t\bar{t}$	Acc, $p_T^V+m_{bb}$ shape
Diboson	Overall acc., UE+PS (acc, p_T^V , m_{bb} , 3/2 jet ratio), QCD scale (acc (2, 3 jet, jet veto), p_T^V , m_{bb})
Single top	Acc., $p_T^V+m_{bb}$ shape

Table 4.1: Summary of modeling systematic uncertainties, with background samples listed in order of importance.

889 PDF set²¹. For the 2-lepton case, gluon fusion initiated Higgs production is also considered (ac-
 890 counting for $\sim 14\%$ of the total cross section in this channel), with samples generated with Powheg interfaced
 891 with Pythia8 using the AZNLO tune; the NNPDF2.3 set²⁰ is used for both the ME and UE+PS.

892 Alternate samples $qqVH$ samples are generated using MadGraph5_aMC@NLO¹⁵ for the ME
 893 and Pythia8 for the UE+PS, hadronizaiton and MPI. The NNPDF2.3_5f FFN PDF sets and the
 894 Al4 tune⁶; the latter has variations included. Powheg+Minlo+Herwig⁷ were samples were also
 895 used for systematics.

896 The signal systematics categories are $H \rightarrow bb$ decay cross section, QCD scale, PDF+ α_s scale, and
 897 UE+PS. Additionally, there is the NNLOEWK correction described above. The correction scale factor
 898 is derived using the HAWK MC software. To encapsulate NNLOEW effects the maximum of 1%, the
 899 square of the correction factor, and the photon induced cross section is used as a systematic.

900 Table 4.2, reproduced from⁵⁸, summarizes the signal cross section systematics, which are applied
 901 uniformly across the analysis channels (as applicable).

902 The remaining signal systematics are analysis channel specific and are summarized in 4.3. The
 903 methodologies match those described in 5. The UE+PS systematics were derived using the alternate

Sys Name	source	Norm. effect	applied to
ATLAS_BR_bb	$H \rightarrow bb$ decay uncertainties (HO effects, m_b , α_s)	1.7%	all VH processes
ATLAS_QCDscale_VH	QCD scale uncertainty	0.7%	$qq \rightarrow VH$ processes
ATLAS_QCDscale_ggZH	QCD scale uncertainty	27%	$gg \rightarrow ZH$
ATLAS_pdf_Higgs_VH	PDF+ α_s uncertainty	1.9% 1.6%	$qq \rightarrow WH$ $qq \rightarrow ZH$
ATLAS_pdf_Higgs_ggZH	PDF+ α_s uncertainty	5.0%	$gg \rightarrow ZH$

Table 4.2: Summary of all systematic uncertainties on the VH cross section including their value, source and the corresponding nuisance parameter name.

samples mentioned above; QCD scale uncertainties were derived by varying scales by $1/3$ and 3 ; and PDF uncertainties were derived by comparing the nominal set with the PDF4LHC15_30 PDF set²².

NP name	0L:		1L:		2L:	
	2j	3j	2j	3j	2j	$\geq 3j$
ATLAS_UEPS_VH_hbb	10.0%	10.0%	12.1%	12.1%	13.9%	13.9%
ATLAS_UEPS_VH_hbb_32JR	–	13.0%	–	12.9%	–	13.4%
ATLAS_UEPS_VH_hbb_VPT	shape only			shape+norm		
ATLAS_UEPS_VH_hbb_MBB	shape only					
QCDscale_VH_ANA_hbb_J2	6.9%	–	8.8%	–	3.3%	–
QCDscale_VH_ANA_hbb_J3	-7%	+5%	-8.6%	+6.8%	-3.2%	+3.9%
QCDscale_VH_ANA_hbb_JVeto	–	-2.5%	–	3.8%	–	–
QCDscale_VH_ANA_hbb_VPT	shape only			shape+norm		
QCDscale_VH_ANA_hbb_MBB	shape only					
pdf_HIGGS_VH_ANA_hbb	1.1%	1.1%	1.3%	1.3%	0.5%	0.5%
pdf_VH_ANA_hbb_VPT	shape only			shape+norm		
pdf_VH_ANA_hbb_MBB	shape only					

Table 4.3: Summary of all systematic uncertainties on the VH acceptance and shapes originating from altering the PDF and α_s uncertainties, including their corresponding nuisance parameter name.

906 4.3.2 $V+JETS$

907 Nominal $V+jets$ samples are generated using **SHERPA 2.2.1@NLO**^{¶ 44} for both the ME and PS,
908 interfaced with the NNPDF's and using a five quark flavor scheme, and alternative samples are de-
909 rived using **MADGRAPH5** interfaced with **PYTHIA8**. In order to increase statistics in important
910 regions of phase space, these samples were separated into kinematic slices based on p_T^V and into
911 bins of jet flavor. The kinematic slices were in the quantity $\max(H_T, P_T^V)$ and had the intervals
912 $[070, 70140, 140280, 280500, 5001000, > 1000]$ GeV. The jet flavor slices were made using flavor
913 vetoes and filters:

- 914 • BFilter: at least 1 b-hadron with $|\eta| < 4, p_T > 0$ GeV
- 915 • CFilterBVeto: at least 1 c-hadron with $|\eta| < 3, p_T > 4$ GeV; veto events which pass the
916 BFilter
- 917 • CVetoBVeto: veto events which pass the BFilter or the CFilterBVeto

918 These in turn are related to the main flavor regions used in the analysis, based on the flavor of
919 the two leading jets in an event (based on p_T). These five flavors (with up, down, and strange collec-
920 tively known as “light”) yield six different flavor combinations: bb, bc, bl (these first three collectively
921 known as “heavy flavor” or $V+hf$), cc, cl, ll (or just “light” or l). Ratio systematics are often made
922 with respect to the acceptance in the bb region.

923 $V+jet$ systematics are derived in several steps. The first is to use double ratios of acceptances
924 between analysis regions and nominal versus alternative MC's (so $(\text{Region1-nominal}/\text{Region2-}$
925 $\text{nominal})/(\text{Region1-alternate}/\text{Region2-alternate})$). The main region comparisons are 2 jet versus 3

[¶]SHERPA 2.1 is used for some variations not available in SHERPA 2.2.1.

926 jet (z + jet for z -lepton) and then o -lepton versus z -lepton (i -lepton) for $Z+hf$ ($W+hf$ [¶]). The final
 927 uncertainty contains the sum in quadrature of four effects:

- 928 1. Variation of $o.5$ and z of QCD scales in the SHERPA sample
 929 2. Sum in quadrature of half the variation from different resummation and CKKW merging
 930 scales **
 931 3. Maximal variation between nominal setup and SHERPA 2.2.1 with the MMHT2014nnlo68cl
 932 and CT14nnlo PDF sets
 933 4. Difference between the SHERPA and MADGRAPH5 sets

934 Summaries of the Z +jets uncertainties are provided here; the reader is referred to^{§8} for the W +jets
 935 systematics, as these events are virtually non-existent in the z -lepton case with which this thesis is al-
 936 most exclusively concerned. In Table 4.4, from^{§8} are the normalization systematics.

Process	Name	prior in region					
		z jet			≥ 3 jets		
		zL : low Vpt	zL : high Vpt	oL	zL : low Vpt	zL high Vpt	oL
$Z+1$ $Z+cl$ $Z+hf$	SysZclNorm				18%		
	SysZlNorm				23%		
	norm_Zbb				Floating Normalisation		
$Z+hf$	SysZbbNorm_L2_J3	–	–	–	30%	30%	–
	SysZbbNorm_J3	–	–	–	–	–	17%
	SysZbbNorm_0L	–	–	7%	–	–	7%
	SysZbbPTV	effect on each region obtained from shape rw					

Table 4.4: Effect of modelling systematics on Z +jets normalisation in the z lepton regions. For systematic uncertainties implemented with a prior the effect of $1-\sigma$ variation is reported. The uncertainties labelled as Zbb act on the entire $Z+hf$ background.

937 The flavor composition ratio systematics are in Table 4.5, also from^{§8}.

[¶]The $W+hf$ CR versus the SR is also considered for $W+hf$

** cf.^{§9}, Section 2 for a summary of the CKKW method for different parton multiplicities used in SHERPA

Category	Nuisance Parameter Name	Prior	Applied to
$Z+bc/Z+bb$	SysZbcZbbRatio	40%	$Z+bc$ events (0-Lepton)
		40%	$Z+bc$ events (2-Lepton 2jet)
		30%	$Z+bc$ events (2-Lepton ≥ 3 jet)
$Z+bl/Z+bb$	SysZblZbbRatio	25%	$Z+bl$ events (0-Lepton)
		28%	$Z+bl$ events (2-Lepton 2jet)
		20%	$Z+bl$ events (2-Lepton ≥ 3 jet)
$Z+cc/Z+bb$	SysZccZbbRatio	15%	$Z+cc$ events (0-Lepton)
		16%	$Z+cc$ events (2-Lepton 2jet)
		13%	$Z+cc$ events (2-Lepton ≥ 3 jet)

Table 4.5: The priors on the relative acceptance variations for $Z+hf$. The first column details the flavour components across which the acceptance variation is being considered, the second column lists the names of the corresponding nuisance parameter in the Profile Likelihood Fit, the third contains the value of the prior and the fourth column the processes and categories to which this nuisance parameter is applied.

Finally, the p_T^V and m_{bb} shape systematics are derived using control regions in data^{††}. The functional form for the p_T^V systematic is $\pm 0.2 \log 10(p_T^V/50\text{GeV})$, and that of the m_{bb} systematic is $\pm 0.0005 \times (m_{jj} - 100\text{ GeV})$.

4.3.3 TOP-PAIR PRODUCTION

Nominal $t\bar{t}$ samples are produced with PowHEG at NLO for the ME calculation using the NNPDF3.0 PDF set interfaced with Pythia8.210 using the A14 tune and the NNPDF2.3 PDF set at LO. The parameters `hdamp` (nominal value $1.5 m_{top}$, a resummation damping factor for ME/PS matching that can heuristically thought of as tuning high p_T radiation) in PowHEG and `pThard` (nominal value 0) and `pTdef` (nominal value 2) in Pythia (both control merging with PowHEG) are varied to evaluate certain systematics. Alternative $t\bar{t}$ samples use PowHEG+Herwig7, MadGraph55_aMC@NLO+Pythia8.2,

^{††}These use the same selections as the signal regions except for b -tags (0, 1, and 2 tags are studied), with the added requirement in 2tag regions that m_{bb} not be in the range of 110–140GeV.

and the nominal setup with varied tunes and parameter values. Uncertainties are taken to cover the largest difference between the nominal and any of these alternate configurations.

The overall $t\bar{t}$ normalization is a floating normalization[#], and further systematics attached to the ratio of acceptances between regions (3-to-2 jet, SR-to-WhfCR, and 1-to-0 lepton) are defined using double ratios; these are summarized in Tables 4.6 and 4.7, taken from^{s8}

Systematic	0-lepton		1-lepton			
	2j	3j	WCR 2j	SR 2j	WCR 3j	SR 3j
norm_ttbar	floating normalisation					
SysttbarNorm_L0	8%	8%	–	–	–	–
SysttbarNorm_J2	9%	–	9%	9%	–	–
SysttbarNorm_DWhfCR_L1	–	–	25%	–	25%	–

Table 4.6: Effect of modelling systematics on normalisation in the 0 and 1-lepton analysis region.

	2jet		≥ 3 jets	
	low Vpt [SR/CR]	high Vpt [SR/CR]	low Vpt [SR/CR]	high Vpt [SR/CR]
norm_ttbar_J2_L2	floating normalisation		–	
norm_ttbar_J3_L2	–		floating normalisation	
SysTTbarPTV_L2_L2	effect on each region obtained from shape rw			

Table 4.7: Effect of modelling systematics on normalisation in the 2lepton regions. The SysTTbarPTV_L2_L2 systematic is implemented as a shape systematic over the full $VpT > 75$ GeV range, and as a result has different acceptance effects in the low and high VpT regions.

Shape systematics for p_T^V and m_{bb} are linear and taken to cover the largest difference reasonably well, as described above in 4.2.2. These are summarized in Table 4.8, again taken from^{s8}.

[#]The use of a top $e - \mu$ control region helps constrain this.

Analysis region	Uncertainty	Value	Source	Nuisance Parameter
0,1 lepton	p_T^V shape	shape	fit through largest deviation (aMC@NLO+PYTHIA8)	TTbarPTV
2 lepton	p_T^V shape	norm + shape	fit through largest deviation (aMC@NLO+PYTHIA8)	TTbarPTV_L2
0,1 lepton	$m_{b\bar{b}}$ shape	shape only	fit through largest deviation (aMC@NLO+PYTHIA8)	TTbarMBB
2 lepton	$m_{b\bar{b}}$ shape	shape only	fit through largest deviation (aMC@NLO+PYTHIA8)	TTbarMBB_L2

Table 4.8: Summary of all shape uncertainties for the $t\bar{t}$ process with short descriptions and the name of the corresponding nuisance parameters.

955 4.3.4 DIBOSON PRODUCTION

956 Three diboson production processes (collectively denoted VV) are important for these analyses: ZZ ,
 957 WZ , and WW . Nominal samples are created using **SHERPA 2.2.1** using the NNPDF3.0 PDF set. Al-
 958 ternative samples use **PowHEG+Pythia8** and **PowHEG+Herwig++**. The methodology here is
 959 similar to that of the $t\bar{t}$ systematics, with both overall acceptance and lepton channel specific uncer-
 960 tainties, with the exception that UE+PS and QCD scale are treated separately (PDF+ α_S was found
 961 to be negligible). p_T^V shape systematics are described using linear fits, while $m_{b\bar{b}}$ shape systematics
 962 are described using hyperbolic tangents (third degree polynomials) in the 2 jet (3 jet) regions. Once
 963 again, summary tables from⁵⁸ are reproduced here.

Sys Name	source	Norm. effect	applied to
SysWWNorm	overall cross section uncertainty	25%	WW in all regions
SysWZNorm	overall cross section uncertainty	26%	WZ in all regions
SysZZNorm	overall cross section uncertainty	20%	ZZ in all regions

Table 4.9: Summary of all systematic uncertainties on the diboson cross section including their value, source and the corresponding nuisance parameter name.

NP name	oL:		iL:		zL:	
	2j	3j	2j	3j	2j	$\geq 3j$
SysVZ_UEPS_Acc	5.6%	5.6%	3.9%	3.9%	5.8%	5.8%
SysVZ_UEPS_32JR	–	7.3%	–	10.8%	–	3.1%
SysVZ_UEPS_VPT	shape+norm		shape only		shape+norm	
SysVZ_UEPS_MBB	shape only					
SysVZ_QCDscale_J2	10.3%	–	12.7%	–	11.9%	–
SysVZ_QCDscale_J3	-15.2%	+17.4%	-17.7%	+21.2%	-16.4%	+10.1%
SysVZ_QCDscale_JVeto	–	+18.2%	–	+19.0%	–	–
SysVZ_QCDscale_VPT	shape+norm		shape only		shape+norm	
SysVZ_QCDscale_MBB	shape only					

Table 4.10: Summary of the systematic uncertainties on the VH acceptance in each analysis region and on the p_T^V and $m_{b\bar{b}}$ shapes originating from altering the QCD scale, including their nuisance parameter name.

964 4.3.5 SINGLE TOP PRODUCTION

965 Single top sample are generated separately for the different production channels (s , t , and Wt) us-
 966 ing PowHEG with the CT10 NLO PDF's interfaced with PYTHIA6 using the PERUGIA2012 PS
 967 tune and the corresponding CTEQ6l LO PDF's and PHOTOS (TAUOLA) for QED final state (τ)
 968 decays. Just as with $t\bar{t}$ samples, PowHEG and PYTHIA settings are varied for certain systematics. Al-
 969 ternative samples use PowHEG+PYTHIA6 with Wt diagram subtraction (DS) (instead of “diagram
 970 removal” for the ME calculation) and MADGRAPH5_aMC@NLO+HERWIG++. Systematics are
 971 derived separately in each channel, and are well described in the sumamry Table 4.11 taken from⁵⁸.

No references were given in the note, and this background really isn't that important.

Production	Uncertainty	Value	Source	Nuisance Parameter
s -channel	overall normalisation	4.6%	sum in quadrature of μ_R , μ_F , α_S and PDF uncertainties	<code>stopNorm</code>
t -channel	overall normalisation	4.4%	sum in quadrature of μ_R , μ_F , α_S and PDF uncertainties	<code>stopNorm</code>
t -channel	2 jet region acceptance	17%	sum in quadrature of deviations in alternative generators	<code>stopAcc</code> correlated with overall and 3 jet case
t -channel	3 jet region acceptance	20%	sum in quadrature of deviations in alternative generators	<code>stopAcc</code> correlated with overall and 2 jet case
Wt channel	overall normalisation	6.2%	sum in quadrature of μ_R , μ_F , α_S and PDF uncertainties	<code>stopWtNorm</code>
Wt channel	2 jet region normalisation	35%	sum in quadrature of deviations in alternative generators	<code>stopWtAcc</code> correlated with overall and 3 jet case
Wt channel	3 jet region normalisation	41%	sum in quadrature of deviations in alternative generators	<code>stopWtAcc</code> correlated with overall and 2 jet case
t -channel	p_T^V shape	shape	fit through largest deviation (POWHEG+HERWIG++) $\pm 0.001 \times p_T^V \mp 0.17 + i$	<code>StopPTV</code>
t -channel	$m_{b\bar{b}}$ shape	shape	fit through largest deviation (POWHEG+PYTHIA6 radHi-radLo) $\pm 0.0008 \times m_{b\bar{b}} \mp 0.12 + i$	<code>StopMBB</code>
Wt channel	p_T^V shape	shape	fit through largest deviation (POWHEG+PYTHIA6 with diagram subtraction) $\pm 0.003 \times p_T^V \mp 0.69 + i$	<code>StopWtPTV</code>
Wt channel	$m_{b\bar{b}}$ shape	shape	fit through largest deviation (POWHEG+PYTHIA6 with diagram subtraction) $\pm 0.0036 \times m_{b\bar{b}} \mp 0.52 + i$ ($m_{b\bar{b}} < 275$ GeV) $\mp 0.47 + i$ ($m_{b\bar{b}} \geq 275$ GeV)	<code>StopWtMBB</code>

Table 4.11: Summary of all uncertainties for the single top process with short descriptions and the name of the corresponding nuisance parameters, updated for the winter baseline analysis.

“...what would you do first?”

The Master said, “It would have to be rectifying names.”

Confucius, *The Analects*

5

972

973

Object and Event Reconstruction and

974

Selection

975 IN BREAKING WITH THE STANDARD CONVENTION both object definitions and their associated

976 experimental systematic uncertainties will be defined in this chapter: the hope is that the proximity

977 of these descriptions will illuminate each other. Summary tables are almost exclusively taken from ²⁴
978 or ³⁶. This analysis, like most typical analyses in ATLAS use central object definitions from collabora-
979 tion combined performance (CP) groups using standard analysis tools and recommendations from
980 these groups for the various objects and their accompanying systematic uncertainties.

981 Before proceeding to the objects used in this analysis, we begin with a few remarks on uncertain-
982 ties associated with object reconstruction. Event-level variables and selections will be discussed more
983 in depth in following chapters. As described in Section , systematics quantify the uncertainty asso-
984 ciated with certain effects, and are generally treated in an analysis by saving histograms of discrimi-
985 nating distributions corresponding to the nominal analysis except with the systematic in question
986 varied by plus and minus one standard deviation each (one histogram each). While for modeling
987 systematics this only corresponds to different event weight, for experimental systematics like those
988 described in this chapter (with the exception of flavor tagging and certain trigger systematics), this is
989 done by varying the parameter in question and re-running reconstruction with the systematic varied
990 before recomputing all event level quantities and then saving discriminant values in their appropri-
991 ate distributions. This is, in general, a much more computationally intensive process in the analysis,
992 which is why an entire software framework, the `CxAODFramework`, was created for this analysis (see
993 Section 3 of ²⁴ for more details).

994 5.1 TRIGGERS

995 Tables of the triggers used with the 2015 and 2016 datasets are given in Tables 5.1 and 5.2.

996 The 0-lepton channel uses a \vec{E}_T^{miss} trigger, while 2-lepton channels use single lepton triggers, with

o lep	1 lep	2 lep
HLT_xe70	HLT_xe70 HLT_e24_lhmedium_L1EM20VH OR HLT_e60_lhmedium OR HLT_e120_lhloose	HLT_mu20_iloose_L1MU15 OR HLT_mu40 HLT_e24_lhmedium_L1EM20VH OR HLT_e60_lhmedium OR HLT_e120_lhloose

Table 5.1: Summary table of triggers used in 2015 Data.

period	o lep	1 lep	2 lep
A	HLT_xe90_mht_L1XE50	HLT_xe90_mht_L1XE50 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0	HLT_mu24_ilosse(data) HLT_mu24_ilosse_L1MU15(MC) OR HLT_mu40 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0
B-D ₃	HLT_xe90_mht_L1XE50	HLT_xe90_mht_L1XE50 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0	HLT_mu24_ivarmedium OR HLT_mu50 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0
D ₄ -E ₃	HLT_xe110_mht_L1XE50	HLT_xe110_mht_L1XE50 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0	HLT_mu24_ivarmedium OR HLT_mu50 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0
$\geq F_1$	HLT_xe110_mht_L1XE50	HLT_xe110_mht_L1XE50 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0	HLT_mu26_ivarmedium OR HLT_mu50 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0

Table 5.2: Summary table of triggers used in 2016 Data.

⁹⁹⁷ the 1-lepton analysis using both. Since the 0- and 1-lepton channels are largely beyond the scope of
⁹⁹⁸ this thesis, the discussion here will be limited to the single lepton triggers; the interested reader is
⁹⁹⁹ directed towards ²⁴ and its cited sources for an in-depth discussion of the use of the \vec{E}_T^{miss} trigger.

¹⁰⁰⁰ The efficiency of triggers is in general different on simulated datasets than in actual data collected
¹⁰⁰¹ in ATLAS, so a scale factor to correct for this difference in efficiency must be applied to simulation
¹⁰⁰² events to correct for this difference. This scale factor is given by the muon CP group for muons for
¹⁰⁰³ both the 1- and 2-lepton cases and from the electron CP group for the 1-lepton case. For the two
¹⁰⁰⁴ electron case, this was calculated by the analysis team as (details in the ²⁴):

$$\frac{1 - (1 - \epsilon_{\text{MC}}^{\text{e1}} \times \text{SF}^{\text{e1}}) \times (1 - \epsilon_{\text{MC}}^{\text{e2}} \times \text{SF}^{\text{e2}})}{1 - (1 - \epsilon_{\text{MC}}^{\text{e1}}) \times (1 - \epsilon_{\text{MC}}^{\text{e2}})} \quad (5.1)$$

¹⁰⁰⁵ There are also systematic uncertainties associated with these trigger efficiencies. The single elec-
¹⁰⁰⁶ tron trigger efficiency systematic uncertainty is encapsulated in a single systematic, `EL_EFF_Trigger_-`
¹⁰⁰⁷ `Total_1NPCOR_PLUS_UNCOR`, while the single muon trigger efficiency has two components, one
¹⁰⁰⁸ each for the sample statistics, `MUON_EFF_TrigStatUncertainty`, and systematic uncertainties
¹⁰⁰⁹ `MUON_EFF_TrigSystUncertainty` associated with that efficiency's measurement.

¹⁰¹⁰ While the momentum associated with the lowest un-prescaled single lepton triggers changes
¹⁰¹¹ depending on data-taking conditions (the numbers associated with the triggers in the tables can
¹⁰¹² be thought of as nominal p_T values for trigger level objects), the lowest typical value is ~ 25 GeV.
¹⁰¹³ In order to maintain this triggering capability on low p_T muons in the higher luminosity environ-
¹⁰¹⁴ ment of the Run 3 LHC and beyond, trigger-capable detectors will be installed in upgraded New

1015 Small Wheels (NSW) of the ATLAS muon detector during the Phase I upgrade. Detailed studies in
1016 simulation of the trigger algorithm performance under nominal and misaligned conditions for the
1017 Micromegas detectors to be installed in the NSW may be found in Appendix A.

1018 **5.2 ELECTRONS**

1019 Electrons in ATLAS are reconstructed using a combination of the ATLAS electromagnetic calorime-
1020 ter (ECAL) and inner detector (ID). Reconstruction begins by searching for so-called “seed clusters”
1021 in the ECAL. The ECAL is divided into a 200×256 tower grid in the $\eta - \phi$ plane, with each tower
1022 having a size of 0.025 square in η and ϕ , corresponding to the granularity of the ECAL in its middle
1023 layer, with all energy in a tower summed longitudinally. A “sliding window” of 3×5 cells in the
1024 plane is then used to identify EM clusters associated with electrons based on criteria detailed in ⁴¹.
1025 This comparatively simple algorithm (in contrast to jet reconstruction detailed below) is effective
1026 since electromagnetic showers have a well defined behavior and shape.

1027 Once seed clusters have been formed, they are associated with tracks in the inner detector. Com-
1028 bined cluster-tracks pairs form electron candidates. In order for a electron candidate to be consid-
1029 ered a suitable electron for analysis, it must pass certain quality requirements, based on a cut on the
1030 value of a likelihood-based (LH) discriminant (cf. ⁷ for details). This discriminant is given by:

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}, \quad \mathcal{L}_S(\vec{x}) = \prod_{i=1}^n P_{s,i}(x_i) \quad (5.2)$$

1031 where the s and S (b and B) subscripts refer to distributions in fiducial signal (background) distribu-

tions in bins of $|\eta|$ and E_T . The $P(x_i)$ are probability distributions functions (pdf)'s for input variables. Several input sets of input variables exist for increasingly stringent quality requirements on electrons; this analysis uses Loose LH electrons as the base for electron selection, with the input variables relating to leakage into the hadronic calorimeter (HCAL), shower and energy deposits in each of the ECAL layers, track quality requirements, TRT hits, and track-cluster matching. This analysis adds a LooseTrackOnly isolation requirement (the p_T sum of tracks within a certain distance of the candidate track must be below a certain value), impact parameter significance cuts, and an explicit B-layer hit requirement. The ZH -signal electrons must further pass a $27 \text{ GeV} p_T$ cut ($1.05 \times p_T^{\text{trigger}}$). These requirements are summarized in Table 5.3.

Electron Selection	η	ID	d_{ϕ}^{sig}	$ \Delta z_{\phi}^{\text{BL}} \sin \theta $	Isolation	
$VH - \text{loose}$	$>7 \text{ GeV}$	$ \eta < 2.47$	LH Loose + B-layer cut	< 5	$< 0.5 \text{ mm}$	LooseTrackOnly
$ZH - \text{signal}$	$>27 \text{ GeV}$	$ \eta < 2.47$	LH Loose + B-layer cut	< 5	$< 0.5 \text{ mm}$	LooseTrackOnly
$WH - \text{signal}$	$>27 \text{ GeV}$	$ \eta < 2.47$	LH Tight	< 5	$< 0.5 \text{ mm}$	FixedCutHighPtCaloOnly

Table 5.3: Electron selection requirements.

5.2.1 ELECTRON SYSTEMATICS

The electron CP group has tabulated standard systematic uncertainties to be associated with the use of reconstructed electrons in ATLAS analyses in two main categories. The first category is related to efficiency corrections and is broken into three components: identification (`EL_EFF_ID_TotalCorrUncertainty`), reconstruction (`EL_EFF_Reco_TotalCorrUncertainty`), and isolation (`EL_EFF_Iso_TotalCorrUncertainty`). The second category deals with electron energy scale (roughly, the uncertainty in taking the energy deposits in an EM cluster and turning them into an

1048 electron energy) and energy resolution (the width associated with this). This is in practice a very
1049 complicated procedure, with over 60 systematics associated, but this analysis is not at all sensitive to
1050 these effects and so a simplified model of two systematics, EG_RESOLUTION_ALL and EG_SCALE-
1051 _ALL, is used.

1052 **5.3 MUONS**

1053 This analysis uses the standard CP muon collection in an event, though these muons in ATLAS
1054 are constructed in a variety of ways; for full details see¹⁸ and¹⁹. Most muons are constructed using
1055 tracks in the chambers of the muon spectrometer (MS), with a variety of algorithms available. MS
1056 tracks are sufficient to reconstruct a muon (a fit on these tracks can be used to point back to an in-
1057 teraction point for vertex matching, for example) and, in the $|\eta| \in (2.5, 2.7)$ interval where there
1058 is no tracking, these standalone (SA) muons are the default. The most common and robust form
1059 of muon reconstruction combines tracks in the MS with tracks in the ID (more precisely, a global
1060 refit with hits from both subsystems is typically done) to form combined (CB) muons. CB and SA
1061 muons automatically pass the loose reconstruction requirements for the Loose muons used in this
1062 analysis. Additionally, since there is a gap in the $|\eta| < 0.1$ range in the MS to make room for cabling
1063 and other detector services, there are two further muon types used in this range: the segment tagged
1064 (ST) muons that match ID tracks to segments in the MDT or CSC chambers and the calorimeter
1065 tagged (CT) muons that match ID tracks to calorimeter clusters consistent with minimum ionizing
1066 particles (which muons in ATLAS generally are).

1067 Further quality requirements are imposed on Loose muons for the different muon categories

1068 used in this analysis. Isolation requirements similar to the electrons in corresponding categories are
 1069 imposed, and impact parameter requirements are also imposed. The ZH signal muons also have a
 1070 p_T cut at 27 GeV and a requirement that the muon fall within the $|\eta|$ range of the ID.

Muon Selection	η	ID	d_o^{sig}	$ \Delta z_o^{\text{BL}} \sin \theta $	Isolation	
$VH - \text{loose}$	$> 7 \text{ GeV}$	$ \eta < 2.7$	Loose quality	< 3	$< 0.5 \text{ mm}$	LooseTrackOnly
$ZH - \text{signal}$	$> 27 \text{ GeV}$	$ \eta < 2.5$	Loose quality	< 3	$< 0.5 \text{ mm}$	LooseTrackOnly
$WH - \text{signal}$	$> 25 \text{ GeV}$	$ \eta < 2.5$	Medium quality	< 3	$< 0.5 \text{ mm}$	FixedCutHighPtTrackOnly

Table 5.4: Muon selection requirements.

1071 5.3.1 MUON SYSTEMATICS

1072 Similar to the treatment of systematic uncertainties associated with the electrons, muons have CP de-
 1073 fined systematics. The muon momentum scale and resolution systematics are divided into three cat-
 1074 egories associated one for uncertainties related to ID tracks (MUONS_ID), one for MS tracks (MUONS-
 1075 _MS), one for the overall scale (MUONS_SCALE), and two for charge dependent momentum scales
 1076 (MUON_SAGITTA_RHO and MUON_SAGITTA_RESBIAS). The remaining systematics have a STAT
 1077 and SYS component corresponding to the sample statistics and systematic uncertainties for their
 1078 individual components. For Efficiency scale factors use different standard candles in different p_T
 1079 ranges (J/ψ 's (Z 's) below (above) 15 GeV), and so these systematics are broken up into two categories
 1080 (MUON_EFF_STAT and MUON_EFF_SYS; MUON_EFF_STAT_LOWPT and MUON_EFF_SYS_LOWPT).
 1081 There are also isolation systematics (MUON_ISO_STAT, MUON_ISO_SYS) and track to vertex associa-
 1082 tion systematics (MUON_TTVA_STAT, MUON_TTVA_SYS).

1083 5.4 MISSING TRANSVERSE ENERGY

1084 High precision performance of \vec{E}_T^{miss} is not so crucial to the 2-lepton analysis (though it is very im-
1085 portant to the other channels), so the interested reader is referred to¹⁹ and its references for an in-
1086 depth discussion. \vec{E}_T^{miss} in ATLAS is the negative vectoral sum of physics objects (in this analysis just
1087 jets and leptons, though in principle also including τ 's and γ 's) and a so-called track based soft term
1088 (TST). The TST is comprised of valid ID tracks not associated with any physics objects in an event.
1089 These tracks must be associated to an event's primary vertex, have a $p_T > 0, 4 \text{ GeV}$, and pass other
1090 quality requirements.

1091 The \vec{E}_T^{miss} systematic uncertainties relevant to this analysis are related to track based energy scale
1092 and resolutions in both the soft term and in the jets and are: MET_SoftTrk_ResoPara, MET_Soft-
1093 Trk_ResoPerp, MET_SoftTrk_ScaleDown, MET_SoftTrk_ScaleUp, MET_JetTrk_Scale-
1094 Down, and MET_JetTrk_ScaleUp.

1095 5.5 JETS

1096 5.5.1 JET ALGORITHMS

1097 The hadronic nature of jets makes jet reconstruction a lot more complicated than electron or photon
1098 reconstruction, where a regular shower shape and track matching (or lack thereof in the case of the
1099 chargeless photon) provide a fairly straightforward and robust approach. The interested reader is
1100 referred to⁶¹ for an excellent survey, from which this discussion is greatly abbreviated.

1101 Looking at an event like the one in Figure 5.1, the jets are particularly easy to identify, more or less

1102 popping out of the plane plot, but this is not always the case.

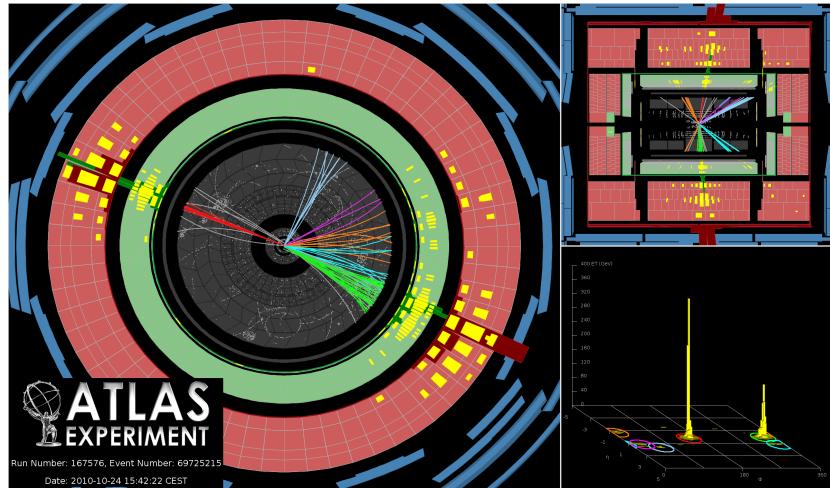


Figure 5.1: A clean ATLAS dijet event.

1103 Two general methods of turning particles/calorimeter towers into jets exist: cone-based and se-
1104 quential recombination. The general theme of the former is to find a hard (energetic) particle and
1105 draw a circle around it in the plane in an intelligent manner, while the theme of the latter is to find
1106 some metric of distance between particles and then to cluster pairs based on this distance into jets in
1107 an intelligent way. Cone-based algorithms are simple (and therefore generally quite fast) but gener-
1108 ally lack some nice properties of the sequentially recombined jets (though there are notable excep-
1109 tions like SISCone). These jets are important for trigger level objects in ATLAS, though since no jet
1110 triggers are used in this analysis, they will not be discussed any further here.

1111 The general drawback of cone-based algorithms is that they are not infrared and collinear (IRC)
1112 safe. That is, the emission of a soft (IR) quark or gluon during hadronization should not change
1113 the final jet collection in an event, and neither should the collinear splitting of hard particles during

₁₁₁₄ hadronization. These are fairly common edge cases and can lead to certain pathologies in QCD cal-
₁₁₁₅ culations. Infrared and collinear safety are diagrammed schematically in Figures 5.2 (a) and (b), taken
₁₁₁₆ from ⁶¹.

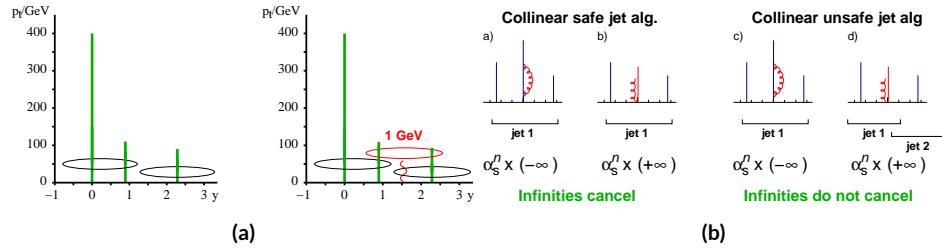


Figure 5.2: Infrared (a) and collinear (b) safety.

₁₁₁₇ The sequential recombination algorithms are generally safe from these effects, as these edge cases
₁₁₁₈ are very “close” to each other by construction. A sequential recombination algorithm proceeds as
₁₁₁₉ follows

- ₁₁₂₀ 1. Evaluate the set of distances d_{ij} (for pairs of objects) and d_{iB} (the “beam distance” for each
₁₁₂₁ individual object)

$$d_{ij} = \min \left(p_{Ti}^{2p}, p_{Tj}^{2p} \right) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{Ti}^{2p} \quad (5.3)$$

- ₁₁₂₂ 2. Find the minimum distance
₁₁₂₃ 3. If the minimum distance is:

- ₁₁₂₄ • A d_{ij} : cluster these objects together, and go to step 1
 - ₁₁₂₅ • A d_{iB} : call the i^{th} object a jet, remove it from the set of objects to be clustered, and go to
₁₁₂₆ step 1
- ₁₁₂₇ 4. Repeat until all objects are clustered into jets

1128 The choices one must make in sequential recombination are the size parameter R , akin to a cone

1129 radius in cone-based algorithms, and the momentum power p . Common choices and their trade-offs

1130 are:

- 1131 • +1: the k_t algorithm; favors the softer particles in an event, so the cluster sequence gives a
1132 history of hadronization, but jet shapes are irregular
- 1133 • 0: the Cambridge-Aachen algorithm: a pure distance metric; less substructure but jets are less
1134 “weird” looking
- 1135 • -1: the anti- k_t algorithm: clustering begins with hardest particles in an event; regular, localized
1136 jet shapes, but virtually no substructure in clustering history

1137 As shown in Figure 5.3

1138 All three algorithms have uses for different applications in ATLAS, with anti- $k_t R = 0.4$ jets

1139 being the default jet collection. These are the jets used in this analysis.

1140 If the choice of jet algorithm seems a little arbitrary, it is. There is no one-size-fits-all jet collection

1141 perfect for every application, and analyzers have to make these choices for themselves. One interest-

1142 ing choice is the jet size parameter, R . A large R jet will contain more of the radiation coming from a

1143 final state object, but its large size makes it susceptible to contamination from the underlying event

1144 and pileup, with small R jets having the opposite features. $R = 0.4$ is a fairly middle-of-the-road

1145 choice. A natural question to ask is whether there needs to be just one jet collection in an analysis.

1146 Might there not be more information to be gained from looking at more jet sizes or clusterings? Pre-

1147 liminary studies point to this answer being yes and are addressed in Appendix B.

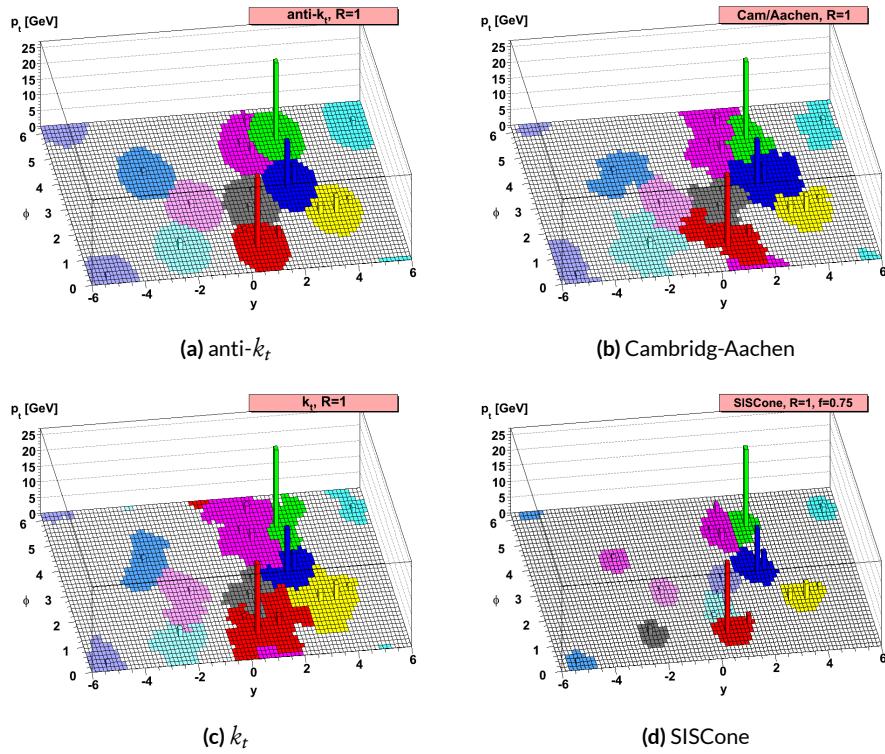


Figure 5.3: Different jet algorithms used on the same event. IC:⁶¹

1148 5.5.2 STANDARD ATLAS HBB JETS

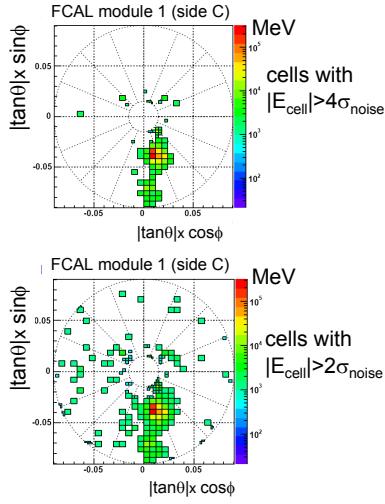
1149 There are a few considerations that arise with jets in physical detectors. The first is what type of
1150 object to use as the stand in for “particles” since ATLAS measures energy deposits, not pions. The
1151 approach ATLAS has settled upon are calorimeter topological clusters (or CaloTopoClusters for
1152 short)⁴⁹. Unlike the sliding window algorithm used for electron clusters, CaloTopoClusters use
1153 a noise significance based approach in the “4-2-1” algorithm. Each cell in the electromagnetic and
1154 hadronic calorimeters (ECAL and HCAL, respectively) has associated with it a characteristic noise
1155 level, with this noise level in each channel, it is possible to construct a “significance” for the regis-
1156 tered energy deposit in the event by dividing the measured value by its characteristic noise. Groups
1157 of cells having a significance of 4 are taken as the centers of clusters in the plane. The second layer
1158 in a cluster includes all neighboring cells to the central layer with significance of 2, and the final layer
1159 includes all the nearest neighbors to the second layer. This is described in Figure 5.4 from⁵².

1160 Once CaloTopoClusters have been formed and clustered into jets, they are calibrated using the
1161 electromagnetic (EM) scale (the scale for clusters coming form EM showers). Further details may be
1162 found in⁹.

1163 Jets in this analysis fall into two categories, “signal” and “forward” jets, and are required to pass
1164 certain quality requirements, described in Table 5.5. All jets must pass a series of jet cleaning require-
1165 ments using calorimeter level variables to eliminate jets coming from problematic calorimeter cells
1166 and certain backgrounds. Some signal jet candidates also make use of a Jet Vertex Tagger (JVT) that
1167 uses primary vertex and jet and track p_T information to decide whether certain soft jets are likely

Local Hadronic Calibration: Clusters

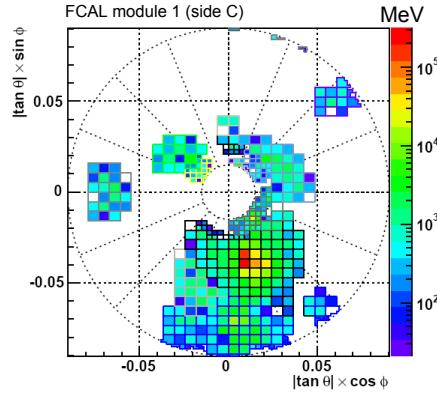
Sven Menke



□ Topological clustering

❖ 4,2,0 clusters in FCal

➢ jets with $p_T > 50 \text{ GeV}$



3rd Hadronic Calibration Workshop, Milan, Italy, 26-27 April, 2007

M. Lefebvre, P. Loch

33

Figure 5.4: A description of the 4-2-0 clustering algorithm.

- ¹¹⁶⁸ to have come from the primary (hard scatter) vertex in an event or are to be considered pileup.
¹¹⁶⁹ Further details on JVT may be found in ¹⁷. Jets are further corrected using standard CP tools and a
¹¹⁷⁰ dedicated PtReco correction, all outlined in Section 7.3 of ²⁴.

Jet Category	Selection Requirements
Forward Jets	jet cleaning $p_T > 30 \text{ GeV}$ $2.5 \leq \eta < 4.5$
Signal Jets	$p_T > 20 \text{ GeV}$ and $ \eta < 2.5$ jet cleaning $\text{JVT} \geq 0.59$ if ($p_T < 60 \text{ GeV}$ and $ \eta < 2.4$)

Table 5.5: AntiKt4EMTopoJets selection requirements. The jet cleaning is applied via the JetCleaningTool, that removes events in regions corresponding to hot calorimeter cells.

- ¹¹⁷¹ Overlap removal in this analysis is done according to the following precedence, taken from ²⁴

¹¹⁷² with further steps only taken into account if an object survives previous steps:

¹¹⁷³ • tau-electron: If $(\tau, e) < 0.2$, the τ lepton is removed.

¹¹⁷⁴ • tau-muon: If $(\tau, \mu) < 0.2$, the τ lepton is removed, with the exception that if the τ lepton has
¹¹⁷⁵ > 50 GeV and the muon is not a combined muon, then the τ lepton is not removed.

¹¹⁷⁶ • electron-muon: If a combined muon shares an ID track with an electron, the electron is re-
¹¹⁷⁷ moved.

¹¹⁷⁸ If a calo-tagged muon shares an ID track with an electron, the muon is removed.

¹¹⁷⁹ • electron-jet: If $(jet, e) < 0.2$ the jet is removed.

¹¹⁸⁰ For any surviving jets, if $(jet, e) < \min(0.4, 0.04 + 10 \text{ GeV}/e)$, the electron is removed.

¹¹⁸¹ • muon-jet If $(jet, \mu) < 0.2$ or the muon ID track is ghost associated to the jet, then the jet is re-
¹¹⁸² moved if the jet has less than three associated tracks with > 500 MeV ($\text{NumTrkPt}500\text{PV}^{\text{jet}} < 3$)
¹¹⁸³ or both of the following conditions are met: the ratio of the muon and jet is larger than 0.5 ($\mu/\text{jet} >$
¹¹⁸⁴ 0.5) and the ratio of the muon to the sum of tracks with > 500 MeV associated to the jet is larger
¹¹⁸⁵ than 0.7 (${}^{\text{muon}}/\text{SumPtTrkPt}500\text{PV}^{\text{jet}} > 0.7$).

¹¹⁸⁶ For any surviving jets, if $(jet, \mu) < \min(0.4, 0.04 + 10 \text{ GeV}/\mu)$, the muon is removed.

¹¹⁸⁷ • tau-jet: If $(\tau, jet) < 0.2$, the jet is removed.

¹¹⁸⁸ • electron-fat jet: If $(e, \text{fat jet}) < 1.2$, the fat jet is removed.

¹¹⁸⁹ Jets are corrected using a muon-in-jet correction and then a kinematic fitter (Appendix D of⁵⁶)
¹¹⁹⁰ for the 2-lepton case (PtReco correction for the 0- and 1- lepton case). The muon-in-jet correction
¹¹⁹¹ is designed for b -jets. Since the decay of a b -quark to a c -quark and finally to a light quark (these are
¹¹⁹² the multiple vertices for which JetFitter searches) involves two weak decays, there are two W -bosons

¹¹⁹³ involved in the decay. Some of these will decay semileptonically, and, while electron and τ energy
¹¹⁹⁴ will be captured by the calorimeters, semileptonic μ 's will only be registered in the MS, which occurs
¹¹⁹⁵ in some 44% of all decays from a theoretical standpoint, which amounts to about 12% in practice
¹¹⁹⁶ (due to track isolation requirements for the leptons). This value is about 1–2% for electrons, which
¹¹⁹⁷ deposit their energy in the calorimeter and so require no correction; any jet with a valid lepton associ-
¹¹⁹⁸ ated to it is deemed semileptonic (all others are called hadronic). Any jet with muons associated with
¹¹⁹⁹ it has the closest muon's 4-vector (in the plane) added to it.

¹²⁰⁰ The PtReco correction is a scale factor on the muon-in-jet corrected jet's 4-vector based on the
¹²⁰¹ jet's p_T and whether the jet is hadronic or semileptonic. This correction factor is based on particle
¹²⁰² level studies done on a TruthWZ sample. As the 0- and 1-lepton cases are not the focus of this thesis,
¹²⁰³ the interested reader is directed to Section 7.3 of²⁴.

¹²⁰⁴ The kinematic fitter used in 2-lepton events with two or three jets takes as its input 12 fit parame-
¹²⁰⁵ ters,

- ¹²⁰⁶ • energies of 2 electron or of 2 muons
- ¹²⁰⁷ • energies of 2 b -jets
- ¹²⁰⁸ • η, ϕ of 2 leptons and 2 jets
- ¹²⁰⁹ • p_X and p_Y of $\ell\ell b\bar{b}$ system.
- ¹²¹⁰ • $m_{\ell\ell}$

¹²¹¹ and 3 constraints for the variation of these parameters,

- ¹²¹² • parameters : Gaussian (b-jet energy : Transfer Functions (TF); these are denoted L , with an
¹²¹³ L_{truth} as a prior) (the ϕ parameters)

¹²¹⁴ • p_X and p_Y of $\ell\ell b\bar{b}$ system : zero with a width of 9 GeV obtained from ZH signal MC.

¹²¹⁵ • m_{ll} : Breit-Wigner (BW) distribution of Z boson (final term, leptons denoted Ω)

¹²¹⁶ which leads to test statistic from the usual likelihood formalism to be minimized in each event:

$$\begin{aligned} -2 \ln \mathcal{L} = & \sum_{i=j} \left(\frac{(\phi_i^n - \phi_i^o)^2}{\sigma_\phi^2} \right) + \left(\frac{(\Omega_l^n - \Omega_l^o)^2}{\sigma_\Omega^2} \right) - 2 \ln (L^j) - 2 \ln (L_{truth}^j) \\ & + \sum_{i=x,y} \frac{(\sum p_i^n - \sum p_i^o)^2}{\sigma_{\sum p_i}^2} + 2 \ln ((m_{\ell\ell} - M_X^2)^2 + M_X^2 \Gamma^2) \quad (5.4) \end{aligned}$$

¹²¹⁸ 5.5.3 JET SYSTEMATICS

¹²¹⁹ As with the electron systematics, jet energy scale (JES) and resolution (JER) are the two principal

¹²²⁰ considerations for systematic uncertainties, with even more standard. JER, as with the electron en-

¹²²¹ ergy resolution, is a single systematic uncertainty, `JET_JER_SINGLE_NP`. There is also a single JVT

¹²²² efficiency `JET_JvtEfficiency` systematic uncertainty. There are 88 nominal JES systematics, and

¹²²³ this analysis is sufficiently sensitive to these variations that a single systematic is grossly insufficient.

¹²²⁴ Nevertheless, some simplification is possible, with the 75 of these nuisance parameters (mostly sta-

¹²²⁵ tistical uncertainties related to the Z +jet and γ +jet calibrations) being reduced to 8, and 13 explicitly

¹²²⁶ named nuisance parameter. These remaining 13 are related to the η intercalibration used to extrap-

¹²²⁷ olate standard calibrations to other jet η regions, uncertainties related to the flavor composition of

¹²²⁸ principle background samples (W/Z +jets, top, and diboson), a single systematic for the b -jet energy

1229 scale, a high p_T jet energy scale systematic, and one for jets that punch through the HCAL to leave
1230 energy deposits in the MS. These are listed explicitly in Table 5.7.

1231 **5.5.4 FLAVOR TAGGING**

1232 Given that the final state in this analysis involves pairs of jets originating from b -quarks, deploying
1233 effective flavor tagging algorithms is imperative. While flavor tagging in general can be used to isolate
1234 any flavor (b , c , or light (u , d , s , or gluon-initiated jets)), this analysis exclusively looks for b -jets, so
1235 this discussion will focus on b 's. At truth-level in simulation this is fairly straightforward: one need
1236 only look at the particles contained within a jet and seeing if any include a b -quark (sometimes a B
1237 hadron) in the decay chain.

1238 **DIRECT TAGGING**

1239 One of the most distinctive features of b -jets is the presence of secondary vertices, as illustrated in
1240 Figure 5.5. While most partons created in particle collisions will hadronize promptly, b -quarks will
1241 first hadronize into B -hadrons, which have lifetimes of about a picosecond. This small but finite
1242 lifetime means that these particles will travel about half a millimeter or so before decaying into a jet
1243 in much the usual way, and the tracks from this decay will point back to this displaced, secondary
1244 vertex.

1245 There are various secondary vertex algorithms used as inputs to the nominal b -tagging algo-
1246 rithm⁸, with three main types of algorithms used as inputs
1247 i. Track impact parameter based algorithms: I_2 PD (signed transverse only; more pileup robust),
1248 I_3 PD (signed transverse and longitudinal)

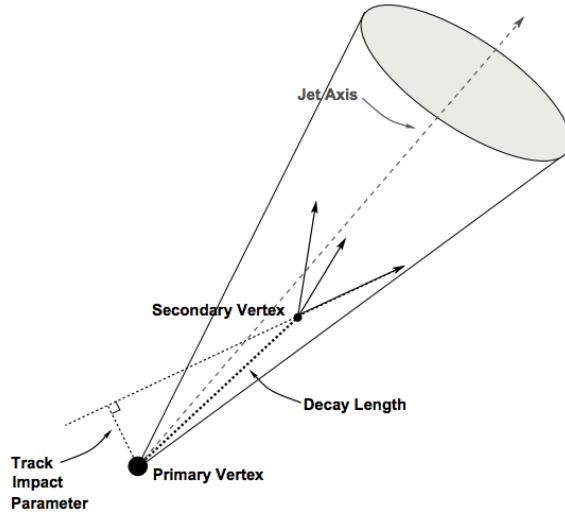


Figure 5.5: An illustration of a secondary vertex in a b -jet. Image credit:⁴⁶

- 1249 2. Inclusive secondary vertex reconstruction: SV1 (start with two track vertex pairs and con-
- 1250 struct a secondary vertex)
- 1251 3. Multiple vertex reconstruction (decay chain): JetFitter ($PV \rightarrow b \rightarrow c$ decay chain using Kalman
- 1252 filter)

1253 All of these are combined into a boosted decision tree (BDT) and trained on five million $t\bar{t}$ events
 1254 with an 90%/10% c /light jet background to form the MV_{2c10} algorithm, with 10 referring to the
 1255 percentage of charm samples in the training background. The 10% charm ratio was found to be a
 1256 good balance between increased charm rejection capability (as opposed to MV_{2c00}, which has no
 1257 charm in the background training) and loss in light jet rejection (compared to MV_{2c20}, which has

1258 20

1259 ANALYSIS SPECIFIC CONCERNS AND SYSTEMATIC UNCERTAINTIES In addition to specifying
 1260 the tagging algorithm, the working point efficiency must be specified. As with selection algorithms

1261 in general, there is a trade off between efficiency (identifying all the b -jets, minimizing type II error)
 1262 and purity (making sure all jets positively identified are in fact b -jets, minimizing type I error). Nom-
 1263 inal efficiency working points have been calibrated by the flavor tagging CP group and are outlined
 1264 in Table 5.6.

name	MV2c10 weight cut	b -tagging efficiency [%]	c RR	light RR
FixedCutBEff_60	0.9349	60.03	34.54	1538.78
FixedCutBEff_70	0.8244	69.97	12.17	381.32
FixedCutBEff_77	0.6459	76.97	6.21	134.34
FixedCutBEff_85	0.1758	84.95	3.10	33.53

Table 5.6: b -tagging working points available for MV2c10 for AntiKt4EMTopoJets. RR is the rejection rate (the inverse of efficiency).

1265 These values are aggregate figures, as both the jet's p_T and η are inputs to the MV2c10 discrim-
 1266 inant. The working point chosen for this analysis is the 70% FixedCutBEff_70 working point,
 1267 with “fixed cut” referring to the fact that this particular usage of the MV2c10 BDT value is a simple
 1268 cut value.

1269 Just as with the trigger and lepton identification efficiencies, efficiencies differ from their nom-
 1270 inal values somewhat depending on what simulation or data sample is being used. To account for
 1271 this difference, just as in the other cases, scale factors are applied to simulation event weights. It is
 1272 through these event weights, as with the modeling systematics, that the flavor tagging systematic
 1273 uncertainties are applied. Given that there are 24 input variables to MV2c10 and that flavor tagging
 1274 is in general a very difficult problem, it is not surprising that, as with the JES, there are very many
 1275 systematic uncertainties associated with flavor tagging. However, as with JES, the CP group has com-
 1276 pacted the full systematic set into a reduced set of 13 systematic uncertainties: 3 each associated with

1277 c and light jets, ς for b -jets (with the naming convention `FT_EFF_Eigen_(B|C|Light)N`), one
1278 for the extrapolation of scale factors to different jet p_T regimes (`FT_EFF_Eigen_extrapolation`),
1279 and one for the charm to bottom extrapolation (`FT_EFF_Eigen_extrapolation_from_charm`)⁶⁷.
1280 This schematic is a middle-of-the-road “Medium” set of systematics.

1281 TRUTH TAGGING

1282 Since imposing a z b -tag requirement overwhelmingly rejects events dominated by c - and light jets,
1283 statistics in such MC samples are very low. In order to circumvent this problem and restore full MC
1284 statistics, the tag rate function, or “truth-tagging” procedure (in contrast to the standard or “direct
1285 tagging” procedure) is applied, in which all events are kept but given a weight that preserves the
1286 overall shape and normalization of underlying distributions. Intuitively, this is done by giving events
1287 with real b -jets in MC a much higher weight than events having only c - or light jets. Truth-tagging is
1288 applied to all samples when conducting MVA training in order to maximize statistics and reduce the
1289 risk of overtraining. Truth-tagging is also used for data-MC comparison plots in z -tag regions and
1290 for $V + cc$, $V + c\ell$, $V + \ell$, and WW samples used in the final likelihood fit. A detailed description of
1291 the truth-tagging process is provided below.

1292 Each jet in a given event has associated with it a b -tagging efficiency, denoted ε , that is a function
1293 of its p_T , η , and real flavor (b , c , or light) from truth-level information in MC. Intuitively, this effi-
1294 ciency can be thought of as the likelihood that a given jet will be b -tagged. Hence, b -jets have a much
1295 higher b -tagging efficiency than c -jets, which in turn have a higher b -tagging efficiency than light jets.
1296 We define a truth-tag weight for a given combination of tagged and untagged jets as the product of

1297 the efficiencies of the tagged jets times the product of the complement of the efficiencies of the un-
 1298 tagged jets. For example, for an event with three jets, labeled 1, 2, and 3, if jets 1 and 2 are tagged, and
 1299 jet 3 is untagged, the truth-tag weight associated with this combination is

$$\varepsilon_1 \varepsilon_2 (1 - \varepsilon_3) \quad (5.5)$$

1300 In order to obtain a truth-tag weight for an event, one takes the sum of the weights for each pos-
 1301 sible tag combination. The current analysis requires that all events have exactly two b -tagged jets, so
 1302 the truth-tag weight is the sum of all the weights of all possible pairs of tagged jets (events with fewer
 1303 than two jets are discarded). Going back to the three jet example, one has the possible combinations:
 1304 jets 1+2 as tagged and jet 3 as untagged; jets 1+3 as tagged and jet 2 as untagged; and finally jets 2+3 as
 1305 tagged and jet 1 as untagged, which yields a total event weight of

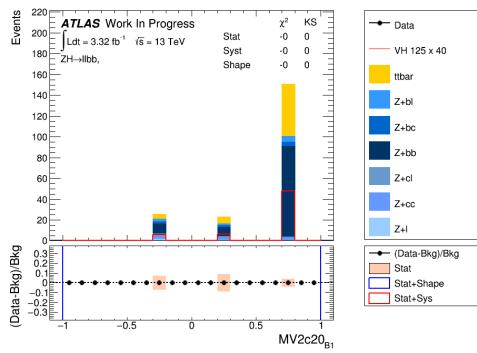
$$w_{tot} = \varepsilon_1 \varepsilon_2 (1 - \varepsilon_3) + \varepsilon_1 \varepsilon_3 (1 - \varepsilon_2) + \varepsilon_2 \varepsilon_3 (1 - \varepsilon_1) \quad (5.6)$$

1306 For some applications (e.g. in order to use variables like pTB1, the p_T of the harder b -tagged jet
 1307 in an event, in MVA training), it is necessary to choose a combination of jets in an event as “tagged.”
 1308 This combination is chosen randomly, with the probability for a given combination to be chosen
 1309 being proportional to its truth-tag weight. In the three jet example, the probability of tagging jets
 1310 1+2 is:

$$\frac{\varepsilon_1 \varepsilon_2 (1 - \varepsilon_3)}{w_{tot}} = \frac{\varepsilon_1 \varepsilon_2 (1 - \varepsilon_3)}{\varepsilon_1 \varepsilon_2 (1 - \varepsilon_3) + \varepsilon_1 \varepsilon_3 (1 - \varepsilon_2) + \varepsilon_2 \varepsilon_3 (1 - \varepsilon_1)} \quad (5.7)$$

1311 Though not used in the current analysis, functionality exists for generic truth-tagging require-
 1312 ments (i.e. an arbitrary number of tags on an arbitrary number of jets) through the logical combina-
 1313 toric extension and for so-called “pseudo-continuous tagging,” where a b -tag score is generated for
 1314 each jet in a given event. Since a random combination of jets is set by hand to pass the b -tagging cuts
 1315 regardless of its b -tag score, a new score must be generated if this information is to be used in further
 1316 analysis. Under current settings, jets that are tagged are assigned a random b -tag score that is sampled
 1317 from the MV2c10 cumulative distribution above the 70% efficiency working point cut. All other
 1318 jets in the event are assigned a random b -tag score below the 70% working point cut. Since these dis-
 1319 tributions are discrete, the scores are not truly continuous (cf. example distributions in Figure ??),
 hence the “pseudo-continuous” nomenclature.

