

1 An MVA Theme and Variations on
2 $ZH \rightarrow llbb$ with the ATLAS Detector at
3 $\sqrt{s} = 13$ 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 thesis describes variations on the two lepton channel of the Run-2 search for the SM Higgs
23 boson produced in association with a vector boson using different variable sets for MVA training.
24 The three variable sets in question are the set of variables from the fiducial analysis, a set based on
25 the Lorentz Invariants (LI) concept, and a set based on a combination of masses and decay angles
26 derived using the RestFrames (RF) package. Aside from the variable sets used for MVA training and
27 discriminant distributions, the analysis is otherwise identical to the fiducial analysis. Both the LI
28 and RF sets perform competitively on the basis of significances, with the RF set showing a $\sim 3.5\%$
29 improvement in expected fits to Asimov and data, though neither set boosts observed significance.
30 Both sets also reduce the observed error on $\hat{\mu}$, with the LI set reducing the error due to systematics
31 by 7.5% and the RF set doing so by 16%.

Contents

32

33	o	INTRODUCTION	I
34	I	THE LARGE HADRON COLLIDER AND THE ATLAS DETECTOR	4
35	1.1	The CERN Accelerator Complex	5
36	1.2	The Large Hadron Collider	5
37	1.3	ATLAS at a Glance	II
38	1.4	The Inner Detector	16
39	1.5	The ATLAS Calorimeters	20
40	1.6	The Muon Spectrometer	28
41	2	THE STANDARD MODEL HIGGS AND COLLIDER EVENT VARIABLES	34
42	2.1	The Standard Model Higgs Boson	35
43	2.2	Higgs Boson Production and Decay at the Large Hadron Collider	38
44	2.3	Collider Events and Event Level Variables	41
45	2.4	Characterization with Event-Level Variables	43
46	2.5	Lorentz Invariants	45
47	2.6	RestFrames Variables	47
48	2.7	Extensions to the τ and μ Lepton Channels	49
49	3	DATA AND SIMULATED SAMPLES	53
50	4	SIGNAL AND BACKGROUND MODELING	55
51	4.1	Event Generation In a Nutshell	56
52	4.2	Description of Modeling Uncertainty Categories	58
53	4.3	Process Specific Systematic Summaries	67
54	5	OBJECT AND EVENT RECONSTRUCTION AND SELECTION	77
55	5.1	Triggers	78
56	5.2	Electrons	81
57	5.3	Muons	83
58	5.4	Missing Transverse Energy	85
59	5.5	Jets	85
60	5.6	Flavor Tagging	95
61	5.7	Miscellania and Systematics Summary	106
62	5.8	Event Selection and Analysis Regions	106

63	6 MULTIVARIATE ANALYSIS CONFIGURATION	109
64	6.1 Training Samples and Variable Selection	110
65	6.2 MVA Training	121
66	6.3 Statistics Only BDT Performance	126
67	7 STATISTICAL FIT MODEL AND VALIDATION	132
68	7.1 The Fit Model	133
69	7.2 Fit Inputs	136
70	7.3 Systematic Uncertainties Review	137
71	7.4 The VZ Validation Fit	139
72	7.5 Nuisance Parameter Pulls	143
73	7.6 Postfit Distributions	151
74	7.7 VH Fit Model Validation	158
75	8 FIT RESULTS	173
76	9 MEASUREMENT COMBINATIONS	179
77	9.1 The Combined Fit Model	180
78	9.2 Combined Fit Results	195
79	10 CLOSING THOUGHTS	203
80	APPENDIX A MICROMEGAS TRIGGER PROCESSOR SIMULATION	206
81	A.1 Algorithm Overview	207
82	A.2 Monte Carlo Samples	212
83	A.3 Nominal Performance	212
84	A.4 Fit Quantities	213
85	A.5 Efficiencies	219
86	A.6 Incoherent Background	222
87	A.7 BCID	227
88	A.8 Charge Threshold	227
89	A.9 Misalignments and Corrections	231
90	A.10 Individual Cases	236
91	A.11 $ds \neq 0$	237
92	A.12 $dz \neq 0$	237
93	A.13 $dt \neq 0$	237
94	A.14 $\alpha \neq 0$	238
95	A.15 $\beta \neq 0$	238
96	A.16 $\gamma \neq 0$	238
97	A.17 Simulation Correction of the Algorithm Under Nominal Conditions	241
98	A.18 Translation Misalignments Along the Horizontal Strip Direction (Δs)	244

99	A.19 Translation Misalignments Orthogonal to the Beamline and Horizontal Strip Direction (Δz)	246
100	A.20 Translation Misalignments Parallel to the Beamline (Δt)	248
101	A.21 Chamber Tilts Towards and Away from the IP (γ_s Rotation)	249
102	A.22 Rotation Misalignments Around the Wedge Vertical Axis (β_z)	251
103	A.23 Rotation Misalignments Around the Axis Parallel to the Beamline (α_t)	252
104	A.24 Conclusion	254
106	APPENDIX B TELESCOPING JETS	256
107	B.1 Monte Carlo Simulation	257
108	B.2 Jet Reconstruction and Calibration	258
109	B.3 Event Reconstruction and Selection	259
110	B.4 Validation of Jet Calibration	260
111	B.5 Truth-Level Analysis	265
112	B.6 Errors on Telescoping Significances	268
113	B.7 Counting	268
114	B.8 Multiple Event Interpretations	269
115	B.9 $t(z) = z$	270
116	B.10 $t(z) = \rho_S(z) / \rho_B(z)$	271
117	B.11 Reconstructed-Level Analysis	276
118	B.12 Conclusions and Prospects	282
119	REFERENCES	289

Acknowledgments

¹²² THIS THESIS WOULD NOT HAVE BEEN POSSIBLE without large amounts of espresso.

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

Zhuangzi

0

123

124

Introduction

125 SINCE THE DISCOVERY of a Standard Model (SM) like Higgs boson at the LHC in 2012⁷⁶, one of
126 the main outstanding physics goals of the LHC has been to observe the primary SM Higgs decay
127 mode, $H \rightarrow b\bar{b}$, with efforts primarily targeted at searching for Higgs bosons produced in associa-
128 tion with a leptonically decaying vector (W or Z , denoted generically as V) boson. As the integrated

¹²⁹ luminosity of data collected at the LHC increases, $H \rightarrow b\bar{b}$ searches will increasingly become limited
¹³⁰ by the ability to constrain systematic uncertainties, with the latest result from ATLAS at $\sqrt{s} = 13$
¹³¹ TeV using 36.1 fb^{-1} of pp collision data already approaching this regime, having a $VH(b\bar{b})$ signal
¹³² strength of $1.20^{+0.24}_{-0.23}(\text{stat.})^{+0.34}_{-0.28}(\text{syst.})$ at $m_H = 125 \text{ GeV}$ ⁴².

¹³³ While this effort will likely require a combination of several different methods at various differ-
¹³⁴ ent stages in the analysis chain, one possible avenue forward is to revise the multivariate analysis
¹³⁵ (MVA) discriminant input variables used, as various schemes offer the promise of reducing system-
¹³⁶ atic uncertainties through more efficient use of both actual and simulated collision data. This thesis
¹³⁷ discusses two such alternate MVA schemes, the RestFrames (RF) and Lorentz Invariants (LI) vari-
¹³⁸ ables, in the context of the 2-lepton channel of the Run 2 analysis in⁴² and⁶⁵, henceforth referred to
¹³⁹ as the “fiducial analysis,” before a brief discussion of combinations across channels and datasets.

¹⁴⁰ Data and simulation samples used are described in Chapter 3. Signal and background modeling
¹⁴¹ with accompanying systematics are defined in Chapter 4. Object and event reconstruction definitions
¹⁴² and event selection requirements are outlined in Chapter 5. The multivariate analysis, including a de-
¹⁴³ scription of the LI and RF variable sets and a summary of performance in the absence of systematic
¹⁴⁴ uncertainties, is described in Section 6. The statistical fit model and systematic uncertainties are de-
¹⁴⁵ scribed in Section 7, and the fit results may be found in Chapter 8. Combining channels and datasets
¹⁴⁶ at different \sqrt{s} values is discussed in the context of the Run 1 + Run 2 SM $VH(b\bar{b})$ combination in
¹⁴⁷ Chapter 9. Finally, conclusions and closing thoughts are presented in Chapter 10.

¹⁴⁸ Editorial notes:

¹⁴⁹ i. pdf will be *probability distribution* function

- 150 2. PDF will be *parton* distribution function
- 151 3. Unless otherwise stated, ATLAS and LHC/CERN images are from public available material
- 152 from experiment webpages. Copyright terms may be found here <https://atlas.cern/>
- 153 **copyright**.

Noli turbare circulos meos

Archimedes

1

¹⁵⁴

¹⁵⁵ The Large Hadron Collider and the ATLAS

Detector

¹⁵⁶

¹⁵⁷ THE CERN ACCELERATOR COMPLEX AND ITS EXPERIMENTS stand as a testament to human in-
¹⁵⁸ genuity and its commitment to the pursuit of fundamental knowledge. In this chapter, we give a

¹⁵⁹ cursory overview of the CERN accelerator complex, including the Large Hadron Collider (LHC),
¹⁶⁰ before moving on to a more detailed review of the ATLAS detector.

¹⁶¹ I.I THE CERN ACCELERATOR COMPLEX

¹⁶² The journey of protons from hydrogen canister to high energy collisions through the CERN ac-
¹⁶³ celerator complex, illustrated in Figure I.I, is also one through the history of CERN’s accelerator
¹⁶⁴ program. After being ionized in an electric field, protons are first accelerated in a linear accelera-
¹⁶⁵ tor, LINAC 2^{*}, to a kinetic energy of 50 MeV. From there, they are fed into the Proton Synchotron
¹⁶⁶ Booster[†], which further accelerates them to 1.4 GeV and, as its name implies, feeds them to the 628
¹⁶⁷ m Proton Synchotron (PS, 1959⁸) and up to 25 GeV. The penultimate stage is the 7 km Super Pro-
¹⁶⁸ ton Synchotron (SPS, 1976; responsible for the discovery of the W and Z bosons and the 1983 Nobel
¹⁶⁹ Prize¹⁰), which accelerates the protons to a kinetic energy of 450 GeV. Finally, these 450 GeV protons
¹⁷⁰ are injected into the LHC¹⁰, a proton-proton collider housed in the 27 km circumference tunnel
¹⁷¹ that housed the Large Electron Positron Collider (LEP) before its operations ceased in 2000.

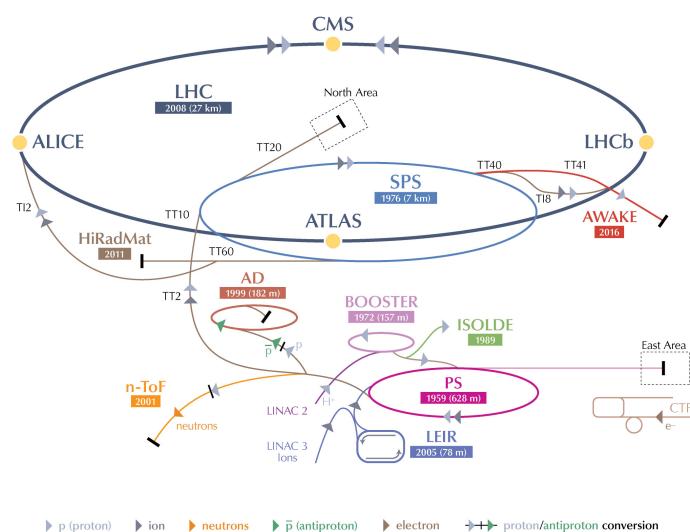
¹⁷² I.2 THE LARGE HADRON COLLIDER

¹⁷³ The LHC was designed to function primarily as a proton-proton collider with a center of mass en-
¹⁷⁴ 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
¹⁷⁵ of producing heavy ion (Pb-Pb) collisions, which it does for approximately one month in a typical

* 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 once⁹, 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 Device
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 Complex⁶⁴

¹⁷⁶ year of physics collisions. Owing to an accident at the beginning of the LHC's initial run, the accelerator has operated at center of mass energies of 7, 8, and now 13 TeV.

¹⁷⁸ The limited size of the LEP tunnel (\sim 3.6 m) means that it is impractical to have separate rings
¹⁷⁹ and magnet systems for each proton beam (proton-antiproton colliders like the Tevatron do not face
¹⁸⁰ this complication and can have both beams circulating in the same beam pipe), so the LHC magnets
¹⁸¹ are coupled in a "twin bore" design. The LHC magnets make use of superconducting NbTi cables
¹⁸² and are cooled using superfluid helium to a temperature of 2 K, which allows for operational field
¹⁸³ strengths in excess of 8 T. The layout of an LHC dipole magnet is shown in Figure 1.2. These dipole
¹⁸⁴ magnets are responsible for bending the LHC's proton beams, and their strength is the principal
¹⁸⁵ limiting factor in the center of mass energy achievable at a circular collider.

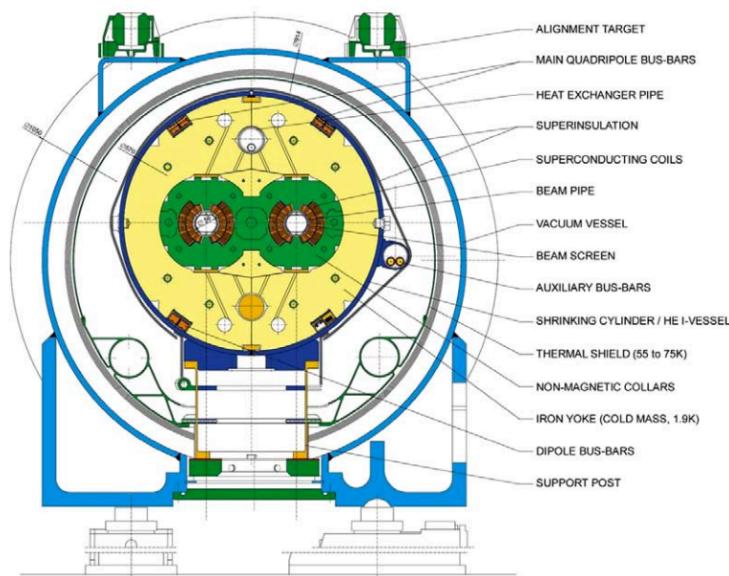


Figure 1.2: Schematic drawing of an LHC dipole magnet and cryogenics system.

¹⁸⁶ In addition to the dipole magnets, there are quadrupole magnet assemblies in the short straight
¹⁸⁷ sections (for beam focusing), as well as quadrupole, octupole, and sextupole magnets interspersed
¹⁸⁸ throughout the length of the LHC ring for beam stabilization and other higher order corrections.

¹⁸⁹ The interior of the LHC beam pipe operates at a nominal pressure of $\sim 10^{-7}$ Pa, famously more
¹⁹⁰ rarefied than outer space.

¹⁹¹ The LHC ring itself is between 45 m and 170 m below ground and has a 1.4% incline towards Lac
¹⁹² Léman with eight arcs and eight straight sections. In the middle of each of the eight straight sections,
¹⁹³ there are potential interaction points (each colloquially referred to by its number as “Point N ”),
¹⁹⁴ with each point housing either accelerator infrastructure or an experiment. A schematic of the con-
¹⁹⁵ tents of each component, as well as a more detailed view of the infrastructure in the LHC ring, can
¹⁹⁶ be found in Figure 1.3.

6

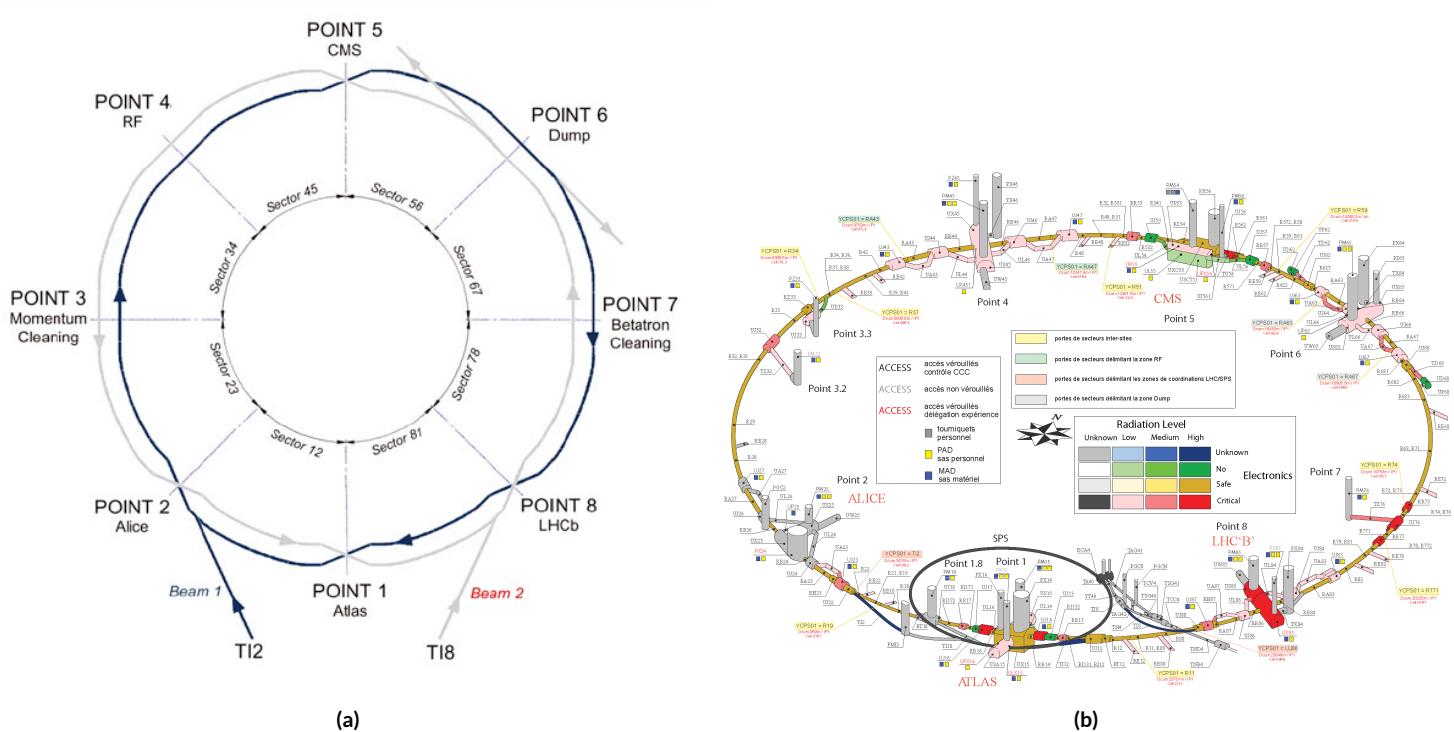


Figure 1.3: Schematic and detailed views of the LHC ring. IC.^{35 72},

197 Points 1, 2, 5, and 8 house the LHC’s experiments, ATLAS (*A Toroidal LHC ApparatuS*, one
198 of the two general purpose detectors, discussed in detail below), ALICE (A Large Ion Collider Ex-
199 periment, a dedicated heavy ion experiment), CMS (Compact Muon Solenoid, the other general
200 purpose detector), and LHCb (LHC beauty, a *B* physics experiment), respectively. Point 3 houses
201 a series of collimators that scatter and absorb particles in the beam with a large momentum devia-
202 tion from other particles in the beam (“momentum cleaning”), while Point 7 has a similar setup to
203 remove particles with large betatron amplitudes (“betatron cleaning”). Betatron amplitudes are re-
204 lated to how well focused beams are and can be thought of as giving a characteristic size for a beam;
205 just as one wants to screen out particles deviating in physical space, one also wants protons in the
206 beam to have nearly identical momentum. Well focused beams in both position and momentum
207 space are crucial to high quality collisions. Point 4 contains the LHC’s RF (radio frequency; 400
208 MHz) acceleration system, responsible for taking protons from their injection energy of 450 GeV to
209 their collision energy of 3.5, 4, 6.5, or 7 TeV. Point 6 is where the energetic ionizing radiation of cir-
210 culating beams can be safely taken out of the collider into a block of absorbing material, either at the
211 end of a data-taking run or in the event of an emergency (in the event of irregular behavior, it is es-
212 sential to do this as quickly as possible to minimize damage to the accelerator and to experiments);
213 this is known as a “beam dump.”

214 1.3 ATLAS AT A GLANCE

215 1.3.1 COORDINATES AND DISTANCES IN THE ATLAS DETECTOR

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

221 The ATLAS coordinate system is shown in Figure 1.4. It is a right-handed coordinate system cen-
222 tered at the nominal collision point, with the x axis pointing towards the center of the LHC ring,
223 the z axis pointing up, and the y axis completing the right-handed coordinate system.

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

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

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

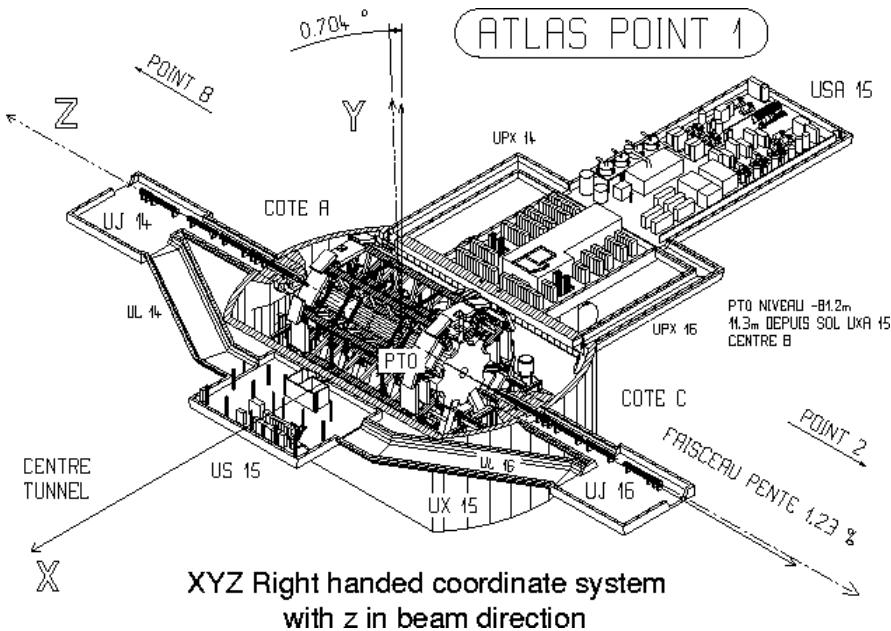


Figure 1.4: The ATLAS coordinate system. "A" side is the airport, and "C" side is "Charlie's," a pub in Saint-Genis, France.

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

233 rapidity of a particle:

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

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

235 is the relativistic analog of a rotation angle; boosts can be added in a manner similar to rotations⁴,

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

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

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

⁴Generally, 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.

239 momentum is still conserved in the transverse plane, so quantities like transverse momentum \vec{p}_T
 240 or energy (E_T)^{||} are often very useful in analysis. However, in the massless limit^{**}, we can take $E =$
 241 $\sqrt{\vec{p}_T^2 + p_z^2}$. Hence, with θ taken as the zenith angle and o corresponding to the $+z$ direction, for a
 242 massless particle, $p_z = E \cos \theta$. Using the usual half angle formula $\cos \theta = (1 - \tan^2 \theta) / (1 + \tan^2 \theta)$

243

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

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

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

245 Lower values of $|\eta|$ (1.3) correspond to more central areas of the detector known as the “barrel,”
 246 with the typical layout here being concentric, cylindrical layers. Larger values of $|\eta|$ (to ~ 2.5 for
 247 some systems and up to as much as $\sim 4.5 - 5$ for others) are known as the “end caps,” where ma-
 248 terial is typically arranged as disks of equal radius centered on the beam pipe stacked to ever greater
 249 values of $|z|$. This terminology will be useful when discussing the various subsystems of the ATLAS
 250 detector. Since decay products from a collision propagate radially (in the calorimeter portions of

^{||}Energy is not a vector quantity, but one can take the scalar or vectorial 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 ~ 10 's of GeV, making for a boost of roughly 100.

251 the detector with no magnetic field), the radial coordinate is not so important for composite physics
252 objects like electrons or jets, which are typically expressed as momentum 4-vectors. Hence, η and ϕ
253 are often the only useful spatial coordinates. Distances between objects are often expressed not as a
254 difference in solid angle, but as a distance, ΔR , in the $\eta - \phi$ plane, where

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

255 Two important concepts when discussing particles traveling through matter (e.g. particle detec-
256 tors) are radiation lengths and (nuclear) interaction lengths, which characterize typical lengths for
257 the energy loss of energetic particles traveling through materials. In general, the energy loss is mod-
258 eled as an exponential

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

259 where E_0 is the initial energy, and L is a characteristic length. These lengths depend both on the in-
260 cident particle and the material through which they pass. In the case of uniform, composite mate-
261 rials, the length may be found by calculating the reciprocal of the sum of mass fraction weighted
262 reciprocal characteristic lengths of the components. This formula works quite well for modeling the
263 very regular behavior of electromagnetic showers (energetic photons convert into electron/positron
264 pairs, which emit photons...). In this case, L is denoted X_0 ; this is the radiation length. Hadronic
265 showers are far more complicated, with shower multiplicity and makeup being much more vari-
266 able^{††}. Nevertheless, a characteristic length can be tabulated for a standard particle type, typically

^{††}Different initial hadrons will shower very differently, and hadronic showers will have phenomena like

²⁶⁷ pions, and is called the nuclear interaction length.

²⁶⁸ **I.3.2 GENERAL LAYOUT OF ATLAS**

²⁶⁹ The ATLAS detector and its main components are shown in Figure 1.5. ATLAS is designed as a
²⁷⁰ largely hermetic detector, offering full coverage in ϕ and coverage in $|\eta|$ up to 4.7. The multiple sub-
²⁷¹ systems allow for good characterization of the decay products from collisions in the LHC. The in-
²⁷² nermost system is the inner detector (ID); composed primarily of silicon pixels and strips immersed
²⁷³ in a magnetic field, it is designed to reconstruct the curved trajectories of charged particles produced
²⁷⁴ in collisions while taking up as little material as possible.

²⁷⁵ Surrounding the ID is the liquid argon based electromagnetic calorimeter (ECAL), which is de-
²⁷⁶ signed to capture all of the energy of the electromagnetic showers produced by electrons and pho-
²⁷⁷ tons coming from particle collisions. The ECAL is in turn encapsulated by a scintillating tile and
²⁷⁸ liquid argon based hadronic calorimeter (HCAL) that captures any remaining energy from the jets
²⁷⁹ produced by hadronizing quarks and gluons.

²⁸⁰ The outermost layer of ATLAS is the muon spectrometer (MS), which has its own magnetic field
²⁸¹ produced by toroidal magnets. Muons are highly penetrating particles that escape the calorimeters
²⁸² with most of their initial momentum, so the MS and its magnets are designed to curve these charged
²⁸³ particles and measure their trajectories to measure their outgoing momenta. Each of these detector
²⁸⁴ systems has several principal subsystems and performance characteristics, which will be described in
²⁸⁵ turn below.

neutral pions converting to photons (which then shower electromagnetically), making them much trickier to deal with.

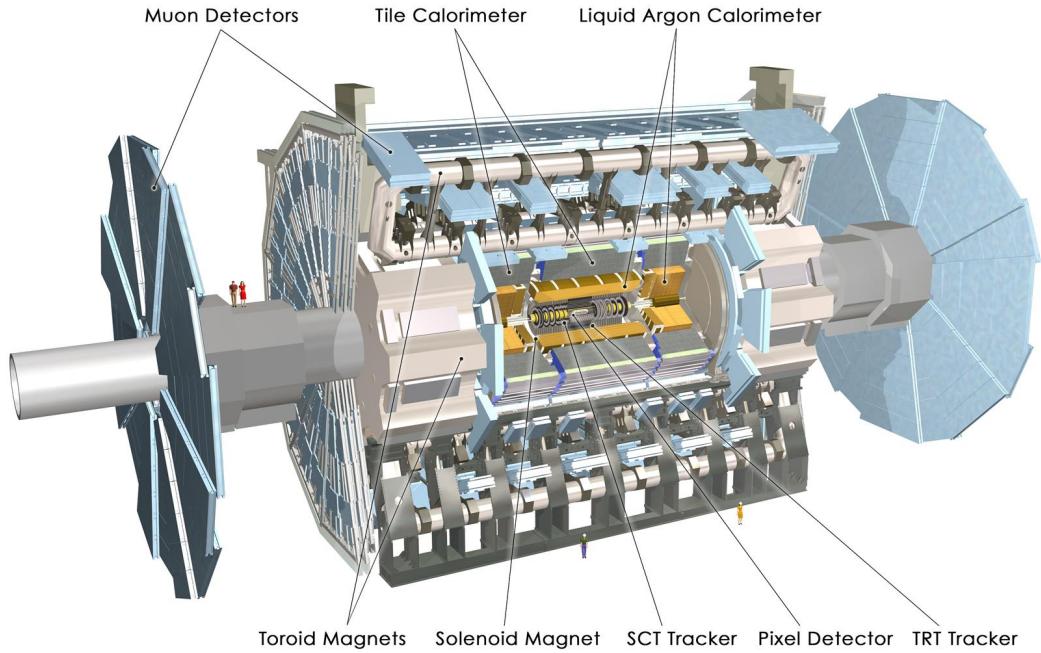


Figure 1.5: The ATLAS detector with principal subsystems shown.

²⁸⁶ 1.4 THE INNER DETECTOR

²⁸⁷ ATLAS's inner detector (ID) is surrounded by a 2 T superconducting solenoid that is cryogenically
²⁸⁸ cooled to a temperature of 4.5 K. The ID uses two silicon detector subsystems (the Pixel and Semi-
²⁸⁹ Conductor (strip) Tracker (SCT)) to track the curved trajectories of charged particles emanating
²⁹⁰ from particle collisions and a Transition Radiation Tracker (TRT) composed of gas straw detectors
²⁹¹ with filaments for e/π discrimination, as shown in Figure 1.9. The ID offers full coverage in ϕ and
²⁹² extends to an $|\eta|$ of 2.5.

²⁹³ Since the components of the ID do not provide an energy measurement, it is desirable for a track-
²⁹⁴ ing system to have as small a material budget as possible so that more accurate energy measurements

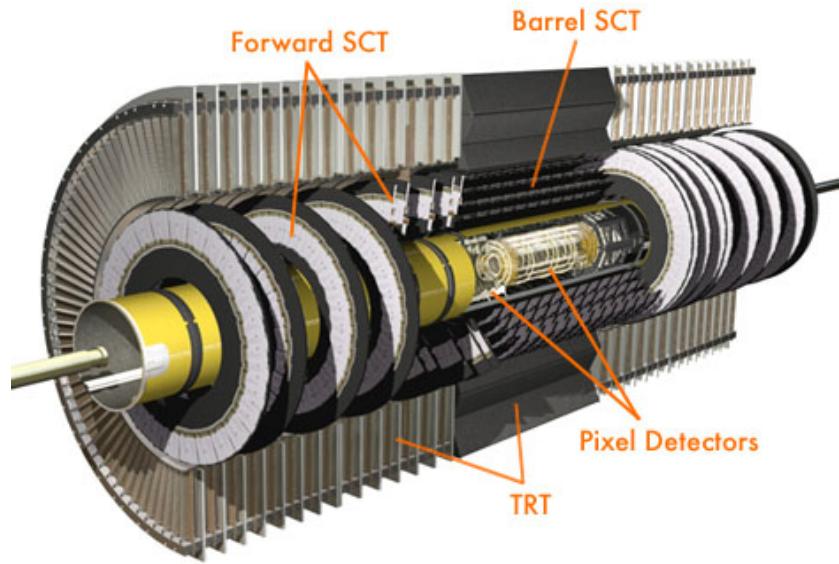


Figure 1.6: The ATLAS inner detector. IC:⁴⁴

295 may be done in the calorimeters. Generally, there are two radiation lengths in the inner detector (the
 296 precise figure varies with η); the full material budget, with the layout of the individual layers in each
 297 subsystem, can be seen in Figure 1.8.

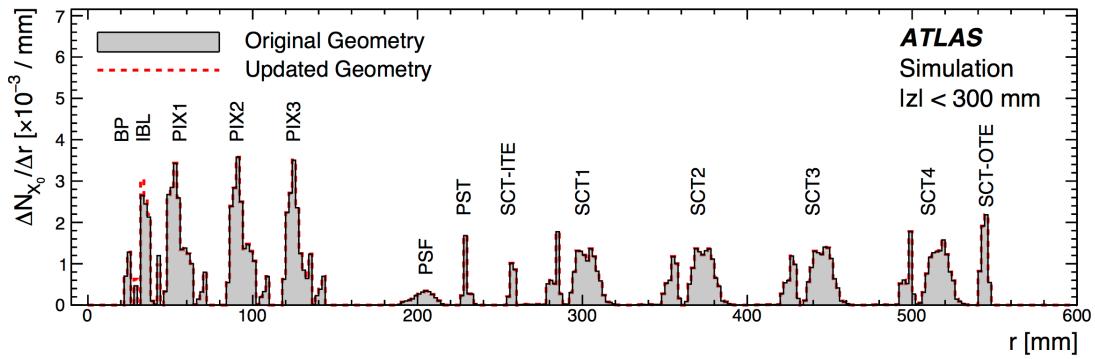


Figure 1.7: The ID material budget. IC:⁴³

²⁹⁸ **I.4.1 THE PIXEL DETECTOR**

²⁹⁹ The innermost part of ATLAS is the Pixel Detector, which, as the name suggests, is comprised of
³⁰⁰ four layers of silicon pixels in the barrel at 32, 51, 89, and 123 mm from the beam pipe, and three lay-
³⁰¹ ers in the end caps at 495, 580, and 650 mm from the beam pipe, with over 80 million channels total.

³⁰² The innermost layer of pixels, the insertable *B* layer (IBL) was installed during the 2013–14 LHC
³⁰³ shutdown. The pixels are cooled to a temperature of $\sim -5^\circ\text{C}$, with N_2 gas and operate at 150–600
³⁰⁴ V. The pixels themselves come in two sizes $50 \times 400(600) \times 250 \mu\text{m}$, with the larger pixels in the
³⁰⁵ outer layers. They provide nominal resolution of $10(115) \mu\text{m}$ resolution in $r - \phi(z)$ direction.

³⁰⁶ In order to improve total coverage in the detector and prevent any gaps, pixels are not installed
³⁰⁷ flush with each other. Pixels in the barrel are tilted at about 20° , with an overlap in $r - \phi$, as shown
³⁰⁸ in Figure I.8. The disks of the ID end caps are rotated with respect to each other by 3.75° .

³⁰⁹ **I.4.2 THE SILICON MICROSTRIP DETECTOR (SCT)**

³¹⁰ The layout of the SCT is similar to that of the Pixel detector, except that, for cost considerations, the
³¹¹ SCT uses silicon strips. These strips are also cooled to $\sim -5^\circ\text{C}$ with N_2 gas and operate from 150–
³¹² 350 V. Strip dimensions are $80 \times 6000 \times 285 \mu\text{m}$, and provide nominal $17(580) \mu\text{m}$ resolution in
³¹³ $r - \phi(z)$. Barrel strips feature an 11° tilt and come in four layers at 299, 371, 443, and 514 mm. There
³¹⁴ are nine end cap disks on each side at z values varying from 934–2720 mm.

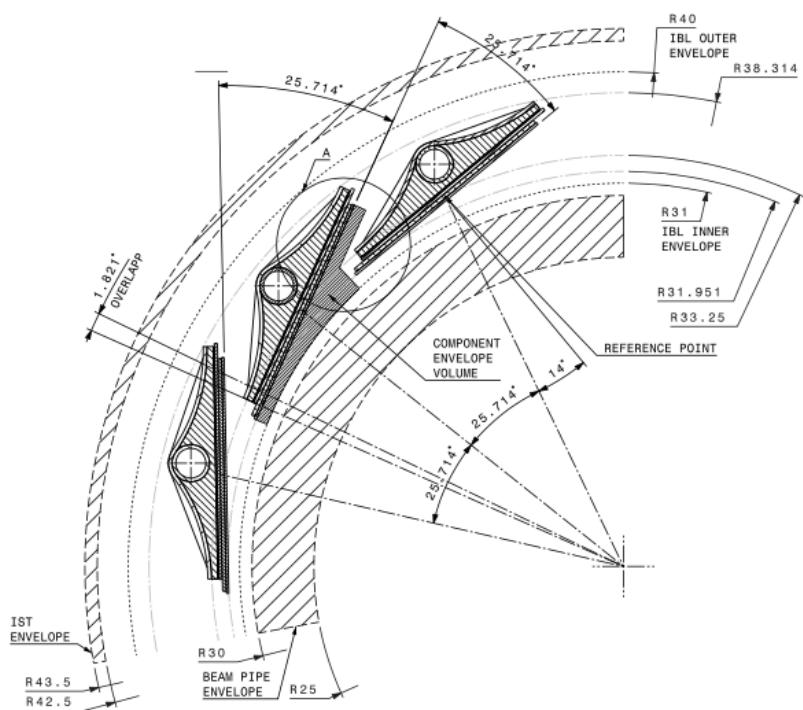


Figure 1.8: Arrangement of pixels in the barrel. IC:³⁴

315 1.4.3 TRANSITION RADIATION TRACKER (TRT)

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

326 1.5 THE ATLAS CALORIMETERS

327 ATLAS has four main calorimeter systems: the liquid argon based Electromagnetic Calorimeter
328 (ECAL), the Hadronic End Cap (HEC), the Forward Calorimeters (FCAL), and the scintillating
329 tile based hadronic Tile Calorimeter in the barrel. Their layout and material budget in interaction
330 lengths can be seen in Figure 1.10.

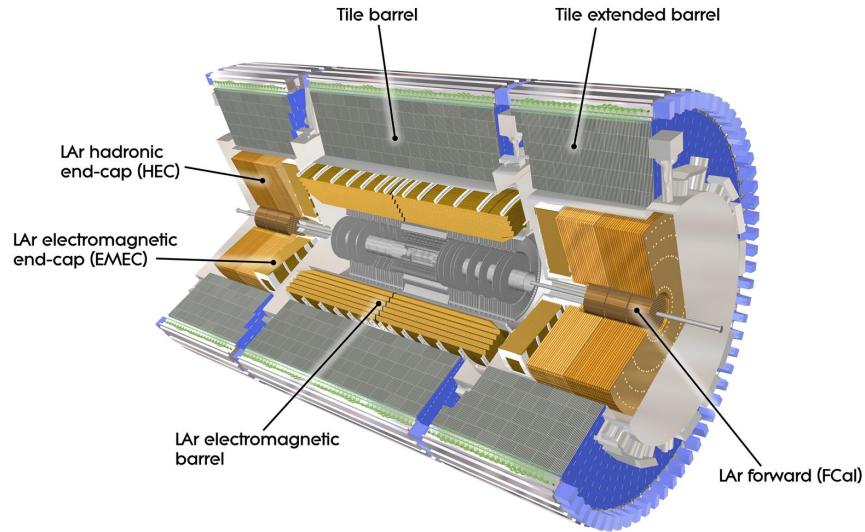


Figure 1.9: The ATLAS calorimeters.

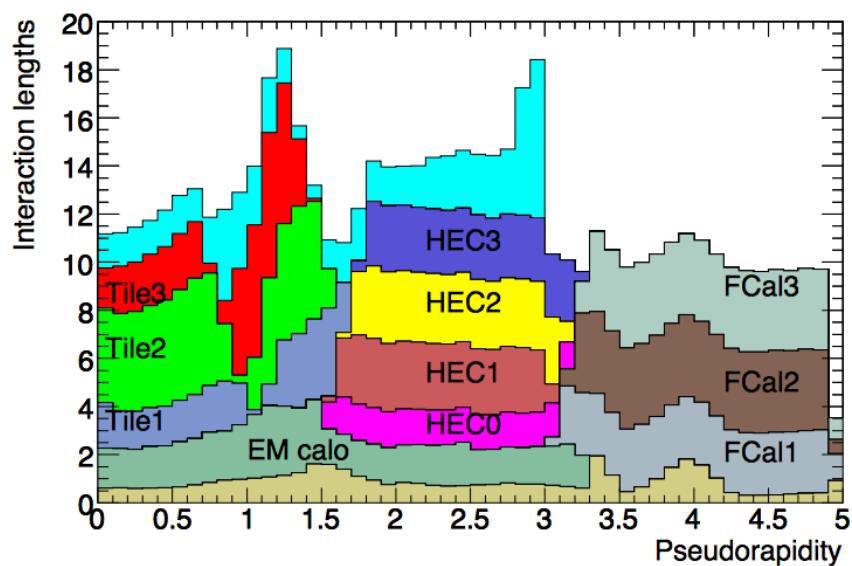


Figure 1.10: Material depth of the ATLAS calorimeters. IC:⁴⁴

331 1.5.1 CALORIMETER RESOLUTION

332 Before diving into the specifics of each of the ATLAS calorimeters, we review some aspects of calorime-
333 ter energy resolution performance. A calorimeter’s relative energy resolution (a ratio) can be broken
334 up into three orthogonal components, as shown in Equation 1.6.