Figure 5.6: An example of a pseudo-continuous b -tagging distribution



1320
 1321 A number of closure tests were performed on both the nominal and several systematics cases. In
 1322 the plots that follow, truth (solid) and direct (dashed) tagging distributions for m_{bb} and $\Delta R(b_1, b_2)$
 1323 in different p_T^V regimes for 2 lepton, 2 jet events. Agreement between the truth and direct tagging
 1324 cases is generally very good, an example of which can be seen in Figure ?? for a signal qqZllH125 sam-

ple, and the overall benefit of truth-tagging can be somewhat dramatically seen in the corresponding
 plots $Z + \ell$ samples in Figure ???. At high p_T^V ($p_T^V > 200$ GeV), however, in events with two real
 b -jets, there is a much greater likelihood that the b -jets will merge into a single jet, which render the
naïve assumption that jets remain discrete invalid. While this does not appear to be a problem in
most samples (cf. $\bar{t}t$ in Figure ??), there is a mismodelling effect at low m_{bb} and low $\Delta R(b_1, b_2)$ at
 $p_T^V > 200$ GeV for $W/Z + bb$ samples where truth-tagging overestimates the number of events in
this merged regime, as can be seen in Figure ???

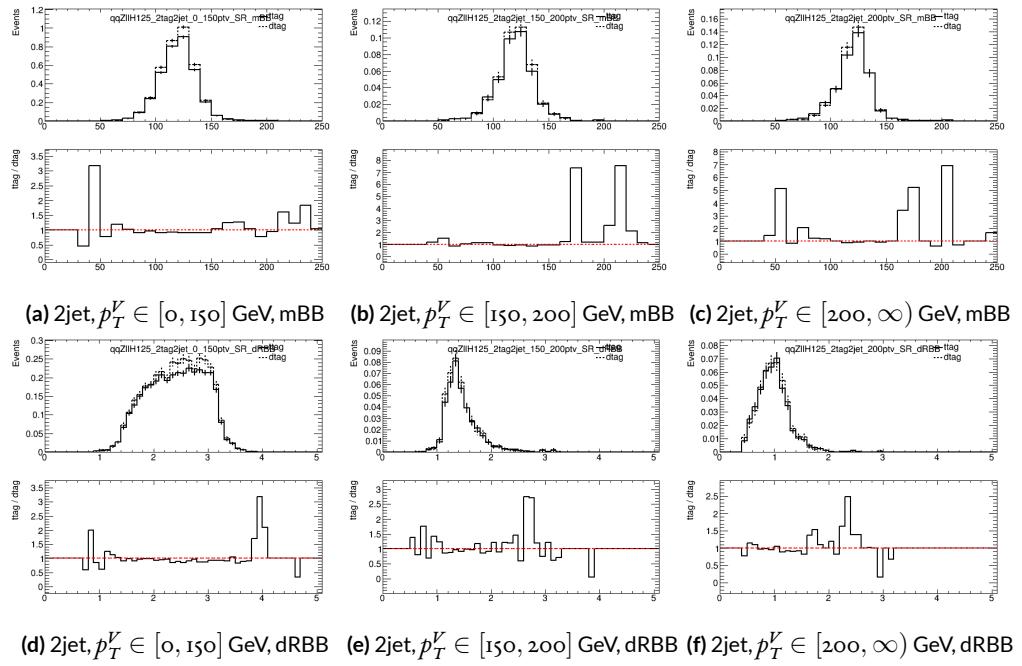


Figure 5.7: Truth-tagging closure tests for 2 lepton, 2 jet qqZIIH125 samples in three different p_T^V regions.

1331

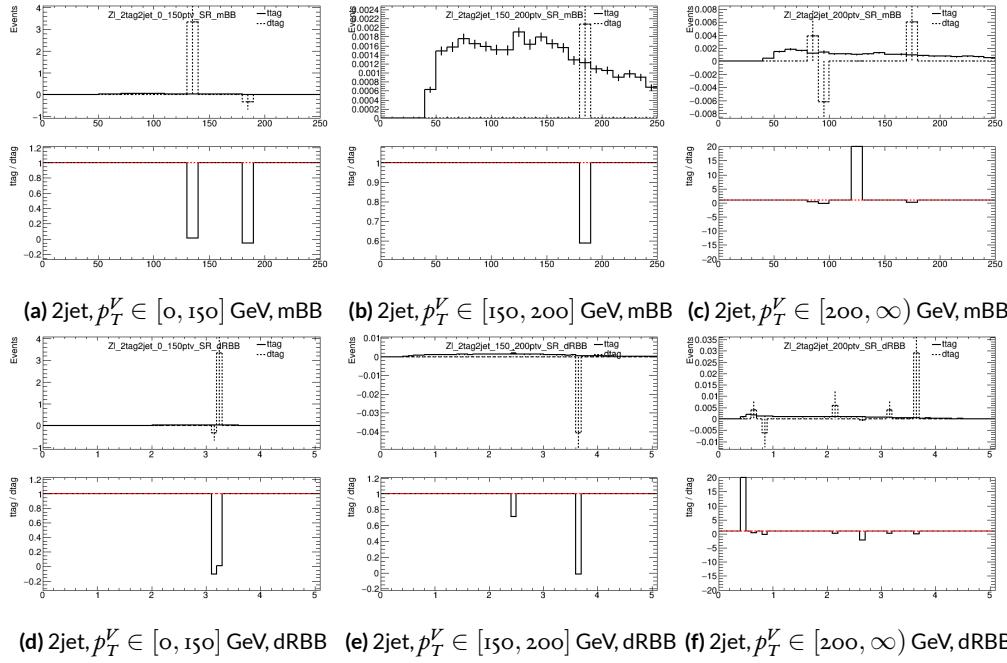


Figure 5.8: Truth-tagging closure tests for 2 lepton, 2 jet $Z + \ell$ samples in three different p_T^V regions.

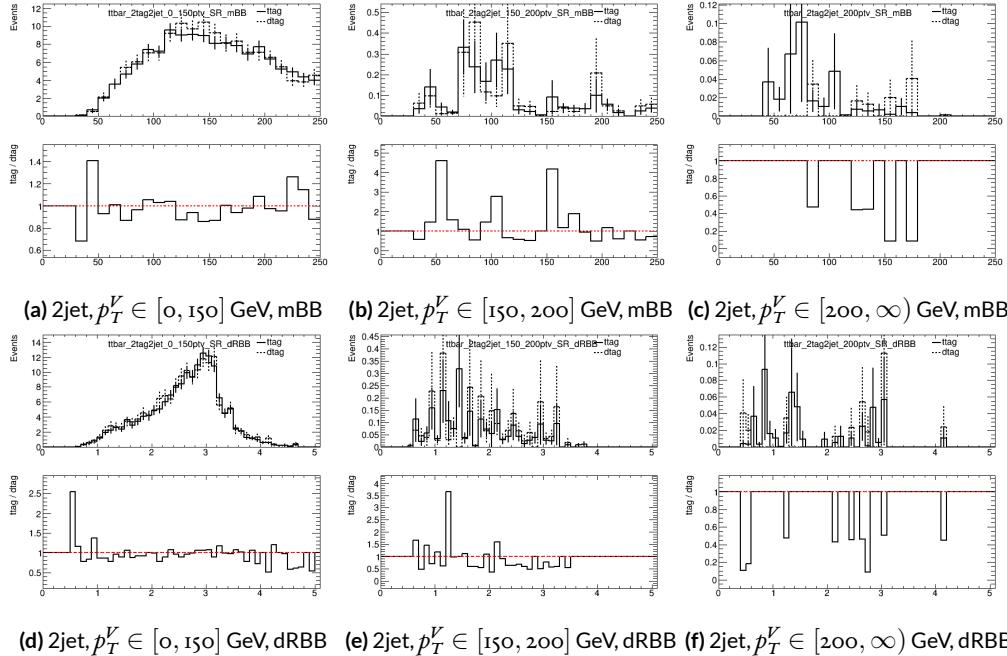


Figure 5.9: Truth-tagging closure tests for 2 lepton, 2 jet $t\bar{t}$ samples in three different p_T^V regions.

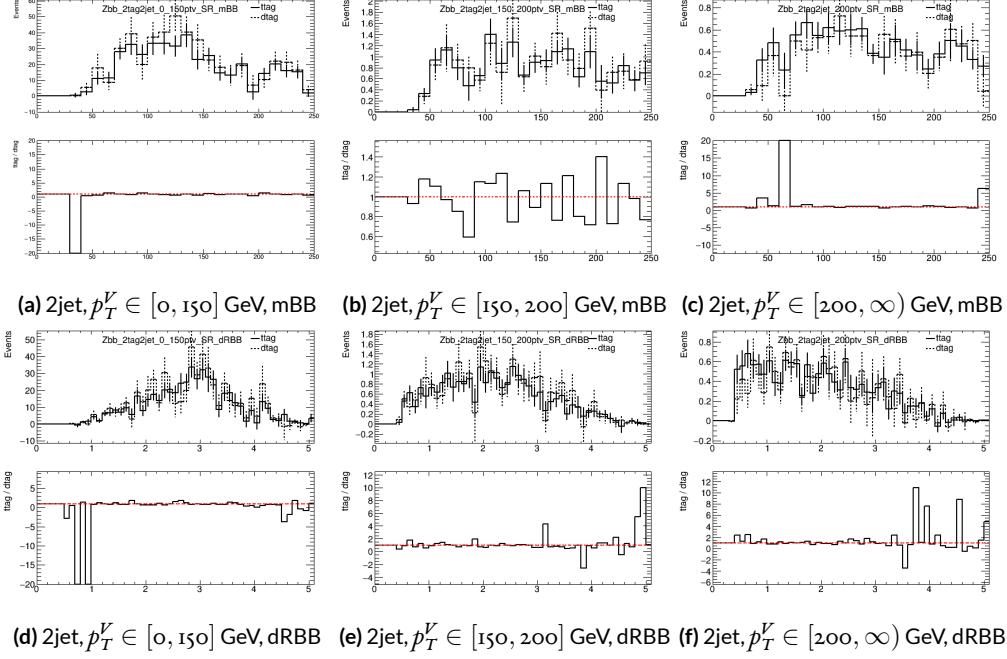


Figure 5.10: Truth-tagging closure tests for 2 lepton, 2 jet $Z + bb$ samples in three different p_T^V regions.

1332 5.6 MISCELLANIA AND SYSTEMATICS SUMMARY

1333 A summary of all experimental systematics, taken from⁵⁶, may be found below. In addition to the
1334 systematics discussed above, there are also two further systematics, on the total integrated luminosity
1335 and on the event reweighting factor used to account for pileup, both included in Table 5.7

1336 5.7 EVENT SELECTION AND ANALYSIS REGIONS

1337 With object and event reconstruction described, it is now time to address which events are actually
1338 selected for use in analysis. This analysis focuses specifically on the 2-lepton channel of the fiducial
1339 analysis, with the event selection and analysis region definitions being identical. Common to all
1340 lepton channels in the fiducial analysis is the set of requirements on the jets in a given event. There
1341 must be at least two central jets and exactly two signal jets that have been “*b*-tagged” according to the
1342 MV2c10 algorithm⁷, with at least one of these *b*-jets having $p_T > 45$ GeV. For MVA training and
1343 certain background samples, a process known as “truth-tagging” is applied instead of the standard *b*-
1344 tagging to boost sample statistics and stabilize training/fits (cf.⁵⁶ Section 4.2 for details). After event
1345 selection, the *muon-in-jet* and *PtReco* corrections, described in²⁴ 6.3.3-4, are applied to the *b*-jets.

1346 In addition to the common selections, there are 2-lepton specific selections. All events are re-
1347 quired to pass an un-prescaled single lepton trigger, a full list of which may be found in Tables 5
1348 and 6 of²⁴ with the requirement that one of the two selected leptons in the event must have fired
1349 the trigger. There must be 2 VH-loose leptons, and at least one of these must be a ZH-signal lepton
1350 (cf. Tables 5.3 and 5.4 for definitions). This lepton pair must have an invariant mass between 81 and

Systematic uncertainty	Short description	Reference
Event		
Luminosity	uncertainty on total integrated luminosity	Section 11.1 in Ref. ²⁴
Pileup Reweighting	uncertainty on pileup reweighting	Section 11.1 in Ref. ²⁴
Electrons		
EL_EFF_Trigger_Total_iNPCTOR_PLUS_UNCOR	trigger efficiency uncertainty	Section 11.2.2. in Ref. ²⁴
EL_EFF_Reco_Total_iNPCTOR_PLUS_UNCOR	reconstruction efficiency uncertainty	Section 11.3.1. in Ref. ²⁴
EL_EFF_ID_Total_iNPCTOR_PLUS_UNCOR	ID efficiency uncertainty	Section 11.3.1. in Ref. ²⁴
EL_EFF_Iso_Total_iNPCTOR_PLUS_UNCOR	isolation efficiency uncertainty	Section 11.3.1. in Ref. ²⁴
EG_SCALE_ALL	energy scale uncertainty	Section 11.3.2. in Ref. ²⁴
EG_RESOLUTION_ALL	energy resolution uncertainty	Section 11.3.2. in Ref. ²⁴
Muons		
MUON_EFF_TrigStatUncertainty	trigger efficiency uncertainty	Section 11.2.2. in Ref. ²⁴
MUON_EFF_TrigSystUncertainty	reconstruction and ID efficiency uncertainty for muons with > 15 GeV	Section 11.4.1. in Ref. ²⁴
MUON_EFF_STAT	reconstruction and ID efficiency uncertainty for muons with < 15 GeV	Section 11.4.1. in Ref. ²⁴
MUON_EFF_SYS	isolation efficiency uncertainty	Section 11.4.1. in Ref. ²⁴
MUON_EFF_STAT_LOWPT	track-to-vertex association efficiency uncertainty	Section 11.4.1. in Ref. ²⁴
MUON_EFF_SYST_LOWPT	momentum resolution uncertainty from inner detector	Section 11.4.2. in Ref. ²⁴
MUON_ISO_STAT	momentum resolution uncertainty from muon system	Section 11.4.2. in Ref. ²⁴
MUON_ISO_SYS	momentum scale uncertainty	Section 11.4.2. in Ref. ²⁴
MUON_TTVA_STAT	charge dependent momentum scale uncertainty	Section 11.4.2 in Ref. ²⁴
MUON_TTVA_SYS		
MUON_ID		
MUON_MS		
MUON_SCALE		
MUON_SAGITTA_RHO		
MUON_SAGITTA_RESBIAS		
Jets		
JET_2iNP_JET_EffectiveNP_1	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EffectiveNP_2	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EffectiveNP_3	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EffectiveNP_4	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EffectiveNP_5	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EffectiveNP_6	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EffectiveNP_7	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EffectiveNP_8restTerm	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EtaIntercalibration_Modeling	energy scale uncertainty on eta-intercalibration (modeling)	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EtaIntercalibration_TotalStat	energy scale uncertainty on eta-intercalibrations (statistics/method)	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_EtaIntercalibration_NonClosure	energy scale uncertainty on eta-intercalibrations (non-closure)	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Pileup_OffsetMu	energy scale uncertainty on pile-up (mu dependent)	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Pileup_OffsetNPV	energy scale uncertainty on pile-up (NPV dependent)	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Pileup_PtTerm	energy scale uncertainty on pile-up (pt term)	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Pileup_RhoTopology	energy scale uncertainty on pile-up (density ρ)	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Flavor_Composition_Zjets	energy scale uncertainty on $Z+jets$ sample's flavour composition	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Flavor_Composition_Wjets	energy scale uncertainty on $W+jets$ sample's flavour composition	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Flavor_Composition_top	energy scale uncertainty on top sample's flavour composition	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Flavor_Composition	energy scale uncertainty on VV and VH sample's flavour composition	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_Flavor_Response	energy scale uncertainty on samples' flavour response	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_BJES_Response	energy scale uncertainty on b-jets	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_PunchThrough_MC15	energy scale uncertainty for punch-through jets	Section 11.5.1. in Ref. ²⁴
JET_2iNP_JET_SingleParticle_HighPt	energy scale uncertainty from the behaviour of high-pT jets	Section 11.5.1. in Ref. ²⁴
JET_JER_SINGLE_NP	energy resolution uncertainty	Section 11.5.1. in Ref. ²⁴
JET_JvtEfficiency	JVT efficiency uncertainty	Section 11.5.1 in Ref. ²⁴
FT_EFF_Eigen_B	b -tagging efficiency uncertainties (“BTAG_MEDIUM”): 3 components for b jets, 3 for c jets and 5 for light jets	Section 11.7. in Ref. ²⁴
FT_EFF_Eigen_C		
FT_EFF_Eigen_L		
FT_EFF_Eigen_extrapolation	b -tagging efficiency uncertainty on the extrapolation to high- jets	Section 11.7. in Ref. ²⁴
FT_EFF_Eigen_extrapolation_from_charm	b -tagging efficiency uncertainty on tau jets	Section 11.7. in Ref. ²⁴
MET		
METTrigStat	trigger efficiency uncertainty	Section 11.2.1. in Ref. ²⁴
METTrigTop/Z	track-based soft term related longitudinal resolution uncertainty	Section 11.6. in Ref. ²⁴
MET_SoftTrk_ResoPara	track-based soft term related transverse resolution uncertainty	Section 11.6. in Ref. ²⁴
MET_SoftTrk_ResoPerp	track-based soft term related longitudinal scale uncertainty	Section 11.6. in Ref. ²⁴
MET_SoftTrk_Scale	track MET scale uncertainty due to tracks in jets	Section 11.6. in Ref. ²⁴

Table 5.7: Summary of the experimental systematic uncertainties considered. Details on the individual systematic uncertainties can be found in the given Sections of Ref.²⁴.

¹³⁵¹ 101 GeV. In addition to the jet corrections described above, a kinematic fitter is applied to the leptons
¹³⁵² and two leading corrected jets in an event with three or fewer jets^{*} to take advantage of the fact that
¹³⁵³ the 2-lepton final state is closed (cf.¹³²); these objects are only used for MVA training/fit inputs.

¹³⁵⁴ In order to increase analysis sensitivity, the analysis is split into orthogonal regions based on the
¹³⁵⁵ number of jets and the transverse momentum of the Z candidate (the vectoral sum of the lepton
¹³⁵⁶ pair; this p_T is denoted p_T^V): 2 and ≥ 3 jets; p_T^V in $[75, 150], [150, \infty)$ GeV. In addition to the signal
¹³⁵⁷ regions where the leptons are required to be the same flavor (e or μ), there are top $e - \mu$ control
¹³⁵⁸ regions used to constrain the top backgrounds.

¹³⁵⁹ All of these requirements are summarized in ??.

Category	Requirement
Trigger	un-prescaled, single lepton
Jets	≥ 2 central jets; 2 b -tagged signal jets, harder jet with $p_T > 45$ GeV
Leptons	2 VH-loose leptons (≥ 1 ZH-signal lepton); same (opp) flavor for SR (CR)
$m_{\ell\ell}$	$m_{\ell\ell} \in (81, 101)$ GeV
p_T^V regions (GeV)	$[75, 150], [150, \infty)$

Table 5.8: Event selection requirements

¹³⁶⁰ It should be noted that the use of ≥ 3 jet events is a 2-lepton specific selection. These regions are
¹³⁶¹ exclusive 3 jet regions in the 0- and 1-lepton channels, but the fiducial 2-lepton analysis was found to
¹³⁶² see a $\sim 4\%$ gain in sensitivity in studies by including ≥ 4 jet events⁵⁶.

*The gain from using the kinematic fitter is found to be smeared out in events with higher jet multiplicities.

If it's stupid but it works, it isn't stupid.

Conventional Wisdom

6

1363

1364

Multivariate Analysis Configuration

1365 IN ORDER TO fully leverage the descriptive power of the 13 TeV dataset, this analysis makes use of a
1366 multivariate (MVA) discriminant. Where traditionally event counts or single discriminating vari-
1367 ables per region of phase space have been fed to fits, MVA discriminants seek to integrate additional
1368 information not captured in the conventional phase space cuts plus dijet invariant mass distribu-

1369 tions. Formulating the MVA discriminant is an exercise in supervised learning to construct a binary
1370 classifier, where one uses labeled “signal” and “background” MC events to train optimize the param-
1371 eters of a statistical model—in this case a boosted decision tree (BDT) with some set of physically
1372 motivated variables (or “factors”). The interested reader is directed to the standard references on
1373 machine learning for further details. Sample and variable selection, including the RestFrames and
1374 Lorentz Invariants variable concepts, are discussed in Section 6.1; MVA training is treated in Sec-
1375 tion 6.2; and the data statistics only (no systematics) performance of the three MVA discriminants is
1376 explored in Section .

1377 6.1 TRAINING SAMPLES AND VARIABLE SELECTION

1378 A subset of samples described in Section 3 was used for multivariate analysis training, with $qqZH \rightarrow$
1379 $\ell\ell b\bar{b}$ and $ggZH \rightarrow \ell\ell b\bar{b}$ used as signal samples and $Z+jets$, $t\bar{t}$, and VV used as background samples.
1380 Truth-tagging is used on all samples in MVA training to improve training statistics and stability. All
1381 figures quoted in this section scale distributions to a luminosity of 36.1 fb^{-1} .

1382 6.1.1 STANDARD VARIABLES

1383 The standard set of variables taken as a baseline is the same as used in the fiducial analysis. The vari-
1384 ables fall into several main categories: energy/momenta scales of composite objects (m_{bb} , m_{bbj} ,
1385 p_T^V , $m_{\ell\ell}$), angles ($\Delta R(b_1, b_2)$, $\Delta\phi(V, H)$, $\Delta\eta(V, H)$), transverse momenta of the jets in the event
1386 ($p_T^{b_1}$, $p_T^{b_2}$, $p_T^{j_3}$), and E_T^{miss} . Input distributions for these variables in all the 2 (≥ 3 jet) analysis signal
1387 regions may be found in Figure 6.1 (6.2). The “kf” at the end of variable names denotes that these

are derived using 4-vectors that are the result of the kinematic fitter. The distributions in the figure

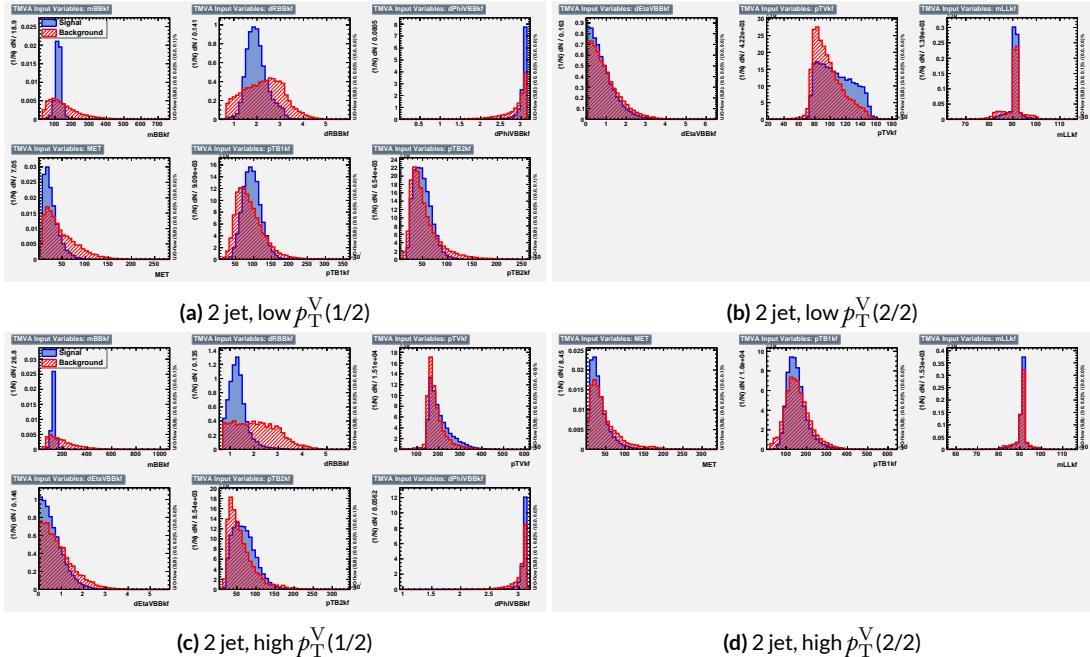


Figure 6.1: Input variables in 2 jet signal regions for the standard variable set. Signal distributions are in red, and background distributions are in blue.

1388

1389 are those used as inputs for one of the two k-folded final discriminants, and the order of the distri-
 1390 butions is the hyperparameter optimized order for feeding into the BDT; what precisely this means
 1391 will be discussed in following sections. While variables in the analysis regions are generally similar,
 1392 there are some notable exceptions. p_T^V and the correlated $\Delta R(b_1, b_2)$ have different shapes, by con-
 1393 struction for the former and by correlation for the latter, at low and high p_T^V . * The ≥ 3 jet regions
 1394 also have variables that are not applicable to the 2 jet regions; the inclusion of m_{BBJ} (the invariant
 1395 mass of the two b -jets and leading untagged jet) in particular is of note and suggests an avenue for

*Recall that higher p_T^V means, in a balanced final state like $ZH \rightarrow \ell\ell b\bar{b}$, the b -jet pair will have higher p_T and hence be more collimated (lower $\Delta R(b_1, b_2)$); this is not necessarily the case for background events, as the distributions show.

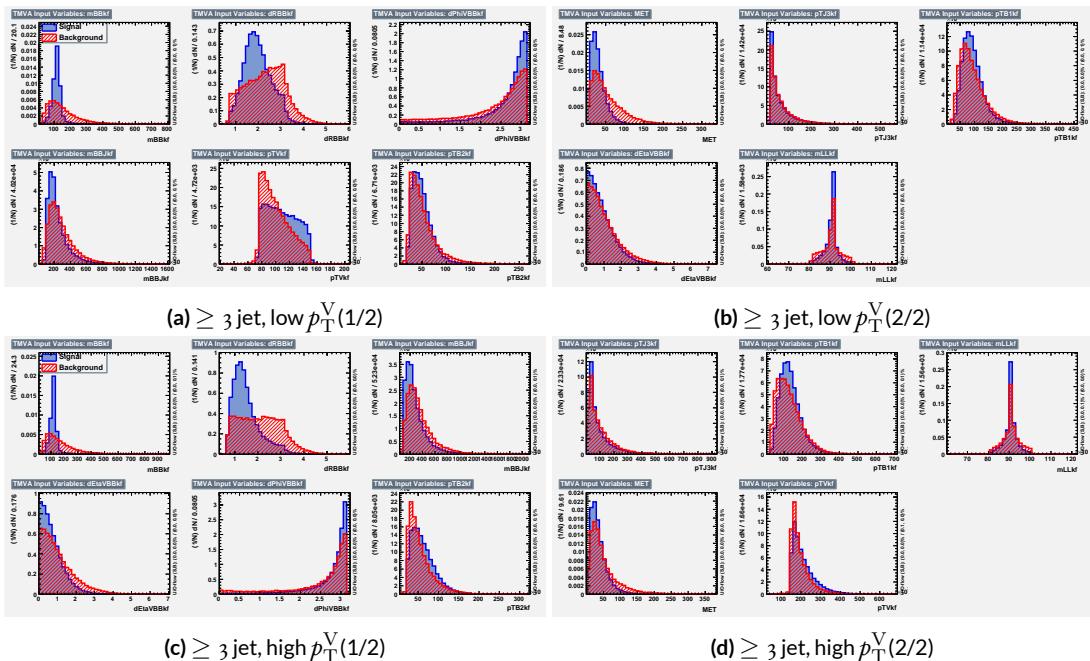


Figure 6.2: Input variables in ≥ 3 jet signal regions for the standard variable set. Signal distributions are in red, and background distributions are in blue.

refinements of the non-standard variables.

Looking at the correlation matrices for the standard variables in Figure 6.3, it is easy to see that

there are large number of non-trivial correlations

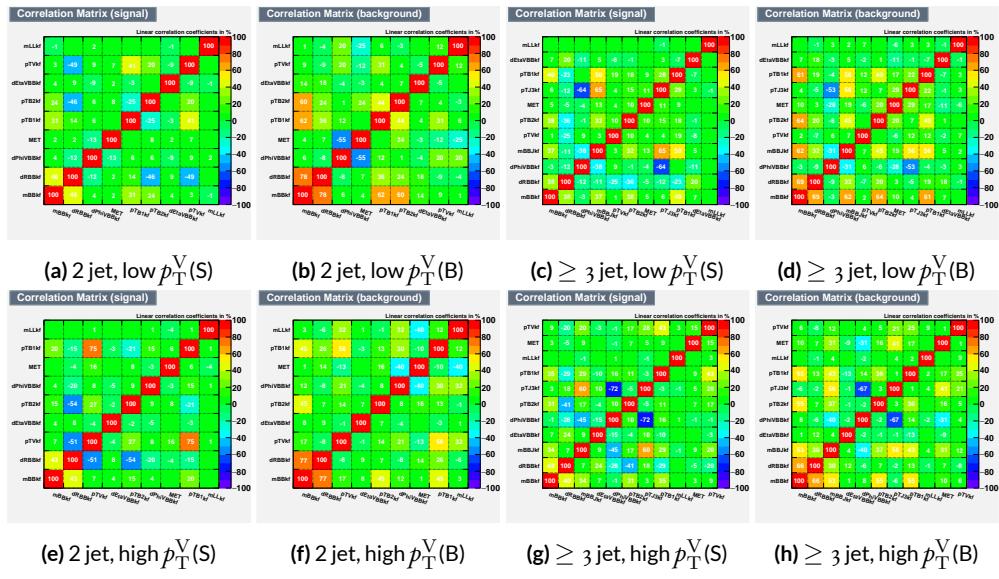


Figure 6.3: Signal and background variable correlations for the standard variable set.

1398

1399 6.1.2 LORENTZ INVARIANTS

1400 In choosing the set of variables used for a set of Lorentz Invariants based discriminants, we decided
 1401 to use S. Hagebeck's set from⁴⁵ and related studies. Distributions of these variables in the same ar-
 rangement as with the standard variables may be seen in Figures 6.4 and 6.5. One thing to note

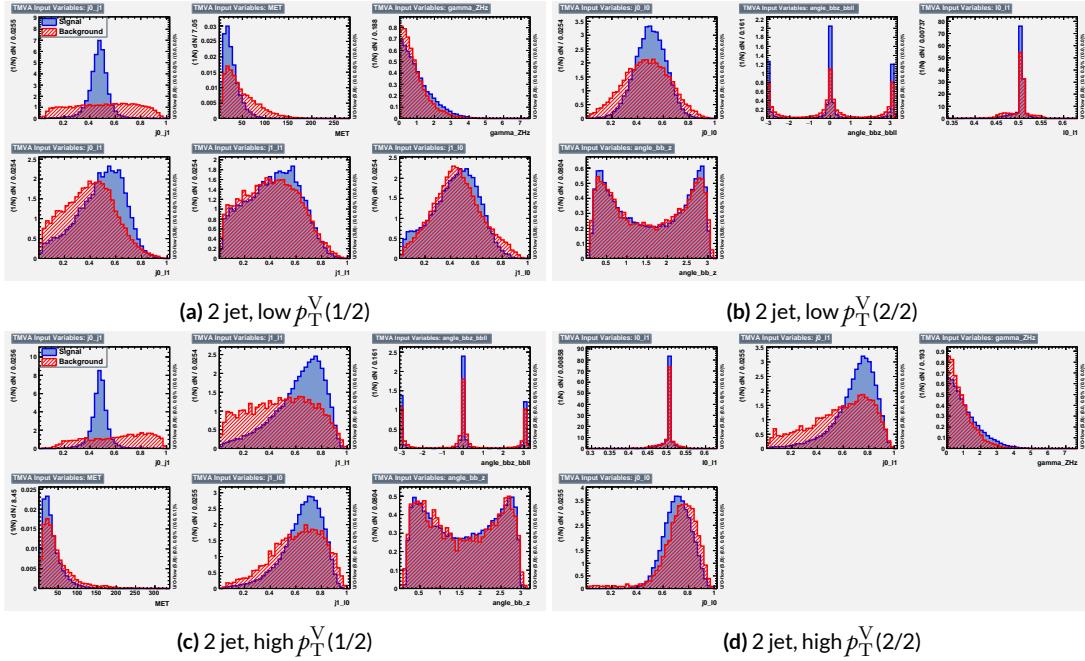


Figure 6.4: Input variables in 2 jet signal regions for the LI variable set. Signal distributions are in red, and background distributions are in blue.

1402

1403 about the variable set chosen here is that \vec{E}_T^{miss} has been added to the standard LI set. Since the LI
 1404 construction assumes that this quantity is zero, there is no obvious way to include it. Nevertheless,
 1405 as the correlation matrices for the LI variables show in Figure 6.6, there is actually very little corre-
 1406 lation between \vec{E}_T^{miss} and the other variables (with this being slightly less the case for the background
 1407 correlations, as to be expected since $t\bar{t}$, a principal background, is \vec{E}_T^{miss} -rich). Hence, if including

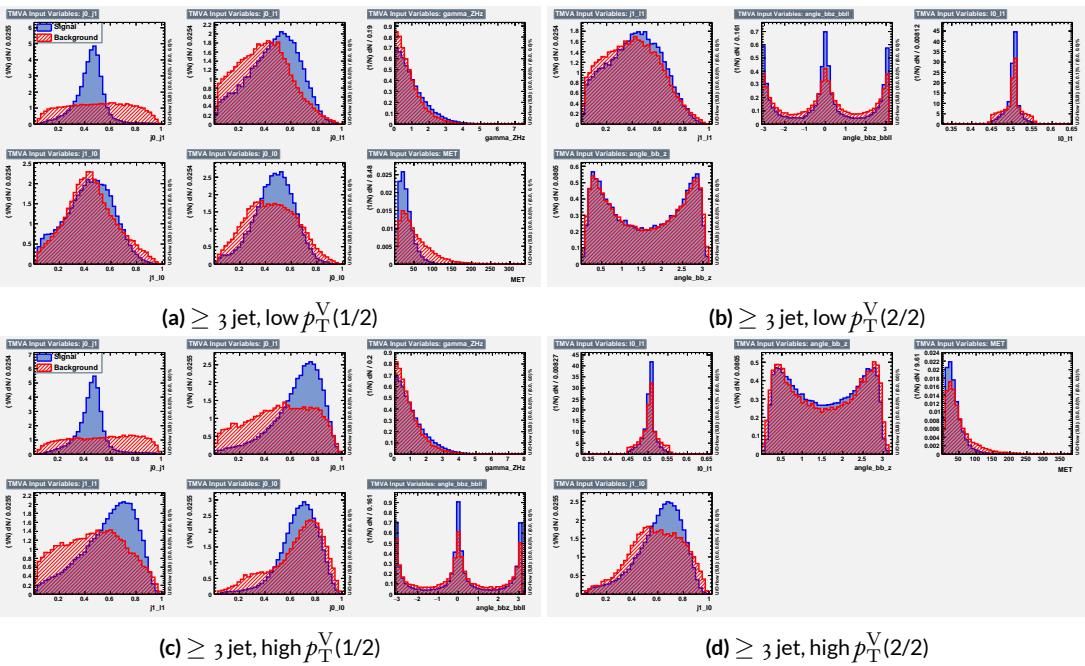


Figure 6.5: Input variables in ≥ 3 jet signal regions for the LI variable set. Signal distributions are in red, and background distributions are in blue.

1408 \vec{E}_T^{miss} violates the spirit somewhat of the LI variables, it does not break terribly much with the aim
of having a more orthogonal set.

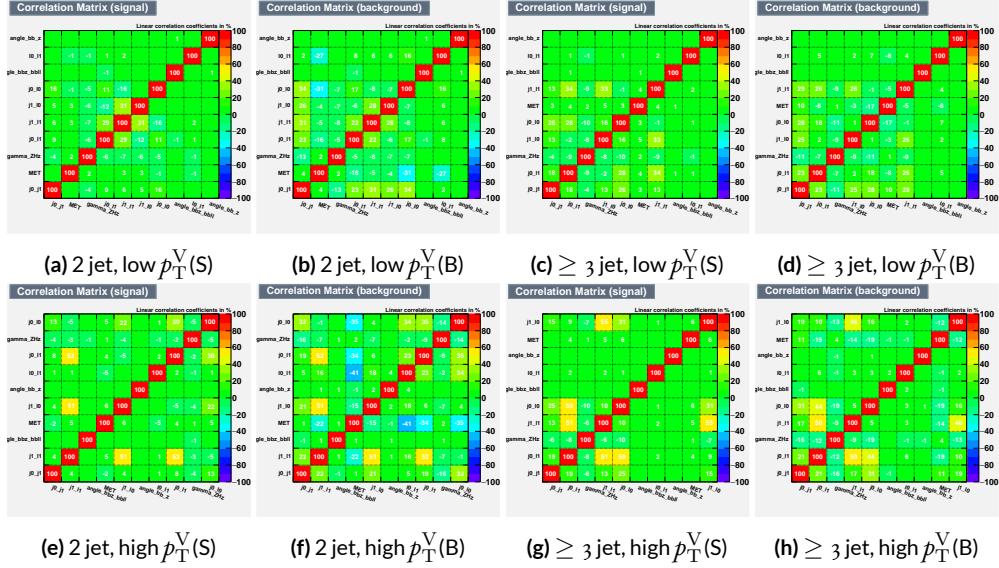


Figure 6.6: Signal and background variable correlations for the LI variable set.

1410 6.1.3 RESTFRAMES VARIABLES

1411 There is no precedent for using the RestFrames variables in the $ZH \rightarrow \ell\ell b\bar{b}$ analysis, so a subset
1412 of possible RF variables had to be selected as the basis of a discriminant. The masses and cosines of
1413 boost angles from parent frames for the CM, Z , and H frames gives six variables, and it was decided
1414 that it would be good to match the LI in terms of variable number and treatment (i.e. no special
1415 treatment of the third jet), which leaves four more variables. In addition to the cosines, there are also
1416 the $\Delta\phi$ angles. Furthermore, there are the event-by-event scaled momentum ratios, both longitudi-
1417 nal and transverse. There is also both a $\Delta\phi$ and an event ratio for the \vec{E}_T^{miss} . All of these variables
1418 were included in a ranking using slightly different training settings as the main hyperparameter
1419 optimization variable rankign described below. The goal of this study was not to develop a discir-
1420 iminant, as the number of variables is too high, but rather to see which ones are generally useful.
1421 Table 6.1 shows the results of this study. Percent gains (losses) at each step by adding the variable
1422 with biggest gain (smallest loss) are shown in green (red). The final row shows an aggregate ranking,
1423 calculated simply by adding up a variables ranks in all bins and ordering the variables smallest to
1424 greatest. This simple aggregation does not take into account which regions are potentially more sen-
1425 sitive and so where taken simply to give an idea of how variables generally performed. With this in
1426 mind, the RF variables were chosen to be the masses M_{CM} , M_H , and M_Z , the angles \cos_{CM} , \cos_H , \cos_Z ,
1427 $d\phi_{CMH}$, and the ratios R_{pt} , R_{pz} , and R_{met} . Their distributions may be seen in Figures ?? and ??.

1428 Correlations for the chosen RF variables are shown in Figure 6.9. These correlations are much
1429 lower than for the standard case but still slightly higher than for the LI case. Given the generally bet-

Region	Variable Chain
2jet pTVbin1	Rpt (6.8%), Rpz (29.0%), cosZ (11.4%), MZ (-1.75%), dphiCMH (7.26%), cosCM (3.95%), cosH (0.142%), MCM (2.18%), dphiCMZ (-2.3%), dphiCMMet (-0.236%), dphiLABCM (0.404%), Rmet (-4.04%)
3jet pTVbin1	Rpt (5.8%), Rpz (15.6%), MZ (14.8%), cosZ (5.08%), MCM (3.79%), dphiCMH (3.24%), cosH (0.755%), dphiCMMet (1.04%), Rmet (-1.03%), cosCM (5.31%), dphiCMZ (-1.27%), dphiLABCM (-2.88%), pTJ3 (-1.27%)
2jet pTVbin2	Rpt (52.0%), Rpz (13.8%), cosZ (16.9%), cosH (6.49%), MCM (1.71%), cosCM (6.21%), Rmet (4.25%), dphiCMMet (-1.53%), dphiLABCM (-0.757%), dphiCMH (0.213%), MZ (-0.788%), dphiCMZ (-2.39%)
3jet pTVbin2	Rpt (31.5%), Rpz (21.6%), cosH (8.97%), cosZ (1.42%), cosCM (1.3%), dphiCMZ (-2.84%), MCM (8.17%), dphiCMH (-0.841%), dphiLABCM (-0.00318%), dphiCMMet (-2.6%), pTJ3 (-3.21%), MZ (-1.8%), Rmet (-6.29%)
Aggregate	Rpt (o,o,o,o), Rpz (1,1,1,1), cosZ (2,3,2,3), cosH (6,6,3,2), MCM (7,4,4,6), MZ (3,2,10,11), dphiCMH (4,5,9,7), cosCM (5,9,5,4), dphiCMMet (9,7,7,9), dphiCMZ (8,10,11,5), Rmet (11,8,6,12), dphiLABCM (10,11,8,8)

Table 6.1: Full RF variable ranking study summary. Green (red) percentages represent gains (losses) in a validation significance at each step.

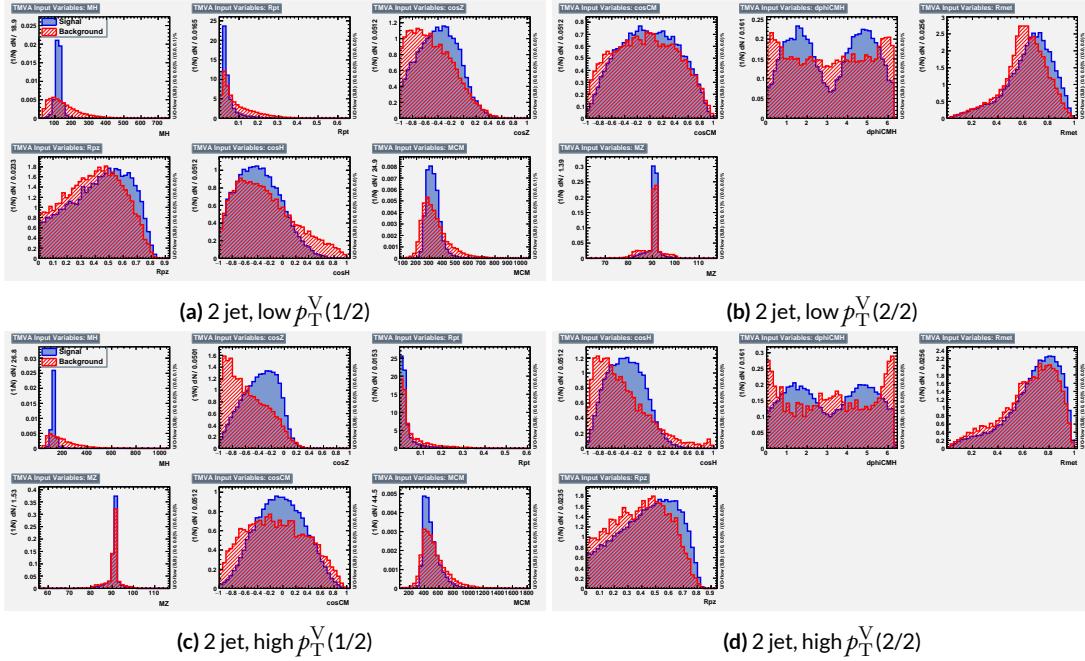


Figure 6.7: Input variables in 2 jet signal regions for the RF variable set. Signal distributions are in red, and background distributions are in blue.

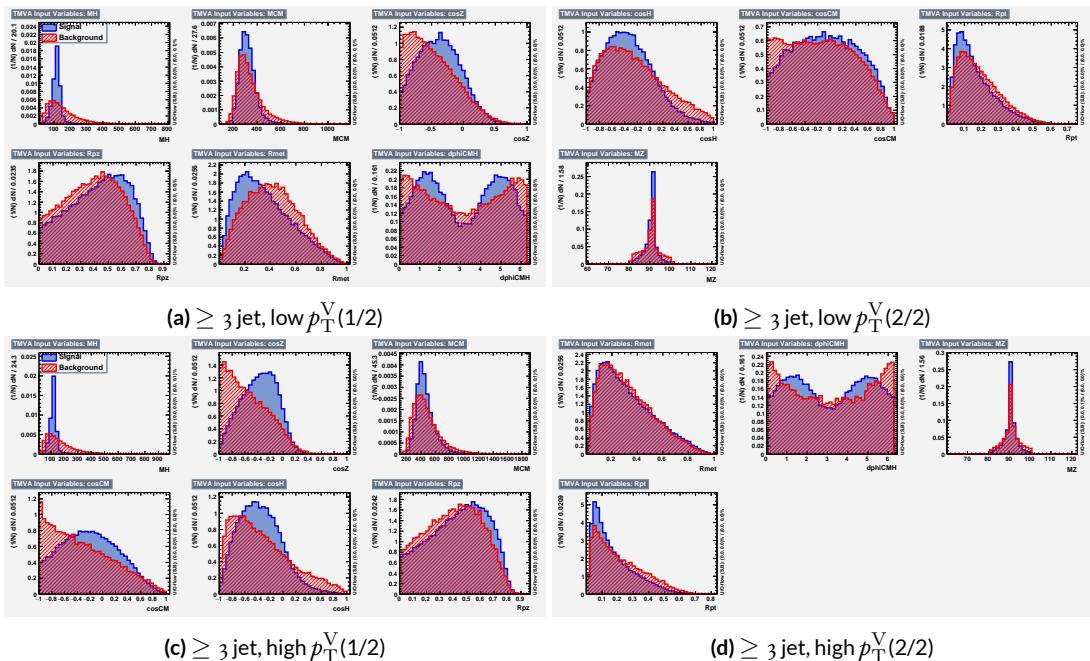


Figure 6.8: Input variables in ≥ 3 jet signal regions for the RF variable set. Signal distributions are in red, and background distributions are in blue.

1430 ter performance of the RF sets, as we shall see in following sections and chapters, this slight tradeoff
 1431 is likely an aesthetic one, with the main benefits of a more orthogonal basis likely realized at this level
 1432 of correlation.

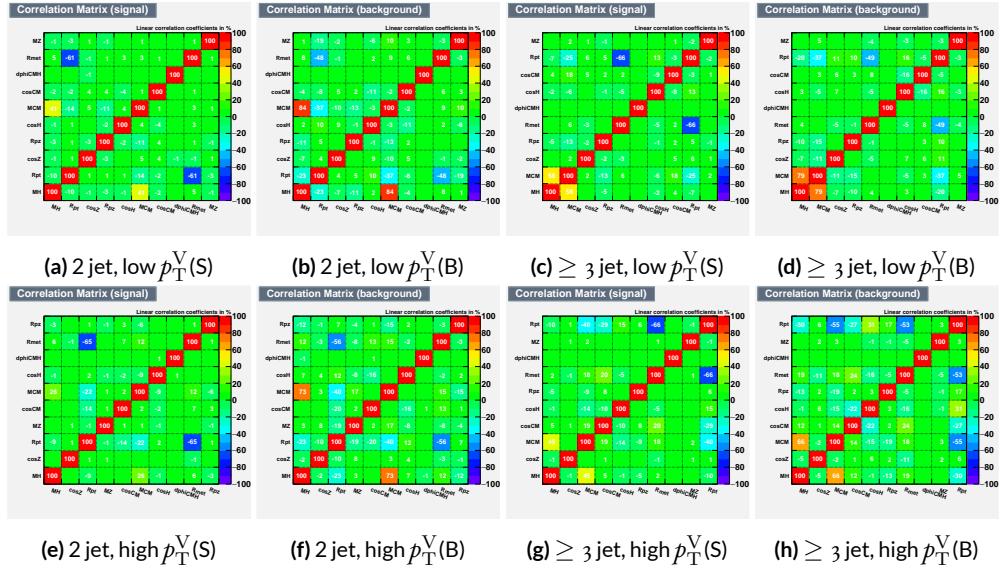


Figure 6.9: Signal and background variable correlations for the RF variable set.

1433 A summary of the variables used in the three cases is given in ??.

Variable Set	Variables
Standard	mBB, mLL, (mBBJ), pTV, pTB1, pTB2, (pTJ3), dRBB, dPhiVBB, dEtaVBB, MET 9 (11) vars
Lorentz Invariants	j0_j1, j0_l1, l0_l1, j1_l1, j0_l0, j1_l0, gamma_ZHz, angle_bbz_bbll angle_bb_z, MET 10 vars
RestFrames	MH, MCM, MZ, cosH, cosCM, cosZ, Rpz, Rpt, dPhiCMH, Rmet 10 vars

Table 6.2: Variables used in MVA training. Variables in parentheses are only used in the ≥ 3 jet regions.

1434 6.2 MVA TRAINING

1435 With variables chosen the MVA discriminants must be trained and optimized. MVA training and
1436 hyperparameter optimization (in this case, just the order in which variables are fed into the MVA) is
1437 conducted using the “holdout” method. In this scheme, events are divided into three equal portions
1438 (in this case using `EventNumber%3`), with the first third (the “training” set) being used for the initial
1439 training, the second third (the “validation” set) being used for hyperparameter optimizaiton, and
1440 the final third (the “testing” set) used to evaluate the performanace of the final discriminants in each
1441 analysis region.

1442 The MVA discriminant used is a boosted descision tree (BDT). Training is done in TMVA using
1443 the training settings of the fiducial analysis^{36†}. For the purposes of hyperparameterization and test-
1444 ing, transformation D with $z_s = z_b = 10$ is applied to the BDT distributions, and the cumulative
1445 sum of the significance $S/\sqrt{S + B}$ in each bin is calculated for each pair of distributions.

1446 Transformation D is a histogram transformation, developed during the Run 1 VHbb search, de-
1447 signed to reduce the number of bins in final BDT distributions (to reduce the effect of statistical
1448 fluctuations in data) while also maintaining sensitivity. Such an arbitrary transformation may be
1449 expressed as:

$$Z(I[k, l]) = Z(z_s, n_s(I[k, l]), N_s, z_b, n_b(I[k, l]), N_b) \quad (6.1)$$

1450 where

[†]Namely, `!H:!V:BoostType=AdaBoost:AdaBoostBeta=0.15:SeparationType=GiniIndex:-PruneMethod=NoPruning:NTrees=200:MaxDepth=4:nCuts=100:nEventsMin=5%`

- $I[k, l]$ is an interval of the histograms, containing the bins between bin k and bin l ;
- N_s is the total number of signal events in the histogram;
- N_b is the total number of background events in the histogram;
- $n_s(I[k, l])$ is the total number of signal events in the interval $I[k, l]$;
- $n_b(I[k, l])$ is the total number of background events in the interval $I[k, l]$;
- z_s and z_b are parameters used to tune the algorithm.

Transformation D uses:

$$Z = z_s \frac{n_s}{N_s} + z_b \frac{n_b}{N_b} \quad (6.2)$$

Rebinning occurs as follow:

1. Begin with the highest valued bin in the original pair of distributions. Call this the “last” bin and use it as l , and have k be this bin as well.
2. Calculate $Z(I[k, l])$
3. If $Z \leq 1$, set $k \rightarrow k - 1$ and return to step 2. If not, rebin bins $k-l$ into a single bin and name $k - 1$ the new “last” bin l .
4. Continue until all bins have been iterated through; if $Z \leq 1$ for any remaining n of the lowest-valued bins (as is often the case), simply rebin these as a single bin.

Variable ranking is done iteratively in each analysis region. In each set, the validation significance of a BDT using an initial subset of variables is calculated (dRBB and mBB for the standard set; j0_j1 for the LI set; and MH for the RF set). Each of the remaining unranked variables are then added separately, one at a time, to the BDT. The variable yielding the highest validation significance is then added to the set list of ranked variables and removed from the list of unranked variables. This process is repeated until no variables remain. These rankings are shown in Figures 6.10–6.12. Rankings tend to be fairly stable.

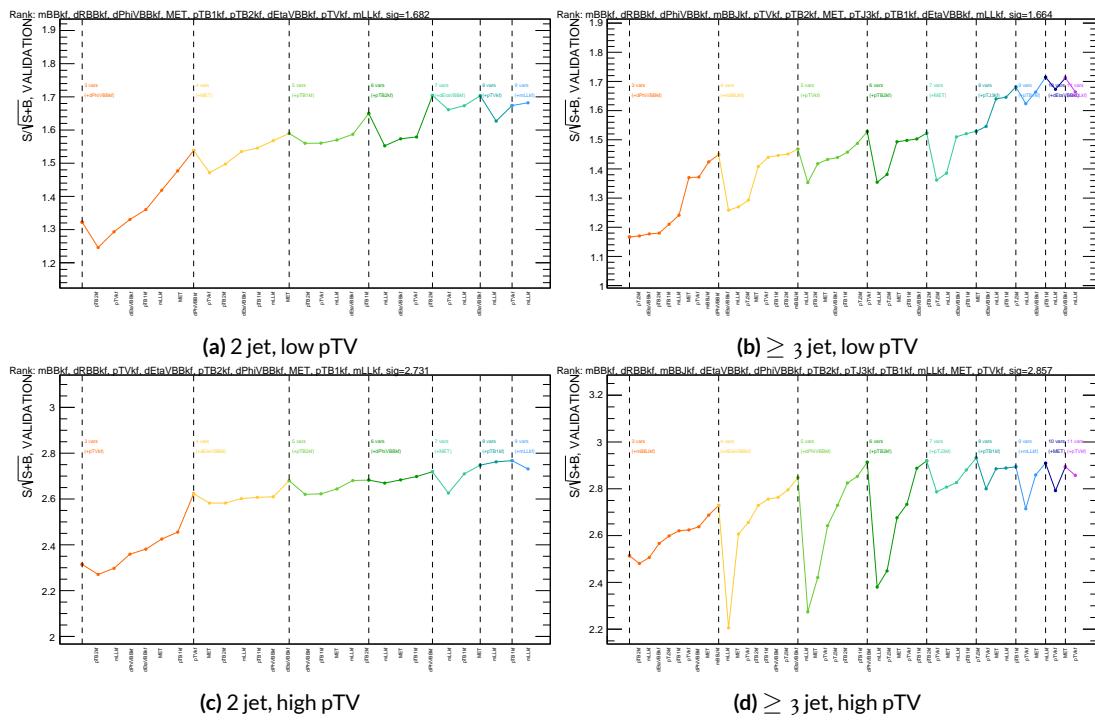


Figure 6.10: Rankings for the standard variable set.

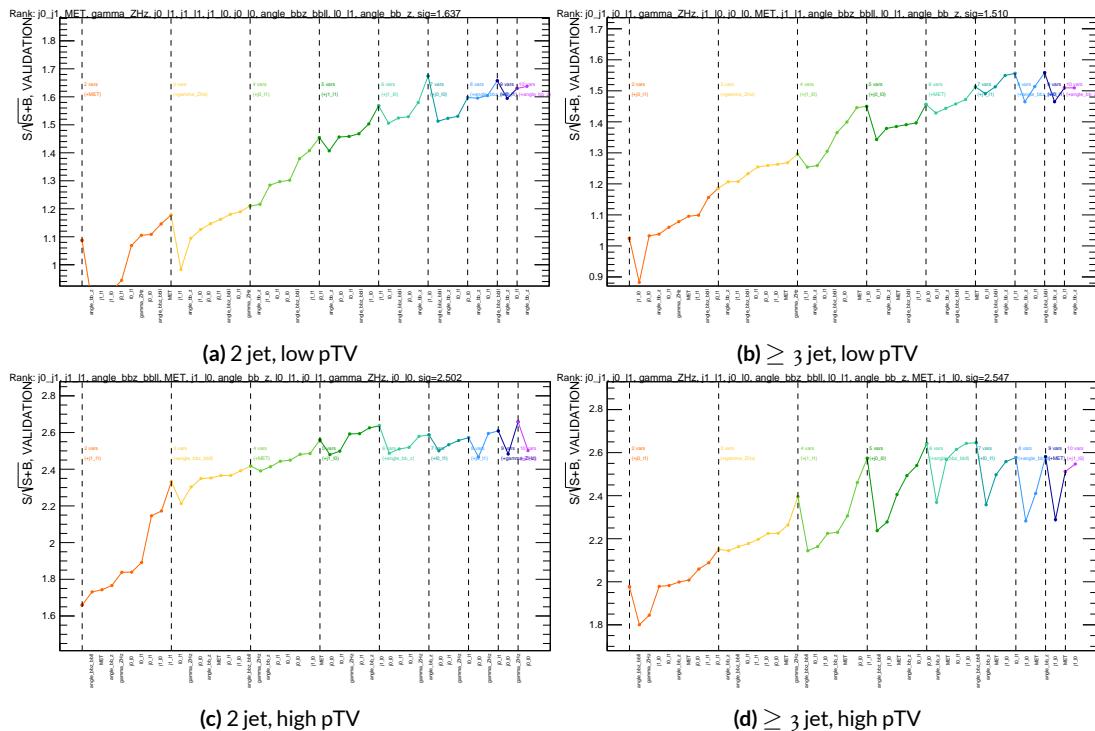


Figure 6.11: Rankings for the L1 variable set.

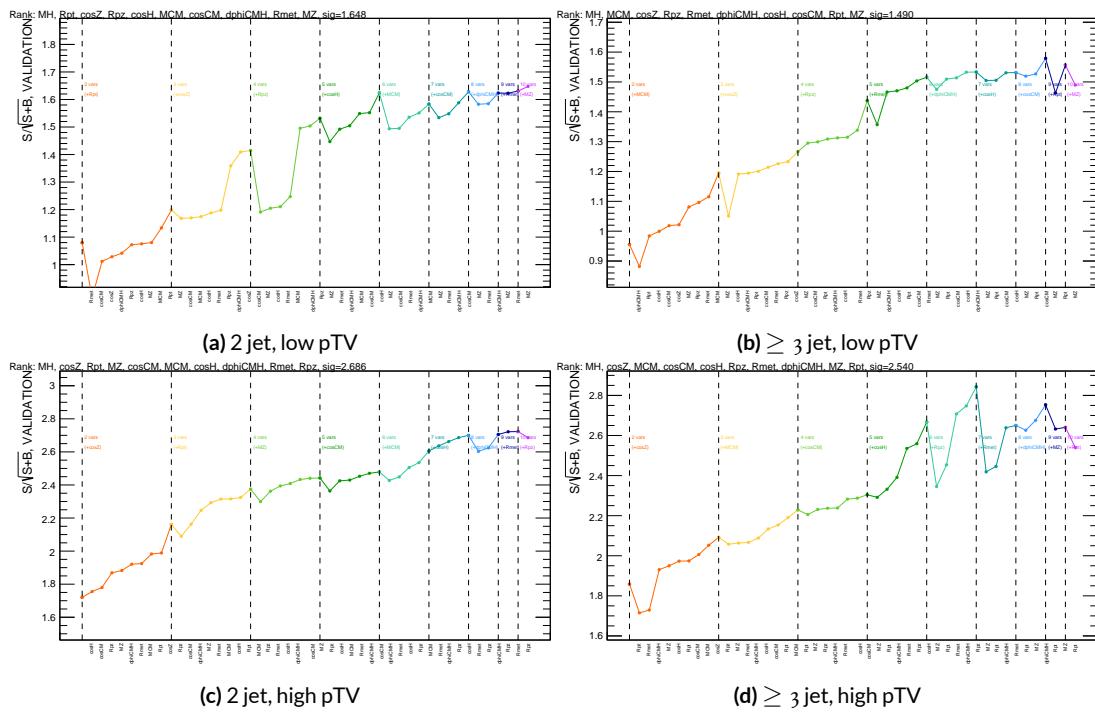


Figure 6.12: Rankings for the RF variable set.

1473 Once variables have been ranked, the BDT may be used both to evaluate performance in a simpli-
1474 fied analysis scenario in the absence of systematic uncertainties (described below in Section 6.3) and
1475 to create xml files for the production of fit inputs for an analysis including systematics. Following
1476 the approach taken in the fiducial analysis, BDT discriminants using two “k-folds” are produced to
1477 prevent overtraining, since the samples used for training are the same as those used to produce in-
1478 puts for the full profile likelihood fit. In this scheme, a BDT trained on events with an even (odd)
1479 `EventNumber` are used to evaluate events with an odd (even) `EventNumber`.

1480 6.3 STATISTICS ONLY BDT PERFORMANCE

1481 As described above, cumulative significances can be extracted from pairs of signal and background
1482 BDT output distributions in a given region. In order to evaluate performance of variable sets in
1483 the absence of systematic uncertainties, such pairs can be constructed by evaluating BDT score on
1484 the testing set of events using the optimal variable rankings in each region. We show two versions
1485 of each testing distribution for each variable set in each signal region in Figures ??–???. The training
1486 distribution is always shown as points. The plots with block histograms with numbers of bins that
1487 match (do not match) the training distribution do not (do) have transformation D applied. Trans-
1488 formation D histograms are included to show the distributions actually used for significance evalu-
1489 ation, while the untransformed histograms are included to illustrate that the level of overtraining is
1490 not too terrible. For better comparison of the distributions, all histograms have been scaled to have
1491 the same normalization.

1492 As can be seen in the summary of cumulative significances for each of these analysis regions and

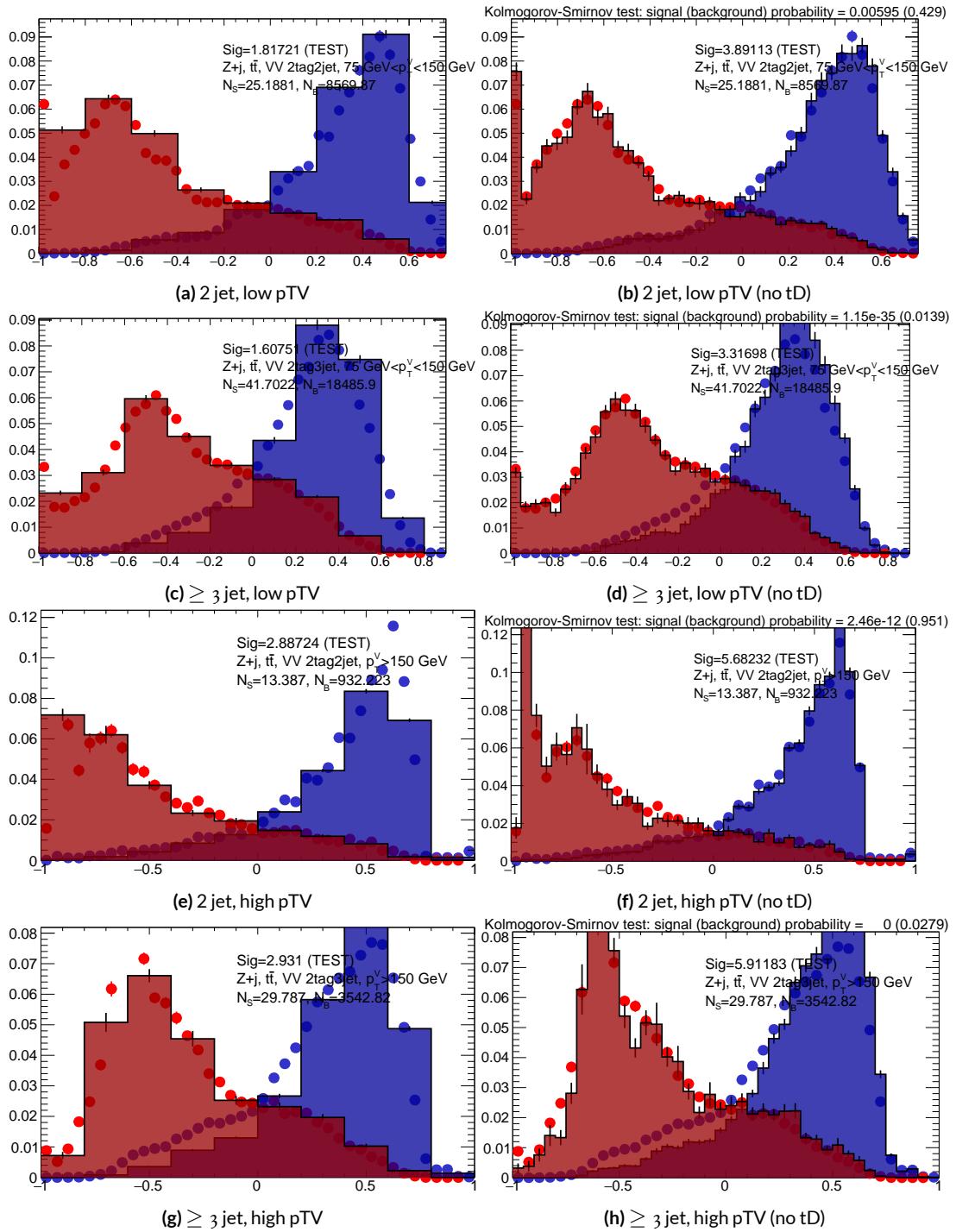


Figure 6.13: Training (points) and testing (block histogram) MVA distributions used for stat only testing for the standard variable set.

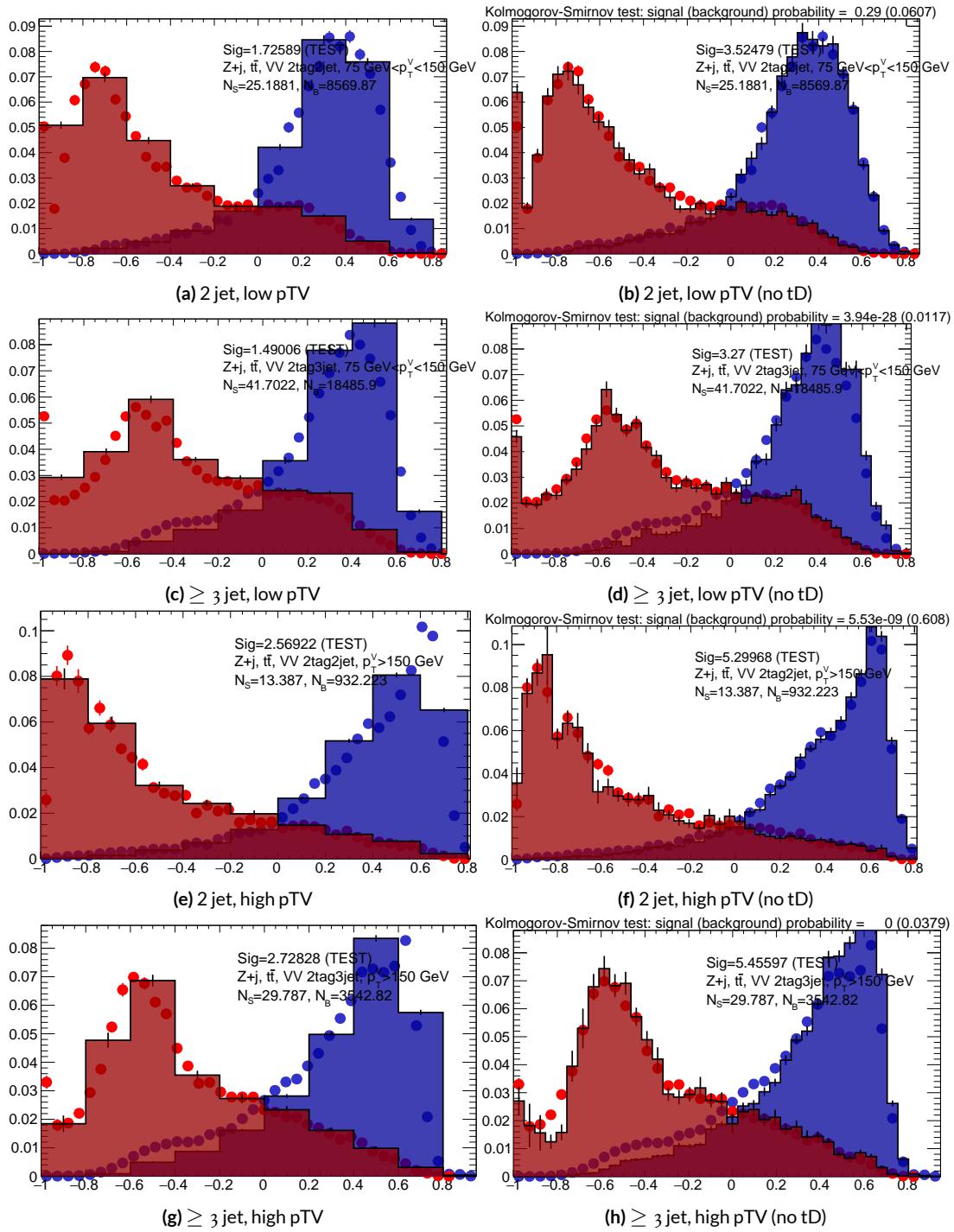


Figure 6.14: Training (points) and testing (block histogram) MVA distributions used for stat only testing for the LI variable set.

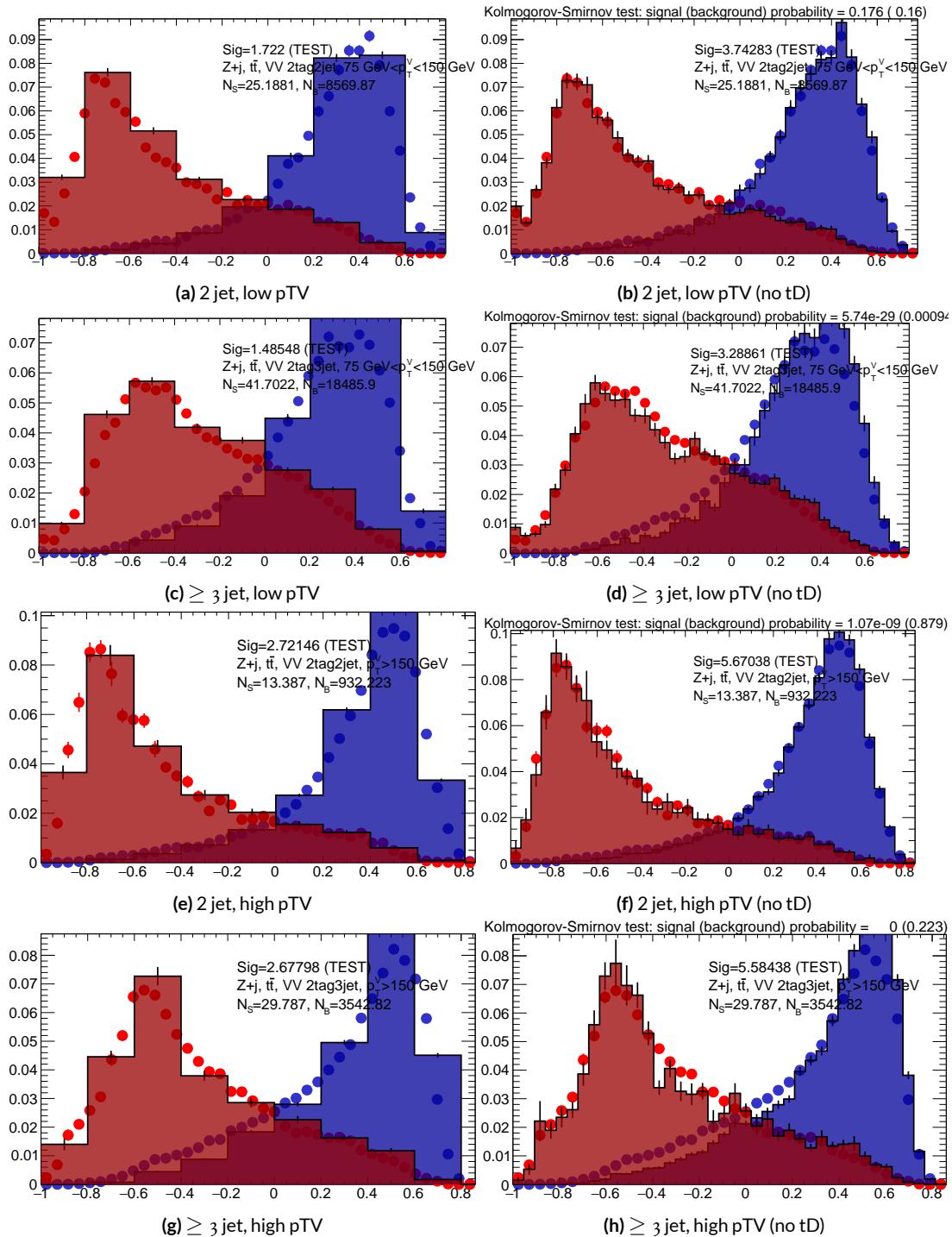


Figure 6.15: Training (points) and testing (block histogram) MVA distributions used for stat only testing for the RF variable set.

¹⁴⁹³ variable sets in Figure 6.16, the performance of each of the variable sets is quite similar. The standard
¹⁴⁹⁴ set performs best, with the LI (RF) set having a cumulative significance that is 7.9% (6.9%) lower.
¹⁴⁹⁵ This suggests that the LI and RF variables, in the $ZH \rightarrow \ell\ell b\bar{b}$ closed final state, have no more in-
¹⁴⁹⁶ trinsic descriptive power than the standard set. That these figures are all relatively high (~ 4.5) is
¹⁴⁹⁷ due largely to the absence of systematics and possibly in part due to the fact that many of the most
¹⁴⁹⁸ significant bins occur at high values of the BDT output, which, as can be seen in any of the testing
¹⁴⁹⁹ distributions, contain a small fraction of background events. An interesting feature to note in Fig-
¹⁵⁰⁰ ure 6.16 is that while the standard set does perform better in all regions, the gap is larger in the ≥ 3
¹⁵⁰¹ jet regions, suggesting that further optimization in the ≥ 3 jet case could be useful. Moreover, as
¹⁵⁰² discussed at the end of Chapter 5, the choice of ≥ 3 jet and not exclusive 3 jet regions is a 2-lepton
¹⁵⁰³ specific choice and may not be justified for the non-standard variable sets.

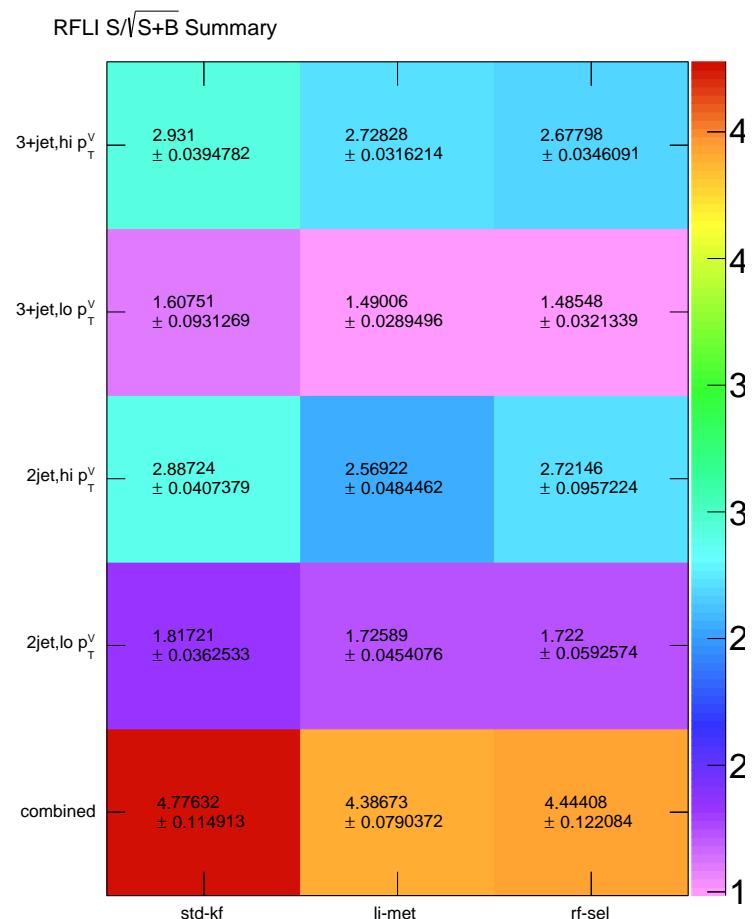


Figure 6.16: Results of testing significances sorted by analysis region and variable set.

Multivac picked you as most representative this year.

*Not the smartest, or the strongest, or the luckiest, but just
the most representative.*

Isaac Asimov, “Franchise”

1504

7

1505

Statistical Fit Model and Validation

1506 THE ULTIMATE GOAL of an analysis like the search for SM VHbb decay is to say with as much preci-
1507 sion as possible with the ATLAS collision data whether or not the SM-like Higgs observed in other
1508 decay modes also decays to b -quarks and, if so, whether this rate matches the SM prediction. In the
1509 limit of perfect modeling of both background processes and detector/reconstruction, the only free

parameter is this production rate, referred to typically as a “signal strength,” denoted μ , with $\mu = 1$ corresponding to the SM prediction.

To get a better sense of what this would look like, take a look at the example discriminant distribution in Figure 7.1. The black points are data (with statistical error bars), and the colored block histograms have size corresponding to the number of predicted events for each process in each bin of the final BDT. In the limit of perfect understanding, a fit would correspond to a constant scale factor on the red, signal histogram, where one would choose a best fit μ value, denoted $\hat{\mu}$, that would minimize the sum in quadrature of differences between the number of observed data events and $\mu s_i + b_i$, where s_i and b_i are the predicted number of signal and background events in each bin.

The only source of uncertainty would be due to data statistics, so for an infinitely large dataset with

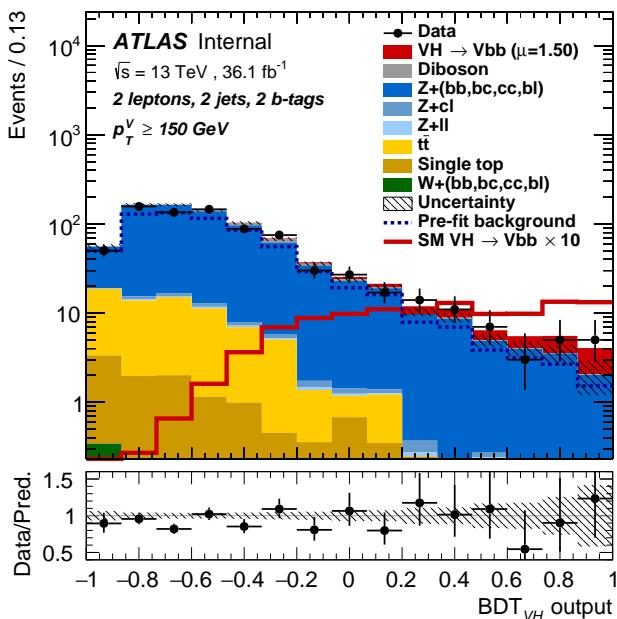


Figure 7.1: An example postfit distribution. The reason this looks different from postfit distributions later in this chapter is that this is a log plot.

1519

1520 perfect understanding, μ could be fitted to arbitrary precision. This, of course, is not the case since
 1521 there are very many sources of systematic uncertainty, discussed in previous chapters. This chapter
 1522 will first describe how systematic uncertainties are integrated into the statistical fit of this analysis
 1523 before describing two sets of cross checks on both a validation VZ fit and on the fit for the VH fit of
 1524 interest.

1525 7.1 THE FIT MODEL

1526 In order to derive the strength of the signal process $ZH \rightarrow \ell\ell b\bar{b}$ and other quantities of interest
 1527 while taking into account systematic uncertainties or nuisance parameters (NP's, collectively de-
 1528 noted θ), a binned likelihood function is constructed as the product over bins of Poisson distribu-
 1529 tions:

$$\mathcal{L}(\mu, \theta) = \text{Pois}(n | \mu S + B) \left[\prod_{i \in \text{bins}} \frac{\mu s_i + b_i}{\mu S + B} \right] \prod_{j \in \text{NP's}} \mathcal{N}_{\theta_j}(\theta_j, \sigma_j^2 | o, i) \quad (7.1)$$

1530 where n is the total number of events observed, s_i and b_i are the number of expected signal and back-
 1531 ground events in each bin, and S and B are the total expected signal and background events. The
 1532 signal and background expectations generally are functions of the NP's θ . NP's related to the nor-
 1533 malization of signal and background processes fall into two categories. The first set is left to float
 1534 freely like μ while the second set are parametrized as log-normally distributed to prevent negative
 1535 predicted values. All other NP's are parametrized with Gaussian priors. This results in a "penalty"
 1536 on the NLL discussed below of $(\hat{\alpha} - \mu_\alpha)^2 / \sigma_\alpha^2$, for NP α , normally parametrized with mean μ_α
 1537 and variance σ_α^2 for an MLE of $\hat{\alpha}$.

1538 One can maximize^{*} the likelihood in Equation 7.1 for a fixed value of μ to derive estimators for
 1539 the NP's θ ; values of θ so derived are denoted $\hat{\theta}_\mu$ to emphasize that these are likelihood maximizing
 1540 for a given θ . The profile likelihood technique finds the likelihood function's maximum by com-
 1541 paring the values of the likelihood over all possible values of μ using these "profiles" and picking
 1542 the one with the greatest $\mathcal{L}(\mu, \hat{\theta}_\mu)$ value; these values of μ and θ are denoted $\hat{\mu}$ and $\hat{\theta}$. The profile
 1543 likelihood can further be used to construct a test statistic[†]

$$q_\mu = -2 \left(\log \mathcal{L}(\mu, \hat{\theta}_\mu) - \log \mathcal{L}(\hat{\mu}, \hat{\theta}) \right) \quad (7.2)$$

1544 This statistic can be used to derive the usual significance (p value), by setting $\mu = 0$ to find the
 1545 compatibility with the background-only hypothesis³⁷. If there is insufficient evidence for the signal
 1546 hypothesis, the CL_s method can be used to set limits¹¹.

1547 In order to both validate the fit model and study the behavior of fits independent of a given
 1548 dataset, a so-called "Asimov" dataset can be constructed for a given fit model; this dataset has each
 1549 bin equal to its expectation value for assumed values of the NP's and a given μ value (in this case,
 1550 $\mu = 0$, the SM prediction).

^{*}Maximization is mathematically identical to finding the minimum of the negative logarithm of the likelihood, which is numerically an easier problem. This is what is done.

[†]The factor of -2 is added so that this statistic gives, in the asymptotic limit of large N , a χ^2 distribution.

1551 7.2 FIT INPUTS

1552 Inputs to the binned likelihood are distributions of the BDT outputs described in Chapter 6 for
1553 the signal regions and of m_{bb} for the top $e - \mu$ control regions. These regions split events accord-
1554 ing to their p_T^V and number of jets. All events are required to have two b -tags, as well as pass the
1555 other event selection requirements summarized in Table ??; the only difference between the signal
1556 and control region selections is that the same flavor requirement (i.e. leptons both be electrons or
1557 muons) is flipped so that events in the control region have exactly one electron and one muon. The
1558 BDT outputs are binned using transformation D, while the m_{bb} distributions have 50 GeV bins,
1559 with the exception of the 2 jet, high p_T^V region, where a single bin is used due to low statistics.

1560 Input distributions in MC are further divided according to their physics process. The signal pro-
1561 cesses are divided based on both the identity of associated V and the number of leptons in the final
1562 state; $ZH \rightarrow \ell\ell b\bar{b}$ events are further separated into distributions for qq and gg initiated processes.
1563 $V+jets$ events are split according to V identity and into the jet flavor bins described in Chapter 3.
1564 Due to the effectiveness of the 2 b -tag requirement suppressing the presence of both c and l jets,
1565 truth-tagging is used to boost MC statistics in the cc , cl , and ll distributions.[‡] For top backgrounds,
1566 single top production is split according to production mode (s , t , and Wt), with $t\bar{t}$ as single category.
1567 Diboson background distributions are also split according to the identity of the V 's (ZZ , WZ , and
1568 WW). Fit input segmentation is summarized in Table 7.1.

[‡]Since WW is not an important contribution to the already small total diboson background, no truth-tagging was applied here, in contrast to the fiducial analysis.

Category	Bins
# of Jets	2, 3+
p_T^V Regions (GeV)	$[75, 150], [150, \infty)$
Sample	data, signal $[(W, qqZ, ggZ) \times n_{lep}]$, $V+\text{jet} [(W, Z) \times (bb, bc, bl, cc, cl, l)]$, $t\bar{t}$, diboson (ZZ, WW, WZ), single top (s, t, Wt)

Table 7.1: Fit input segmentation.

¹⁵⁶⁹ 7.3 SYSTEMATIC UNCERTAINTIES REVIEW

¹⁵⁷⁰ Tables 7.2 and 7.3 summarize modeling (Chapter 4) and experimental (Chapter 5) systematic uncer-
¹⁵⁷¹ tainties considered in this analysis, respectively. In addition to these, simulation statistics uncertain-
¹⁵⁷² ties (“MC stat errors”) are also included in the fit model. There is one distribution per systematic
¹⁵⁷³ (one each for up and down) per sample per region. The $\pm 1\sigma$ variation for a systematic is calculated
¹⁵⁷⁴ as the difference in the integrals between the nominal and up/down varied distributions.

Process	Systematics
Signal	$H \rightarrow bb$ decay, QCD scale, PDF+ α_S scale, UE+PS (acc., $p_T^V, m_{bb}, 3/2$ jet ratio)
$Z+\text{jets}$	Acc, flavor composition, $p_T^V+m_{bb}$ shape
$t\bar{t}$	Acc, $p_T^V+m_{bb}$ shape
Single top	Acc., $p_T^V+m_{bb}$ shape
Diboson	Overall acc., UE+PS (acc, $p_T^V, m_{bb}, 3/2$ jet ratio), QCD scale (acc (2, 3 jet, jet veto), p_T^V, m_{bb})

Table 7.2: Summary of modeling systematic uncertainties.

¹⁵⁷⁵ The systematics distributions undergo processes known as “smoothing” and “pruning” before
¹⁵⁷⁶ being combined into the final likelihood used in minimization.
¹⁵⁷⁷ The difference between systematics varied distributions and nominal distributions approaches

Process	Systematics
Jets	21 NP scheme for JES, JER as single NP
E_T^{miss}	trigger efficiency, track-based soft terms, scale uncertainty due to jet tracks
Flavor Tagging	Eigen parameter scheme (CDI File: 2016-20_7-13TeV-MC15-CDI-2017-06-07_v2)
Electrons	trigger eff, reco/ID eff, isolation eff, energy scale/resoltuion
Muons	trigger eff, reco/ID eff, isolation eff, track to vertex association, momentum resolution/scale
Event	total luminosity, pileup reweighting

Table 7.3: Summary of experimental systematic uncertainties.

1578 some stable value in the limit of large simulation statistics, but if the fluctuations due to simulation
 1579 statistics in a distribution are large compared to the actual physical effect (whether this is because
 1580 the actual effect is small or if the actual distribution is derived from a small number of simulation
 1581 events), then systematic uncertainty will be overestimated by, in effect, counting the MC stat error
 1582 multiple times. Smoothing is designed to mitigate these effects by merging adjacent bins in some
 1583 input distributions. Smoothing happens in two steps (the full details of smoothing algorithms may
 1584 be found in ⁵⁶ and in the `WSMaker` code):
 1585 Merge bins iteratively where bin differences are smallest in input distributions until no local extrema
 1586 remain (obviously, a single peak or valley is allowed to remain)
 1587 Sequentially merge bins (highest to lowest, like transformation D) until the statistical uncertainty in
 1588 a given bin is smaller than 5% of merged bin content
 1589 Not all systematic uncertainties defined are included in the final fit. Pruning Systematics are
 1590 subject “pruning” (individually in each region/sample: there are two histograms per systematic

₁₅₉₁ (up/down) per region per sample, so pruning just consists of removing the histograms from the
₁₅₉₂ set of distributions included in the likelihood) if they are do not have a significant impact, defined as
₁₅₉₃ follows.

₁₅₉₄ • Normalization/acceptance systematics are pruned away if either:

₁₅₉₅ – The variation is less than 0.5%

₁₅₉₆ – Both up and down variations have the same sign

₁₅₉₇ • Shape systematics pruned away if either:

₁₅₉₈ – Not one single bin has a deviation over 0.5% after the overall normalisation is removed

₁₅₉₉ – If only the up or the down variation is non-zero

₁₆₀₀ • Shape+Normalisation systematics are pruned away if the associated sample is less than 2% of
₁₆₀₁ the total background and either:

₁₆₀₂ – If the predicted signal is $< 2\%$ of the total background in all bins and the shape and
₁₆₀₃ normalisation error are each $< 0.5\%$ of the total background

₁₆₀₄ – If instead at least one bin has a signal contribution $> 2\%$ of the total background, and
₁₆₀₅ only in each of these bins, the shape and normalisation error are each $< 2\%$ of the
₁₆₀₆ signal yield

1607 7.4 THE VZ VALIDATION FIT

1608 One of the primary validation cross-checks for the fiducial analysis was a VZ fit—that is, conducting
1609 the entire analysis but looking for $Z \rightarrow b\bar{b}$ decays instead of the Higgs. The idea here is that the Z is
1610 very well understood and so “rediscovering” the Z is taken as a benchmark of analysis reliability since
1611 the complexity of the fit model precludes the use of orthogonal control regions for validation as is
1612 done in other analyses (generally, if there is a good control region, one prefers to use it to constrain
1613 backgrounds and improve the fit model). To do this, a new MVA discriminant is made by keeping
1614 all hyperparameter configurations the same (e.g. variable ranking) but using diboson samples as
1615 signal. For the 2-lepton case, this means using $ZZ \rightarrow \ell\ell b\bar{b}$ as the signal sample. This new MVA is
1616 used to make the inputs described in Section 7.2, and the fit is then run as for the VH fit (again, with
1617 ZZ as signal).

1618 The VZ fit sensitivities for the standard, LI, and RF fits are summarized in Table 7.4. The ex-
1619 pected significances are all fairly comparable and about what was the case in the fiducial analysis.
1620 The observed significance for the standard set matches fairly well with the expected value on data,
1621 but the LI and RF observed significances are quite a bit lower.

	Standard	LI	RF
Expected (Asimov)	3.83	3.67	3.72
Expected (data)	3.00	2.95	3.11
Observed (data)	3.17	1.80	2.09

Table 7.4: Expected (for both data and Asimov) and observed $VZ \rightarrow \ell\ell b\bar{b}$ sensitivities for the standard, LI, and RF variable sets.

1622 These values, however, are consistent with the observed signal strength values, which can be seen
 1623 in Figure 7.2 (b), with both the LI and RF fits showing a deficit of signal events with respect to the
 1624 SM expectation, though not by much more than one standard deviation (a possible explanation is
 1625 explored in the following section). Just as in the VH fits, errors arising systematic uncertainties are
 1626 lower in the fits to the observed dataset. That the effect is not noticeable in Asimov fits is not too
 1627 surprising, since this analysis (and these variable configurations in particular), is not optimized for
 1628 VZ .

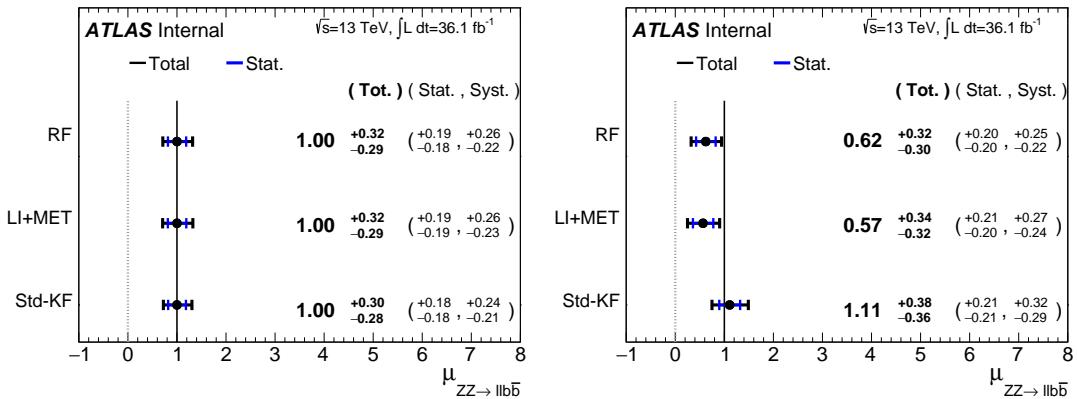


Figure 7.2: μ summary plots for the standard, LI, and RF variable sets. The Asimov case (with $\mu = 1$ by construction) is in (a), and $\hat{\mu}$ best fit values and error summary are in (b).

1629 7.4.I 2 AND ≥ 3 JET FITS

1630 While the treatment of simply ignoring any additional jets in the event seems adequate for the VH
 1631 analysis (discussed below), the potential shortcoming of this treatment appears in the VZ analysis
 1632 when the 2 and ≥ 3 jet cases are fit separately[§], as can be seen in Figure 7.3. Compared to the stan-
 1633 dard fit, the LI and RF fits have lower $\hat{\mu}_{\geq 3 \text{ jet}}$ values, consistent with the interpretation that the addi-

[§]standalone fits, with half the regions each, not 2 POI fits

1634 tional information in the ≥ 3 jet regions for the standard case is important for characterizing events
1635 in these regions for VZ fits.

1636 A natural question to ask is why this would be an issue for the VZ but not the VH case. One
1637 potential answer is that at high transverse boosts, there is a greater probability for final state
1638 radiation in the hadronically decaying Z , so there are more events where the third jet should be in-
1639 cluded in the calculation of variables like $m_{b\bar{b}}$ or for angles involving the $b\bar{b}$ system (e.g. \cosh in the
1640 RF case). While the absolute scale at which the low and high p_T^V regions are separated remains the
1641 same does not change from the VH to the VZ analysis, 150 GeV, the implicit cutoff on the transverse
1642 boost of the hadronically decaying boson does. For the Higgs, with a mass of 125 GeV, the p_T^V cutoff
1643 corresponds to $\gamma \sim 1.56 - 6.74$, but for the Z , with a mass of 91 GeV, this is $\gamma \sim 1.93 - 9.21$, about
1644 23–37% higher.

1645 If either the LI or RF schemes were to be used in a mainstream analysis, these validation fits sug-
1646 gest that the third jet ought to be included in variable schemes (e.g. by adding the third jet to the
1647 Higgs in the high p_T^V case). On the issue of whether or not ≥ 4 jet events should be included, the
1648 RF set shows very little sensitivity to this change (a 2 jet and 3 jet only fit moves $\hat{\mu}$ to 0.64, while
1649 doing so for the LI set moves it to 0.40, so this, like the addition of the third jet into the variable
1650 sets, would have to be addressed individually. Nevertheless, this optimization is beyond the scope of
1651 this thesis, which aim to preserve as much of the fiducial analysis as possible for as straightforward a
1652 comparison as possible.

1653 For completeness, we include the full set of fit validation results for the VZ fit, explaining them in
1654 turn.

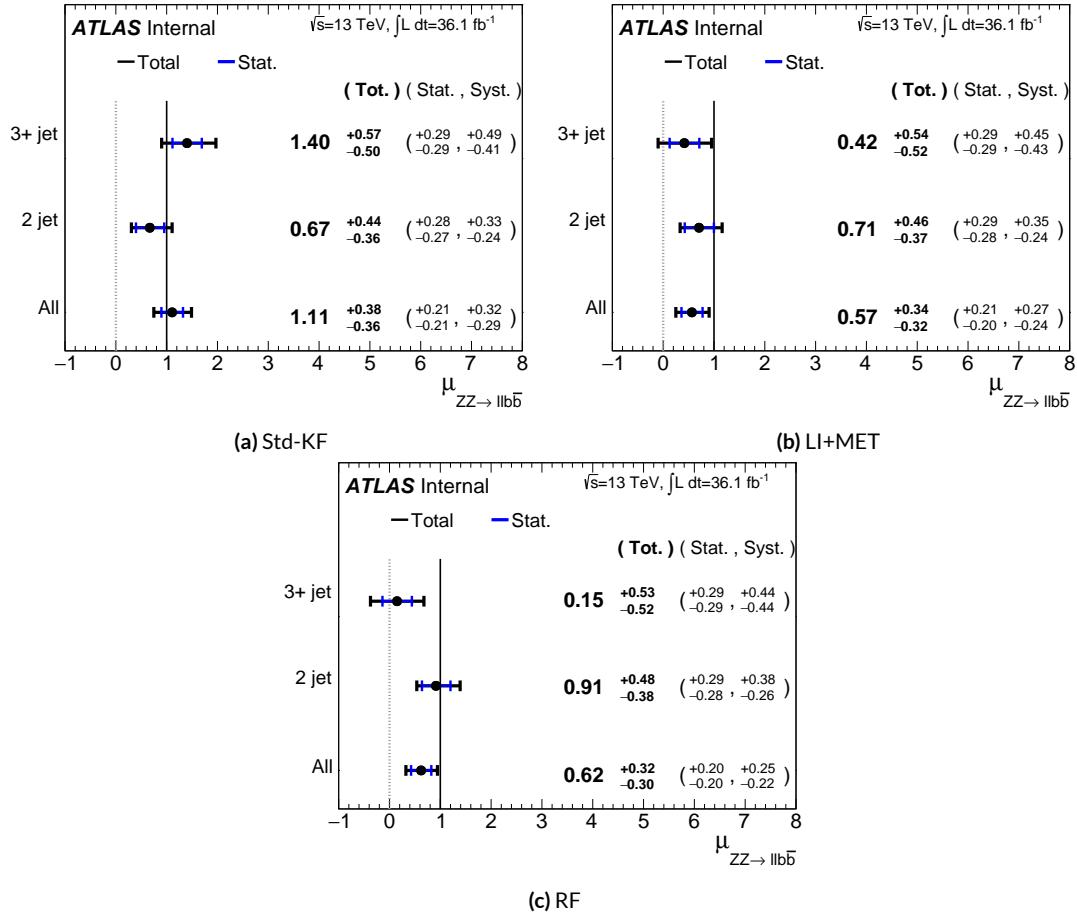


Figure 7.3: $\hat{\mu}$ summary plots with standalone fits for the different n_{jet} regions for the standard, LI, and RF variable sets.

1655 7.5 NUISANCE PARAMETER PULLS

1656 The first set of plots statistical fit experts will want to look at are the “pulls” and “pull comparisons.”

1657 In these plots, the best fit (nominal) values and one standard deviation error bars are shown for ob-

1658 served (Asimov) pull plots, with the green and yellow bands corresponding to $\pm 1, 2\sigma$, respectively.

1659 These plots are divided by NP category for readability. ¶ In pull comparisons, these pulls are over-

1660 layed and color-coded. Pull comparisons here have the following color code: black is the standard

1661 variable set, red is the LI set, and blue is the RF set.

1662 A well-behaved fit has pulls close to nominal values (“closeness” should be interpreted in the

1663 context of pull value divided by pull error). As can be seen in Figures 7.4–7.8, the fits for the three

1664 different variable sets are fairly similar from a NP pull perspective, though the $Z+jets$ m_{bb} and p_T^V

1665 NP’s and the jet energy resolution NP are heavily pulled (a handful of poorly behaved pulls is not

1666 uncommon, though typically warrants further investigation). As a general note, these pull plots

1667 calculate pulls using a simultaneous HESSE matrix inversion, which is fine for relatively small fits,

1668 but the more reliable MINOS result, which calculates the impact of each NP on its own, should be

1669 cross-checked for significant pulls. The ranking plots below do this.

1670 Nuisance parameter correlation matrices (for correlations with magnitude at least 0.25) for all

1671 three variable set fits can be found in Figures 7.10–7.12. These are useful for seeing which NP’s move

1672 together (if there is no physical argument for them to do so, this is a potential indicator that further

1673 investigation is warranted).

¶Over 100 non-MC stat NP’s survive pruning in these 2-lepton only Run 2 fits; well over 500 survive in the Run 1+Run 2 combined fit.

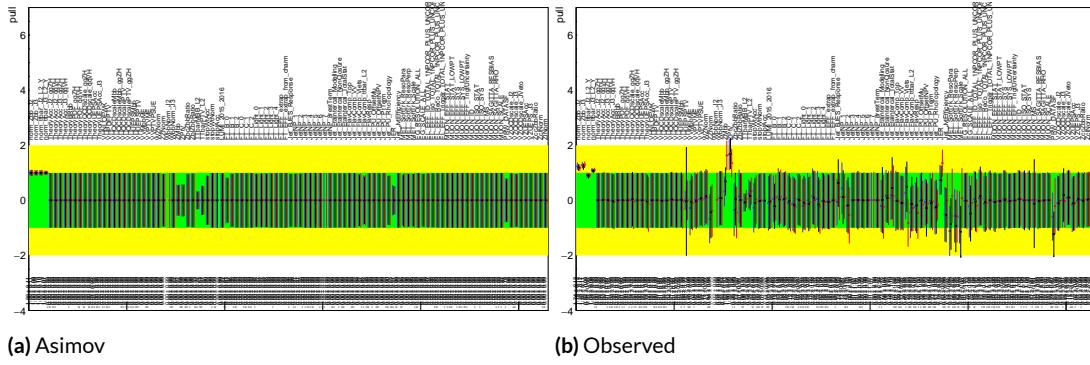


Figure 7.4: Pull comparison for all NP's but MC stats.

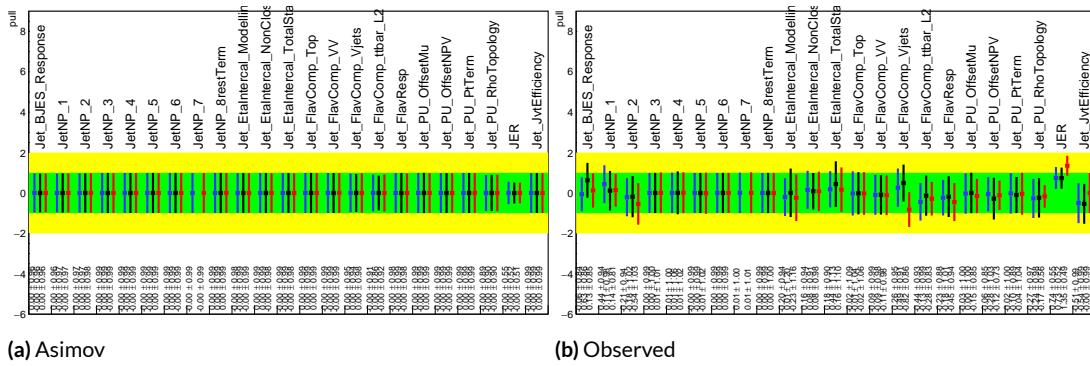


Figure 7.5: Pull comparison for jet NP's.

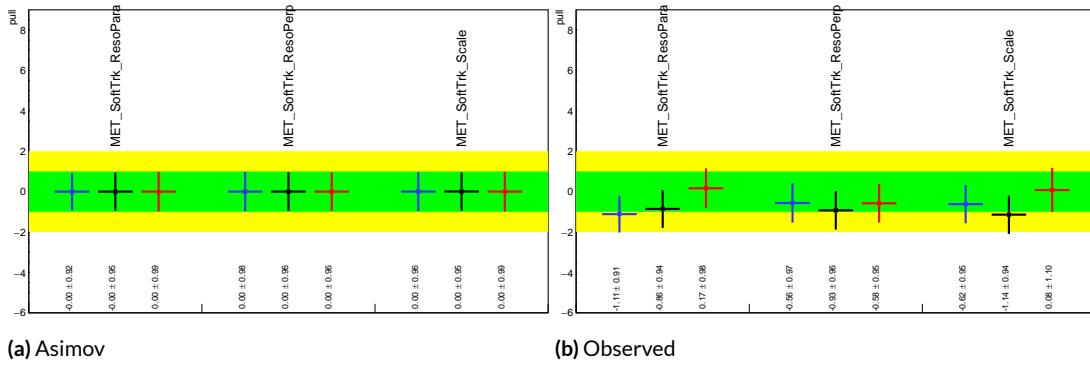


Figure 7.6: Pull comparison for MET NP's.

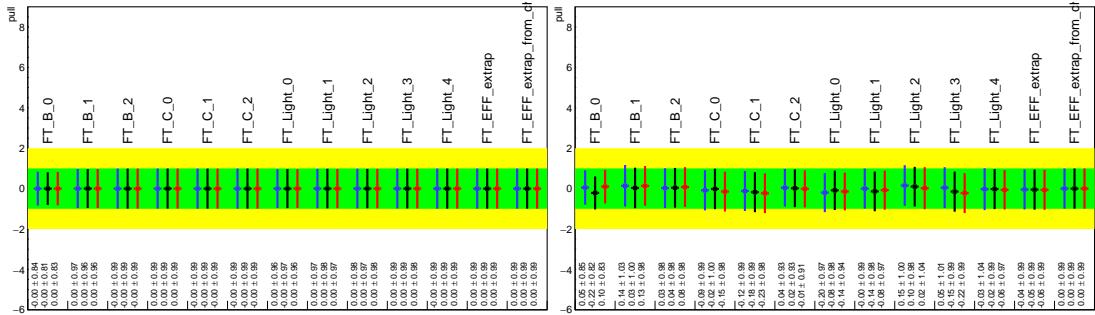


Figure 7.7: Pull comparison for Flavour Tagging NP's.

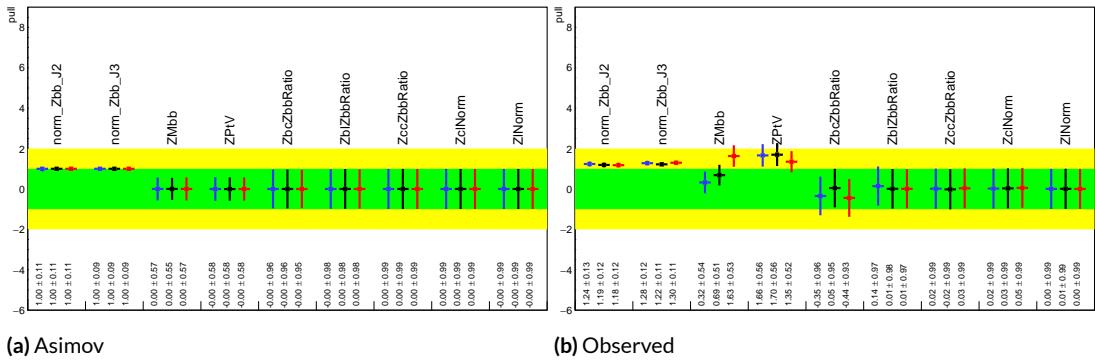


Figure 7.8: Pull comparison for $Z + \text{jets}$ NP's.

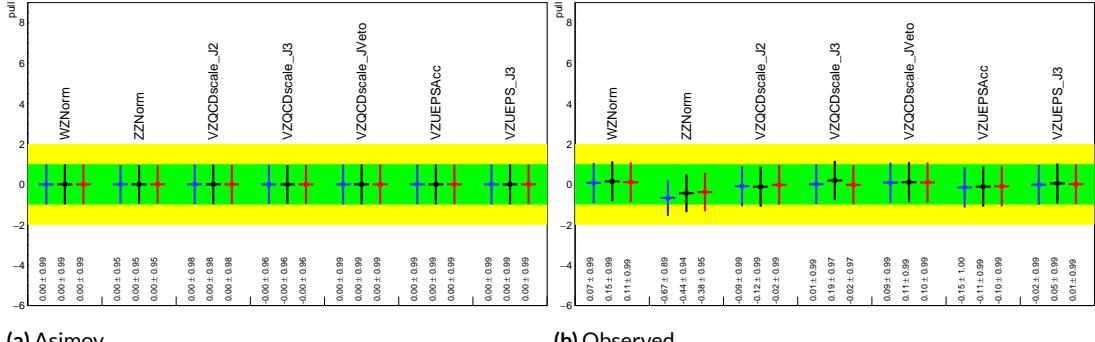


Figure 7.9: Pull comparison for signal process modeling NP's.

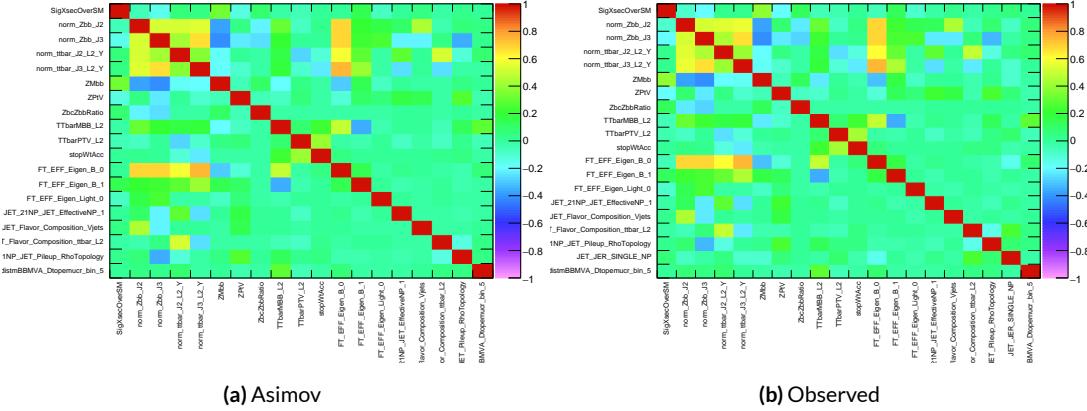


Figure 7.10: NP correlations for standard variable fits.

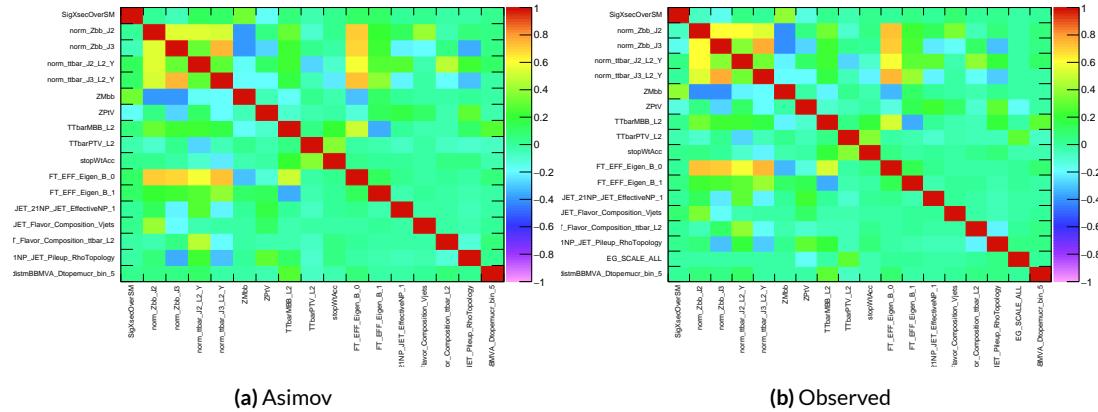


Figure 7.11: NP correlations for L1 variable fits.

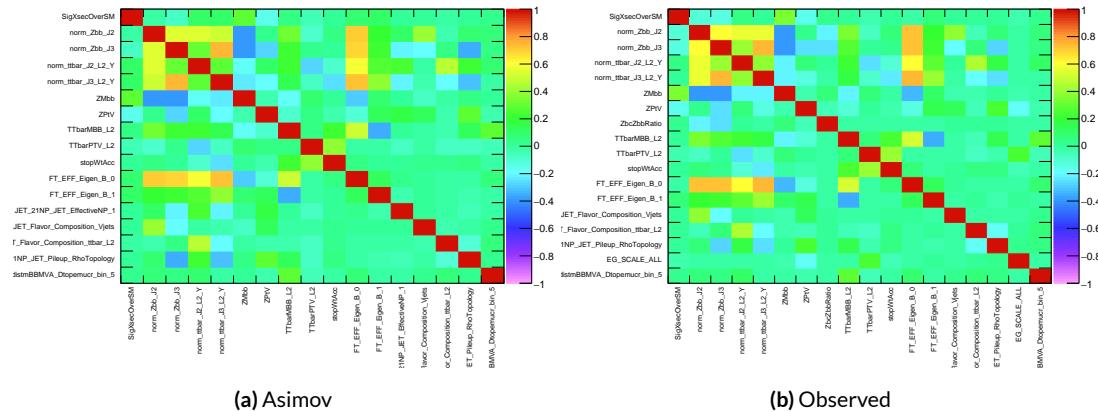


Figure 7.12: NP correlations for RF variable fits.

1674 7.5.1 NUISANCE PARAMETER RANKING PLOTS AND BREAKDOWNS

1675 The next set of fit results that is used to diagnose the quality of a fit is the impact of different nui-
 1676 sance parameters on the total error on μ , both individually and as categories. Figure 7.13 shows the
 1677 top 25 nuisance parameters ranked by their postfit impact on $\hat{\mu}$; these plots use the aforementioned
 1678 more reliable MINOS approach. This set of rankings is fairly similar, with $Z+jets$ systematics being
 1679 particularly prominent. The advantage of seeing individual nuisance parameter rankings, as op-
 1680 posed to impacts of categories in aggregate, is that particularly pathological NP's are easier to see; in
 1681 particular, jet energy resolution and $Z+jets p_T^V$ systematic from the pull comparison plots show up
 with high rankings. Yellow bands are pre-fit impact on μ .

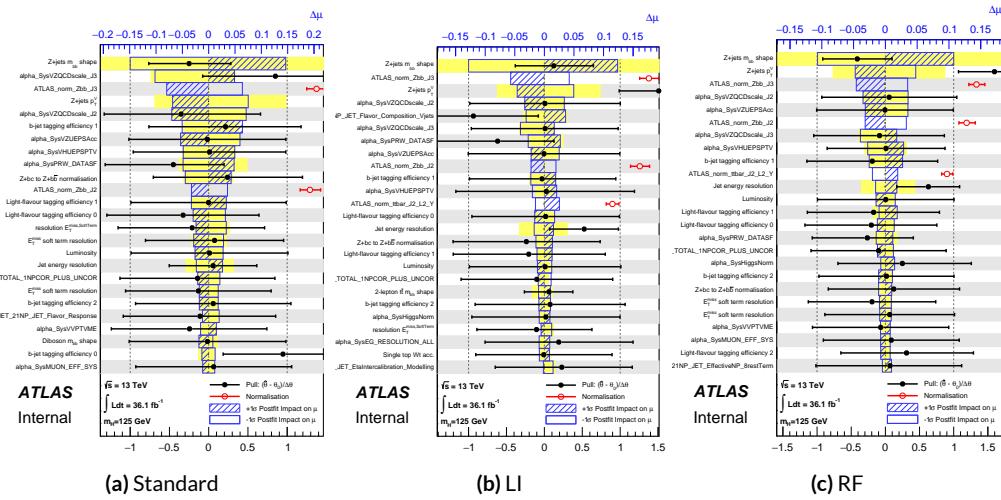


Figure 7.13: Plots for the top 25 nuisance parameters according to their postfit impact on $\hat{\mu}$ for the standard (a), LI (b), and RF (c) variable sets.

1682

1683 This is consistent with the picture of NP's taken in aggregate categories in Tables 7.5 and 7.6,
 1684 known as “breakdowns,” with $Z+jets$ in particular featuring prominently. Of particular interest is

1685 also the lower impact of MC stats in the observed fit.

	Std-KF	LI+MET	RF
Total	+0.305 / -0.277	+0.324 / -0.292	+0.319 / -0.288
DataStat	+0.183 / -0.179	+0.190 / -0.186	+0.188 / -0.184
FullSyst	+0.244 / -0.212	+0.262 / -0.226	+0.258 / -0.221
Floating normalizations	+0.092 / -0.084	+0.098 / -0.079	+0.094 / -0.076
All normalizations	+0.093 / -0.084	+0.098 / -0.079	+0.094 / -0.076
All but normalizations	+0.214 / -0.179	+0.229 / -0.188	+0.224 / -0.182
Jets, MET	+0.052 / -0.043	+0.041 / -0.034	+0.047 / -0.037
Jets	+0.034 / -0.029	+0.033 / -0.028	+0.032 / -0.026
MET	+0.035 / -0.027	+0.015 / -0.012	+0.020 / -0.016
BTag	+0.064 / -0.051	+0.063 / -0.031	+0.059 / -0.032
BTag b	+0.053 / -0.041	+0.061 / -0.028	+0.055 / -0.025
BTag c	+0.011 / -0.010	+0.006 / -0.005	+0.007 / -0.006
BTag light	+0.030 / -0.027	+0.016 / -0.013	+0.022 / -0.019
Leptons	+0.021 / -0.012	+0.022 / -0.014	+0.023 / -0.014
Luminosity	+0.039 / -0.022	+0.039 / -0.022	+0.040 / -0.022
Diboson	+0.049 / -0.028	+0.047 / -0.026	+0.047 / -0.026
Model Zjets	+0.106 / -0.105	+0.113 / -0.110	+0.102 / -0.099
Zjets flt. norm.	+0.039 / -0.053	+0.024 / -0.029	+0.021 / -0.031
Model Wjets	+0.000 / -0.000	+0.000 / -0.000	+0.000 / -0.000
Wjets flt. norm.	+0.000 / -0.000	+0.000 / -0.000	+0.000 / -0.000
Model ttbar	+0.015 / -0.013	+0.032 / -0.017	+0.030 / -0.016
Model Single Top	+0.004 / -0.003	+0.009 / -0.008	+0.005 / -0.004
Model Multi Jet	+0.000 / -0.000	+0.000 / -0.000	+0.000 / -0.000
Signal Systematics	+0.003 / -0.003	+0.003 / -0.003	+0.003 / -0.003
MC stat	+0.097 / -0.094	+0.108 / -0.103	+0.107 / -0.104

Table 7.5: Summary of impact of various nuisance parameter categories on the error on μ for Asimov fits for the standard, LI, and RF variable sets.

	Std-KF	LI+MET	RF
$\hat{\mu}$	1.1079	0.5651	0.6218
Total	+0.381 / -0.360	+0.339 / -0.316	+0.322 / -0.299
DataStat	+0.214 / -0.211	+0.210 / -0.205	+0.201 / -0.197
FullSyst	+0.315 / -0.292	+0.267 / -0.241	+0.252 / -0.225
Floating normalizations	+0.120 / -0.122	+0.095 / -0.089	+0.082 / -0.079
All normalizations	+0.121 / -0.123	+0.095 / -0.090	+0.082 / -0.079
All but normalizations	+0.279 / -0.254	+0.228 / -0.200	+0.213 / -0.184
Jets, MET	+0.076 / -0.065	+0.045 / -0.043	+0.038 / -0.033
Jets	+0.047 / -0.040	+0.044 / -0.041	+0.027 / -0.024
MET	+0.055 / -0.046	+0.015 / -0.015	+0.012 / -0.010
BTag	+0.083 / -0.079	+0.041 / -0.031	+0.041 / -0.035
BTag b	+0.063 / -0.059	+0.032 / -0.022	+0.031 / -0.026
BTag c	+0.018 / -0.017	+0.008 / -0.007	+0.010 / -0.009
BTag light	+0.051 / -0.046	+0.024 / -0.021	+0.025 / -0.022
Leptons	+0.022 / -0.011	+0.015 / -0.008	+0.019 / -0.008
Luminosity	+0.044 / -0.022	+0.026 / -0.006	+0.027 / -0.008
Diboson	+0.049 / -0.026	+0.025 / -0.013	+0.027 / -0.017
Model Zjets	+0.156 / -0.162	+0.133 / -0.133	+0.115 / -0.117
Zjets flt. norm.	+0.061 / -0.089	+0.041 / -0.064	+0.028 / -0.056
Model Wjets	+0.000 / -0.001	+0.000 / -0.001	+0.000 / -0.001
Wjets flt. norm.	+0.000 / -0.001	+0.000 / -0.001	+0.000 / -0.001
Model ttbar	+0.015 / -0.024	+0.018 / -0.005	+0.017 / -0.009
Model Single Top	+0.005 / -0.003	+0.010 / -0.008	+0.007 / -0.004
Model Multi Jet	+0.000 / -0.001	+0.000 / -0.001	+0.000 / -0.001
Signal Systematics	+0.005 / -0.004	+0.009 / -0.006	+0.005 / -0.006
MC stat	+0.140 / -0.143	+0.132 / -0.131	+0.128 / -0.129

Table 7.6: Summary of impact of various nuisance parameter categories on the error on $\hat{\mu}$ for observed fits for the standard, LI, and RF variable sets.

1686 7.6 POSTFIT DISTRIBUTIONS

1687 Finally, postfit distributions for the MVA discriminant (m_{bb}) distribution in the signal (top $e - \mu$
1688 control) region for the standard, Lorentz Invariant, and RestFrames variable sets are shown. It is
1689 generally considered good practice to check the actual postfit distributions of discriminating quan-
1690 tities used to make sure there is good agreement.[¶] It should be noted that agreement is not always
1691 great when “eyeballing” a distribution, as fits are messy and $V+hf$ modeling is notoriously hard to
1692 get correct. This is particularly true in the VZ fit since normalizations for $Z+hf$ in particular are de-
1693 rived using VH optimized sidebands. This is also why a lot of these plots are presented as log plots
1694 (which hide disagreement better; the general argument goes that one has the ratio plots on the bot-
1695 tom).

1696 7.7 VH FIT MODEL VALIDATION

1697 We now move onto the fit validation distributions and numbers for the VH fit of interest.

1698 7.7.1 NUISANCE PARAMETER PULLS

1699 As can be seen in Figures 7.20–7.24, the fits for the three different variable sets are fairly similar from
1700 a NP pull perspective. Again, black is the standard variable set, red is the LI set, and blue is the RF
1701 set. The possible exception is the signal UE+PS p_T^V systematic, which looks very different for all
1702 three cases (underconstrained for the standard, but overconstrained for the novel variable cases),

1702 [¶]Sometimes distributions of input variables (MC histograms scaled by their postfit normalizations) are also used.

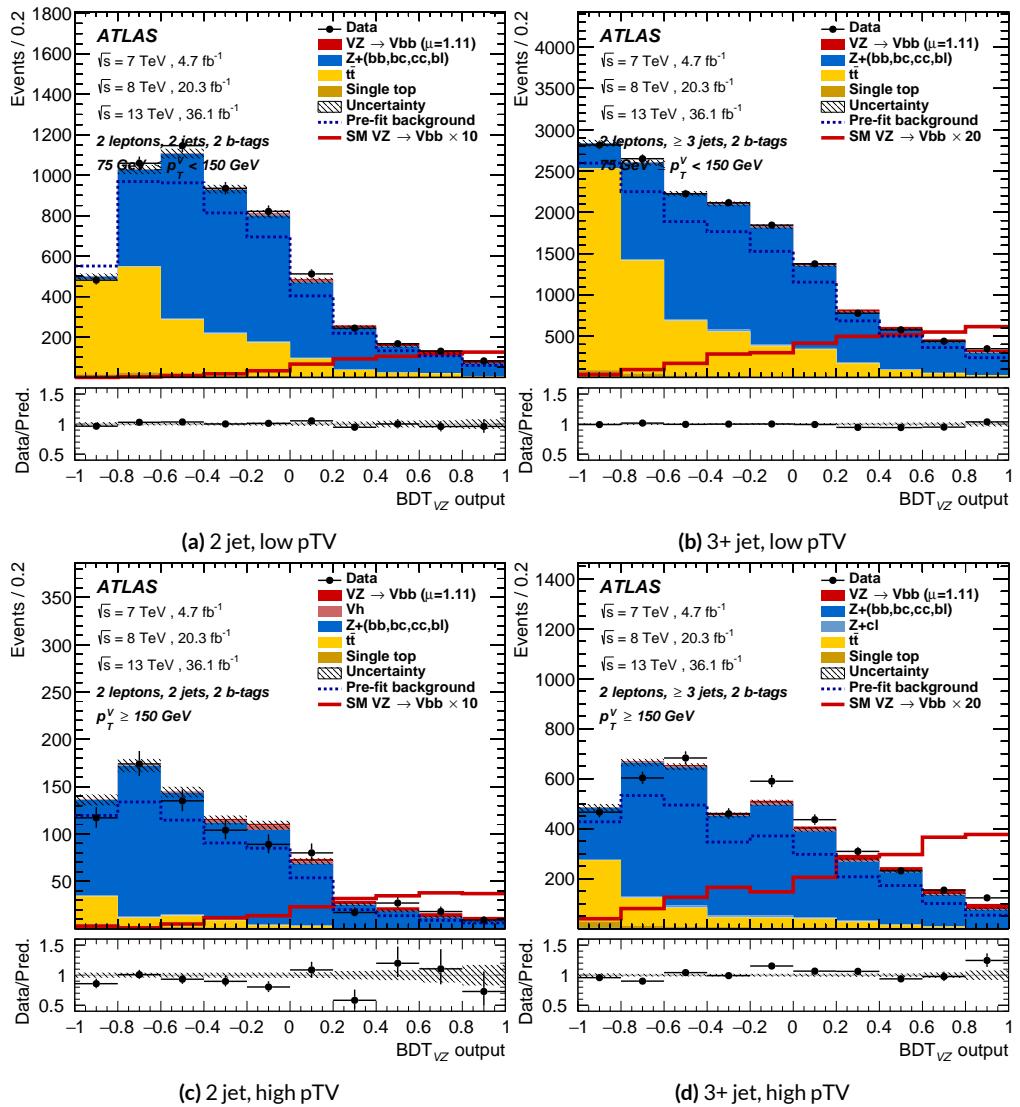


Figure 7.14: Postfit BDT_{VZ} plots in the signal region for the standard variable set.