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

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

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

348 to this is the CMS crystal-based calorimeter; ATLAS is a more conservative design), so well-behaved
349 absorbers like lead or iron are necessary to ensure all the energy is contained within a calorimeter.
350 Another factor is cost; things like liquid argon are expensive. Finally, most active media are unsuit-
351 able for structural support, so sturdy absorbing materials help relieve engineering constraints.

352 1.5.2 THE ELECTROMAGNETIC CALORIMETER (ECAL)

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

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

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

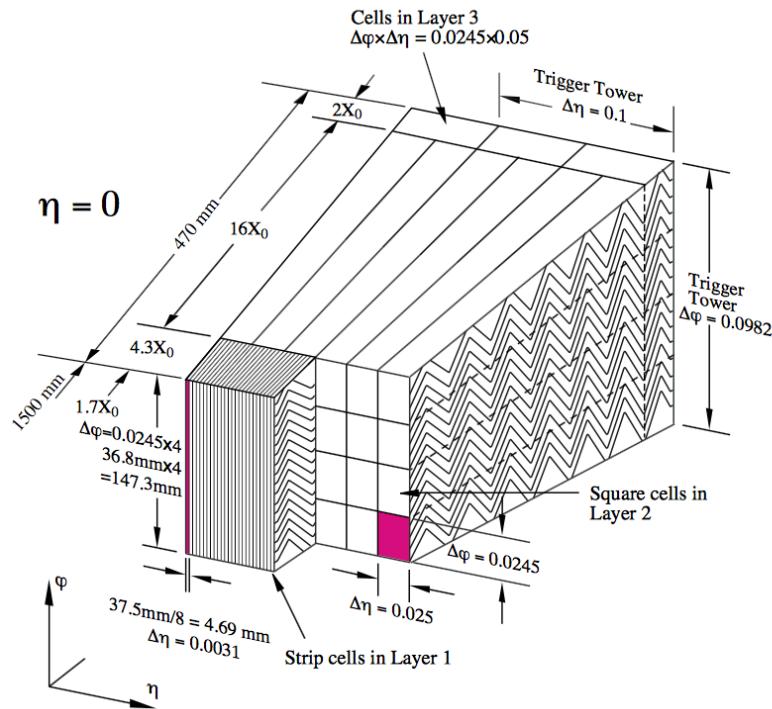


Figure 1.11: The accordion geometry of the LAr electromagnetic calorimeters is prominently shown in this illustration of an ECAL barrel module. IC:⁴⁴

³⁶⁸ 1.5.3 HADRONIC END CAPS (HEC)

³⁶⁹ The HEC covers an $|\eta|$ range of 1.5 to 3.2. Like the ECAL end caps, the HEC consists of two identi-
³⁷⁰ cal wheels out to a distance from the beam axis of 2030 mm; its layout is shown in Figure 1.12. The
³⁷¹ HEC also has LAr as the active material, but instead has flat copper plates as absorbers for sampling
³⁷² fraction of 4.4% and 2.2% in the first and second wheels, respectively. Its granularity in $\eta - \phi$ is
³⁷³ 0.1×0.1 for $|\eta|$ up to 2.5 and 0.2×0.2 in the more forward regions.

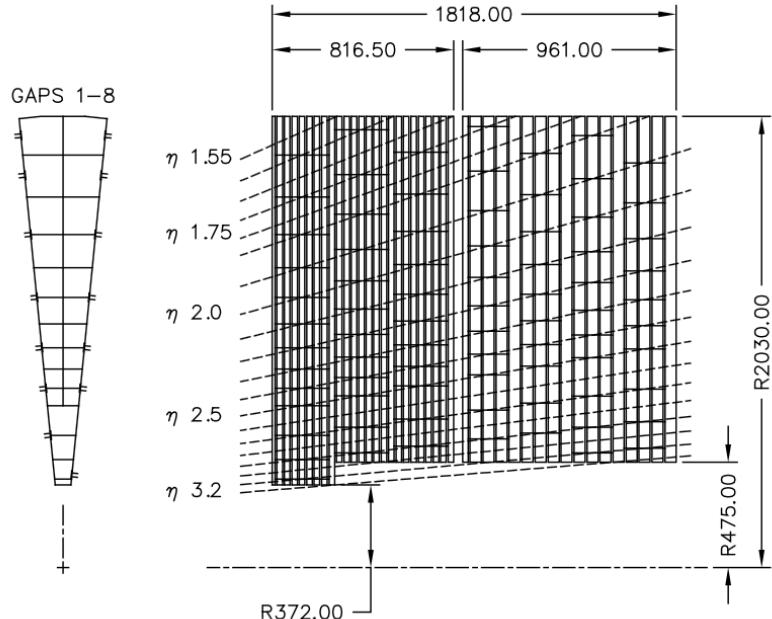


Figure 1.12: The layout of the HEC in $r - \phi$ and $r - z$; dimensions are in millimeters. IC.⁴⁴

³⁷⁴ 1.5.4 THE FORWARD CALORIMETER (FCAL)

³⁷⁵ The FCAL covers an $|\eta|$ range from 3.1 to 4.9, again using LAr as the active material in gaps between
³⁷⁶ rods and tubes in a copper-tungsten matrix, as shown in Figure 1.13. These system has characteristic
³⁷⁷ performance corresponding to stochastic term of $S \approx 1 \text{ GeV}^{-1/2}$. There are three modules in the
³⁷⁸ FCAL: one electromagnetic and two hadronic, with the latter two featuring a higher tungsten con-
³⁷⁹ tent for a larger absorption length.

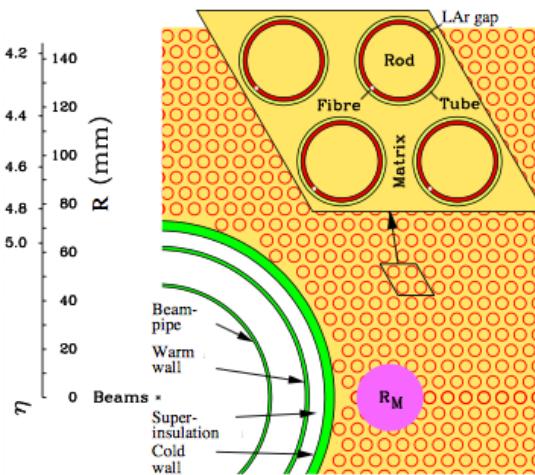


Figure 1.13: The material layout for a typical section of the FCAL in the transverse plane. IC:⁴⁴

³⁸⁰ 1.5.5 THE HADRONIC TILE CALORIMETER

³⁸¹ The tile calorimeter, covering an $|\eta|$ of up to 1.7 is made up of 64 modules in the barrel (each cover-
³⁸² ing $\Delta\phi$ of $360/64 = 5.625^\circ$), each with a layout as in Figure 1.14. It is designed to be self-supporting
³⁸³ for structural reasons, and so is the only calorimeter without LAr as a an active medium, with a stag-
³⁸⁴ gered matrix of active scintillating polystyrene and supporting steel. It operates at 1800 V with a 400

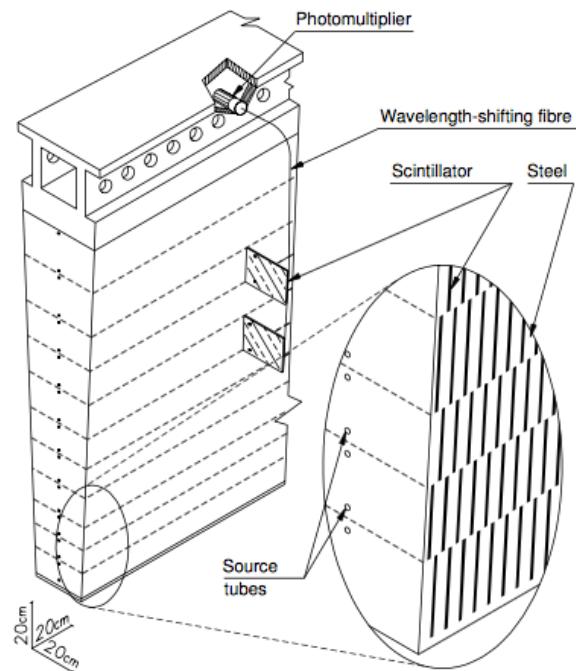


Figure 1.14: The material layout for a typical section of the hadronic tile calorimeter. IC.⁴⁴

385 ns dead time and has a thickness corresponding to 10–20 interaction lengths (2.28–4.25 m). Its cells
386 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
387 performance corresponds to $S = 0.5 \text{ GeV}^{-1/2}$ and $C = 0.05$ (0.03 after calibration).

388 1.6 THE MUON SPECTROMETER

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

$$\frac{\sigma_{p_T}}{p_T} = c_0 \oplus c_1 \cdot p_T \quad (1.7)$$

395 The c_0 term represents a degradation in resolution due to multiple scattering, and is typically 0.5–
396 2%⁷⁷. The c_1 term describes the phenomenon of, holding magnetic field constant, higher momen-
397 tum muons curving less. This term has typical values of $10^{-3} - 10^{-4} \text{ GeV}^{-1}$. At very high p_T val-
398 ues, this is of particular concern since a very small curvature can result in charge misidentification.

399 A cross-sectional view (in $r-z$) of the muon spectrometer with station names, detector types, and
400 layouts is shown in Figure 1.15. There are three layers of muon detectors in both the barrel (at 5 000,
401 7 500, and 10 000 mm) and end cap (at 7 000 (11 000), 13 500, and 21 000 mm), with the innermost

402 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.^{†‡}

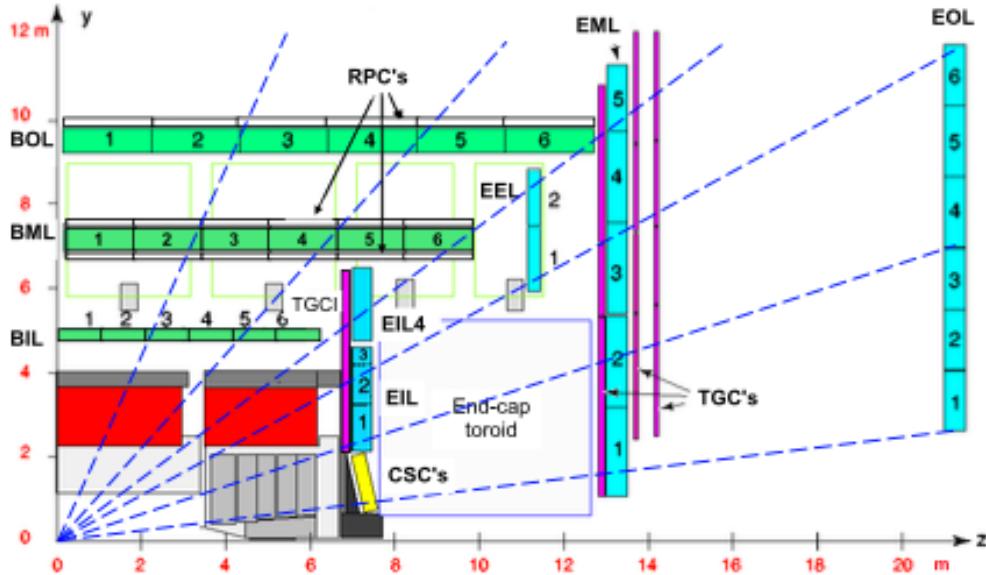


Figure 1.15: The ATLAS muon spectrometer. Naming of the MDT stations obeys the following convention [BE] (barrel or end cap) [IEM0] (inner, inner extended (end cap only), middle, or outer layer) [1-6] (increasing in z (r) for the barrel (end cap)), so EI1 is the station in the inner most end cap layer closest to the beam pipe. IC:⁴⁴

403
 404 The MS can reconstruct muons with transverse momenta from 5 GeV up to 3 TeV (with 10%
 405 resolution at 1 TeV (3% at 100 GeV)). Detectors in the MS fall into two broad headings, precision
 406 detectors and trigger detectors, both described below. Nominal performance of the current detec-
 407 tor types in the MS is summarized in Figure 1.16, a table taken from⁴⁴. It should be noted that $|\eta|$
 408 ranges quoted below, where applicable, do not include the range 0-0.1, where this a gap in the MS to
 409 allow for cabling and other services to the ATLAS detector; for a discussion of compensatory mea-

^{†‡}This will change with the New Small Wheel Phase I Upgrade. cf. Appendix A

⁴¹⁰ 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:⁴⁴

⁴¹¹ I.6.1 PRECISION DETECTORS

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

425 In this region, the precision detectors are the CSC's, which have a much lower dead time of ~ 40
426 ns. These are multiwire proportional chambers with cathode planes that have orthogonal sets of
427 strips, allowing for a measurement in both principal directions. CSC detector sizes also vary by sta-
428 tion, coming in both small and large chambers. The CSC strip pitch is 5.31 (5.56) mm for the large
429 (small) chambers, with position determined from the induced charge distribution in the strips. This
430 corresponds to a nominal resolution of 60 (5 000) μm per plane in the bending (non-bending) direc-
431 tion. These are slated to be replaced by Micromegas detectors in the NSW.

432 **I.6.2 TRIGGER DETECTORS**

433 Trigger detectors have a fundamentally different role than the precision detectors, instead needing to
434 deliver “good enough” approximate values of muon track positions and p_T values. The MS has two
435 types of trigger detectors: Resistive Plate Chambers (RPC’s) in the barrel and Thin Gap Chambers
436 (TGC’s) in the end caps. They collectively cover an $|\eta|$ range to 2.4, and their arrangement is shown
437 in Figure I.17.

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

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

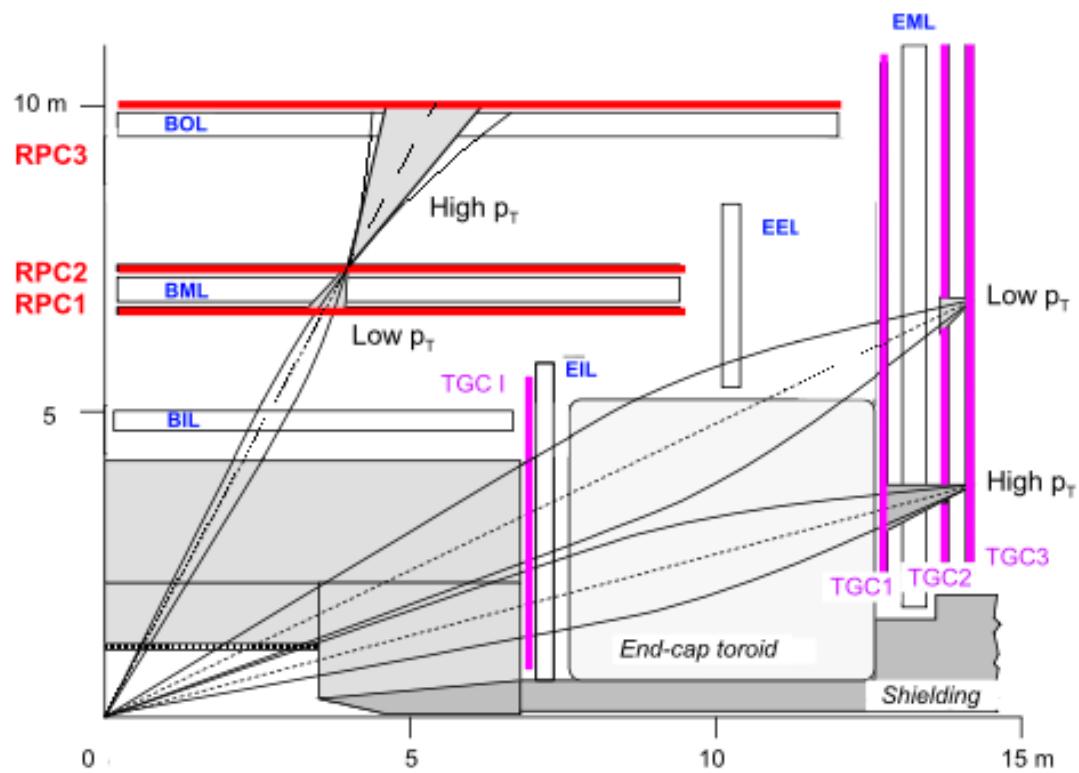


Figure 1.17: ATLAS MS trigger detector arrangement. IC:⁴⁴

⁴⁴⁵ For more details on how detector level trigger objects work in the ATLAS MS, see Appendix A

⁴⁴⁶ 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

447

448

The Standard Model Higgs and Collider

449

Event Variables

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

⁴⁵² be addressed here. This discussion follows⁶⁶ Chapter 11 in both content and notation. We then
⁴⁵³ move onto the treatment of kinematic variables in collider events, including the two novel schemes
⁴⁵⁴ considered in this thesis, the Lorentz Invariants (LI) and RestFrames (RF) concepts.

⁴⁵⁵ 2.1 THE STANDARD MODEL HIGGS BOSON

⁴⁵⁶ The generic scalar Lagrangian potential (the kinetic term will be addressed later) for a scalar in the
⁴⁵⁷ SM is:

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

⁴⁵⁸ where Φ is the Higgs field, a complex scalar doublet under $SU(2)$. Its four degrees of freedom are
⁴⁵⁹ typically decomposed as follows:

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

⁴⁶⁰ ϕ^+ is the complex charged component of the Higgs doublet, and ϕ^0 and α^0 are the CP-even and
⁴⁶¹ CP-odd neutral components, respectively.

⁴⁶² If the sign of $m^2 \Phi^\dagger \Phi$ is negative, Φ acquires a *vacuum expectation value* or VEV:

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

⁴⁶³ with this value typically denoted $v = \sqrt{2m^2/\lambda} = (\sqrt{2}G_F)^{-1/2} \approx 246$ GeV (with the coupling

⁴⁶⁴ of the 4-Fermi effective theory of weak interactions measured through experiments involving muon
⁴⁶⁵ decay), and ϕ^0 is rewritten as $\phi^0 = H + v$.

⁴⁶⁶ This non-zero VEV induces spontaneous symmetry breaking in the SM's gauge (local) symme-
⁴⁶⁷ try group of $SU(3)_C \times SU(2)_L \times U(1)_Y$ since the VEV does not respect the $SU(2)_L \times U(1)_Y$
⁴⁶⁸ symmetry of the Lagrangian (i.e. $\langle \Phi \rangle$ is not invariant under a gauge transformation of this group).

⁴⁶⁹ Three of the four generators of this subgroup are spontaneously broken, which implies the existence
⁴⁷⁰ of three massless Goldstone bosons, which are in turn “eaten” by linear combinations of the W^a and
⁴⁷¹ B bosons to form the longitudinal components of the familiar W^\pm and Z bosons, with the last gen-
⁴⁷² erator giving rise to the usual, unbroken $U(1)_{EM}$ symmetry and its massless photon, A , as well as
⁴⁷³ the scalar Higgs boson H . To see this, one starts with the full Higgs SM Lagrangian (kinetic minus
⁴⁷⁴ potential only)

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

⁴⁷⁵ One simply plugs in the reparametrized Φ with $\phi^0 = H + v$, collects the terms involving v together
⁴⁷⁶ with the appropriate W and B kinetic terms to extract:

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

⁴⁷⁷ This is left as an exercise for the reader; this exercise also makes manifest that the Higgs couples with
⁴⁷⁸ the W^\pm and Z with strength quadratic in the gauge boson masses. Since the Higgs field also respects

⁴⁷⁹ the $SU(3)_C$ color symmetry, the eight gluons are also left massless, and the H is left interacting with
⁴⁸⁰ photons and gluons primarily through heavy quark loops (i.e. no tree-level interactions).

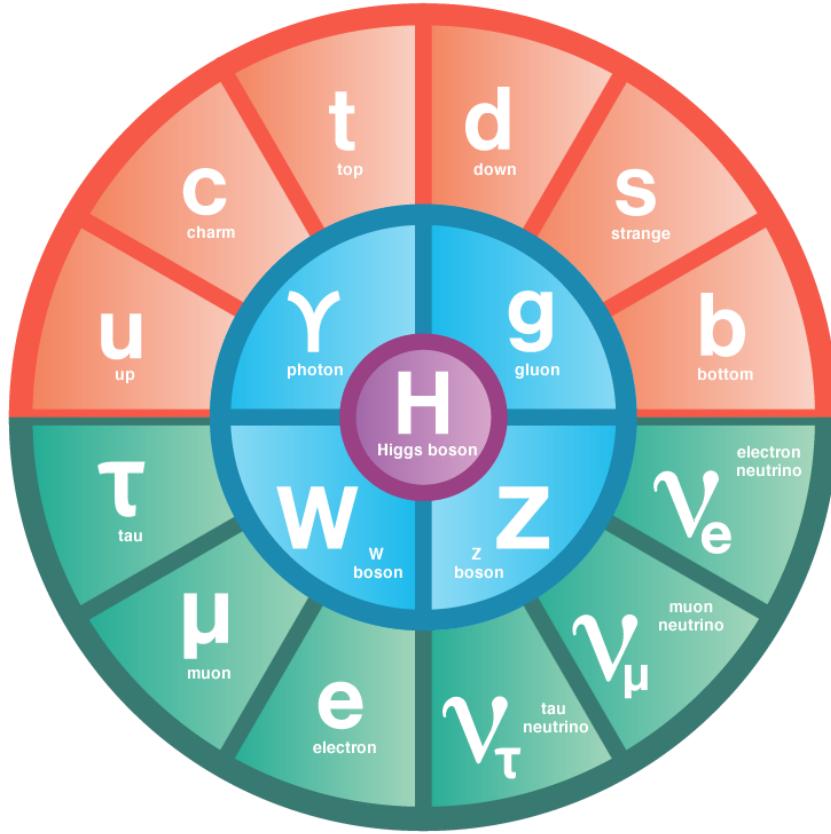


Figure 2.1: The fundamental particles of the Standard Model. IC:⁷⁴

⁴⁸¹ The Higgs is often introduced to the public at large as the mechanism through which fundamen-
⁴⁸² tal fermions (enumerated in Figure 2.1) acquire mass—this is through the Yukawa interactions of the
⁴⁸³ Higgs:

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

⁴⁸⁴ 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

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

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

495 2.2 HIGGS BOSON PRODUCTION AND DECAY AT THE LARGE HADRON COLLIDER

496 The leading order Feynman diagrams for the four dominant modes of Higgs production at the LHC
497 are shown in Figure 2.2, each described briefly in turn. The dominant process, accounting for some
498 87% of Higgs production at the nominal LHC center of mass energy of 14 TeV, is gluon-gluon fu-
499 sion (ggF), shown at top left in Figure 2.2. At high center of mass energies, most of a proton's mo-
500 mentum is predominantly carried by sea gluons (as opposed to the constituent valence quarks asso-
501 ciated with the hadron's identity). This, along with the difficulties associated with high luminosity
502 antiproton beam production, is why the LHC was designed as a proton-proton collider instead of a
503 proton-antiproton collider (like the Tevatron or once planned SSC). As mentioned above, the Higgs
504 does not couple directly to gluons but must instead be produced through the fermion loop shown

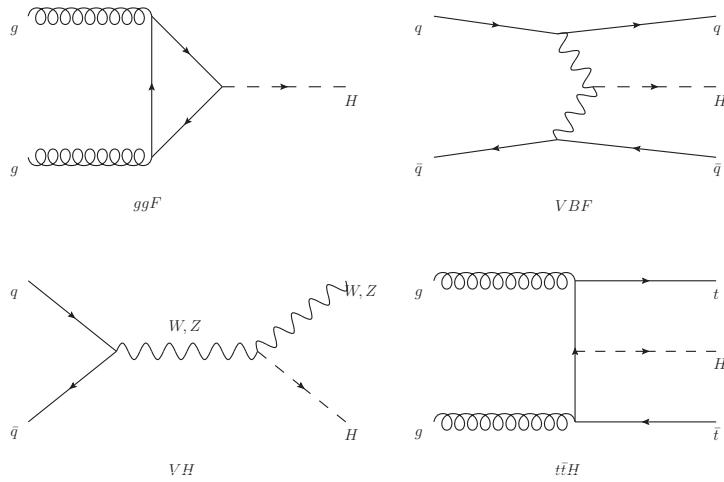


Figure 2.2: Dominant Higgs production modes.

in the figure. The heaviest fundamental fermion by far is the top quark, with $m_t = 173$ GeV, so top loops dominate this process. While not particularly relevant for this thesis, about 14% of events in the 2-lepton channel of the $H \rightarrow b\bar{b}$ analysis are ggF initiated.

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

The third leading process is “Higgsstrahlung” or Higgs production in association with a vector boson, often simply VH production. In this process, a quark-antiquark pair in the colliding protons forms an energetic vector boson, which then radiates a Higgs (this is similar to photon emission of accelerating electrons, called “Bremsstrahlung,” hence the name). Some fraction of the time (about 21% of the time for WH and 6.7% of the time for ZH), the energetic V will decay leptonically (i.e.

517 into a decay involving an electron or a muon), which provides a unique and triggerable signature
 518 for this process. Another 20% of the time for ZH production, the Z will decay to neutrinos, which
 519 are not absorbed by detectors and show up as missing transverse energy (\vec{E}_T^{miss}), another triggerable
 520 signature. This ability to trigger on leptons and \vec{E}_T^{miss} and the requirement that this leptonic signa-
 521 ture be consistent with a V allow one to significantly reduce the impact of multijet background (a
 522 very common generic processes at the LHC) on analysis. Hence, this is the process of primary impor-
 523 tance to this thesis.

524 The final important Higgs production process is $t\bar{t}H$ production, the box diagram in the lower
 525 right of Figure 2.2. Again, the top pair provides a useful signature for analysis. This, like VBF, is also
 526 not considered in this thesis.

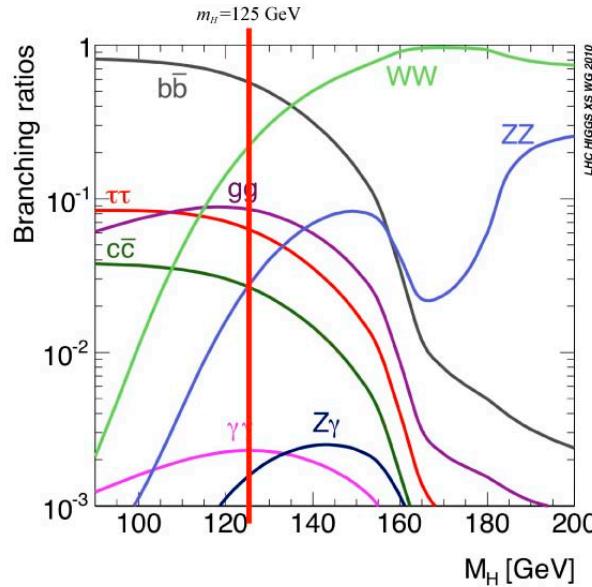


Figure 2.3: Higgs decay modes as a function of its mass; a line has been drawn at the observed Higgs mass of 125 GeV.

527 Once the Higgs has been produced, it can decay in a number of ways, as shown in Figure 2.3. By

528 far the most dominant decay mode of the Higgs is to $b\bar{b}$ at 58% of all decays. This b -quark pair then
 529 hadronizes into two b -jets (for a more thorough discussion of jets and b -jets in particular, see Sec-
 530 tion 5.5). However, many processes at the LHC create pairs of b -jets with invariant masses consistent
 531 with the Higgs and have much higher production rates ($t\bar{t}$ production at the LHC is in the neighbor-
 532 hood of hundreds of pb, compared to Higgs cross sections of a few pb), so a clear process signature
 533 is necessary to study $H \rightarrow b\bar{b}$ production at the LHC. This is why the bulk of search efforts have fo-
 534 cused on VH production. A summary of Higgs production cross sections and simple extrapolations
 535 to raw numbers of Higgs bosons produced for VH for leptonically decaying V is shown in Table 2.1

\sqrt{s} (TeV)	ZH	WH	ggF	total σ	$N_{V \rightarrow \ell^{\pm} H}$
7	$0.34^{+4\%}_{-4\%}$	$0.58^{+3\%}_{-3\%}$	$15.3^{+10\%}_{-10\%}$	17.5	$4.7 \text{ fb}^{-1} \rightarrow 589$
8	$0.42^{+5\%}_{-5\%}$	$0.70^{+3\%}_{-3\%}$	$19.5^{+10\%}_{-11\%}$	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 (in pb) for processes important to the SM VH ($b\bar{b}$) 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.

536 2.3 COLLIDER EVENTS AND EVENT LEVEL VARIABLES

537 Collision data in experiments like ATLAS is structured using what is known as the *event data model*.
 538 In this model, one collision corresponds to one event. The raw data, the various tracks, energy de-
 539 posits, and hits in the detector, undergo reconstruction (described at length in Chapter 5) both through
 540 automated, experiment-wide, standardized production and through analysis-specific level selec-
 541 tions, corrections, and calibrations. The result of this considerable effort is a collection of labeled

542 4-vectors, representing the final state objects. This is shown in Figure 2.4.

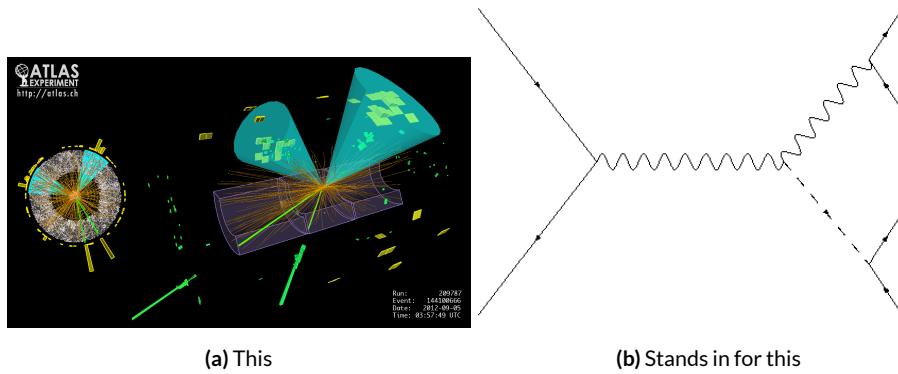


Figure 2.4: Reconstruction in a nutshell

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

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

- 554 two b -jets in selected events with the greatest transverse momenta, m_{bb} , is used as a validation), or
555 something more complicated like a multivariate analysis (MVA) discriminant.

556 **2.4 CHARACTERIZATION WITH EVENT-LEVEL VARIABLES**

557 Traditionally, particle physicists have favored the approach of using distributions of physical vari-
558 ables since it is easier to develop “physical intuition” for what these distributions should “look like”
559 during validation, so it is no surprise that as many LHC analyses have transitioned to using MVA
560 techniques that these variables form the basis of many very robust physics results. These variables do
561 quite well summarize many of the main physics features of an event for the signal topology, certainly
562 much better than feeding all 18–22 4-vector components directly into some machine learning algo-
563 rithm. In $ZH \rightarrow \ell\ell b\bar{b}$ events, for example, one wishes to characterize the ZH system by using the
564 lepton pair as a stand-in for the Z and the b -jet pair as a stand-in for the H , and composite variables
565 like m_{bb} and $m_{\ell\ell}$ can be used to check whether events are consistent with these objects. There are
566 also variables like p_T^V that characterize the momentum scale of the event, angles like $\Delta R(b_1, b_2)$ and
567 $\Delta\phi(V, H)$ that can be further used to characterize the overall “shape” of these events, and variables
568 like \vec{E}_T^{miss} that can discriminate against backgrounds like $t\bar{t}$ that do not have a closed topology.

569 Nevertheless, the intuition based approach, with incremental addition of variables as they prove
570 useful in the lifetime of an analysis’s iterations, does beg the question of whether there is a more sys-
571 tematic way to treat this information. There are clearly patterns to which variables are useful: these
572 correspond to important information about the hypothesized physics objects and their relation-
573 ships, and there have been many attempts to systematize the way these variables are found. Such

574 systematic, top-down approaches often promise to increase performance in two ways. The first is by
575 having higher descriptive power, often through some sophisticated treatment of the missing trans-
576 verse energy in an event, \vec{E}_T^{miss} . \vec{E}_T^{miss} is just a single quantity, and if there is just one invisible object
577 in a desired event topology, using \vec{E}_T^{miss} on its own often provides sufficient sensitivity. In more com-
578 plicated topologies with multiple invisible particles in the final state, for example in many supersym-
579 metry searches, a more careful treatment of the missing energy is often necessary.

580 The second means of improvement is through using a more orthogonal basis of description,
581 which allows one to more efficiently use data and simulation samples. A more orthogonal basis im-
582 plies that variables contain less overlapping information with each other and so allow for a more
583 efficient exploration of parameter space. This means one can gain higher sensitivity from equivalent
584 datasets using a more orthogonal basis. To see why this might be the case, take an MVA discrimi-
585 nant 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 $ZH \rightarrow$
586 $\ell\ell b\bar{b}$ topology, the transverse mass of the Z and H (and hence the lepton pair and jet pair) are equiv-
587 alent. This means that at higher p_T^V the p_T of b -jets will also be higher, which in turn implies that
588 they will have a smaller angle of separation and hence a smaller $\Delta R(b_1, b_2)$. This correlation is not
589 unity—each variable still does have information the other does not—but is still very high. Hence,
590 when training an MVA, which in principle knows nothing about these variables other than some
591 set limits, an undue number of training events will be wasted converging upon relations that could
592 be known *a priori*, and while this might be easy to hard code in for a two variable toy example, the
593 dimensionality of any real discriminant makes this prohibitive. An MVA that uses data (both ac-
594 tual and simulated) more efficiently will also tend to be have lower variance, offering a potential av-

595 enue for reduction in the error on quantities of interest due to systematic uncertainties. Details of
596 how this plays out in a likelihood fit will be deferred to the discussion of the fit model used in the
597 $VH(b\bar{b})$ search in Chapter 7.

598 Many of these novel schemes are designed to explicitly address the first issue of invisibles in the
599 final state in channels where it is of paramount importance while having the second issue as some-
600 thing of a fringe benefit. However, as the amount of data taken at the LHC grows, analyses will in-
601 creasingly become systematics limited, so an exploration to the veracity of the second claim has great
602 potential for the high luminosity era of the LHC. The $ZH \rightarrow \ell\ell b\bar{b}$ process offers a great setting
603 for investigating this issue on its own since its closed topology largely mitigates any improvement
604 from more sophisticated treatments of \vec{E}_T^{miss} . We introduce two of these more top-down approaches
605 to event-level variables below: the “Lorentz Invariant” (LI)⁵³ and “RestFrames inspired” (RF)⁵⁶
606 variable schemes. A broad overview of the concepts behind these schemes will be given here, with
607 a more in-depth discussion of their implementation deferred until Chapter 6.

608 2.5 LORENTZ INVARIANTS

609 The LI variables, first put forth by S. Hagebeck and others⁵³, are based upon the fact that once the
610 4-vectors of an event are determined, all of the information in an event are encoded into their inner
611 products (Lorentz invariant quantities, hence the name) and the angles between them. This makes
612 for 16 quantities in all: the ten inner products of the 4-vectors, the three Euler angles, and the three
613 parameters specifying the boost of the ZH system. The masses of the four final state objects are not
614 considered very useful and so can be removed to leave six meaningful inner products (the ${}_4C_2$ com-

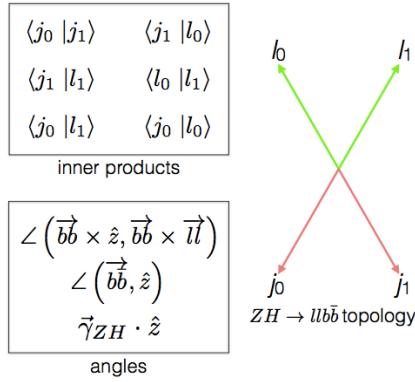


Figure 2.5: Summary of LI variables in the $ZH \rightarrow \ell\ell b\bar{b}$ topology.

binations between distinct final state 4-vectors). Since these inner products can have an ill-defined physical interpretation and in order to help MVA training, each inner product is scaled by:

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

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

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

626 These variables do contain a lot of information similar to the usual set: there are mass equivalents
 627 ($j_0 \cdot j_1 \leftrightarrow m_{bb}$, and $l_0 \cdot l_1 \leftrightarrow m_{\ell\ell}$) and angles. Instead of individual final state object scales, there
 628 are the four jet-lepton inner products, though this correspondence (and indeed any physical inter-
 629 pretation) is far from clear. An important advantage of the LI variable set is that all of the variables
 630 are in it are orthogonal in the signal case by construction. A drawback of this framework in a com-
 631 pletely closed final state is that there is no way to treat E_T^{miss} in a Lorentz invariant way.

632 There is also no prescription for any additional jets in the event beyond the two b -tagged jets.
 633 They are simply ignored in these variable calculations since the fiducial analysis requirement of ex-
 634 actly two b -tagged jets eliminates any combinatoric ambiguity, and additional, untagged jets are as-
 635 sumed (not entirely rigorously) to be unrelated to the signal-like hard scatter.

636 2.6 RESTFRAMES VARIABLES

637 The RestFrames variables⁶⁶, calculated using the software package of the same name, is based upon
 638 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-

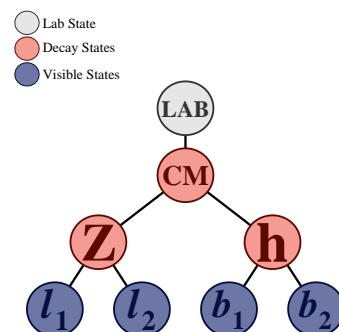


Figure 2.6: The $ZH \rightarrow \ell\ell b\bar{b}$ decay tree.

639 erally, one does not typically have enough information to determine exactly each of the intermediate
 640 rest frames or the boosts between the frames, but in a completely closed final state like $ZH \rightarrow \ell\ell b\bar{b}$,
 641 this can be done in the usual way by adding the 4-vectors of the final state objects and solving the
 642 usual equations from special relativity (`RestFrames` does this automatically for each event).

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

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

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

655 denoted `Rpt`, `Rpz`, and `Rmet`. These can be thought of as behaving like significance based variables
 656 in particle physics, like METHT or impact parameter significances, or event level defined versions

657 of the scalings applied to the LI inner products. These are used instead of the final state object scales
658 and standard E_T^{miss} of the standard variable set.

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

668 2.7 EXTENSIONS TO THE 1 AND 0 LEPTON CHANNELS

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

674 The LI concept was in fact initially formulated to improve sensitivity in the 1-lepton channel,
675 with the same orthogonality of variables described in the 2-lepton case being the main draw. The LI
676 approach to estimating the neutrino longitudinal momentum is outlined in ⁵³, which we reproduce

677 here. We first guess the neutrino energy in its rest frame and then boost to the lab frame:

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

678 Finally, assuming energy and momentum in aggregate are equally shared among final state con-

679 stituents, we arrive at

$$\langle p_z^\nu \rangle = \frac{1}{4} \times \frac{4}{3} (p_z^l + p_z^{j0} + p_z^{j1}) \quad (2.10)$$

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

681 As alluded to in the 2-lepton discussion, when there is missing information in the final state from

682 invisible particles and/or combinatoric ambiguities, recursive jigsaw reconstruction (RJR) offers a

683 standard toolkit for deriving estimated boosts between rest frames by analytically minimizing on

684 unknown quantities. While in more exotic final states with multiple invisible particles and com-

685 binatoric ambiguities the choice of jigsaw rule can be subjective, the case of W is well-studied and

686 outlined in detail in Section V.A. of⁶. It reproduces the usual transverse mass of the W in place of

687 MZ in the 2-lepton case. Not surprisingly, the underlying calculation is also much the same as the LI

688 case (where rest frames and boost were explicitly invoked); again, information is the same, only its

689 decomposition is different.

690 The 0-lepton channel would appear to present some difficulty as two neutrinos in the final state

691 introduce extra degrees of freedom, but both concepts may be extended by treating the invisibly de-

692 caying Z as a single invisible particle and requiring the Z to be on-shell. Both of these requirements

⁶⁹³ may be folded into the 1-lepton framework to produce similar sets of variables.

Variable	Name	0-lepton	1-lepton	2-lepton
\vec{p}_T^V	pTV		✓	✓
\vec{E}_T^{miss}	MET	✓	✓	✓
\vec{p}_T^{jet1}	pTB ₁	✓	✓	✓
\vec{p}_T^{jet2}	pTB ₂	✓	✓	✓
$\text{MV}_{2\text{C1O}}(\text{jet}_1)^*$	$\text{MV}_{2\text{C1O}}\text{B}_1$	✓	✓	✓
$\text{MV}_{2\text{C1O}}(\text{jet}_2)^*$	$\text{MV}_{2\text{C1O}}\text{B}_2$	✓	✓	✓
m_{jj}	mBB	✓	✓	✓
$\Delta R(\text{jet}_1, \text{jet}_2)$	dRBB	✓	✓	✓
$ \Delta\eta(\text{jet}_1, \text{jet}_2) $	dEtaBB	✓		
$\Delta\phi(V, H)$	dPhiVBB	✓	✓	✓
$\Delta\eta(V, H)$	dEtaVBB			✓
$M_{\text{eff}}(M_{\text{eff}3})$	HT	✓		
$\min(\Delta\phi(\ell, \text{jet}))$	dPhiLBmin		✓	
m_T^W	mTW		✓	
m_{ll}	mLL			✓
$\Delta Y(W, H)$	dYWH		✓	
m_{top}	mTop		✓	
Only in 3 Jet Events				
\vec{p}_T^{jet3}	pTJ ₃	✓	✓	✓
$\text{MV}_{2\text{C1O}}(\text{jet}_3)^*$	$\text{MV}_{2\text{C1O}}\text{B}_3$	✓	✓	✓
m_{jjj}	mBBJ	✓	✓	✓

Table 2.2: Variables used to train the multivariate discriminant. Starred variables (b -tag scores) are not included in current versions of the standard discriminants, but have traditionally been included and most likely will be reintroduced as soon as their accompanying systematics are available.

⁶⁹⁴ While the precise variables that would be included in 0- and 1-lepton LI and RF MVA discrimi-
⁶⁹⁵ nants is beyond the scope of this thesis, looking at Table 2.2, we can see the dimensionality and in-
⁶⁹⁶ puts of the discriminants of the fiducial analysis. The correspondence for LI/RF variables and stan-
⁶⁹⁷ dard variables extends nicely to the other lepton channels. The reduction in multiplicity of variables

⁶⁹⁸ owing the lower number of degrees of freedom provided by treating the Z as a single invisible particle in the o-lepton channel would likely not be an issue, as one would just be able to use a greater
⁶⁹⁹ fraction of available variables in the MVA discriminant.
⁷⁰⁰

⁷⁰¹ *Maybe do the o-lep calculation and some cute RF diagrams*

What do you read, my lord?

Words, words, words.

Hamlet, 2:2

3

702

703

Data and Simulated Samples

704 THE DATA AND Monte Carlo simulation (MC) samples used in this thesis are the same as in the fidu-
705 cial analysis. The data corresponds to 36.1 fb^{-1} of pp collision data collected in 2015+16 at the AT-
706 LAS detector at $\sqrt{s} = 13 \text{ TeV}$. Details of the Run 1 analysis referenced in Chapter 9, may be found
707 in ²⁰. Only events recorded with all systems in ATLAS in good working order and passing certain

708 quality requirements, according to a Good Run List (GRL), are analyzed.

709 Details about MC samples may be found in ⁶⁷, and signal and background modeling are discussed
710 in the next. The $ZH \rightarrow \ell\ell b\bar{b}$ process is considered for both multivariate analysis (MVA) optimiza-
711 tion and the final statistical analysis, while $WH \rightarrow \ell\nu b\bar{b}$ and $ZH \rightarrow \nu\nu b\bar{b}$ production are included
712 in the final statistical analysis only. Signal MC samples were generated separately for qq and gg ini-
713 tiated VH processes. $qqVH$ samples were generated with PowHEG MiNLO + PYTHIA8 with the
714 AZNLO tune set and NNPDF3.0 PDF. Nominal $ggZH$ samples were generated using PowHEG for
715 the matrix element (ME) and PYTHIA8 for the parton shower (PS), underlying event (UE), and mul-
716 tiple parton interactions (MPI), again applying the AZNLO tune and NNPDF3.0 PDF set.⁶⁰

717 The background processes considered in these studies are $Z+jets$, $t\bar{t}$, and diboson production for
718 both MVA optimization and the final statistical analysis with single top production and $W+jets$
719 only considered in the final statistical analysis. $V+jets$ samples are generated using SHERPA 2.2.1⁴⁰
720 for both the ME and PS. These samples are generated in different groups, according to the identity
721 of the V , the max (H_T, p_T^V) of events, with further subdivisions according to the flavor of the two
722 leading jets in an event, b , c , or l , for a total of six categories. $t\bar{t}$ samples are generated using PowHEG
723 with the NNPDF3.0 PDF set interfaced with PYTHIA8 using the NNPDF2.3 PDF's and the A14
724 tune⁴⁶. Single top samples use PowHEG with the CT10 PDF's interfaced with PYTHIA6 using the
725 CTEQ6L1 PDF's^{21,57}. Diboson samples are generated with SHERPA 2.2.1 interfaced with the NNPDF3.0
726 NNLO PDF set normalized to NLO cross sections³³.

*There are certain calculations one simply doesn't do in
public.*

Alan Blaer

4

727

728

Signal and Background Modeling

729 THIS CHAPTER summarizes the modeling of the dominant signal and background processes in
730 this analysis, including corrections and systematic uncertainties (systematic uncertainty, also called
731 nuisance parameter (NP), titles are set in **this** font) related to each process. Further details on the
732 specifics of these topics, including in-depth studies for the derivation and definitions of some of the

733 quantities cited, may be found in⁶⁷. We start with a general discussion of modeling and associated
734 major categories of uncertainties before addressing each of the physics processes.

735 **4.1 EVENT GENERATION IN A NUTSHELL**

736 Before diving into the specifics of modeling and systematic uncertainties associated with each ma-
737 jor set of physics processes considered in this analysis, we review at a schematic level* the problem
738 of simulation event generation. Once a physics processes of interest has been determined, how one
739 simulates an ensemble of particle collisions to model the process in question. This is illustrated in
740 Figure 4.1. Note that the scope of this problem does not include how these generated collision prod-
741 ucts propagate through one's detector. This problem is left for Chapter 5.

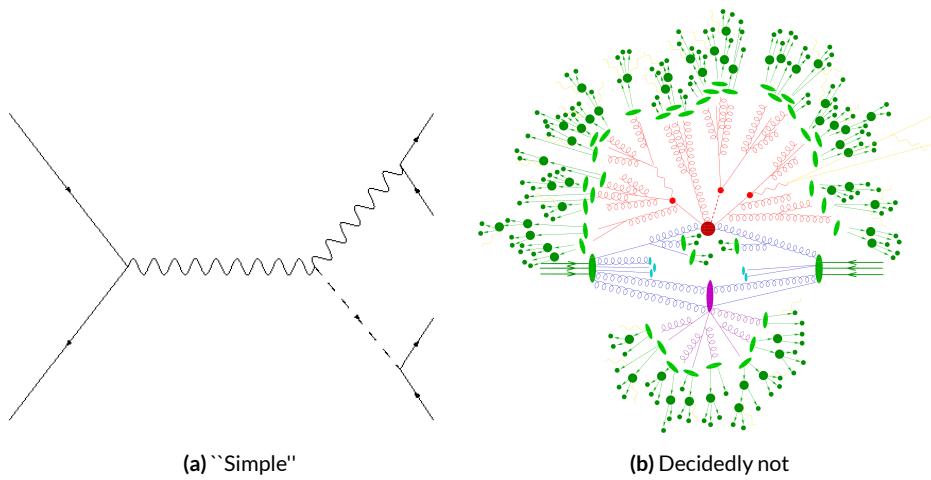


Figure 4.1: The problem here is how to get from (a) to (b).

742 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.

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

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

Typically, it takes several steps and tools to accomplish this. The hard scatter itself is often modelled with a dedicated event generator like PowHEG⁶⁸ or MADGRAPH²³, with events generated then interfaced with a tool like PYTHIA⁷¹ for the PS, UE, and MPI, though there are exceptions (SHERPA⁵¹, for example, can do both the hard scatter and hadronization/ for some processes).

764 4.2 DESCRIPTION OF MODELING UNCERTAINTY CATEGORIES

765 Each of the steps in event generation described above has associated uncertainties. Some uncertain-
766 ties are inherent in the calculations themselves. The choice of which order in perturbation theory
767 to do a calculation, for example, comes with it an implicitly defined level of precision[†]. Extrapolat-
768 ing from one energy/momentum scale to another also introduces uncertainty. Furthermore, there
769 is no *a priori* correct choice to make at each step in event generation, so each choice (the choice of
770 generator, PDF set, parton shower calculator, all of their configurable parameters, etc.) implies an
771 additional layer of uncertainty.

772 In order to quantify these choices, each source of systematic uncertainty is treated separately and
773 given a unique name. To make this more concrete, take the specific example of the uncertainty asso-
774 ciated with the $H \rightarrow b\bar{b}$ branching ratio of 58%, called `ATLAS_BR_bb`, which encapsulates a num-
775 ber of effects (higher order terms, the mass of the b quark, and choice of α_S). The quoted (in prin-
776 ciple asymmetric) uncertainty on the Higgs BR is not itself a direct input into the analysis model.
777 Instead, the effect of varying the branching ratio up and down by one standard deviation is propa-
778 gated to simulated collision events and recorded (i.e. the analysis is run with the Higgs branching
779 ratio at $\pm 1\sigma$, and the results are recorded alongside the nominal result). The nominal and “up” and
780 “down” variations are then typically taken to define a normally distributed, freely floating param-
781 eter in the statistical fit model. Since these parameters associated with systematic uncertainties are
782 not typically considered interesting quantities, they are often referred to as “nuisance parameters”

[†]though this is less well-defined in QCD calculations than for electroweak calculations since they don’t converge

783 (NP's). The terms “systematic,” “systematic uncertainty,” and “nuisance parameter” are often used
784 interchangeably.

785 The specifics of exactly how the effects of variations are saved and propagated to the full fit model
786 are deferred to Chapter 7. The discussion here is confined to how systematic uncertainties for signal
787 and background modeling and their accompanying variations are defined. Modeling systematics are
788 derived separately for each physics process (simulation sample). Sometimes, all of the variation for
789 a given process is encapsulated in a single systematic, but oftentimes the variations from multiple
790 considerations are distinct enough to be treated separately. Furthermore, each of these separate sys-
791 tematics for a given sample/process may be treated in a number of ways (e.g. 0-lepton events may
792 be treated differently from 2-lepton events). An additional subtlety is that a continuous parameter
793 like a branching ratio lends itself quite naturally to defining Gaussian $\pm 1\sigma$ variations, while for dis-
794 crete variations, like choice of PDF set for parton showers, how to proceed is less obvious. This is
795 addressed on a case-by-case basis, as described below.

796 Before enumerating each of the principal physics processes and their systematics, we begin by
797 describing considerations and choices that must be addressed for every physics process in order to
798 make the discussion of individual samples and systematics both clearer and less repetitive.

799 **4.2.1 PHYSICS CONSIDERATIONS**

800 In general, evaluating the uncertainties arising from the many choices in event generation entails
801 producing alternate samples of events, which practically means tuning parameters in the various soft-
802 ware packages and/or using alternate packages/libraries to make new samples. Once these samples

803 have been created, they are compared at truth-level (particle level) using a package called Rivet³⁰ in-
804 stead of using the full ATLAS detector reconstruction for computational considerations. Given the
805 nature of the problem and the tools, there are generally three main categories of physics issues, each
806 described below.

807 **UNDERLYING EVENT AND PARTON SHOWER**

808 The modeling of the underlying event (UE) and parton shower (PS) are usually handled by the same
809 package and so are usually treated together. The typical nominal choice in the fiducial analysis is
810 PYTHIA8. One approach to modeling these uncertainties is simply to see what happens when a
811 different model is used and then compare this alternate set of events to the nominal set, taking the
812 difference as the (implicitly one standard deviation) scale of variation. Another approach is to vary
813 some parameter within a given model, for example, using different tunes in the A14 set for PYTHIA8
814 with their accompanying variations, to characterize the scale of variation.

815 A natural question is how to treat these two approaches on the same footing. When examining
816 a set of potential variations related to the same process or effect, oftentimes the largest single varia-
817 tion in a set is picked as defining the scale for the systematic uncertainty. Another approach is to use
818 the average over a set of variations.[‡] The ATLAS_UEPS_VH_hbb systematic, for example, uses the
819 Pythia8 + A14 tunes approach to determine the scale of UE variation and compares Pythia8 with
820 Herwig7 to characterize the PS variation. Each of the A14 tunes comes with an up and down varia-
821 tion, and the difference between each of these variations and a nominal setup may be expressed as a

[‡]Generally, the maximum is used if it is much larger than other variations, and the average is used if scales are comparable. In general, the historical preference is to be conservative.

822 ratio, R , of total events.

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

$$\sum_{tunes} \max \left(|R_{up} - R_{down}| \right) \oplus \sigma_{PS} \quad (4.1)$$

826 QCD SCALE

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

833 PARTON DISTRIBUTION FUNCTIONS AND α_S

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

839 4.2.2 MODELING SYSTEMATIC TYPES

840 With the concept of what type of effect is taken as a single systematic uncertainty and how its varia-
841 tions are generally evaluated, it is now time to turn to the issue of what exactly is being varied.

842 ACCEPTANCE/NORMALIZATION

843 The most basic type of modeling uncertainty is a normalization uncertainty, often called an accep-
844 tance uncertainty. This simply denotes the uncertainty on the number of predicted events for a
845 given process in a given region of phase space (usually delineated by the number of leptons in the
846 final state and sometimes also by the number of and jets the p_T^V [§] of an event) and is usually expressed
847 as a percent.

848 As an example, the uncertainty on the theoretical prediction of the $H \rightarrow b\bar{b}$ branching ratio,
849 denoted `ATLAS_BR_bb` (it is an ATLAS-wide systematic), is expressed as a normalization system-
850 atic with a value of 1.7%, affecting all VH processes. Now imagine we have an event in a VH sample
851 with weight 1.0. The nominal histograms for this region gets filled with this event's relevant informa-
852 tion with weight 1.0, while the `ATLAS_BR_bb__1up` (`_1do`) histograms get filled with weight 1.017
853 (0.983).

[§]This is the transverse mass of the lepton pair for 2-lepton events, the vectorial sum of the single lepton
and \vec{E}_T^{miss} for 1-lepton events, and the \vec{E}_T^{miss} for 0-lepton events.

854 SHAPE SYSTEMATICS

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

$$w_{event} = A_{region} \times f_{region}(event)$$

858 where w_{event} is the simulated event’s weight, A_{region} is the overall normalization (in principle includ-
859 ing any systematics), and $f_{region}(event)$ is some function of event-level variables, usually a single vari-
860 able, like p_T^V or m_{bb} . The purpose of these systematics is to take into account (in the case of a system-
861 atic) or correct (in the case of a correction applied to the event weight) the non-trivial dependence
862 of a normalization on one of these quantities. Some of these are taken from histograms while others
863 are parametric functions (in this analysis, usually linear ones).

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

867 An example of the latter case is the systematic associated with the m_{bb} dependence of the the
868 $t\bar{t}$ normalization for 2 jet, $p_T^V \in [75, 150]$ GeV, 2 lepton events. In this case, a variety of effects are
869 studied (ME, PS, UE), as shown in Figure 4.3. The top half of the plot is the m_{bb} plot for this re-
870 gion, with the black bars representing the nominal spectrum and spectra generated with different
871 ME, PS, and UE choices. The ratio plot in the bottom half of the figure shows the scale of varia-
872 tion normalized to bin content. From this ratio plot, it is clear that the choice of ME (pink points)

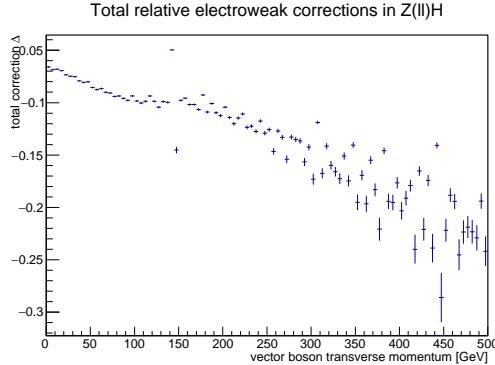


Figure 4.2: The δ_{EW} correction term for 2-lepton $qqZH$.

873 was seen to have the largest effect on normalization. The linear fit in the plot reasonably envelopes

874 this maximum variation was done, and so is taken as the systematic variation. Hence, in this case,

875 $f_{\text{region}}(\text{event})$ is a linear function of m_{bb} , with positive (negative) slope for the up (down) variation.

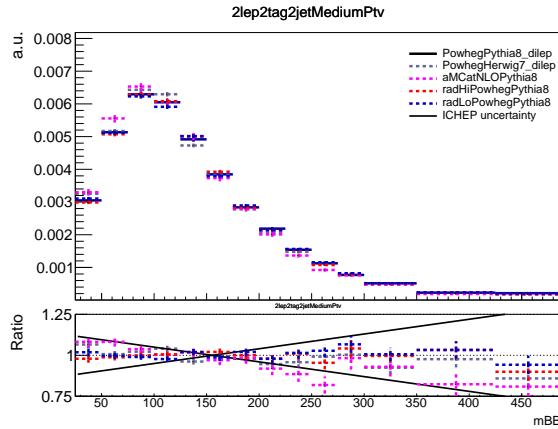


Figure 4.3: The derivation of the 2-lepton $t\bar{t} m_{bb}$ shape systematic.

876 DIVIDING MODELING UNCERTAINTIES: ACCEPTANCE RATIOS

877 In addition to uncertainties on absolute normalizations (both inclusive and region specific), mod-

878 eling uncertainties are sometimes introduced for the ratio of normalizations between different re-

gions. While these can be simple ratios, evaluating a systematic's effect between regions means evaluating nominal and alternate choices between regions, so the so-called “double ratio” is often taken as the scale of variation (plus one). The ATLAS_UEPS_VH_hbb systematic mentioned above, for example, has associated with it, ATLAS_UEPS_VH_hbb_32JR. This systematic is evaluated by dividing the 3 jet to 2 jet ratio in the nominal setup by the same ratio in an alternate setup. Such a ratio generally looks like:

$$\frac{\text{Acceptance}[\text{Category}_A(\text{nominalMC})]}{\text{Acceptance}[\text{Category}_B(\text{nominalMC})]} \Big/ \frac{\text{Acceptance}[\text{Category}_A(\text{alternativeMC})]}{\text{Acceptance}[\text{Category}_B(\text{alternativeMC})]} \quad (4.3)$$

The three main categories are ratios between different flavor regions, ratios between different lepton channels[¶], and ratios between regions with different numbers of jets, n_{jet} . The first category is only relevant for $V+jets$ systematics and will be treated in that process’s dedicated section below. As this thesis is primarily concerned with the 2-lepton channel, the second category will not be treated in detail, though the treatment is much the same as other ratio systematics.^{||} In order to discuss the n_{jet} ratios in systematics (e.g. the ratios in the double ratio example), we must first describe how exclusive n_{jet} cross section calculations are done.

THEORETICAL ASIDE: STEWART-TACKMANN A way to calculate uncertainties on processes in regions with different numbers of jets was developed by Stewart and Tackmann and is implicitly used

[¶]e.g. $Z+heavy$ flavor jets (at least one b -jet in the event; often denoted “hf” normalizations in 0- and 2-lepton events

^{||}Such ratios allow for information in one channel to help constrain other channels, particularly for hard to model processes like $Z+hf$. This helps to reduce final overall uncertainties in combined fits. For a discussion of the interplay of nuisance parameters in combined fits, cf. Chapter 9.

⁸⁹⁴ for most n_{jet} ratio systematics⁷³. The problem is how to calculate the cross section and associated
⁸⁹⁵ uncertainty for a process with exclusively N jets in the final state. Generically:

$$\sigma_{\geq N} = \sigma_N + \sigma_{\geq N+1} \quad (4.4)$$

⁸⁹⁶ The physical interpretation of one parton to one jet is an idealized case. In order to demarcate
⁸⁹⁷ between jets, one has some quantity that is used as a cutoff in an integral that defines the border be-
⁸⁹⁸ tween jet regions.

$$\sigma_{\geq N} = \int_0^{p_{cut}} \frac{d\sigma_N}{dp} + \int_{p_{cut}} \frac{d\sigma_{\geq N+1}}{dp} \quad (4.5)$$

⁸⁹⁹ Since these cutoffs (not necessarily constant, etc.) can make calculations more complicated, inclu-
⁹⁰⁰ sive cross sections tend to be easier to calculate. Hence, it is usually much easier to evaluate the two
⁹⁰¹ inclusive cross sections and find the uncertainties on these by varying α_S in the usual way (cf. Sec-
⁹⁰² tion 4.2.1). One then assumes the inclusive uncertainties are uncorrelated, for a covariance matrix for
⁹⁰³ $\{\sigma_{\geq N}, \sigma_N, \sigma_{\geq N+1}\}$ of (with Δ_x^2 as the variance associated with x):

$$\Sigma = \begin{pmatrix} \Delta_{\geq N}^2 & \Delta_{\geq N}^2 & 0 \\ \Delta_{\geq N}^2 & \Delta_{\geq N}^2 + \Delta_{\geq N+1}^2 & -\Delta_{\geq N+1}^2 \\ 0 & -\Delta_{\geq N+1}^2 & \Delta_{\geq N+1}^2 \end{pmatrix} \quad (4.6)$$

⁹⁰⁴ These calculations contain Sudakov double logs of p/Q , where Q corresponds to the scale of the
⁹⁰⁵ hard scatter process (m_H), and p_{cut} is usually something like a p_T cutoff. The $N + 1$ term in the co-

906 variance matrix is an uncertainty associated with the cutoff, but the Sudakov double logs will domi-
907 nate any higher order terms. Stewart and Tackmann give the following reasoning:

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

915 The above considerations are important for this analysis since phase space is separated into 2 and
916 ≥ 3 jet regions, and the uncertainties for these regions are anti-correlated.

917 4.3 PROCESS SPECIFIC SYSTEMATIC SUMMARIES

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

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.

4.3.1 SIGNAL PROCESSES

Nominal signal $q\bar{q}VH$ samples are generated using Powheg with the MiNLO (multiscale improved NLO)⁶² procedure applied interfaced with Pythia8 using the AZNLO tune¹⁹ and NNPDF3.0 PDF set²⁸. For the 2-lepton case, gluon fusion initiated Higgs production is also considered (accounting for $\sim 14\%$ of the total cross section in this channel), with samples generated with Powheg interfaced with Pythia8 using the AZNLO tune. The NNPDF2.3 set²⁷ is used for both the ME and UE+PS.

Alternate samples $q\bar{q}VH$ samples are generated using MADGRAPH5_aMC@NLO²² for the ME and Pythia8 for the UE+PS, hadronization and MPI. The NNPDF2.3 5f FFN PDF sets and the Al4 tune¹³; the latter has variations included. Powheg+MiNLO+Herwig7 were samples were also used for systematics.

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

⁹³⁹ Table 4.2, reproduced from ⁶⁷, summarizes the signal cross section systematics, which are applied
⁹⁴⁰ uniformly across the analysis channels (as applicable).

Sys Name	source	Norm. effect	applied to
ATLAS_BR_bb	$H \rightarrow bb$ dec. unc, (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.

⁹⁴¹ The remaining signal systematics are analysis channel specific and are summarized in Table 4.3.
⁹⁴² The methodologies match those described in Section 4.2. The UE+PS systematics were derived us-
⁹⁴³ ing the alternate samples mentioned above; QCD scale uncertainties were derived by varying scales
⁹⁴⁴ by 1/3 and 3; and PDF uncertainties were derived by comparing the nominal set with the PDF4LHC15_30
⁹⁴⁵ PDF set²⁹.

⁹⁴⁶ 4.3.2 $V + \text{jets}$

⁹⁴⁷ Nominal $V + \text{jets}$ samples are generated using **SHERPA 2.2.1@NLO**^{** 52} for both the ME and PS, in-
⁹⁴⁸ terfaced with the NNPDF's and using a five quark flavor scheme, and alternative samples are derived
⁹⁴⁹ using **MADGRAPH5** interfaced with **PYTHIA8**. In order to increase statistics in important regions
⁹⁵⁰ of phase space, these samples were separated into kinematic slices based on p_T^V and into bins of jet fla-
⁹⁵¹ vor. The kinematic slices were in the quantity $\max(H_T, P_T^V)$ and had the intervals [0 – 70, 70 – 140, 140 – 280, 280 – 50
⁹⁵² GeV. The jet flavor slices were made using flavor vetoes and filters:

^{**}SHERPA 2.1 is used for some variations not available in SHERPA 2.2.1.

NP name	oL:		iL:		zL:	
	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			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			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			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.

- BFfilter: at least 1 b-hadron with $|\eta| < 4, p_T > 0$ GeV
- CFilterBVeto: at least 1 c-hadron with $|\eta| < 3, p_T > 4$ GeV; veto events which pass the BFfilter
- CVetoBVeto: veto events which pass the BFfilter and/or the CFilterBVeto

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

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

965 and then 0-lepton versus 2-lepton (1-lepton) for $Z+hf$ ($W+hf^{\dagger\dagger}$). The final uncertainty contains the
 966 sum in quadrature of four effects:

- 967 1. Variation of 0.5 and 2 of QCD scales in the SHERPA sample
 968 2. Sum in quadrature of half the variation from different resummation and CKKW merging
 969 scales ^{††}
 970 3. Maximal variation between nominal setup and SHERPA 2.2.1 with the MMHT_{2014nnlo68cl}
 971 and CT_{14nnlo} PDF sets
 972 4. Difference between the SHERPA and MADGRAPH₅ sets

973 Summaries of the Z +jets uncertainties are provided here; the reader is referred to ⁶⁷ for the W +jets
 974 systematics, as these events are virtually non-existent in the 2-lepton case with which this thesis is al-
 975 most exclusively concerned. In Table 4.4, from ⁶⁷ are the normalization systematics.

Process	Name	prior in region					
		2jet			(\geq)3jets		
		2L: low Vpt	2L: high Vpt	oL	2L: low Vpt	2L high Vpt	oL
$Z+l$	SysZclNorm				18%		
	SysZlNorm				23%		
	norm_Zbb				Floating Normalization		
$Z+hf$	SysZbbNorm_L2_J3	-	-	-	30%	30%	-
	SysZbbNorm_J3	-	-	-	-	-	17%
	SysZbbNorm_OL	-	-	7%	-	-	7%
	SysZbbPTV				effect on each region obtained from shape rw		

Table 4.4: Effect of modelling systematics on Z +jets normalization in the 2lepton 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.

976 The flavor composition ratio systematics are in Table 4.5, also from ⁶⁷.

^{††}The $W+hf$ CR versus the SR is also considered for $W+hf$

^{‡‡}cf. ⁵⁹, 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 flavor 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.

977 Finally, the p_T^V and m_{bb} shape systematics are derived using control regions in data. The func-
 978 tional 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 $\pm 0.0005 \times$
 979 $(m_{jj} - 100\text{ GeV})$.

980 4.3.3 TOP-PAIR PRODUCTION

981 Nominal $t\bar{t}$ samples are produced with Powheg at NLO for the ME calculation using the NNPDF3.0
 982 PDF set interfaced with Pythia8.210 using the A14 tune and the NNPDF2.3 PDF set at LO. The
 983 parameters `hdamp` (nominal value $1.5 m_{top}$, a resummation damping factor for ME/PS matching
 984 that can heuristically thought of as tuning high p_T radiation) in Powheg and `pThard` (nominal
 985 value 0) and `pTdef` (nominal value 2) in Pythia (both control merging with Powheg) are varied
 986 to evaluate certain systematics. Alternative $t\bar{t}$ samples use Powheg+Herwig7, MadGraph5_aMC@NLO-

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.

⁹⁸⁷ +PYTHIA8.2, and the nominal setup with varied tunes and parameter values. Uncertainties are taken
⁹⁸⁸ to cover the largest difference between the nominal and any of these alternate configurations.

⁹⁸⁹ The overall $t\bar{t}$ normalization is a floating normalization, and further systematics attached to the
⁹⁹⁰ ratio of acceptances between regions (3-to-2 jet, SR-to-WhfCR, and 1-to-0 lepton) are defined using
⁹⁹¹ double ratios; these are summarized in Tables 4.6 and 4.7, taken from⁶⁷.

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

Table 4.6: Effect of modelling systematics on normalization 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 normalization		-	
norm_ttbar_J3_L2	-		floating normalization	
SysTTbarPTV_L2_L2	effect on each region obtained from shape rw			

Table 4.7: Effect of modelling systematics on normalization 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.

⁹⁹² Shape systematics for p_T^V and m_{bb} are linear and taken to cover the largest difference reasonably
⁹⁹³ well, as described above in 4.2.2. These are summarized in Table 4.8, again taken from⁶⁷.

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.

994 4.3.4 DIBOSON PRODUCTION

995 Three diboson production processes (collectively denoted VV) are important for these analyses: ZZ ,
 996 WZ , and WW . Nominal samples are created using **SHERPA 2.2.1** using the NNPDF3.0 PDF set. Al-
 997 ternative samples use PowHEG+PYTHIA8 and PowHEG+HERWIG++. The methodology here is
 998 similar to that of the $t\bar{t}$ systematics, with both overall acceptance and lepton channel specific uncer-
 999 tainties, with the exception that UE+PS and QCD scale are treated separately (PDF+ α_s was found
 1000 to be negligible). p_T^V shape systematics are described using linear fits, while $m_{b\bar{b}}$ shape systematics
 1001 are described using hyperbolic tangents (third degree polynomials) in the 2 jet (3 jet) regions. Once
 1002 again, summary tables from⁶⁷ 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.

1003 4.3.5 SINGLE TOP PRODUCTION

1004 Single top sample are generated separately for the different production channels (s , t , and Wt) using
 1005 Powheg with the CT10 NLO PDF's interfaced with Pythia6 using the PERUGIA2012 PS tune
 1006 and the corresponding CTEQ6L1 LO PDF's and PHOTOS (TAUOLA) for QED final state (τ) de-
 1007 cays.

Production	Uncertainty	Value	Source	Nuisance Parameter
s -channel	overall normalization	4.6%	sum in quadrature of μ_R , μ_F , α_S and PDF uncertainties	<code>stopNorm</code>
t -channel	overall normalization	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 normalization	6.2%	sum in quadrature of μ_R , μ_F , α_S and PDF uncertainties	<code>stopWtNorm</code>
Wt channel	2 jet region normalization	35%	sum in quadrature of deviations in alternative generators	<code>stopWtAcc</code> correlated with overall and 3 jet case
Wt channel	3 jet region normalization	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++) $\pm 0.001 \times p_T^V \mp 0.17 + 1$	<code>StopPTV</code>
t -channel	$m_{b\bar{b}}$ shape	shape	fit through largest deviation (POWHEG+PYTHIA6 radHi-radLo) $\pm 0.0008 \times m_{b\bar{b}} \mp 0.12 + 1$	<code>StopMBB</code>
Wt channel	p_T^V shape	shape	fit through largest deviation (POWHEG+PYTHIA6 with diagram subtraction) $\pm 0.003 \times p_T^V \mp 0.69 + 1$	<code>StopWtPTV</code>
Wt channel	$m_{b\bar{b}}$ shape	shape	fit through largest deviation (POWHEG+PYTHIA6 with diagram subtraction) $\pm 0.0036 \times m_{b\bar{b}} \mp 0.52 + 1$ ($m_{b\bar{b}} < 275$ GeV) $\mp 0.47 + 1$ ($m_{b\bar{b}} \geq 275$ 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*

1008

5

1009

Object and Event Reconstruction and

1010

Selection

1011 IN BREAKING WITH THE STANDARD CONVENTION both object definitions and their associated

1012 experimental systematic uncertainties will be defined in this chapter: the hope is that the proximity

1013 of these descriptions will allow them to elucidate each other. Summary tables are almost exclusively
1014 taken from³¹ or⁶⁵. This analysis, like most typical analyses in ATLAS, use central object definitions
1015 from collaboration combined performance (CP) groups using standard analysis tools and recom-
1016 mendations from these groups for the various objects and their accompanying systematic uncertain-
1017 ties.

1018 Before proceeding to the objects used in this analysis, we begin with a few remarks on uncertain-
1019 ties associated with object reconstruction. Event-level variables and selections are discussed more in
1020 depth in Chapters 2 and 6. As described in Section 4.2, systematics quantify the uncertainty asso-
1021 ciated with certain effects, and are generally treated in an analysis by saving histograms of discrimi-
1022 nating distributions corresponding to the nominal analysis except with the systematic in question
1023 varied by plus and minus one standard deviation each (one histogram each). While for modeling
1024 systematics this only corresponds to different event weights, for experimental systematics like those
1025 described in this chapter (with the exception of flavor tagging and certain trigger systematics), this is
1026 done by varying the parameter in question and re-running reconstruction with the systematic varied
1027 before recomputing all event level quantities and then saving discriminant values in their appropri-
1028 ate distributions. This is, in general, a much more computationally intensive process in the analysis,
1029 which is why an entire software framework, the `CxAODFramework`, was created for this analysis (see
1030 Section 3 of³¹ for more details).

1031 **5.1 TRIGGERS**

1032 Tables of the triggers used with the 2015 and 2016 datasets are given in Tables 5.1 and 5.2.

o lep	i lep	z 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	i lep	z 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.

1033 The 0-lepton channel uses a \vec{E}_T^{miss} trigger, while the 2-lepton channel uses single lepton triggers,
 1034 with the 1-lepton analysis using both. Since the 0- and 1-lepton channels are largely beyond the scope
 1035 of this thesis, the discussion here will be limited to the single lepton triggers; the interested reader is
 1036 directed towards ³¹ and its cited sources for an in-depth discussion of the use of the \vec{E}_T^{miss} trigger.

1037 The efficiency of triggers is in general different on simulated datasets than in actual data collected
 1038 in ATLAS, so a scale factor to correct for this difference in efficiency must be applied to simulation
 1039 events. This scale factor is given by the muon CP group for muons for both the 1- and 2-lepton cases
 1040 and from the electron CP group for the 1-lepton case. For the two electron case, this was calculated
 1041 by the analysis team as (details in ³¹):

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

1042 There are also systematic uncertainties associated with these trigger efficiencies. The single elec-
 1043 tron trigger efficiency systematic uncertainty is encapsulated in a single systematic, `EL_EFF_Trigger_-`
 1044 `Total_1NPCOR_PLUS_UNCOR`, while the single muon trigger efficiency has two components, one
 1045 each for the sample statistics, `MUON_EFF_TrigStatUncertainty`, and systematic uncertainties
 1046 `MUON_EFF_TrigSystUncertainty` associated with that efficiency's measurement.

1047 While the momentum associated with the lowest un-prescaled single lepton triggers changes de-
 1048 pending on data-taking conditions (the numbers associated with the triggers in the tables can be
 1049 thought of as nominal p_T values for trigger level objects), the lowest typical value is ~ 25 GeV. In
 1050 order to maintain this triggering capability on low p_T muons in the higher luminosity environment

1051 of the Run 3 LHC and beyond, trigger-capable detectors will be installed in upgraded New Small
1052 Wheels (NSW) of the ATLAS muon detector during the Phase I upgrade. Detailed studies in sim-
1053 ulation of the trigger algorithm performance under nominal and misaligned conditions for the Mi-
1054 cromegas detectors to be installed in the NSW may be found in Appendix A.

1055 **5.2 ELECTRONS**

1056 Electrons in ATLAS are reconstructed using a combination of the ATLAS electromagnetic calorime-
1057 ter (ECAL) and inner detector (ID). Reconstruction begins by searching for so-called “seed clusters”
1058 in the ECAL. The ECAL is divided into a 200×256 tower grid in the $\eta - \phi$ plane, with each tower
1059 having a size of 0.025 square in η and ϕ , corresponding to the granularity of the ECAL in its middle
1060 layer, with all energy in a tower summed longitudinally. A “sliding window” of 3×5 cells in the
1061 $\eta - \phi$ plane is then used to identify EM clusters associated with electrons based on criteria detailed
1062 in ⁴⁹. This comparatively simple algorithm (in contrast to jet reconstruction detailed below) is effec-
1063 tive since electromagnetic showers have a well defined behavior and shape.

1064 Once seed clusters have been formed, they are associated with tracks in the inner detector. Com-
1065 bined cluster-tracks pairs form electron candidates. In order for a electron candidate to be consid-
1066 ered a suitable electron for analysis, it must pass certain quality requirements, based on a cut on the
1067 value of a likelihood-based (LH) discriminant (cf. ¹⁴ for details). This discriminant is given by:

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

where the s and S (b and B) subscripts refer to distributions in fiducial signal (background) distributions in bins of $|\eta|$ and E_T . The $P(x_i)$ are probability distributions functions (pdf)'s for input variables. Several 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 $\eta - \phi$ 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 GeV p_T cut ($1.05 \times p_T^{\text{trigger}}$). These requirements are summarized in Table 5.3.

Electron Selection		η	ID	d_0^{sig}	$ \Delta z_0^{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

1084 (roughly, the uncertainty in taking the energy deposits in an EM cluster and turning them into an
1085 electron energy) and energy resolution (the width associated with this). This is in practice a very
1086 complicated procedure, with over 60 systematics associated, but this analysis is not at all sensitive
1087 to these effects and so a simplified model of two systematics, EG_RESOLUTION_ALL and EG_SCALE-
1088 _ALL, is used.

1089 **5.3 MUONS**

1090 This analysis uses the standard CP muon collection in an event, though these muons in ATLAS are
1091 constructed in a variety of ways; for full details see ²⁵ and ¹⁷. Most muons are constructed using tracks
1092 in the chambers of the muon spectrometer (MS), with a variety of algorithms available. MS tracks
1093 are sufficient to reconstruct a muon (a fit on these tracks can be used to point back to an interaction
1094 point for vertex matching, for example) and, in the $|\eta| \in (2.5, 2.7)$ interval where there is no track-
1095 ing, these standalone (SA) muons are the default. The most common and robust form of muon
1096 reconstruction combines tracks in the MS with tracks in the ID (more precisely, a global refit with
1097 hits from both subsystems is typically done) to form combined (CB) muons. CB and SA muons
1098 automatically pass the loose reconstruction requirements for the Loose muons used in this analy-
1099 sis. Additionally, since there is a gap in the $|\eta| < 0.1$ range in the MS to make room for cabling and
1100 other detector services, there are two further muon types used in this range: the segment tagged (ST)
1101 muons that match ID tracks to segments in the MDT or CSC chambers and the calorimeter tagged
1102 (CT) muons that match ID tracks to calorimeter clusters consistent with minimum ionizing parti-
1103 cles (which muons in ATLAS generally are).

1104 Further quality requirements are imposed on Loose muons for the different muon categories
 1105 used in this analysis. Isolation requirements similar to the electrons in corresponding categories are
 1106 imposed, and impact parameter requirements are also imposed. The ZH signal muons also have a
 1107 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_0^{sig}	$ \Delta z_0^{\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.

1108 5.3.1 MUON SYSTEMATICS

1109 Similar to the treatment of systematic uncertainties associated with the electrons, muons have CP de-
 1110 fined systematics. The muon momentum scale and resolution systematics are divided into three cat-
 1111 egories associated one for uncertainties related to ID tracks (`MUONS_ID`), one for MS tracks (`MUONS-`
 1112 `_MS`), one for the overall scale (`MUONS_SCALE`), and two for charge dependent momentum scales
 1113 (`MUON_SAGITTA_RHO` and `MUON_SAGITTA_RESBIAS`). The remaining systematics have a STAT and
 1114 SYS component corresponding to the sample statistics and systematic uncertainties for their individ-
 1115 ual components. Efficiency scale factors use different standard candles in different p_T ranges (J/ψ 's
 1116 (Z 's) below (above) 15 GeV), and so these systematics are broken up into two categories (`MUON_EFF-`
 1117 `_STAT` and `MUON_EFF_SYS`; `MUON_EFF_STAT_LOWPT` and `MUON_EFF_SYS_LOWPT`). There are also
 1118 isolation systematics (`MUON_ISO_STAT`, `MUON_ISO_SYS`) and track to vertex association systematics
 1119 (`MUON_TTVA_STAT`, `MUON_TTVA_SYS`).

1120 5.4 MISSING TRANSVERSE ENERGY

1121 High precision performance of \vec{E}_T^{miss} is not so crucial to the 2-lepton analysis (though it is very im-
1122 portant to the other channels), so the interested reader is referred to²⁶. \vec{E}_T^{miss} in ATLAS is the neg-
1123 ative vectoral sum of physics objects (in this analysis just jets and leptons, though in principle also
1124 including τ 's and γ 's) and a so-called track based soft term (TST). The TST is comprised of valid
1125 ID tracks not associated with any physics objects in an event. These tracks must be associated to an
1126 event's primary vertex, have a $p_T > 0.4$ GeV, and pass other quality requirements.