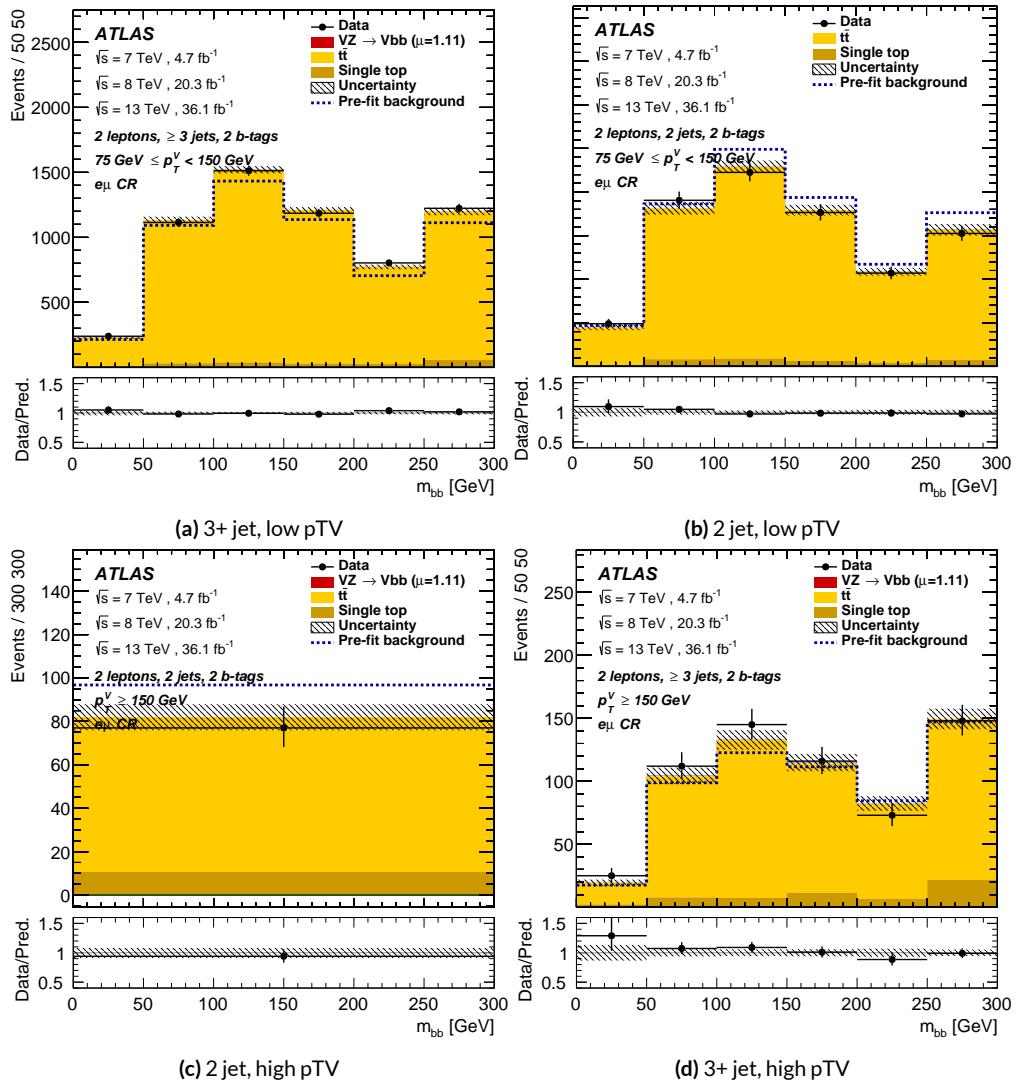


Figure 7.15: Postfit m_{bb} plots in the top $e - \mu$ CR for the standard variable set.

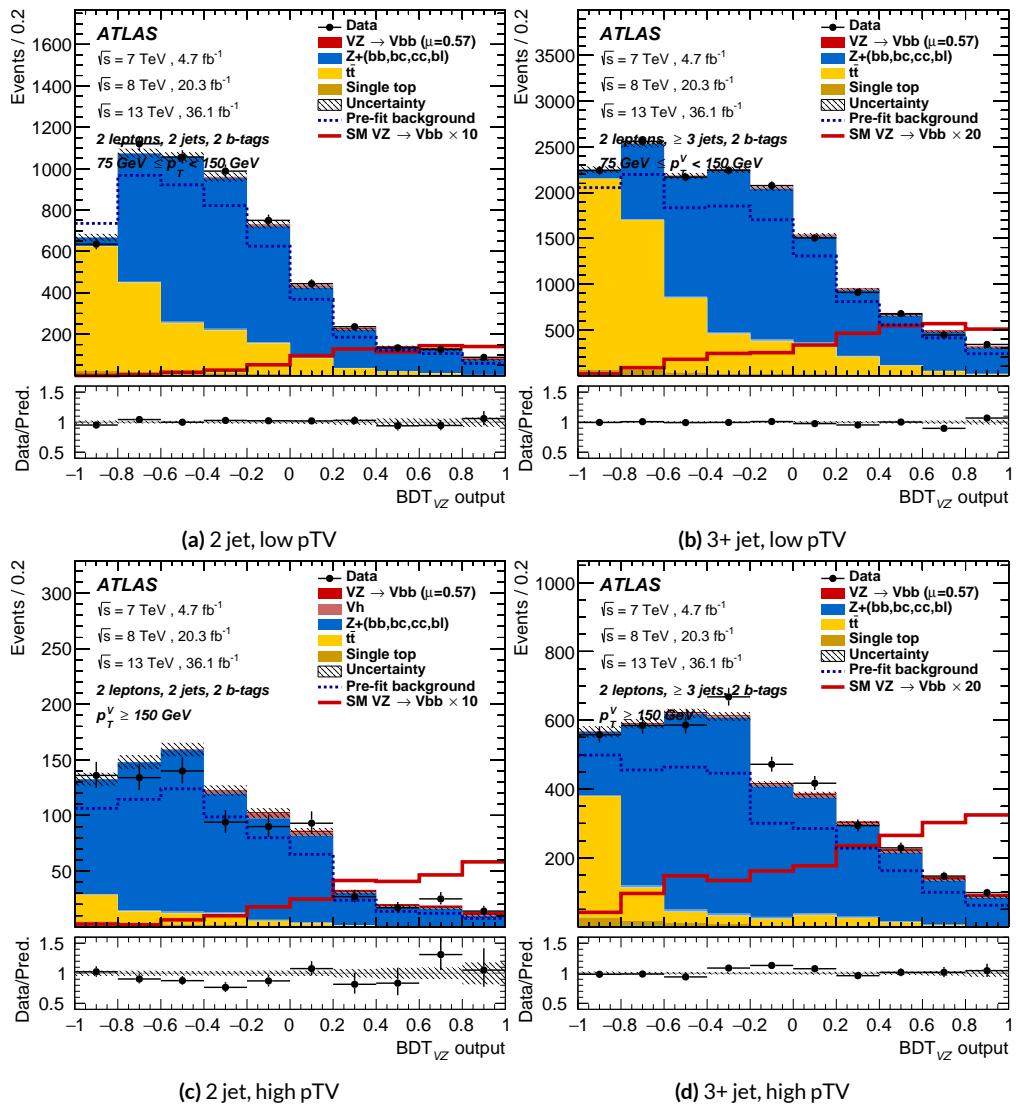


Figure 7.16: Postfit BDT_{VZ} plots in the signal region for the LI variable set.

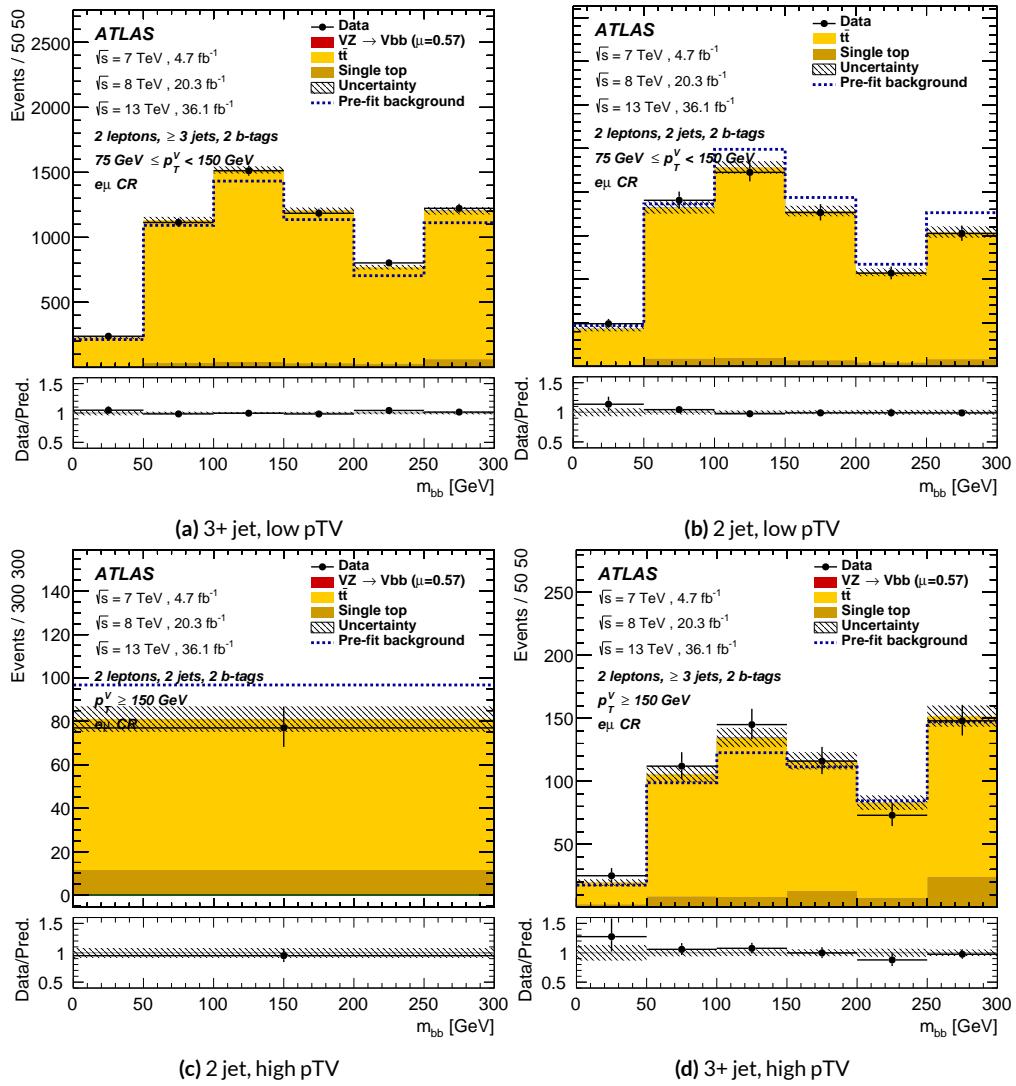


Figure 7.17: Postfit m_{bb} plots in the top $e - \mu$ CR for the LI variable set.

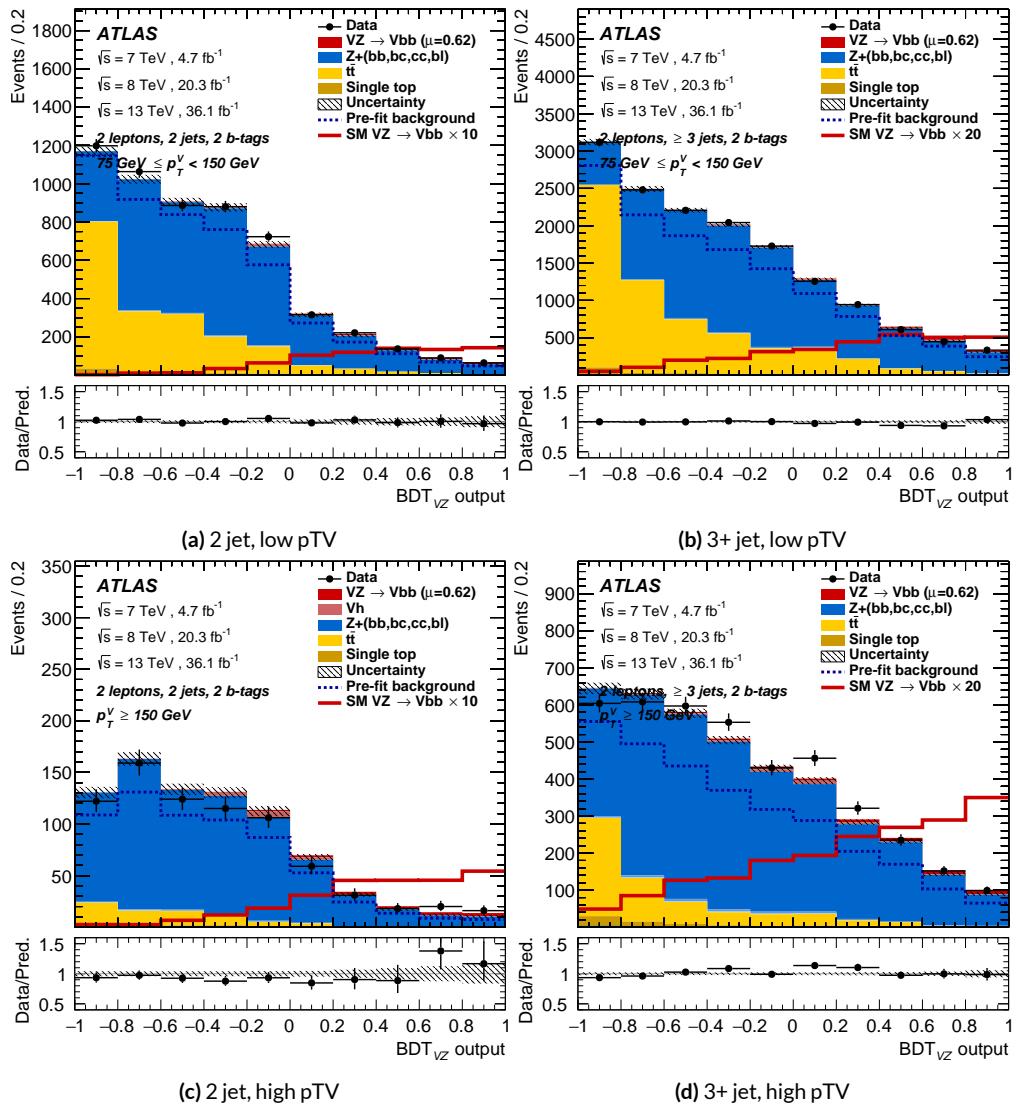


Figure 7.18: Postfit BDT_{VZ} plots in the signal region for the RF variable set.

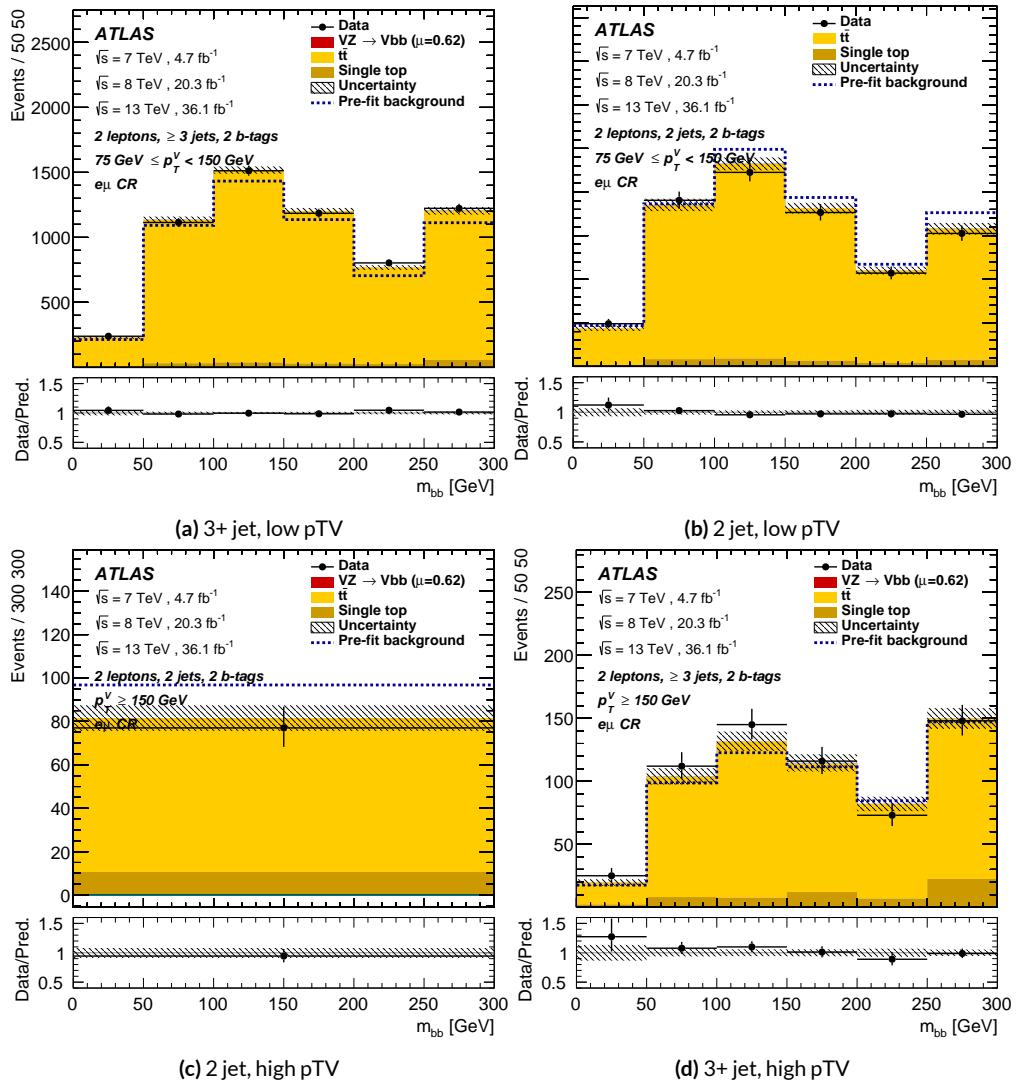


Figure 7.19: Postfit m_{bb} plots in the top $e - \mu$ CR for the RF variable set.

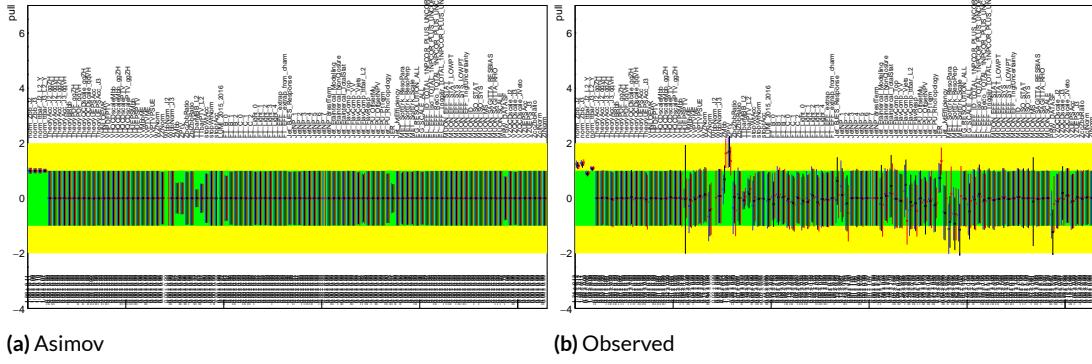


Figure 7.20: Pull comparison for all NP's but MC stats.

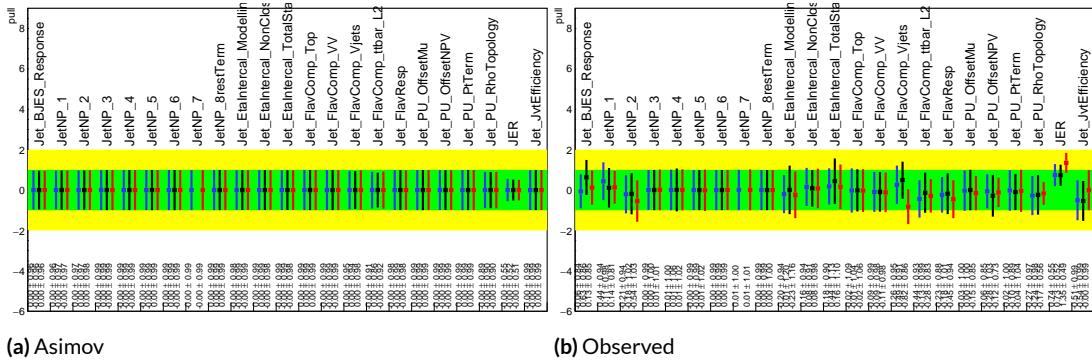


Figure 7.21: Pull comparison for jet NP's.

1703 though this difference goes away in the ranking plot.

1704 Nuisance parameter correlation matrices (for correlations with magnitude at least 0.25) for all

1705 three variable set fits can be found in Figures 7.26–7.28.

1706 7.7.2 FULL BREAKDOWN OF ERRORS

1707 A postfit ranking of nuisance parameters according to their impact on $\hat{\mu}$ for the different variable

1708 sets may be found in Figure 7.29, with rankings being fairly similar. In particular, the signal UE+PS

1709 p_T^V systematic is top-ranked for all three variable sets and also looks very similar, unlike in the pull

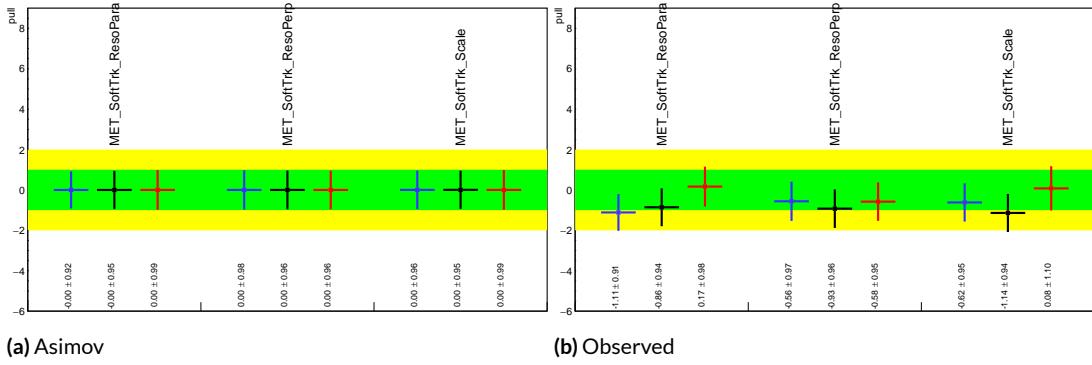


Figure 7.22: Pull comparison for MET NP's.

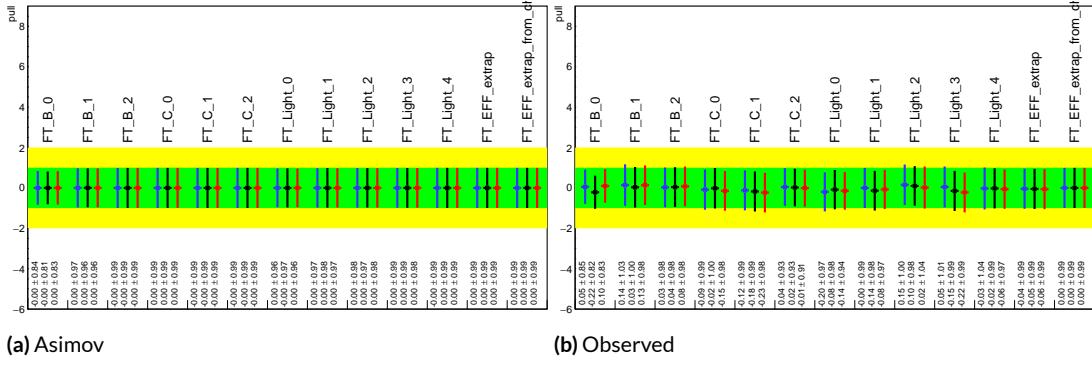


Figure 7.23: Pull comparison for Flavour Tagging NP's.

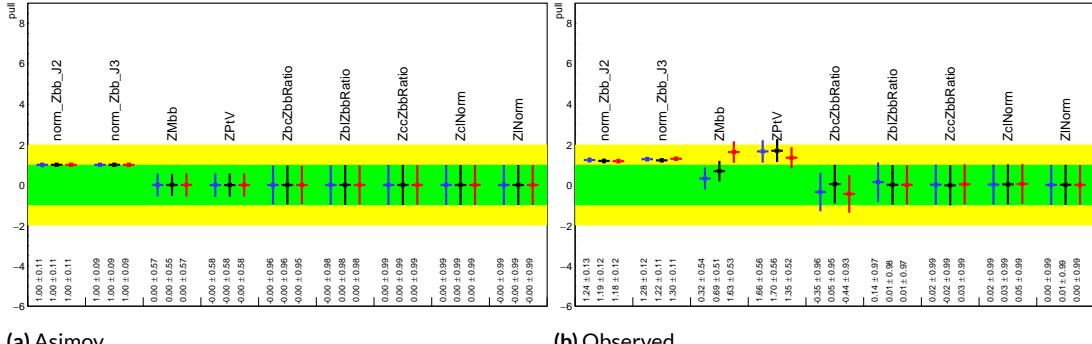
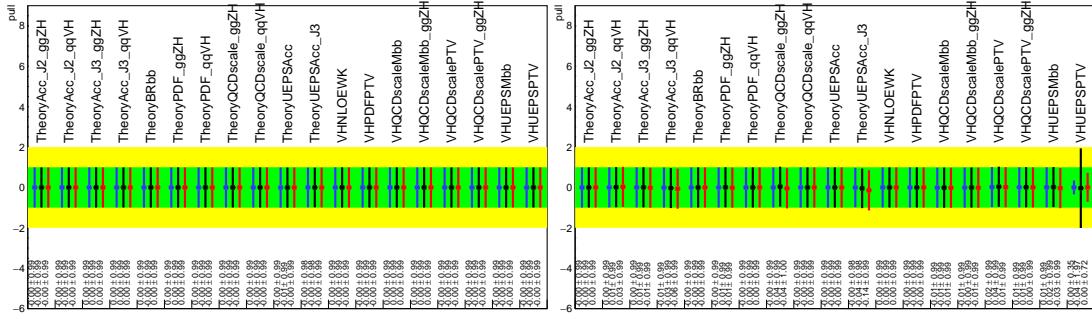


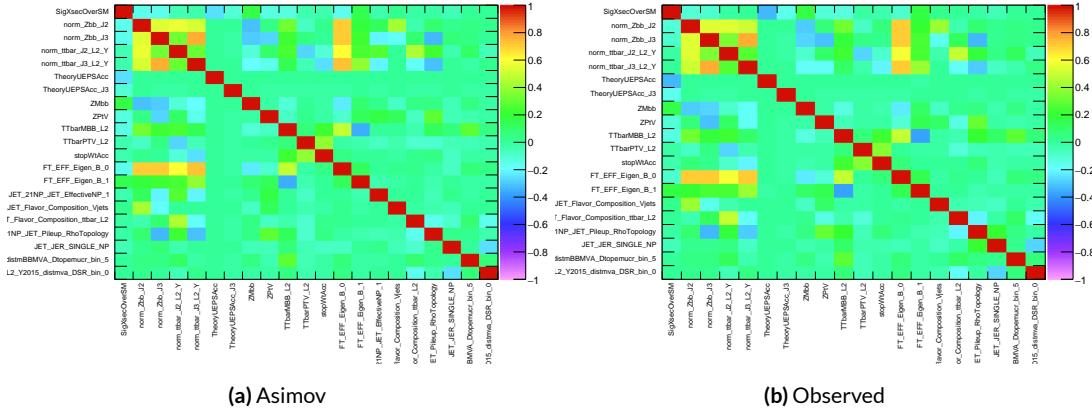
Figure 7.24: Pull comparison for Z +jets NP's.



(a) Asimov

(b) Observed

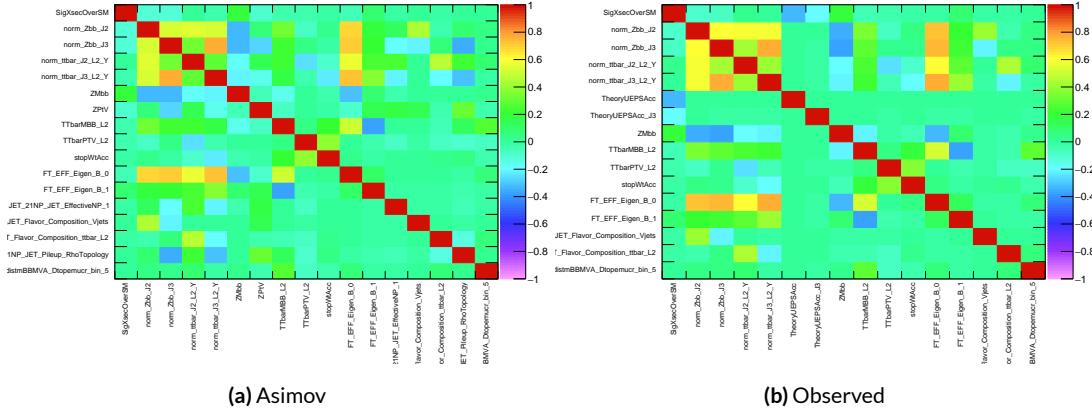
Figure 7.25: Pull comparison for signal process modeling NP's.



(a) Asimov

(b) Observed

Figure 7.26: NP correlations for standard variable fits.



(a) Asimov

(b) Observed

Figure 7.27: NP correlations for L1 variable fits.

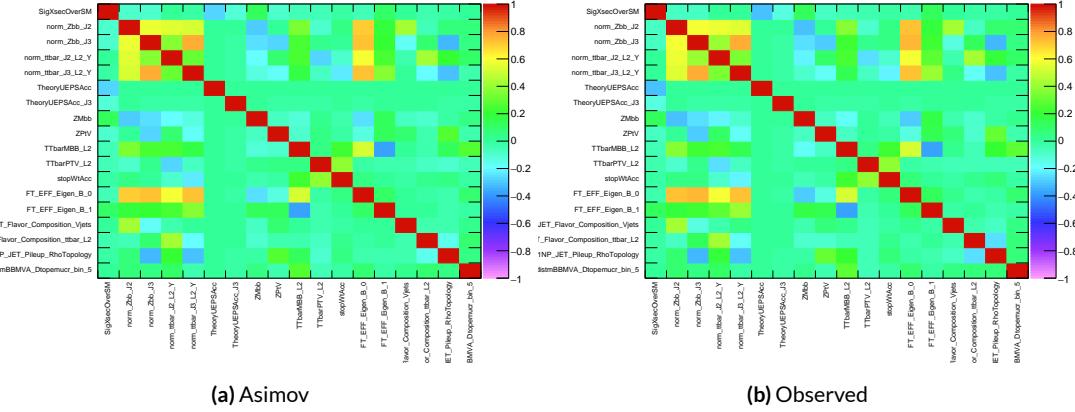


Figure 7.28: NP correlations for RF variable fits.

comparison plot, reiterating the importance of evaluating individually the impact of highly ranked NP's. The $Z+jets p_T^V$ is highly pulled in all three cases, though this is less severe for the non-standard set (it is off the scale for the standard). The RF discriminant mitigates the effect of poorly modeled jet energy resolution better than the other sets.

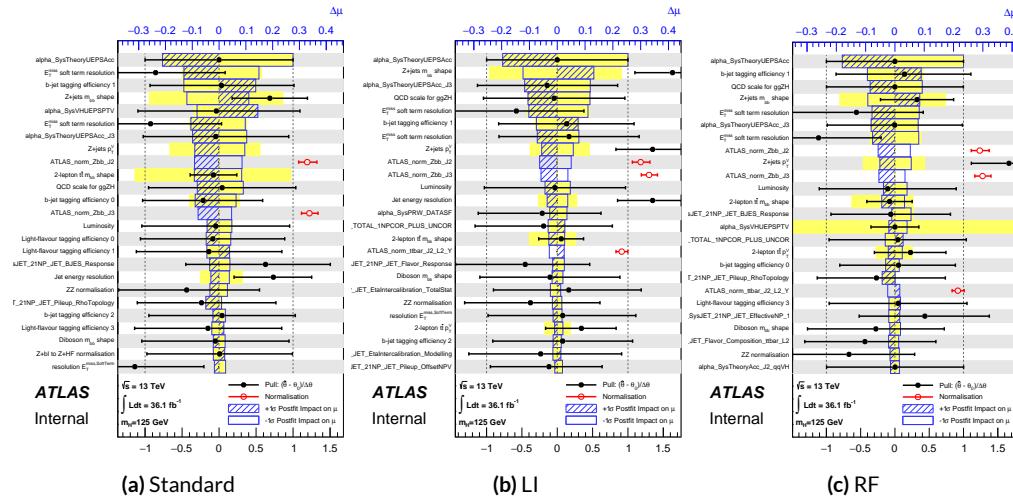


Figure 7.29: Plots for the top 25 nuisance parameters according to their postfit impact on $\hat{\mu}$ for the standard (a), LI (b), and RF (c) variable sets.

¹⁷¹⁴ The Asimov (Table 7.7) and observed (Table 7.8) breakdowns both consistently suggest that the
¹⁷¹⁵ LI variable set does a better job of constraining systematic uncertainties than the standard set and
¹⁷¹⁶ that the RF set does better still. It is also not surprising that the gain is more substantial in the ob-
¹⁷¹⁷ served fit than in the Asimov fits, as in the latter there are the “penalty” terms from pulls in addition
¹⁷¹⁸ to the overall broadening in the likelihood.

	Std-KF	LI+MET	RF
Total	+0.608 / -0.511	+0.632 / -0.539	+0.600 / -0.494
DataStat	+0.420 / -0.401	+0.453 / -0.434	+0.424 / -0.404
FullSyst	+0.440 / -0.318	+0.441 / -0.319	+0.425 / -0.284
Floating normalizations	+0.122 / -0.125	+0.110 / -0.111	+0.093 / -0.089
All normalizations	+0.128 / -0.129	+0.112 / -0.112	+0.099 / -0.092
All but normalizations	+0.403 / -0.274	+0.387 / -0.250	+0.382 / -0.227
Jets, MET	+0.180 / -0.097	+0.146 / -0.079	+0.122 / -0.083
Jets	+0.051 / -0.030	+0.044 / -0.035	+0.025 / -0.042
MET	+0.173 / -0.091	+0.140 / -0.074	+0.117 / -0.063
BTag	+0.138 / -0.136	+0.069 / -0.071	+0.076 / -0.078
BTag b	+0.125 / -0.125	+0.067 / -0.070	+0.073 / -0.075
BTag c	+0.018 / -0.016	+0.004 / -0.004	+0.005 / -0.005
BTag light	+0.057 / -0.051	+0.020 / -0.014	+0.009 / -0.018
Leptons	+0.013 / -0.012	+0.029 / -0.026	+0.012 / -0.023
Luminosity	+0.052 / -0.020	+0.050 / -0.016	+0.050 / -0.019
Diboson	+0.043 / -0.039	+0.035 / -0.031	+0.038 / -0.029
Model Zjets	+0.119 / -0.117	+0.124 / -0.127	+0.095 / -0.086
Zjets flt. norm.	+0.080 / -0.106	+0.052 / -0.092	+0.026 / -0.072
Model Wjets	+0.001 / -0.001	+0.001 / -0.001	+0.000 / -0.001
Wjets flt. norm.	+0.000 / -0.000	+0.000 / -0.000	+0.000 / -0.000
Model ttbar	+0.076 / -0.080	+0.025 / -0.035	+0.025 / -0.040
Model Single Top	+0.015 / -0.015	+0.002 / -0.004	+0.021 / -0.007
Model Multi Jet	+0.000 / -0.000	+0.000 / -0.000	+0.000 / -0.000
Signal Systematics	+0.262 / -0.087	+0.272 / -0.082	+0.290 / -0.088
MC stat	+0.149 / -0.136	+0.168 / -0.154	+0.153 / -0.136

Table 7.7: Expected error breakdowns for the standard, LI, and RF variable sets

	Std-KF	LI+MET	RF
$\hat{\mu}$	1.7458	1.6467	1.5019
Total	+0.811 / -0.662	+0.778 / -0.641	+0.731 / -0.612
DataStat	+0.502 / -0.484	+0.507 / -0.489	+0.500 / -0.481
FullSyst	+0.637 / -0.451	+0.591 / -0.415	+0.533 / -0.378
Floating normalizations	+0.153 / -0.143	+0.128 / -0.118	+0.110 / -0.109
All normalizations	+0.158 / -0.147	+0.130 / -0.119	+0.112 / -0.110
All but normalizations	+0.599 / -0.402	+0.544 / -0.354	+0.486 / -0.318
Jets, MET	+0.218 / -0.145	+0.198 / -0.113	+0.167 / -0.106
Jets	+0.071 / -0.059	+0.065 / -0.047	+0.036 / -0.051
MET	+0.209 / -0.130	+0.190 / -0.102	+0.152 / -0.077
BTag	+0.162 / -0.166	+0.093 / -0.070	+0.115 / -0.099
BTag b	+0.142 / -0.147	+0.090 / -0.066	+0.110 / -0.094
BTag c	+0.022 / -0.021	+0.006 / -0.006	+0.007 / -0.007
BTag light	+0.074 / -0.072	+0.025 / -0.022	+0.031 / -0.029
Leptons	+0.039 / -0.029	+0.035 / -0.031	+0.034 / -0.030
Luminosity	+0.079 / -0.039	+0.073 / -0.034	+0.069 / -0.032
Diboson	+0.047 / -0.043	+0.031 / -0.028	+0.029 / -0.028
Model Zjets	+0.164 / -0.152	+0.141 / -0.143	+0.101 / -0.105
Zjets flt. norm.	+0.070 / -0.109	+0.041 / -0.086	+0.033 / -0.083
Model Wjets	+0.001 / -0.001	+0.001 / -0.000	+0.001 / -0.001
Wjets flt. norm.	+0.000 / -0.000	+0.000 / -0.000	+0.000 / -0.000
Model ttbar	+0.067 / -0.102	+0.029 / -0.040	+0.040 / -0.048
Model Single Top	+0.015 / -0.020	+0.001 / -0.005	+0.004 / -0.006
Model Multi Jet	+0.000 / -0.000	+0.000 / -0.000	+0.000 / -0.000
Signal Systematics	+0.434 / -0.183	+0.418 / -0.190	+0.364 / -0.152
MC stat	+0.226 / -0.201	+0.221 / -0.200	+0.212 / -0.189

Table 7.8: Observed signal strengths, and error breakdowns for the standard, LI, and RF variable sets

1719 7.7.3 POSTFIT DISTRIBUTIONS AND S/B PLOTS

1720 Postfit distributions for the MVA discriminant (m_{bb}) distribution in the signal (top $e - \mu$ control) re-
1721 gion for the standard, Lorentz Invariant, and RestFrames variable sets. Here, as in the VZ fit, agree-
1722 ment is reasonable. In a combined fit with all three channels, $Z + hf$ normalizations in particular
1723 would be correlated across the 0- and 2-lepton channels, which might help to better constrain this
1724 mismodeling (and perhaps as a result some of the $Z + jets$ systematics as well).

1725 One final type of plot presented as a result is the binned $\log_{10} (S/B)$ in signal regions distribu-
1726 tions may be found in Figure 7.36. For these plots, one fills a histogram with the $\log_{10} (S/B)$ ratio
1727 in each postfit distribution bin weighted by the total number of events. In this case, a log plot is
1728 helpful because otherwise the highest bins would be invisible on a linear plot. These distributions
1729 are allegedly useful for seeing where most of one's sensitivity lies. Practically, it is problematic if the
1730 pull (from the null hypothesis) is higher at lower S/B values, which may indicate a poorly optimized
1731 discriminant.

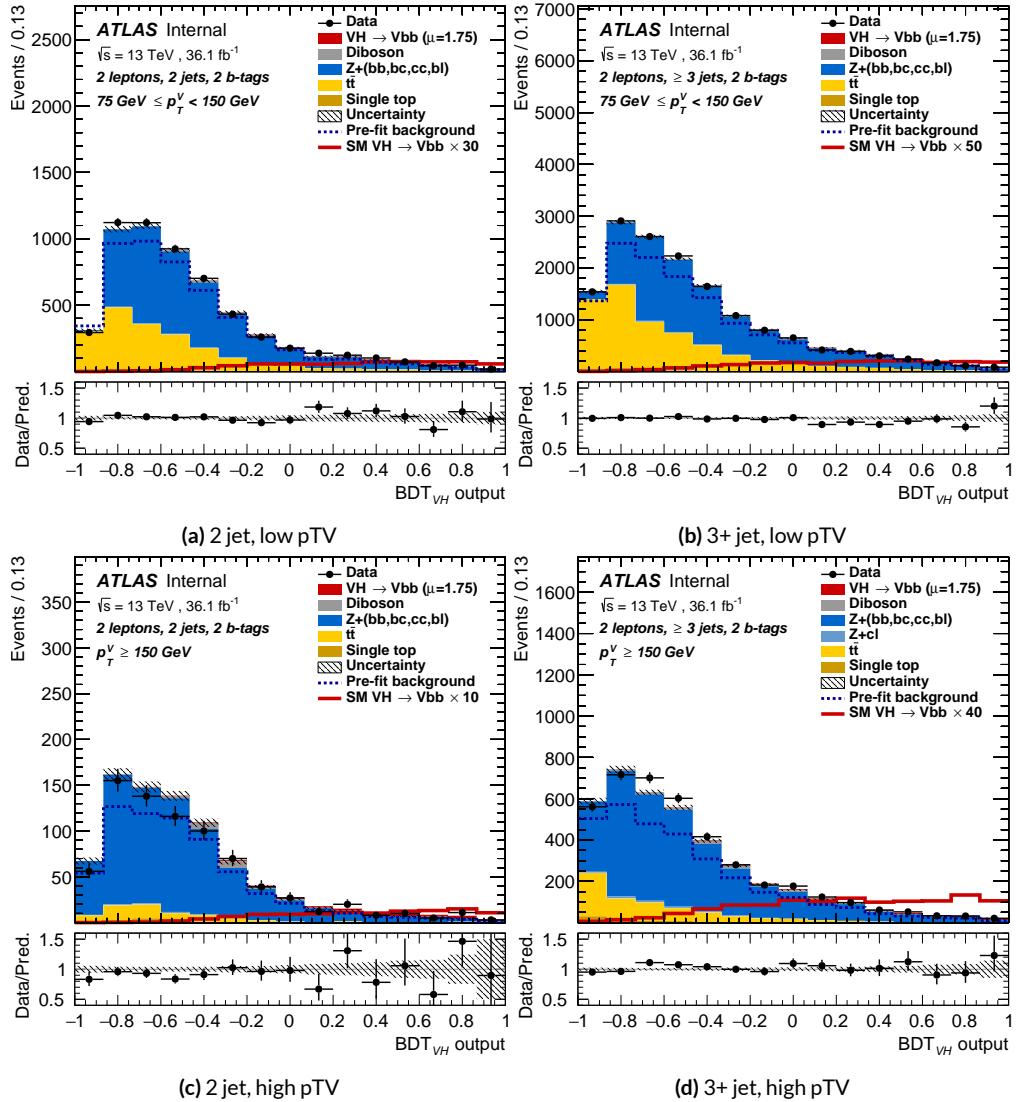


Figure 7.30: Postfit BDT_{VH} plots in the signal region for the standard variable set.

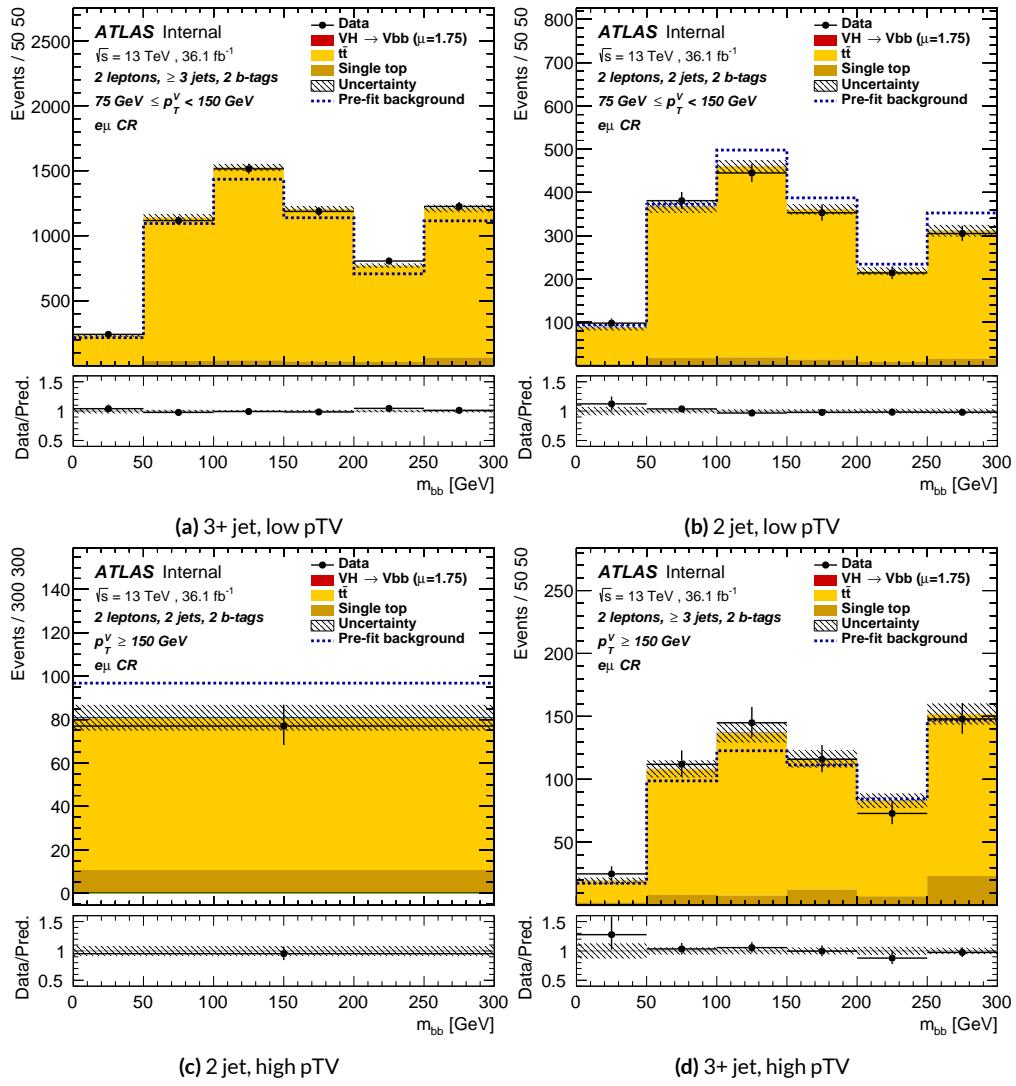


Figure 7.31: Postfit m_{bb} plots in the top $e - \mu$ CR for the standard variable set.

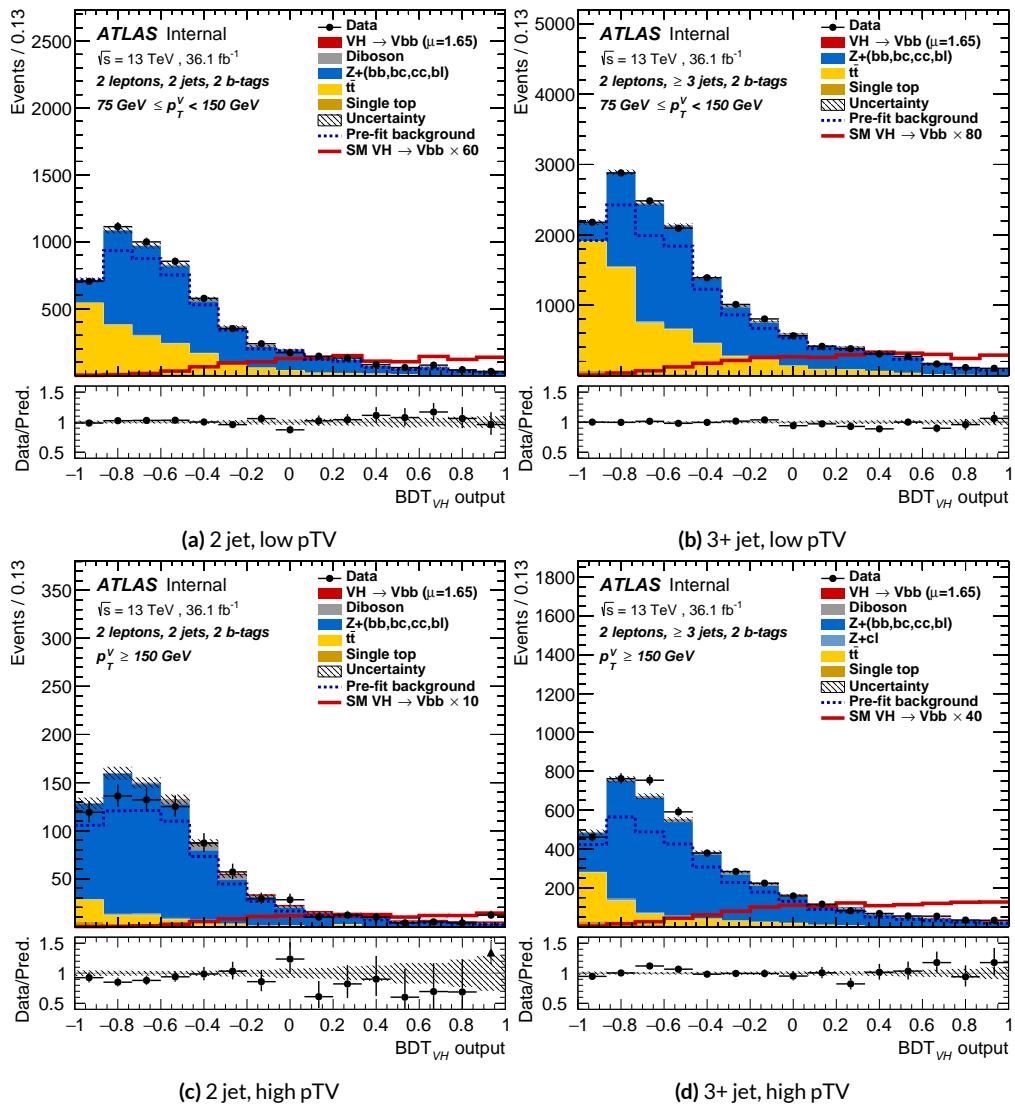


Figure 7.32: Postfit BDT_{VH} plots in the signal region for the LI variable set.

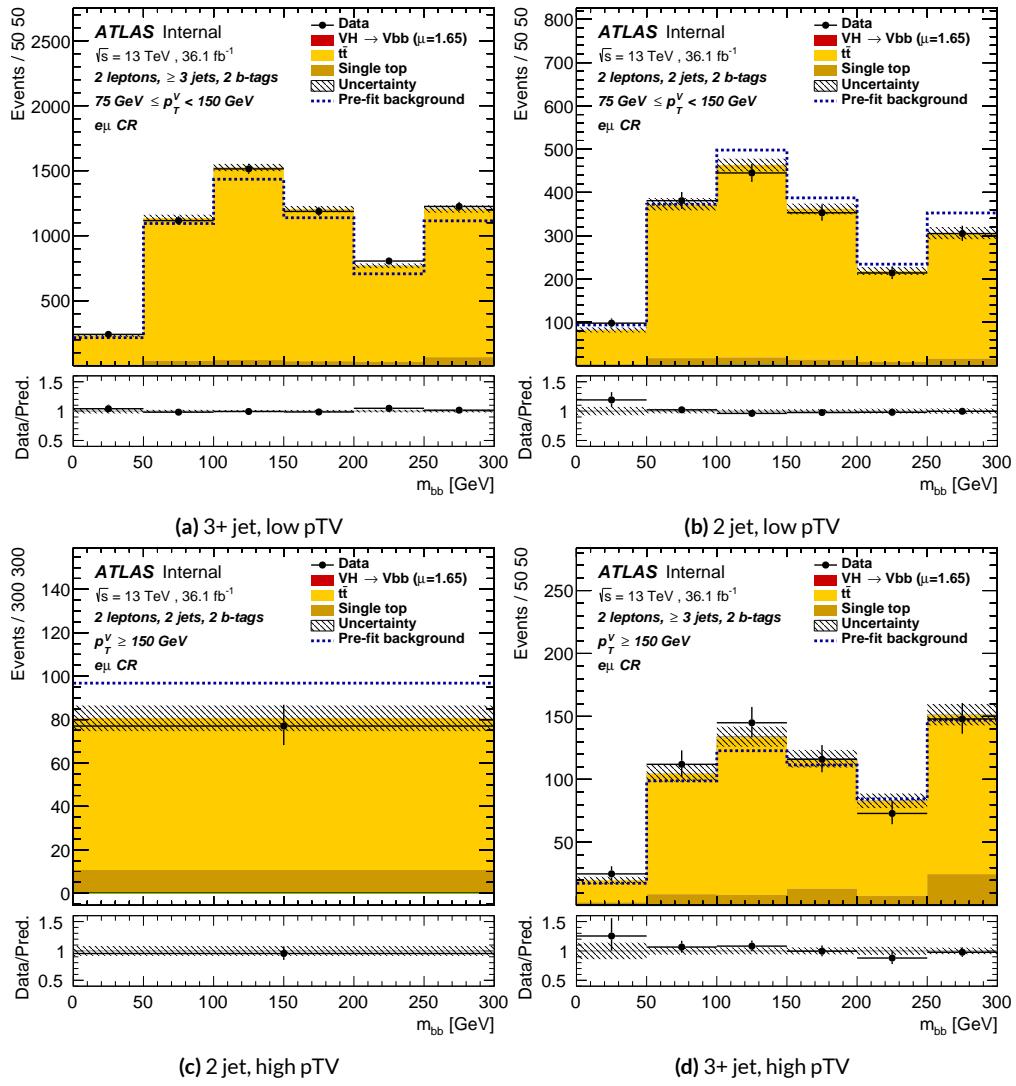


Figure 7.33: Postfit m_{bb} plots in the top $e - \mu$ CR for the LI variable set.

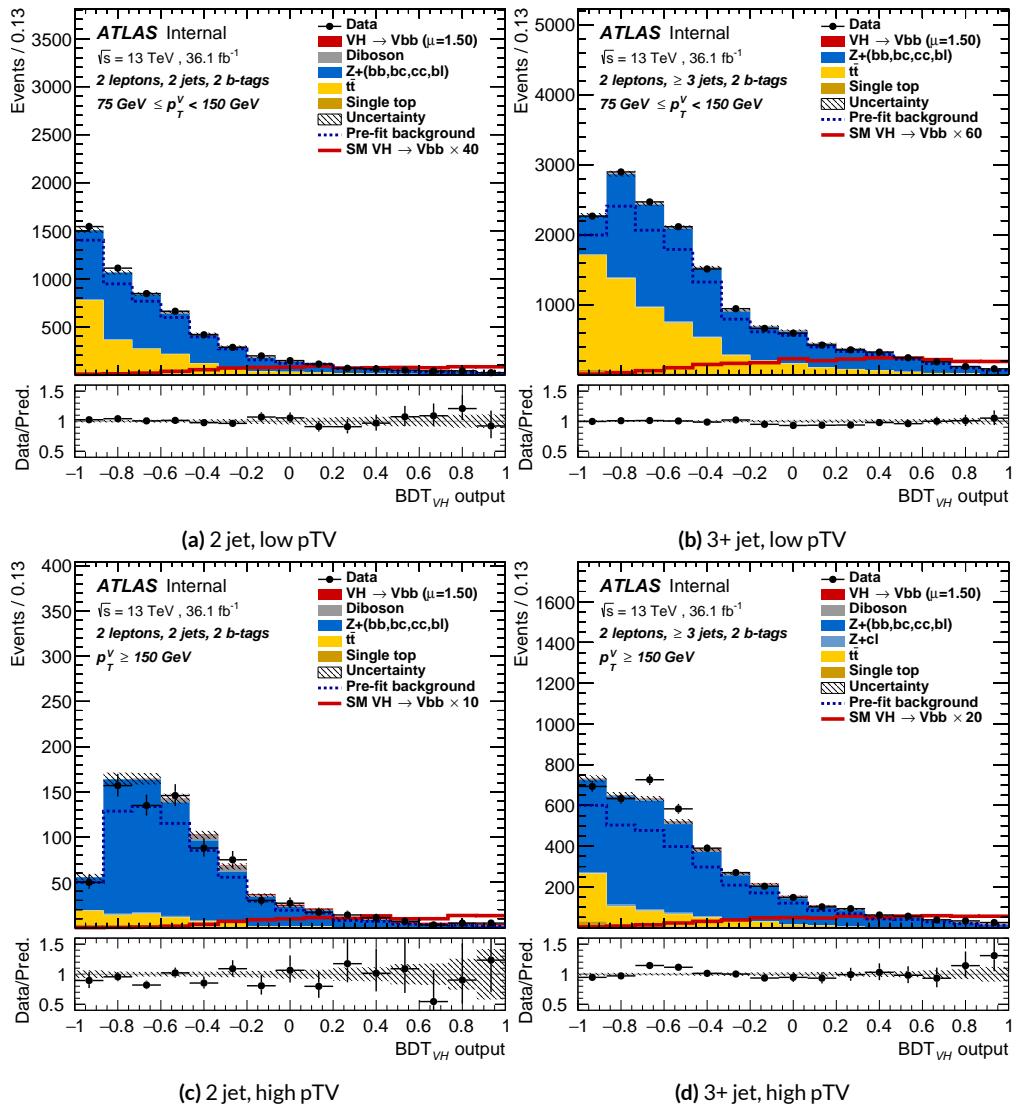


Figure 7.34: Postfit BDT_{VH} plots in the signal region for the RF variable set.

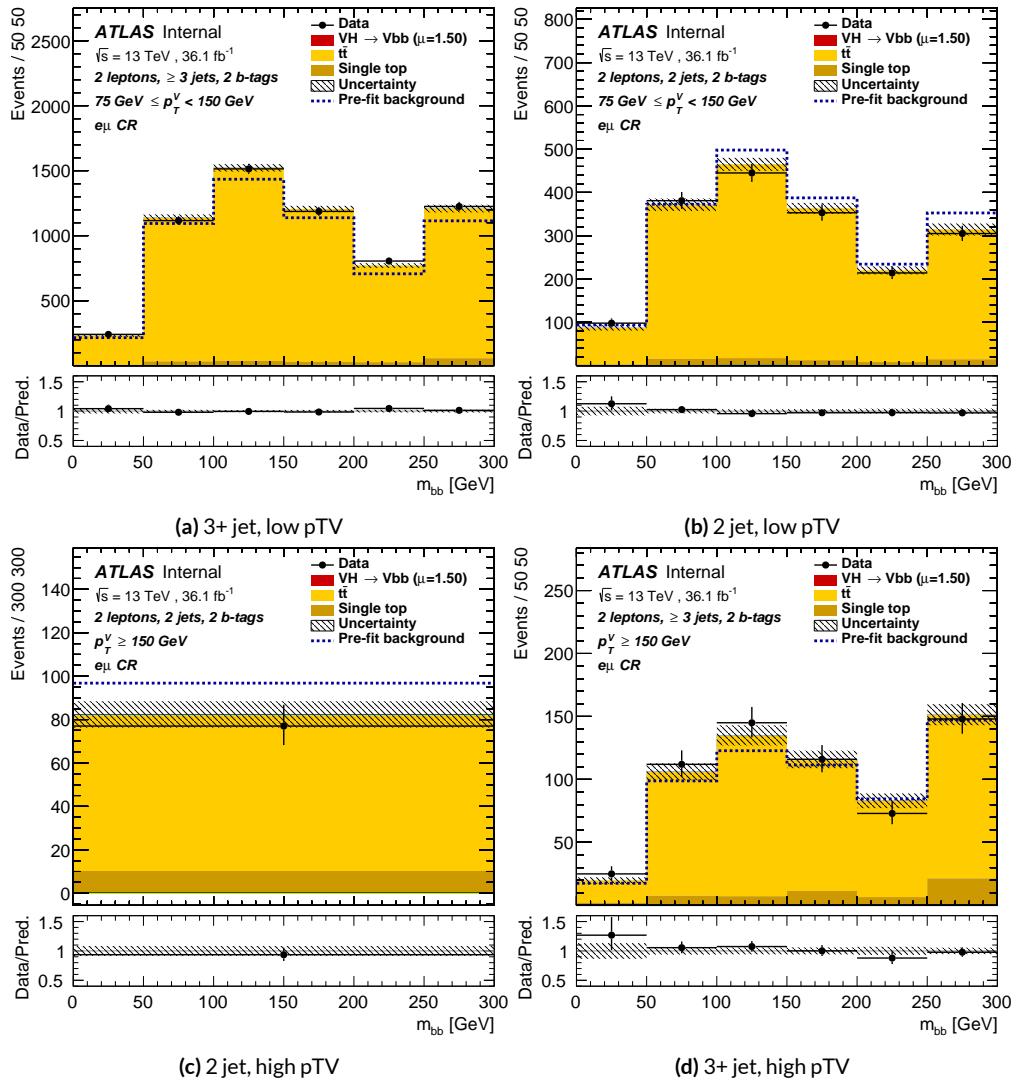


Figure 7.35: Postfit m_{bb} plots in the top $e - \mu$ CR for the RF variable set.

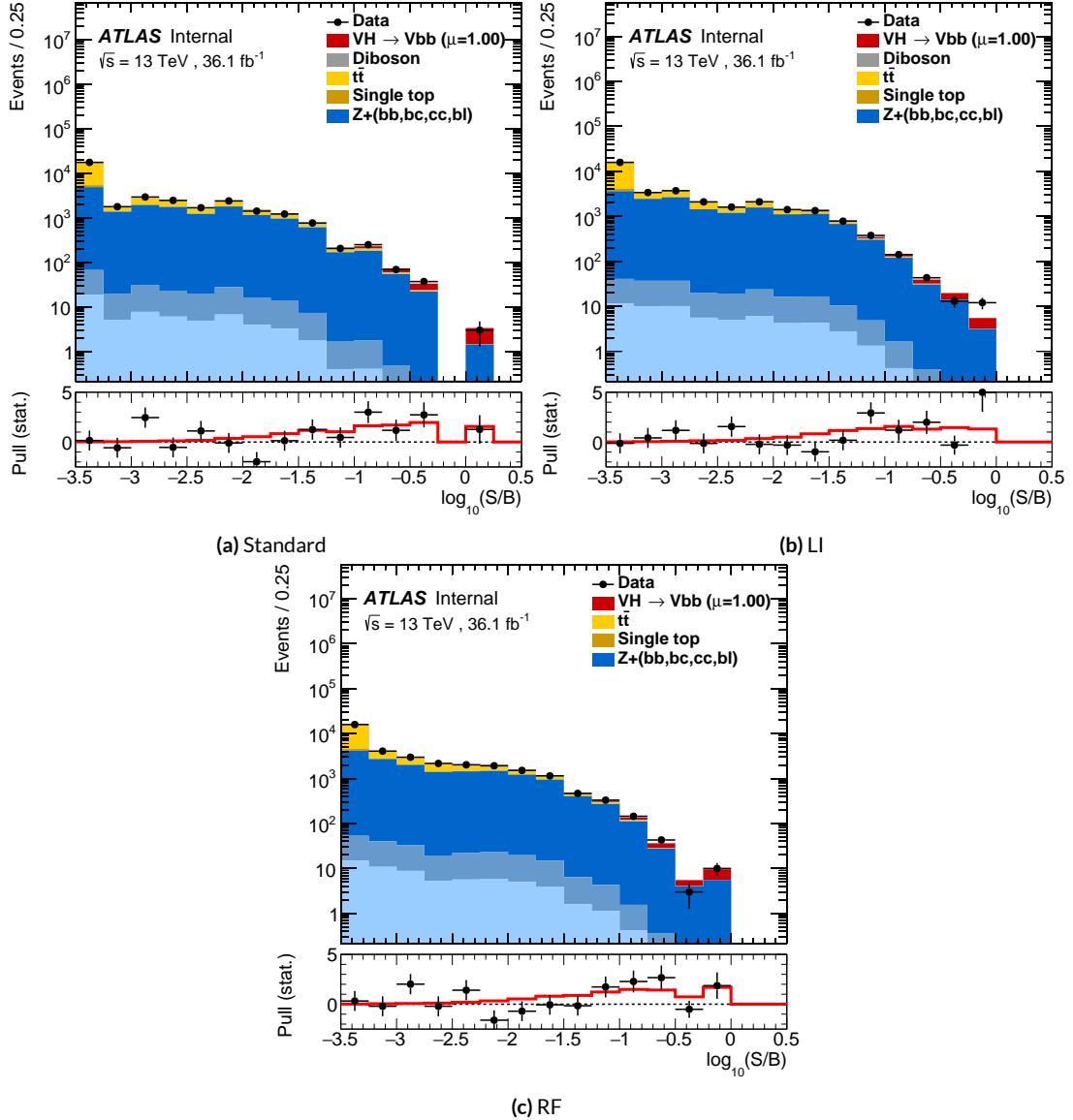


Figure 7.36: Binned S/B plots for the standard (a), LI (b), and RF (c) variable sets. Signal is weighted to $\mu = 1$ for comparison to the SM prediction.

*Kein Operationsplan reicht mit einiger Sicherheit
über das erste Zusammentreffen mit der feindlichen
Hauptmacht hinaus.*

Helmuth von Moltke

8

1732

1733

Fit Results

1734 THE RESULTS IN THIS CHAPTER were first reported in ²⁹ and describes how the three different fit models detailed and validated Chapter 7, corresponding to the standard, RF, and LI variable sets described in Chapter 6 perform on actual VH fits. In particular sensitivities, nuisance parameter impacts, and signal strengths on expected fits to Asimov datasets and both expected and observed

1738 fits on the actual dataset are compared.

1739 Expected and observed sensitivities for the different variable sets may be found in Table 8.1. The
1740 RF fits feature the highest expected sensitivities, outperforming the standard set by 3.5% and 3.4%
1741 for fits to Asimov and observed datasets, respectively. The LI variable has a lower significance than
1742 both for expected fits to both Asimov and data with a 6.7% (1.7%) significance than the standard set
1743 for the Asimov (observed) dataset. While the fit using standard variables does have a higher observed
1744 significance than both the LI and RF fits, by 2.8% and 8.6%, respectively, these numbers should be
1745 viewed in the context of the best fit $\hat{\mu}$ values, discussed below. That is, the standard set may yield the
1746 highest sensitivity for this particular dataset, but this is not necessarily (and likely is not) the case for
1747 any given dataset.

	Standard	LI	RF
Expected (Asimov)	2.06	1.92	2.13
Expected (data)	1.76	1.73	1.80
Observed (data)	2.87	2.79	2.62

Table 8.1: Expected (for both data and Asimov) and observed significances for the standard, LI, and RF variable sets.

1748 A summary of fitted signal strengths and errors for both the Asimov (a) and observed (b) datasets
1749 are shown in Figure 8.1.* A summary of error breakdowns is given in Tables 8.2 (Asimov) and 8.3
1750 (observed) for total error, data statistics contributions, total systematic error contributions, and
1751 categories for which the total impact is ≥ 0.1 for the standard fit. As is to be expected for both the
1752 Asimov and observed dataset fits, the contribution to the total error on μ arising from data statistics

*For reference, the standalone 2-lepton fit from the fiducial analysis is $2.11^{+0.50}_{-0.48}$ (stat.) $^{+0.64}_{-0.47}$ (syst.)

¹⁷⁵³ is nearly identical, since each set of fits uses the same selections and data.[†]

	Std-KF	LI+MET	RF
Total	+0.608 / -0.511	+0.632 / -0.539	+0.600 / -0.494
DataStat	+0.420 / -0.401	+0.453 / -0.434	+0.424 / -0.404
FullSyst	+0.440 / -0.318	+0.441 / -0.319	+0.425 / -0.284
Signal Systematics	+0.262 / -0.087	+0.272 / -0.082	+0.290 / -0.088
MET	+0.173 / -0.091	+0.140 / -0.074	+0.117 / -0.063
Flavor Tagging	+0.138 / -0.136	+0.069 / -0.071	+0.076 / -0.078
Model Zjets	+0.119 / -0.117	+0.124 / -0.127	+0.095 / -0.086

Table 8.2: Summary of error impacts on total μ error for principal categories in the Asimov standard, LI, and RF fits.

	Std-KF	LI+MET	RF
Total	+0.811 / -0.662	+0.778 / -0.641	+0.731 / -0.612
DataStat	+0.502 / -0.484	+0.507 / -0.489	+0.500 / -0.481
FullSyst	+0.637 / -0.451	+0.591 / -0.415	+0.533 / -0.378
Signal Systematics	+0.434 / -0.183	+0.418 / -0.190	+0.364 / -0.152
MET	+0.209 / -0.130	+0.190 / -0.102	+0.152 / -0.077
Flavor Tagging	+0.162 / -0.166	+0.093 / -0.070	+0.115 / -0.099
Model Zjets	+0.164 / -0.152	+0.141 / -0.143	+0.101 / -0.105

Table 8.3: Summary of error impacts on total $\hat{\mu}$ error for principal categories in the observed standard, LI, and RF fits.

¹⁷⁵⁴ The contribution from systematic uncertainties, however, does vary considerably across the vari-
¹⁷⁵⁵ able sets. The Asimov fits are a best case scenario in the sense that, by construction, all NP's are equal
¹⁷⁵⁶ to their predicted values (and so no "penalty" is paid for pulls on Gaussian NP's). The systematics
¹⁷⁵⁷ error from the LI fit is slightly higher (subpercent) than that from the standard fit, and 4.6% higher
¹⁷⁵⁸ error overall due to differences in data stats. The RF Asimov fit, however, has a 6.5% lower total er-

[†]Though not exactly identical. Since the BDT's are different for the different variable sets, the binning (as determined by transformation D) and bin contents in each set are generally different, leading to slightly different data statistics errors.

ror from systematics than the standard Asimov fit (and a 2.2% lower error overall). Moreover, for
 both the LI and RF sets, errors are markedly smaller for the MET and Flavor Tagging categories,
 with the RF fit also featuring a smaller errors on $Z+jets$ modeling; the only notable exception to this
 trend in Asimov fits are the signal systematics.

These trends are more pronounced in the observed fits. As can be seen in Table 8.3, both the
 LI and RF fits have smaller errors from systematic uncertainties, both overall and in all principal
 categories, with the LI and RF fits having 7.5% (3.7%) and 16% (8.8%) lower systematics (total) error
 on $\hat{\mu}$, respectively.

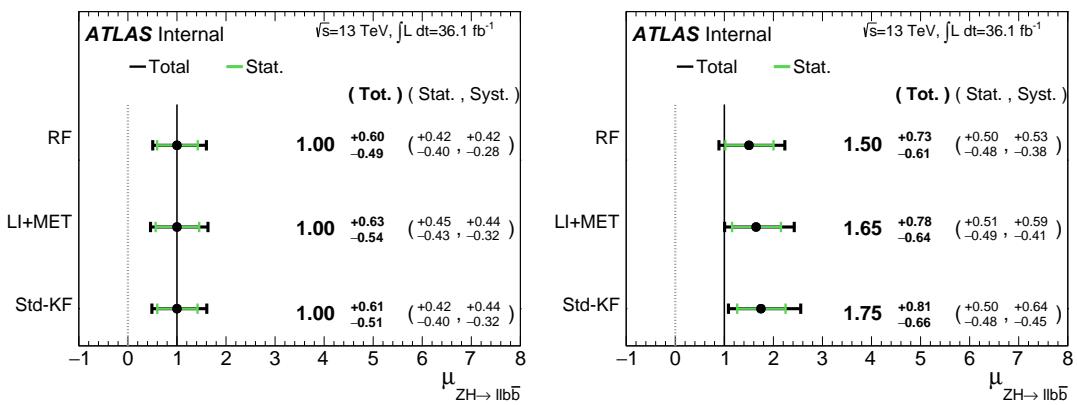


Figure 8.1: μ summary plots for the standard, LI, and RF variable sets. The Asimov case (with $\mu = 1$ by construction) is in (a), and $\hat{\mu}$ best fit values and error summary are in (b).

Studying the performance of the Lorentz Invariant and RestFrames variable sets at both a data statistics only context and with the full fit model in the $ZH \rightarrow \ell\ell b\bar{b}$ channel of the $VH(b\bar{b})$ analysis suggests that these variables may offer a potential method for better constraining systematic uncertainties in $VH(b\bar{b})$ searches as more orthogonal bases in describing the information in collision events.

₁₇₇₂ The marginally worse performance of the LI and RF variables (7.9% and 6.9%, respectively) with
₁₇₇₃ respect to the standard variable at a stats only level illustrates that neither variable set has greater
₁₇₇₄ intrinsic descriptive power in the absence of systematics in this closed final state. Hence, any gains
₁₇₇₅ from either of these variable sets in a full fit come from improved treatment of systematic uncertain-
₁₇₇₆ ties.

₁₇₇₇ With full systematics, the LI variable set narrows the sensitivity gap somewhat, with lower signif-
₁₇₇₈ icances by 6.7% (1.7%) on expected fits to Asimov (data) and by 2.8% on observed significances. The
₁₇₇₉ RF variable set outperforms the standard set in expected fits with 3.5% (3.4%) higher significance
₁₇₈₀ on Asimov (data), but has an 8.6% lower observed significance, though the observed significances
₁₇₈₁ should be viewed in the context of observed $\hat{\mu}$ values.

₁₇₈₂ Moreover, the LI and RF variable sets generally perform better in the context of the error on μ .
₁₇₈₃ The LI fit is comparable to the standard set on Asimov data and has a 7.5% lower total systematics
₁₇₈₄ error on $\hat{\mu}$ on observed data, while the RF fit is lower in both cases, with systematics error being
₁₇₈₅ 6.5% (16%) lower on Asimov (observed) data.

₁₇₈₆ These figures of merit suggest that both the LI and RF variables are more orthogonal than the
₁₇₈₇ standard variable set used in the fiducial analysis. Moreover, the RF variable set does seem to con-
₁₇₈₈ sistently perform better than the LI set. Furthermore, both variable sets have straightforward exten-
₁₇₈₉ sions to the one lepton channel in the $VH(b\bar{b})$ analysis, and the RF set has a straightforward ex-
₁₇₉₀ tension to the zero lepton channel as well. The magnitude of any gain from the more sophisticated
₁₇₉₁ treatment of E_T^{miss} in these extensions is beyond the scope of these studies, but the performance in
₁₇₉₂ this closed final state do suggest that there is some value to be had in these non-standard descriptions

¹⁷⁹³ independent of these considerations.

*If I have seen further, it is by standing on ye shoulders of
giants.*

Isaac Newton

9

1794

1795

Measurement Combinations

1796 WHILE THE DISCUSSION thus far has focused on improvements looking towards future in just
1797 the $ZH \rightarrow \ell\ell b\bar{b}$ channel, any actual result for SM $VH(b\bar{b})$ combines all channels and all avail-
1798 able datasets. Using additional channels at a given center of mass energy is straightforward since
1799 the fit model is designed with this combination in mind. Combining dataset results (known as

1800 “workspaces”) from different center of mass energies is not so simple an exercise since both the un-
1801 derlying physics (and its associated modeling) and the treatment of key experimental considerations,
1802 like flavor tagging, and their associated systematics change from dataset to dataset. A combined fit
1803 model must take these considerations into account, and the formulation of the fit model combin-
1804 ing the Run 1 ($\sqrt{s} = 7$ TeV with 4.7 fb^{-1} of data, and $\sqrt{s} = 8$ TeV with 20.3 fb^{-1} of data) and Run 2
1805 ($\sqrt{s} = 13$ TeV with) SM $VH(b\bar{b})$ results is the topic of Section 9.1. Its results, as reported in ³⁴, are
1806 given in 9.2.

1807 **9.1 THE COMBINED FIT MODEL**

1808 It is clear the signal strength parameter of interest should be fully correlated among the different
1809 datasets. Some signal modeling systematics were left unchanged from Run 1 through Run 2 and/or
1810 were designed to be explicitly correlated. Beyond these two special cases, it is not immediately clear
1811 what level of correlation should be imposed. The general methodology for settling upon a correla-
1812 tion scheme is as follows:

- 1813 1. Identify which NP categories have significant impacts on μ
- 1814 2. Of these NP’s, identify which have one-to-one correspondences or established correlation
1815 schemes among \sqrt{s} values
- 1816 3. Test whether correlation has a sizeable impact on expected fit quantities

1817 The only two sizeable experimental NP categories are jet energy scale (JES) and flavor tagging
1818 systematics. Correlation schemes of varying degrees of completeness exist for these categories, so ex-
1819 plicit NP correlations can be tested for these two categories. As these studies were conducted before

1820 unblinding, “sizeable impact” was judged by comparing fit results (sensitivities, pull comparisons,
1821 and breakdowns) on combined workspaces using the unblinded and public Run 1 $\mu = 0.51$ result
1822 and Asimov data for the Run 2 result. These are treated in Sections 9.1.1 and 9.1.2.

1823 Modeling systematics require a slightly different treatment, and are explored in 9.1.3.

1824 As a general note when looking at pull comparison plots for combined workspaces, the error bars
1825 in these plots are calculated using a simultaneous HESSE matrix inversion, which can fail to give
1826 sensible values for high dimensional models (the combined workspaces have well over 500 NP’s).

1827 This is not true of the nuisance parameter ranking plots, which use a MINOS based approach to test
1828 the effect of each NP individually. This is much slower but much more rigorous, which is why only
1829 ranking plots appear outside of supporting material and pull comparisons are considered “diagnos-
1830 tic” plots.

1831 9.1.1 JET ENERGY SCALE SYSTEMATICS

1832 Fortunately for the case of jet energy scale systematics, the JetEtMiss group provides two recom-
1833 mended “strong” and “weak” correlation schemes between Run 1 and Run 2. These were used as a
1834 point of departure for the JES combination correlation scheme. However, the JES NP’s in both the
1835 Run 1 and Run 2 workspaces are a reduced set of NP’s, with some 56 (75) NP’s reduced to 6 (8) for
1836 Run 1 (2). In order to restore the full set of JES NP’s, the effective NP’s in each workspace are un-
1837 folded using maps detailing the linear combinations of unfolded NP’s that form the effective NP’s.

The linear combinations used to unfold the effective JES NP's were calculated as follows:

$$NP_{i,eff} = \frac{\sum_j A_{ij} |NP_{j,unf}| NP_{j,unf}}{\sqrt{\sum_j A_{ij}^2 |NP_{j,unf}|^2}} \quad (9.1)$$

1839 where *eff* and *unf* are for effective and unfolded NP's, respectively, the A_{ij} 's are scalar coefficients

1840 taken from raw maps, and $|NP_{j,unf}|$ are the amplitudes of the unfolded NP's. The raw A_{ij} and
1841 scaled maps for Run 1 and Run 2 may be found in Figure 9.1.

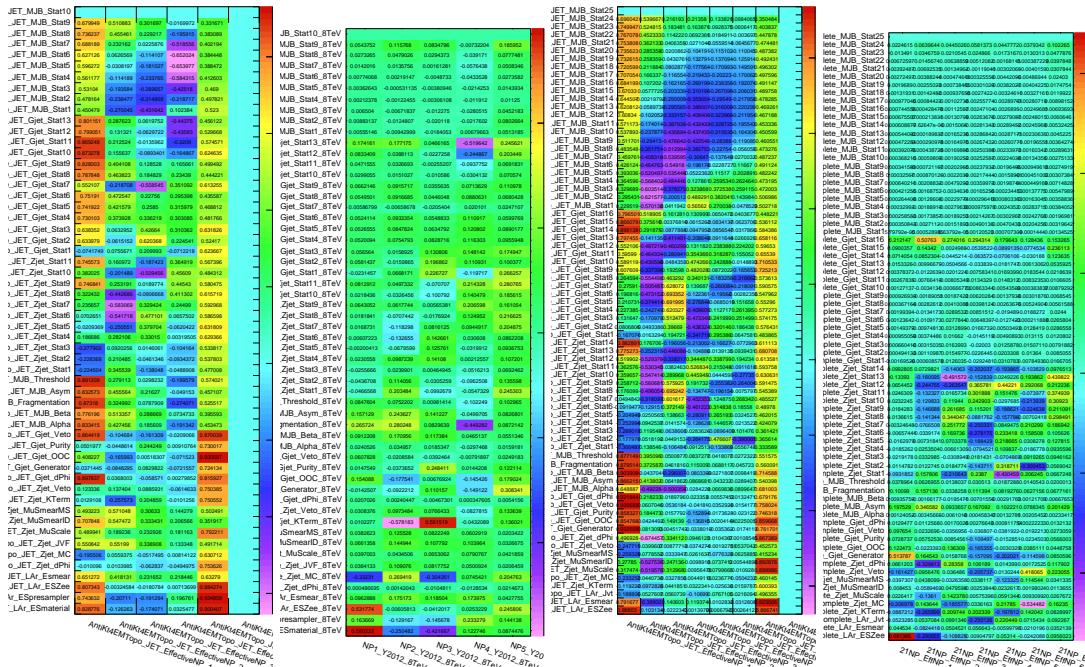


Figure 9.1: The raw and scaled coefficients for unfolding Run 1 (a and b) and Run 2 (c and d), respectively

Unfolding was found to have very little effect on both expected sensitivities and errors, as can be

1843 seen in Tables ?? and ??.

	R ₁ Unfold	R ₁ Eff	R ₂ Unfold	R ₂ Eff	Comb Unfold	Comb Eff
Exp. Sig.	2.604	2.606	3.014	3.014	4.005	3.998
Obs. Sig.	1.369	1.374	3.53	3.53	3.581	3.571
Exp. Limit	0.755 ^{+0.296} _{-0.211}	0.755 ^{+0.296} _{-0.211}	0.732 ^{+0.287} _{-0.205}	0.732 ^{+0.287} _{-0.205}	0.512 ^{+0.201} _{-0.143}	0.51 ^{+0.2} _{-0.143}
Obs. Limit	1.21	1.21	1.94	1.94	1.36	1.37

Table 9.1: Expected and observed sensitivities for Run 1, Run 2, and combined workspaces with effective and unfolded JES NP's.

	R ₁ Unfold	R ₁ Eff
$ \Delta\hat{\mu} $	0.0018	
$\hat{\mu}$	0.5064	0.5082
Total	+0.400 / -0.373	+0.401 / -0.373
DataStat	+0.312 / -0.301	+0.312 / -0.301
FullSyst	+0.250 / -0.220	+0.251 / -0.220
Jets	+0.060 / -0.051	+0.060 / -0.052
BTag	+0.094 / -0.079	+0.095 / -0.079
	+0.119 / -0.106	+0.119 / -0.106
	+0.076 / -0.076	+0.077 / -0.076

Table 9.2: Error on signal strength breakdowns for Run 1 workspaces with effective and unfolded JES NP's.

	R ₁ Unfold	R ₁ Eff
R ₂ Unfold	R ₂ Eff	Comb Unfold
Comb Eff		
$ \Delta\hat{\mu} $	0.0	
$\hat{\mu}$	1.2051	1.2052
Total	+0.421 / -0.366	+0.421 / -0.366
DataStat	+0.239 / -0.234	+0.239 / -0.234
FullSyst	+0.346 / -0.282	+0.346 / -0.282
Jets	+0.066 / -0.047	+0.066 / -0.047
BTag	+0.119 / -0.106	+0.119 / -0.106

Table 9.3: Error on signal strength breakdowns for Run 2 workspaces with effective and unfolded JES NP's.

	Comb Unfold	Comb Eff
$ \Delta\hat{\mu} $	0.0006	
$\hat{\mu}$	0.8992	0.8985
Total	+0.278 / -0.261	+0.278 / -0.261
DataStat	+0.185 / -0.181	+0.185 / -0.181
FullSyst	+0.208 / -0.187	+0.208 / -0.188
Jets	+0.040 / -0.044	+0.041 / -0.036
BTag	+0.076 / -0.076	+0.077 / -0.076

Table 9.4: Error on signal strength breakdowns for combined workspaces with effective and unfolded JES NP's.

1844 It was also found that fit sensitivities and breakdowns were similarly indifferent to the use of
1845 either the strong or weak JES correlation schemes, as shown in Tables 9.5 and 9.6.

	JES Weak Unfold	JES Weak Eff	JES Strong Unfold	JES Strong Eff
Exp. Sig.	3.57	3.57	3.59	3.59
Exp. Limit	0.493 ^{+0.193} _{-0.138}	0.494 ^{+0.193} _{-0.138}	0.493 ^{+0.193} _{-0.138}	0.493 ^{+0.193} _{-0.138}

Table 9.5: Expected sensitivities for both effective and unfolded combined workspaces using the strong and weak JES correlation schemes.

	Comb Unfold	Comb Eff	Strong Unfold	Strong Eff
$\Delta\hat{\mu}$	0.0009		0.0025	
Total	+0.269 -0.254	+0.27 -0.255	+0.27 -0.255	+0.27 -0.255
DataStat	+0.181 -0.177	+0.181 -0.177	+0.181 -0.177	+0.181 -0.178
FullSyst	+0.199 -0.183	+0.2 -0.183	+0.2 -0.183	+0.201 -0.183
Jets	+0.0387 -0.032	+0.041 -0.0337	+0.0425 -0.0329	+0.0432 -0.0338
BTag	+0.0975 -0.0933	+0.098 -0.0936	+0.0979 -0.0935	+0.098 -0.0936

Table 9.6: Error on signal strength breakdowns for both effective and unfolded combined workspaces using the strong and weak JES correlation schemes.

1846 Comparisons of top ranked nuisance parameters in Figures 9.2–9.4 and for the complete JES pull

comparisons in Figures 9.5–9.8

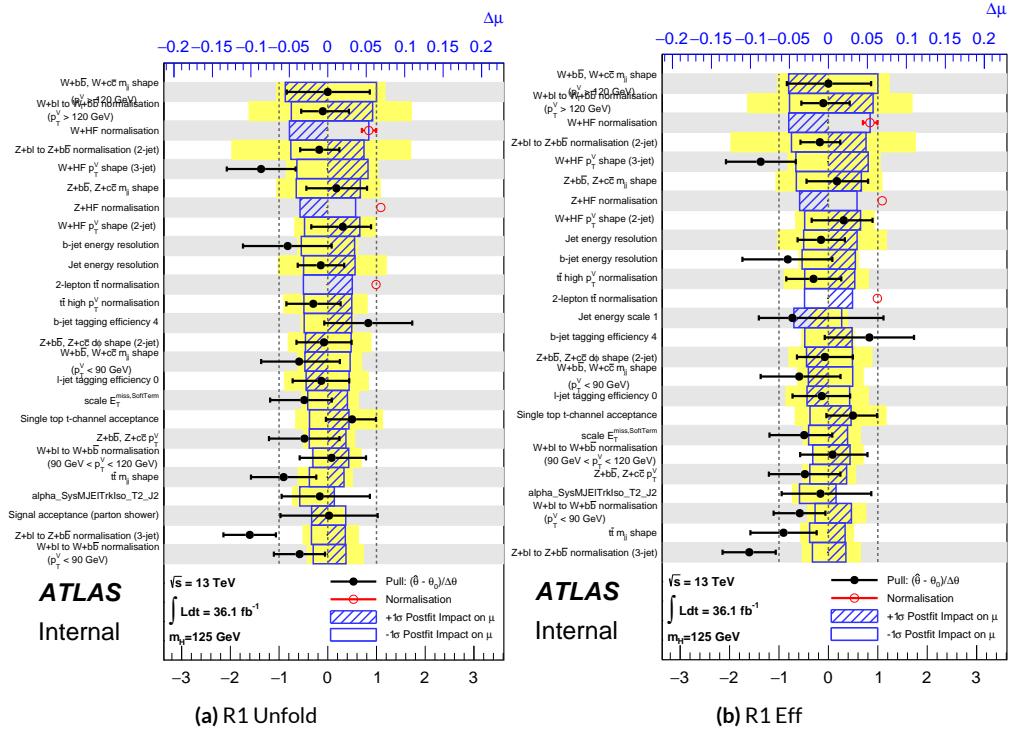


Figure 9.2: Ranks for the effective and unfolded JES NP Run1 combined workspaces.

1847

1848 As a result of these studies, the weak JES correlation scheme with uncorrelated effective JES NP's

1849 (i.e. just the b -jet energy scale NP) has been chosen as the treatment of JES in the Run 1 + Run 2

1850 combined fit.

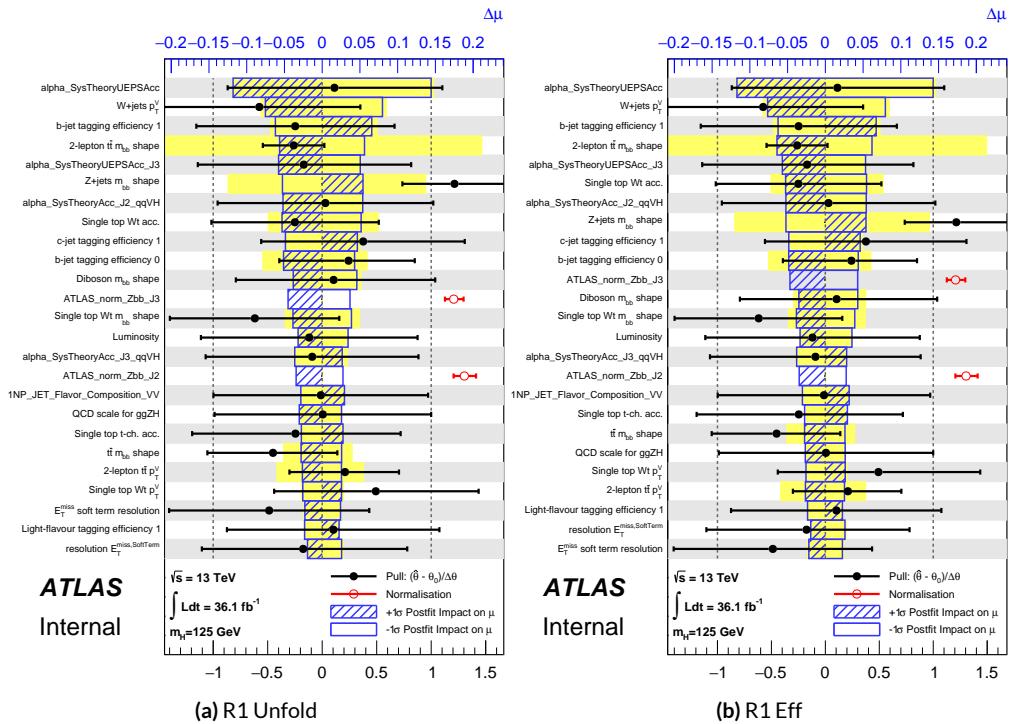


Figure 9.3: Ranks for the effective and unfolded JES NP Run2 combined workspaces.

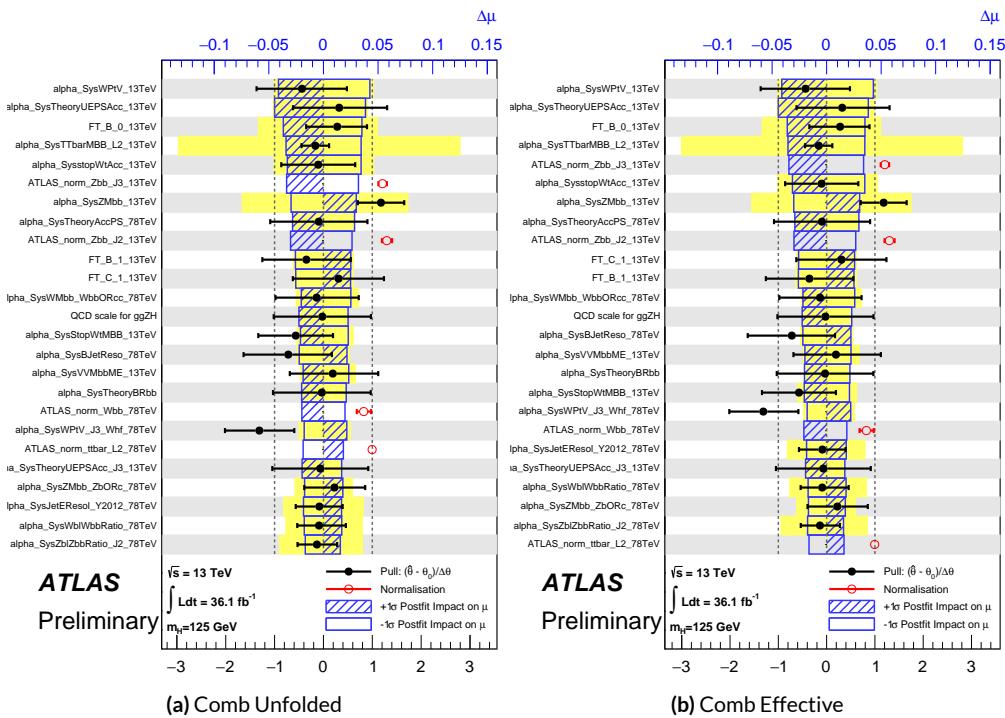


Figure 9.4: Ranks for the effective and unfolded JES NP Run1+Run2 combined workspaces.

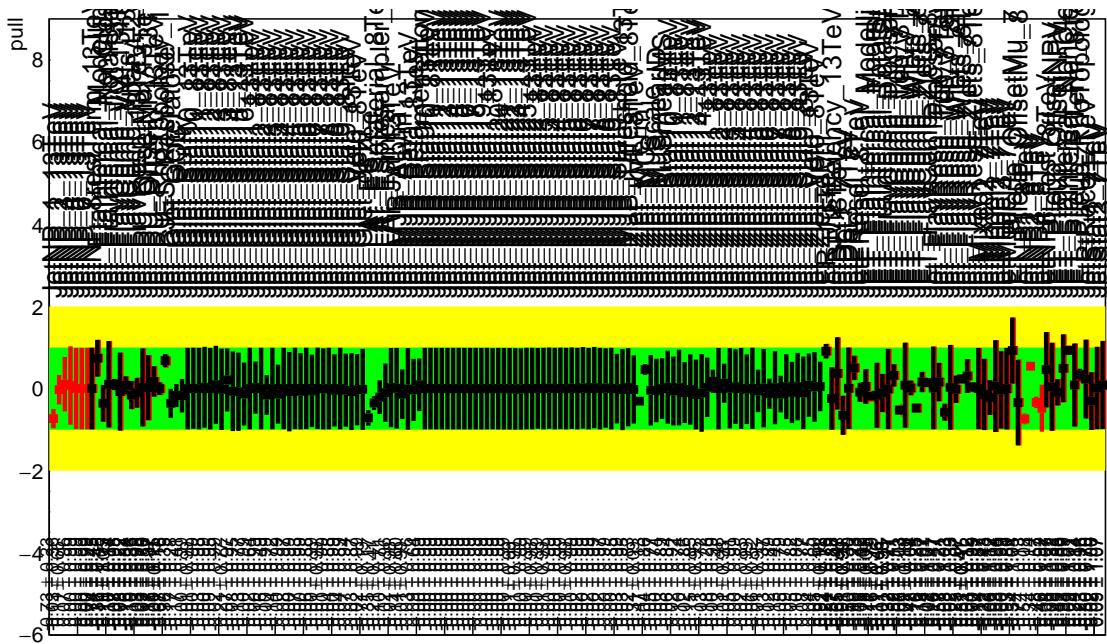


Figure 9.5: Pull Comparisons: jesu---Jet Comb Unfold, Comb Eff, Strong Unfold, Strong Eff

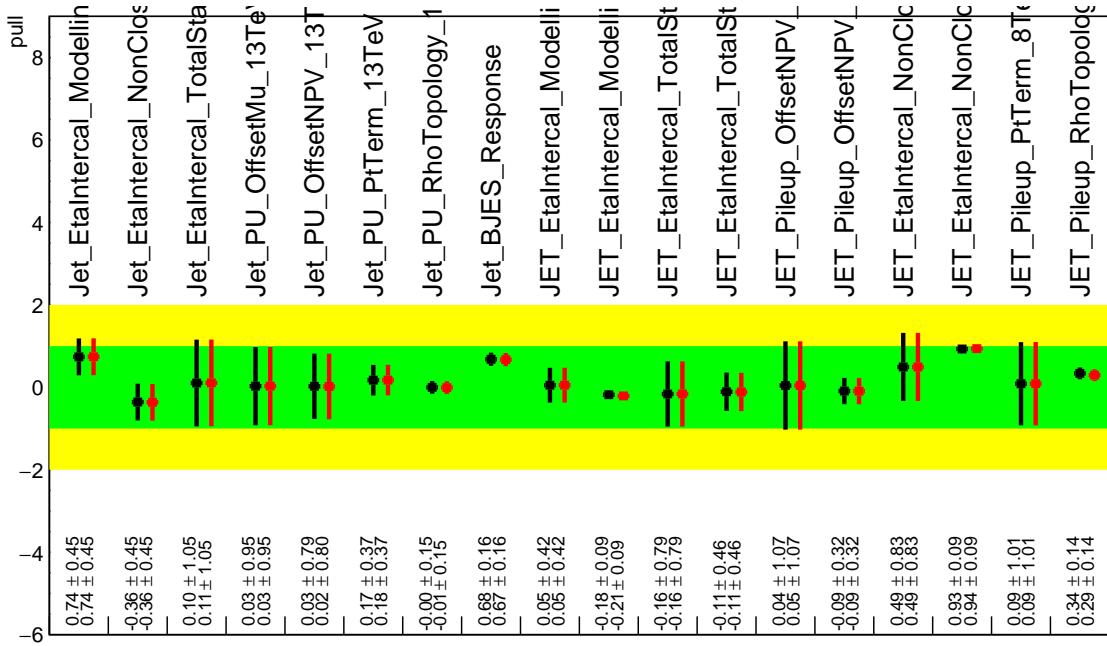


Figure 9.6: Pull Comparisons: jesu---JetMatched Comb Unfold, Comb Eff, Strong Unfold, Strong Eff

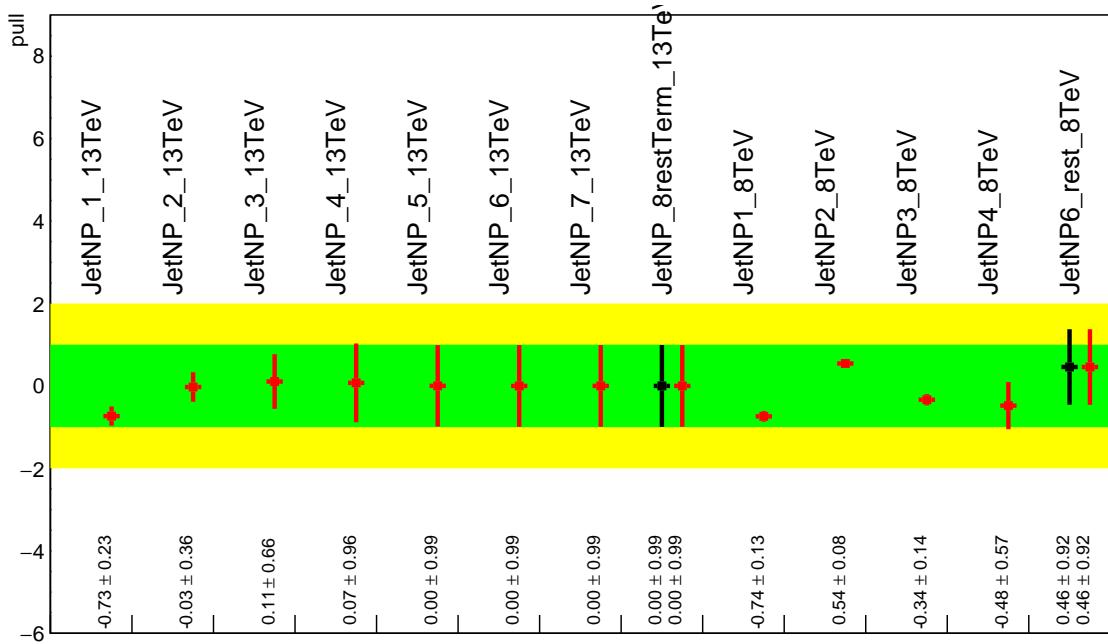


Figure 9.7: Pull Comparisons: jesu---JetEff Comb Unfold, **Comb Eff**, Strong Unfold, **Strong Eff**

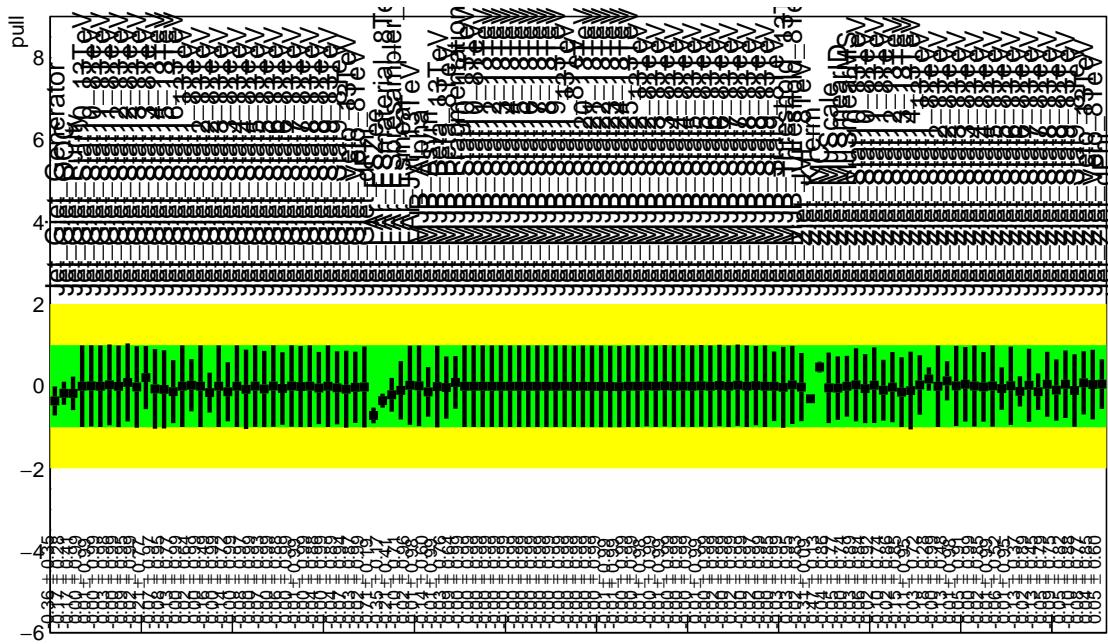


Figure 9.8: Pull Comparisons: jesu---JetUnfold Comb Unfold, **Comb Eff**, Strong Unfold, **Strong Eff**

1851 9.1.2 FLAVOR TAGGING

1852 Unfortunately, the ATLAS Flavor Tagging group did not provide any recommendations for cor-
1853 relating Run 1 and Run 2 NP's, though given the high ranking of these NP's in the Run 2, result,
1854 some studies at least were crucial. Nevertheless, great improvements and changes to the treatment
1855 of flavor tagging between Run 1 and Run 2 does weaken the argument for any strong flavor tagging
1856 correlation scheme.

1857 Given that c -tagging changed significantly between Run 1 and Run 2 and that light tagging NP's
1858 are very lowly ranked, these sets of NP's are left uncorrelated. Moreover, the change in the physical
1859 meaning of the effective b -tagging NP's means a full correlation of such NP's (insomuch as they exist
1860 in each result) is one of limited utility. Hence, it was decided to leave flavor tagging NP's uncorre-
1861 lated. However, since the meaning of the leading b -tagging NP's is approximately constant across
1862 years and since Run 2 b -tagging NP's are very highly ranked in both the Run 2 only and combined
1863 fits, tests correlating these NP's were conducted, the results of which can be seen below. It should be
1864 noted that the leading B NP at 8 TeV, `SysBTagB0Effic_Y2012_8TeV`, has an opposite effect on $t\bar{t}$
1865 normalization than the 7 and 13 TeV NP's, and so must be flipped using a similar strategy as for JES
1866 unfolding. Initial studies of flavor tagging correlations did not flip this NP, and so results for this
1867 scheme (labeled "Bo 8TeV Not Flipped") have also been included for comparison.

1868 It is clear from these results that correlating the leading effective Eigen NP associated with b 's can
1869 have a noticeable effect on final fit results and that the 8 TeV Bo NP is the most important compo-
1870 nent of a combined Bo NP. It is also not so surprising that the 8 TeV result should drive the com-

	Comb Eff	BTag Bo	Bo 8TeV Flipped
Exp. Sig.	3.998	4.127	3.921
Obs. Sig.	3.571	3.859	3.418
Exp. Limit	$0.51^{+0.2}_{-0.143}$	$0.5^{+0.196}_{-0.14}$	$0.517^{+0.202}_{-0.144}$
Obs. Limit	1.37	1.41	1.35

Table 9.7: Expected and observed sensitivities for a combination featuring the weak JES scheme, combination with the weak JES scheme + leading b NP's correlated, and the b correlation with the 8 TeV NP with sign unflipped.

	Comb Eff	BTag Bo	Bo 8TeV Not Flipped
$ \Delta\hat{\mu} $	—	0.0446	0.0268
$\hat{\mu}$	0.8985	0.9431	0.8717
Total	+0.278 / -0.261	+0.275 / -0.256	+0.282 / -0.263
DataStat	+0.185 / -0.181	+0.180 / -0.177	+0.189 / -0.186
FullSyst	+0.208 / -0.188	+0.207 / -0.186	+0.209 / -0.186
BTag	+0.077 / -0.076	+0.071 / -0.068	+0.079 / -0.075
BTag b	+0.062 / -0.059	+0.055 / -0.049	+0.064 / -0.060

Table 9.8: Breakdowns of the impact of different NP sets on total error on \hat{m}_u for a combination featuring the weak JES scheme and a combination with the weak JES scheme + leading b NP's correlated.

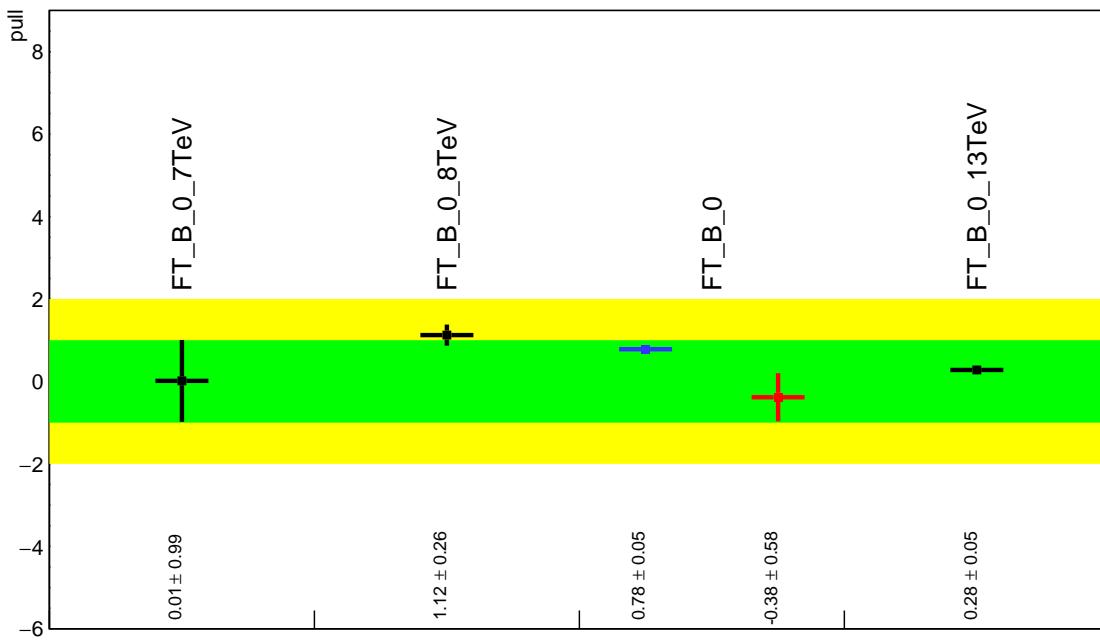


Figure 9.9: Pull Comparisons: btag-b---BTagBO Comb Eff, BTag BO

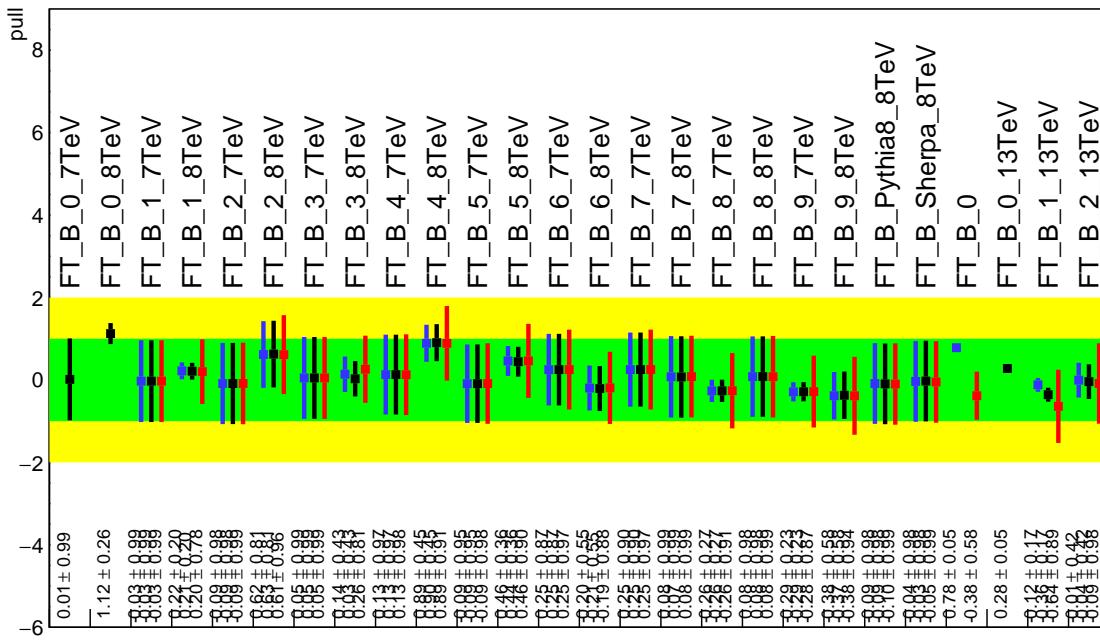


Figure 9.10: Pull Comparisons: btag-b---BTagB Comb Eff, BTag BO

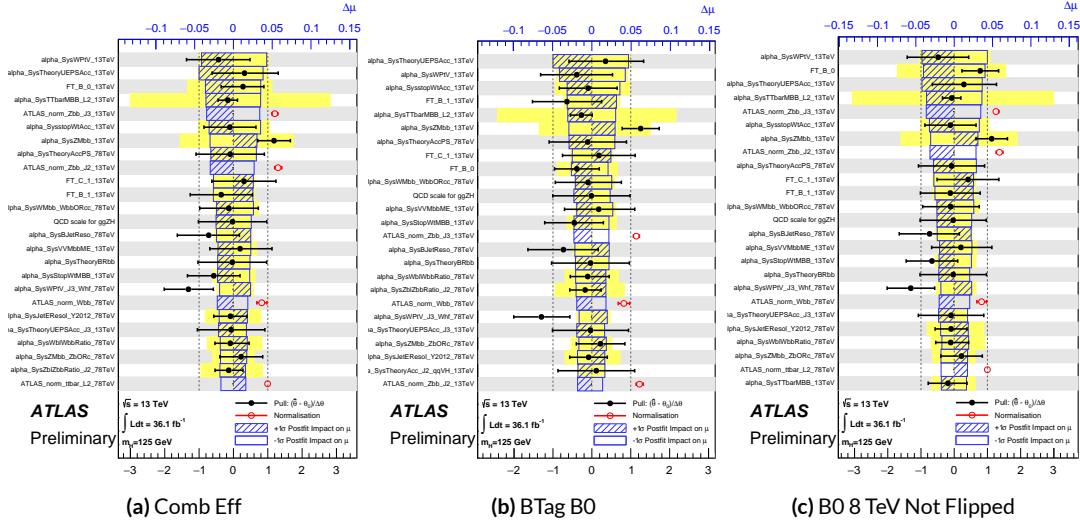


Figure 9.11: NP rankings for a combination featuring the weak JES scheme and a combination with the weak JES scheme + leading b NP's correlated.

1871 bined nuisance parameter since it is the only result to make use of both pseudocontinuous tagging-
 1872 based and b -tag regions into the final fit, implicitly yielding much more information about b 's. The
 1873 13 TeV fit has neither of these regions. What is less clear is whether there are sufficient grounds for
 1874 doing so (i.e. does the correspondence of these NP's across years warrant a full correlation). While
 1875 there are no current plans to do so, this matter warrants careful scrutiny if Run 1 is to be included in
 1876 future results.

1877 9.1.3 MODELING SYSTEMATICS

1878 Another principal systematic category is modeling uncertainties. The effect of correlating groups
 1879 of systematics was estimated using the same strategy employed by the ATLAS/CMS SM $VH(b\bar{b})$
 1880 combination for Run 1. This extrapolation can be used to estimate the impact of correlations on
 1881 the estimated signal strength, the total error on the signal strength, and the χ^2 of the result. The

¹⁸⁸² impact of such correlations is no more than a few percent effect, as the following tables demonstrate,
¹⁸⁸³ beginning with the category with the greatest shift, W+jets modeling, in Table ??.

	$ \Delta\mu $	σ	$ \Delta\sigma $	χ^2
$\rho=-1$	0.0024	0.2448	0.011 (4.3%)	0.95
$\rho=-0.6$	0.0015	0.2493	0.00654 (2.55%)	0.9804
$\rho=-0.3$	0.0008	0.2526	0.00325 (1.27%)	1.0045
$\rho=0$	—	0.2558	—	1.0298
$\rho=0.3$	0.0008	0.259	0.0032 (1.25%)	1.0564
$\rho=0.6$	0.0017	0.2622	0.00636 (2.49%)	1.0844
$\rho=1$	0.0029	0.2664	0.0105 (4.11%)	1.1242

Table 9.9: Run 1 + Run 2 W+jets modeling correlation projections

¹⁸⁸⁴ 9.1.4 FINAL CORRELATION SCHEME

¹⁸⁸⁵ The final Run 1 + Run 2 correlation scheme is shown in Table 9.10. As detailed above, neither JES
¹⁸⁸⁶ nor modeling systematics had any demonstrable effect on combined fit results. Hence, only signal
¹⁸⁸⁷ NP's and the b -jet energy scale are correlated. While the effect of flavor tagging correlations is less
¹⁸⁸⁸ clear, the result physical arguments for correlation are less strong; the size of effect was discovered
¹⁸⁸⁹ rather late in the analysis process; and has no nuisance parameter unfolding maps exist for flavor
¹⁸⁹⁰ tagging as they do for JES, so it was decided to leave these uncorrelated as well.

7 TeV NP	8 TeV NP	13 TeV NP
	ATLAS_BR_bb	SysTheoryBRbb
	SysTheoryQCDscale_ggZH	SysTheoryQCDscale_ggZH
	SysTheoryQCDscale_qqVH	SysTheoryQCDscale_qqVH
—	SysTheoryPDF_ggZH_8TeV	SysTheoryPDF_ggZH
—	SysTheoryPDF_qqVH_8TeV	SysTheoryPDF_qqVH
—	SysTheoryVHPt_8TeV	SysVHNLOEWK
SysJetFlavB_7TeV	SysJetFlavB_8TeV	SysJET_21NP_JET_BJES_Response

Table 9.10: A summary of correlated nuisance parameters among the 7, 8, and 13 TeV datasets.

1891 **9.2 COMBINED FIT RESULTS**

1892 **9.2.1 COMBINED FIT MODEL VALIDATION**

1893 Before moving onto the final results, we present the rest of the validation for the Run 1 + Run 2
1894 combined fits, beginning with impacts of ranked individual nuisance parameters in Figure 9.12 and
1895 for all nuisance parameter categories in Table ???. Both of these sets of results point to the most im-
1896 portant nuisance parameters being signal systematics, b -tagging, and V +jets modeling systematics,
1897 with Run 2 NP's generally being higher ranked. That some NP's are strongly pulled is not unusual
1898 as the fit model has so many NP's; V +jets modeling in particular has been historically difficult.

1899 In addition to looking at the behaviors of nuisance parameters to gauge fit model performance
1900 and stability, fits are conducted using multiple parameters of interest. Typical divisions are Run 1
1901 vs. Run 2, lepton channels, and WH vs ZH . As mentioned in Chapter 7, the profile likelihood test
1902 statistic given in Equation 7.2 is, in the limit of large sample statistics, a χ^2 distribution with degrees
1903 of freedom equal to the number of nuisance parameters plus nuisance parameters. Thus, changing
1904 the number of interest parameters and leaving the rest of the fit model unchanged means that the

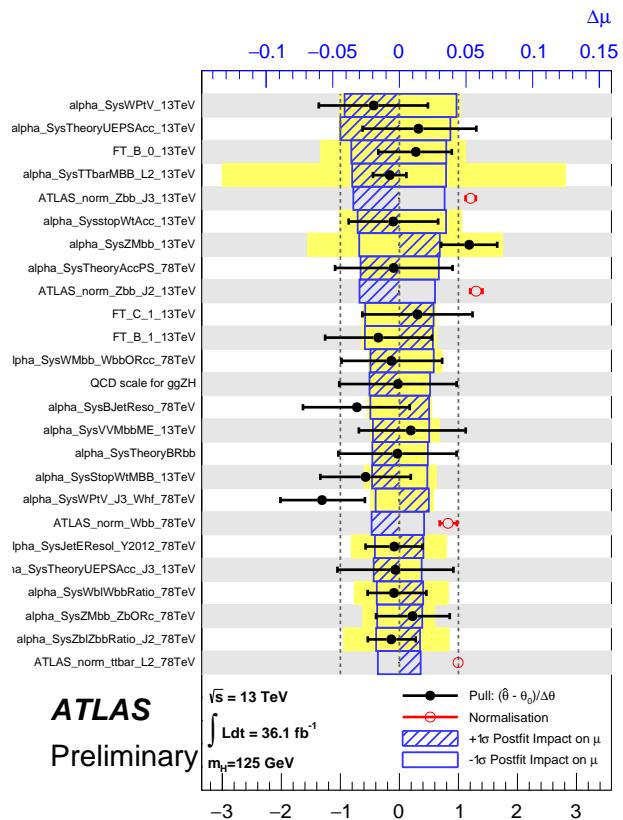


Figure 9.12: Ranked nuisance parameters for the Run1+Run2 combination.

Total	+0.278 / -0.261
DataStat	+0.185 / -0.181
FullSyst	+0.208 / -0.188
Floating normalizations	+0.055 / -0.056
All normalizations	+0.068 / -0.069
All but normalizations	+0.192 / -0.172
Jets, MET	+0.046 / -0.040
Jets	+0.041 / -0.036
MET	+0.023 / -0.018
BTag	+0.077 / -0.076
BTag b	+0.062 / -0.059
BTag c	+0.033 / -0.032
BTag light	+0.028 / -0.028
Leptons	+0.008 / -0.008
Luminosity	+0.026 / -0.014
Diboson	+0.030 / -0.027
Model Zjets	+0.049 / -0.050
Zjets flt. norm.	+0.032 / -0.040
Model Wjets	+0.082 / -0.083
Wjets flt. norm.	+0.031 / -0.027
Model ttbar	+0.047 / -0.046
ttbar flt. norm.	+0.025 / -0.026
Model Single Top	+0.047 / -0.045
Model Multi Jet	+0.027 / -0.038
Signal Systematics	+0.098 / -0.052
MC stat	+0.080 / -0.084

Table 9.11: Summary of the impact of different nuisance parameter categories on the total error on $\hat{\mu}$ for the combined Run1+Run2 fit.

1905 difference between the nominal fit and a fit with more parameters of interest ought to also be dis-
 1906 tributed as a χ^2 distribution with degrees of freedom equivalent to the number of extra parameters
 1907 of interest. This difference can then be interpreted as a compatibility between the two results using
 1908 the standard tables for this distribution, giving another gauge of fit performance. These are shown
 1909 in Table 9.12.

Fit	Compatibility
Leptons (3 POI)	1.49%
WH/ZH (2 POI)	34.2%
Run 1/Run 2 (2 POI)	20.8%
Run 1/Run 2 \times Leptons (6 POI)	7.10%
Run 1/Run 2 \times WH/ZH (4 POI)	34.6%

Table 9.12: Summary of multiple POI compatabilities. The well-known Run 1 7 TeV 0-lepton deficit is responsible for the low compatibility with the 6 and 3 POI fits.

1910 The low compatabilities associated with treating the lepton channels as separate parameters of
 1911 interest are a symptom of the low signal strengths associated with the Run 1 0-lepton channel, in par-
 1912 ticular the 7 TeV result. Given the relatively small amount of data associated with the 7 TeV result,
 1913 this should not be a cause for alarm. Signal strength summary plots for the fits treating Run 1 and
 1914 Run 2 separately are shown in Figures 9.13-9.15, where the effect of the Run 1 parameters can be seen
 1915 graphically.

1916 9.2.2 FINAL RESULTS

1917 The combined results yields an observed (expected) significance of 3.57 (4.00) and an observed (ex-
 1918 pected) limit of 1.37 ($0.510^{+0.200}_{-0.143}$), with a signal strength of $\hat{\mu} = 0.898^{+0.278}_{-0.261}$.

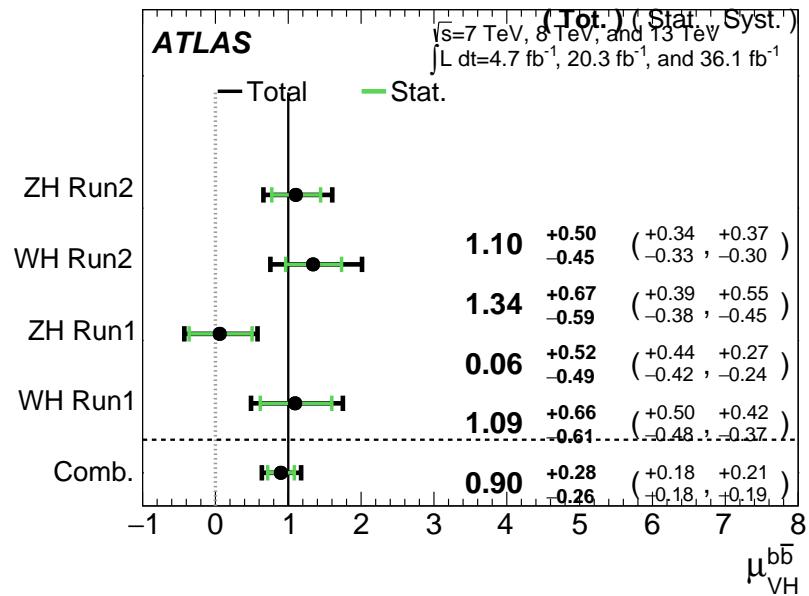


Figure 9.13: $\hat{\mu}$ summary plot for a four parameter of interest fit.

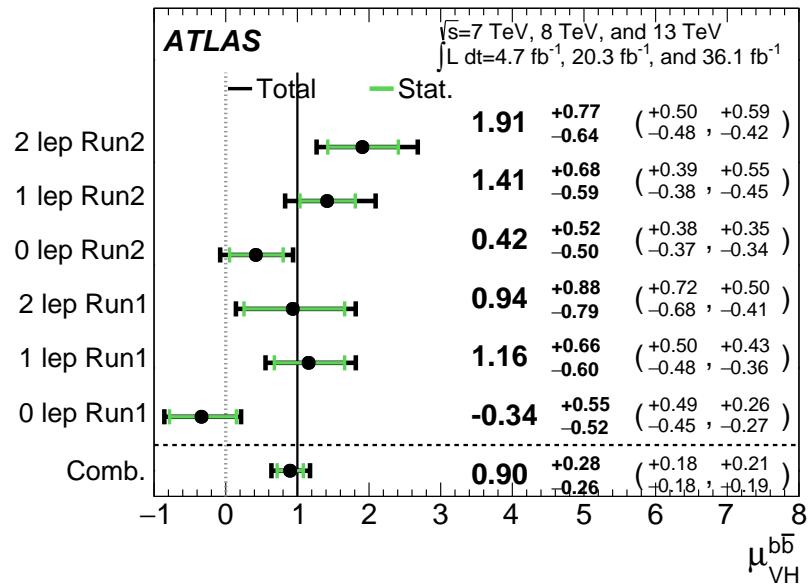


Figure 9.14: $\hat{\mu}$ summary plot for a six parameter of interest fit.