1127 The \vec{E}_T^{miss} systematic uncertainties relevant to this analysis are related to track based energy scale
1128 and resolutions in both the soft term and in the jets and are: MET_SoftTrk_ResoPara, MET_Soft-
1129 Trk_ResoPerp, MET_SoftTrk_ScaleDown, MET_SoftTrk_ScaleUp, MET_JetTrk_Scale-
1130 Down , and MET_JetTrk_ScaleUp.

1131 5.5 JETS

1132 Unlike leptons, all analyses considered in this thesis are sensitive to factors regarding jet reconstruc-
1133 tion and associated systematic uncertainties. A general discussion of jets precedes jet reconstruction
1134 in ATLAS and associated systematics relevant to this thesis.

1135 5.5.1 JET ALGORITHMS

1136 The hadronic nature of jets makes jet reconstruction a lot more complicated than electron or photon
1137 reconstruction, where a regular shower shape and track matching (or lack thereof in the case of the

1138 chargeless photon) provide a fairly straightforward and robust approach. The interested reader is

1139 referred to⁷⁰ for an excellent survey, from which this discussion is greatly abbreviated.

1140 Looking at an event like the one in Figure 5.1, unambiguous individual jets are particularly easy to

1141 identify, more or less popping out of the $\eta - \phi$ plane plot, but this is not always the case.

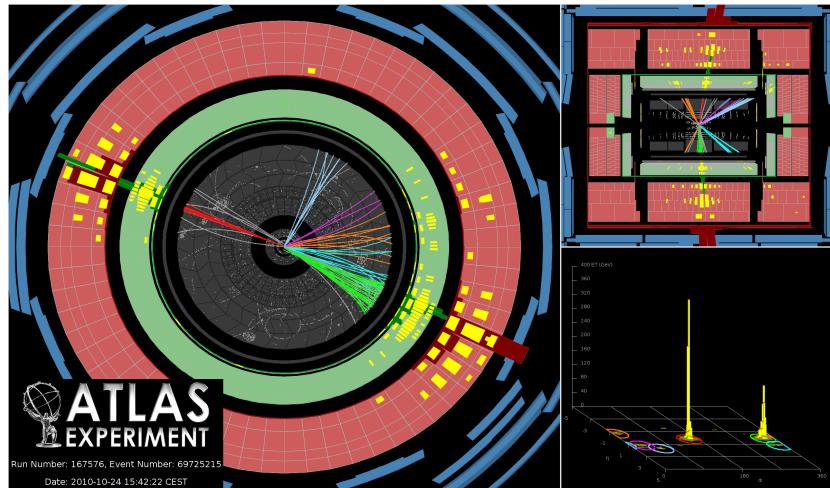


Figure 5.1: A clean ATLAS dijet event.

1142 Two general methods of turning particles/calorimeter towers into jets exist: cone-based and se-

1143 quential recombination. The general theme of the former is to find a hard (energetic) particle and

1144 draw a circle around it in the $\eta - \phi$ plane in an intelligent manner, while the theme of the latter

1145 is to find some metric of distance between particles and then to cluster pairs based on this distance

1146 into jets in an intelligent way. Cone-based algorithms are simple (and therefore generally quite fast)

1147 but generally lack some properties of the sequentially recombined jets (though there are notable ex-

1148 ceptions like SISCone). Cone algorithm reconstructed jets are important for trigger level objects in

1149 ATLAS, though since no jet triggers are used in this analysis, they will not be discussed any further

1150 here.

1151 The general drawback of cone-based algorithms is that they are not infrared and collinear (IRC)
1152 safe. That is, neither the emission of a soft (IR) quark or gluon during hadronization nor the collinear
1153 splitting of hard particles during hadronization should not change the final jet collection in an event.
1154 These are fairly common edge cases and can lead to certain pathologies in QCD calculations. In-
1155frared and collinear safety are diagrammed schematically in Figures 5.2 (a) and (b), taken from⁷⁰.

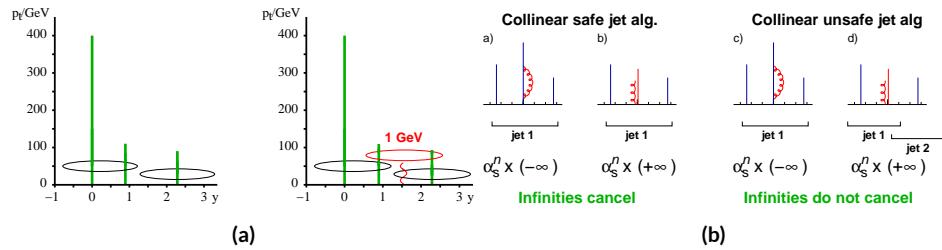


Figure 5.2: Infrared (a) and collinear (b) safety.

1156 Sequential recombination algorithms are generally safe from these effects, as these edge cases are
1157 very “close” to each other by construction. A sequential recombination algorithm proceeds as fol-
1158 lows

- 1159 1. Evaluate the set of distances d_{ij} (for pairs of objects) and d_{iB} (the “beam distance” for each
1160 individual object)

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

- 1161 2. Find the minimum distance
1162 3. If the minimum distance is:

- 1163 • A d_{ij} : cluster these objects together, and go to step 1
1164 • A d_{iB} : call the i^{th} object a jet, remove it from the set of objects to be clustered, and go to
1165 step 1

1166 4. Repeat until all objects are clustered into jets

1167 The choices one must make in sequential recombination are the size parameter R , akin to a cone
1168 radius in cone-based algorithms, and the momentum power p . Common choices and their trade-offs
1169 are:

- 1170 • +1: the k_t algorithm; favors the softer particles in an event, so the cluster sequence gives a history
1171 of hadronization, but jet shapes are irregular (i.e. not circular in the $\eta - \phi$ plane)
- 1172 • 0: the Cambridge-Aachen algorithm: a pure distance metric; less substructure but jets tend to
1173 be more circular
- 1174 • -1: the anti- k_t algorithm: clustering begins with hardest particles in an event; regular, localized
1175 jet shapes, but virtually no substructure in clustering history

1176 Jet reconstruction using all three algorithms on the same event, as well as SISCone, are shown in

1177 Figure 5.3.

1178 All three algorithms have uses for different applications in ATLAS, with anti- $k_t R = 0.4$ jets
1179 being the default jet collection.* These are the jets used in this analysis.

1180 If the choice of jet algorithm seems a little arbitrary, it is. There is no one-size-fits-all jet collection
1181 perfect for every application, and analyzers have to make these choices for themselves. One interesting
1182 choice is the jet size parameter, R . A large R jet will contain more of the radiation coming from a
1183 final state object, but its large size makes it susceptible to contamination from the underlying event
1184 and pileup (as well as other analysis objects if R is sufficiently large or objects sufficiently boosted),
1185 with small R jets having the opposite features. $R = 0.4$ is a fairly middle-of-the-road choice. A natural
1186 question to ask is whether there needs to be just one jet collection in an analysis. Might there not

*The other collections find their primary uses in jet substructure techniques. For an example, cf. the discussion of jet trimming in Appendix B.

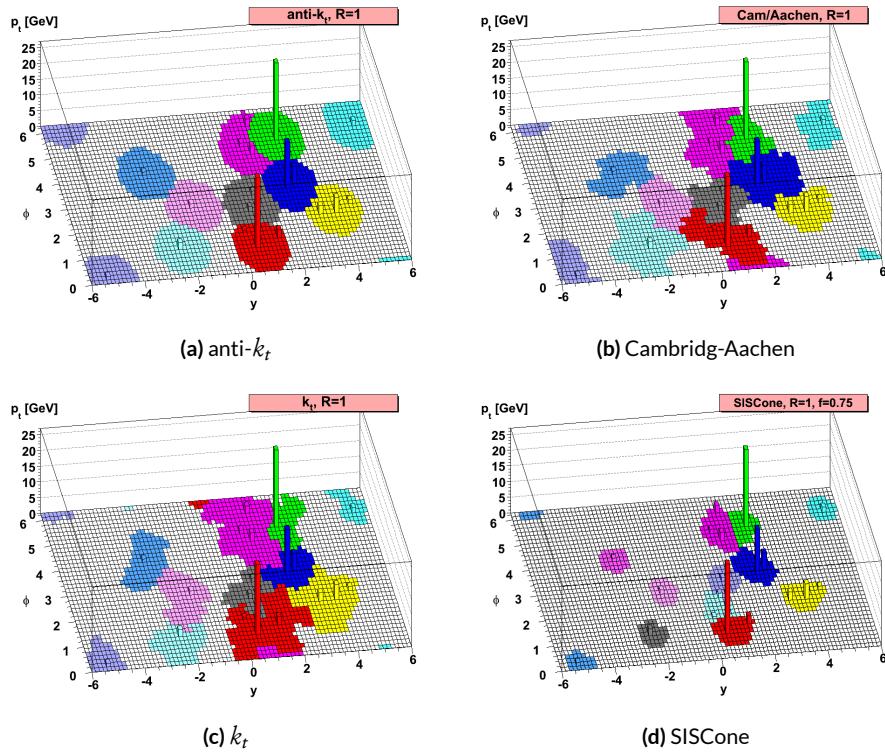


Figure 5.3: Different jet algorithms used on the same event. IC:⁷⁰

1187 be more information to be gained from looking at more jet sizes or clusterings? Preliminary studies
1188 point to this answer being yes and are addressed in Appendix B.

1189 5.5.2 STANDARD ATLAS HBB JETS

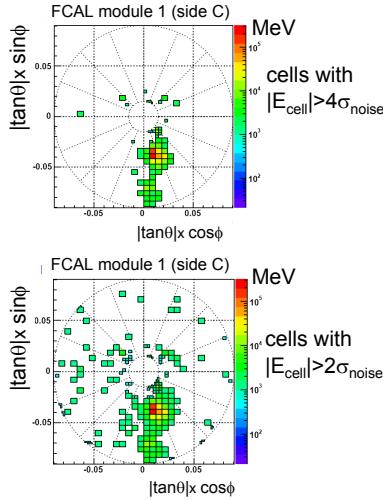
1190 There are a few considerations that arise with jets in physical detectors. The first is what type of ob-
1191 ject to use as the stand in for “particles” since ATLAS measures energy deposits, not pions. The
1192 approach ATLAS has settled upon are calorimeter topological clusters (or CaloTopoClusters for
1193 short)⁵⁸. Unlike the sliding window algorithm used for electron clusters, CaloTopoClusters use a
1194 noise significance based approach in the “4-2-0” algorithm. Each cell in the electromagnetic and
1195 hadronic calorimeters (ECAL and HCAL, respectively) has associated with it a characteristic noise
1196 level (N in Equation 1.6), with this noise level in each channel, it is possible to construct a “signifi-
1197 cance” for the registered energy deposit in a given channel for a given event by dividing the measured
1198 value by its characteristic noise. Groups of cells having a significance of 4 are taken as the centers of
1199 clusters in the $\eta - \phi$ plane. The second layer in a cluster includes all neighboring cells to the central
1200 layer with significance of at least 2, and the final layer includes all the nearest neighbors to the second
1201 layer. This is described in Figure 5.4 from⁶¹.

1202 Once CaloTopoClusters have been formed and clustered into jets, they are calibrated using the
1203 electromagnetic (EM) scale (the scale for clusters coming form EM showers). Further details may be
1204 found in¹⁶.

1205 Jets in this analysis fall into two categories, “signal” and “forward” jets, and are required to pass
1206 certain quality requirements, described in Table 5.5. All jets must pass a series of jet cleaning require-

Local Hadronic Calibration: Clusters

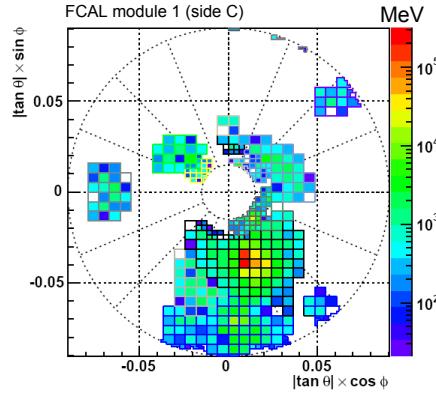
Sven Menke



□ 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.

1207 ments using calorimeter level variables to eliminate jets coming from problematic calorimeter cells
 1208 and certain backgrounds. Some signal jet candidates also make use of a Jet Vertex Tagger (JVT) that
 1209 uses primary vertex and jet and track p_T information to decide whether certain soft jets are likely to
 1210 have come from the the primary (hard scatter) vertex in an event or are to be considered pileup. Fur-
 1211 ther details on JVT may be found in ²⁴. Jets are further corrected using standard CP tools and a dedi-
 1212 cated PtReco correction, all outlined in Section 7.3 of ³¹.

1213 Overlap removal in this analysis is done according to the following precedence, taken from ³¹ with
 1214 further steps only taken into account if an object survives previous steps:

- 1215 • tau-electron: If $(\tau, e) < 0.2$, the τ lepton is removed.
- 1216 • tau-muon: If $(\tau, \mu) < 0.2$, the τ lepton is removed, with the exception that if the τ lepton has

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.

- ₁₂₁₇ $> 50 \text{ GeV}$ and the muon is not a combined muon, then the τ lepton is not removed.
- ₁₂₁₈ • electron-muon: If a combined muon shares an ID track with an electron, the electron is re-
- ₁₂₁₉ moved.
- ₁₂₂₀ If a calo-tagged muon shares an ID track with an electron, the muon is removed.
- ₁₂₂₁ • electron-jet: If $(\text{jet}, e) < 0.2$ the jet is removed.
- ₁₂₂₂ For any surviving jets, if $(\text{jet}, e) < \min(0.4, 0.04 + 10 \text{ GeV}/e)$, the electron is removed.
- ₁₂₂₃ • muon-jet If $(\text{jet}, \mu) < 0.2$ or the muon ID track is ghost associated to the jet, then the jet is re-
- ₁₂₂₄ moved if the jet has less than three associated tracks with $> 500 \text{ MeV}$ ($\text{NumTrkPt500PV}^{\text{jet}} < 3$) or
- ₁₂₂₅ both of the following conditions are met: the ratio of the muon and jet is larger than 0.5 ($\mu/\text{jet} >$
- ₁₂₂₆ 0.5) and the ratio of the muon to the sum of tracks with $> 500 \text{ MeV}$ associated to the jet is
- ₁₂₂₇ larger than 0.7 ($\text{muon}/\text{SumPtTrkPt500PV}^{\text{jet}} > 0.7$).
- ₁₂₂₈ For any surviving jets, if $(\text{jet}, \mu) < \min(0.4, 0.04 + 10 \text{ GeV}/\mu)$, the muon is removed.
- ₁₂₂₉ • tau-jet: If $(\tau, \text{jet}) < 0.2$, the jet is removed.
- ₁₂₃₀ • electron-fat jet: If $(e, \text{fat jet}) < 1.2$, the fat jet is removed.

1231 Jets are corrected using a muon-in-jet correction and then a kinematic fitter (Appendix D of⁶⁵)
1232 for the 2-lepton case (PtReco correction for the 0- and 1- lepton case). The muon-in-jet correction
1233 is designed for b -jets. Since the decay of a b -quark to a c -quark and finally to a light quark (these are
1234 the multiple vertices for which JetFitter in Section 5.6.1 searches) involves two weak decays, there are
1235 two W -bosons involved in the decay. Some of these will decay semileptonically, and, while electron
1236 and τ energy will be captured by the calorimeters, semileptonic μ 's will only be registered in the MS,
1237 which occurs in some 44% of all decays from a theoretical standpoint, which amounts to about 12%
1238 in practice (due to track isolation requirements for the leptons). This value is about 1–2% for elec-
1239 trons, which deposit their energy in the calorimeter and so require no correction; any jet with a valid
1240 lepton associated to it is deemed semileptonic (all others are called hadronic). Any jet with muons
1241 associated with it has the closest muon's 4-vector (in the $\eta - \phi$ plane) added to it.

1242 The PtReco correction is a scale factor on the muon-in-jet corrected jet's 4-vector based on the
1243 jet's p_T and whether the jet is hadronic or semileptonic. This correction factor is based on particle
1244 level studies done on a TruthWZ sample. As the 0- and 1-lepton cases are not the focus of this thesis,
1245 the interested reader is directed to Section 7.3 of³¹.

1246 The kinematic fitter used in 2-lepton events with two or three jets takes as its input 12 fit parame-
1247 ters,

- 1248 • energies of 2 electron or of 2 muons
- 1249 • energies of 2 b -jets
- 1250 • η, ϕ of 2 leptons and 2 jets
- 1251 • p_X and p_Y of $\ell\ell b\bar{b}$ system.

1252 • $m_{\ell\ell}$

1253 and 3 constraints for the variation of these parameters,

- 1254 • parameters : Gaussian (b-jet energy : Transfer Functions (TF); these are denoted L , with an
1255 L_{truth} as a prior) (the ϕ parameters)
- 1256 • p_X and p_Y of $\ell\ell b\bar{b}$ system : zero with a width of 9 GeV obtained from ZH signal MC.
- 1257 • m_{ll} : Breit-Wigner (BW) distribution of Z boson (final term, leptons denoted Ω)

1258 which leads to test statistic from the usual likelihood formalism to be minimized in each event:

$$-2 \ln \mathcal{L} = \sum_{i=j} \left(\frac{(\phi_i^n - \phi_i^0)^2}{\sigma_\phi^2} \right) + \left(\frac{(\Omega_l^n - \Omega_l^0)^2}{\sigma_\Omega^2} \right) - 2 \ln(L^j) - 2 \ln(L_{truth}^j)$$

1259 $+ \sum_{i=x,y} \frac{(\sum p_i^n - \sum p_i^0)^2}{\sigma_{\sum p_i}^2} + 2 \ln((m_{\ell\ell} - M_X^2)^2 + M_X^2 \Gamma^2) \quad (5.4)$

1260 5.5.3 JET SYSTEMATICS

1261 As with the electron systematics, jet energy scale (JES) and resolution (JER) are the two principal
1262 considerations for systematic uncertainties, with even more standard. JER, as with the electron en-
1263 ergy resolution, is a single systematic uncertainty, `JET_JER_SINGLE_NP`. There is also a single JVT
1264 efficiency `JET_JvtEfficiency` systematic uncertainty. There are 88 nominal JES systematics, and
1265 this analysis is sufficiently sensitive to these variations that a single systematic is grossly insufficient.
1266 Nevertheless, some simplification is possible, with the 75 of these nuisance parameters (mostly statis-

1267 tical uncertainties related to the Z +jet and γ +jet calibrations) being reduced to 8, and several explic-
1268 itly named nuisance parameter. These remaining named NP's are: 3 NP's related to the η intercali-
1269 bration used to extrapolate standard calibrations to other jet η regions, 4 NP's related to the flavor
1270 composition of principle background samples (W/Z +jets, top, and diboson), 4 pileup systematics, a
1271 single NP for the b -jet energy scale, a high p_T jet energy scale systematic, and one for jets that punch
1272 through the HCAL to leave energy deposits in the MS. These are listed explicitly in Table 5.7.

1273 5.6 FLAVOR TAGGING

1274 Given that the final state in this analysis involves pairs of jets originating from b -quarks, deploying
1275 effective flavor tagging algorithms is imperative. While flavor tagging in general can be used to isolate
1276 any flavor (b , c , or light (u , d , s , or gluon-initiated jets)), this analysis exclusively looks for b -jets, so
1277 this disucssion will focus on b 's. At truth-level in sumlation, this is fairly straightforward: one need
1278 only look at the particles contained within a jet and seeing if any include a b -quark (sometimes a B
1279 hadron) in the decay chain.

1280 5.6.1 DIRECT TAGGING

1281 One of the most distinctive features of b -jets is the presence of secondary vertices, as illustrated in Fig-
1282 ure 5.5. While most partons created in particle collisions will hadronize promptly, b -quarks will first
1283 hadronize into B -hadrons, which have lifetimes of about a picosecond. This small but finite lifetime
1284 means that these particles will travel about half a millimeter or so before decaying into a jet in much
1285 the usual way, and the tracks from this decay will point back to this displaced, secondary vertex.

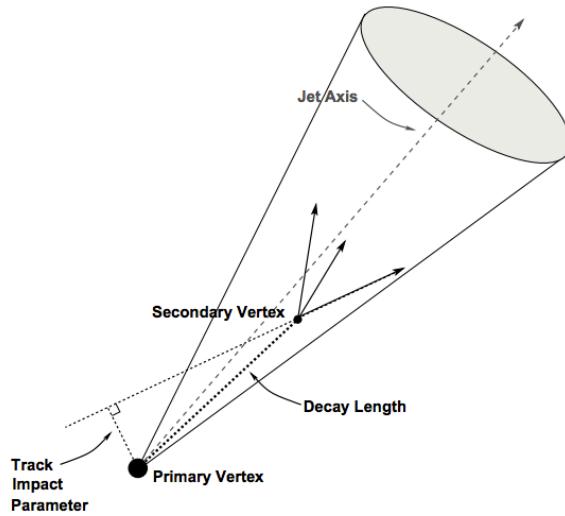


Figure 5.5: An illustration of a secondary vertex in a b -jet. Image credit: ⁵⁴

¹²⁸⁶ There are various secondary vertex algorithms used as inputs to the nominal b -tagging algorithm¹⁵,

¹²⁸⁷ with three main types of algorithms used as inputs

- ¹²⁸⁸ 1. Track impact parameter based algorithms: I₂PD (signed transverse only; more pileup ro-
¹²⁸⁹ bust), I₃PD (signed transverse and longitudinal)
- ¹²⁹⁰ 2. Inclusive secondary vertex reconstruction: SV1 (start with two track vertex pairs and con-
¹²⁹¹ struct a secondary vertex)
- ¹²⁹² 3. Multiple vertex reconstruction (decay chain): JetFitter (PV $\rightarrow b \rightarrow c$ decay chain using Kalman
¹²⁹³ filter)

¹²⁹⁴ All of these are combined into a boosted decision tree (BDT) and trained on five million $t\bar{t}$ events

¹²⁹⁵ with an 90%/10% light/ c jet background to form the MV_{2c10} algorithm, with 10 referring to the per-

¹²⁹⁶ centage of charm events in the training background. The 10% charm ratio was found to be a good

¹²⁹⁷ balance between increased charm rejection capability (as opposed to MV_{2c00}, which has no charm

1298 in the background training samples) and loss in light jet rejection (compared to MV2c20, which has
1299 20% charm events in background training samples).

1300 ANALYSIS SPECIFIC CONCERNS AND SYSTEMATIC UNCERTAINTIES In addition to specifying
1301 the tagging algorithm, the working point efficiency must be specified. As with selection algorithms
1302 in general, there is a trade off between efficiency/recall (identifying all the b -jets, minimizing type
1303 II error) and purity/precision (making sure all jets positively identified are in fact b -jets, minimiz-
1304 ing type I error). Nominal efficiency working points have been calibrated by the flavor tagging CP
1305 group and are outlined 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).

1306 These values are aggregate figures, as both the jet’s p_T and η are inputs to the MV2c10 discrim-
1307 inant. The working point chosen for this analysis is the 70% **FixedCutBEff_70** working point,
1308 with “fixed cut” referring to the fact that this particular usage of the MV2c10 BDT value is a simple
1309 cut value.

1310 Just as with the trigger and lepton identification efficiencies, flavor tagging efficiencies differ from
1311 their nominal values somewhat depending on what simulation or data sample is being used. To
1312 account for this difference, just as in the other cases, scale factors are applied to simulation event

1313 weights. It is through these event weights, as with the modeling systematics, that the flavor tagging
1314 systematic uncertainties are applied. Given that there are 24 input variables to MV2c10 and that fla-
1315 vor tagging is in general a very difficult problem, it is not surprising that, as with the JES, there are
1316 very many systematic uncertainties associated with flavor tagging. However, as with JES, the CP
1317 group has compacted the full systematic set into a reduced set of 13 systematic uncertainties: 3 each
1318 associated with c and light jets, 5 for b -jets (with the naming convention FT_EFF_Eigen_(B|C|Light)N),
1319 one for the extrapolation of scale factors to different jet p_T regimes (FT_EFF_Eigen_extrapolation),
1320 and one for the charm to bottom extrapolation (FT_EFF_Eigen_extrapolation_from_charm)⁷⁹.
1321 This schematic is a middle-of-the-road “Medium” set of systematics.

1322 5.6.2 TRUTH TAGGING

1323 Since imposing a 2 b -tag requirement overwhelmingly rejects events dominated by c - and light jets,
1324 statistics in such MC samples are very low. In order to circumvent this problem and restore full MC
1325 statistics, the tag rate function, or “truth-tagging” procedure (in contrast to the standard or “direct
1326 tagging” procedure) is applied, in which all events are kept but given a weight that preserves the over-
1327 all shape and normalization of underlying distributions. Intuitively, this is done by giving events with
1328 real b -jets in MC a much higher weight than events having only c - or light jets. Truth-tagging is ap-
1329 plied to all samples when conducting MVA training in order to maximize statistics and reduce the
1330 risk of overtraining. Truth-tagging is also used for data-MC comparison plots in 2-tag regions and
1331 for $V + cc$, $V + c\ell$, $V + \ell$, and WW samples used in the final likelihood fit. A detailed description of
1332 the truth-tagging process is provided below.

1333 Each jet in a given event has associated with it a b -tagging efficiency, denoted ε , that is a function
 1334 of its p_T , η , and real flavor (b , c , or light) from truth-level information in MC. Intuitively, this effi-
 1335 ciency can be thought of as the likelihood that a given jet will be b -tagged. Hence, b -jets have a much
 1336 higher b -tagging efficiency than c -jets, which in turn have a higher b -tagging efficiency than light jets.
 1337 We define a truth-tag weight for a given combination of tagged and untagged jets as the product of
 1338 the efficiencies of the tagged jets times the product of the complement of the efficiencies of the un-
 1339 tagged jets. For example, for an event with three jets, labeled 1, 2, and 3, if jets 1 and 2 are tagged, and
 1340 jet 3 is untagged, the truth-tag weight associated with this combination is

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

1341 In order to obtain a truth-tag weight for an event, one takes the sum of the weights for each pos-
 1342 sible tag combination. The current analysis requires that all events have exactly two b -tagged jets, so
 1343 the truth-tag weight is the sum of all the weights of all possible pairs of tagged jets (events with fewer
 1344 than two jets are discarded). Going back to the three jet example, one has the possible combinations:
 1345 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
 1346 tagged and jet 1 as untagged, which yields a total event weight of

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

1347 For some applications (e.g. in order to use variables like pTB_1 , the p_T of the harder b -tagged jet in

1348 an event, in MVA training), it is necessary to choose a combination of jets in an event as “tagged.”

1349 This combination is chosen randomly, with the probability for a given combination to be chosen

1350 being proportional to its truth-tag weight. In the three jet example, the probability of tagging jets

1351 1+2 is:

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

1352 Though not used in the current analysis, functionality exists for generic truth-tagging require-

1353 ments (i.e. an arbitrary number of tags on an arbitrary number of jets) through the logical combina-

1354 toric extension and for so-called “pseudo-continuous tagging,” where a b -tag score is generated for

1355 each jet in a given event. Since a random combination of jets is set by hand to pass the b -tagging cuts

1356 regardless of its b -tag score, a new score must be generated if this information is to be used in further

1357 analysis. Under current settings, jets that are tagged are assigned a random b -tag score that is sampled

1358 from the MV2C10 cumulative distribution above the 70% efficiency working point cut. All other

1359 jets in the event are assigned a random b -tag score below the 70% working point cut. Since these dis-

1360 tributions are discrete, the scores are not truly continuous (cf. example distributions in Figure 5.6),

1361 hence the “pseudo-continuous” nomenclature.

1362 A number of closure tests were performed on both the nominal and several systematics cases. In

1363 the plots that follow, truth (solid) and direct (dashed) tagging distributions for m_{bb} and $\Delta R(b_1, b_2)$

1364 in different p_T^V regimes for 2 lepton, 2 jet events. Agreement between the truth and direct tagging

1365 cases is generally very good, an example of which can be seen in Figure 5.7 for a signal qqZllH125

1366 sample, and the overall benefit of truth-tagging can be somewhat dramatically seen in the corre-

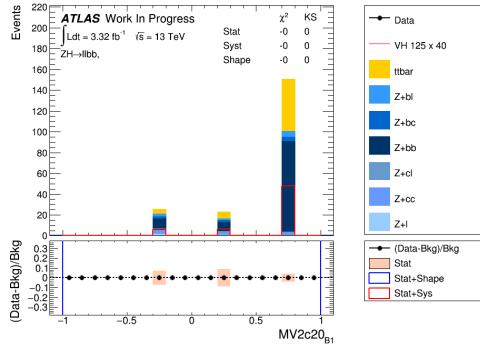


Figure 5.6: An example of a pseudo-continuous b -tagging distribution

1367 sponding plots $Z+\ell$ samples in Figure 5.8. At high p_T^V ($p_T^V > 200 \text{ GeV}$), however, in events with two
 1368 real b -jets, there is a much greater likelihood that the b -jets will merge into a single jet, which render
 1369 the naïve assumption that jets remain discrete invalid. While this does not appear to be a problem in
 1370 most samples (cf. $t\bar{t}$ in Figure 5.9), there is a mismodelling effect at low m_{bb} and low $\Delta R(b_1, b_2)$ at
 1371 $p_T^V > 200 \text{ GeV}$ for $W/Z + bb$ samples where truth-tagging overestimates the number of events in
 1372 this merged regime, as can be seen in Figure 5.10

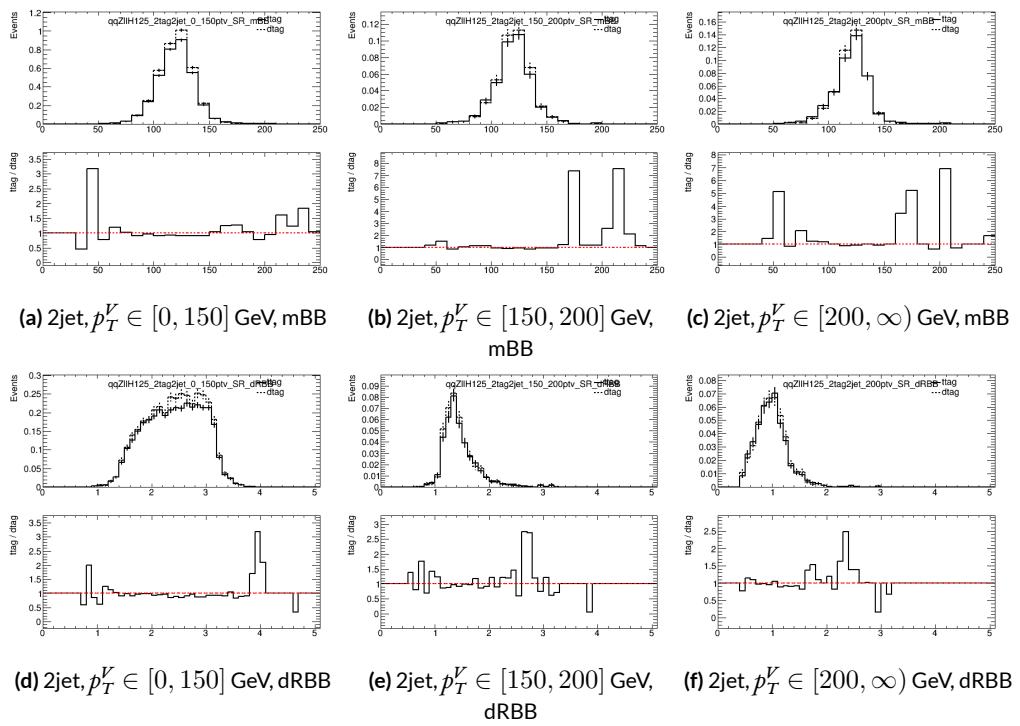


Figure 5.7: Truth-tagging closure tests for 2 lepton, 2 jet qqZIIH125 samples in three different p_T^V regions.

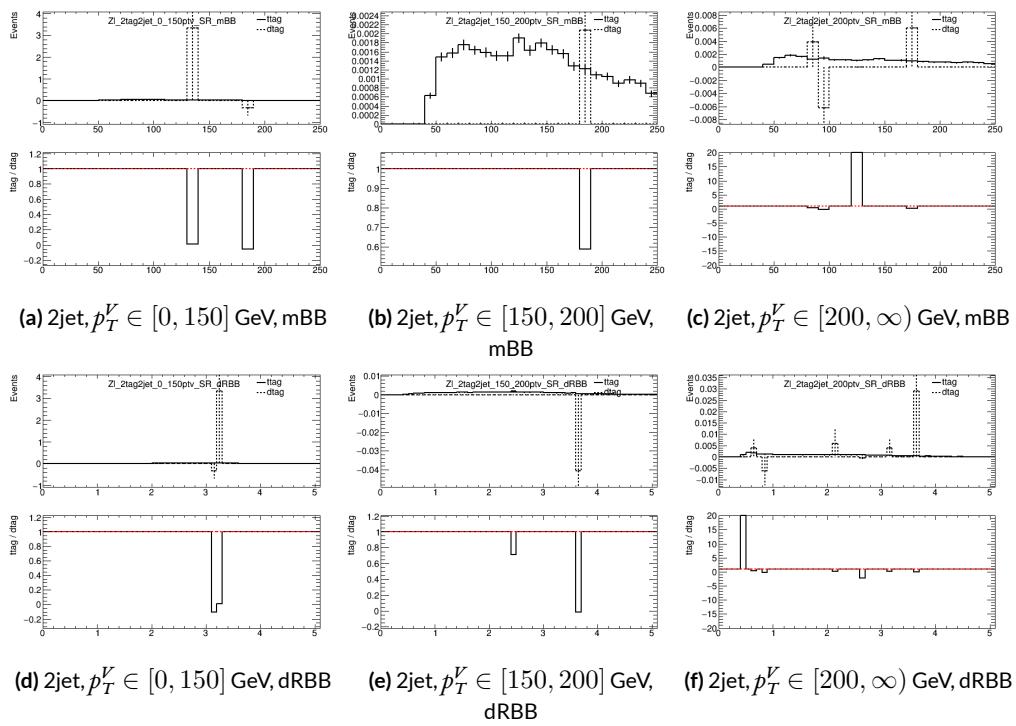


Figure 5.8: Truth-tagging closure tests for 2 lepton, 2 jet $Z + \ell$ samples in three different p_T^V regions.

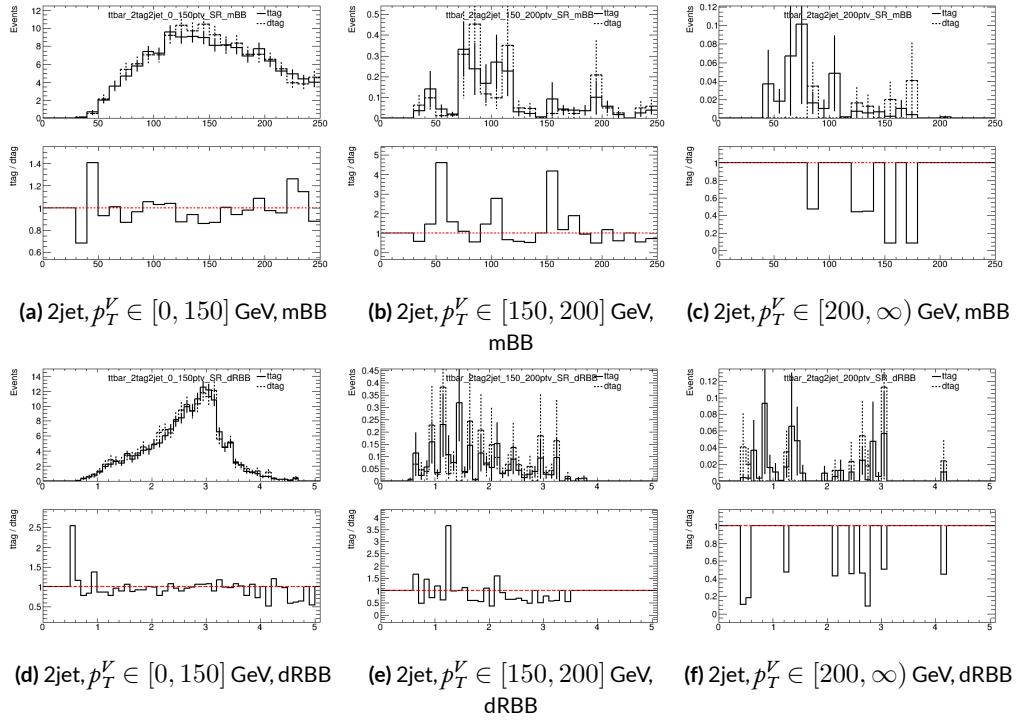


Figure 5.9: Truth-tagging closure tests for 2 lepton, 2 jet $t\bar{t}$ samples in three different p_T^V regions.

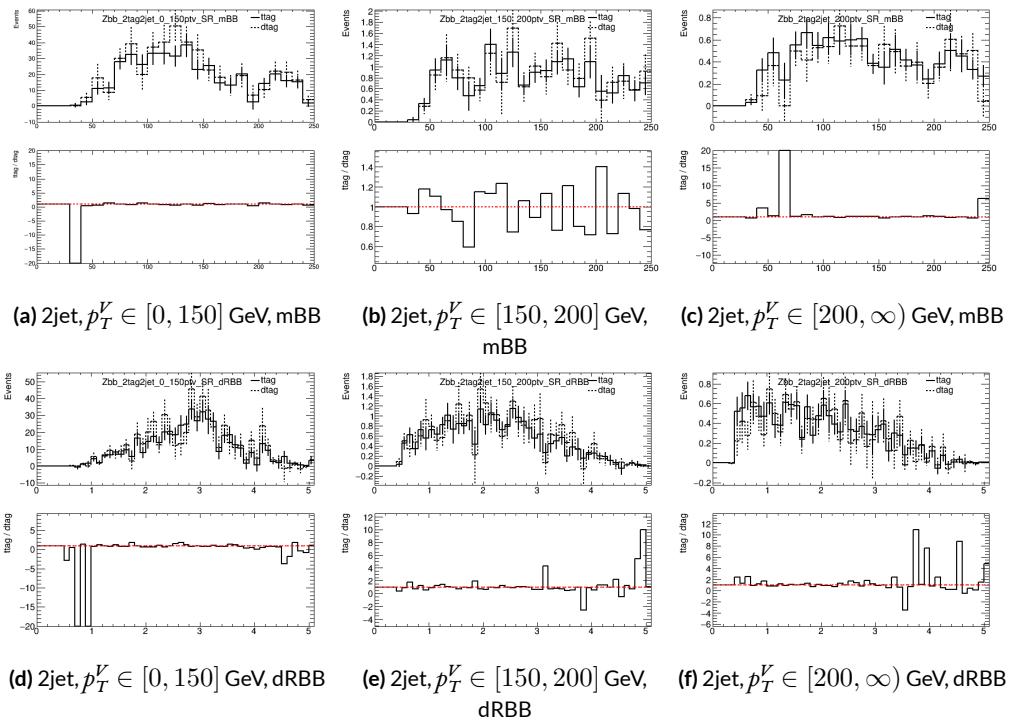


Figure 5.10: Truth-tagging closure tests for 2 lepton, 2 jet $Z + bb$ samples in three different p_T^V regions.

1373 5.7 MISCELLANIA AND SYSTEMATICS SUMMARY

1374 A summary of all experimental systematics, taken from ^{6s}, may be found below. In addition to the
1375 systematics discussed above, there are also two further systematics, on the total integrated luminosity
1376 and on the event reweighting factor used to account for pileup, both included in Table 5.7.

1377 5.8 EVENT SELECTION AND ANALYSIS REGIONS

1378 With object and event reconstruction described, it is now time to address which events are actually
1379 selected for use in analysis. This analysis focuses specifically on the 2-lepton channel of the fiducial
1380 analysis, with the event selection and analysis region definitions being identical. Common to all
1381 lepton channels in the fiducial analysis is the set of requirements on the jets in a given event. There
1382 must be at least two central jets and exactly two signal jets that have been “*b*-tagged” according to the
1383 MV2c10 algorithm ¹⁵, with at least one of these *b*-jets having $p_T > 45$ GeV. For MVA training and
1384 certain background samples, a process known as “truth-tagging” is applied instead of the standard *b*-
1385 tagging to boost sample statistics and stabilize training/fits (cf. ^{6s} Section 4.2 for details). After event
1386 selection, the *muon-in-jet* and *PtReco* corrections, described in ^{3t} 6.3.3-4, are applied to the *b*-jets.

1387 In addition to the common selections, there are 2-lepton specific selections. All events are re-
1388 quired to pass an un-prescaled single lepton trigger, a full list of which may be found in Tables 5
1389 and 6 of ^{3t} with the requirement that one of the two selected leptons in the event must have fired
1390 the trigger. There must be 2 VH-loose leptons, and at least one of these must be a ZH-signal lepton
1391 (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. ³¹
Pileup Reweighting	uncertainty on pileup reweighting	Section 11.1 in Ref. ³¹
	Electrons	
EL_EFF_Trigger_Total_iNPFCOR_PLUS_UNCOR	trigger efficiency uncertainty	Section 11.2.2. in Ref. ³¹
EL_EFF_Reco_Total_iNPFCOR_PLUS_UNCOR	reconstruction efficiency uncertainty	Section 11.3.1. in Ref. ³¹
EL_EFF_ID_Total_iNPFCOR_PLUS_UNCOR	ID efficiency uncertainty	Section 11.3.1. in Ref. ³¹
EL_EFF_Iso_Total_iNPFCOR_PLUS_UNCOR	isolation efficiency uncertainty	Section 11.3.1. in Ref. ³¹
EG_SCALE_ALL	energy scale uncertainty	Section 11.3.2. in Ref. ³¹
EG_RESOLUTION_ALL	energy resolution uncertainty	Section 11.3.2. in Ref. ³¹
	Muons	
MUON_EFF_TrigStatUncertainty	trigger efficiency uncertainty	Section 11.2.2. in Ref. ³¹
MUON_EFF_TrigSystUncertainty		
MUON_EFF_STAT	reconstruction and ID efficiency uncertainty for muons with > 15 GeV	Section 11.4.1. in Ref. ³¹
MUON_EFF_SYS	reconstruction and ID efficiency uncertainty for muons with < 15 GeV	Section 11.4.1. in Ref. ³¹
MUON_EFF_STAT_LOWPT		
MUON_EFF_SYST_LOWPT		
MUON_ISO_STAT	isolation efficiency uncertainty	Section 11.4.1. in Ref. ³¹
MUON_ISO_SYS		
MUON_TTVA_STAT	track-to-vertex association efficiency uncertainty	Section 11.4.1. in Ref. ³¹
MUON_TTVA_SYS		
MUON_ID	momentum resolution uncertainty from inner detector	Section 11.4.2. in Ref. ³¹
MUON_MS	momentum resolution uncertainty from muon system	Section 11.4.2. in Ref. ³¹
MUON_SCALE	momentum scale uncertainty	Section 11.4.2. in Ref. ³¹
MUON_SAGITTA_RHO		
MUON_SAGITTA_RESBIAS	charge dependent momentum scale uncertainty	Section 11.4.2 in Ref. ³¹
Jets		
JET_2iNP_JET_EffectiveNP_1	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EffectiveNP_2	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EffectiveNP_3	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EffectiveNP_4	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EffectiveNP_5	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EffectiveNP_6	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EffectiveNP_7	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EffectiveNP_8restTerm	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EtaIntercalibration_Modeling	energy scale uncertainty on eta-intercalibration (modeling)	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EtaIntercalibration_TotalStat	energy scale uncertainty on eta-intercalibrations (statistics/method)	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_EtaIntercalibration_NonClosure	energy scale uncertainty on eta-intercalibrations (non-closure)	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Pileup_OffsetMu	energy scale uncertainty on pile-up (mu dependent)	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Pileup_OffsetNPV	energy scale uncertainty on pile-up (NPV dependent)	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Pileup_PtTerm	energy scale uncertainty on pile-up (pt term)	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Pileup_RhoTopology	energy scale uncertainty on pile-up (density ρ)	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Flavor_Composition_Zjets	energy scale uncertainty on $Z+jets$ sample's flavour composition	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Flavor_Composition_Wjets	energy scale uncertainty on $W+jets$ sample's flavour composition	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Flavor_Composition_top	energy scale uncertainty on top sample's flavour composition	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Flavor_Composition	energy scale uncertainty on VV and VH sample's flavour composition	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_Flavor_Response	energy scale uncertainty on samples' flavour response	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_BJES_Response	energy scale uncertainty on b-jets	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_PunchThrough_MC15	energy scale uncertainty for punch-through jets	Section 11.5.1. in Ref. ³¹
JET_2iNP_JET_SingleParticle_HighPt	energy scale uncertainty from the behaviour of high-pT jets	Section 11.5.1. in Ref. ³¹
JET_JER_SINGLE_NP	energy resolution uncertainty	Section 11.5.1. in Ref. ³¹
JET_JvtEfficiency	JVT efficiency uncertainty	Section 11.5.1 in Ref. ³¹
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. ³¹
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. ³¹
FT_EFF_Eigen_extrapolation_from_charm	b -tagging efficiency uncertainty on tau jets	Section 11.7. in Ref. ³¹
MET		
METTrigStat	trigger efficiency uncertainty	Section 11.2.1. in Ref. ³¹
METTrigTop/Z		
MET_SoftTrk_ResoPara	track-based soft term related longitudinal resolution uncertainty	Section 11.6. in Ref. ³¹
MET_SoftTrk_ResoPerp	track-based soft term related transverse resolution uncertainty	Section 11.6. in Ref. ³¹
MET_SoftTrk_Scale	track-based soft term related longitudinal scale uncertainty	Section 11.6. in Ref. ³¹
MET_JetTrk_Scale	track MET scale uncertainty due to tracks in jets	Section 11.6. in Ref. ³¹

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. ³¹.

¹³⁹² 101 GeV. In addition to the jet corrections described above, a kinematic fitter is applied to the leptons
¹³⁹³ and two leading corrected jets in an event with three or fewer jets[†] to take advantage of the fact that
¹³⁹⁴ the 2-lepton final state is closed (cf. ²⁰); these objects are only used for MVA training/fit inputs.

¹³⁹⁵ In order to increase analysis sensitivity, the analysis is split into orthogonal regions based on the
¹³⁹⁶ number of jets and the transverse momentum of the Z candidate (the vectoral sum of the lepton
¹³⁹⁷ 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
¹³⁹⁸ regions where the leptons are required to be the same flavor (e or μ), there are top $e - \mu$ control
¹³⁹⁹ regions used to constrain the top backgrounds.

¹⁴⁰⁰ All of these requirements are summarized in 5.8.

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

¹⁴⁰¹ It should be noted that the use of ≥ 3 jet events is a 2-lepton specific selection. These regions are
¹⁴⁰² exclusive 3 jet regions in the 0- and 1-lepton channels, but the fiducial 2-lepton analysis was found to
¹⁴⁰³ see a $\sim 4\%$ gain in sensitivity in studies by including ≥ 4 jet events⁶⁹.

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

猛き者も遂には滅びぬ、
偏に風の前の塵に同じ。

Heike monogatari

6

1404

1405

Multivariate Analysis Configuration

1406 IN ORDER TO fully leverage the descriptive power of the 13 TeV dataset, this analysis makes use of a
1407 multivariate (MVA) discriminant. Where traditionally event counts or single discriminating vari-
1408 ables per region of phase space have been fed to fits, MVA discriminants seek to integrate additional
1409 information not captured in the conventional phase space cuts plus dijet invariant mass distribu-

1410 tions. Formulating the MVA discriminant is an exercise in supervised learning to construct a binary
1411 classifier, where one uses labeled “signal” and “background” MC events to optimize the parameters
1412 of a statistical model—in this case a boosted decision tree (BDT) with some set of physically moti-
1413 vated variables (or “factors”). The interested reader is directed to the standard references on machine
1414 learning for further details. Sample and variable selection, including variables derived using the the
1415 RestFrames and Lorentz Invariants concepts introduced in Sections 2.5–2.7, are discussed in Section
1416 6.1; MVA training is treated in Section 6.2; and the data statistics only (no systematics) performance
1417 of the three MVA discriminants is explored in Section 6.3.

1418 6.1 TRAINING SAMPLES AND VARIABLE SELECTION

1419 A subset of samples described in Chapter 3 is used for multivariate analysis training, with $qqZH \rightarrow$
1420 $\ell\ell b\bar{b}$ and $ggZH \rightarrow \ell\ell b\bar{b}$ used as signal samples and $Z+jets$, $t\bar{t}$, and VV used as background samples.
1421 Truth-tagging (Section 5.6.2) is used on all samples in MVA training to improve training statistics
1422 and stability. All figures quoted in this section scale distributions to a luminosity of 36.1 fb^{-1} .

1423 6.1.1 STANDARD VARIABLES

1424 The standard set of variables taken as a baseline is the same as used in the fiducial analysis. The vari-
1425 ables fall into several main categories: energy/momenta scales of composite objects (m_{bb} , m_{bbj} ,
1426 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
1427 ($p_T^{b_1}$, $p_T^{b_2}$, $p_T^{j_3}$), and E_T^{miss} . Input distributions for these variables in all the 2 (≥ 3 jet) analysis signal
1428 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

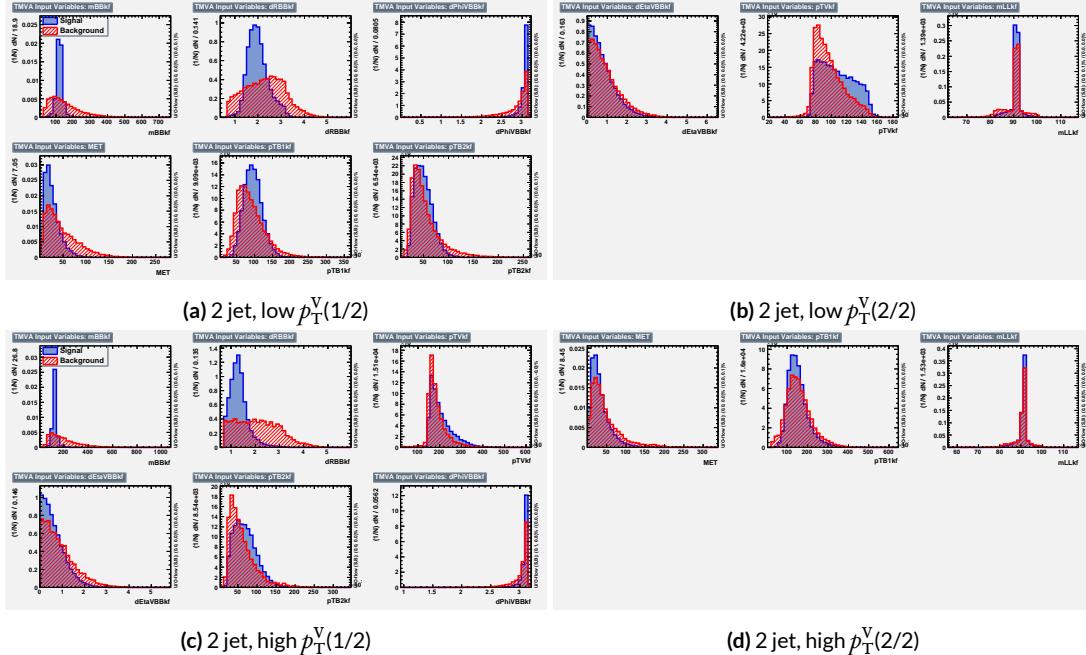


Figure 6.1: Input variables in 2 jet signal regions for the standard variable set. Signal distributions are in red, and background distributions are in blue.

1429

1430 are used as inputs for one of the two k-folded final discriminants, and the order of the distributions
 1431 is the hyperparameter optimized order for feeding into the BDT; what precisely this means will be
 1432 discussed in following sections. While variables in the analysis regions are generally similar, there are
 1433 some notable exceptions. p_T^V and the correlated $\Delta R(b_1, b_2)$ have different shapes, by construction
 1434 for the former and by correlation for the latter, at low and high p_T^V . * The ≥ 3 jet regions also have
 1435 variables that are not applicable to the 2 jet regions; the inclusion of m_{BBJ} (the invariant mass of the
 1436 two b -jets and leading untagged jet) in particular is of note and suggests a potential avenue forward

*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.

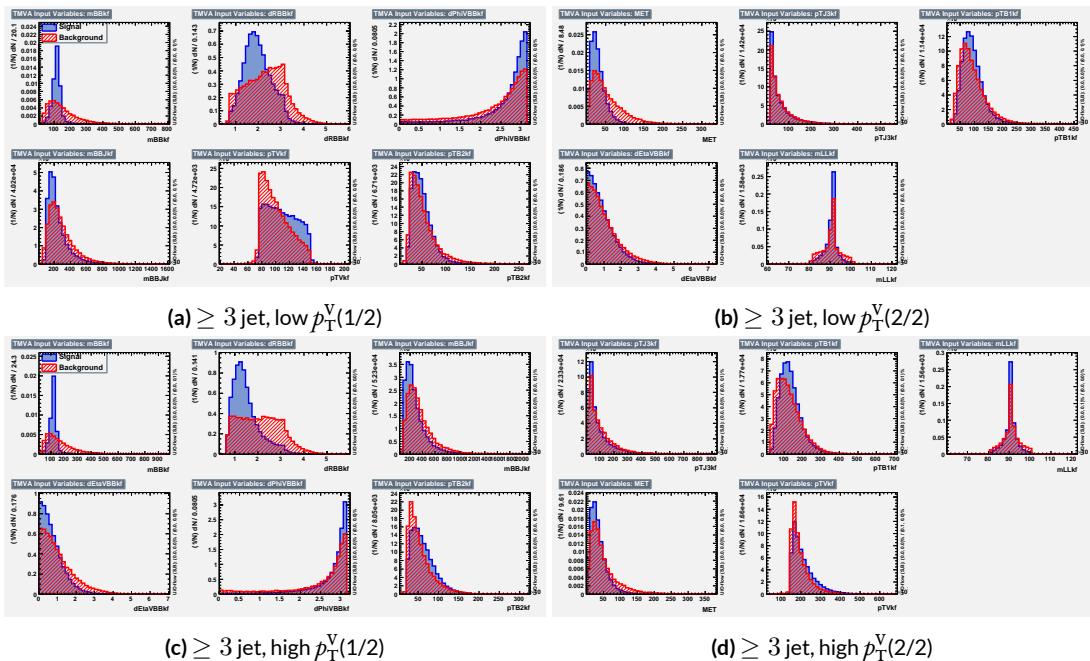


Figure 6.2: Input variables in ≥ 3 jet signal regions for the standard variable set. Signal distributions are in red, and background distributions are in blue.

1437 for refinements of the non-standard variables.

1438 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

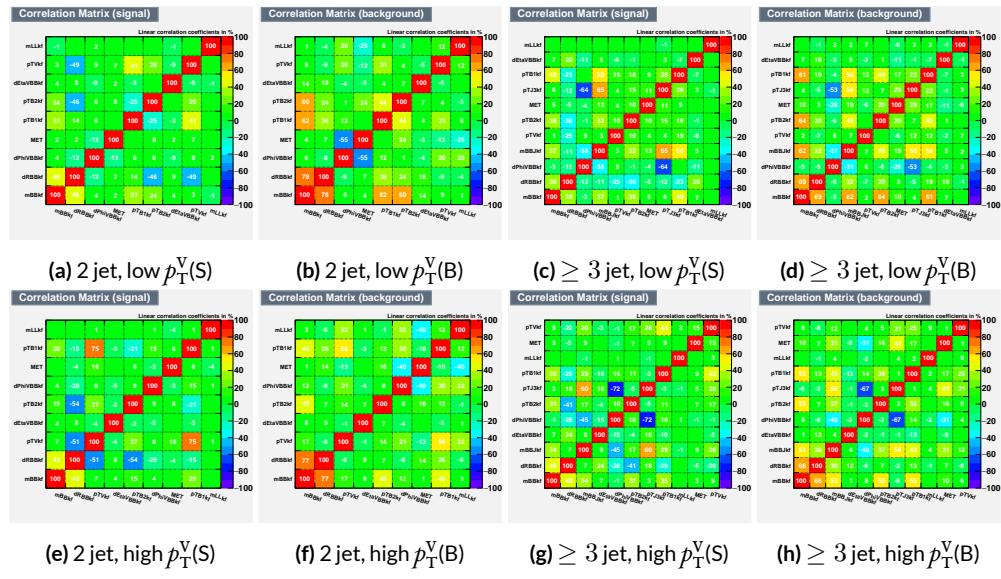


Figure 6.3: Signal and background variable correlations for the standard variable set.

1439

1440 6.1.2 LORENTZ INVARIANTS

1441 In choosing the set of variables used for a set of Lorentz Invariants based discriminants, we decided
 1442 to use S. Hagebeck's set from⁵³ 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

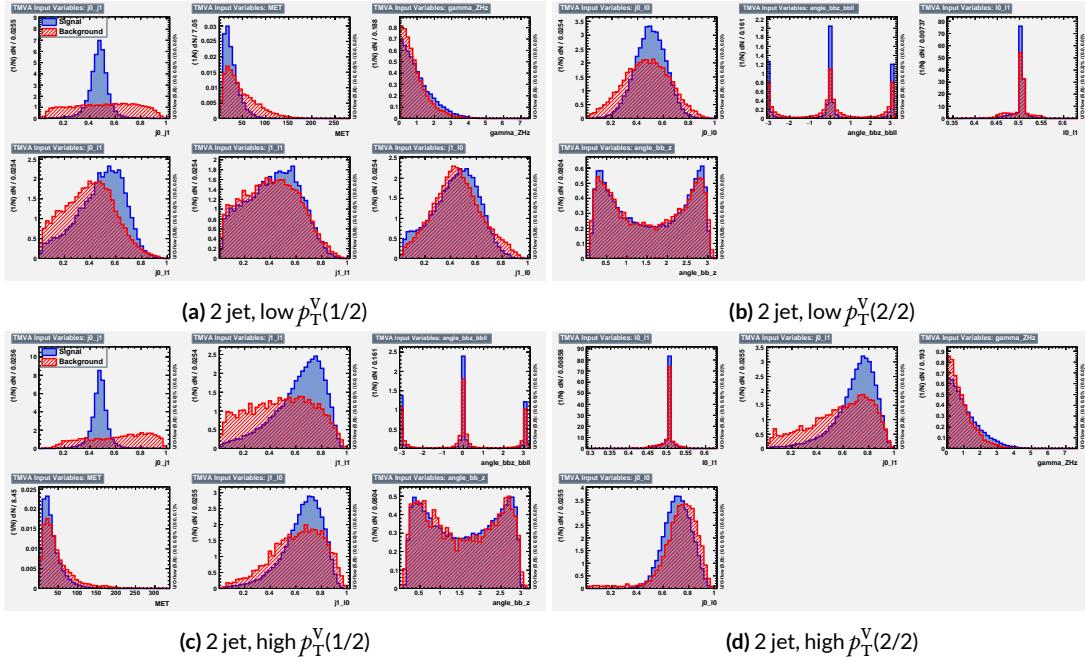


Figure 6.4: Input variables in 2 jet signal regions for the LI variable set. Signal distributions are in red, and background distributions are in blue.

1443

1444 about the variable set chosen here is that \vec{E}_T^{miss} has been added to the standard LI set. Since the LI
 1445 construction assumes that this quantity is zero, there is no obvious way to include it. Nevertheless,
 1446 as the correlation matrices for the LI variables show in Figure 6.6, there is actually very little corre-
 1447 lation between \vec{E}_T^{miss} and the other variables (with this being slightly less the case for the background
 1448 correlations, as to be expected since $t\bar{t}$, a principal background, is \vec{E}_T^{miss} -rich). Hence, if including

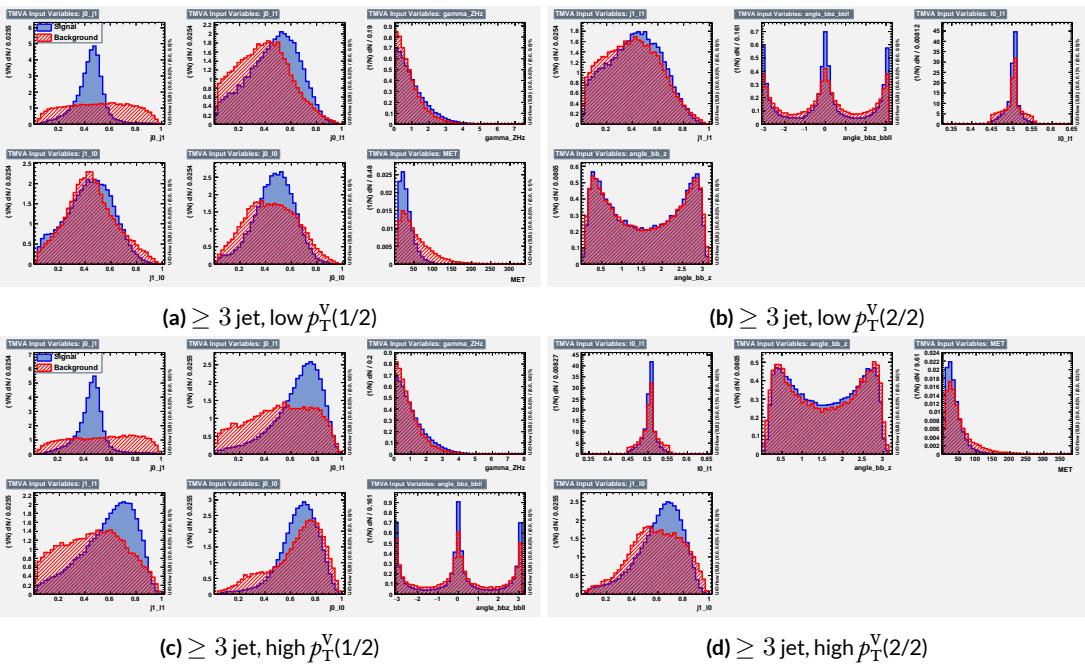


Figure 6.5: Input variables in ≥ 3 jet signal regions for the LI variable set. Signal distributions are in red, and background distributions are in blue.

1449 \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.

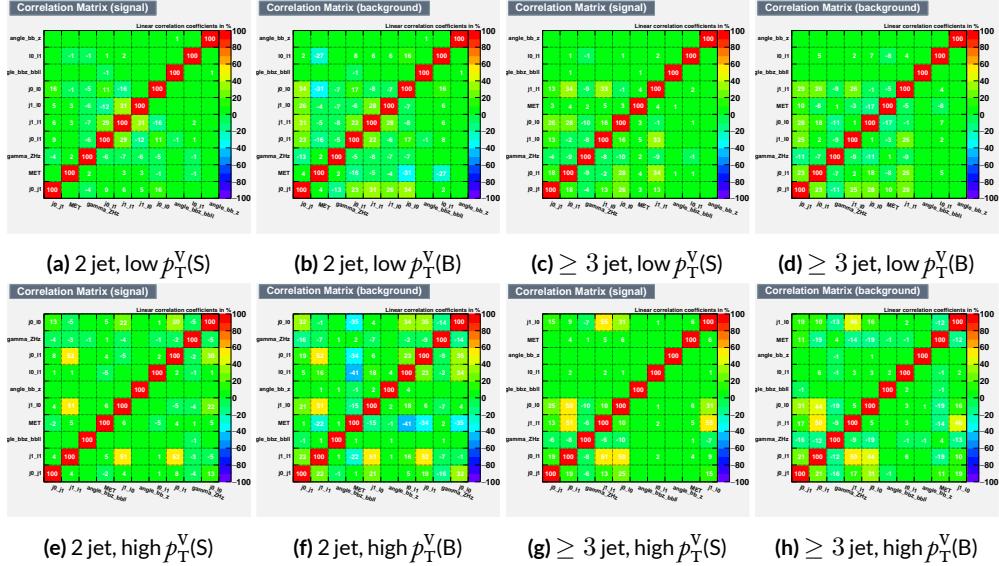


Figure 6.6: Signal and background variable correlations for the LI variable set.

1450

1451 6.1.3 RESTFRAMES VARIABLES

1452 There is no precedent for using the RestFrames variables in the $ZH \rightarrow \ell\ell b\bar{b}$ analysis, so a subset
1453 of possible RF variables had to be selected as the basis of a discriminant. The masses and cosines of
1454 boost angles from parent frames for the CM, Z , and H frames gives six variables, and it was decided
1455 that it would be good to match the LI in terms of variable number and treatment (i.e. no special
1456 treatment of the third jet), which leaves four more variables. In addition to the cosines, there are
1457 also the $\Delta\phi$ angles. Furthermore, there are the event-by-event scaled momentum ratios, both lon-
1458 gitudinal and transverse. There is also both a $\Delta\phi$ and an CM-scaled ratio for the \vec{E}_T^{miss} . All of these
1459 variables were included in a ranking using slightly different training settings as the main hyperpa-
1460 rameter optimization variable ranking described below. The goal of this study was not to develop a
1461 discriminant, as the number of variables is too high, but rather to see which ones are generally use-
1462 ful. Table 6.1 shows the results of this study. Percent gains (losses) at each step by adding the variable
1463 with biggest gain (smallest loss) are shown in green (red). The final row shows an aggregate rank-
1464 ing, calculated simply by adding up a variables ranks in all bins and ordering the variables smallest
1465 to greatest. This simple aggregation does not take into account which regions are potentially more
1466 sensitive and so where taken simply to give an idea of how variables generally performed. With this
1467 in mind, the RF variables were chosen to be the masses M_{CM} , M_H , and M_Z , the angles $\cos CM$, \cosh ,
1468 $\cos Z$, $\cos \phi CMH$, and the ratios R_{pt} , R_{pz} , and R_{met} . Their distributions may be seen in Figures 6.7
1469 and 6.8.

1470 Correlations for the chosen RF variables are shown in Figure 6.9. These correlations are much

Region	Variable Chain
2jet pTVbin1	Rpt (65.8%), 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 (50.8%), Rpz (15.6%), MZ (14.8%), cosZ (3.08%), 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 (31.3%), Rpz (21.6%), cosH (8.97%), cosZ (1.42%), cosCM (11.3%), 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 (i,i,i,i), 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.

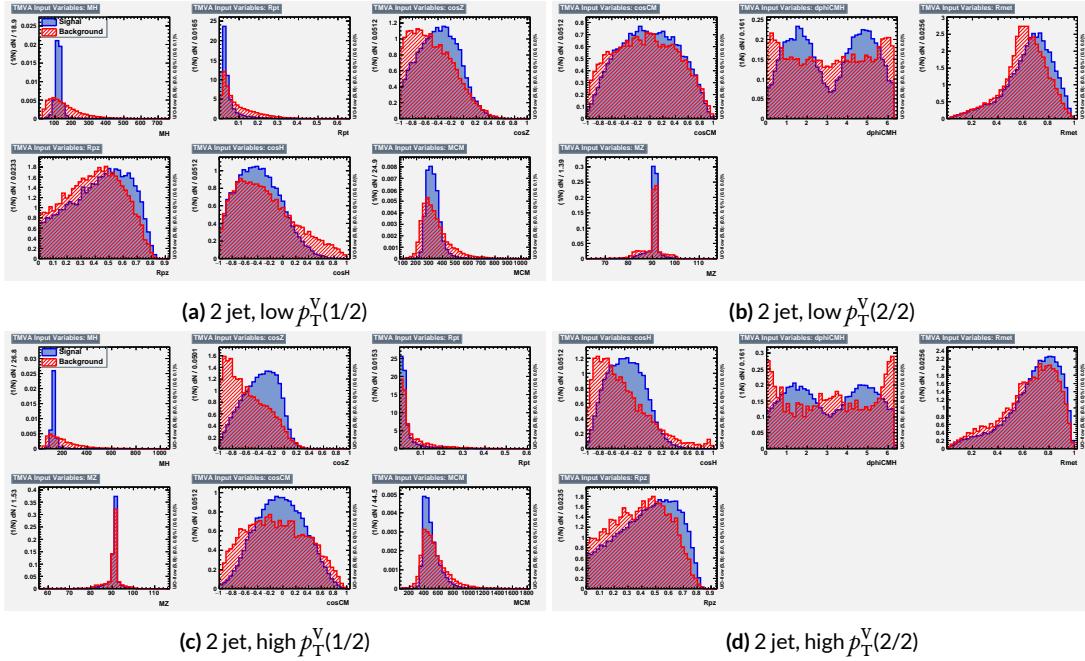


Figure 6.7: Input variables in 2 jet signal regions for the RF variable set. Signal distributions are in red, and background distributions are in blue.

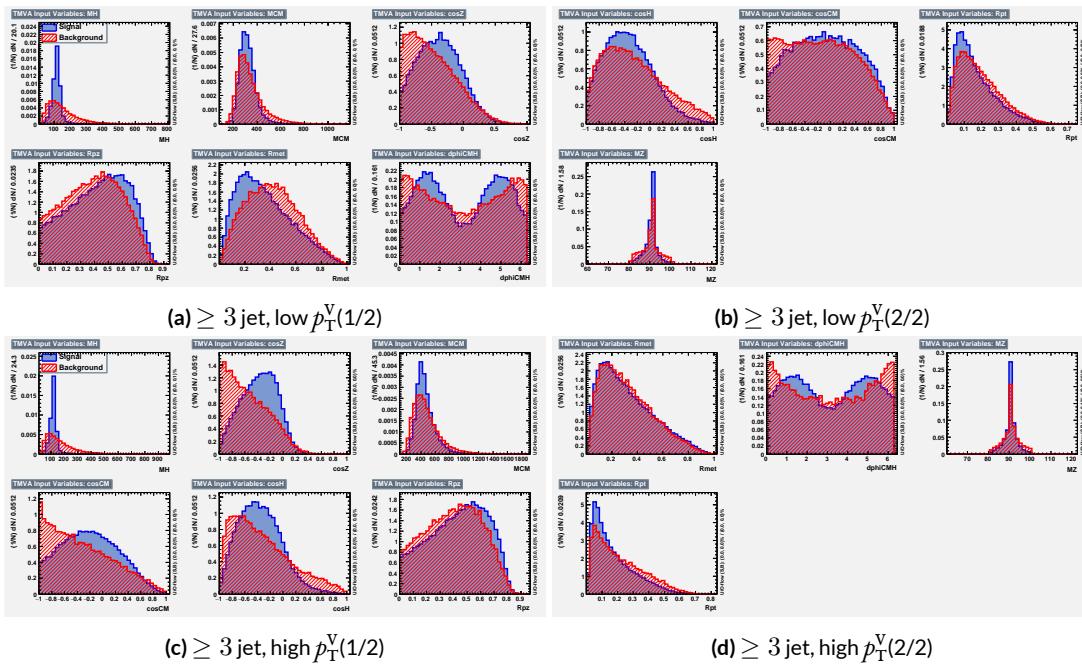


Figure 6.8: Input variables in ≥ 3 jet signal regions for the RF variable set. Signal distributions are in red, and background distributions are in blue.

1471 lower than for the standard case but still slightly higher than for the LI case. Notably, many strong
 1472 correlations that exist for signal events do not exist in background events and vice versa, so what is
 1473 lost in orthogonality may very well be recuperated in greater separation[†]. Given the generally better
 1474 performance of the RF sets, as we shall see in following sections and chapters, this slight tradeoff is
 1475 likely an aesthetic one, with the main benefits of a more orthogonal basis likely realized at this level
 1476 of correlation.

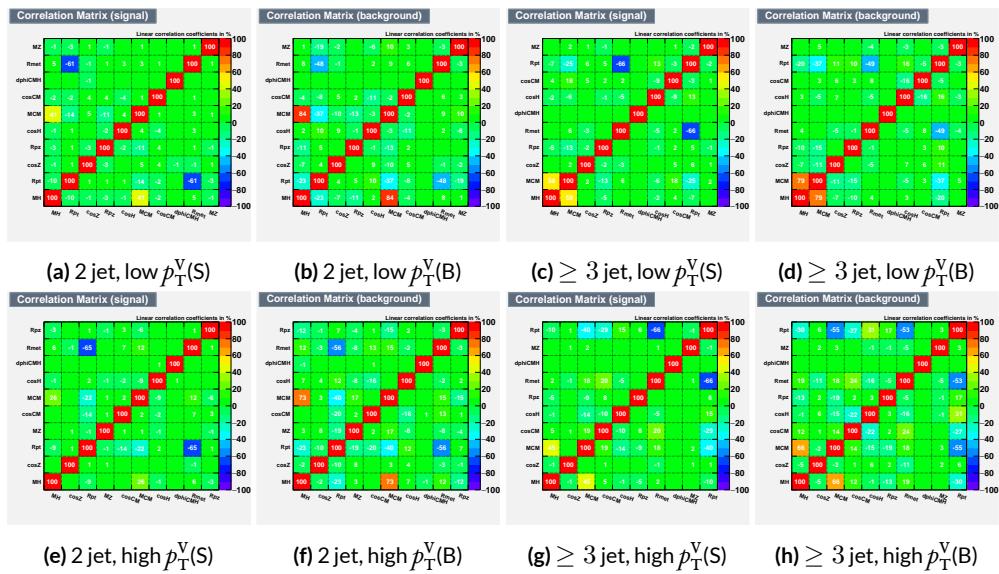


Figure 6.9: Signal and background variable correlations for the RF variable set.

1477 A summary of the variables used in the three cases is given in 6.2.

[†]It is very hard to say for certain whether this is the case for MVA discriminants, and such dedicated studies might make worthwhile future studies.

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.

1478 6.2 MVA TRAINING

1479 With variables chosen, the MVA discriminants must be trained and optimized. MVA training and
 1480 hyperparameter optimization (in this case, just the order in which variables are fed into the MVA) is
 1481 conducted using the “holdout” method. In this scheme, events are divided into three equal portions
 1482 (in this case using `EventNumber%3`), with the first third (the “training” set) being used for the initial
 1483 training, the second third (the “validation” set) being used for hyperparameter optimization, and
 1484 the final third (the “testing” set) used to evaluate the performance of the final discriminants in each
 1485 analysis region.

1486 The MVA discriminant used is a boosted decision tree (BDT). Training is done in TMVA using
 1487 the training settings of the fiducial analysis⁶[‡]. For the purposes of hyperparameterization and test-
 1488 ing, transformation D with $z_s = z_b = 10$ is applied to the BDT distributions, and the cumulative
 1489 sum of the significance $S/\sqrt{S + B}$ in each bin is calculated for each pair of distributions.

1490 Transformation D is a histogram transformation, developed during the Run 1 SM $VH(b\bar{b})$ search,

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

¹⁴⁹¹ designed to reduce the number of bins in final BDT distributions and thereby mitigate the effect of
¹⁴⁹² statistical fluctuations in data while also maintaining sensitivity. Such an arbitrary transformation
¹⁴⁹³ may be expressed as:

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

¹⁴⁹⁴ where

- ¹⁴⁹⁵ • $I[k, l]$ is an interval of the histograms, containing the bins between bin k and bin l ;
- ¹⁴⁹⁶ • N_s is the total number of signal events in the histogram;
- ¹⁴⁹⁷ • N_b is the total number of background events in the histogram;
- ¹⁴⁹⁸ • $n_s(I[k, l])$ is the total number of signal events in the interval $I[k, l]$;
- ¹⁴⁹⁹ • $n_b(I[k, l])$ is the total number of background events in the interval $I[k, l]$;
- ¹⁵⁰⁰ • z_s and z_b are parameters used to tune the algorithm.

¹⁵⁰¹ Transformation D uses:

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

¹⁵⁰² Rebinning occurs as follow:

- ¹⁵⁰³ 1. Begin with the highest valued bin in the original pair of distributions. Call this the “last” bin
¹⁵⁰⁴ and use it as l , and have k be this bin as well.
- ¹⁵⁰⁵ 2. Calculate $Z(I[k, l])$
- ¹⁵⁰⁶ 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
¹⁵⁰⁷ $k - 1$ the new “last” bin l .
- ¹⁵⁰⁸ 4. Continue until all bins have been iterated through; if $Z \leq 1$ for any remaining n of the
¹⁵⁰⁹ lowest-valued bins (as is often the case), simply rebin these as a single bin.

1510 Variable ranking is done iteratively (greedily) in each analysis region. In each set, the validation
 1511 significance of a BDT using an initial subset of variables is calculated ($dRBB$ and mBB for the stan-
 1512 dard set; $j_0_j_1$ for the LI set; and MH for the RF set). Each of the remaining unranked variables
 1513 are then added separately, one at a time, to the BDT. The variable yielding the highest validation
 1514 significance is then added to the set list of ranked variables and removed from the list of unranked
 1515 variables. This process is repeated until no variables remain. These rankings are shown in Figures
 1516 6.10–6.12. Rankings tend to be fairly stable.

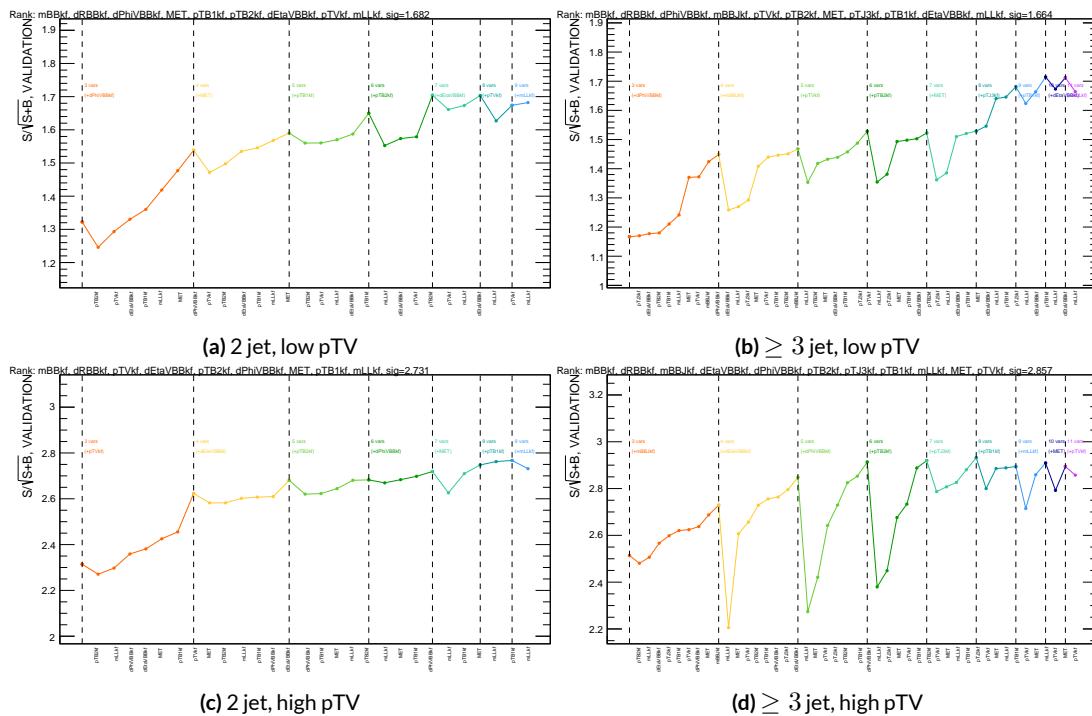


Figure 6.10: Rankings for the standard variable set.

1517 Once variables have been ranked, the BDT may be used both to evaluate performance in a simpli-
 1518 fied analysis scenario in the absence of systematic uncertainties (described below in Section 6.3) and

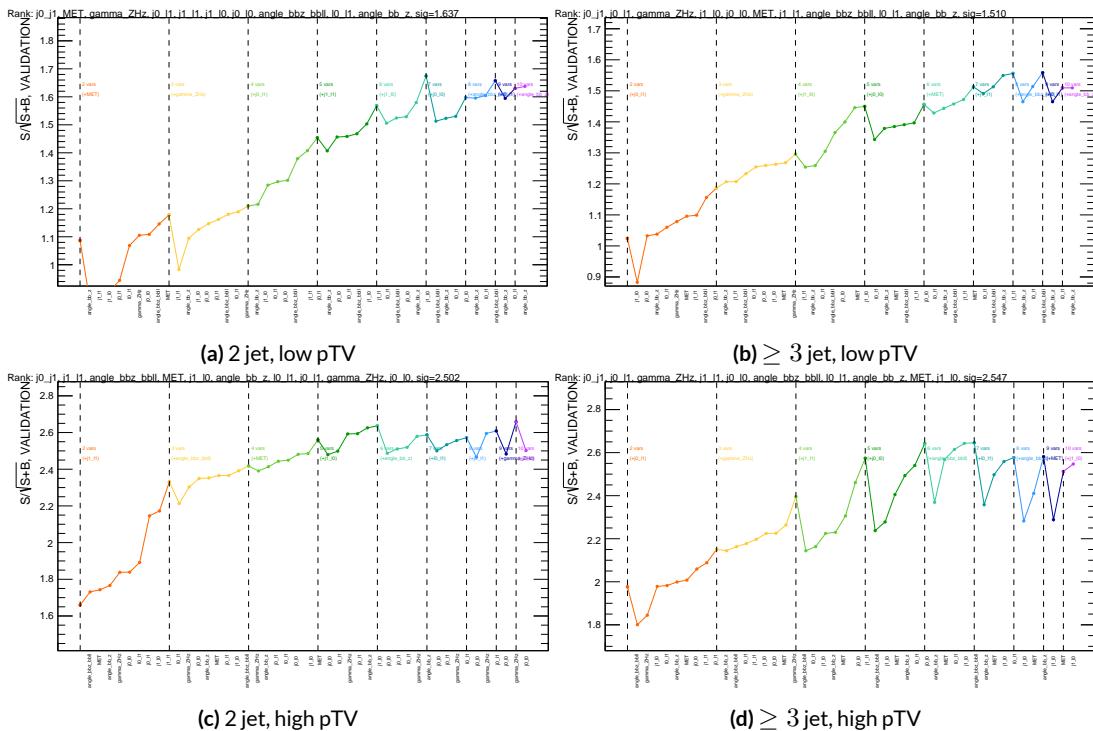


Figure 6.11: Rankings for the LI variable set.