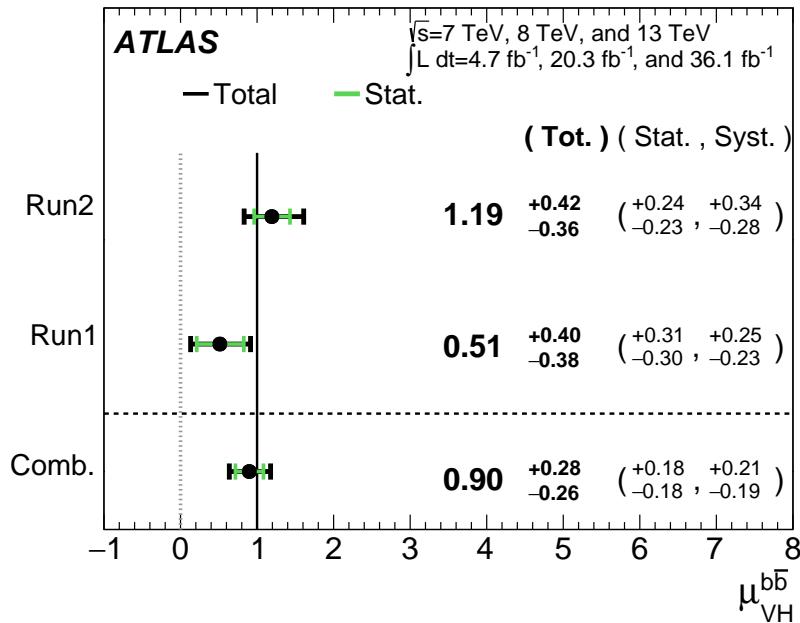


Figure 9.15: $\hat{\mu}$ summary plot for a two parameter of interest (Run 1 and Run 2) values.

1919 The two and three parameter of interest fit signal strength summary plots, as well as a summary
 1920 of the historical values of the 7, 8, and 13 TeV results may be found in Figures 9.16-9.18. The main
 1921 results for Run 1, Run 2, and the combination may be found in Table 9.13. These results were collec-
 1922 tively noted as the first ever experimental evidence for SM $VH(b\bar{b})$ in³⁴.

Dataset	$\hat{\mu}$	Total Error in $\hat{\mu}$	Obs. (Exp.) Significance
Run 1	0.51	$+0.40 / -0.37$	1.4 (2.6)
Run 2	1.20	$+0.42 / -0.36$	3.54 (3.03)
Combined	0.90	$+0.28 / -0.26$	3.57 (4.00)

Table 9.13: A summary of main results for the Run 1, Run 2, and combined fits.

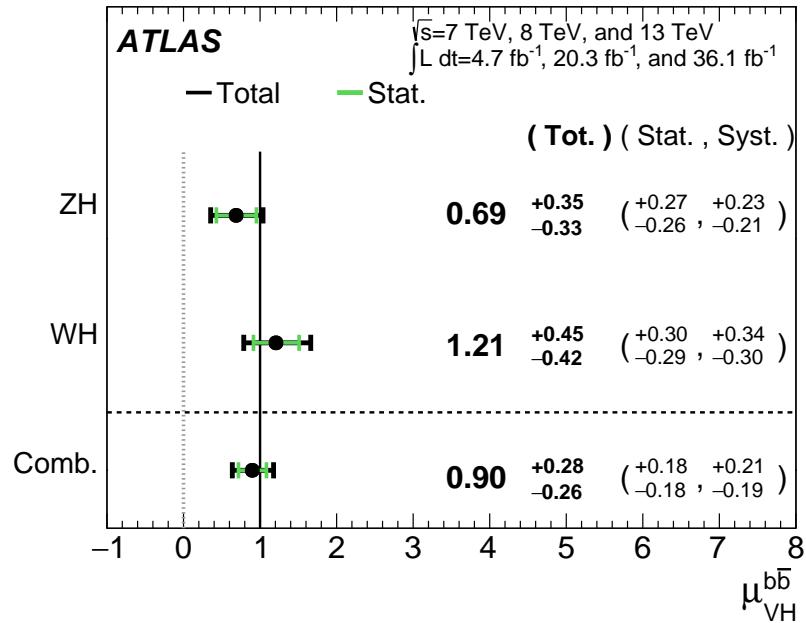


Figure 9.16: $\hat{\mu}$ summary plot for a two parameter of interest fit.

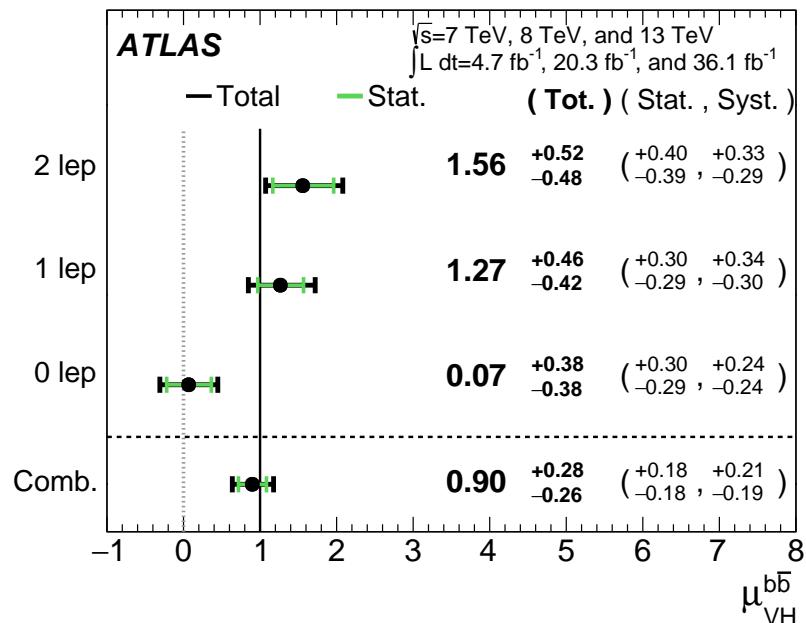


Figure 9.17: $\hat{\mu}$ summary plot for a three parameter of interest fit.

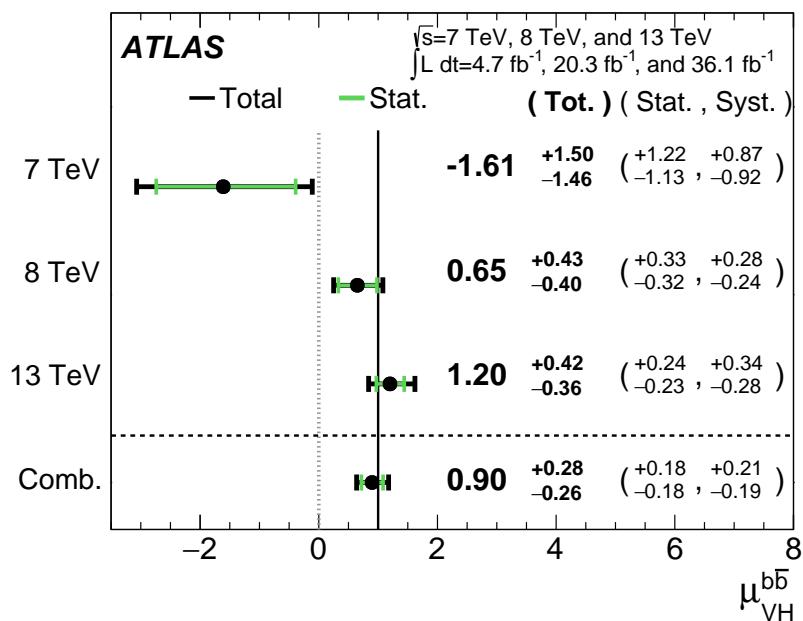


Figure 9.18: $\hat{\mu}$ summary plot for different \sqrt{s} values.

Vanitas vanitatum, omnis vanitas

Ecclesiastes 1:2

1923

10

1924

Closing Thoughts

1925 WITH BOTH THE LHC and ATLAS performing very well, it is only a matter of time before the
1926 evidence for SM $VH(b\bar{b})$ passes the 5 Gaussian standard deviation threshold necessary for discovery.
1927 Depending the latter two years of ATLAS data from Run 2 of the LHC (2017 and 2018), this may
1928 come less than a year after reports of first evidence and may not even require a combination with

¹⁹²⁹ Run 1 data.

¹⁹³⁰ It is entirely natural to ask, then, how essential the techniques and results described in this disser-
¹⁹³¹ tation will prove to be moving forward. Neither the LI/RF multivariate techniques nor combina-
¹⁹³² tion with Run 1 datasets and their accompanying low signal strength values are necessary for discov-
¹⁹³³ ery, and the latter may not even be essential to timely^{*} discovery of SM $VH(b\bar{b})$. Nevertheless, both
¹⁹³⁴ sets of results hold great potential as key parts of a concerted ensemble of efforts towards precision
¹⁹³⁵ Higgs physics.

¹⁹³⁶ With the final major center of mass energy increase at the energy frontier for the foreseeable fu-
¹⁹³⁷ ture behind us, it is becoming increasingly likely that any new fundamental physics at colliders will
¹⁹³⁸ require the use of results of systematics limited analyses. This is the regime where the techniques
¹⁹³⁹ described in this dissertation will be most useful.

¹⁹⁴⁰ As the LHC and its experiments undergo successive stages of upgrades and operate in evermore
¹⁹⁴¹ extreme environments, the statistical fit models used to describe LHC data will continue to evolve in
¹⁹⁴² complexity and diverge from their predecessors. The techniques described in Chapter 9 will become
¹⁹⁴³ increasingly more vital to producing the best physics results possible. The improvement in precision
¹⁹⁴⁴ from $1.20^{+0.24}_{-0.23}$ (stat.) $^{+0.34}_{-0.28}$ (syst.) to $0.90^{+0.18}_{-0.18}$ (stat.) $^{+0.21}_{-0.19}$ (syst.) is just the beginning.

¹⁹⁴⁵ The best methods for reduction of systematic uncertainties will naturally depend in part on the
¹⁹⁴⁶ state of the art for both fundamental physics process and detector modeling, but techniques that
¹⁹⁴⁷ can reduce systematic uncertainties independent of fit model, dataset, and physics process provide a
¹⁹⁴⁸ promising avenue forward. The improvements in systematic uncertainties using the Lorentz Invari-

*i.e. before or coincident with CMS

ant and RestFrames variable techniques in the $ZH \rightarrow \ell\ell b\bar{b}$ analysis, summarized in Table 10.1, show
 that a smarter and more orthogonal decomposition of information in a collision event provides ben-
 efits independent of any clever treatment of \vec{E}_T^{miss} (which both schemes also provide). Both tech-
 niques are readily extendible to other analysis channels, with the RestFrames concept demonstrating
 stronger performance and greater flexibility to nearly completely generic final states.

	Standard	LI	RF
$\hat{\mu}$	$1.75^{+0.24}_{-0.23}(\text{stat.})^{+0.34}_{-0.28}(\text{syst.})$	$1.65^{+0.24}_{-0.23}(\text{stat.})^{+0.34}_{-0.28}(\text{syst.})$	$1.50^{+0.24}_{-0.23}(\text{stat.})^{+0.34}_{-0.28}(\text{syst.})$
Asi. $\Delta err(\mu)$	—	$< 1\%, +4.6\%$	$-6.5\%, -2.2\%$
Obs. $\Delta err(\hat{\mu})$	—	$-7.5\%, -3.7\%$	$-16\%, -8.8\%$
Stat only sig.	4.78	4.39 (-7.9%)	4.44 (-6.9%)
Exp. (Asi.) sig.	2.06	1.92 (-6.7%)	2.13 (+3.5%)
Exp. (data) sig.	1.76	1.73 (-1.7%)	1.80 (+3.4%)
Obs. (data) sig.	2.87	2.79 (-2.8%)	2.62 (-8.6%)

Table 10.1: Summary of performance figures for the standard, LI, and RF variable sets. In the case of the latter two, % differences are given where relevant. Differences in errors on μ are on full systematics and total error, respectively.

Critical work remains to be done refining and extending the treatment of both the LI and RF
 techniques in $VH(b\bar{b})$ analyses and their fit models, and completely independent techniques, like
 the use of multiple event interpretations addressed in Appendix B promise further improvements
 still.

No one can say for certain what the future of the energy frontier of experimental particle physics
 may hold, but more nuanced treatments of the information in collision events born of meaningful
 physical insight are sure to light the way.

If it's stupid but it works, it isn't stupid.

Conventional Wisdom

A

1961

1962 Micromegas Trigger Processor Simulation

1963 IN ORDER TO PRESERVE key physics functionality by maintaining the ability to trigger on low p_T

1964 muons, the Phase I Upgrade to ATLAS includes a New Small Wheel (NSW) that will supply muon

1965 track segments to the Level 1 trigger. These NSW trigger segments will combine segments from the

1966 sTGC and Micromegas (MM) trigger processors (TP). This note will focus in particular on the algo-

¹⁹⁶⁷ rithm for the MMTP, described in detail with initial studies in³³. The goal of this note is to describe
¹⁹⁶⁸ the MMTP algorithm performance under a variety of algorithm settings with both nominal and
¹⁹⁶⁹ misaligned chamber positions, as well as addressing a number of performance issues.

¹⁹⁷⁰ This note is organized as follows: the algorithm and its outputs are briefly described in Section
¹⁹⁷¹ A.1; Monte Carlo samples used are in Section B.1; nominal algorithm performance and certain quan-
¹⁹⁷² tities of interest are described in Section A.3; algorithm performance under misalignment, misalign-
¹⁹⁷³ ment corrections, and corrected performance are shown in Section A.9; and conclusions are pre-
¹⁹⁷⁴ sented in Section A.24.

¹⁹⁷⁵ A.1 ALGORITHM OVERVIEW

¹⁹⁷⁶ The MMTP algorithm is shown schematically in Figure A.1, taken from³³, where a more detailed
¹⁹⁷⁷ description may be found. The algorithm begins by reading in hits, which are converted to slopes.
¹⁹⁷⁸ These slopes are calculated under the assumption that the hit originates from the IP; slopes calcu-
¹⁹⁷⁹ lated under this assumption are denoted by a superscript g for global in order to distinguish them
¹⁹⁸⁰ from local slopes calculated using only hits in the wedge. In the algorithm simulation, events are
¹⁹⁸¹ screened at truth level to make sure they pass certain requirements. The track's truth-level coor-
¹⁹⁸² dinates must place it with the wedge since some generated tracks do not reach the wedge. These
¹⁹⁸³ hits are stored in a buffer two bunch crossings (BCs) in time deep that separates the wedge into
¹⁹⁸⁴ so-called "slope-roads." If any given slope-road has sufficient hits to pass what is known as a coin-
¹⁹⁸⁵ cidence threshold, a fit proceeds. A coincidence threshold is a requirement for an event expressed as
¹⁹⁸⁶ $aX+bUV$, which means that an slope-road must have at least a hits in horizontal (X) planes and at

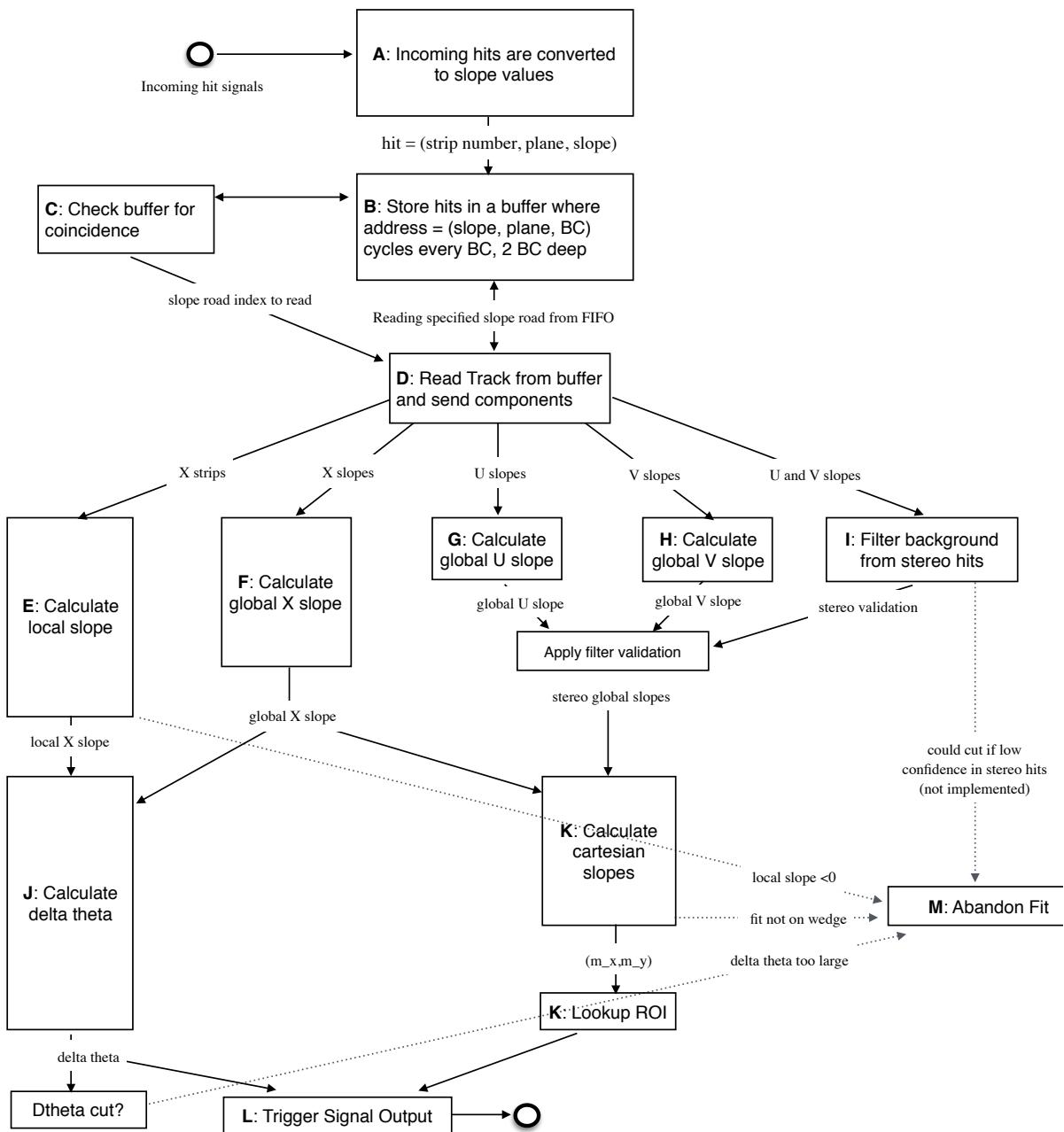


Figure A.1: A flow chart describing the algorithm steps, taken from ³³.

1987 least b hits in stereo (U or V (corresponding to positive and negative stereo rotations)) planes. For
 1988 coincidence thresholds with a $2X$ hit requirement there is the extra requirement that, in the case of
 1989 only $2X$ hits, one be on each quadruplet in order to ensure an adequate lever arm for the $\Delta\theta$ calcu-
 1990 lation. Note that less stringent (lower hit) coincidence thresholds are inclusive; i.e. a slope-road pass-
 1991 ing a $4X+4UV$ cut automatically passes $2X+1UV$. The coincidence threshold, size of the slope-roads
 1992 (denoted b), and the number of slope-roads into which each horizontal and stereo hits get written
 1993 centered upon their nominal value are configurable parameters of the algorithm.

1994 An individual hit's slope is calculated as shown in Equation A.1, where y_{base} is the local y coordi-
 1995 nate (orthogonal to the beamline and direction of the horizontal strips) of a station's base, w_{str} is the
 1996 strip pitch, n_{str} is the hit's strip number, and z_{plane} is the location of the hit's plane along the beam-
 1997 line.

$$M_{hit} = \frac{y}{z} = \frac{y_{base}}{z_{plane}} + \frac{w_{str}}{z_{plane}} \times n_{str} \quad (\text{A.1})$$

1998 In the fit, individual hit slopes in a slope-road are used to calculate global slopes associated with each
 1999 plane type, which are averages (e.g. M_X^{ℓ} for the average slope of horizontal planes). These in turn are
 2000 used to calculate the three composite slopes: slopes associated with the horizontal (m_x) and vertical
 2001 coordinates (m_y) and the local slope of hits in the horizontal planes (M_X^l), all of which are shown in
 2002 Equation A.4. Note that the expression for M_X^l differs but is equivalent to the expression given in³³.
 2003 This is due to a procedural change in the algorithm. The local X slope is expressed in³³ as:

$$M_X^{local} = A_k \sum_i y_i z_i - B_k \sum_i y_i, \quad B_k = \frac{1}{n} \sum_i z_i A_k = \bar{z} A_k \quad (\text{A.2})$$

2004 Procedurally, this entails doing the sums over y_i and $y_i z_i$, multiplying the sums by A_k , B_k , and then
 2005 subtracting both of these numbers, $\mathcal{O}(10^3)$, to get local slopes, $\mathcal{O}(10^{-1})$, while requiring precision
 2006 on these numbers on the order of $\mathcal{O}(10^{-3})$. This requires precision in the sums $\mathcal{O}(10^{-7})$, and with
 2007 32 bit fixed point numbers, there are deviations with respect to the floating point calculations at the
 2008 level of $\mathcal{O}(10^{-5})$, which is enough to introduce a significant bias in the $\Delta\theta$ calculation.

2009 In order to prevent these errors, we do the subtraction first

$$M_X^{local} = A_k \sum_i y_i z_i - B_k \sum_i y_i = A_k \sum_i (y_i z_i - y_i \bar{z}) = B_k \sum_i y_i \left(\frac{z_i}{\bar{z}} - 1 \right) \quad (\text{A.3})$$

2010 Thus, we change the order of operations and store $1/\bar{z}$ instead of A_k in addition to B_k . We also
 2011 change the units of y_i and z_i in the calculation by dividing the millimeter lengths by 8192.* With
 2012 these changes, a 32 bit fixed point based algorithm has essentially identical performance to that of an
 2013 algorithm based on the usual C++ 32 floating point numbers. Future work includes converting the
 2014 32 bit fixed point arithmetic to 16 bit where possible in the algorithm. While introducing 16 bit num-
 2015 bers uniformly might seem preferable, since simple 16-bit operations in the firmware can be done in
 2016 a single clock tick, and a larger number of bits increases the algorithm latency, some numbers in the
 2017 algorithm will require a larger number of bits, in particular in the local slope calculation, which is
 2018 the single calculation in the algorithm requiring the largest numeric range.

2019 In Equation A.4, θ_{st} is the stereo angle of 1.5 degrees; the sums are over relevant planes; \bar{z} is the
 2020 average position in z of the horizontal planes; and y_i and z_i in the local slope expression refer to the y

*Chosen since it is a perfect power of 2 and of order the length scale of z in millimeters

2021 and z coordinates of hits in X planes.

$$m_x = \frac{1}{2} \cot \theta_{st} (\mathcal{M}_U^g - \mathcal{M}_V^g), \quad m_y = \mathcal{M}_X^g, \quad M_X^l = \frac{\bar{z}}{\sum_i z_i^2 - 1/n (\sum_i z_i)^2} \sum_i y_i \left(\frac{z_i}{\bar{z}} - 1 \right) \quad (\text{A.4})$$

2022 From these composite slopes, the familiar expressions for the fit quantities θ (the zenith), ϕ (the az-
2023 imuth[†]), and $\Delta\theta$ (the difference in θ between the direction of the segment extrapolated back to the
2024 interaction point and its direction when entering the detector region; the following is an approxima-
2025 tion) may be calculated, as noted in³³:

$$\theta = \arctan \left(\sqrt{m_x^2 + m_y^2} \right), \quad \phi = \arctan \left(\frac{m_x}{m_y} \right), \quad \Delta\theta = \frac{M_X^l - \mathcal{M}_X^g}{1 + M_X^l \mathcal{M}_X^g} \quad (\text{A.5})$$

2026 Looking at Equations A.4 and A.5, the dependence of fit quantities on input hit information be-
2027 comes clear. $\Delta\theta$ relies exclusively on information from the horizontal (X) planes. Both θ and ϕ rely
2028 on both horizontal and stereo slope information. However, the sum in quadrature of m_x and m_y in
2029 the arctangent for θ means that θ is less sensitive to errors in stereo hit information than ϕ . Given
2030 that θ_{st} is small, $\cot \theta_{st}$ is large (~ 38), so m_x multiplies small differences in \mathcal{M}_U and \mathcal{M}_V , where m_y
2031 is simply an average over slopes. This means that while errors in horizontal hit information will af-
2032 fect all three fit quantities, comparable errors in stereo hits will have a proportionately larger effect
2033 on θ and particularly on ϕ . The $\Delta\theta$ cut after step J in Figure A.1 has been implemented, requiring
2034 all fits to have $|\Delta\theta| < 16$ mrad. This requirement ensures good quality fits but also slightly reduces

[†]Defined with respect to the center (y) axis and *not* the axis of the strips (x) as is sometimes typical, so a hit along the center of the wedge has $\phi = 0$

2035 algorithm efficiency.

2036 A.2 MONTE CARLO SAMPLES

2037 The Monte Carlo (MC) samples used for these studies were generated in Athena release 20.1.0.2 us-
2038 ing simulation layout ATLAS-R2-2015-01-01-00 with muon GeoModel override version MuonSpectrometer-
2039 R.07.00-NSW and modifications to have two modules per multiplet and xxuvuvxx geometry with a
2040 stereo angle of 1.5 degrees. Muons of a single p_T were generated around the nominal IP with a smear-
2041 ing of 50 mm along the beam line and 0.015 mm orthogonal to it; these muons were pointed toward
2042 a single, large sector of the NSW. Each event consists of one muon fired towards the single NSW
2043 wedge separated by effectively infinite time from other events.

2044 A.3 NOMINAL PERFORMANCE

2045 In order to evaluate algorithm performance, a number of quantities are evaluated, including the fit
2046 quantities θ , ϕ , and $\Delta\theta$ as well as algorithm efficiency. Unless otherwise stated, that algorithm is run
2047 with a 4X+4UV coincidence threshold, slope-road size of 0.0009, an X tolerance of two slope-roads
2048 (i.e. hits in horizontal planes are written into the two slope-roads closest to the hits' value), a UV
2049 tolerance of four slope-roads[‡], and a charge threshold requirement on hits of 1 (measured in units
2050 of electron charge) for a sample of 30 000 events with a muon p_T of 100 GeV. Samples were also
2051 generated for p_T values of 10 GeV, 20 GeV, 30 GeV, 50 GeV, and 200 GeV, which were used in some

[‡]The larger tolerance on stereo hits takes into account the particulars of the m_x calculation mentioned in Section A.1.

2052 of the following studies.

2053 A.4 FIT QUANTITIES

2054 In order to evaluate the performance of the algorithm's fit quantities θ , ϕ , and $\Delta\theta$, fit values are com-
2055 pared to truth-level MC values. The residual of the three fit quantities, $\theta_{fit} - \theta_{tru}$, $\phi_{fit} - \phi_{tru}$, and
2056 $\Delta\theta_{fit} - \Delta\theta_{tru}$, are recorded for every fitted track. The distributions of these quantities, in particular
2057 their biases and standard deviations, are then used to evaluate performance. In most cases, follow-
2058 ing³³, the mean and standard deviation of a 3σ Gaussian fit are quoted, as they capture the main
2059 features of the algorithm and generally behave like the raw mean and rms. Nevertheless, discussion
2060 of the raw quantities will be included when their behavior deviates markedly from that of the 3σ fit
2061 quantities.

2062 The truth-level quantities used in residual distribution are taken from information in the MC.

2063 These come directly from the MC for θ , ϕ , and $\Delta\theta$. These quantities, along with the geometry of
2064 the (large) wedge, are then in turn used to calculate truth-level values for any intermediate quantities
2065 used in the algorithm. $m_{x,tru}$, for instance, is given by $\tan \theta_{tru} \sin \phi_{tru}$.

2066 Residual distributions for fit quantities under the previously described default settings of the al-
2067 gorithm are shown in Figure A.2. Both the $\theta_{fit} - \theta_{tru}$ and $\Delta\theta_{fit} - \Delta\theta_{tru}$ distributions feature a
2068 mostly Gaussian shape with more pronounced tails. The mean bias for these distributions is negligi-
2069 ble at under one tenth of a milliradian, and the fitted (raw) rms values are 0.349 (0.614) mrad for θ
2070 and 1.03 (2.55) mrad for $\Delta\theta$. The case of the $\phi_{fit} - \phi_{tru}$ distribution is less straightforward, with both
2071 the shape and bias arising from the xxuvuvxx geometry and relatively large extent of one of the two

2072 η -stations, as explained in Appendix B of³⁰. The fitted (raw) rms for the ϕ distribution is 8.67 (16.6)
 2073 mrad.

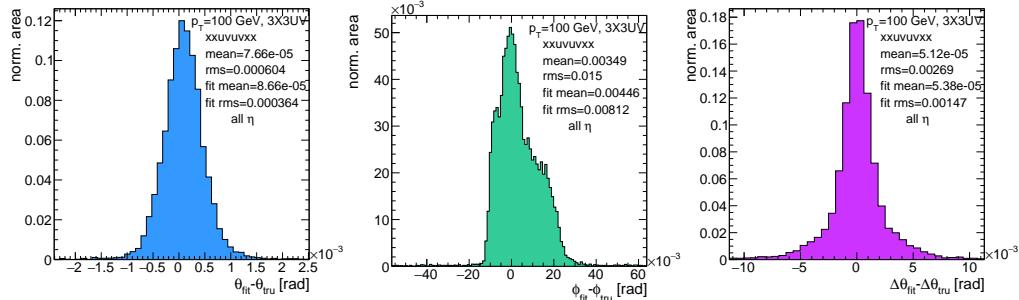


Figure A.2: Nominal residual plots; $\theta, \phi, \Delta\theta$ for $p_T = 100 \text{ GeV}$ muons

2074 Both increasing muon p_T and increasing muon η for a fixed p_T imply increasing muon energy. As
 2075 muons become more energetic, two effects compete in affecting the quality of fit. On the one hand,
 2076 higher energy muons are deflected less by the ATLAS magnetic field, which should tend to improve
 2077 the quality of the fit, since the fitted θ (upon which $\Delta\theta$ also relies) and ϕ values are calculated under
 2078 the infinite momentum muon (straight track) assumption. However, as muon energy increases, the
 2079 likelihood that the muon will create additional secondaries increases, which creates extra hits that
 2080 degrade the quality of the fit. While the geometry of the multiplet is such that there is very good res-
 2081 olution in the direction orthogonal to the horizontal strip direction, the shallow stereo angle of 1.5
 2082 degrees means that early hits caused by secondaries can have an outsize impact on m_x . $\Delta\theta$, which
 2083 does not rely upon stereo information should feel the effect of secondaries the least and benefit from
 2084 straighter tracks the most and hence benefit from higher muon energies; ϕ , relying upon stereo in-
 2085 formation the most, would be most susceptible to secondaries and benefit the least from straighter

tracks and hence least likely to benefit from higher muon energy; θ relies upon both horizontal and vertical slope information, though small errors are less likely to seriously affect the calculation, so the two effects are most likely to be in conflict for this fit quantity.

The interplay of these effects on the residual standard deviations can be seen in their dependences on η (Figure A.3; note that the final point in each of these plots is the rms of the distribution overall η) and p_T (Figure A.4). For $p_T = 100$ GeV muons, $\Delta\theta$ performance increases with η (energy), and ϕ performance decreases, as expected;[§] for θ , the two effects appear to compete, with performance first increasing with η until the effects of secondaries begins to dominate. Integrated over all η , the effects are less clearly delineated. Both $\Delta\theta$ and θ performance increases with increasing p_T , suggesting straighter tracks with increasing energy are the dominant effect for these quantities, while ϕ performance appears to improve and then deteriorate (the slight improvement at high p_T is due to the addition of the $\Delta\theta$ cut into the algorithm, which filters out very poor quality fits).

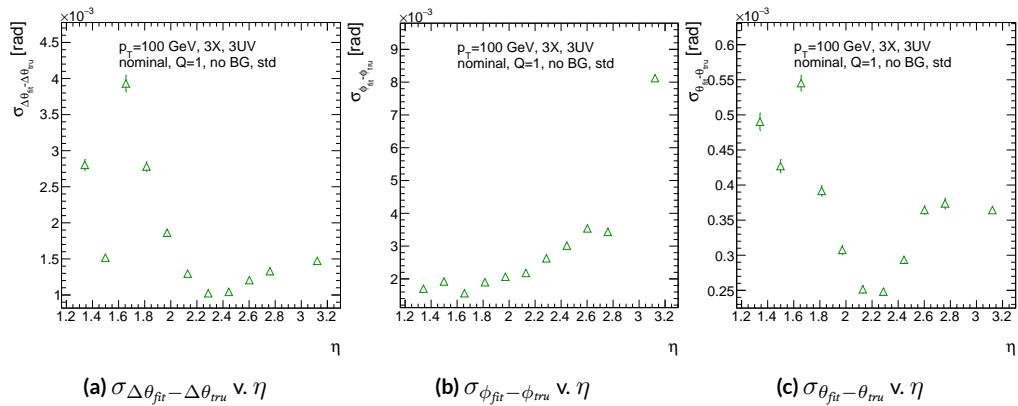


Figure A.3: The rms distributions of $\Delta\theta$, ϕ , and θ as a function of η for $p_T = 100$ GeV; the final point in each plot is the rms obtained from a fit to the full distribution including all η bins.

[§]The much worse overall performance for ϕ is due to the η dependent bias and other effects

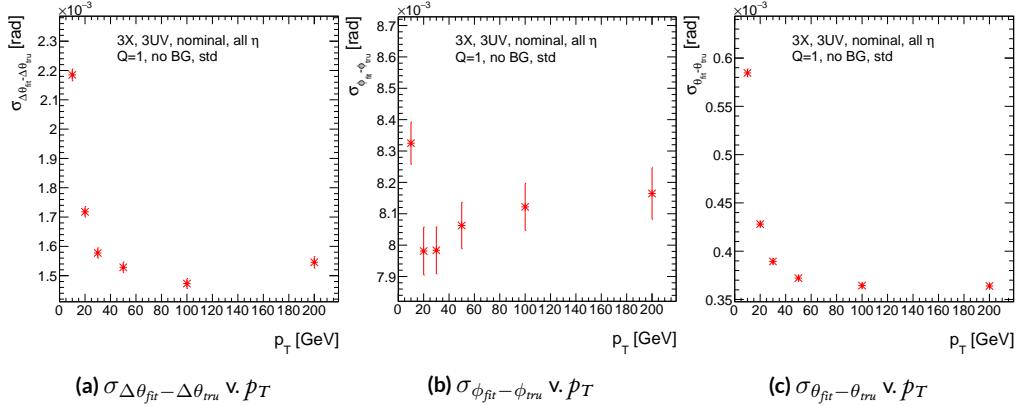


Figure A.4: The rms distributions of $\Delta\theta$, ϕ , and θ as a function of p_T .

2098 The rms of the three benchmark quantities as a function of algorithm set (i.e. slope-road) coinci-
 2099 dence threshold are shown in Figure A.5 using Gaussian fits and in Figure A.6 for the raw quantities.
 2100 The fitted σ 's for θ and ϕ are fairly stable across coincidence threshold. $\Delta\theta$, on the other hand, per-
 2101 forms better particularly for the most stringent coincidence threshold; this is a result of the fact that
 2102 additional information for more hits greatly improves the quality of the local slope fit calculation.
 2103 The raw rms is a different story. Naïvely, one would expect the performance to get better with more
 2104 stringent coincidence threshold, but this is not the case in Figure A.6. As the coincidence thresh-
 2105 old gets more stringent, fewer and fewer tracks are allowed to be fit. When moving from 2X hits to
 2106 3X hits, the tracks that get vetoed populate the tails of the distribution outside the 3σ fit range but
 2107 are not in the very extremes of the distribution. While tracks with 2X hits are of lower quality than
 2108 those with 3 and 4 X hits, tracks with the very worst fit values pass even the most stringent coinci-
 2109 dence threshold requirements (e.g. as a result of many hits arising from a shower of secondaries).
 2110 This is best illustrated when comparing the 2X+1UV $\Delta\theta$ residual distribution with the 4X+4UV

distribution in Figure A.7. As both the overlayed normalized curves and ratio distribution show,
 while the most central regions are fairly similar, the $\omega X + 1$ UV distribution is much more prominent
 in the tails but not the extreme tails, which means that, though the overall $\omega X + 1$ UV raw rms goes
 down, the overall quality of algorithm fits is worse.

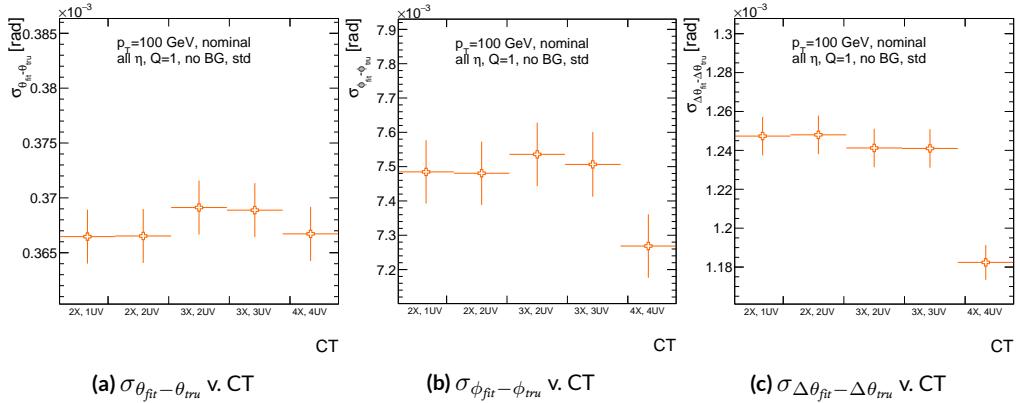


Figure A.5: The fitted rms of residual distributions for θ , ϕ , and $\Delta\theta$ as a function of coincidence threshold for $p_T = 100$ GeV.

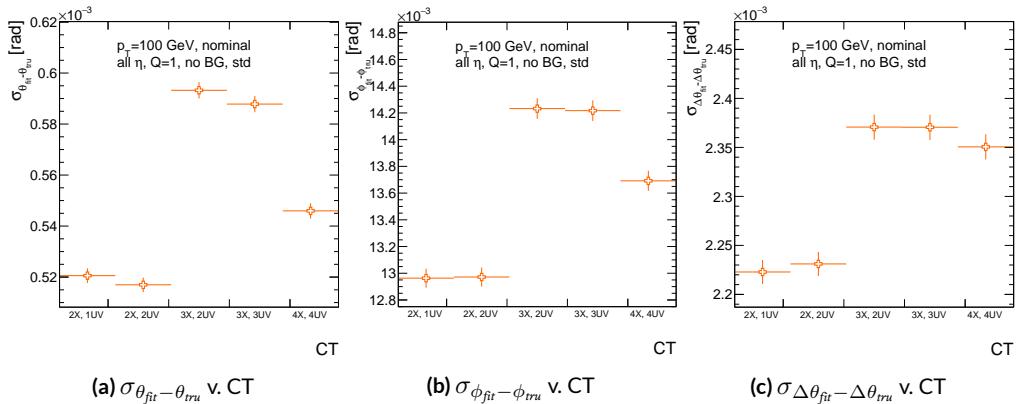


Figure A.6: The raw rms of residual distributions for θ , ϕ , and $\Delta\theta$ as a function of coincidence threshold for $p_T = 100$ GeV.

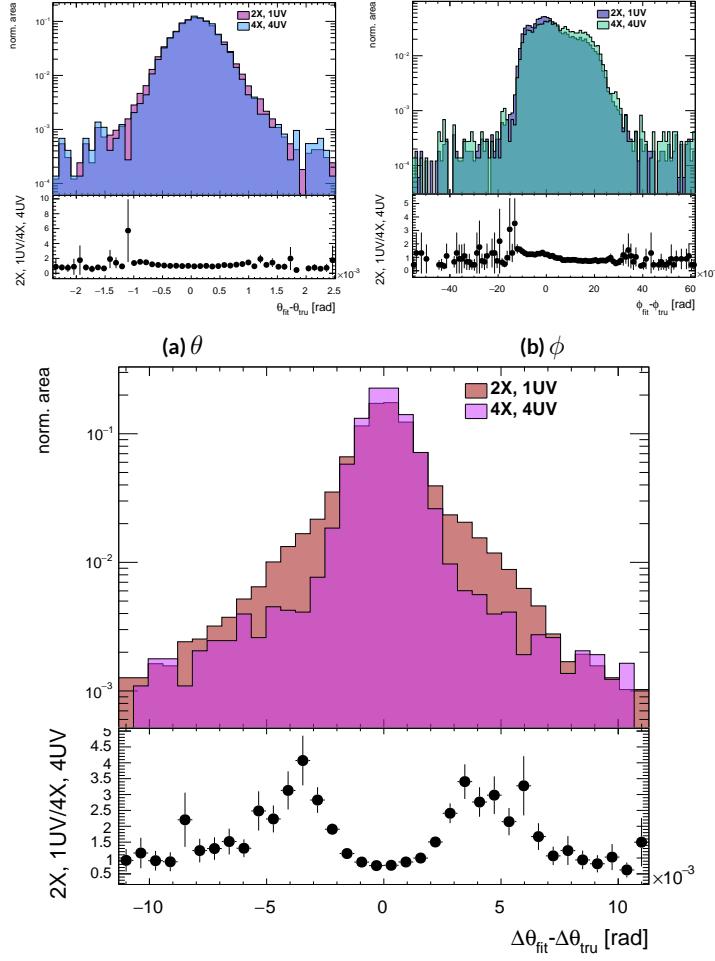


Figure A.7: Nominal $\Delta\theta$ residual distribution for $p_T = 100$ GeV muons with coincidence thresholds 2X+1UV and 4X+4UV normalized to the same area and plotted together (top) as well as the ratio of the 2X+1UV distribution and the 4X+4UV per bin.

2115 A.5 EFFICIENCIES

2116 Two general efficiencies have been formulated to study the performance of the MMTP algorithm.
 2117 The first, denoted ε_{alg} , is the fraction of tracks that pass some (slope-road) coincidence threshold
 2118 configuration that are successfully fit. An event that passes a slope-road coincidence but does not fit
 2119 fails because some of the hits included are of sufficiently poor quality to throw off the fit. This effi-
 2120 ciency answers the question of how often the algorithm performs fits when technically possible, giv-
 2121 ing a measure of overall algorithm performance for a given configuration. For example, $\varepsilon = 95\%$ for
 2122 $3X+2UV$ means that 95% of tracks that produce at least $3X$ hits and $2UV$ hits in at least one slope-
 2123 road will be successfully fitted 95% of the time. The performance of this efficiency as a function of
 2124 coincidence threshold, η (with the final point once again being the efficiency integrated over all η),
 2125 and p_T is shown in Figure A.8. ε_{alg} is fairly constant in η and decreases with increased p_T , which can
 2126 be attributed to the increased likelihood of secondaries introducing lower quality hits that cause the
 2127 fit to fail.

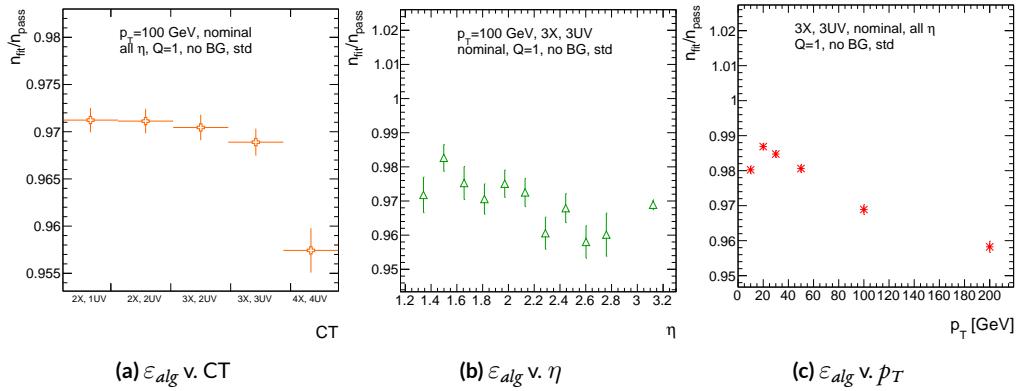


Figure A.8: ε_{alg} and as a function of coincidence threshold, η (final point is ε_{alg} integrated over all η), and p_T .

2128 The second efficiency type, denoted ε_{fit} , is the fraction of tracks that enter the wedge whose fits
 2129 (if any) satisfy a given coincidence threshold. This efficiency can be used to help establish an optimal
 2130 coincidence threshold setting in the algorithm, balancing the improved overall fit quality of higher
 2131 thresholds with the greater number of fits for lower thresholds. Hence, an ε_{fit} of 95% at 3X+2UV
 2132 means that 95% of tracks entering the wedge are fit and that these fits include at least 3X and 2UV
 2133 hits. ε_{fit} as a function of coincidence threshold is shown in Figure A.9 (a), which shows that the
 2134 majority of fits having at most 3X+3UV hits. That there is a marked drop to 4X+4UV is not sur-
 2135 prising, as there is a substantial population outside the 4X+4UV bin in Figure A.10. The behavior
 2136 of ε_{fit} with η in Figure A.9 (b) (with the final point once again being the efficiency integrated over
 2137 all η) is much more varied, with geometric effects of detector acceptance coming into play. The per-
 2138 formance of ε_{fit} as a function of p_T , shown in Figure A.9 (c), is similar to that of ε_{alg} coincidence
 2139 threshold, again consistent with the effects of secondaries at higher energies.

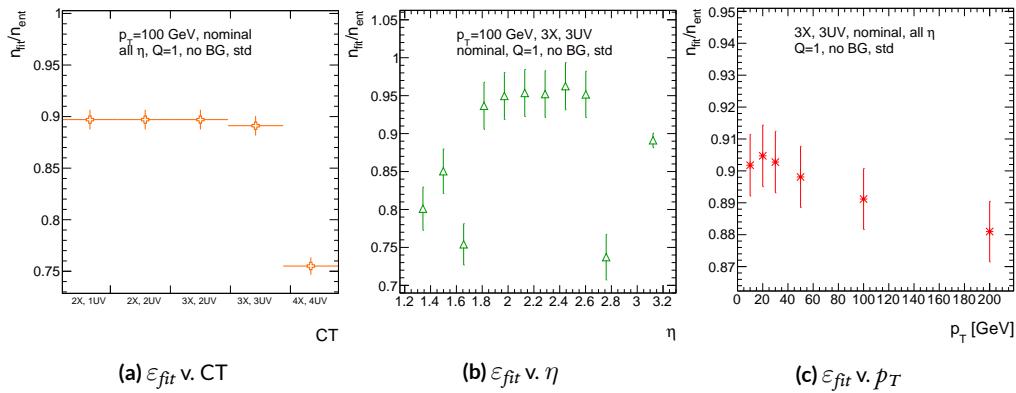


Figure A.9: ε_{fit} and as a function of coincidence threshold, η (final point is ε_{fit} integrated over all η), and p_T .

2140 In order to better understand efficiency behavior with coincidence threshold, the distribution

2141 of highest slope-road coincidence thresholds in events is shown in Figure A.10, with the o,o bin
 2142 containing events that did not meet requirements for the minimum $2X+1UV$ coincidence threshold
 2143 for a fit to occur. That the efficiency is lower at higher coincidence threshold suggests that most
 2144 of the fits that fail have high hit multiplicity (i.e. a similar number fails in each of the coincidence
 2145 threshold bins in Figure A.8 (a)), which is consistent with the interpretation that the primary source
 2146 of fit failures is bad hits originating from secondaries created by higher energy muons.

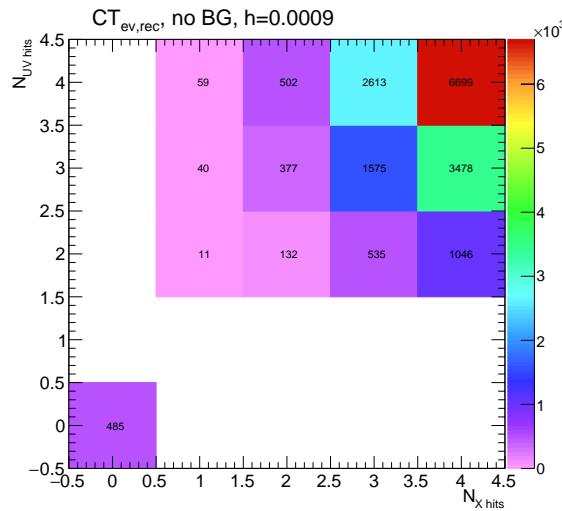


Figure A.10: The distribution of highest slope-road coincidence thresholds in events; the 0,0 bin is the number of events passing selection requirements that fail to form the minimum $2X+1UV$ coincidence threshold necessary for a fit.

2147 A.6 INCOHERENT BACKGROUND

2148 The default slope-road size and tolerances associated with horizontal and stereo hits used in the
2149 above studies were configured to optimize algorithm performance, similar to studies in³³. In order to
2150 evaluate algorithm performance under conditions with more limited resources, as might be expected
2151 at run-time, additional studies were conducted with the slope-road size and hit tolerances set equiv-
2152 alent to the sensitive area of a single VMM chip[¶] both with and without generation of incoherent
2153 background.

2154 Incoherent background is generated based on the assumption that the intensity only varies as a
2155 function of the distance from a point to the beamline, r . The number of hits per unit area per unit
2156 time as a function of r is given in Equation A.6 and taken from³³.

$$I = I_o (r/r_o)^{-2.125} \quad (\text{A.6})$$

2157 where $r_o = 1000$ mm and $I_o = 0.141$ kHz/mm²

2158 Background generation happens per event as follows:

- 2159 1. Determine the total number of hits to be generated in this event according to a Poisson distri-
2160 bution
- 2161 2. Assign a time to hits uniformly in $[t_{start} - t_{VMM}, t_{end}]$ where start and end are for the event
2162 clock and t_{VMM} is the VMM chip deadtime (100 ns)
- 2163 3. Assign a plane to hits uniformly
- 2164 4. Assign a ϕ value to hits uniformly

[¶]One VMM is assumed to cover 64 MM strips at 0.445 mm each.

2165 5. Assign an r to hits according to Equation A.6

2166 6. Calculate hit information according to these values.

2167 The expectation value for the Poisson distribution is determined by integrating Equation A.6

2168 over the surface area of the wedge to get the total hit rate for the wedge, Γ , and then multiplying this

2169 by the length of the time window over which hits may be generated. With $H = 982$ mm, $b_t = 3665$

2170 mm, and $\theta_w = 33\pi/180$, we find¹¹:

$$\Gamma = 2I_0 r_o^{2.125} \int_0^{\theta_w/2} d\phi \int_{H \sec \phi}^{(H+b_t) \sec \phi} r dr r^{-2.125} = 98.6657 \text{ MHz} \quad (\text{A.7})$$

2171 In this case, we have taken the nominal values of the MM sector geometry for H (wedge base), b_t

2172 (the wedge height), and θ_w (the wedge opening angle).

2173 The effects of incoherent background and larger slope road size are summarized in Figure A.11 for

2174 efficiencies and in Figure A.13 and Table A.1 for residual of fit quantities.

2175 Figure A.11 show the effect of both wider slope-roads and the introduction of background on ef-

2176 ficiencies. The introduction of wider slope-roads increases the chance that an early errant hit (either

2177 from secondaries/ionization or background) will be introduced into the fit, and the presence of in-

2178 coherent background greatly increases the number of such errant hits. Both wider slope-roads and

2179 background drive down the number of fits (numerator) in both efficiencies, and background can

2180 artificially inflate the denominator of ε_{alg} , a reco-level, slope-road coincidence threshold. The shape

2181 of the ε_{fit} versus coincidence threshold distributions remains fairly constant with each complicating

¹¹Using Mathematica and the extra factor of r from the volume element

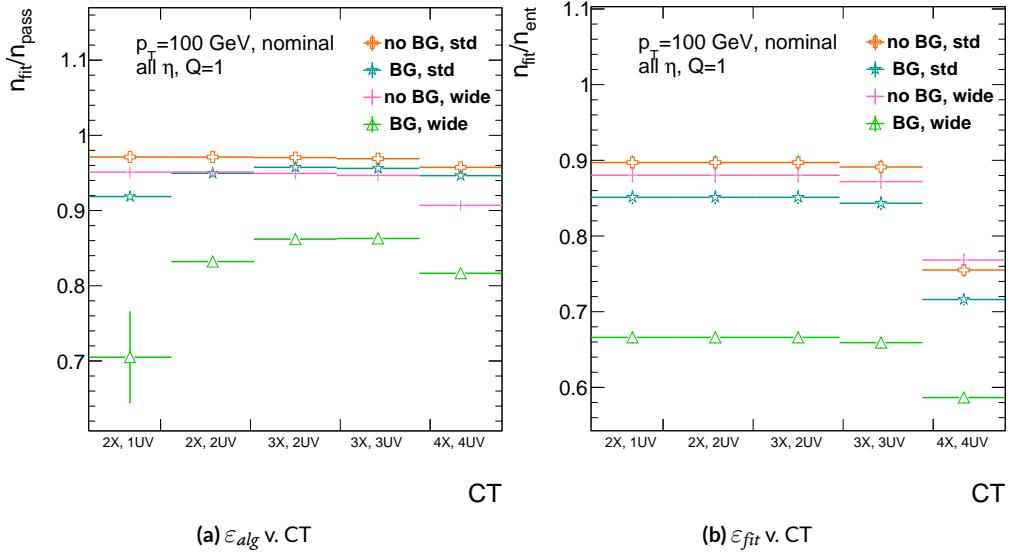


Figure A.11: The algorithm and total efficiencies as a function of coincidence threshold for different background settings and slope-road sizes (standard and wide (one slope road as 1 VMM chip)).

2182 factor (standard, wider slope-roads, background, both wider slope-roads and background), suggest-
 2183 ing many muons will simply not be fit with any number of hits; ε_{fit} does not take into account the
 2184 coincidence threshold of tracks that are not fit, so the effect appears uniform across coincidence
 2185 threshold. The effects seen for ε_{alg} , which are not uniform across coincidence threshold can be bet-
 2186 ter understood when examining the distribution of event highest coincidence thresholds, shown for
 2187 wide slope-roads both without and with background in Figure A.12. Take, for example the 2X+1UV
 2188 case. The 2X+1UV bin in particular has a marked increase when background is introduced. No
 2189 new, good tracks are introduced between the no backgrond and background cases, so the increase is
 2190 entirely due to bad, background hits; hence, these events do not (and should not) fit, causing the
 2191 particularly pronounced drop in this bin between these two cases in Figure A.11.

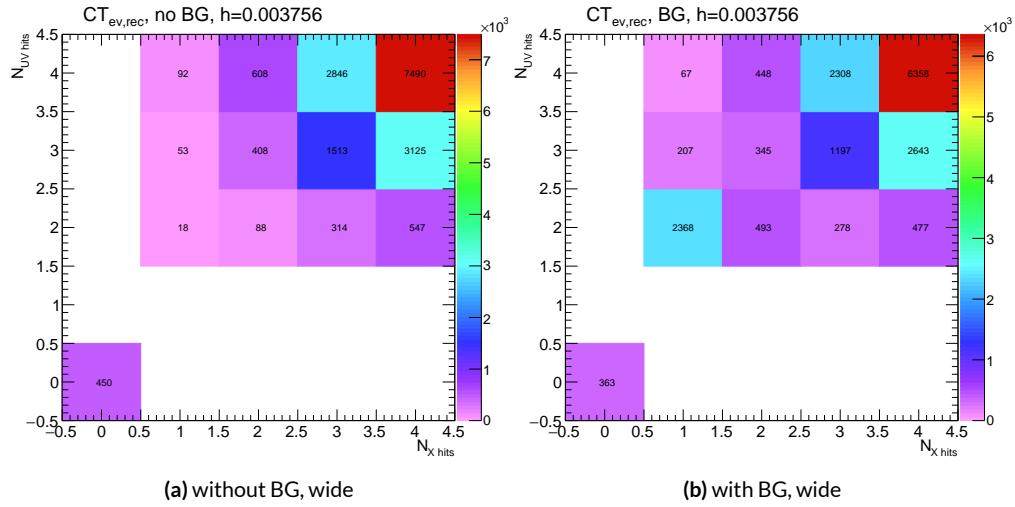


Figure A.12: The distribution of highest slope-road coincidence thresholds in events for the algorithm with wide slope-roads (width of 1 VMM) both without (a) and with (b) incoherent background; the 0,0 bin is the number of events passing selection requirements that fail to form the minimum $2X+1UV$ coincidence threshold necessary for a fit.

The effect of increasing slope-road size and incoherent background on fit quantity residual rms values as a function of p_T is shown in Figure A.13. As the figure shows, the fitted rms values are fairly robust against increased slope-road size and background. This does not hold for all of the raw rms values, however, as shown in Table A.1. Just as with the efficiencies, the introduction of background has a larger effect than that of increased slope-road size, which does not seem to have an overly large impact on any of the fit quantities on its own. While $\Delta\theta$ remains robust to both increased slope-road size and background (likely due to the $\Delta\theta$ cut of 16 mrad built into the algorithm), θ shows some degradation in performance, and the ϕ residual raw rms shows a very large increase upon the introduction of background. Nevertheless, the contrasting behavior of the fitted and raw rms values suggests that tracks that drive up the raw rms values already had very poor fit quality even before the introduction of background, so the impact on fit quantities should remain fairly limited.

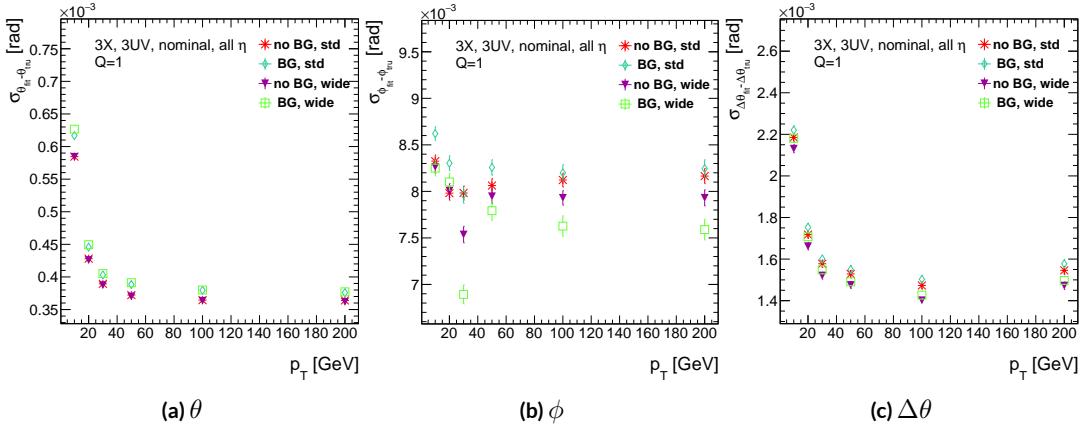


Figure A.13: The three fit quantity residual rms values as a function of p_T for different background settings and slope-road sizes (standard and wide (one slope road as 1 VMM chip)).

	No BG, std	No BG, wide	BG, std	BG, wide
θ	0.364 (0.604)	0.363 (0.542)	0.379 (0.886)	0.380 (1.07)
ϕ	8.12 (15.0)	7.93 (13.2)	8.20 (24.6)	7.63 (24.8)
$\Delta\theta$	1.47 (2.69)	1.40 (2.66)	1.50 (2.89)	1.43 (2.90)

Table A.1: The fitted (absolute) σ of fit quantity residuals in mrad under different algorithm settings.

2203 As Table A.1 shows, rms values appear to be robust to an increase in slope-road size. Neverthe-
2204 less, though the fitted σ residual values are also fairly robust to the introduction of background, the
2205 raw rms values are not. While the raw $\Delta\theta$ rms stays stable, both θ and ϕ suffer noticeable degra-
2206 dation, which suggests that the introduction of background has a detrimental effect on horizontal
2207 slope residual (i.e. on stereo strips in particular). This level of degradation is likely acceptable for θ ,
2208 though further steps may need to be taken to address ϕ .

2209 **A.7 BCID**

2210 A fitted track's BCID is determined by the most common BCID associated with its hits. Concerns
2211 were raised that this might cause incorrect BCID association for fitted tracks. In order to address this,
2212 the rate of successful BCID association for fitted tracks was recorded. Figure A.14 shows the depen-
2213 dence of this success rate as a function of p_T and coincidence threshold in the different background
2214 and resource conditions used in the previous section. The successful BCID identification rate is al-
2215 ways over 99.5%, demonstrating that this issue is not a concern with the state-of-the-art detector
2216 simulation.

2217 **A.8 CHARGE THRESHOLD**

2218 The MMTP uses the first hits registered passing a charge threshold requirement given in units of
2219 electron charge. In principle, it would be beneficial to be able to use any hits that are registered re-
2220 gardless of deposited charge, but in the high rate environment envisioned for the NSW, this require-
2221 ment might need to be raised. Nominal algorithm settings have this charge threshold requirement

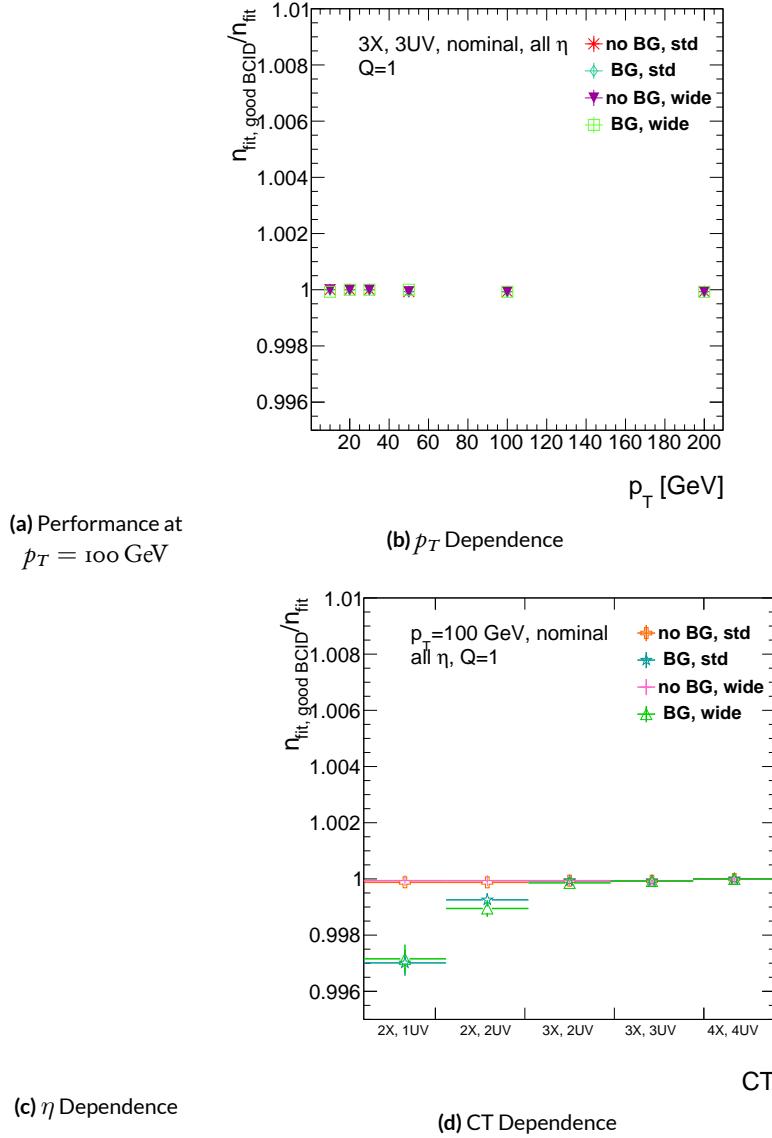


Figure A.14: The rate of good BCID association based majority hit BCID as a function of p_T and coincidence threshold.