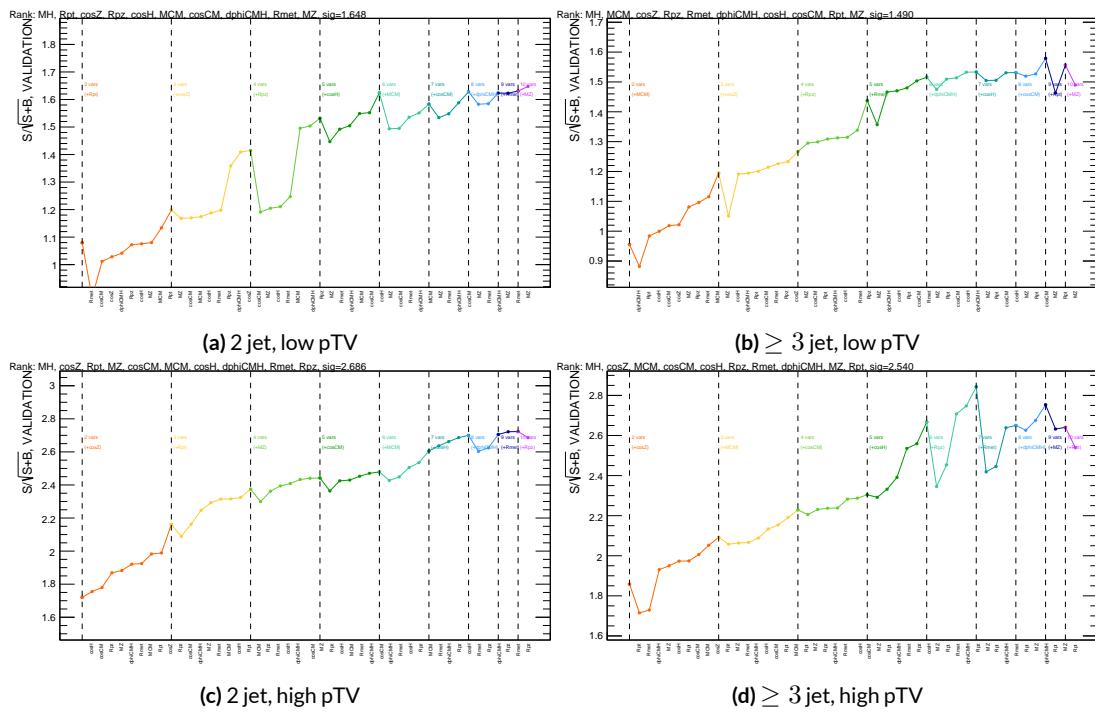


Figure 6.12: Rankings for the RF variable set.

1519 to create xml files for the production of fit inputs for an analysis including systematics. Following
1520 the approach taken in the fiducial analysis, BDT discriminants using two “k-folds” are produced to
1521 prevent overtraining, since the samples used for training are the same as those used to produce in-
1522 puts for the full profile likelihood fit. In this scheme, a BDT trained on events with an even (odd)
1523 `EventNumber` are used to evaluate events with an odd (even) `EventNumber`.

1524 6.3 STATISTICS ONLY BDT PERFORMANCE

1525 As described above, cumulative significances can be extracted from pairs of signal and background
1526 BDT output distributions in a given region. In order to evaluate performance of variable sets in the
1527 absence of systematic uncertainties, such pairs can be constructed by evaluating BDT score on the
1528 testing set of events using the optimal variable rankings in each region. We show two versions of
1529 each testing distribution for each variable set in each signal region in Figures 6.13–6.15. The training
1530 distribution is always shown as points. The plots with block histograms with numbers of bins that
1531 match (do not match) the training distribution do not (do) have transformation D applied. Trans-
1532 formation D histograms are included to show the distributions actually used for significance evalu-
1533 ation, while the untransformed histograms are included to illustrate that the level of overtraining is
1534 not too terrible[§]. For better comparison of the distributions, all histograms have been scaled to have
1535 the same normalization.

1536 As can be seen in the summary of cumulative significances for each of these analysis regions and
1537 variable sets in Figure 6.16, the performance of each of the variable sets is quite similar. The standard

[§]The raw distributions include a K-S test statistic for signal (background) distributions.

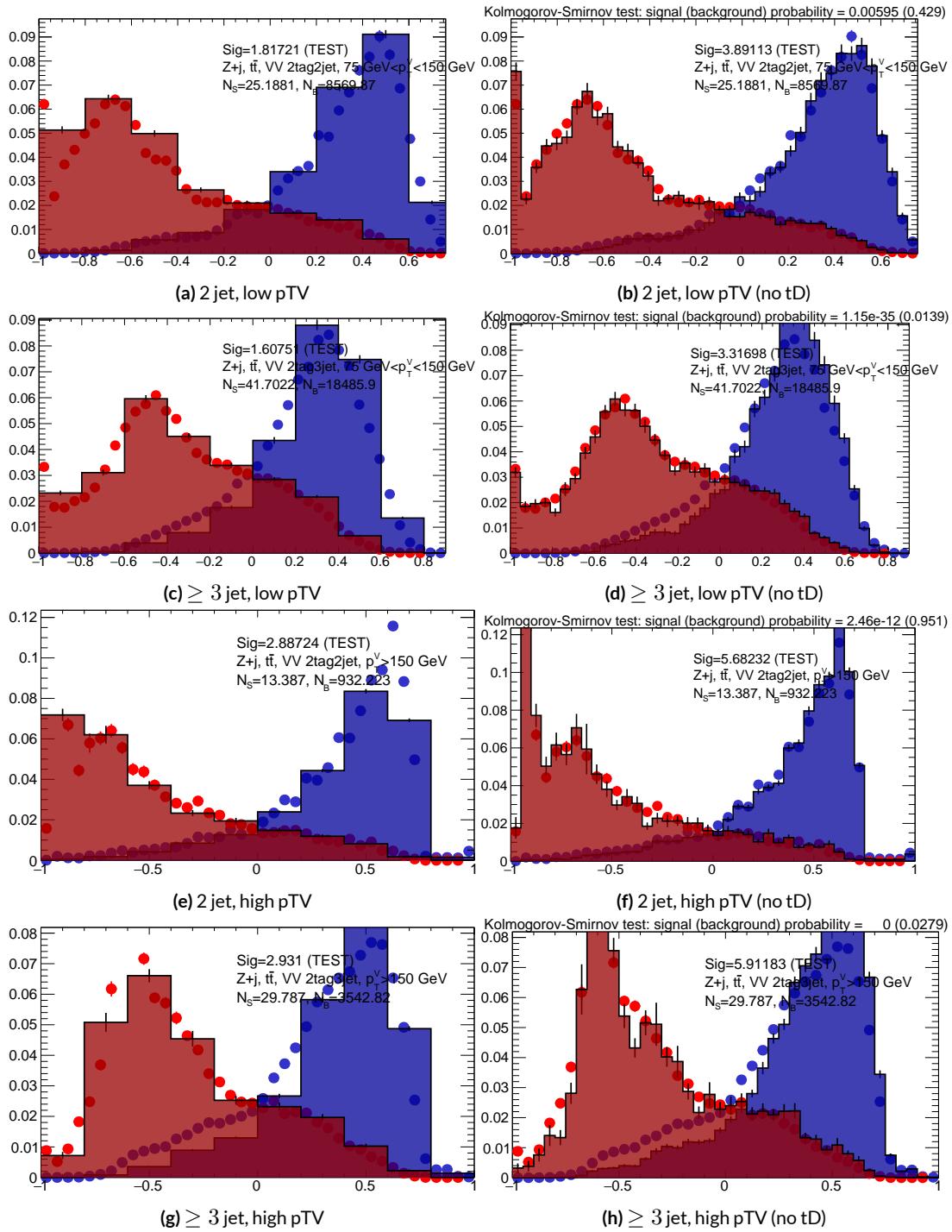


Figure 6.13: Training (points) and testing (block histogram) MVA distributions used for stat only testing for the standard variable set.

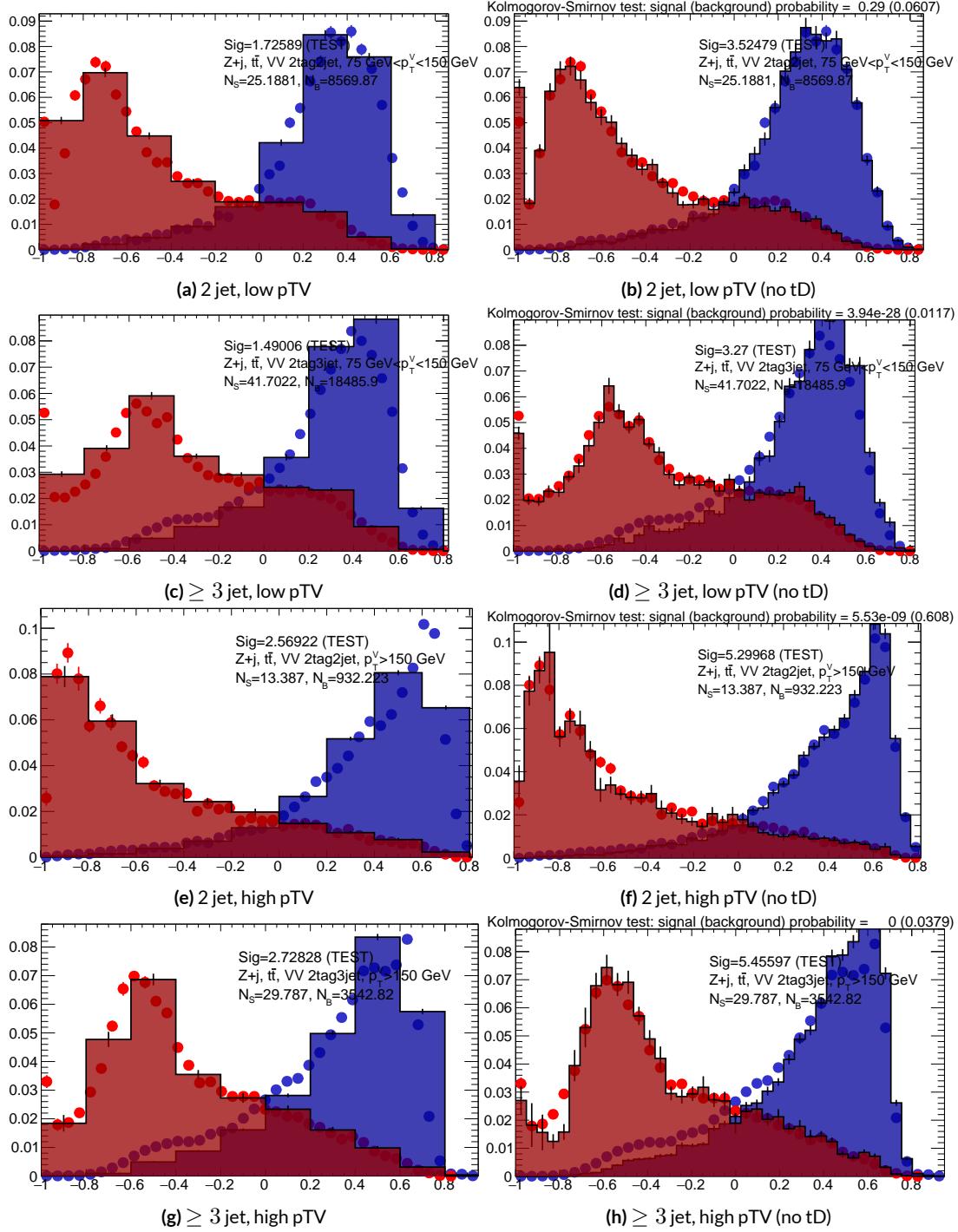


Figure 6.14: Training (points) and testing (block histogram) MVA distributions used for stat only testing for the LI variable set.

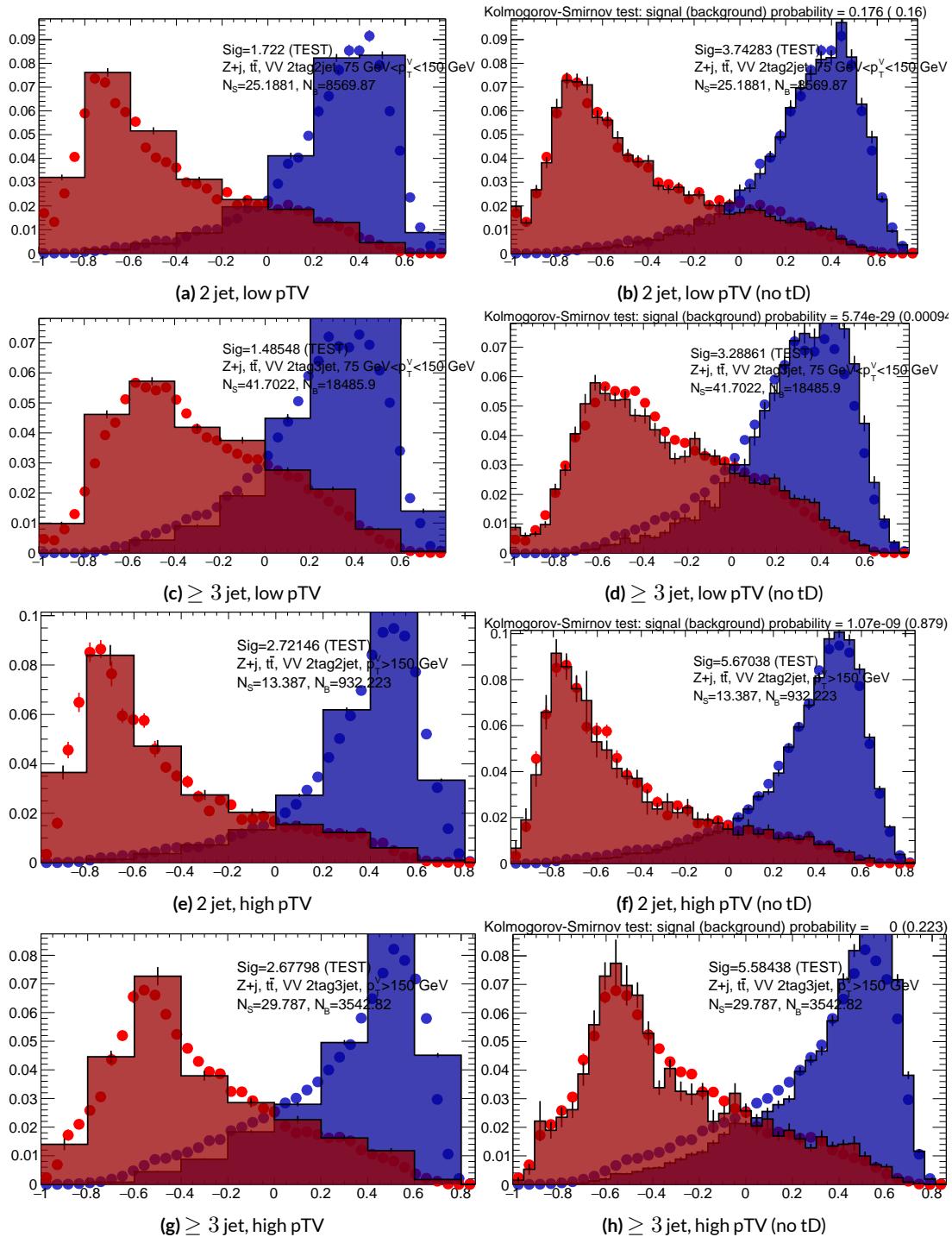


Figure 6.15: Training (points) and testing (block histogram) MVA distributions used for stat only testing for the RF variable set.

1538 set performs best, with the LI (RF) set having a cumulative significance that is 7.9% (6.9%) lower.
 1539 This suggests that the LI and RF variables, in the $ZH \rightarrow \ell\ell b\bar{b}$ closed final state, have no more in-
 1540 trinsic descriptive power than the standard set. That these figures are all relatively high (~ 4.5) is
 1541 due largely to the absence of systematics and possibly in part due to the fact that many of the most
 1542 significant bins occur at high values of the BDT output, which, as can be seen in any of the testing
 distributions, contain a small fraction of background events. An interesting feature to note in Fig-

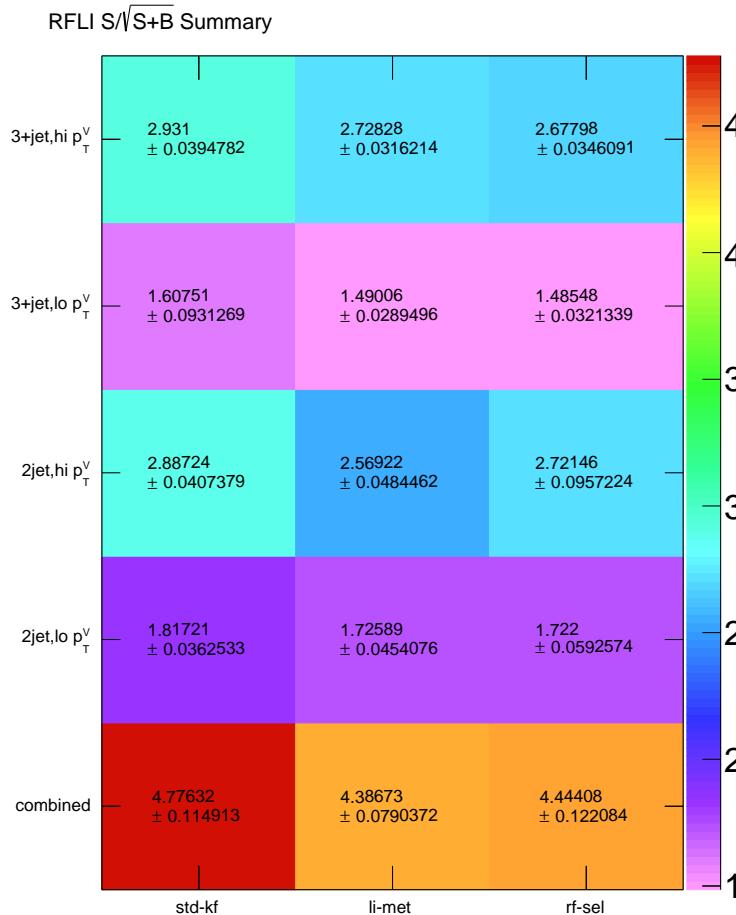


Figure 6.16: Results of testing significances sorted by analysis region and variable set.

1543

¹⁵⁴⁴ ure 6.16 is that while the standard set does perform better in all regions, the gap is larger in the ≥ 3
¹⁵⁴⁵ jet regions, suggesting that further optimization in the ≥ 3 jet case could be useful. Moreover, as
¹⁵⁴⁶ discussed at the end of Chapter 5, the choice of ≥ 3 jet and not exclusive 3 jet regions is a 2-lepton
¹⁵⁴⁷ specific choice and may not be justified for the non-standard variable sets.

Multivac picked you as most representative this year.

Not the smartest, or the strongest, or the luckiest, but

just the most representative. Now we don't question

Multivac, do we?

Isaac Asimov, "Franchise"

7

1548

1549

Statistical Fit Model and Validation

1550 THE ULTIMATE GOAL of an analysis like the search for SM $VH(b\bar{b})$ decay is to say with as much
1551 justified precision as possible with the ATLAS collision data whether or not the SM-like Higgs ob-
1552 served in other decay modes also decays to b -quarks and, if so, whether this rate is consistent with
1553 the SM prediction. In the limit of perfect modeling of both background processes and detector/reconstruction,

1554 the only free parameter is this production rate, referred to typically as a “signal strength,” denoted μ ,
1555 with $\mu = 1$ corresponding to the SM prediction and $\mu = 0$ corresponding to the SM with no
1556 Higgs.

1557 To get a better sense of what this might look like, take a look at the example discriminant distri-
1558 bution in Figure 7.1. The black points are data (with statistical error bars), and the colored block
1559 histograms have size corresponding to the number of predicted events for each process in each bin of
1560 the final BDT. In the limit of perfect understanding, a fit would correspond to a constant scale fac-
1561 tor on the red, signal histogram, where one would choose a best fit μ value, denoted $\hat{\mu}$, that would
1562 minimize the sum in quadrature of differences between the number of observed data events and
1563 $\mu s_i + b_i$, where s_i and b_i are the predicted number of signal and background events in each bin.

1564 The only source of uncertainty would be due to data statistics, so for an infinitely large dataset with
1565 perfect understanding, μ could be fitted to arbitrary precision. This, of course, is not the case since
1566 there is a finite amount of data and very many sources of systematic uncertainty, discussed in pre-
1567 vious chapters. This chapter will first describe how systematic uncertainties are integrated into the
1568 statistical fit of this analysis before describing two sets of cross checks on both a validation VZ fit and
1569 on the fit for the VH fit of interest.

1570 7.1 THE FIT MODEL

1571 In order to derive the strength of the signal process $ZH \rightarrow \ell\ell b\bar{b}$ and other quantities of interest
1572 while taking into account systematic uncertainties or nuisance parameters (NP’s, collectively de-
1573 noted θ), a binned likelihood function is constructed as the product over bins of Poisson distribu-

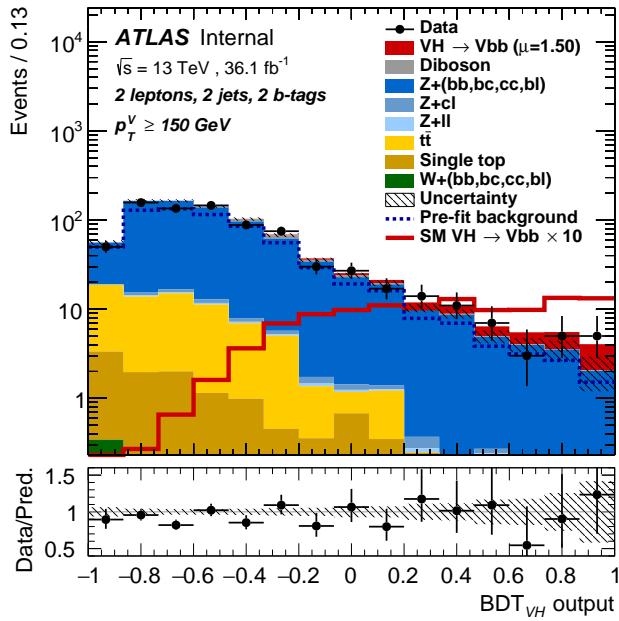


Figure 7.1: An example postfit distribution. The reason this looks different from postfit distributions later in this chapter is that this is a log plot.

1574 tions:

$$\mathcal{L}(\mu, \theta) = \text{Pois}(n | \mu S + B) \left[\prod_{i \in \text{bins}} \frac{\mu s_i + b_i}{\mu S + B} \right] \prod_{j \in \text{NP's}} \mathcal{N}_{\theta_j}(\theta_j, \sigma_j^2 | 0, 1) \quad (7.1)$$

1575 where n is the total number of events observed, s_i and b_i are the number of expected signal and back-

1576 ground events in each bin, and S and B are the total expected signal and background events. The

1577 signal and background expectations generally are functions of the NP's θ . NP's related to the nor-

1578 malization of signal and background processes fall into two categories. The first set is left to float

1579 freely like μ while the second set are parametrized as log-normally distributed to prevent negative

1580 predicted values. All other NP's are parametrized with Gaussian priors. This results in a “penalty”

1581 on the NLL discussed below of $(\hat{\alpha} - \mu_\alpha)^2 / \sigma_\alpha^2$, for NP α , normally parametrized with mean μ_α

1582 (corresponding to the nominal prediction) and variance σ_α^2 (derived as discussed in Chapters 4 and

1583 5) for an MLE of $\hat{\alpha}$.

1584 One can maximize^{*} the likelihood in Equation 7.1 for a fixed value of μ to derive estimators for
1585 the NP's θ ; values of θ so derived are denoted $\hat{\theta}_\mu$ to emphasize that these are likelihood maximizing
1586 for a given μ . The profile likelihood technique finds the likelihood function's maximum by compar-
1587 ing the values of the likelihood over all possible values of μ using these "profiles" and picking the
1588 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 like-
1589 lihood can further be used to construct a test statistic[†]

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

1590 This statistic can be used to derive the usual significance (p value), by setting $\mu = 0$ to find the com-
1591 patability with the background-only hypothesis⁴⁵. If there is insufficient evidence for the signal hy-
1592 pothesis, the CL_s method can be used to set limits¹⁸.

1593 In order to both validate the fit model and study the behavior of fits independent of a given dataset,
1594 a so-called "Asimov"[‡] dataset can be constructed for a given fit model; this dataset has each bin equal
1595 to its expectation value for assumed values of the NP's and a given μ value (in this case, $\mu = 1$, the
1596 SM prediction).

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

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

[‡]A reference to the short story quoted at the beginning of this chapter in which a computer picks a single voter to stand in for the views of the entire American electorate.

1597 7.2 FIT INPUTS

1598 Inputs to the binned likelihood are distributions of the BDT outputs described in Chapter 6 for the
1599 signal regions and of m_{bb} for the top $e - \mu$ control regions. These regions split events according
1600 to their p_T^V and number of jets. All events are required to have two b -tagged jets, as well as pass the
1601 other event selection requirements summarized in Table 5.8; the only difference between the signal
1602 and control region selections is that the same flavor requirement (i.e. leptons both be electrons or
1603 muons) is flipped so that events in the control region have exactly one electron and one muon. The
1604 BDT outputs are binned using transformation D, while the m_{bb} distributions have 50 GeV bins,
1605 with the exception of the 2 jet, high p_T^V region, where a single bin is used due to low statistics.

1606 Input distributions in MC are further divided according to their physics process. The signal pro-
1607 cesses are divided based on both the identity of associated V and the number of leptons in the final
1608 state; $ZH \rightarrow \ell\ell b\bar{b}$ events are further separated into distributions for qq and gg initiated processes.
1609 $V+jets$ events are split according to V identity and into the jet flavor bins described in Chapter 3.
1610 Due to the effectiveness of the 2 b -tag requirement suppressing the presence of both c and l jets,
1611 truth-tagging is used to boost MC statistics in the cc , cl , and ll distributions.[§] For top backgrounds,
1612 single top production is split according to production mode (s , t , and Wt), with $t\bar{t}$ as single category.
1613 Diboson background distributions are also split according to the identity of the V 's (ZZ , WZ , and
1614 WW). Fit input segmentation is summarized in Table 7.1.

1615 [§]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 $[(W, qqZ, ggZ)] \times n_{lep}$, $V+jets [(W, Z)] \times (bb, bc, bl, cc, cl, ll)]$, $t\bar{t}$, diboson (ZZ, WW, WZ), single top (s, t, Wt)

Table 7.1: Fit input segmentation.

1615 7.3 SYSTEMATIC UNCERTAINTIES REVIEW

1616 Tables 7.2 and 7.3 summarize modeling (Chapter 4) and experimental (Chapter 5) systematic uncer-
 1617 tainties considered in this analysis, respectively. In addition to these, simulation statistics uncertain-
 1618 ties (“MC stat errors”) are also included in the fit model. There is one distribution per systematic
 1619 (one each for up and down) per sample per region. The $\pm 1\sigma$ variation for a systematic is calculated
 1620 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)
$Z+jets$	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.

1621 The systematics distributions undergo processes known as “smoothing” and “pruning” before
 1622 being combined into the final likelihood used in minimization.
 1623 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.

1624 some stable value in the limit of large simulation statistics, but if the fluctuations due to simulation
 1625 statistics in a distribution are large compared to the actual physical effect (whether this is because
 1626 the actual effect is small or if the actual distribution is derived from a small number of simulation
 1627 events), then systematic uncertainty will be overestimated by, in effect, counting the MC stat error
 1628 multiple times. Smoothing is designed to mitigate these effects by merging adjacent bins in some
 1629 input distributions. Smoothing happens in two steps (the full details of smoothing algorithms may
 1630 be found in ⁶⁵ and in the `WSMaker` code):
 1631 1. Merge bins iteratively where bin differences are smallest in input distributions until no local
 1632 extrema remain (obviously, a single peak or valley is allowed to remain)
 1633 2. Sequentially merge bins (highest to lowest, like transformation D) until the statistical uncer-
 1634 tainty in a given bin is smaller than 5% of merged bin content
 1635 Not all systematic uncertainties defined are included in the final fit. Systematics are subject “prun-
 1636 ing” (individually in each region/sample: there are two histograms per systematic (up/down) per
 1637 region per sample, so pruning just consists of removing the histograms from the set of distributions
 1638 included in the likelihood) if they are do not have a significant impact, defined as follows:

- Normalization/acceptance systematics are pruned away if either:
 - The variation is less than 0.5%
 - Both up and down variations have the same sign

- Shape systematics pruned away if either:
 - Not one single bin has a deviation over 0.5% after the overall normalisation is removed
 - If only the up or the down variation is non-zero

- Shape+Normalisation systematics are pruned away if the associated sample is less than 2% of the total background and either:
 - If the predicted signal is < 2% of the total background in all bins and the shape and normalisation error are each < 0.5% of the total background
 - If instead at least one bin has a signal contribution > 2% of the total background, and only in each of these bins, the shape and normalisation error are each < 2% of the signal yield

7.4 THE VZ VALIDATION FIT

One of the primary validation cross-checks for the fiducial analysis was a VZ fit—that is, conducting the entire analysis but looking for $Z \rightarrow b\bar{b}$ decays instead of the Higgs. The idea here is that the Z is very well understood and so “rediscovering” Z decay to b 's is taken as a benchmark of analysis reliability since the complexity of the fit model precludes the use of orthogonal control regions for validation as is done in other analyses (generally, if there is a good control region, one prefers to use it

1658 to constrain backgrounds and improve the fit model). To do this, a new MVA discriminant is made
 1659 by keeping all hyperparameter configurations the same (e.g. variable ranking) but using diboson
 1660 samples as signal. For the 2-lepton case, this means using $ZZ \rightarrow \ell\ell b\bar{b}$ as the signal sample. This
 1661 new MVA is used to make the inputs described in Section 7.2, and the fit is then run as for the VH
 1662 fit (again, with ZZ as signal). VH samples are considered background in these diboson fits.

1663 The VZ fit sensitivities for the standard, LI, and RF fits are summarized in Table 7.4. The ex-
 1664 pected significances are all fairly comparable and about what was the case in the fiducial analysis.
 1665 The observed significance for the standard set matches fairly well with the expected value on data,
 1666 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.

1667 These values, however, are consistent with the observed signal strength values, which can be seen
 1668 in Figure 7.2 (b), with both the LI and RF fits showing a deficit of signal events with respect to the
 1669 SM expectation, though not by much more than one standard deviation (a possible explanation is
 1670 explored in the following section). Just as in the VH fits, errors arising systematic uncertainties are
 1671 lower in the fits to the observed dataset. That the effect is not noticeable in Asimov fits is not too
 1672 surprising, since this analysis (and these variable configurations in particular), is not optimized for
 1673 VZ .

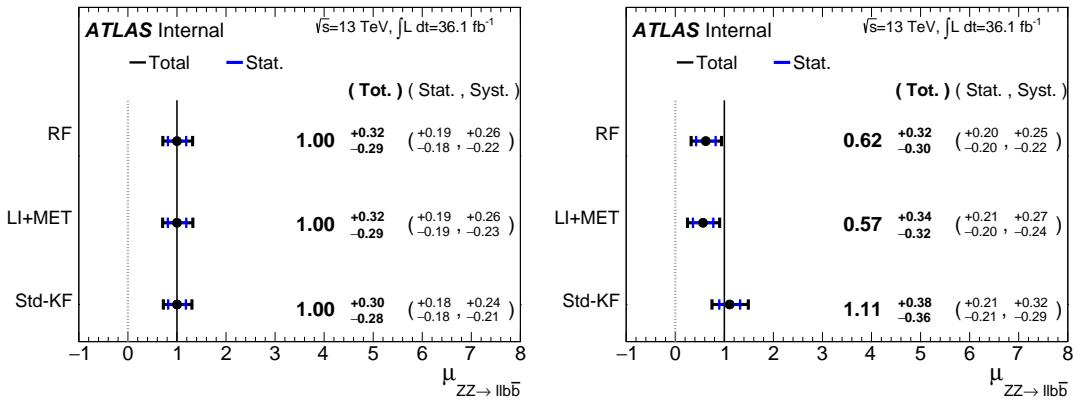


Figure 7.2: μ summary plots for the standard, LI, and RF variable sets. The Asimov case (with $\mu = 1$ by construction) is in (a), and $\hat{\mu}$ best fit values and error summary are in (b).

1674 7.4.I 2 AND ≥ 3 JET FITS

1675 While the treatment of simply ignoring any additional jets in the event seems adequate for the VH
 1676 analysis (discussed below), the potential shortcoming of this treatment appears in the VZ analysis
 1677 when the 2 and ≥ 3 jet cases are fit separately[¶], as can be seen in Figure 7.3. Compared to the stan-
 1678 dard fit, the LI and RF fits have lower $\hat{\mu}_{\geq 3 \text{ jet}}$ values, consistent with the interpretation that the ad-
 1679 ditional information from the third jet in the ≥ 3 jet regions for the standard case is important for
 1680 characterizing events in these regions for VZ fits.

1681 A natural question to ask is why this would be an issue for the VZ but not the VH case. One
 1682 potential answer is that at high transverse boosts, there is a greater probability for final state
 1683 radiation in the hadronically decaying Z , so there are more events where the third jet should be in-
 1684 cluded in the calculation of variables like $m_{b\bar{b}}$ or for angles involving the $b\bar{b}$ system (e.g. $\cos\theta$ in the
 1685 RF case). While the absolute scale at which the low and high p_T^V regions are separated remains the

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

same does not change from the VH to the VZ analysis, 150 GeV, the implicit cutoff on the transverse boost of the hadronically decaying boson does. For the Higgs, with a mass of 125 GeV, the p_T^V cutoff 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 23–37% higher.

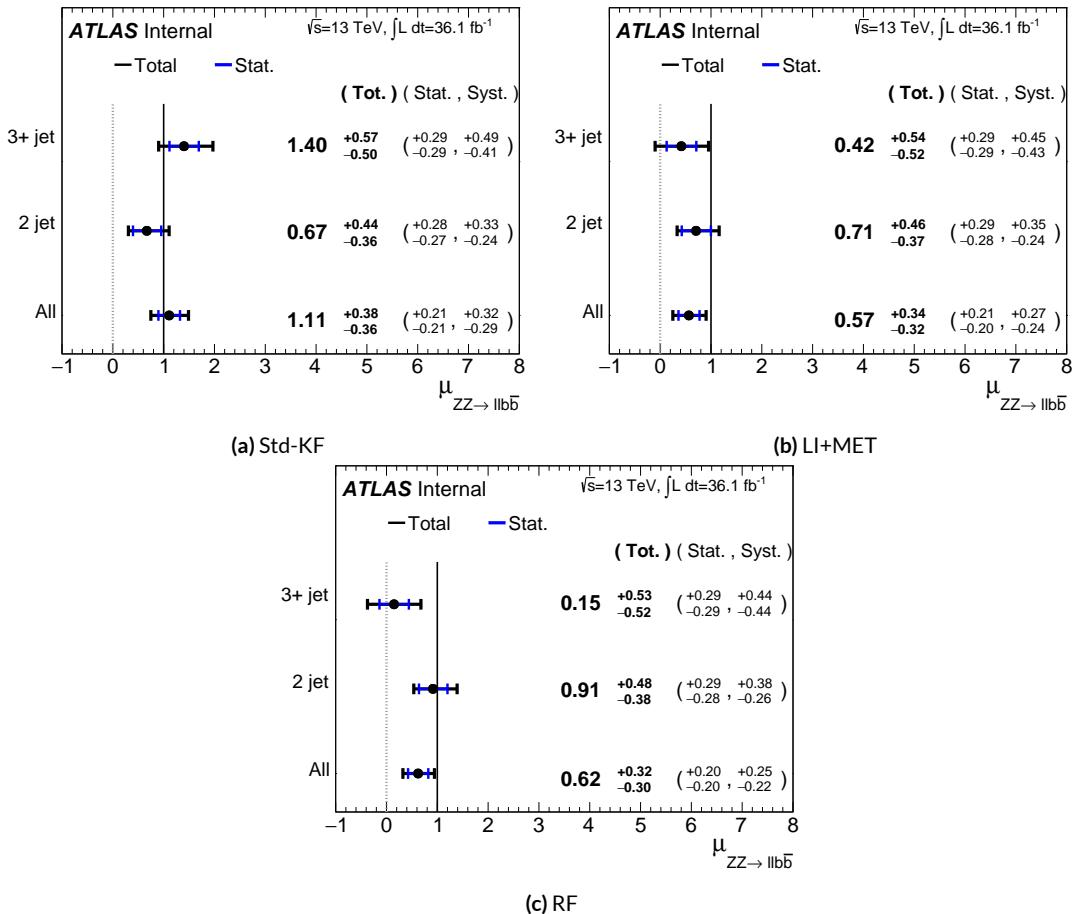


Figure 7.3: $\hat{\mu}$ summary plots with standalone fits for the different n_{jet} regions for the standard, LI, and RF variable sets.

If either the LI or RF schemes were to be used in a mainstream analysis, these validation fits suggest that the third jet ought to be included in variable schemes (e.g. by adding the third jet to the

1692 Higgs in the high p_T^V case). On the issue of whether or not ≥ 4 jet events should be included, the
1693 RF set shows very little sensitivity to this change (a 2 jet and 3 jet only fit moves $\hat{\mu}$ to 0.64, while
1694 doing so for the LI set moves it to 0.40), so this, like the addition of the third jet into the variable
1695 sets, would have to be addressed individually. Nevertheless, this optimization is beyond the scope of
1696 this thesis, which aims to preserve as much of the fiducial analysis as possible for as straightforward a
1697 comparison as possible.

1698 For completeness, we include the full set of fit validation results for the VZ fit, explaining them in
1699 turn.

1700 7.5 NUISANCE PARAMETER PULLS

1701 The first set of plots statistical fit experts will want to look at are the “pulls” and “pull comparisons.”
1702 In these plots, the best fit (nominal) values and one standard deviation error bars are shown for ob-
1703 served (Asimov) pull plots, with the green and yellow bands corresponding to $\pm 1, 2\sigma$, respectively.
1704 These plots are divided by NP category for readability. [¶] In pull comparisons, these pulls are over-
1705 layed and color-coded. Pull comparisons here have the following color code: black is the standard
1706 variable set, red is the LI set, and blue is the RF set.

1707 A well-behaved fit has pulls close to nominal values (“closeness” should be interpreted in the
1708 context of pull value divided by pull error). As can be seen in Figures 7.4–7.8, the fits for the three
1709 different variable sets are fairly similar from a NP pull perspective, though the $Z+{\rm jets}$ m_{bb} and p_T^V

1708 [¶]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.

1710 NP's and the jet energy resolution NP are heavily pulled (a handful of poorly behaved pulls is not
 1711 uncommon, though typically warrants further investigation). As a general note, these pull plots cal-
 1712 culate pulls using a simultaneous HESSE matrix inversion, which is fine for relatively small fits, but
 1713 the more reliable MINOS result, which calculates the impact of each NP on its own, should be cross-
 1714 checked for significant pulls**. The ranking plots below do this.

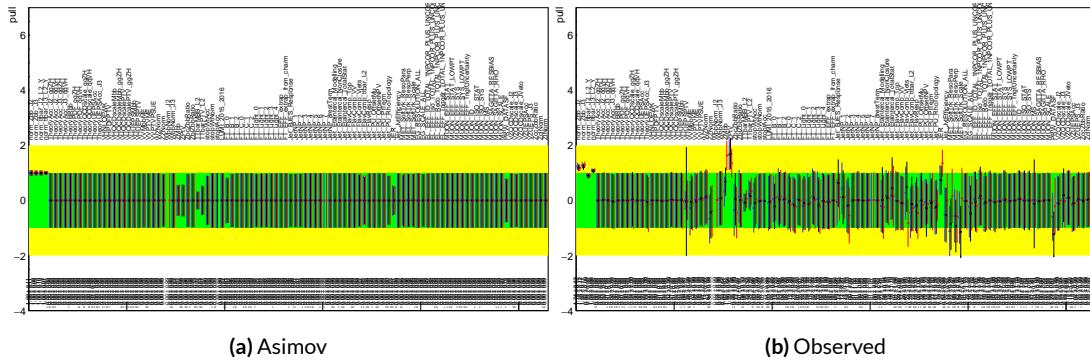


Figure 7.4: Pull comparison for all NP's but MC stats.

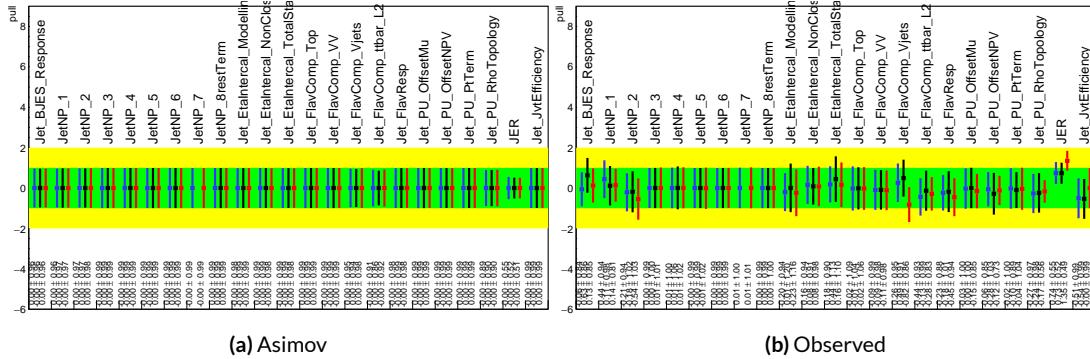


Figure 7.5: Pull comparison for jet NP's.

1715 Nuisance parameter correlation matrices (for correlations with magnitude at least 0.25) for all
 1716 three variable set fits can be found in Figures 7.10–7.12. These are useful for seeing which NP's move

**This becomes more of an issue for very large fits, like the full Run 1 + Run 2 combined fits in Chapter 9.

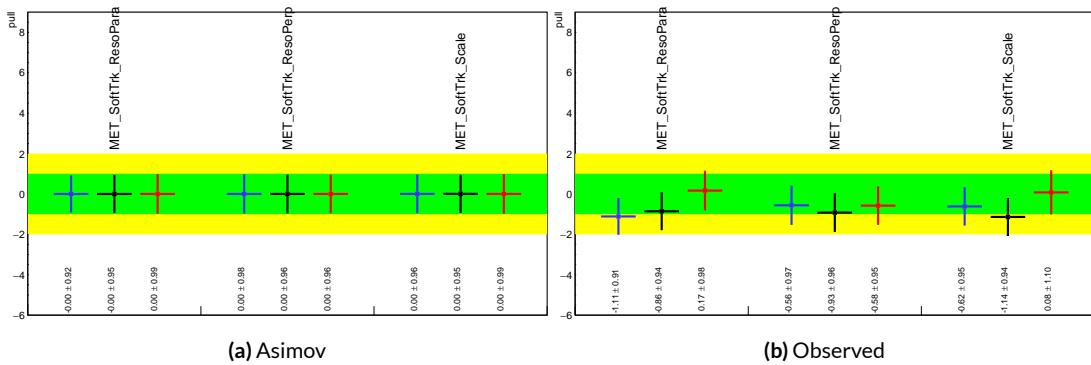


Figure 7.6: Pull comparison for MET NP's.

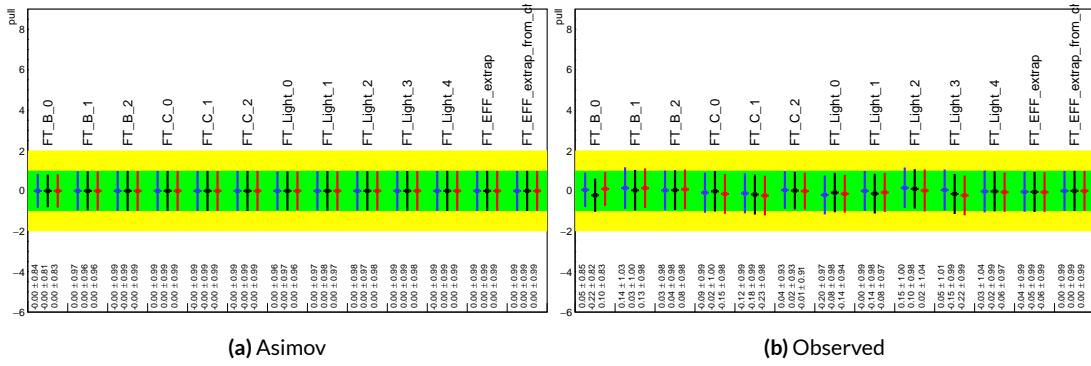


Figure 7.7: Pull comparison for Flavour Tagging NP's

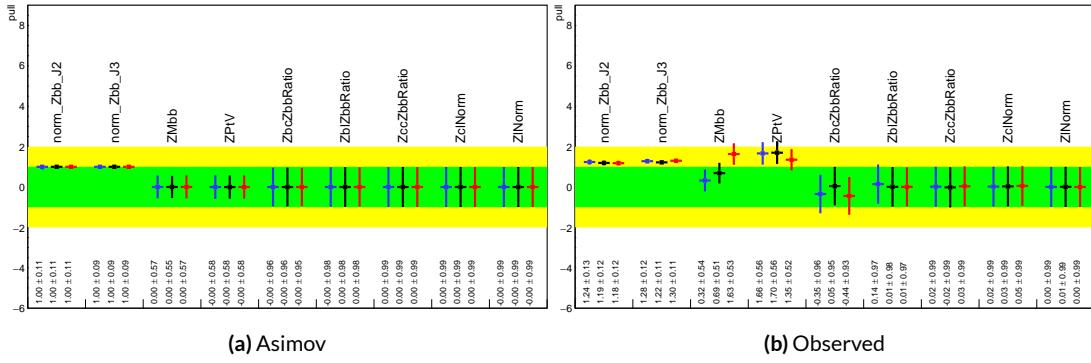


Figure 7.8: Pull comparison for Z +jets NP's

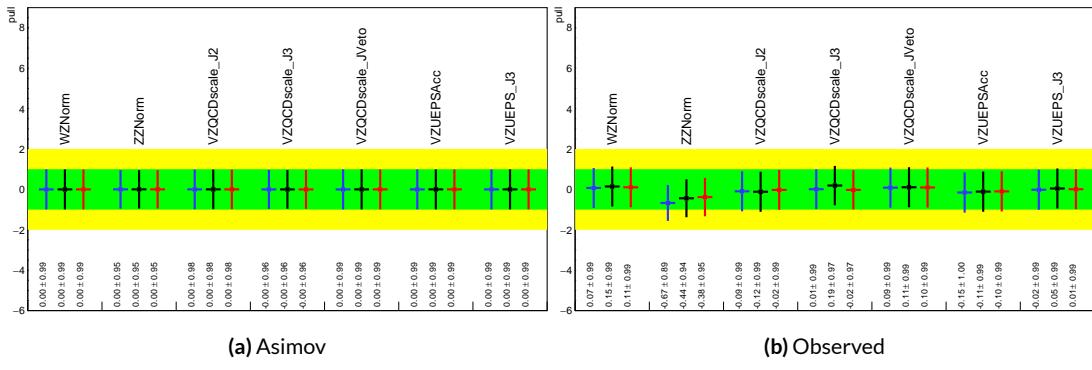


Figure 7.9: Pull comparison for signal process modeling NP's.

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

investigation is warranted).

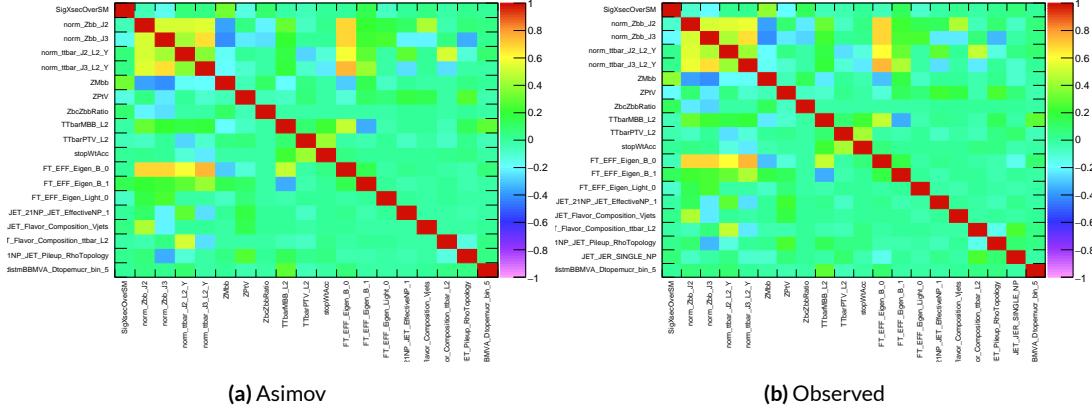


Figure 7.10: NP correlations for standard variable fits.

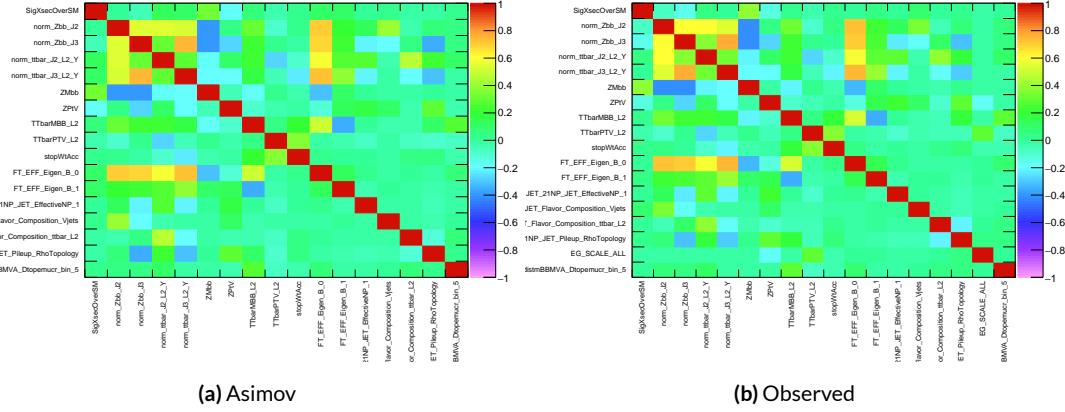


Figure 7.11: NP correlations for LI variable fits.

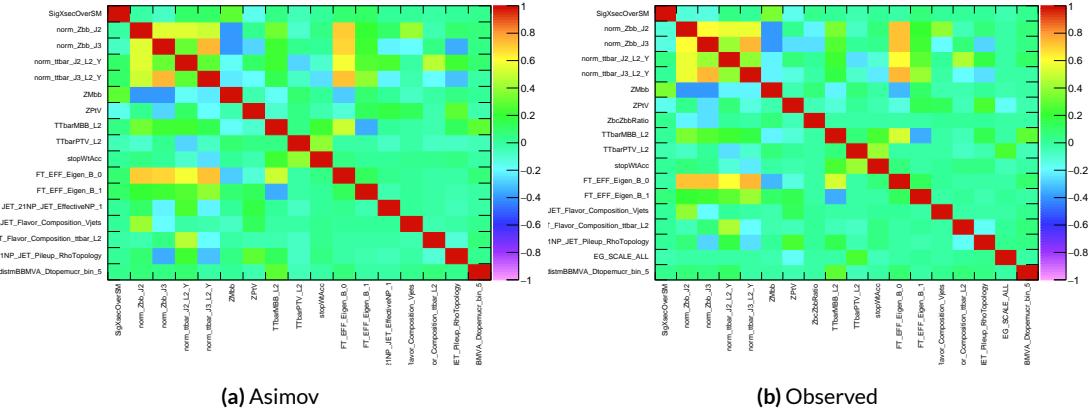


Figure 7.12: NP correlations for RF variable fits.

1719 7.5.1 NUISANCE PARAMETER RANKING PLOTS AND BREAKDOWNS

1720 The next set of fit results that is used to diagnose the quality of a fit is the impact of different nui-
 1721 sance parameters on the total error on μ , both individually and as categories. Figure 7.13 shows the
 1722 top 25 nuisance parameters ranked by their postfit impact on $\hat{\mu}$; these plots use the aforementioned,
 1723 more reliable MINOS approach. This set of rankings is fairly similar, with $Z+jets$ systematics being
 1724 particularly prominent. The advantage of seeing individual nuisance parameter rankings, as op-
 1725 posed to impacts of categories in aggregate, is that particularly pathological NP's are easier to see;
 1726 in 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 .

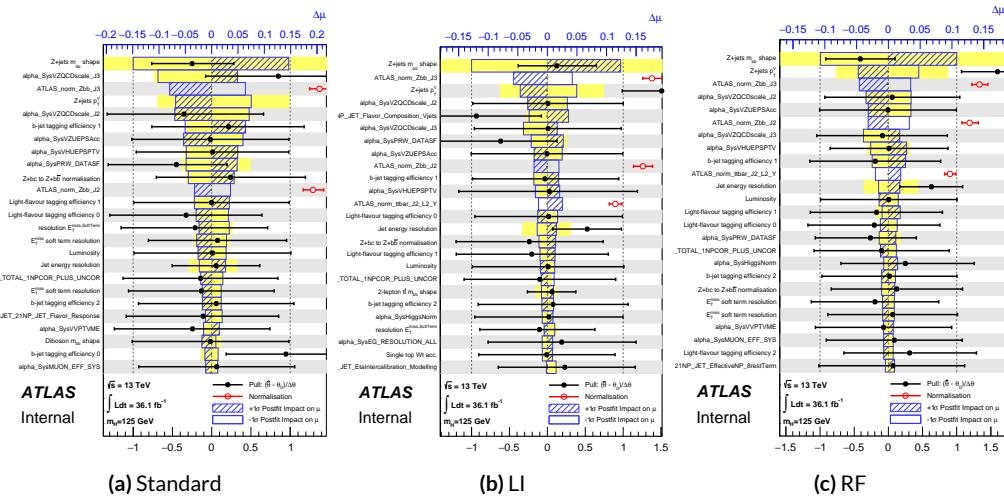


Figure 7.13: Plots for the top 25 nuisance parameters according to their postfit impact on $\hat{\mu}$ for the standard (a), LI (b), and RF (c) variable sets.

1727
 1728 This is consistent with the picture of NP's taken in aggregate categories in Tables 7.5 and 7.6,
 1729 known as “breakdowns,” with $Z+jets$ in particular featuring prominently. Of particular interest

¹⁷³⁰ is 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.

1731 7.6 POSTFIT DISTRIBUTIONS

1732 Finally, postfit distributions for the MVA discriminant (m_{bb}) distribution in the signal (top $e - \mu$
1733 control) region for the standard, Lorentz Invariant, and RestFrames variable sets are shown. It is
1734 generally considered good practice to check the actual postfit distributions of discriminating quan-
1735 tities used to make sure there is good agreement. ^{††} It should be noted that agreement is not always
1736 great when “eyeballing” a distribution, as fits are messy and $V+hf$ modeling is notoriously hard.
1737 This is particularly true in the VZ fit since normalizations for $Z+hf$ in particular are derived using
1738 VH optimized sidebands. This is also why a lot of these plots are presented as log plots (which hide
1739 disagreement better; the general argument goes that one has the ratio plots on the bottom and log
1740 plots allow one to see rare backgrounds in plots).

^{††}Sometimes distributions of input variables (MC histograms scaled by their postfit normalizations) are also used.

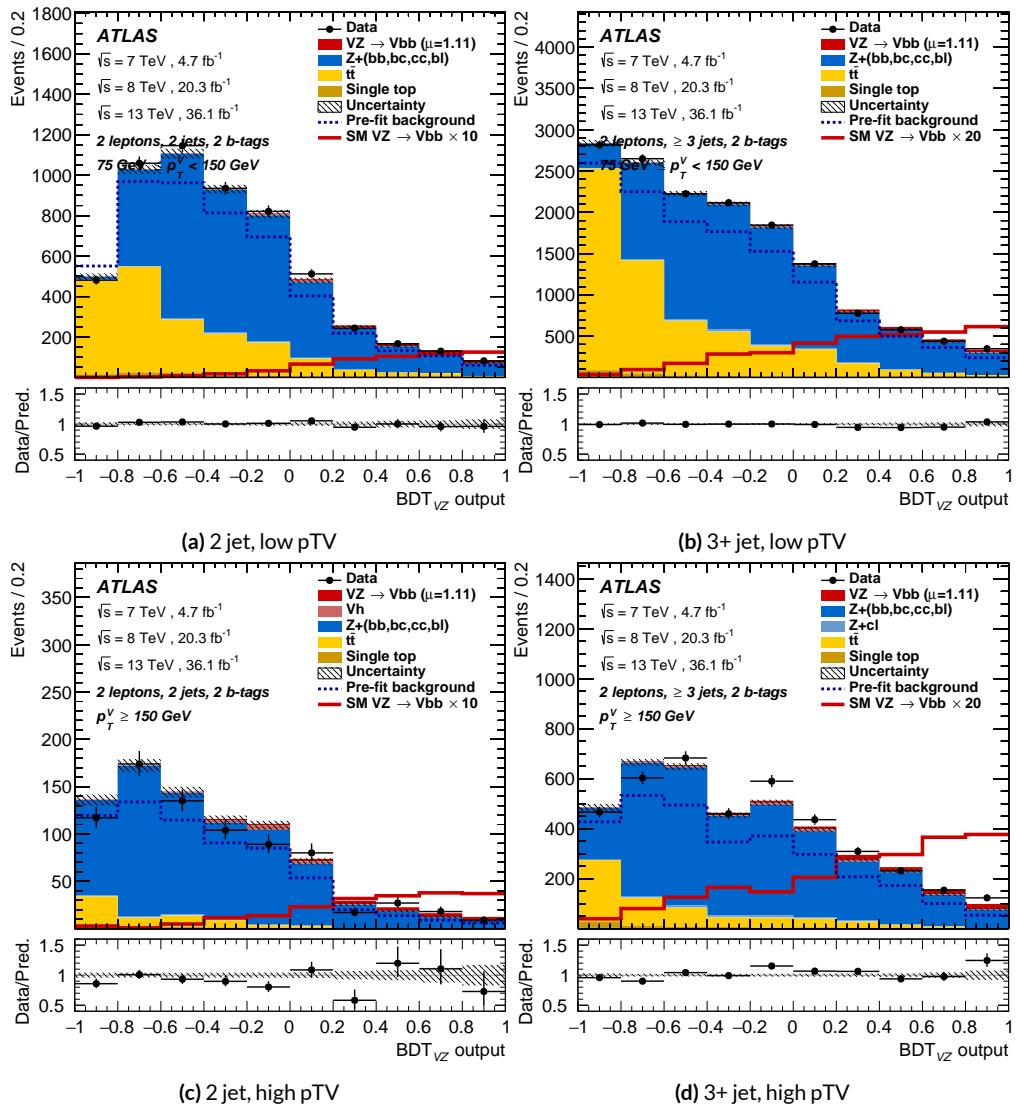


Figure 7.14: Postfit BDT_{VZ} plots in the signal region for the standard variable set.

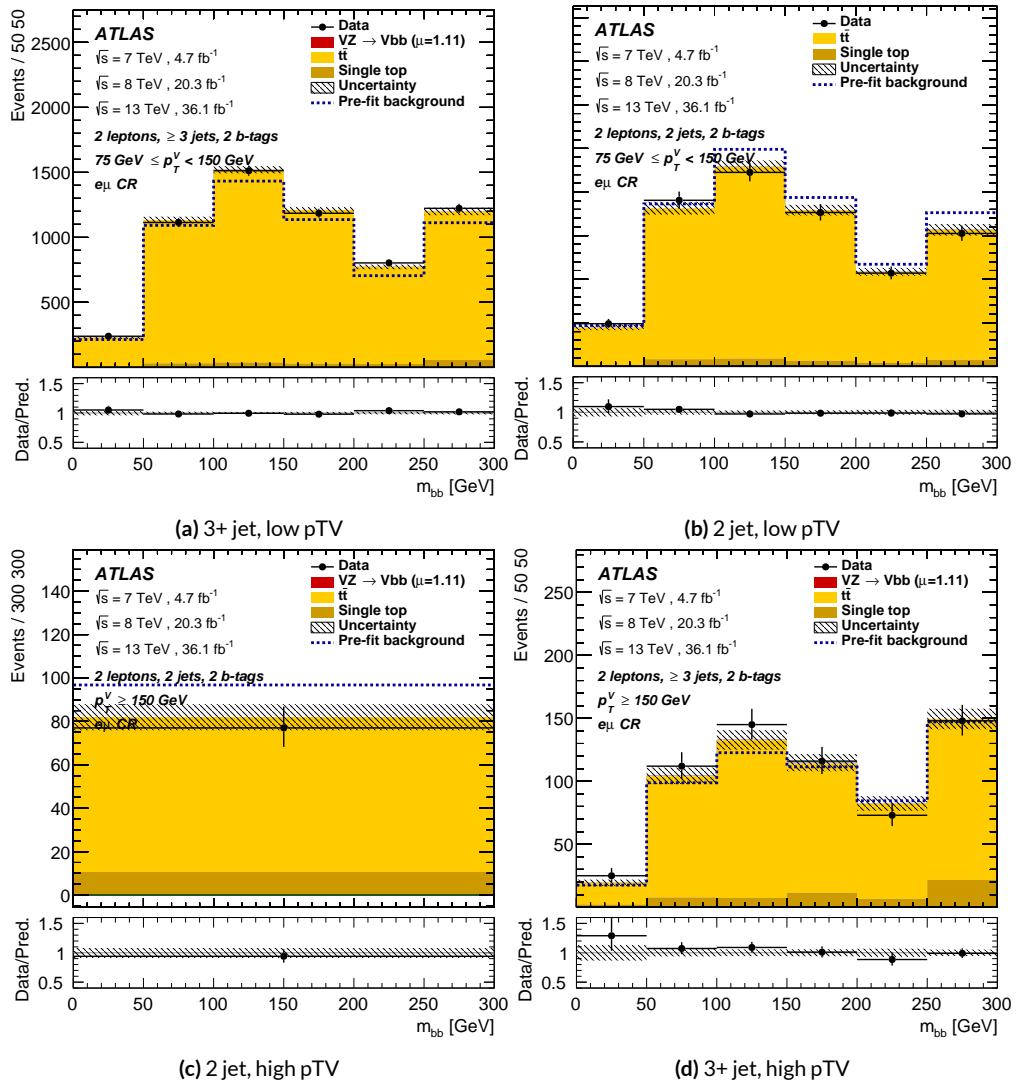


Figure 7.15: Postfit m_{bb} plots in the top $e - \mu$ CR for the standard variable set.

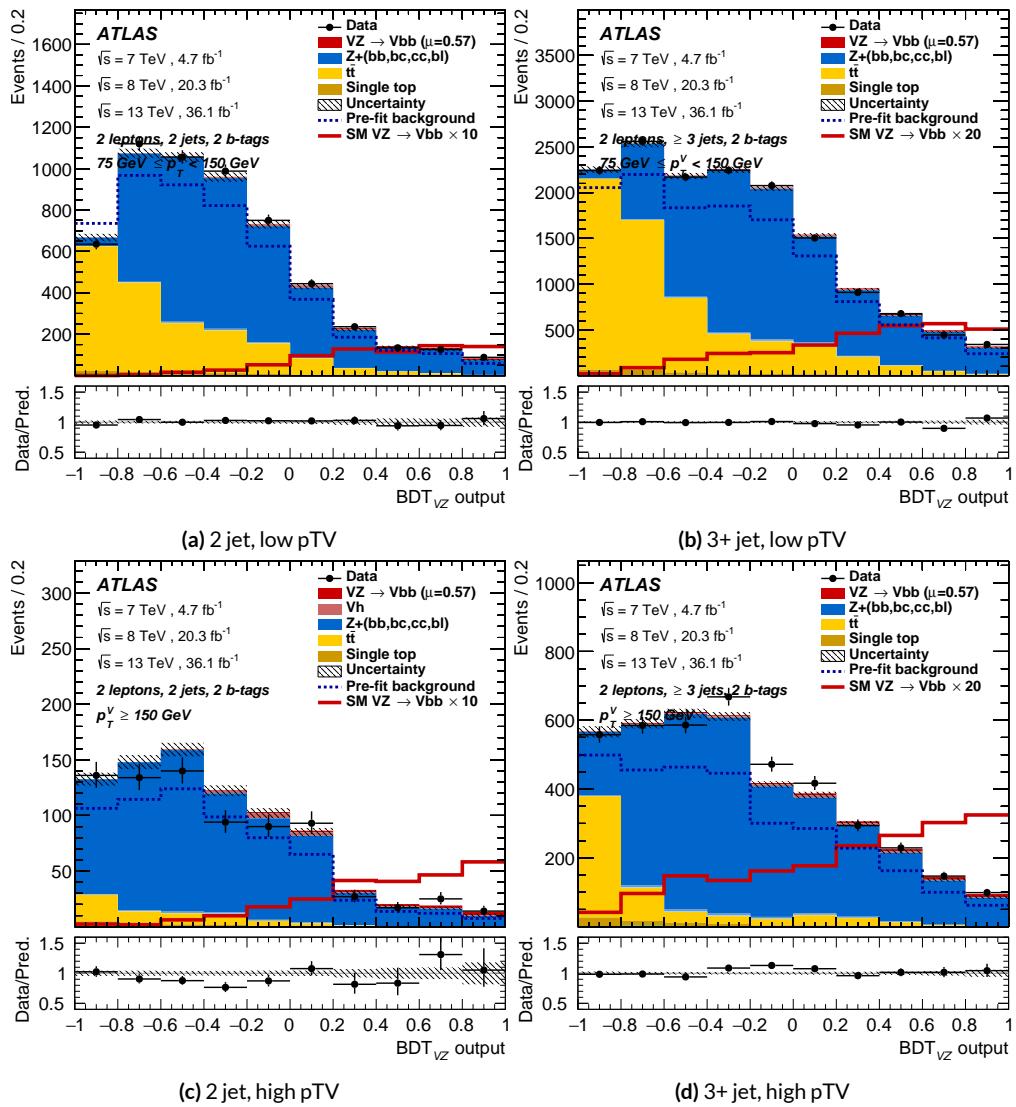


Figure 7.16: Postfit BDT_{VZ} plots in the signal region for the LI variable set.

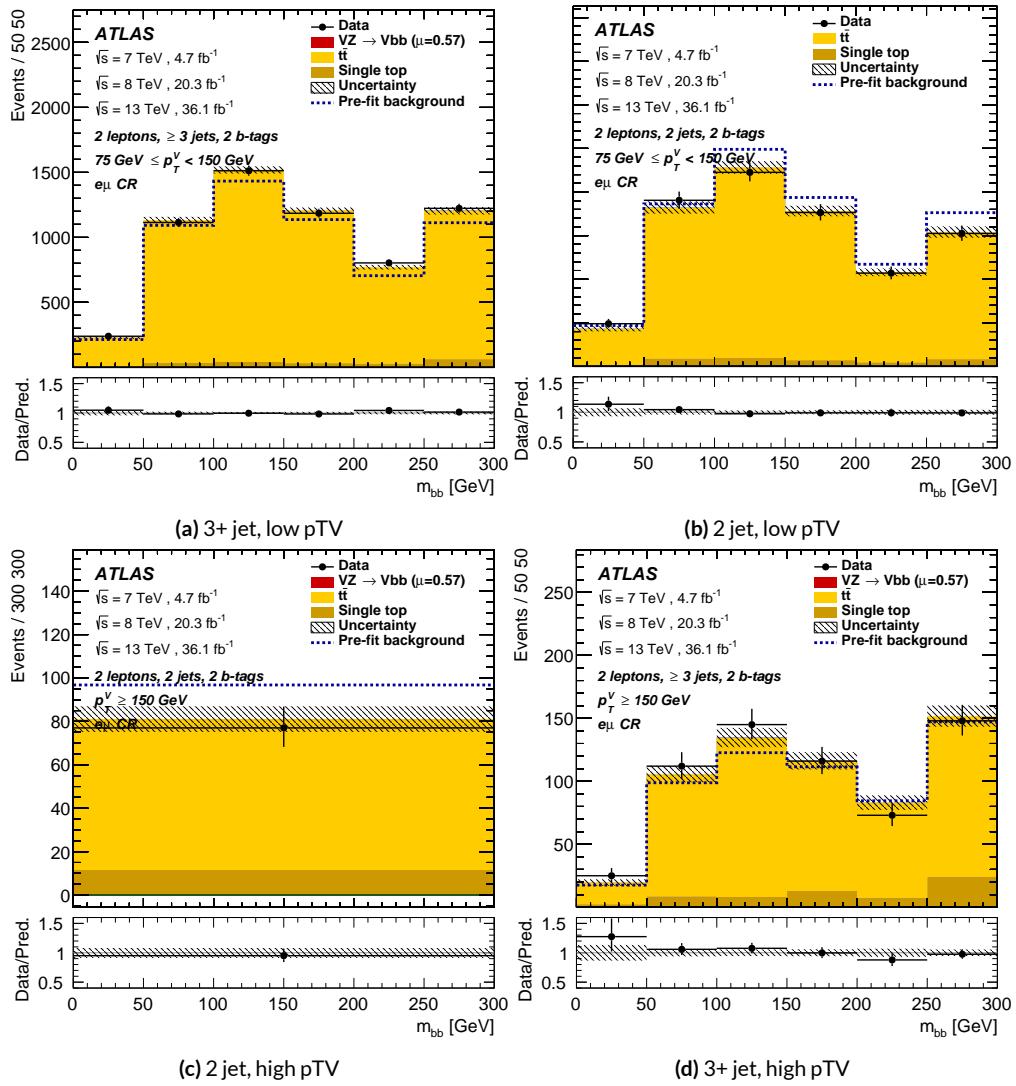


Figure 7.17: Postfit m_{bb} plots in the top $e - \mu$ CR for the LI variable set.

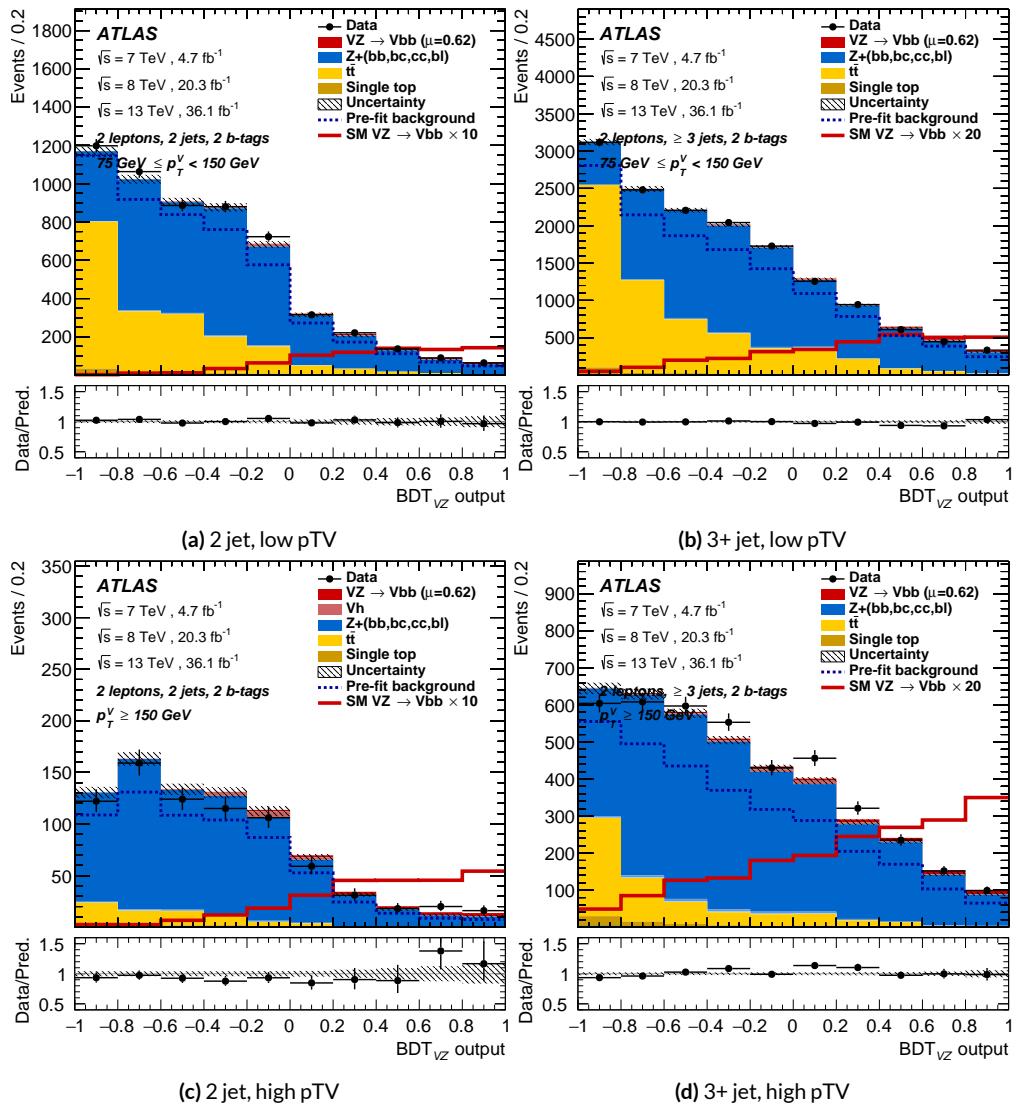


Figure 7.18: Postfit BDT_{VZ} plots in the signal region for the RF variable set.

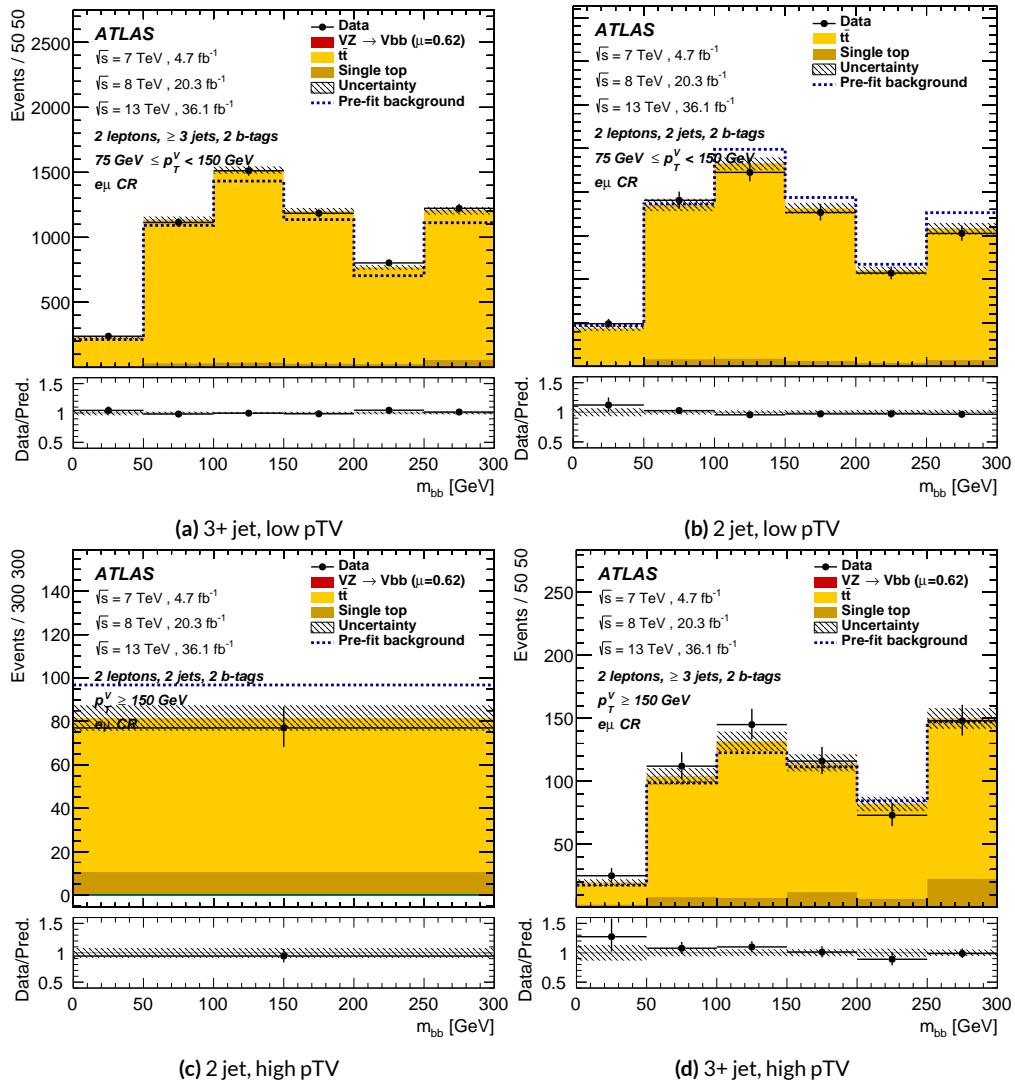


Figure 7.19: Postfit m_{bb} plots in the top $e - \mu$ CR for the RF variable set.

1741 7.7 VH FIT MODEL VALIDATION

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

1743 7.7.1 NUISANCE PARAMETER PULLS

As can be seen in Figures 7.20–7.24, the fits for the three different variable sets are fairly similar from a NP pull perspective. Again, black is the standard variable set, red is the LI set, and blue is the RF set. The possible exception is the signal UE+PS p_T^V systematic, which looks very different for all three cases (underconstrained for the standard, but overconstrained for the novel variable cases), though this difference goes away in the ranking plot, meaning this is almost certainly an unphysical artifice of the faster HESSE inversion used to produce the pull comparison plots.

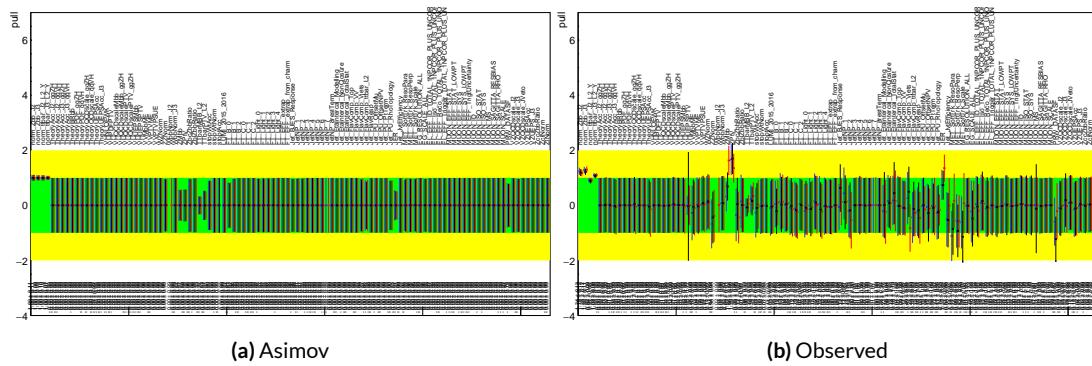


Figure 7.20: Pull comparison for all NP's but MC stats.

¹⁷⁵⁰ Nuisance parameter correlation matrices (for correlations with magnitude at least 0.25) for all
¹⁷⁵¹ three variable set fits can be found in Figures 7.26–7.28.

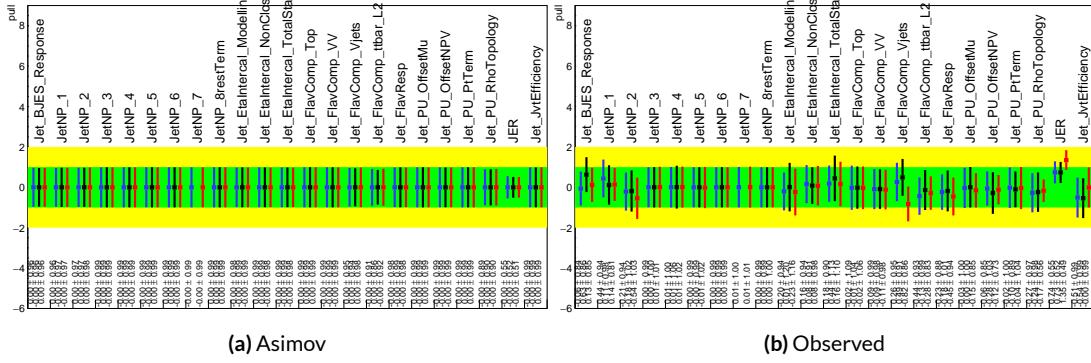


Figure 7.21: Pull comparison for jet NP's.