2222 set to 1, and studies were conducted on algorithm performance for charge threshold values of 0, 1,
 2223 and 2. Efficiencies as a function of coincidence threshold for different charge thresholds are shown
 2224 in Figure A.15. Increasing the charge threshold lowers both efficiencies, particularly at high coinci-
 2225 dence threshold, which suggests that energetic muons with secondaries create both very many hits
 2226 and hits with higher charge. While the shapes of the fit quantity distributions as a function of p_T in
 2227 Figure A.16 are fairly constant across charge threshold, performance is not. θ and $\Delta\theta$ show some im-
 2228 provement with higher charge threshold, particularly at low p_T , suggesting that resolution improves
 2229 in the vertical direction, but ϕ shows degradation at higher charge threshold, which is a symptom
 2230 of more highly charged particles experiencing greater bending in the ATLAS magnetic field in the ϕ
 2231 direction.

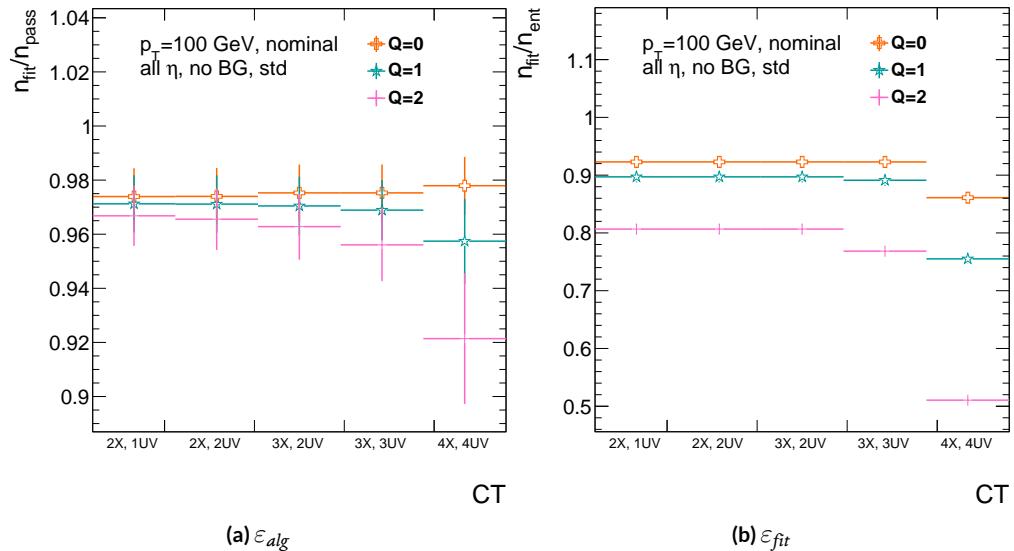


Figure A.15: The efficiencies as a function of coincidence threshold for charge thresholds of 0, 1, and 2.

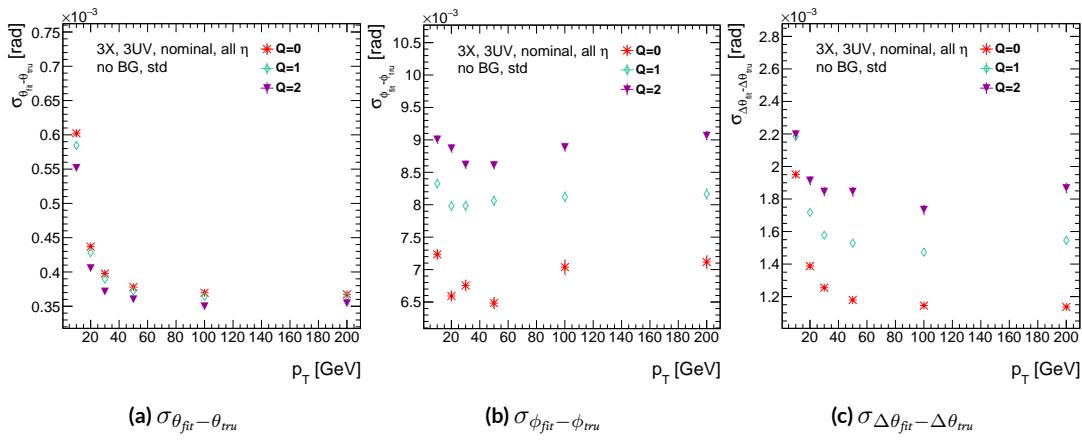


Figure A.16: The fit quantity residual rms values as a function of p_T for charge thresholds of 0, 1, and 2.

2232 A.9 MISALIGNMENTS AND CORRECTIONS

2233 The performance of the trigger algorithm under misalignment has been studied for each of the six
2234 alignment quantities (three translations and three rotations all along the principal axes) described
2235 in[?] and[?], whose convention we will follow here. For the simulated wedge studied here the local co-
2236 ordinates described in[?] are taken to be centered at the center of the base of the wedge^{**}, the local t
2237 axis corresponds to the axis of the beam line, the local z axis corresponds to the direction orthogo-
2238 nal to both the beam line and the horizontal strips, and the local s axis completes the right-handed
2239 coordinate system. The rotation angles α , β , and γ correspond to rotations around the local t , z ,
2240 and s axes, respectively. Note that the local s , z , and $-t$, axes correspond to the usual global x , y , and
2241 z axes. Misalignments were studied in twenty evenly spaced increments from nominal positions
2242 to misalignments of 1.5 mrad for the rotations (-1.5 mrad to +1.5 mrad for the γ case), and of 5 mm
2243 (a roughly corresponding linear shift) for the translations. In all cases, the front quadruplet is mis-
2244 aligned while the rear quadruplet remains in its nominal position. While only the front quadruplet
2245 of a single wedge is misaligned, the framework for misalignment presented below could be used to
2246 study generic local and global misalignments. The six misalignments are schematically represented
2247 in Figure A.17.

2248 Chamber misalignments manifest themselves as altered strips in algorithm input. In order to
2249 simulate the effects of misalignment, the change in the local y coordinate—the distance from the

^{**}Not, as is sometimes the case, the centroid position for simplicity's sake, as the agreed upon geometry of the detector changed several times while studies were in progress; any transformation in a centroid-origin coordinate system can of course be formed by a combination of the six transformations examined.

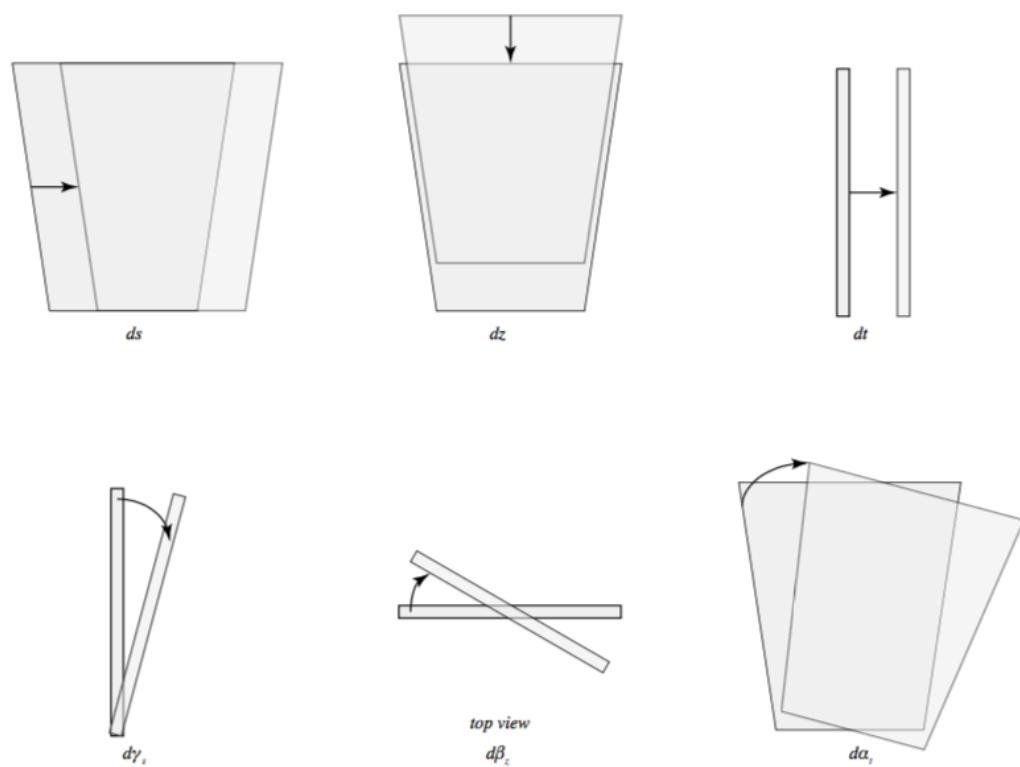


Figure A.17: The different misalignment cases as defined in the AMDB manual.

bottom wedge center in the direction perpendicular to both the beamline and the strip direction—is
 calculated for a track coming straight from the interaction point defined by the truth-level θ and ϕ
 angles for generic misalignment. This displacement in y is then added to input hit information and
 the algorithm is then run normally.

To understand how this displacement is calculated, some notation first needs to be described.

Table A.2: A summary of notation used in this section: note that non-AMDB notation is used in this section.

Symbol	Definition
s_x, s_y, s_z, \vec{s}	Position of the muon hit in ATLAS global coordinates; the infinite momentum muon track
\hat{n}	Vector normal to the plane; taken to be \hat{z} (the beamline) in the nominal case
$\vec{\mathcal{O}}_{IP}^{g,l}$	Position of the interaction point in ATLAS global (g) or wedge local (l) coordinates
$\vec{\mathcal{O}}_{base}^{g,l}$	Position of the plane base in ATLAS global (g) or wedge local (l) coordinates; (o, y_{base}, z_{pl}) ((o, o, o)) for the nominal case in global (local) coordinates
$\vec{\zeta}$	$\vec{s} - \vec{\mathcal{O}}_{base}$
primed quant.	quantities after misalignment

Generically speaking, a hit is the intersection of a line (the muon track) with a plane (the individual plane in the multiplet). We assume the muon moves in a straight line defined by the origin and the truth-level θ_{pos} and ϕ_{pos} (i.e. the infinite momentum limit) and that the MM plane is rigid and defined by a point, which we take to be the center of the bottom edge of the plane, and a normal vector, which we take to be the z axis in the nominal case.

The coordinate axes x, y, z axes used here correspond to the usual AMDB $s, z, -t$ axes. Since the direction does not really matter when studying misalignment or corrections thereof, the major

2262 difference is the choice of origin.

2263 The muon track we denote^{††} \vec{s} , the bottom point of the plane $\vec{\mathcal{O}}_{base}$, and the normal vector \hat{n} .

2264 The muon track will always be given as (the wedge gets moved, not the muon):

$$\vec{s} = \vec{\mathcal{O}}_{IP} + k\hat{s} \quad (\text{A.8})$$

$$\hat{s} = \sin \theta_{pos} \sin \phi_{pos} \hat{x} + \sin \theta_{pos} \cos \phi_{pos} \hat{y} + \cos \theta_{pos} \hat{z} \quad (\text{A.9})$$

$$\vec{s} = k\hat{s} = \frac{z_{pl}}{\cos \theta_{pos}} \hat{s} = z_{pl} (\tan \theta \sin \phi \hat{x} + \tan \theta \cos \phi \hat{y} + \hat{z}) \quad (\text{A.10})$$

2265 where $k \in \mathbb{R}$, along with the unit vector \hat{s} , defines the point where the track intersects the wedge.

2266 Rotations are done before translations, according to the order prescribed in the AMDB guide for

2267 chamber alignment, so the axes the principal axes of the plane are rotated according to the following

2268 matrix (where s , c , and t are the obvious trigonometric substitutions)

$$\begin{aligned} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & c\gamma & -s\gamma \\ 0 & s\gamma & c\gamma \end{pmatrix} \begin{pmatrix} c\beta & 0 & s\beta \\ 0 & 1 & 0 \\ -s\beta & 0 & c\beta \end{pmatrix} \begin{pmatrix} c\alpha & -s\alpha & 0 \\ s\alpha & c\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c\gamma & -s\gamma \\ 0 & s\gamma & c\gamma \end{pmatrix} \begin{pmatrix} c\alpha c\beta & -s\alpha c\beta & s\beta \\ s\alpha c\gamma + c\alpha s\beta s\gamma & c\alpha c\gamma - s\alpha s\beta s\gamma & -c\beta s\gamma \\ s\alpha s\gamma - c\alpha s\beta c\gamma & c\alpha s\gamma + s\alpha s\beta c\gamma & c\beta c\gamma \end{pmatrix} \\ & = \boxed{\begin{pmatrix} c\alpha c\beta & -s\alpha c\beta & s\beta \\ s\alpha c\gamma + c\alpha s\beta s\gamma & c\alpha c\gamma - s\alpha s\beta s\gamma & -c\beta s\gamma \\ s\alpha s\gamma - c\alpha s\beta c\gamma & c\alpha s\gamma + s\alpha s\beta c\gamma & c\beta c\gamma \end{pmatrix}} = \mathcal{A} \end{aligned} \quad (\text{A.11})$$

^{††}Recall ϕ_{pos} is defined with respect to the y axis instead of the x axis, as might otherwise be typical.

2269 The thing that matters is what the new strip hit is—i.e. what the new y value is since this, along
 2270 with a plane number, is all that is fed into the algorithm. To find this, we must solve for the new
 2271 point of intersection with the rotated plane and then apply the effects of translations. The path
 2272 connecting the base of the wedge with the intersection of the muon track will always be orthogonal
 2273 to the normal vector of the plane. Our quantities after misalignment, denoted by primed quantities,
 2274 will look like

$$\mathcal{O}_{base} \rightarrow \mathcal{O}_{base} + ds\hat{x} + dz\hat{y} + dt\hat{z} = \mathcal{O}'_{base}, \quad \hat{n} \rightarrow A\hat{n} = A\hat{z} = \hat{z}', \quad \vec{s} \rightarrow k'\hat{s} + \mathcal{O}_{IP} = \vec{s}' \quad (\text{A.12})$$

2275 so, moving to explicit, global coordinates in the last line so we can do the computation (relying on
 2276 the fact that any vector in the wedge, namely $\vec{\zeta} = \vec{s} - \mathcal{O}$ the local coordinates of the interaction
 2277 point, is necessarily orthogonal to \hat{n}):

$$o = \hat{n} \cdot (\vec{\mathcal{O}}_{base} - \vec{s}) \rightarrow o = A\hat{z}' \cdot (\vec{\mathcal{O}}'_{base} - (k'\hat{s} + \vec{\mathcal{O}}_{IP})) \quad (\text{A.13})$$

$$\rightarrow k' = \frac{s\beta\vec{\mathcal{O}}'_{base-IP,x} - c\beta s\gamma\vec{\mathcal{O}}'_{base-IP,y} + c\beta c\gamma\vec{\mathcal{O}}'_{base-IP,z}}{\hat{s} \cdot \hat{z}'} \quad (\text{A.14})$$

$$= \frac{s\beta ds - c\beta s\gamma(y_{base} + dz) + c\beta c\gamma(z_{pl} + dt)}{s\beta s\theta s\phi - c\beta s\gamma s\theta c\phi + c\beta c\gamma c\theta} \quad (\text{A.15})$$

2278 To find our new y coordinate, we need to evaluate $s'_y = \hat{y}' \cdot k'\vec{s}$ to find the final correction of:

$$\Delta y = \vec{\zeta}' \cdot \hat{y}' - \vec{\zeta} \cdot \hat{y} = (k'\hat{s} - \vec{\mathcal{O}}'_{base}) \cdot \hat{y}' - (s_y - y_{base}) \quad (\text{A.16})$$

²²⁷⁹ The correction will be plane dependent since (denoting the stereo angle ω):

$$\hat{y}_x = \hat{y} \rightarrow \hat{y}'_x = -s\alpha c\beta \hat{x} + (c\alpha c\gamma - s\alpha s\beta s\gamma) \hat{y} + (c\alpha s\gamma + s\alpha s\beta c\gamma) \hat{z} \quad (\text{A.17})$$

²²⁸⁰ and

$$\begin{aligned} \hat{y}_{U,V} = & \pm s\omega \hat{x}' + c\omega \hat{y}'_{U,V} = [\pm c\alpha c\beta s\omega - s\alpha c\beta c\omega] \hat{x} + [\pm (s\alpha c\gamma + c\alpha s\beta s\gamma) s\omega \\ & + (c\alpha c\gamma - s\alpha s\beta s\gamma) c\omega] \hat{y} + [\pm (s\alpha s\gamma - c\alpha s\beta c\gamma) s\omega + (c\alpha s\gamma + s\alpha s\beta c\gamma) c\omega] \hat{z} \end{aligned}$$

²²⁸¹

²²⁸² A.10 INDIVIDUAL CASES

²²⁸³ Currently we only study the cases where one misalignment parameter is not zero. We examine these
²²⁸⁴ in detail below, calculating the most pertinent quantities in the misalignment calculation, k'/k and
²²⁸⁵ the new horizontal and stereo y axes. Before setting out, we simplify the expressions for the trans-
²²⁸⁶ formed \hat{y}' 's, removing any terms with the product of two sines of misalignment angles, which will be
²²⁸⁷ zero.^{†‡}

$$\hat{y}'_x = -s\alpha c\beta \hat{x} + c\alpha c\gamma \hat{y} + c\alpha s\gamma \hat{z} \quad (\text{A.19})$$

²²⁸⁸

$$\hat{y}'_{U,V} = [\pm c\alpha c\beta s\omega - s\alpha c\beta c\omega] \hat{x} + [\pm s\alpha c\gamma s\omega + c\alpha c\gamma c\omega] \hat{y} + [\mp c\alpha s\beta c\gamma s\omega + c\alpha s\gamma c\omega] \hat{z} \quad (\text{A.20})$$

^{†‡}If only one misalignment parameter is non-zero, then two or more sines will contain at least one term will contain $\sin \theta = 0$.

²²⁸⁹ If the translations are zero,

$$k' = \frac{-c\beta s\gamma y_{base} + c\beta c\gamma z_{pl}}{s\beta s\theta s\phi - c\beta s\gamma s\theta c\phi + c\beta c\gamma \theta}, \quad k'/k = \frac{-c\beta s\gamma y_{base}/z_{pl} + c\beta c\gamma}{s\beta t\theta s\phi - c\beta s\gamma t\theta c\phi + c\beta c\gamma} \quad (\text{A.21})$$

²²⁹⁰ A.II $ds \neq 0$

²²⁹¹ $k'/k = 1$ (the point of intersection does not move closer or further from the IP), and only the stereo
²²⁹² planes are affected. Note that only relevant term in Equation A.16, for the stereo strip \hat{y} for $\vec{\mathcal{O}}'_{base} =$
²²⁹³ $ds\hat{x}$ is:

$$\pm \sin \omega ds \approx \pm 0.0261 ds \quad (\text{A.22})$$

²²⁹⁴ meaning that a displacement in x of 17 mm, more than three times the range of misalignments studied,
²²⁹⁵ would be necessary for a shift in the stereo planes corresponding to one strip width.

²²⁹⁶ A.12 $dz \neq 0$

²²⁹⁷ $k'/k = 1$ (the point of intersection does not move closer or further from the IP). This case is the
²²⁹⁸ trivial one (cf. Equation A.16 with $\vec{\mathcal{O}}'_{base} = dz\hat{y}$). y just gets moved in the opposite direction as the
²²⁹⁹ wedge. Correction is an additive constant.

²³⁰⁰ A.13 $dt \neq 0$

²³⁰¹ $k'/k = (z_{pl} + dt) / z_{pl}$. y gets modified by a simple scale factor. Correct by storing changing definitions
²³⁰² of plane positions in algorithm to match the misaligned values.

2303 A.14 $\alpha \neq 0$

2304 $k'/k = 1$ and

$$\hat{y}'_x = -s\alpha\hat{x} + c\alpha\hat{y} \quad (\text{A.23})$$

$$\hat{y}'_{U,V} = [\pm c\alpha s\omega - s\alpha c\omega] \hat{x} + [\pm s\alpha s\omega + c\omega] \hat{y} \quad (\text{A.24})$$

2305 A.15 $\beta \neq 0$

2306 We have $k'/k = (1 + \tan \beta \tan \theta \sin \phi)^{-1}$, and

$$\hat{y}'_x = \hat{y} \quad (\text{A.25})$$

$$\hat{y}'_{U,V} = \hat{y} \pm (c\beta\hat{x} - s\beta\hat{z}) s\omega \quad (\text{A.26})$$

2307 A.16 $\gamma \neq 0$

$$k'/k = \frac{1 - \tan \gamma \frac{y_{base}}{z_{pl}}}{1 - \tan \gamma \tan \theta \cos \phi} \quad (\text{A.27})$$

$$\hat{y}'_x = c\gamma\hat{y} + s\gamma\hat{z} \quad (\text{A.28})$$

$$\hat{y}'_{U,V} = \pm s\omega\hat{x} + c\omega\hat{y} - s\gamma c\omega\hat{z} \quad (\text{A.29})$$

2308 In order to evaluate algorithm performance under misalignment and corrections for misalignment, the absolute means and relative resolutions of the fit quantities θ , ϕ , and $\Delta\theta$ are measured

2310 as a function of misalignment. In the following, results will only be shown for which the effects of
2311 misalignment are significant. “Significant,” for misalignments of 1 mm (0.3 mrad) for translations
2312 (rotations) means more than a 5% degradation in rms and/or bias shifts in θ , ϕ , and $\Delta\theta$ of 0.01 mrad,
2313 1 mrad, and 0.1 mrad, respectively.

2314 While corrections are typically done on a case-by-base basis, they fall under two general categories,
2315 analytic and simulation based. Analytic corrections rely upon specific knowledge of the misalign-
2316 ment, with each case being handled separately; as such, the additional resources required, both extra
2317 constants and operations, if any, vary accordingly. Simulation based corrections are all done in the
2318 same manner. The algorithm is run over a training MC sample (same setup but with $p_T = 200$
2319 GeV instead of the normal 100 GeV sample so as not to overtrain the corrections), and the mean bi-
2320 ases for θ , ϕ , and $\Delta\theta$ are saved for different, equally spaced regions in the $\eta - \phi$ plane over the wedge
2321 based on the fitted θ and ϕ values. Currently, these values are saved for 10 η and 10 ϕ bins (100 η, ϕ
2322 bins total), with the number of bins in each direction being a configurable parameter. When the al-
2323 gorithm runs with simulation based correction, this table of constant corrections is saved in a LUT
2324 before runtime, and corrections are added to final fit quantities based on the (uncorrected) θ and
2325 ϕ fit values. With the settings mentioned, this is 300 extra constants ($10\eta\text{-bins} \times 10\phi\text{-bins} \times 3$ fit
2326 quantities) and two extra operations (a lookup and addition for each quantity done in parallel). The
2327 simulation correction can, in principle, also be applied to the algorithm in nominal conditions with
2328 non-trivial improvements, as detailed below in Section A.17. Depending on the misalignment case in
2329 question, different approaches work better. A summary of correction methods, including resources
2330 necessary for the individual analytic cases, is shown in Table A.3.

	Δs	Δz	Δt	γ_s	β_z	α_t
Analytic Resources	yes+ 11c/2op	yes+ oc/oop	yes+ oc/oop	yes 56c/1op	no —	yes 400c/2n _X op, 32c/12n _X op
Simulation	yes+	no	no	no	yes+	yes+

Table A.3: A summary of corrections with additional constants/operations (written as $n_{const}c/n_{ops}op$; n_X is the number of X hits in a fit) necessary for analytic corrections. Yes means a correction exists but might not entirely remove misalignment effects, while yes+ means a quality of correction is only limited by knowledge of misalignment and memory

2331 A.17 SIMULATION CORRECTION OF THE ALGORITHM UNDER NOMINAL CONDITIONS

2332 In addition to using simulation based correction to counter the effects of several classes of misalign-
 2333 ment, the correction can be applied at to the algorithm under nominal conditions. The main effect
 2334 of this correction is to mitigate the effects of the bias in stereo strips. As such, the correction has a
 2335 larger effect on quantities that rely on the aggregate slope m_y , as can be seen in in Figure A.18, im-
 2336 proving $\sigma_{\theta_{fit} - \theta_{true}}$ resolution by about 25%, and reducing $\sigma_{\phi_{fit} - \phi_{true}}$ by over 50% and restoring a largely
 2337 Gaussian shape. The slight, apparent degradation in $\Delta\theta$ is due to a more mild version of the effect
 2338 seen in Figure A.7.

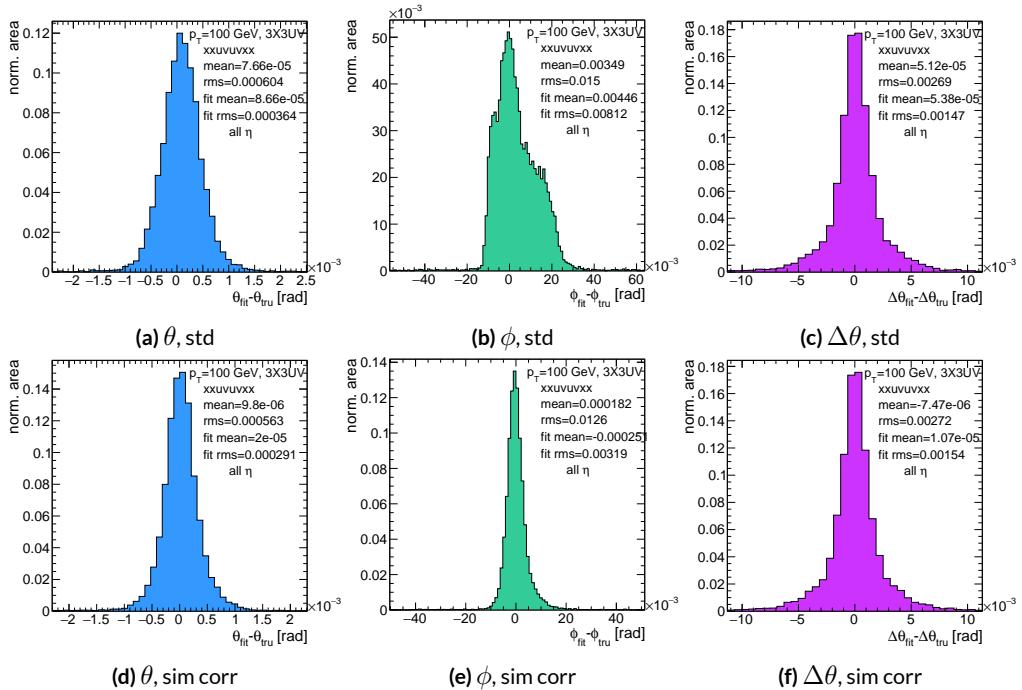


Figure A.18: Nominal residual plots for both uncorrected and simulation corrected cases; $\theta, \phi, \Delta\theta$ for $p_T = 100$ GeV muons

2339 As can be seen in Figure A.19, the simulation based correction also removes the η dependence to
 2340 fit quantity resolution distributions, as expected. One consequence of this is that simulation-based
 2341 corrections applied to the misalignment cases below will restore performance to the “sim” and not
 2342 the “std” distributions of Figure A.18. Hence, when making comparisons between simulation cor-
 2343 rected curves and the nominal performance point, simulation-corrected distributions of benchmark
 quantities versus misalignment will often look generally better.

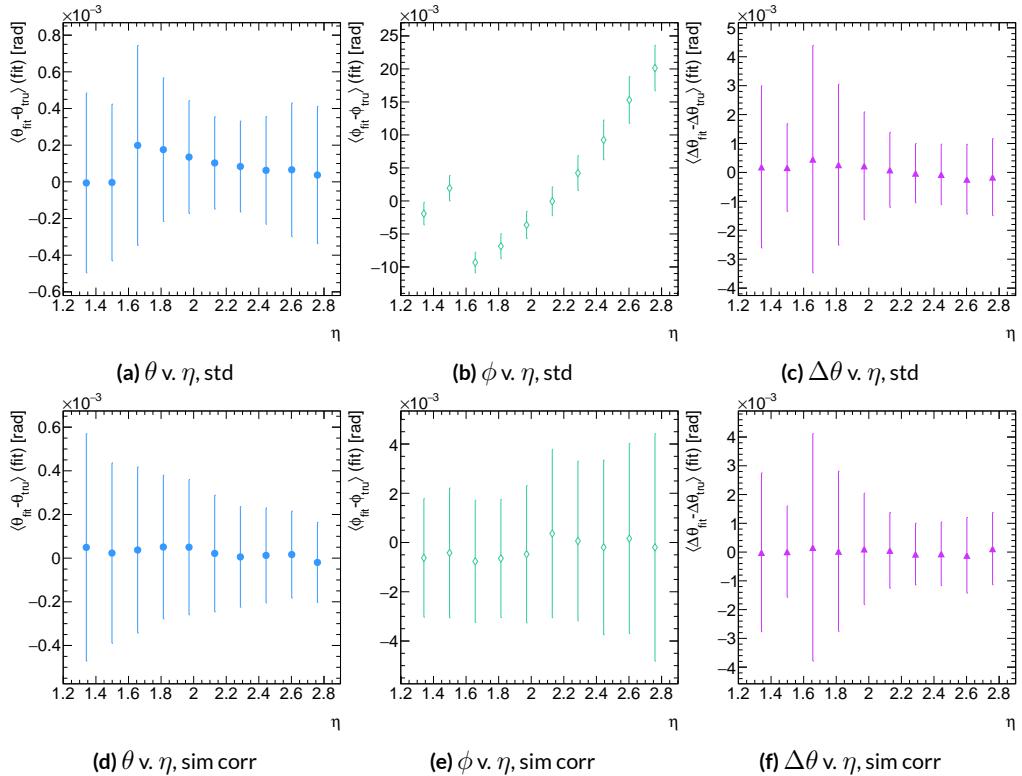


Figure A.19: Nominal residual plots as a function of η with points as means and error bars as rms values in each η bin for the angles $\theta, \phi, \Delta\theta$ for $p_T = 100$ GeV muons in the uncorrected and simulation corrected cases.

2344
 2345 That the improvements from a simulation-based correction improve performance of the algo-

2346 rithm in nominal conditions most for the quantities that depend most on stereo information (ϕ and
 2347 θ) and remove the η dependence of fit quantity resolutions suggests that there could, in principle, be
 2348 analytic corrections that could be applied to the nominal algorithm. One possible solution is to in-
 2349 troduce an additional set of constants, having the y_{base} depend on the strip number, similar to the γ_s
 2350 correction for z_{plane} described in Section A.21, which would add a lookup per hit and $8 \times n_{bins,y}$ extra
 2351 constants that would be optimized as the γ_s correction was.

$$M_{hit} = \frac{\gamma}{z} = \frac{y_{base}}{z_{plane}} (n_{str}) + \frac{w_{str}}{z_{plane}} n_{str} \quad (\text{A.30})$$

2352 The simulation correction residual rms values suggest a limit on the quality of such correction
 2353 and could perhaps be implemented generically on their own regardless of misalignment for rms
 2354 values on fit quantities of 0.291 mrad for θ , 3.19 mrad for ϕ , and 1.54 for $\Delta\theta$, which represent a 20%
 2355 improvement for θ , a 62% improvement for ϕ , and a slight degradation in $\Delta\theta$ of 4.7%, again owing
 2356 to an effect similar to the one in A.7.

2357 A.18 TRANSLATION MISALIGNMENTS ALONG THE HORIZONTAL STRIP DIRECTION (Δs)

2358 A translation in s (i.e. along the direction of a horizontal strip) only affects the stereo strips, and,
2359 since the stereo angle is small, a very large misalignment is necessary for effects to be noticeable (a
2360 misalignment of roughly 17 mm corresponds to one strip's misalignment in the stereo planes). The
2361 only quantity to show any meaningful deviation with misalignments with translations in s is the ϕ
2362 residual bias (a change of 0.4 mrad at $\Delta s = 1$ mm), as can be seen in the uncorrected curve of Figure
2363 A.20.

2364 A translation in s induces a constant shift in the calculated horizontal slope, m_x in Equation A.4.
2365 This constant shift should only depend on which stereo planes included in a fit are misaligned and
2366 how misaligned they are. Hence, the correction to m_x , for a sum over misaligned stereo planes i ,
2367 with their individual misalignments in s and plane positions in z is:

$$\Delta m_x = \frac{1}{N_{\text{stereo}}} \sum_{i, \text{misal stereo}} \frac{\Delta s_i}{z_{i, \text{plane}}} \quad (\text{A.31})$$

2368 Given prior knowledge of misalignment, these corrections to m_x can be performed ahead of time
2369 and saved in a lookup table (LUT), similar to the LUT used for constants in the X local slope (M_x^l)
2370 calculation. The added overhead of this analytic correction is hence eleven constants in memory, a
2371 lookup, and one addition. The correction perfectly corrects the effects of misalignment, as can be
2372 seen in Figure A.20. The simulation based correction described above can also be used to correct
2373 for Δs misalignments, with the results of that correction also shown in Figure A.20. The apparent

2374 discrepancy between the simulated and analytic correction is a natural consequence of the fact that
 2375 the simulation correction, as previously mentioned, restores the ϕ residual distribution to an overall
 2376 more Gaussian shape.

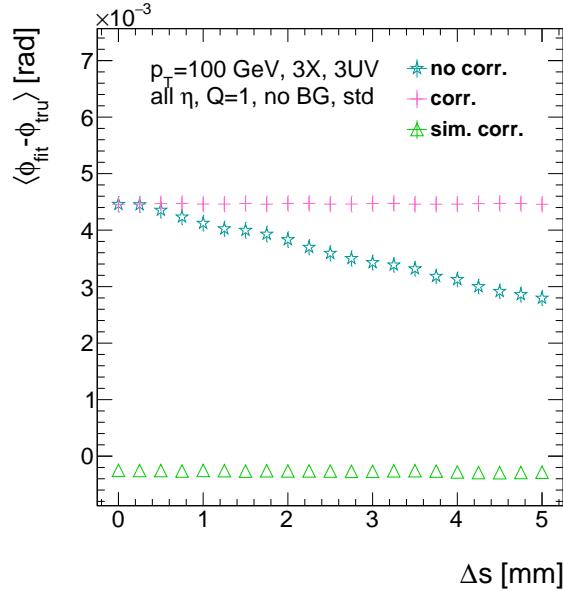


Figure A.20: The mean of the ϕ residual as a function of misalignment for the uncorrected case and the analytic and simulation correction cases.

2377 A.19 TRANSLATION MISALIGNMENTS ORTHOGONAL TO THE BEAMLINE AND HORIZON-
2378 TAL STRIP DIRECTION (Δz)

2379 A translation in AMDB z , the direction orthogonal to both the beamline and the horizontal strip
2380 direction, corresponds to a translation in the y of Equation A.1, affecting all slope calculations. This
2381 has a large impact on the θ residual bias and both the bias and rms of $\Delta\theta$ residual, as can be seen in
2382 Figures A.21 (a)–(c). The marked degradation and non-linear behavior in performance at very high
2383 levels of misalignments is a result of low statistics; there are fewer fits at high level of misalignments
2384 since for $\Delta z \gtrsim 3$ mm, most fits will fail the $\Delta\theta$ cut. The θ bias shifts by about 0.075 mrad at $\Delta z = 1$
2385 mm, and $\Delta\theta$ shifts by about 5 mrad for the same level of misalignment. While the fitted rms of the
2386 $\Delta\theta$ residual remains fairly stable for $\Delta z < 1$ mm or so, between $\Delta z = 2$ mm and $\Delta z = 3$ mm, the
2387 rms increases by 15% before the $\Delta\theta$ cut issue mentioned above intervenes.

2388 Fortunately, these misalignments are straightforward to correct with knowledge of the misalign-
2389 ment. The only modification necessary for this correction is to change the definitions of y_{base} in
2390 Equation A.1 for the individual hit slope addressing. This is done before runtime and adds no over-
2391 head to the algorithm, and the correction quality is only limited by knowledge of the misalignment.
2392 The results of this correction are also shown in Figures A.21 (a)–(c) and restore nominal perfor-
2393 mance.

Since $\Delta\theta = \frac{M_x^l - M_x^r}{1 + M_x^l M_x^r}$ and $M_x^l = B_k \sum y_i (z/\bar{z} - 1)$, a shift Δy translates (with typical slope values of ~ 0.3) to $5B_k (z_1 + z_2)/\bar{z}$ (with B_k in units of inverse mm); set equal to 16 mrad ($\Delta\theta$ is centered at zero), this corresponds to $\Delta y = 2.7$ mm

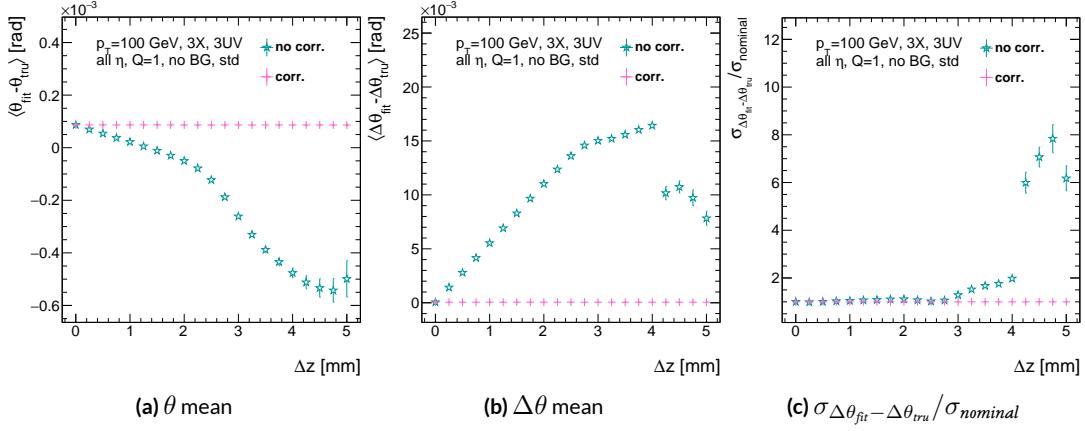


Figure A.21: The affected quantities of Δz misalignments: θ bias, $\Delta\theta$ bias, and $\sigma_{\Delta\theta_{\text{fit}} - \Delta\theta_{\text{tru}}} / \sigma_{\text{nominal}}$ for both the misaligned and corrected cases.

2394 A.20 TRANSLATION MISALIGNMENTS PARALLEL TO THE BEAMLINE (Δt)

2395 The effects of misalignment due to translations in t are very similar to those due to translations in
 2396 z without the complication of the $\Delta\theta$ cut, affecting the z instead of the y coordinate that enters
 2397 into hit slope calculations. Again, θ bias, $\Delta\theta$ bias, and $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}$ are the primarily affected quan-
 2398 tities. For $\Delta t = 1$ mm, θ bias shifts by about 0.02 mrad, $\Delta\theta$ bias shifts by just under 2 mrad, and
 2399 $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}$ degrades by about 20%. The correction for this misalignment once again costs no over-
 2400 head and consists of changing stored constants in the algorithm, in this case the positions along
 2401 the beamline of the misaligned planes, with results similarly limited by knowledge of the misalign-
 2402 ment. The slight improvement with correction to $\Delta\theta$ rms is due to the real effect of a larger lever
 2403 arm. Both the misaligned and corrected distributions of affected quantities of interest are shown in

Figure A.22.

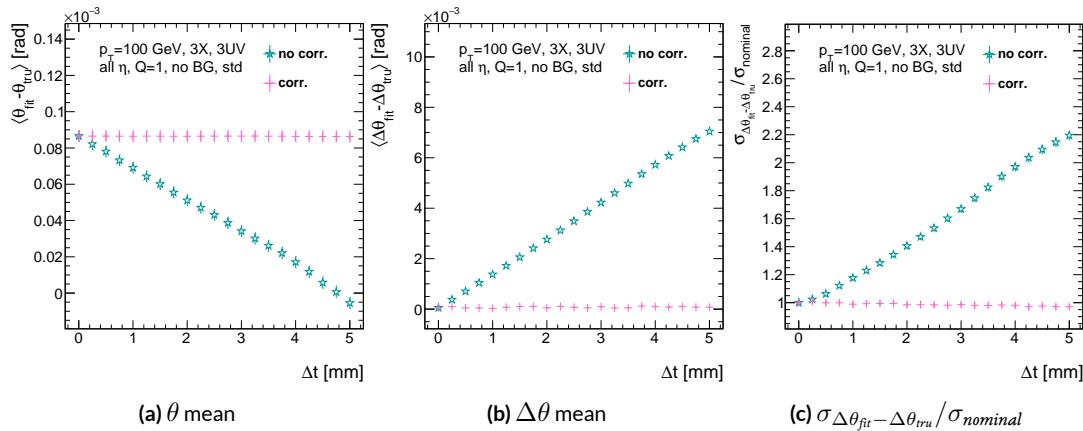


Figure A.22: The affected quantities of Δt misalignments: θ bias, $\Delta\theta$ bias, and $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}} / \sigma_{\text{nominal}}$ for both the misaligned and corrected cases.

2405 A.21 CHAMBER TILTS TOWARDS AND AWAY FROM THE IP (γ_s ROTATION)

2406 Chamber misalignment due to rotations around the s axis act effectively like a translation in t that
2407 depends on strip number. These rotations tilt misaligned chambers away from (towards) the IP for
2408 positive (negative) values of γ_s . Since, unlike for the other two rotation cases that will be studied,
2409 positive and negative rotation values are not symmetric, this misalignment is studied for both posi-
2410 tive and negative γ_s values. The divergent effect at the tails is a result of a large population of fits not
2411 having fit quantities within the cores, and so not appearing in the fit rms. Once again, affected quan-
2412 tities of interest θ bias, $\Delta\theta$ bias, and $\sigma_{\Delta\theta_{fit} - \Delta\theta_{tru}}$. The effects of misalignment can be seen in Figures
2413 A.23 (a)–(c). The relationship between biases and γ_s is roughly linear with $\Delta\gamma_s = 0.3$ mrad (the an-
2414 gular scale corresponding to linear shifts of ~ 1 mm) corresponding to 0.005 mrad (0.12 mrad) for θ
2415 ($\Delta\theta$). For $\sigma_{\Delta\theta_{fit} - \Delta\theta_{tru}}$, degradation is not symmetric. For negative (positive) γ_s , with the quadruplet
2416 tilted towards (away from) the IP, slope-roads are artificially expanded (shrunk), decreasing (increas-
2417 ing) the granularity of the trigger, explaining the asymmetry in Figure A.23 (c), with the degradation
2418 being a 10% (25%) effect for γ_s of $+(-)0.3$ mrad.

2419 Corrections are less simple in this case. In principle, corrections of the same accuracy of the trans-
2420 lations could be calculated per strip, but the overhead of one correction per strip (many thousands
2421 of constants) is prohibitive. Instead, each plane was divided into eight equal segments with a t value
2422 (z in the slope calculation) assigned to strips in each region to correct for the misalignment. This
2423 amounts to 56 extra constants and a 2D instead of a 1D LUT for z positions while the algorithm
2424 runs. The corrected distributions can also be seen in Figures A.23 (a)–(c). The corrections, while not

as effective as for the simple translation cases, are still very effective with the quoted misalignment values for bias shifts down to 0.001 mrad (0.25 mrad) for θ ($\Delta\theta$) and no more than a 2% degradation in $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}$ for $|\gamma_s| = 0.3$ mrad.

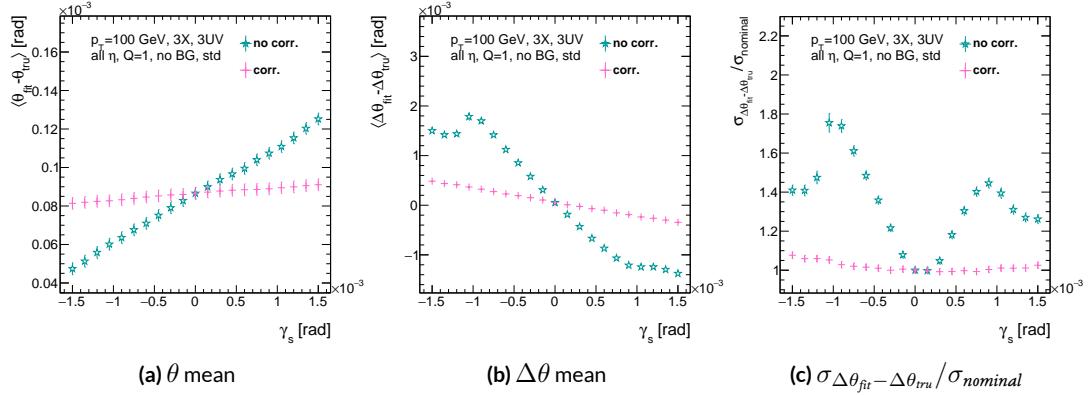


Figure A.23: The noticeable effects of rotations in the s axis and the behavior of these quantities (θ and $\Delta\theta$ bias shifts and $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}/\sigma_{nominal}$) with and without misalignment correction.

²⁴²⁸ A.22 ROTATION MISALIGNMENTS AROUND THE WEDGE VERTICAL AXIS (β_z)

²⁴²⁹ While misalignments coming from rotations around the z axis (the direction orthogonal to both
²⁴³⁰ the beamline and the horizontal strip direction) foreshorten the strips as seen from the IP and add
²⁴³¹ a deviation in t , the long lever arm largely washes out any effects of this misalignment. Only the
²⁴³² $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$ is noticeably affected, though only at severe misalignments, with only about a 1% degra-
²⁴³³ dation in performance at $\beta_z = 0.3$ mrad (corresponding to a linear shift of ~ 1 mm). A simulation
²⁴³⁴ based correction works well to cancel out the effects of this misalignment, and the $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$ as a
²⁴³⁵ function of misalignment with and without corrections are shown in Figure A.24. The apparent
²⁴³⁶ 2% effect in the simulation corrected curve is a result of a more mild version of the effect shown in
²⁴³⁷ Figure A.7.

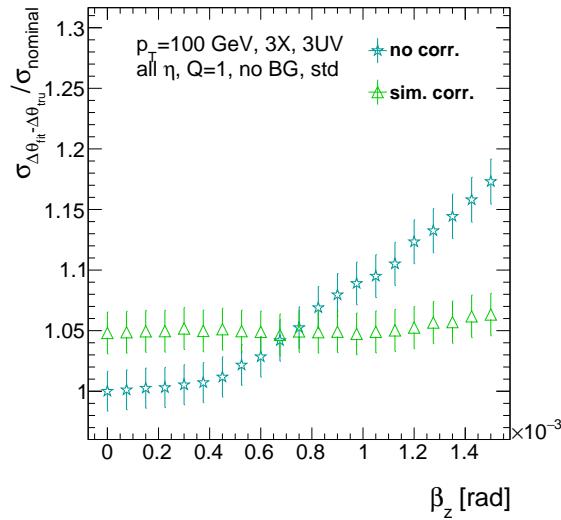


Figure A.24: The effects of rotations in the z axis on $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}} / \sigma_{nominal}$ a function of β_z both with and without misalignment corrections.

2438 A.23 ROTATION MISALIGNMENTS AROUND THE AXIS PARALLEL TO THE BEAMLINE (α_t)

2439 Misalignments arising from rotations around the t axis (parallel to the beamline at the center of
2440 the base of the wedge) are essentially rotations in the ϕ direction. The quantities of interest most
2441 affected are the ϕ bias and $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$, as shown in Figures A.25 (a) and (b), respectively, and cor-
2442 respond to a shift in ϕ bias of 0.2 mrad and a 10% degradation in $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$ for $\alpha_t = 0.3$ mrad
2443 (corresponding to a linear shift of ~ 1 mm). The raw instead of fitted mean ϕ biases is used in Figure
2444 A.25 (a) to better illustrate the effect of misalignment.

2445 Since the effect of misalignment is dependent on horizontal (along the strip direction, \hat{s}) in addi-
2446 tion to vertical information, corrections cannot be applied before a fit takes place. The ϕ bias shift is
2447 uniform over the entire wedge, so a constant additive correction to ϕ based on the level of misalign-
2448 ment can be applied to all fits depending on how many misaligned stereo planes enter in the fit. $\Delta\theta$
2449 is less straightforward, but corrections to the y and z information used in the local slope calculation
2450 in Equation A.4 can be applied once θ_{fit} and ϕ_{fit} are known. These corrections are calculated ahead
2451 of time in bins of uniform η and ϕ as with the simulation corrections using the same framework
2452 as the misalignment calculation. The results of both types of correction can be seen in Figure A.22.
2453 The apparent discrepancy between the simulation and analytic corrections in the ϕ bias happens for
2454 the same reason as in the Δs misalignment correction cases, as simulation correction restores a more
2455 Gaussian shape to the ϕ residual distribution opposed to the uncorrected nominal case, as discussed
2456 in Section A.17.

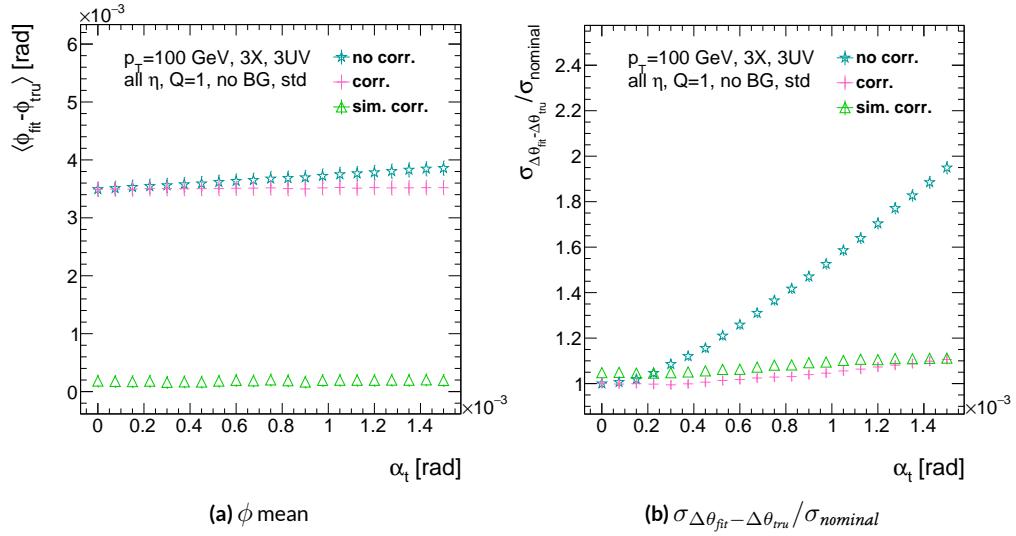


Figure A.25: The effects of rotation misalignments around the t axis for ϕ bias and $\sigma_{\Delta\theta_{\text{fit}} - \Delta\theta_{\text{true}}} / \sigma_{\text{nominal}}$ as a function of misalignment. The uncorrected and both the analytic and simulation correction cases are shown.

2457 A.24 CONCLUSION

2458 The algorithm for Micromegas detectors in the NSW Trigger Processor performs well in a variety of
2459 conditions and has proven robust to a number of effects to deliver measurements on muon tracks
2460 of the three angles θ , ϕ , $\Delta\theta$. Under nominal conditions, the rms values for the residuals of these
2461 quantities are 0.364 mrad for θ , 8.12 mrad for ϕ , and 1.47 mrad for $\Delta\theta$. Algorithm performance was
2462 found to be largely independent of the charge threshold setting, and a hit majority BCID associa-
2463 tion was found to provide proper timing information over 99.7% even in the most relaxed settings
2464 (2X+1UV coincidence threshold requirement+wide slope-road+background). The introduction of
2465 wide slope-roads to better mimic potentially limited algorithm resources at run time and the intro-
2466 duction of incoherent background was found to have a manageable effect on fit quantity residual
2467 rms values and on total algorithm efficiency for sufficiently stringent coincidence threshold. The ef-
2468 ffects of the three translation and three rotation misalignments specified by AMDB convention were
2469 studied, and correction methods for each of the six cases was developed. Simulation-based correc-
2470 tions were found to improve nominal algorithm performance to residual rms value of 0.291 mrad for
2471 θ , 3.19 mrad for ϕ , and 1.54 for $\Delta\theta$, which represent improvements of 20%, 62%, and -4.7%, respec-
2472 tively. Misalignment corrections were found to restore nominal performance for all but the rotation
2473 around the s axis, and a summary of tolerances may be found in Table A.4.

	No Correction	Correction
Δs	4 mm (ϕ bias)	> 5 mm
Δz	0.25 mm ($\Delta\theta$)	> 5 mm
Δt	0.25 mm ($\Delta\theta$)	> 5 mm
γ_s	0.15 mrad ($\Delta\theta$ bias)	0.75 mrad
β_z	0.9 mrad ($\Delta\theta$ rms)	> 1.5 mrad
α_t	0.375 mrad ($\Delta\theta$ rms)	> 1.5 mrad

Table A.4: A summary of levels of misalignment corresponding to a 10% degradation in any residual rms or, for biases shifts of, 0.01 mrad for θ , 1 mrad for ϕ , and 0.25 mrad for $\Delta\theta$ for both the uncorrected and corrected cases; > 5 mm and > 1.5 mrad mean that such a degradation does not occur for the range of misalignment studied. Most affected quantity in parentheses.

If it's stupid but it works, it isn't stupid.

Conventional Wisdom

B

2474

2475

Telescoping Jets

2476 ANOTHER APPROACH TO IMPROVING $ZH \rightarrow \ell\ell b\bar{b}$ is the use of telescoping jets²⁸, which har-
2477 nesses the power of multiple event interpretations. The use of multiple event interpretations was
2478 originally developed with non-deterministic jet algorithms like the Q-jets (“quantum” jets) algo-
2479 rithm⁶⁰. When a traditional or “classical” algorithm, such as the Cambridge-Aachen⁶⁸ and anti- k_t ⁵⁴

249

algorithms, is applied to an event, it produces one set of jets for that event, i.e. a single interpretation of that event. With multiple event interpretations, each event is instead given an ensemble of interpretations. In the case of Q-jets, this ensemble is created through a non-deterministic clustering process for an anti- k_t jet algorithm. With telescoping jets, multiple jet cone radii (the characteristic size parameter, R) around a set of points in the pseudorapidity-azimuth ($\eta - \phi$) plane are used to generate a series of jet collections. Instead of an event passing or not-passing a given set of cuts, a fraction (called the cut-weight, z) of interpretations will pass these cuts. This cut-weight allows for enhanced background suppression and increased significance of observed quantities for a given data set, as detailed in Ref.³⁹. The telescoping jets algorithm provides the benefits of multiple event interpretations without the significant computational overhead of a non-deterministic algorithm like the Q-jets algorithm, and its multiple cone sizes are particularly suited to studying processes like associated production, which suffers from a pronounced low tail in the dijet invariant mass distribution due to final state radiation (FSR) “leaking” outside the relatively narrow jets used for object reconstruction.

MONTE CARLO SIMULATION

The MC simulated samples used in this study are the same as in Ref.[?]. The signal sample used is generated in PYTHIA8[?] with the CTEQ6L1 parton distributions functions (PDFs) and AU2 tune^{???} for the ZH process with $m_H = 125$ GeV (henceforth, ZH_{125}). The primary background processes examined in this study were $Z +$ jets with massive b and c quarks. These samples are generated with version 1.4.1 of the SHERPA generator[?].

2500 B.2 JET RECONSTRUCTION AND CALIBRATION

2501 In order to construct telescoping jets, jet axes must first be found around which to “telescope.” In
2502 the reconstructed-level analysis, the anti- k_t algorithm with $R = 0.4$ is used to reconstruct jets from
2503 topological clusters in the calorimeters. The four vectors of these anti- k_t algorithm with $R = 0.4$ jets
2504 are calibrated to match truth information obtained from simulation and validated in data. To take
2505 into account the effect of pile-up interactions, jet energies are corrected using a jet-area based tech-
2506 nique³, and each jet with $p_T < 50$ GeV and $|\eta| < 2.4$ is subject to a requirement that at least 50% of
2507 the scalar sum of the p_T of tracks matched to this jet be composed of tracks also associated with the
2508 primary vertex. Jet energies are also calibrated using p_T and η -dependent correction factors³. Fur-
2509 thermore, at least two jets must have $|\eta| < 2.5$ in order to be b -tagged. The MV1 algorithm^{?????}
2510 is used for b -tagging. Once jets are reconstructed and b -tag weights have been calculated, the two
2511 hardest, b -tagged jets are used as the telescoping jet axes. Additional details can be found in Ref.³.

2512 After the telescoping jet axes have been established, telescoping jets are constructed using topolog-
2513 ical clusters in the calorimeters at a variety of jet cone sizes. Including the original anti- k_t jets used for
2514 the $R = 0.4$ case, twelve total sets of jets of cone sizes ranging from $R = 0.4\text{--}1.5$ are constructed,
2515 with each successive size having a radius 0.1 larger than the preceding set. For each jet axis, telescop-
2516 ing jets consist of any topological cluster lying within R of the axis. In the event of overlap, clusters
2517 are assigned to the closer jet axis. If a given cluster is equidistant from the two jet axes, the cluster
2518 is assigned to whichever jet axis is associated with the anti- k_t jet with higher p_T . Calibration for the
2519 telescoping jets is conducted using corrections for anti- k_t calorimeter topological cluster jets; the

2520 $R = 0.4$ corrections are used for telescoping $R = 0.5$, and the $R = 0.6$ corrections are used for
2521 telescoping $R \geq 0.6$ (cf. Sec. B.4). The telescoping cone jets ($R \geq 0.5$) at reconstructed level are
2522 trimmed using Cambridge-Aachen jets with $R = 0.3$ and $f_{cut} = 0.05$ with respect to the untrimmed
2523 jet p_T^{40} . Since these jets are trimmed, the active area correction is not applied. In the event a Z can-
2524 didate electron falls within R of the axis of a telescoping jet, its 4-momentum is subtracted from that
2525 of the jet vectorially.

2526 A similar process is used to construct telescoping jets in the truth-level analysis below. Instead of
2527 the two hardest b -tagged anti- k_t with $R = 0.4$ jets reconstructed with calorimeter topological clus-
2528 ters, the two hardest truth b -jets in an event are used. Instead of making a cut on b -tagging weight
2529 to b -tag, truth jets are examined to see whether a b -hadron with $p_T > 5$ GeV is contained within
2530 $\Delta R < 0.4$ of the jet axis; the presence of a b -hadron is used to b -tag truth-level jets. These two jets
2531 again provide the jets for the $R = 0.4$ case and the axes around which telescoping takes place. Stable
2532 truth particles, not including muons and neutrinos, are used in place of calorimeter topological clus-
2533 ters. Z candidate electron-telescoping jet overlap removal is performed at truth level, too. Missing E_T
2534 is calculated using the vector sum of the four momenta of stable truth-level neutrinos. Since there
2535 are no pileup particles stored at truth level, truth-level telescoping jets are not trimmed.

2536 **B.3 EVENT RECONSTRUCTION AND SELECTION**

2537 Events are selected on the basis of a combination of leptonic, jet, and missing E_T requirements,
2538 which are outlined in Table B.1. Leptons are categorized by three sets of increasingly stringent qual-
2539 ity requirements, which include lower limits on E_T , upper limits on $|\eta|$, impact-parameter require-

2540 ments, and track-based isolation criteria. The requirements differ for electrons³ and muons³. Events
2541 are selected with a combination of single lepton, dielectron, and dimuon requirements. Each event
2542 must contain at least one lepton passing medium requirements and at least one other lepton pass-
2543 ing loose requirements. These leptons are used to create a dilepton invariant mass cut to ensure the
2544 presence of a Z boson and suppress multijet backgrounds.

2545 Event selection requirements are also imposed on the anti- k_t with $R = 0.4$ jets. There must be at
2546 least two b -tagged jets in a given event. The p_T of the harder b -tagged jet must be at least 45 GeV, and
2547 the second b -tagged jet must have p_T of at least 20 GeV. There are further topological cuts on the
2548 separation of the two jets $\Delta R(b, \bar{b})$, the distance between the two jets in the (η, ϕ) plane, according
2549 to the transverse momentum of the Z boson, p_T^Z . These are shown in Table B.2.

2550 The truth-level analysis has the same missing E_T , jet p_T , m_{ll} , and additional topological selection
2551 criteria, but the use of truth-level information simplifies the other requirements. Instead of lepton
2552 quality requirements, Z boson candidate leptons' statuses and MC record barcodes are checked to
2553 ensure the leptons are stable.

2554 In the jet calibration validation, the reconstructed level analysis lepton and m_{ll} requirements are
2555 imposed, but neither the missing E_T nor the jet selection requirements are applied so as not to bias
2556 the validation.

2557 B.4 VALIDATION OF JET CALIBRATION

2558 In order to validate the jet energy scale and resolution of jets constructed with this telescoping-jets
2559 algorithm, values of p_T^{rec}/p_T^{tru} are studied for each value of R for the $Z+jets$ MC sample. In a given

Table B.1: A summary of basic event selection requirements. Truth-level b -tagging is done with truth-level information.

Requirement	Reconstructed	Truth	Validation
Leptons	1 medium + 1 loose lepton	2 produced by Z boson	1 medium + 1 loose lepton
b -jet	2 b -tags	2 b -jets	—
p_T jet 1 (jet 2)	> 45 GeV (> 20) GeV		—
Missing E_T		$E_T^{\text{miss}} < 60$ GeV	—
Z boson		$83 < m_{ll} < 99$ GeV	

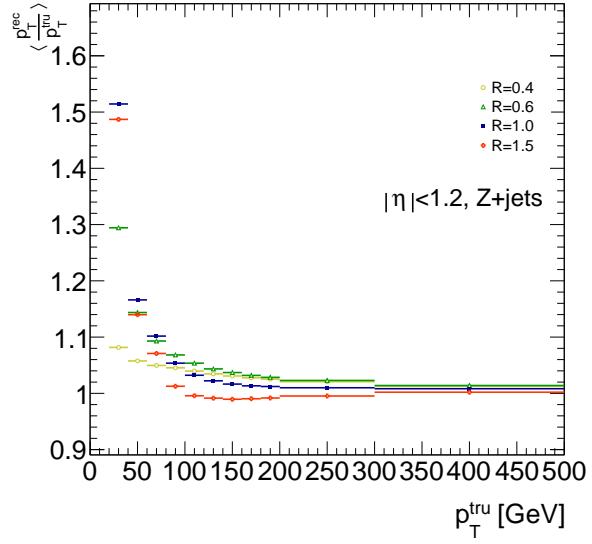
Table B.2: Topological requirements of the event selection.

p_T^Z [GeV]	$\Delta R(b, \bar{b})$
0–90	0.7–3.4
90–120	0.7–3.0
120–160	0.7–2.3
160–200	0.7–1.8
> 200	< 1.4

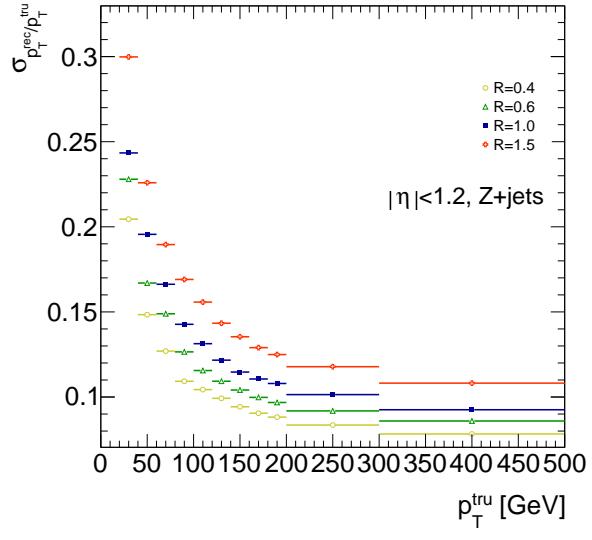
2560 event, all jets, not just the two hardest b -tagged jets, are telescoped. These jets are constructed in the
2561 same way as in the reconstructed and truth-level analyses; reconstructed-level jets are made from
2562 calorimeter topological clusters within R of the anti- k_t with $R = 0.4$ jet axes and then trimmed, and
2563 truth-level jets are made from stable truth particles within R of the anti- k_t with $R = 0.4$ jet axes.
2564 The reconstructed and truth-level telescoping jet ensembles are matched according to the separation
2565 in the (η, ϕ) plane of their corresponding anti- k_t with $R = 0.4$ jets used as seeds. Only jets with
2566 $|\eta| < 1.2$ are examined here, and the results of studies on the ZH_{125} , ZZ , and $t\bar{t}$ samples, as well as
2567 over other $|\eta|$ ranges, are outlined in²⁸. Any reconstructed jets not within $\Delta R = 0.3$ of a truth jet
2568 are discarded. In the event that multiple reconstructed jets are the same distance away from a given
2569 truth jet, the reconstructed jet with the highest p_T gets matched. Matching is retained for all R values
2570 (i.e. telescoping seeds are matched, and telescoping jets are assumed to match if the anti- k_t jets from
2571 which their seeds are derived match).

2572 Once anti- k_t with $R = 0.4$ reconstructed and truth jets are matched, response functions are cre-
2573 ated by generating a series of distributions of p_T^{rec}/p_T^{tru} in 20 GeV bins of p_T^{tru} from 20–200 GeV, one
2574 bin for 200–300 GeV, and one bin for 300–500 GeV for each R , with bins chosen for purposes of
2575 statistics. Ensembles with $p_T^{tru} < 20$ GeV are ignored since no calibration exists for jets with trans-
2576 verse momentum below this value. The values of $\langle p_T^{rec}/p_T^{tru} \rangle$ in each p_T^{tru} bin are calculated by doing
2577 a two sigma gaussian fit on the distribution of p_T^{rec}/p_T^{tru} in that bin and taking the mean of that fit,
2578 and the error on the mean is taken from the error of this parameter in the fit. The resolutions are the
2579 values of the square root of the variance on this fit. As the total response distributions in Figure B.1
2580 show, performance is best for low R values and high values of p_T^{tru} . Figure B.1 shows the $R = 0.4$

2581 (anti k_t) case to show a baseline for performance, $R = 0.6$ to show the deviations with “correct”
2582 calibrations, and $R = 1.0, 1.5$ to show how big those deviations get with larger R jets. The resolu-
2583 tions, $\sigma_{p_T^{rec}}/p_T^{tru}$, as a function of p_T^{tru} are shown in Figure B.1(b). For $p_T^{tru} > 60$ GeV, response is fairly
2584 consistent over various R values. Resolution, as might naïvely be expected, is worse for increasingly
2585 larger values of R . For $p_T^{tru} < 60$ GeV, resolution degrades, and response degrades in particular for
2586 increasing R ; this is likely a result from residual pileup effects.



(a)



(b)

Figure B.1: The mean and resolution of p_T^{rec}/p_T^{tru} for the background $Z + \text{jets}$ sample for $|\eta| < 1.2$ and for $R = 0.4, 0.6, 1.0$, and 1.5 in 20 GeV bins of p_T^{tru} for 20--200 GeV, one bin for 200--300 GeV, and one bin for 300--500 GeV, with bins chosen for purposes of statistics.

2587 **B.5 TRUTH-LEVEL ANALYSIS**

2588 To understand the limits and sources of any potential improvements, a truth-level analysis was con-
2589 ducted on MC samples with a ZH_{125} signal sample and a $Z+jets$ background sample. Distribu-
2590 tions for the dijet invariant mass, m_{bb} , were made for each telescoping radius.* Both signal and back-
2591 ground samples develop more pronounced tails in the high m_{bb} region as R increases, as shown in
2592 Figure B.2. N_{events} is normalized to expected values in data.

2593 One way to take advantage of this information is to make a cut on m_{bb} for two different radii.
2594 This is graphically depicted in Figure B.3 for the optimized combination of $m_{bb,R=0.9}$ (telescoping
2595 cone jets constructed as outlined in Sec. B.2) vs. $m_{bb,R=0.4}$ (anti- k_t jets). At truth-level, the majority
2596 of events in the signal ZH_{125} sample are concentrated in relatively narrow region of parameter space,
2597 where this is certainly not the case for the more diffuse $Z+jets$ background sample.

2598 Another way to take advantage of multiple event interpretations is to make use of an event's cut-
2599 weight, denoted z and defined as the fraction of interpretations in a given event that pass a certain set
2600 of cuts (in this note, a cut on m_{bb}). The distribution of cut-weights for a sample of events is denoted
2601 $\rho(z)$. To enhance the significance of a cut-based analysis, events can be weighted by the cut-weight
2602 or any function $t(z)$ of the cut-weight. Weighting events by $t(z)$ modifies the usual $S/\delta B$ formula
2603 used to calculate significances. In this note, δB is based on Poissonian statistics and is taken as $0.5 +$
2604 $\sqrt{0.25 + N_B}$, where N_B is the number of background events.

*Distributions for m_{bb} at truth and reconstructed level for all telescoping radii studied may be found in Appendix ??

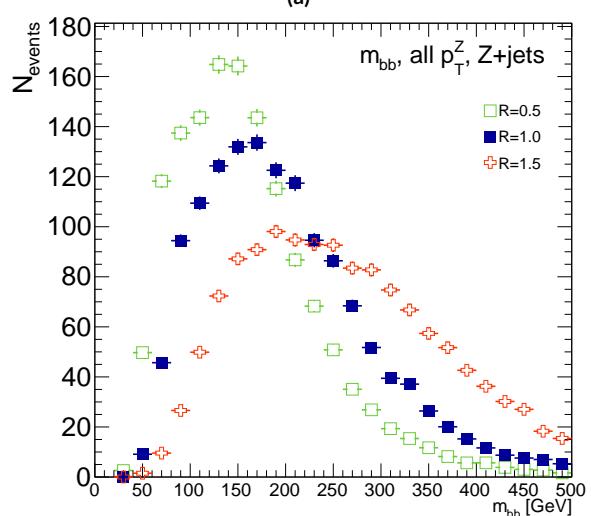
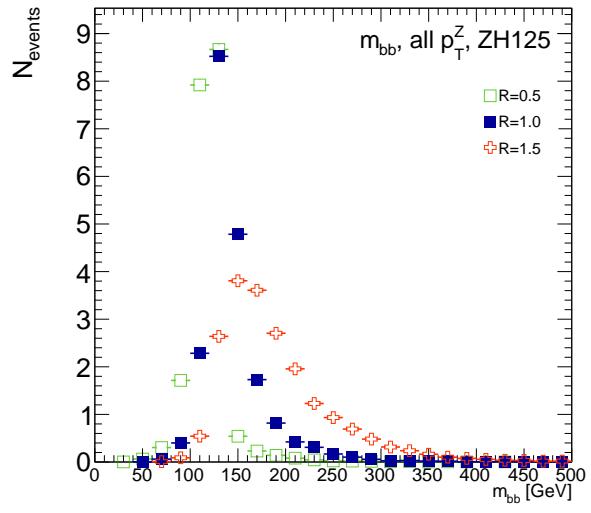
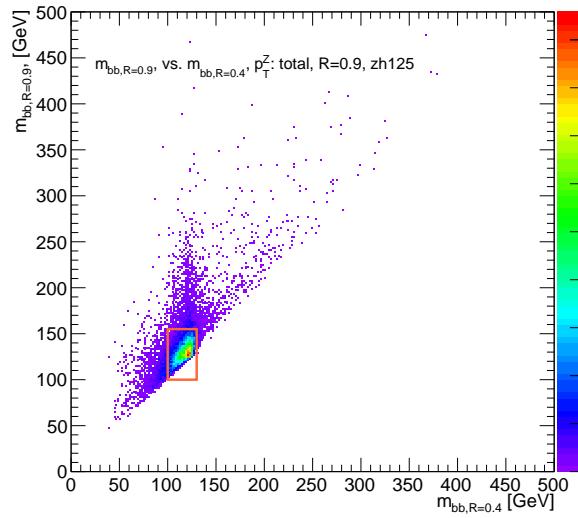
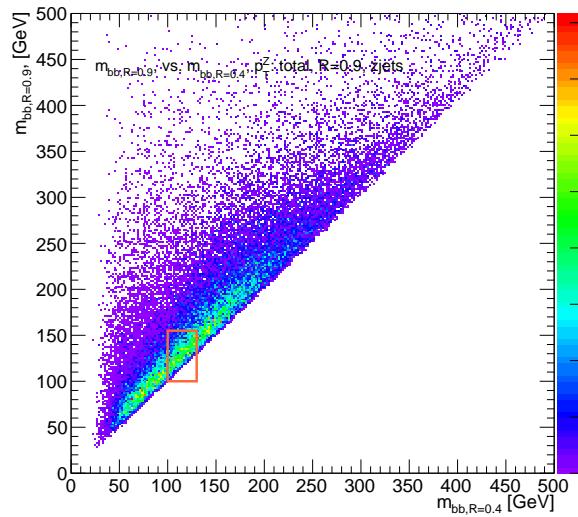


Figure B.2: The m_{bb} distribution for the telescoping jets with $R = 0.5, 1.0$, and 1.5 truth-level jets is shown for the signal and background samples in (a) and (b), respectively.



(a)



(b)

Figure B.3: The 2D distribution of $m_{bb,R=0.9}$ vs. $m_{bb,R=0.4}$ is shown for signal and background truth-level samples in (a) and (b), respectively. The region chosen for the double m_{bb} cut is outlined in orange.

2605 B.6 ERRORS ON TELESCOPING SIGNIFICANCES

2606 Significances of measurements are quoted in units of expected background fluctuations, schemati-
2607 cally, $S/\delta B$. For counting experiments with high numbers of events, we can use Gaussian statistics
2608 and express this as S/\sqrt{B} , which we here denote as \mathcal{S} . However, with lower statistics, it becomes
2609 more appropriate to use Poissonian statistics, and

$$\mathcal{S}_{gaus} = \frac{S}{\sqrt{B}} \rightarrow \mathcal{S}_{pois} = \frac{S}{0.5 + \sqrt{0.25 + B}} \quad (\text{B.1})$$

2610 where $0.5 + \sqrt{0.25 + B}$ is the characteristic upward fluctuation expected in a Poissonian data set
2611 using the Pearson chi-square test⁶⁶.

2612 B.7 COUNTING

2613 The significance is given as above, where $S = N_S$ and $B = N_B$. That is, the signal and background
2614 are just the number of events in signal and background that pass some cuts. The error for the Guas-
2615 sian case is the standard:

$$\Delta\mathcal{S}_{gaus} = \frac{1}{\sqrt{B}} \Delta S \oplus \frac{S}{2B^{3/2}} \Delta B \quad (\text{B.2})$$

2616 The error for the Poissonian case is:

$$\Delta\mathcal{S}_{pois} = \frac{1}{0.5 + \sqrt{0.25 + B}} \Delta S \oplus \frac{S}{2(0.5 + \sqrt{0.25 + B})^2 \sqrt{0.25 + B}} \Delta B \quad (\text{B.3})$$

2617 where \oplus denotes addition in quadrature, and $\Delta S(B)$ is the error on signal (background).

2618 B.8 MULTIPLE EVENT INTERPRETATIONS

2619 Using multiple event interpretations changes the formulae used in with simple counting. That is, S

2620 is not necessarily merely N_S , the number of events passing some signal cuts, and similarly for B and

2621 N_B . Using an event weighting by some function of the cut-weight, z , denoted $t(z)$, $S = N_S \langle t \rangle_{\rho_S}$

2622 and $B = N_B \langle t^2 \rangle_{\rho_B}$. So

$$\mathcal{S}_{t,gaus} = \frac{N_S \langle t \rangle_{\rho_S}}{\sqrt{N_B \langle t^2 \rangle_{\rho_B}}} \rightarrow \mathcal{S}_{t,pois} = \frac{N_S \langle t \rangle_{\rho_S}}{0.5 + \sqrt{0.25 + N_B \langle t^2 \rangle_{\rho_B}}} = \frac{N_S \int_0^1 dz t(z) \rho_S(z)}{0.5 + \sqrt{0.25 + N_B \int_0^1 dz t^2(z) \rho_B(z)}} \quad (\text{B.4})$$

For histograms, everything is done bin-wise. The notation used below is as follows: ρ_i is the

value of $\rho(z)$ at bin i (where the bins run from 0 to n_{tel} , where n_{tel} is the total number of telecop-

ing radii). $t_i = t_i(\rho_{S,i}, \rho_{B,i}, i/n_{tel})$ is the value of $t(z)$ at bin i , which can depend, in principle, on

$\rho_{S,i}$, $\rho_{B,i}$, and i/n_{tel} (the last of which is z in bin i). Explicitly,

$$N = \sum_{i=0}^{n_{tel}} \rho_i, \quad \int_0^1 dz t(z) \rho_S(z) = \sum_{i=0}^{n_{tel}} t_i \rho_{S,i}, \quad \int_0^1 dz t^2(z) \rho_B(z) = \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}$$

2623 For the calculations that follow, let $\xi = \sum_{i=0}^{n_{tel}} t_i \rho_{S,i}$, $\psi = 0.5 + \sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}$,

2624 $\partial_S = \frac{\partial}{\partial \rho_{S,i}}$ (and similarly for B), so $\mathcal{S}_t = N_S \xi / \psi$

2625 Some partial derivatives:

$$\partial_S N_S = 1, \quad \partial_{B,i} N_B = 1$$

$$\partial_S \xi = t_i + (\partial_S t_i) \rho_{S,i}, \quad \partial_B \xi = (\partial_B t_i) \rho_{B,i}$$

$$\begin{aligned}\partial_S \psi &= \frac{N_B t_i (\partial_S t_i) \rho_{B,i}}{\sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}}, & \partial_B \psi &= \frac{\sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i} + N_B (t_i^2 + 2t_i (\partial_B t_i) \rho_{B,i})}{2 \sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}} \\ \partial_S \mathcal{S}_t &= \frac{\xi}{\psi} + \frac{N_S}{\psi} \partial_S \xi - \frac{N_S \xi}{\psi^2} \partial_S \psi, & \partial_B \mathcal{S}_t &= N_S \left(\frac{1}{\psi} \partial_B \xi - \frac{\xi}{\psi^2} \partial_B \psi \right)\end{aligned}$$

2626 Thus,

$$\Delta \mathcal{S}_{t,i} = \left[\frac{\xi}{\psi} + \frac{N_S}{\psi} \partial_S \xi - \frac{N_S \xi}{\psi^2} \partial_S \psi \right] \Delta \rho_{S,i} \oplus N_S \left[\frac{1}{\psi} \partial_B \xi - \frac{\xi}{\psi^2} \partial_B \psi \right] \Delta \rho_{B,i} \quad (\text{B.5})$$

2627 and the total error is given by the sum in quadrature over all bins i of $\Delta \mathcal{S}_{t,i}$.

2628 B.9 $t(z) = z$

2629 With $t(z) = z$, $t_i = i/n_{tel}$, so $\partial_S t_i = \partial_B t_i = 0$. So:

$$\begin{aligned}\partial_S \psi &= \partial_B \xi = 0 \\ \partial_S \xi &= \frac{i}{n_{tel}} \\ \partial_B \psi &= \frac{\sum_i i^2 \rho_{B,i} + N_B i^2}{n_{tel} \sqrt{n_{tel}^2 + N_B \sum_i i^2 \rho_{B,i}}}\end{aligned}$$

²⁶³⁰ so $\Delta\mathcal{S}_{z,i}$ reduces to

$$\Delta\mathcal{S}_{t,i} = \left[\frac{\xi + N_S t_i}{\psi} \right] \Delta\rho_{S,i} \oplus \left[\frac{N_S \xi}{\psi^2} \partial_B \psi \right] \Delta\rho_{B,i} \quad (\text{B.6})$$

²⁶³¹ B.10 $t(z) = \rho_S(z) / \rho_B(z)$

²⁶³² With the likelihood optimized[†] $t^*(z) = \rho_S(z) / \rho_B(z)$, $t_i = \rho_{S,i} / \rho_{B,i}$, so $\partial_S t_i = 1 / \rho_{B,i}$ and

²⁶³³ $\partial_B t_i = -\rho_{S,i} / \rho_{B,i}^2$. So:

$$\begin{aligned} \partial_S \xi &= 2 \frac{\rho_{S,i}}{\rho_{B,i}} = 2t_i \\ \partial_B \xi &= -\frac{\rho_{S,i}}{\rho_{B,i}} = -t_i \\ \partial_S \psi &= \frac{N_B t_i}{\sqrt{0.25 + N_B \sum_i \rho_{S,i}^2 / \rho_{B,i}}} \\ \partial_B \psi &= \frac{\sum_i \rho_{S,i}^2 / \rho_{B,i} - N_B (\rho_{S,i} / \rho_{B,i})^2}{\sqrt{1 + 4N_B \sum_i \rho_{S,i}^2 / \rho_{B,i}}} \end{aligned}$$

²⁶³⁴ simplifying somewhat the terms in the per bin error in Equation B.6.

²⁶³⁵ The new significance figure using multiple event interpretations becomes, with ρ_S and ρ_B denot-
²⁶³⁶ ing the cut-weight distributions in signal and background, respectively

$$\frac{S}{\delta B} = \frac{N_S \langle t \rangle_{\rho_S}}{0.5 + \sqrt{0.25 + N_B \langle t^2 \rangle_{\rho_B}}} \quad (\text{B.7})$$

[†]for the Gaussian statistics case

2637 Of particular interest is the likelihood optimized $t(z)$,[‡] $t^*(z) = \rho_S(z) / \rho_B(z)$. m_{bb} windows are
 2638 chosen separately for each scheme studied to maximize total significances and are summarized in
 2639 Table B.3.

$$\left(\frac{S}{\delta B} \right)_z = \frac{N_S \epsilon_S}{0.5 + \sqrt{0.25 + N_B (\epsilon_B^2 + \sigma_B^2)}} \quad (B.8)$$

2640

$$\left(\frac{S}{\delta B} \right)_{t^*(z)} = \frac{N_S \int_0^1 dz \frac{\rho_S^2(z)}{\rho_B(z)}}{0.5 + \sqrt{0.25 + N_B \int_0^1 dz \frac{\rho_S^2(z)}{\rho_B(z)}}} \quad (B.9)$$

2641 where $\epsilon_{S,B}$ are the means of $\rho_{S,B}(z)$ and σ_B^2 is the variance of $\rho_B(z)$. Further details can be found in
 2642 Refs.^{28,39} and Appendix B.6.

Table B.3: m_{bb} windows studied. These windows were chosen to optimize significances over all p_T^Z .

Analysis Type	$S/\delta B$ Type	Optimal m_{bb} Window
Reconstructed	$anti-k_t R = 0.4$ $t(z) = z$ $t(z) = \rho_S(z) / \rho_B(z)$ $anti-k_t R = 0.4$, telescoping $R = 0.6$	$90-140$ GeV $110-155$ GeV $110-155$ GeV $95-140$ GeV ($R = 0.4$), $105-160$ GeV ($R = 0.6$)
Truth	$anti-k_t R = 0.4$ $t(z) = z$ $t(z) = \rho_S(z) / \rho_B(z)$ $anti-k_t R = 0.4$, telescoping $R = 0.9$	$100-130$ GeV $115-140$ GeV $120-135$ GeV $100-130$ GeV ($R = 0.4$), $100-155$ GeV ($R = 0.9$)

2643 The truth-level distributions $\rho_S(z)$, $\rho_B(z)$, and $\rho_S(z) / \rho_B(z)$ are shown for the m_{bb} window that
 2644 optimizes $(S/\delta B)_{t^*(z)}$ in Figure B.4, and significance improvements as a function of p_T^Z are summa-
 2645 rized in Figure B.5. Uncertainties in Figures B.5 and B.9 are statistical uncertainties. JES systematics

[‡]Derived under the assumption of Gaussian statistics in Ref³⁹

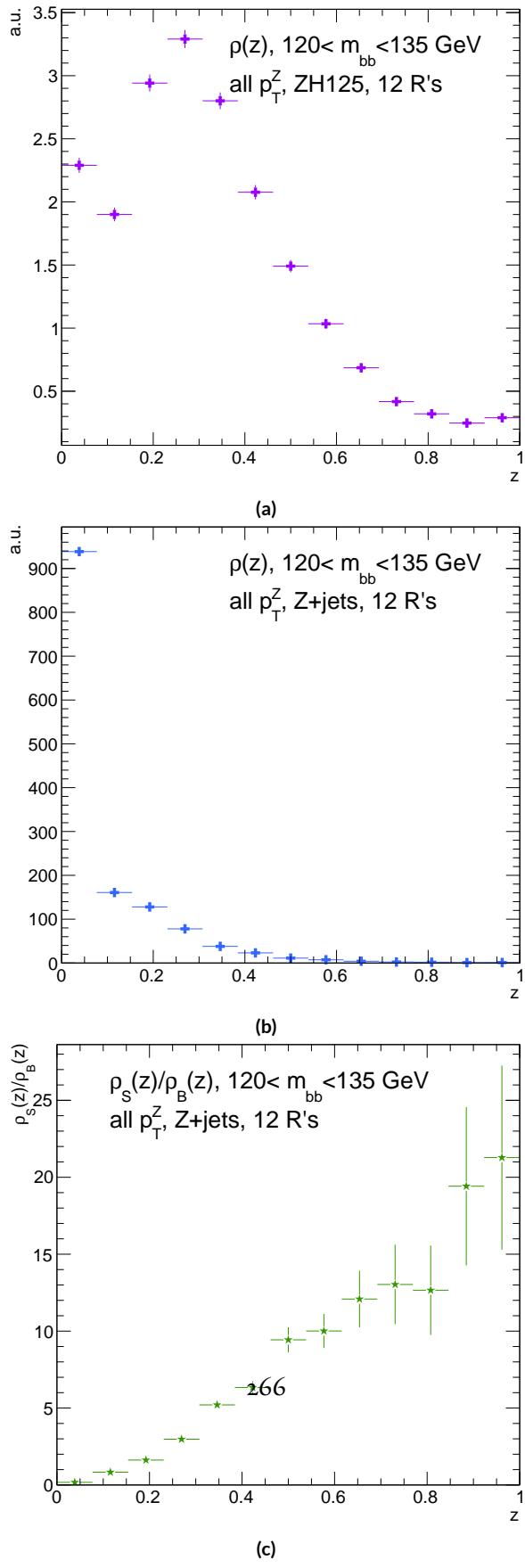


Figure B.4: Truth-level $\rho(z)$ distributions for the m_{bb} window optimizing $(S/\delta B)_{t^*(z)} \cdot \rho_S(z)$ for the signal $ZH125$ sample is shown in (a), and $\rho_B(z)$ for the background $Z+jets$ sample is shown in (b). The distribution of $\rho_S(z) / \rho_B(z)$

will need to be evaluated for different R 's, as modeling uncertainties is an outstanding issue, but
 these systematics will likely be strongly correlated for the different R 's and are not anticipated to be
 a very large contribution to total uncertainties. While the two dimensional m_{bb} cut and $t(z) = z$
 schemes only showed marginal improvement at truth level at 2.87%[§] and 1.45%, respectively, the
 likelihood optimized $t^*(z)$ showed a more substantial 40.7% improvement overall, with a steady in-
 crease in improvement with increasing p_T^Z . Figure B.5 (d) summarizes the improvements with respect
 to p_T^Z for the $t^*(z)$ event weight for five, seven, and twelve telescoping radii (interpretations) per
 event. Improvements increase with a greater number of interpretations and are more pronounced at
 higher p_T^Z for this scheme.[¶] The optimal $120 < m_{bb} < 135$ GeV window for $t^*(z)$ case is among the
 smallest studied. The benefits of this window's narrowness are suggested in Figure B.4. While the
 background cut-weight distribution, $\rho_B(z)$ in Figure B.4 (b) behaves as one might with a marked
 peak at $z = 0$, the signal $\rho_S(z)$ distribution peaks at a relatively modest $z = 0.3$, which indicates
 that much of the gain at truth level comes from background rejection. This is possible at truth level
 since there is both truth-level information available and no smearing and since ρ_S/ρ_B is the relevant
 quantity (as shown in Figure B.4 (c)).

[§]The limited improvement is provably due to the simplified treatment of the 2D case; better performance with a more sophisticated treatment has been observed in Ref.³¹.

[¶]This is not the case for the z event weight, which is more thoroughly treated in Section ??.

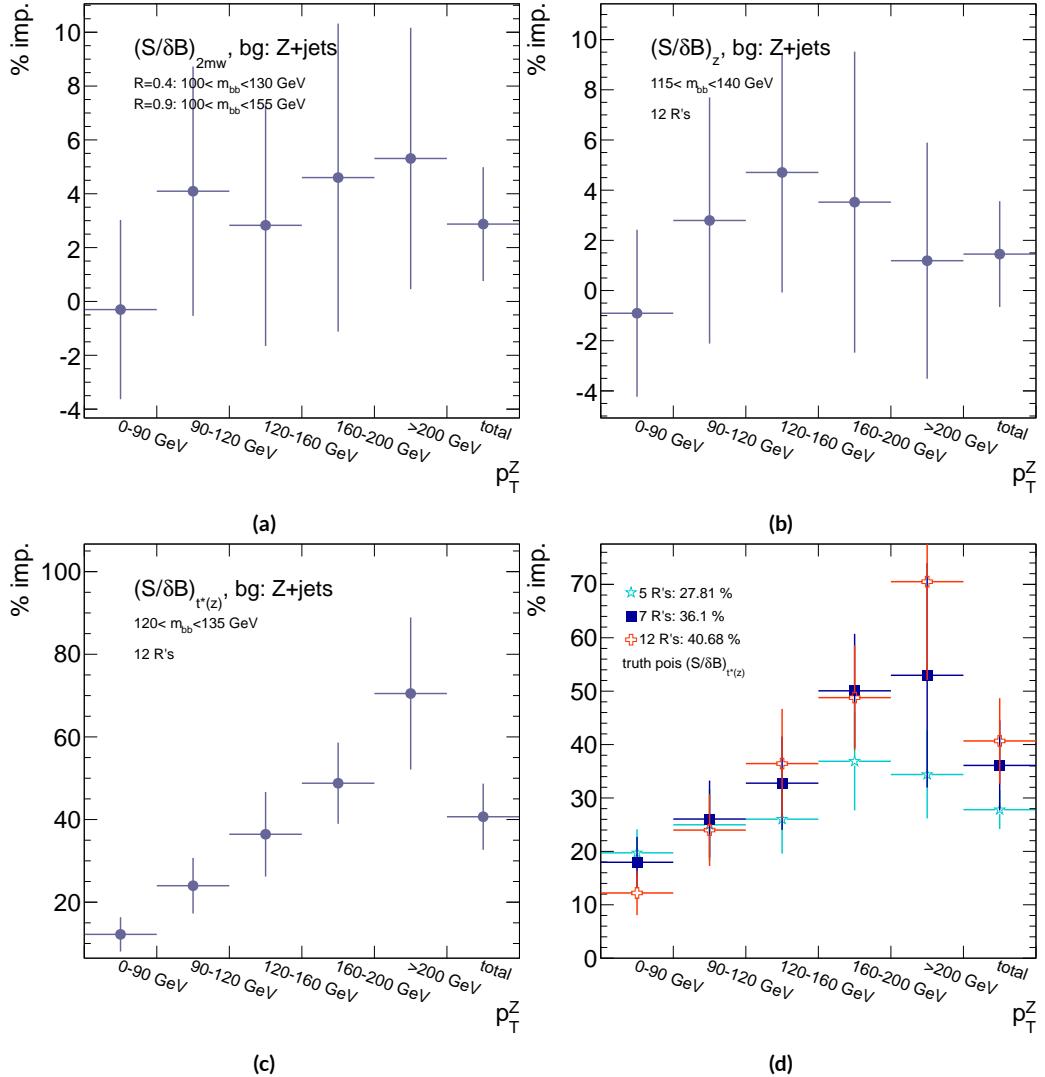


Figure B.5: A summary of the improvements for different truth-level telescoping jet cuts and weights is shown in bins of p_T^Z . The final bin is the total improvement over all p_T^Z . Shown are improvements for the 2D m_{bb} cut (a), $t(z) = z$ (b), $t(z) = t^*(z)$ with 12 radii (c), and $t(z) = t^*(z)$ for various radii (d).

2661 B.II RECONSTRUCTED-LEVEL ANALYSIS

2662 At reconstructed level, the same overall effect of introducing a high tail in m_{bb} distributions with
2663 increasing R is evident in comparing Figures B.2 and B.6. The optimal m_{bb} windows, however, grow
2664 larger, due to the lack of truth-level information.

2665 Total significance gains at reconstructed level for the two dimensional m_{bb} cut and the $t(z) = z$
2666 case are similar, at 2.87% and 1.45%, respectively. The optimal two-dimensional m_{bb} cut at recon-
2667 structed level is $95 < m_{bb,R=0.4} < 140$ GeV, $105 < m_{bb,R=0.6} < 160$ GeV. Just as at truth level, the
2668 $R = 0.4$ m_{bb} cut is comparable to the optimal single $R = 0.4$ m_{bb} cut, and the second m_{bb} cut is at
2669 similar values (cf. Table B.3 and Figures B.3 and B.7). However, the optimal second telescoping ra-
2670 dius is markedly smaller at $R = 0.6$ versus the optimal truth-level second radius of $R = 0.9$, which
2671 suggests that effects like pileup at reconstructed level obscure correlations between the $R = 0.4$
2672 interpretations and limit the usefulness of larger R interpretations in this particular scheme. The
2673 $t(z) = z$ case has a wider optimal window and yields about half the improvement it does at truth
2674 level.^{||}

2675 The optimal m_{bb} window for the $t^*(z)$ case is also markedly wider at reconstructed level, at
2676 $110 < m_{bb} < 155$ GeV in comparison to the truth-level optimal $120 < m_{bb} < 135$ GeV. The $\rho(z)$ dis-
2677 tributions for the signal ZH_{125} and background $Z + \text{jets}$ as well as the $\rho_S(z) / \rho_B(z)$ in this window
2678 are shown in Figure B.8. Compared with the truth-level distributions in Figure B.4, both the signal
2679 and background optimal $\rho(z)$ distributions have higher values at higher z . The peak in $\rho_S(z)$ at

^{||}A fuller treatment of this scheme is given in Section ??.

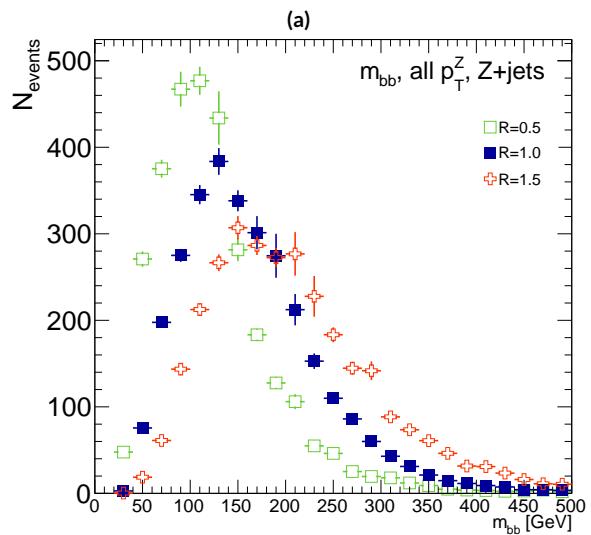
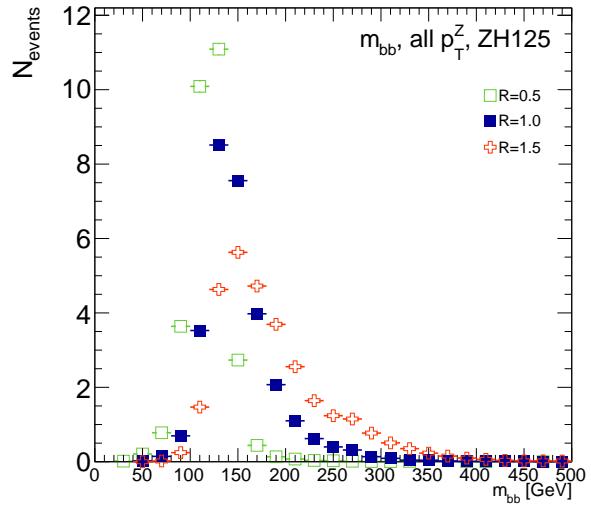


Figure B.6: The m_{bb} distribution for the telescoping jets with $R = 0.5$, $R = 1.0$, and $R = 1.5$ reconstructed-level jets is shown for the signal and background samples in (a) and (b), respectively.

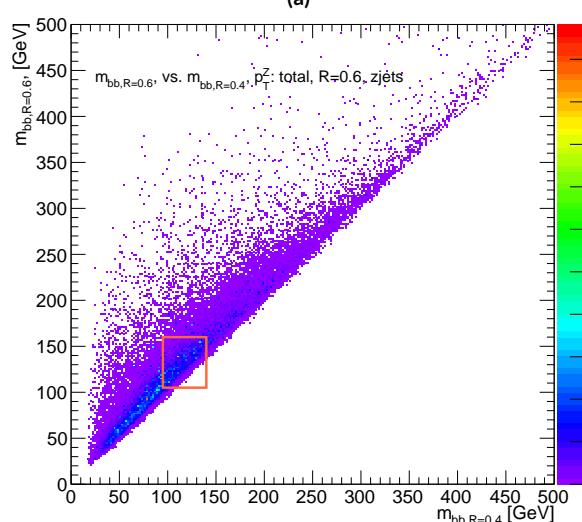
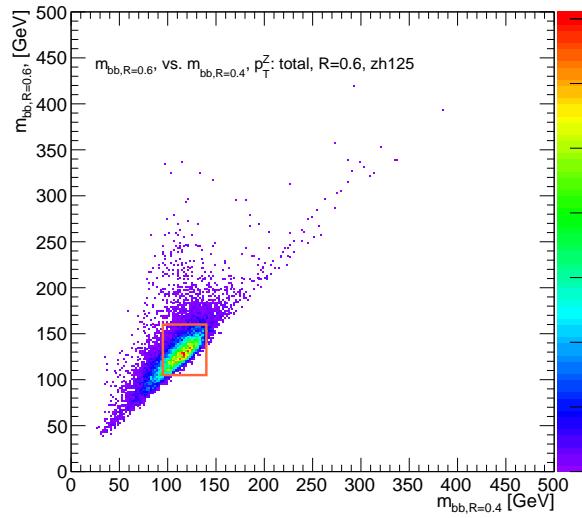


Figure B.7: The 2D distribution of $m_{bb,R=0.8}$ vs. $m_{bb,R=0.4}$ is shown for signal and background reconstructed-level samples in (a) and (b), respectively. The region chosen for the double m_{bb} cut is outlined in orange.

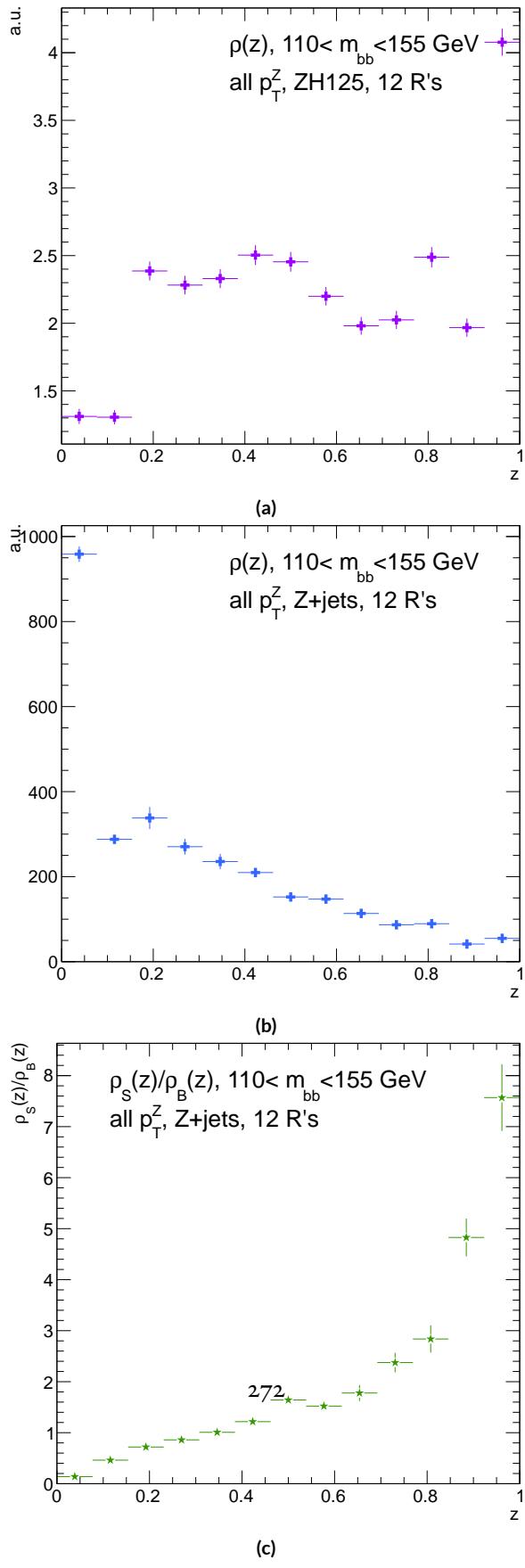


Figure B.8: Reconstructed-level $\rho(z)$ distributions for the m_{bb} window optimizing $(S/\delta B)_{t^*(z)} \cdot \rho_S(z)$ for the signal $Z\text{H}125$ sample is shown in (a), and $\rho_B(z)$ for the background $Z\text{+jets}$ sample is shown in (b). The distribution of $\rho_s(z)/\rho_B(z)$ for these samples is shown in (c).

2680 $z = 1$ suggests that at reconstructed level, maximizing the number of more “signal-like” events is the
 2681 key to optimizing significances, as opposed to the optimal, background suppressing $\rho(z)$ distribu-
 2682 tions at truth level. The use of a greater number of interpretations per event (telescoping radii) does
 2683 appear to result in overall greater improvement as at truth level, as twelve radii performed better
 2684 than five, but this is less clear at reconstructed level, as shown in Figure B.9 (d). The improvement at
 2685 reconstructed level using an event weight of $t^*(z)$ is 20.5%, just over half the improvement at truth
 2686 level but still quite significant. Summaries of improvements as a function of p_T^Z for all three cases
 2687 studied and for the $t^*(z)$ case for different numbers of telescoping radii are shown in Figure B.9.

Table B.4: A summary of significances for different weighting schemes and cuts and for reconstructed and truth jets for a luminosity of 20.3 fb^{-1} .

Type	0–90 GeV	90–120 GeV	120–160 GeV	160–200 GeV	> 200 GeV	total
anti- k_t , $R = 0.4_{rec}$	0.47492	0.28214	0.28339	0.25748	0.37337	0.76887
anti- k_t , $R = 0.4_{tru}$	0.57414	0.30655	0.37309	0.35042	0.53569	0.98619
$2 m_{bb,rec}$	0.48903	0.2858	0.28812	0.25972	0.38297	0.78611
$2 m_{bb,tru}$	0.5724	0.3191	0.38364	0.36655	0.56414	1.0145
z_{rec}	0.50698	0.27962	0.29937	0.25688	0.36846	0.79158
z_{tru}	0.56894	0.31511	0.39065	0.36277	0.54206	1.0005
$t^*(z)_{rec}$	0.55085	0.29931	0.33367	0.30107	0.51321	0.92649
$t^*(z)_{tru}$	0.64425	0.38008	0.50904	0.5214	0.91337	1.3873

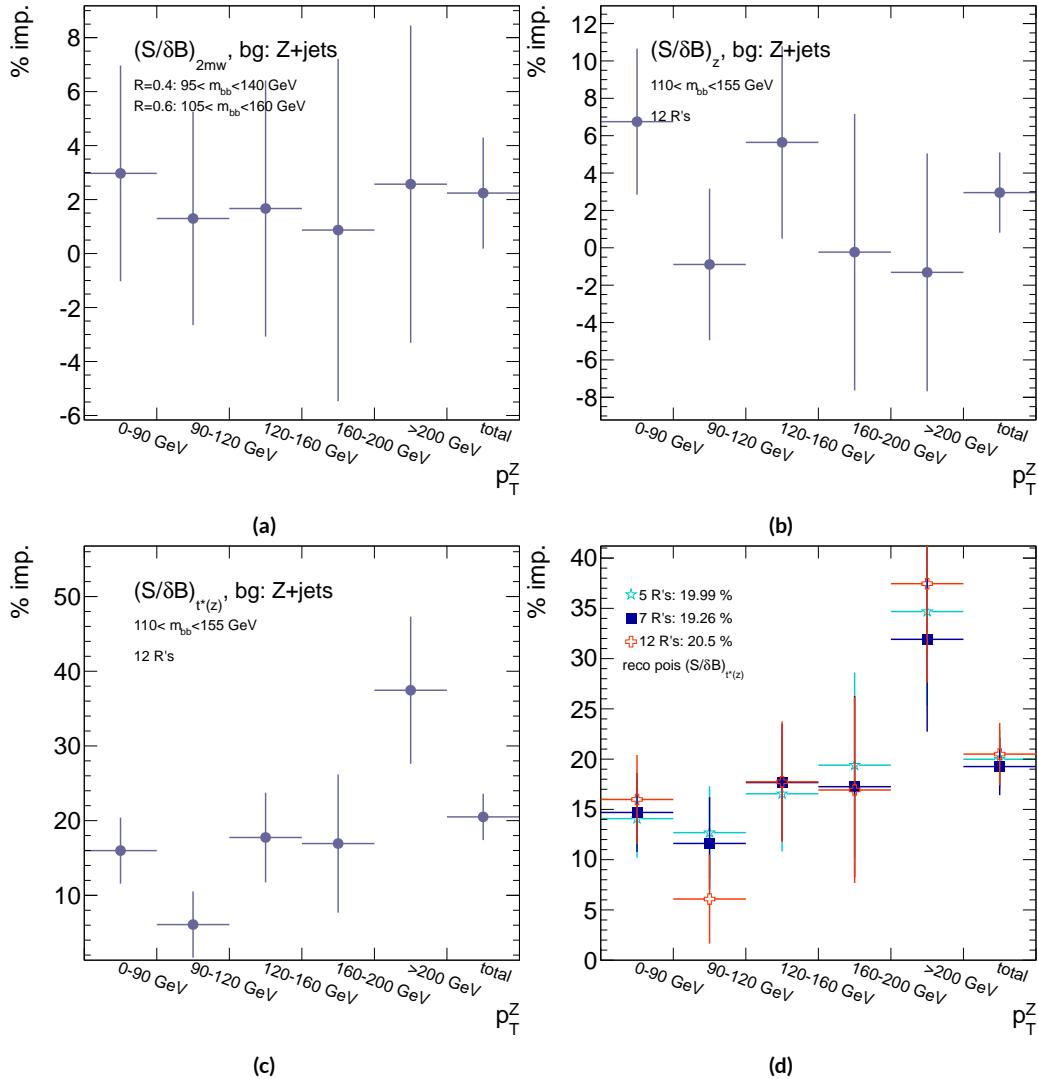


Figure B.9: A summary of the improvements for different reconstructed-level telescoping jet cuts and weights is shown in bins of p_T^Z . The final bin is the total improvement over all p_T^Z . Shown are improvements for the 2D m_{bb} cut (a), $t(z) = z$ (b), $t(z) = t^*(z)$ with 12 radii (c), and $t(z) = t^*(z)$ for various radii (d).

2688 B.12 CONCLUSIONS AND PROSPECTS

2689 The use of telescoping jets to provide multiple event interpretations shows promise as an avenue to
2690 increase significances in the $H \rightarrow b\bar{b}$ search in ATLAS and make an observation in the systematics-
2691 limited environment of early Run 2. A preliminary study using the telescoping jets algorithm with
2692 12 telescoping radii to build 12 event interpretations on 2012 Monte Carlo based on the full, cut-
2693 based Run 1 analysis yielded a 20.5% improvement in $S/\delta B$ over using anti- k_t with $R = 0.4$ alone
2694 at reconstructed level using a likelihood maximized event weighting to study the $ZH \rightarrow llb\bar{b}$ pro-
cess. The jets used in this note at reconstructed level were trimmed in order to guarantee reasonable
2695 resolution in the large- R interpretations. The algorithm, in particular, showed discriminating power
2696 at high p_T^Z , so better performance can be expected in Run 2 with a higher \sqrt{s} and higher numbers
2697 of events with large p_T^Z . Additionally, the many simplifying assumptions regarding jet calibration
2698 and the relatively basic use of information^{**} from multiple invariant masses in this note suggest
2699 that even further improvements than those quoted are possible. While this note did not explore
2700 the correlations between multiple event interpretations and the variables used in the BDT of the
2701 latest multivariate version of the $H \rightarrow b\bar{b}$ analysis¹³, new phenomenological studies suggest that
2702 such correlations are not strong³¹. The corresponding reconstructed-level study, using a BDT, is left
2703 for future work. Also left for future work are better understanding the effects of jet trimming and
2704 which interpretations are the most useful.

2705 **For examples of more sophisticated treatments compared to the treatment in this note, see Ref³¹.

References

- [1] (2012a). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716, 30–61.
- [2] (2012b). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys.Lett.*, B716, 1–29.
- [3] (2012). The Proton Synchrotron.
- [4] (2012). The Proton Synchrotron Booster.
- [5] (2012). The Super Proton Synchrotron.
- [6] (2014). *ATLAS Run 1 Pythia8 tunes*. Technical Report ATL-PHYS-PUB-2014-021, CERN, Geneva.
- [7] (2014). *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*. Technical Report ATLAS-CONF-2014-032, CERN, Geneva.
- [8] (2015). *Expected performance of the ATLAS b-tagging algorithms in Run-2*. Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva.
- [9] (2015). *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13 \text{ TeV}$* . Technical Report ATL-PHYS-PUB-2015-015, CERN, Geneva.
- [10] (2015). *Muon reconstruction performance in early $\sqrt{s}=13 \text{ TeV}$ data*. Technical Report ATL-PHYS-PUB-2015-037, CERN, Geneva.
- [11] A. L. Read (2002). Presentation of search results: the CL_s technique. *J. Phys. G*, 28, 2693–2704.
- [12] Aad, G. et al. (2014). Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7 \text{ TeV}$ with the ATLAS detector. *JHEP*, 09, 145.

- 2728 [13] Ahmadov, F., Alio, L., Allbrooke, B., Bristow, T., Buescher, D., Buzatu, A., Coadou, Y.,
 2729 Debenedetti, C., Enari, Y., Facini, G., Fisher, W., Francavilla, P., Gaycken, G., Gentil, J.,
 2730 Goncalo, R., Gonzalez Parra, G., Grivaz, J., Gwilliam, C., Hageboeck, S., Halladjian, G., Jack-
 2731 son, M., Jamin, D., Jansky, R., Kiuchi, K., Kostyukhin, V., Lohwasser, K., & Lopez Mateos,
 2732 D, e. a. (2014). *Supporting Document for the Search for the bb decay of the Standard Model*
 2733 *Higgs boson in associated (W/Z)H production with the ATLAS detector*. Technical Report
 2734 ATL-COM-PHYS-2014-051, CERN, Geneva.
- 2735 [14] Aliev, M., Lacker, H., Langenfeld, U., Moch, S., Uwer, P., et al. (2011). HATHOR: Hadronic
 2736 Top and Heavy quarks cross section calculator. *Comput.Phys.Commun.*, 182, 1034–1046.
- 2737 [15] Alwall, J., Frederix, R., Frixione, S., Hirschi, V., Maltoni, F., Mattelaer, O., Shao, H. S.,
 2738 Stelzer, T., Torrielli, P., & Zaro, M. (2014). The automated computation of tree-level and
 2739 next-to-leading order differential cross sections, and their matching to parton shower simula-
 2740 tions. *JHEP*, 07, 079.
- 2741 [16] Alwall, J., Herquet, M., Maltoni, F., Mattelaer, O., & Stelzer, T. (2011). Madgraph 5 : Going
 2742 beyond.
- 2743 [17] ATLAS Collaboration (2014). *Tagging and suppression of pileup jets with the ATLAS detec-*
 2744 *tor*. Technical Report ATLAS-CONF-2014-018, CERN, Geneva.
- 2745 [18] ATLAS Collaboration (2015a). MCPAnalysisGuidelinesMC15.
 2746 <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesMC15>.
- 2747 [19] ATLAS Collaboration (2015b). *Performance of missing transverse momentum reconstruction*
 2748 *for the ATLAS detector in the first proton-proton collisions at at $\sqrt{s} = 13 \text{ TeV}$* . Technical
 2749 Report ATL-PHYS-PUB-2015-027, CERN, Geneva.
- 2750 [20] Ball, R. D. et al. (2013). Parton distributions with LHC data. *Nucl. Phys.*, B867, 244–289.
- 2751 [21] Ball, R. D. et al. (2015). Parton distributions for the LHC Run II. *JHEP*, 04, 040.
- 2752 [22] Botje, M. et al. (2011). The PDF4LHC Working Group Interim Recommendations.
- 2753 [23] Buckley, A., Butterworth, J., Grellscheid, D., Hoeth, H., Lonnblad, L., Monk, J., Schulz, H.,
 2754 & Siegert, F. (2010). Rivet user manual.

- 2755 [24] Buzatu, A. & Wang, W. (2016). *Object selections for SM Higgs boson produced in associa-*
 2756 *tion with a vector boson in which $H \rightarrow b\bar{b}$ and V decays leptonically with Run-2 data: Object*
 2757 *support note for $VH(b\bar{b})$ 2015+2016 dataset publication.* Technical Report ATL-COM-PHYS-
 2758 2016-1674, CERN, Geneva. This is a support note for the $VH(b\bar{b})$ SM publication using the
 2759 2015+2016 datasets.
- 2760 [25] Campbell, J. M. & Ellis, R. K. (2010). MCFM for the Tevatron and the LHC. *Nucl. Phys.*
 2761 *Proc. Suppl.*, 205-206, 10–15.
- 2762 [26] Capeans, M., Darbo, G., Einsweiller, K., Elsing, M., Flick, T., Garcia-Sciveres, M., Gemme,
 2763 C., Pernegger, H., Rohne, O., & Vuillermet, R. (2010). *ATLAS Insertable B-Layer Technical*
 2764 *Design Report.* Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19.
- 2765 [27] CERN (2008). LHC first beam: a day to remember.
- 2766 [28] Chan, S., Huth, J., Lopez Mateos, D., & Mercurio, K. (2015a). *ZH → llb̄ Analysis with*
 2767 *Telescoping Jets.* Technical Report ATL-PHYS-INT-2015-002, CERN, Geneva.
- 2768 [29] Chan, S. K.-w. & Huth, J. (2017). *Variations on MVA Variables in the SM ZH → llb̄*
 2769 *Search.* Technical Report ATL-COM-PHYS-2017-1318, CERN, Geneva.
- 2770 [30] Chan, S. K.-w., Lopez Mateos, D., & Huth, J. (2015b). *Micromegas Trigger Processor Al-*
 2771 *gorithm Performance in Nominal, Misaligned, and Corrected Misalignment Conditions.*
 2772 Technical Report ATL-COM-UPGRADE-2015-033, CERN, Geneva.
- 2773 [31] Chien, Y.-T., Farhi, D., Krohn, D., Marantan, A., Mateos, D. L., & Schwartz, M. (2014).
 2774 Quantifying the power of multiple event interpretations.
- 2775 [32] Ciccolini, M., Dittmaier, S., & Kramer, M. (2003). Electroweak radiative corrections to
 2776 associated WH and ZH production at hadron colliders. *Phys. Rev.*, D68, 073003.
- 2777 [33] Clark, B., Lopez Mateos, D., Felt, N., Huth, J., & Oliver, J. (2014). *An Algorithm for Mi-*
 2778 *cromegas Segment Reconstruction in the Level-1 Trigger of the New Small Wheel.* Technical
 2779 Report ATL-UPGRADE-INT-2014-001, CERN, Geneva.
- 2780 [34] Collaboration, A. (2017a). Evidence for the $h \rightarrow b\bar{b}$ decay with the atlas detector.
- 2781 [35] Collaboration, A. (2017b). Study of the material of the atlas inner detector for run 2 of the
 2782 lhc.

- 2783 [36] Collaboration, T. A., Aad, G., Abat, E., Abdallah, J., & A A Abdelalim, e. a. (2008). The atlas
2784 experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08), S08003.
- 2785 [37] Cowan, G., Cranmer, K., Gross, E., & Vitells, O. (2011). Asymptotic formulae for likelihood-
2786 based tests of new physics. *Eur. Phys. J. C*, 71, 1554.
- 2787 [38] Czakon, M., Fiedler, P., & Mitov, A. (2013). Total Top-Quark Pair-Production Cross Section
2788 at Hadron Colliders Through $\alpha(\frac{4}{S})$. *Phys. Rev. Lett.*, 110, 252004.
- 2789 [39] D. Kahawala, D. Kahawala, D. K. & Schwartz, M. D. (2013). Jet sampling: Improving event
2790 reconstruction through multiple interpretations.
- 2791 [40] D. Krohn, D. Krohn, J. T. L. W. (2009). Jet trimming.
- 2792 [41] Delmastro, M., Gleyzer, S., Hengler, C., Jimenez, M., Koffas, T., Kuna, M., Liu, K., Liu, Y.,
2793 Marchiori, G., Petit, E., Pitt, M., Soldatov, E., & Tackmann, K. (2014). *Photon identification*
2794 *efficiency measurements with the ATLAS detector using LHC Run 1 data*. Technical Report
2795 ATL-COM-PHYS-2014-949, CERN, Geneva.
- 2796 [42] Evans, L. & Bryant, P. (2008). Lhc machine. *Journal of Instrumentation*, 3(08), S08001.
- 2797 [43] Gleisberg, T. et al. (2009a). Event generation with SHERPA 1.1. *JHEP*, 02, 007.
- 2798 [44] Gleisberg, T., Höche, S., Krauss, F., Schönherr, M., Schumann, S., Siegert, F., & Winter, J.
2799 (2009b). Event generation with sherpa 1.1. *Journal of High Energy Physics*, 2009(02), 007.
- 2800 [45] Hagebock, S. (CERN, Geneva, 2017). Lorentz Invariant Observables for Measurements of
2801 Hbb Decays with ATLS.
- 2802 [46] Heinemann, B., Hirsch, F., & Strandberg, S. (2010). *Performance of the ATLAS Secondary*
2803 *Vertex b-tagging Algorithm in 7 TeV Collision Data*. Technical Report ATLAS-COM-CONF-
2804 2010-042, CERN, Geneva. (Was originally 'ATL-COM-PHYS-2010-274').
- 2805 [47] Jackson, P. & Rogan, C. (2017). Recursive jigsaw reconstruction: Hep event analysis in the
2806 presence of kinematic and combinatoric ambiguities.
- 2807 [48] Kant, P., Kind, O., Kintscher, T., Lohse, T., Martini, T., Molbitz, S., Rieck, P., & Uwer,
2808 P. (2015). Hathor for single top-quark production: Updated predictions and uncertainty
2809 estimates for single top-quark production in hadronic collisions. *Computer Physics Commu-*
2810 *nifications*, 191, 74 – 89.

- 2811 [49] Lampl, W., Laplace, S., Lelas, D., Loch, P., Ma, H., Menke, S., Rajagopalan, S., Rousseau,
 2812 D., Snyder, S., & Unal, G. (2008). *Calorimeter Clustering Algorithms: Description and*
 2813 *Performance*. Technical Report ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003,
 2814 CERN, Geneva.
- 2815 [50] Lavesson, N. & Lonnblad, L. (2005). W+jets matrix elements and the dipole cascade.
- 2816 [51] LHC Higgs Cross Section Working Group, Dittmaier, S., Mariotti, C., Passarino, G., &
 2817 Tanaka (Eds.), R. (CERN, Geneva, 2011). Handbook of LHC Higgs Cross Sections: 1. In-
 2818 clusiv Observables. *CERN-2011-002*.
- 2819 [52] Loch, Peter and Lefebvre, Michel (2007). Introduction to Hadronic Calibration in ATLAS.
- 2820 [53] Luisoni, G., Nason, P., Oleari, C., & Tramontano, F. (2013). $H_W \pm/hz + o$ and 1 jet at nlo
 2821 with the powheg box interfaced to gosam and their merging within minlo. *Journal of High*
 2822 *Energy Physics*, 2013(10), 83.
- 2823 [54] M. Cacciari, M. Cacciari, G. P. S. & Soyez, G. (2008). The anti- k_t jet clustering algorithm.
 2824 *JHEP*, 0804.
- 2825 [55] Marcastel, F. (2013). CERN's Accelerator Complex. La chaîne des accélérateurs du CERN.
 2826 General Photo.
- 2827 [56] Masubuchi, T., Benitez, J., Bell, A. S., Argyropoulos, S., Arnold, H., Amaral Coutinho, Y.,
 2828 Sanchez Pineda, A. R., Buzatu, A., Calderini, G., & Chan, Stephen Kam-wah, e. a. (2016).
 2829 *Search for a Standard Model Higgs boson produced in association with a vector boson and de-*
 2830 *caying to a pair of b-quarks*. Technical Report ATL-COM-PHYS-2016-1724, CERN, Geneva.
- 2831 [57] Patrignani, C. et al. (2016). Review of Particle Physics. *Chin. Phys.*, C40(10), 100001.
- 2832 [58] Robson, A., Piacquadio, G., & Schopf, E. (2016). *Signal and Background Modelling Stud-*
 2833 *ies for the Standard Model VH, $H \rightarrow b\bar{b}$ and Related Searches: Modelling support note*
 2834 *for VH(bb) 2015+2016 dataset publication*. Technical Report ATL-COM-PHYS-2016-1747,
 2835 CERN, Geneva. This is a support note for the VH(bb) SM publication using the 2015+2016
 2836 datasets.
- 2837 [59] S. Alioli et al. (2009). NLO Higgs boson production via gluon fusion matched with shower
 2838 in POWHEG. *JHEP*, 0904, 002.

- 2839 [60] S. D. Ellis, S. D. Ellis, A. H. T. S. R. D. K. & Schwartz, M. D. (2012). Qjets: A non-
2840 deterministic approach to tree-based jet substructure. *Phys. Rev. Lett.*, 108.
- 2841 [61] Salam, G. P. (2009). Towards jetography.
- 2842 [62] Sjostrand, T., Mrenna, S., & Skands, P. Z. (2008). A Brief Introduction to PYTHIA 8.1.
2843 *Comput.Phys.Commun.*, 178, 852–867.
- 2844 [63] Stancari, G., Previtali, V., Valishev, A., Bruce, R., Redaelli, S., Rossi, A., & Salvachua Fer-
2845 rando, B. (2014). *Conceptual design of hollow electron lenses for beam halo control in the*
2846 *Large Hadron Collider*. Technical Report FERMILAB-TM-2572-APC. FERMILAB-TM-
2847 2572-APC, CERN, Geneva. Comments: 23 pages, 1 table, 10 figures.
- 2848 [64] Stewart, I. W. & Tackmann, F. J. (2011). Theory uncertainties for higgs and other searches
2849 using jet bins.
- 2850 [65] Symmetry Magazine (2015). The Standard Model of Particle Physics.
- 2851 [66] Verkerke, W. & Kirkby, D. (2003). The RooFit toolkit for data modeling. In *2003 Computing*
2852 *in High Energy and Nuclear Physics, CHEP03*.
- 2853 [67] Watts, G., Filthaut, F., & Piacquadio, G. (2015). *Extrapolating Errors for b-tagging*. Technical
2854 Report ATL-COM-PHYS-2015-711, CERN, Geneva. This is for internal information only, no
2855 approval to ever be seen outside of ATLAS.
- 2856 [68] Y. L. Dokshitzer, Y. L. Dokshitzer, G. D. L. S. M. & Webber, B. R. (1997). Better jet cluster-
2857 ing algorithms. *JHEP*, 9708.