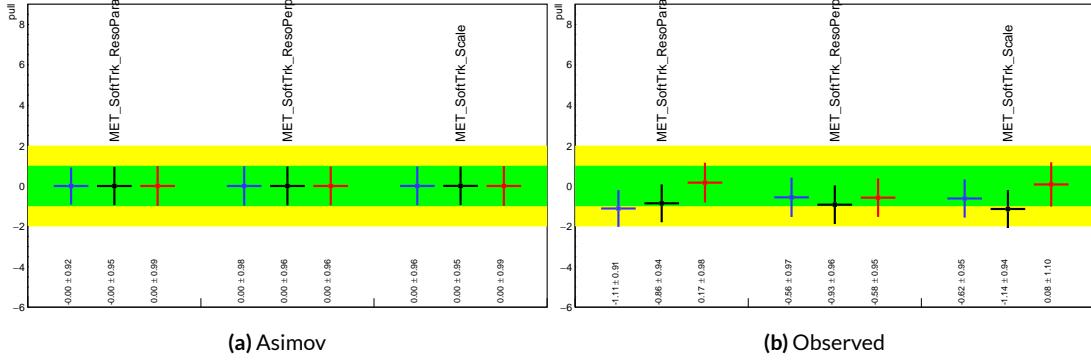


Figure 7.22: Pull comparison for MET NP's.

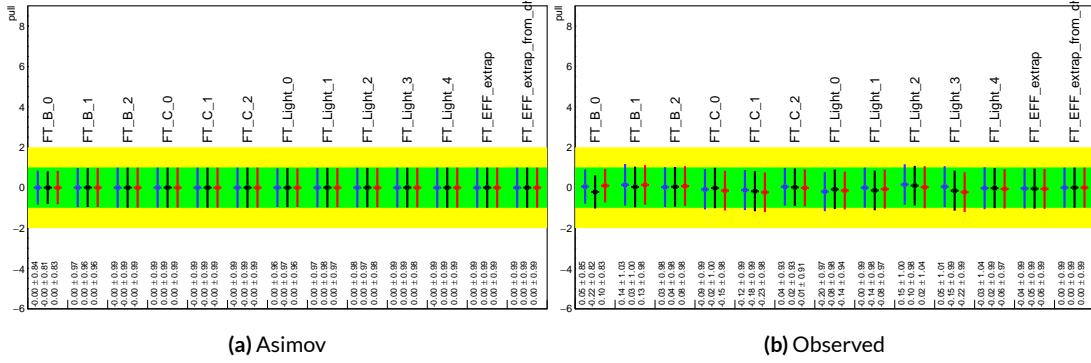


Figure 7.23: Pull comparison for Flavour Tagging NP's.

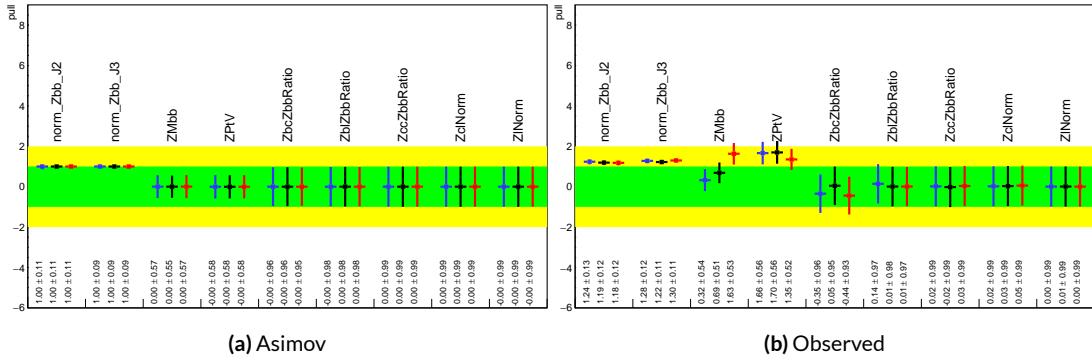


Figure 7.24: Pull comparison for $Z + \text{jets}$ NP's.

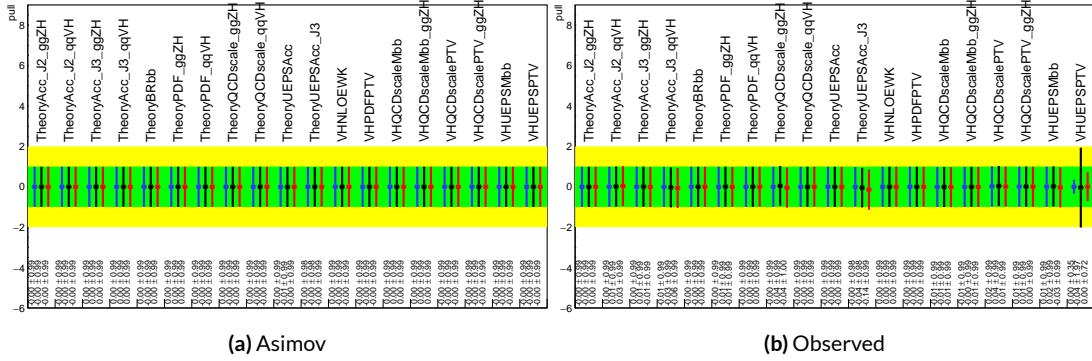


Figure 7.25: Pull comparison for signal process modeling NP's.

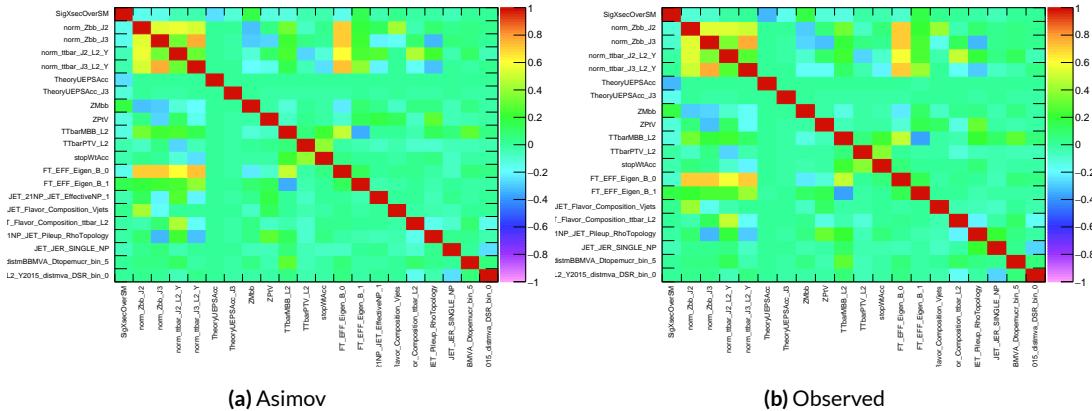


Figure 7.26: NP correlations for standard variable fits.

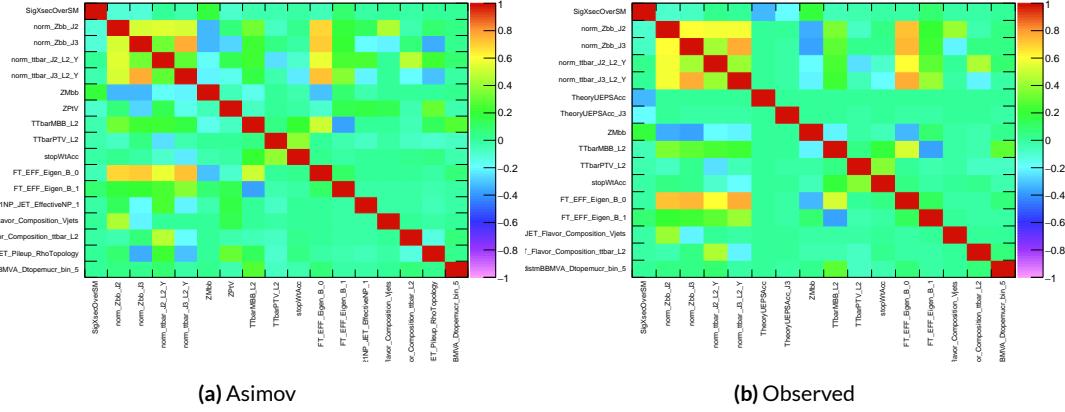


Figure 7.27: NP correlations for LI variable fits.

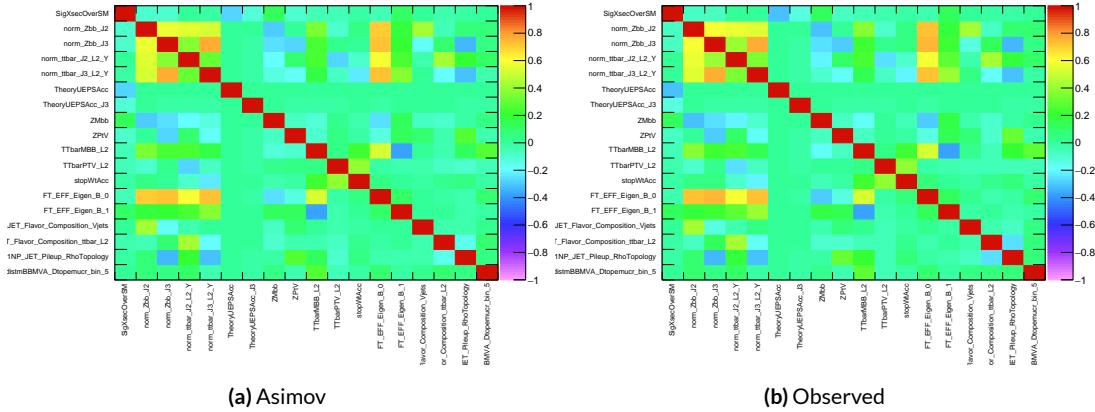


Figure 7.28: NP correlations for RF variable fits.

1752 7.7.2 FULL BREAKDOWN OF ERRORS

1753 A postfit ranking of nuisance parameters according to their impact on $\hat{\mu}$ for the different variable
 1754 sets may be found in Figure 7.29, with rankings being fairly similar. In particular, the signal UE+PS
 1755 p_T^V systematic is top-ranked for all three variable sets and also looks very similar, unlike in the pull
 1756 comparison plot, reiterating the importance of evaluating individually the impact of highly ranked
 1757 NP's. The $Z+jets p_T^V$ is highly pulled in all three cases, though this is less severe for the non-standard
 1758 set (it is off the scale for the standard variable ranking). The RF discriminant mitigates the effect of
 1759 poorly modeled jet energy resolution better than the other sets.

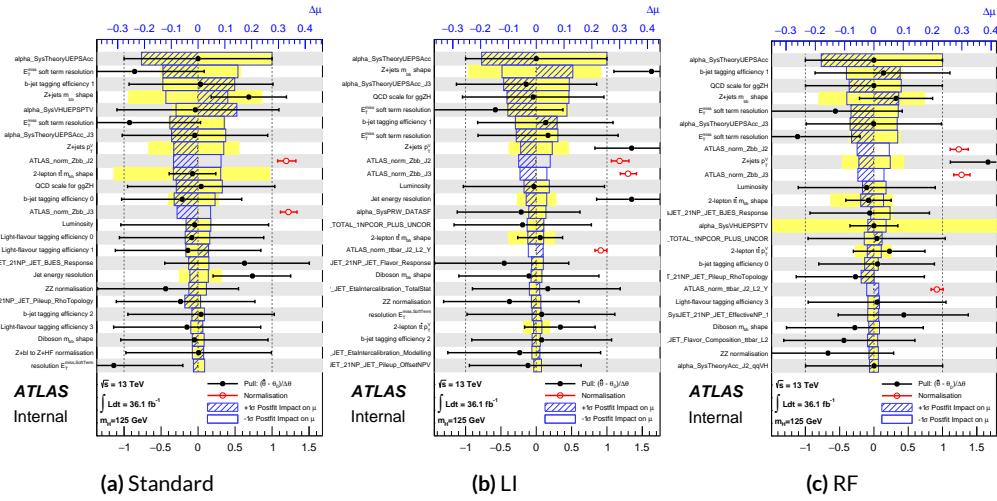


Figure 7.29: Plots for the top 25 nuisance parameters according to their postfit impact on $\hat{\mu}$ for the standard (a), LI (b), and RF (c) variable sets.

1760 The Asimov (Table 7.7) and observed (Table 7.8) breakdowns both consistently suggest that the
 1761 LI variable set does a better job of constraining systematic uncertainties than the standard set and
 1762 that the RF set does better still. It is also not surprising that the gain is more substantial in the ob-

¹⁷⁶³ served fit than in the Asimov fits, as in the latter there are the “penalty” terms from pulls in addition
¹⁷⁶⁴ 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

1765 7.7.3 POSTFIT DISTRIBUTIONS AND S/B PLOTS

1766 Postfit distributions for the MVA discriminant (m_{bb}) distribution in the signal (top $e - \mu$ control)
1767 region for the standard, Lorentz Invariant, and RestFrames variable sets are found in Figures 7.30–
1768 7.35. Here, as in the VZ fit, agreement is reasonable. In a combined fit with all three channels, $Z+hf$
1769 normalizations in particular would be correlated across the 0- and 2-lepton channels, which might
1770 help to better constrain this mismodeling (and perhaps as a result some of the $Z+jets$ systematics as
1771 well).

1772 One final type of plot presented as a result is the binned $\log_{10} (S/B)$ in signal regions distribu-
1773 tions may be found in Figure 7.36. For these plots, one fills a histogram with the $\log_{10} (S/B)$ ratio in
1774 each postfit distribution bin weighted by the total number of events. In this case, a log plot is help-
1775 ful because the highest bins would be invisible on a linear plot. These distributions are allegedly use-
1776 ful for seeing where most of one's sensitivity lies. Practically, it is problematic if the pull (from the
1777 null hypothesis) is higher at lower S/B values, which may indicate a poorly optimized discriminant.

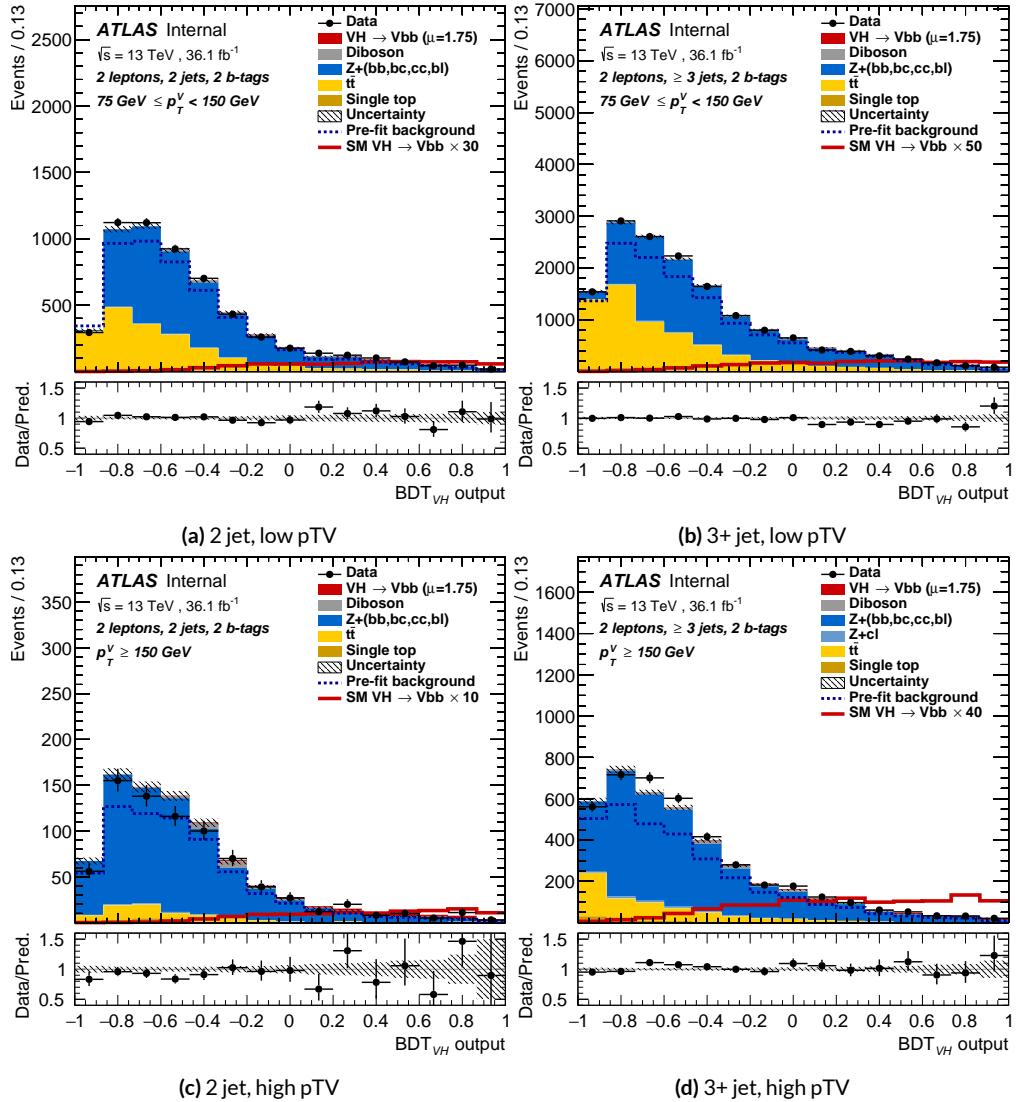


Figure 7.30: Postfit BDT_{VH} plots in the signal region for the standard variable set.

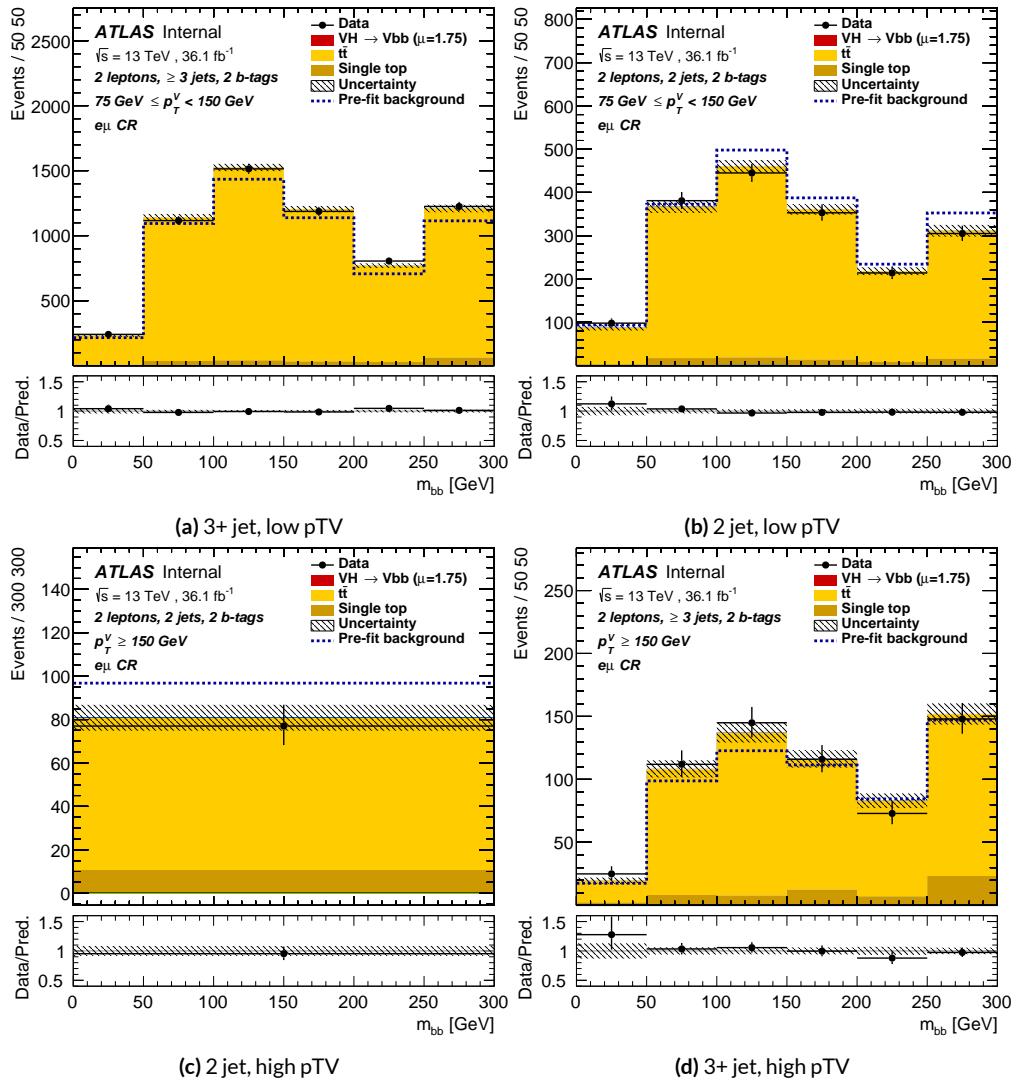


Figure 7.31: Postfit m_{bb} plots in the top $e - \mu$ CR for the standard variable set.

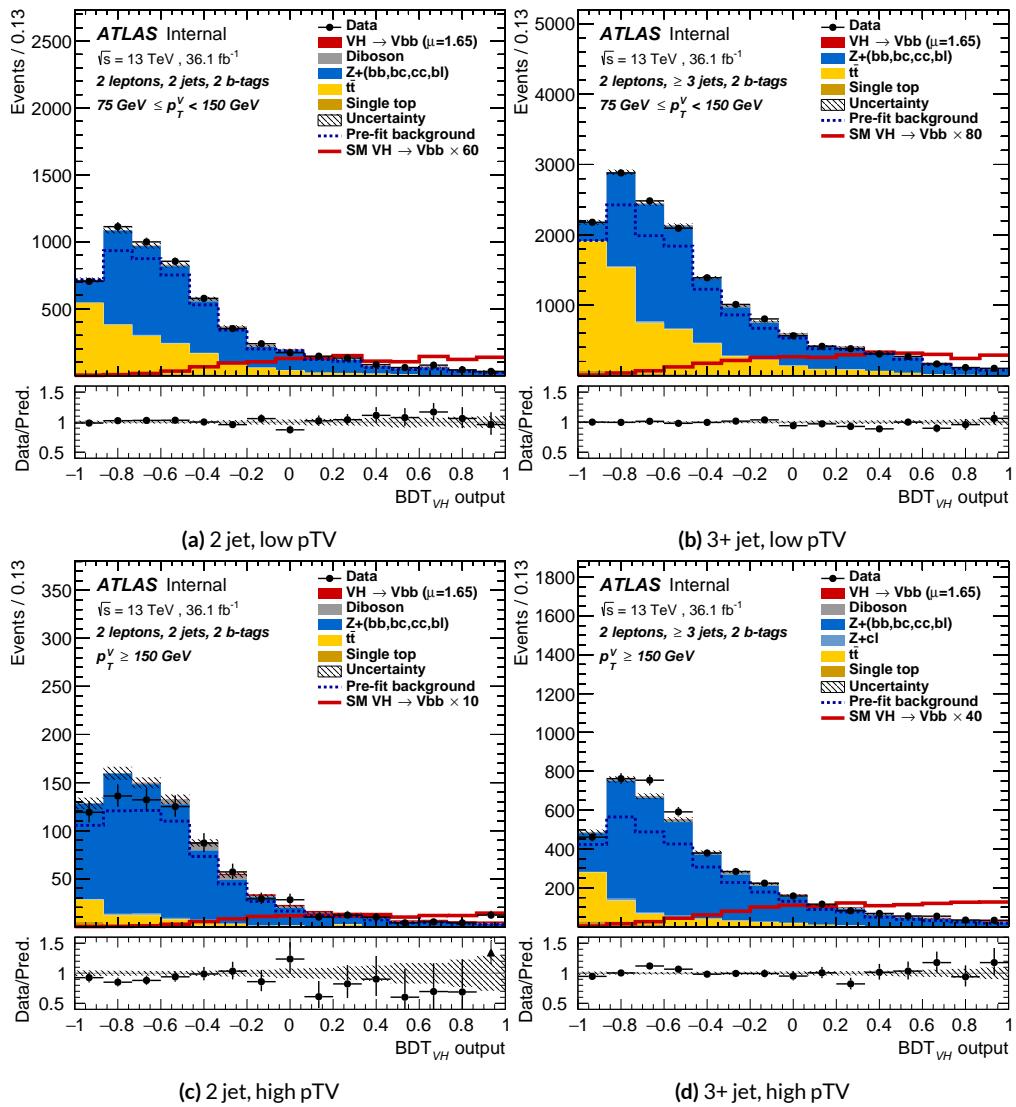


Figure 7.32: Postfit BDT_{VH} plots in the signal region for the LI variable set.

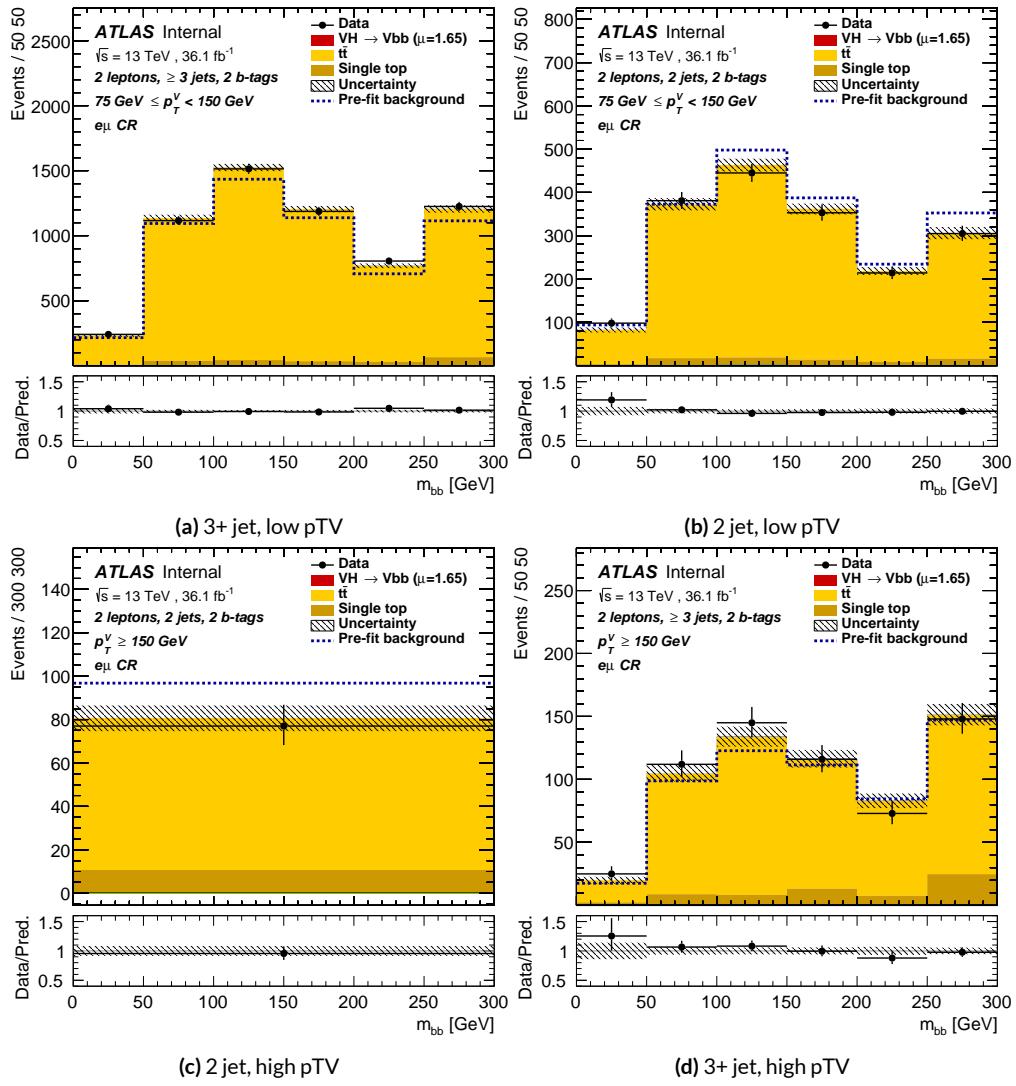


Figure 7.33: Postfit m_{bb} plots in the top $e - \mu$ CR for the LI variable set.

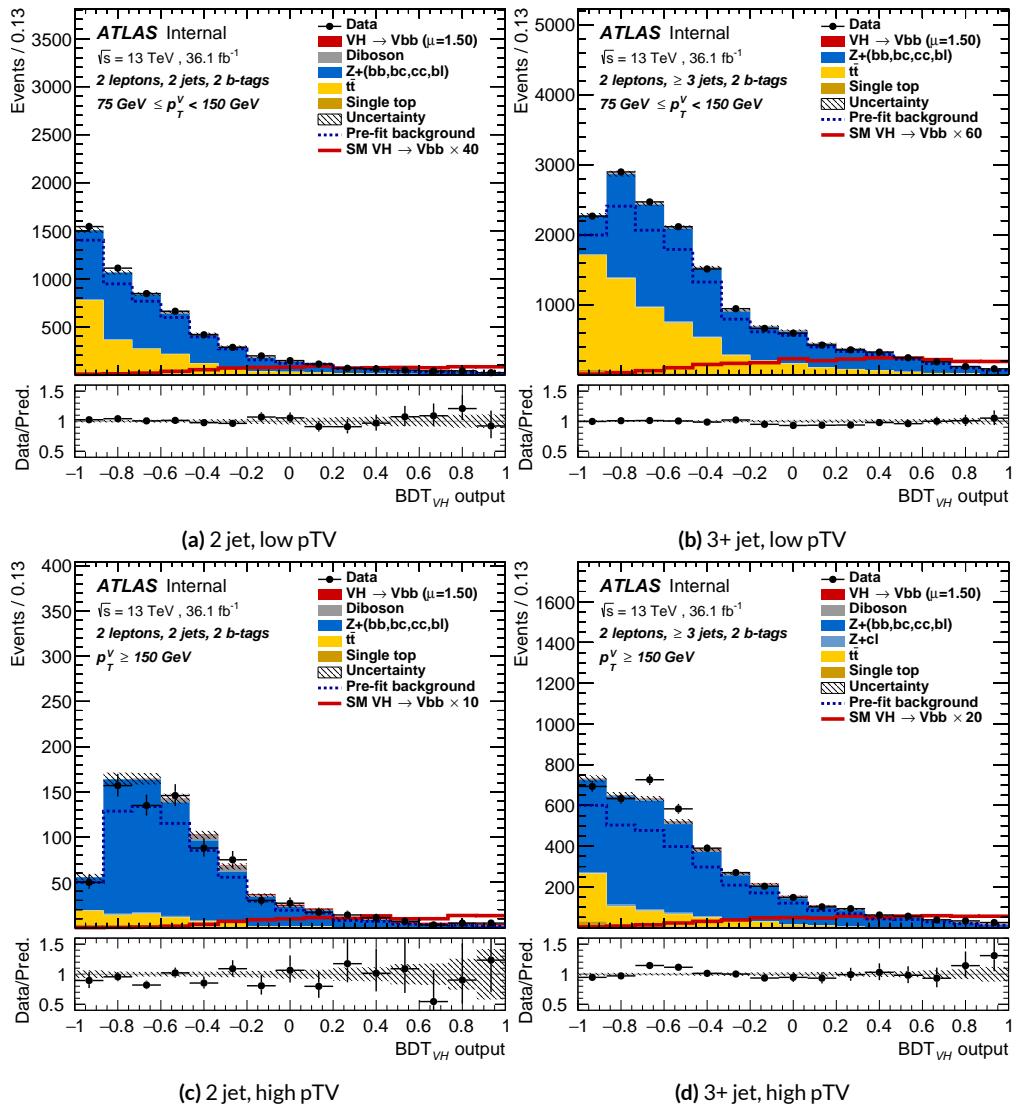


Figure 7.34: Postfit BDT_{VH} plots in the signal region for the RF variable set.

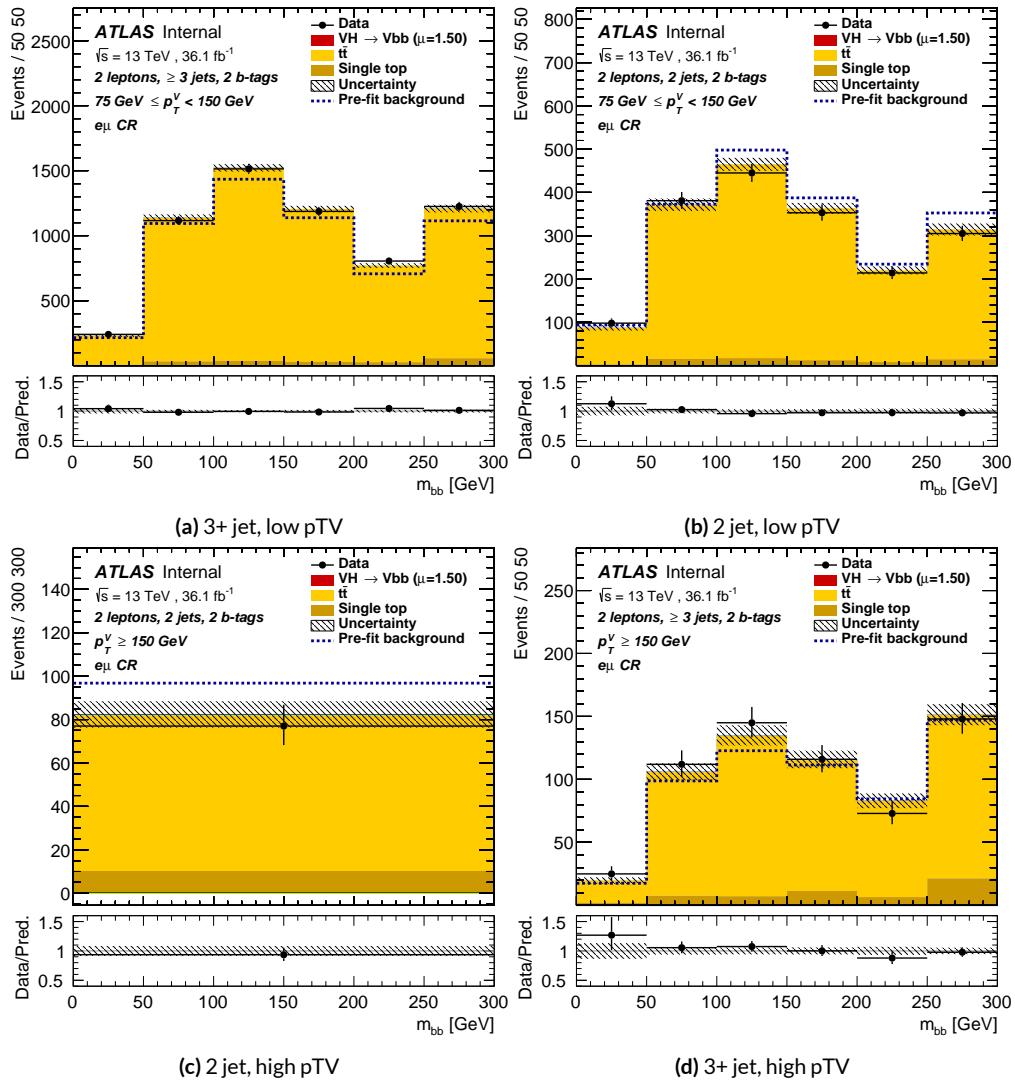


Figure 7.35: Postfit m_{bb} plots in the top $e - \mu$ CR for the RF variable set.

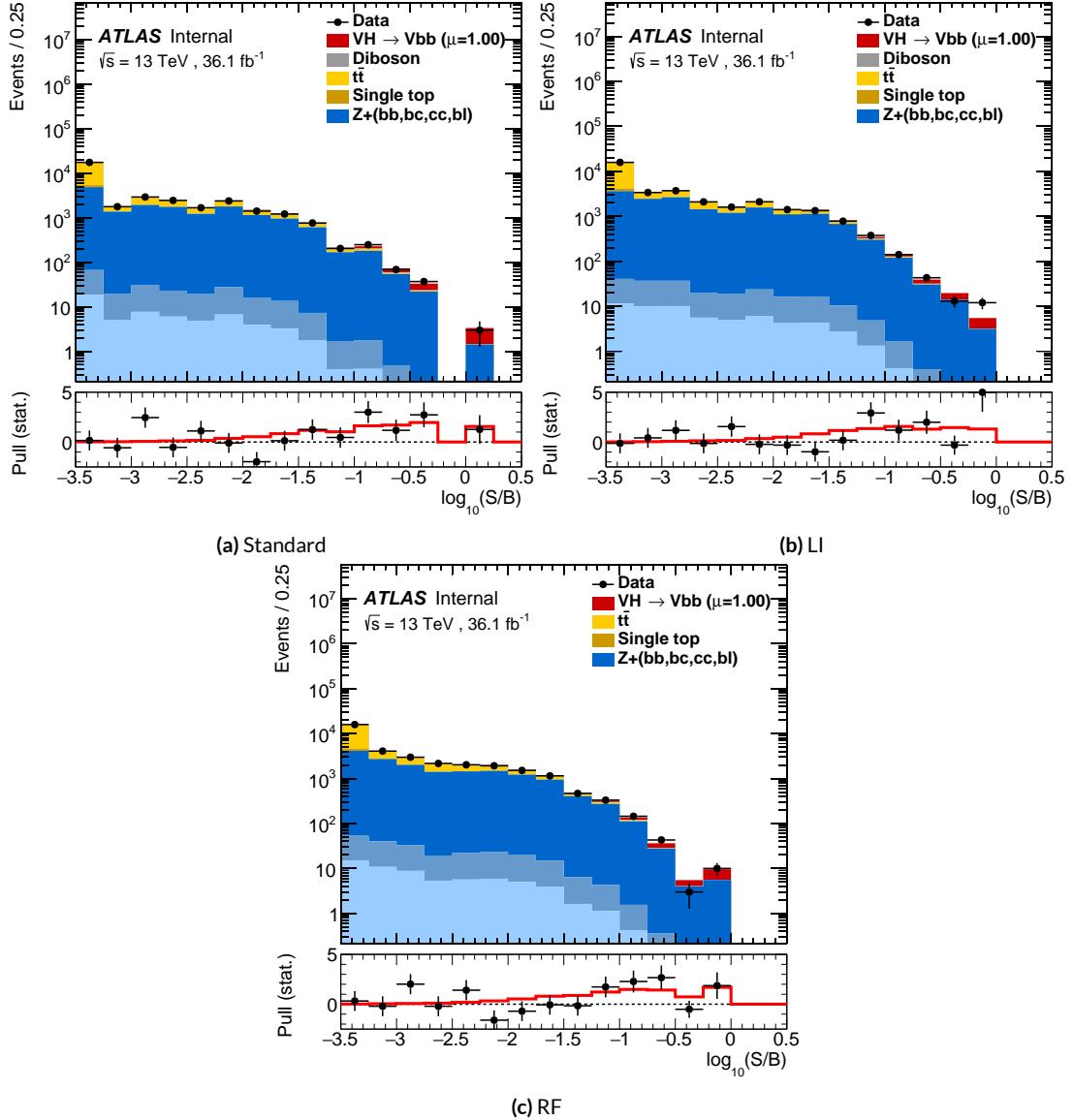


Figure 7.36: Binned S/B plots for the standard (a), LI (b), and RF (c) variable sets. Signal is weighted to $\mu = 1$ for comparison to the SM prediction.

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

Helmuth von Moltke

8

1778

1779

Fit Results

1780 THE RESULTS IN THIS CHAPTER were first reported in³⁷ and describe how the three different fit
1781 models detailed and validated Chapter 7, corresponding to the standard, RF, and LI variable sets
1782 described in Chapter 6 perform on actual VH fits. In particular sensitivities, nuisance parameter
1783 impacts, and signal strengths on expected fits to Asimov datasets and both expected and observed

¹⁷⁸⁴ fits on the actual dataset are compared.

¹⁷⁸⁵ Expected and observed sensitivities for the different variable sets may be found in Table 8.1. The
¹⁷⁸⁶ RF fits feature the highest expected sensitivities, outperforming the standard set by 3.5% and 3.4%
¹⁷⁸⁷ for fits to Asimov and observed datasets, respectively. The LI variable has a lower significance than
¹⁷⁸⁸ both for expected fits to both Asimov and data with a 6.7% (1.7%) significance than the standard set
¹⁷⁸⁹ for the Asimov (observed) dataset. While the fit using standard variables does have a higher observed
¹⁷⁹⁰ significance than both the LI and RF fits, by 2.8% and 8.6%, respectively, these numbers should be
¹⁷⁹¹ viewed in the context of the best fit $\hat{\mu}$ values, discussed below. That is, the standard set may yield the
¹⁷⁹² highest sensitivity for this particular dataset, but this is not necessarily (and likely is not) the case for
¹⁷⁹³ 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.

¹⁷⁹⁴ A summary of fitted signal strengths and errors for both the Asimov (a) and observed (b) datasets
¹⁷⁹⁵ are shown in Figure 8.1.* A summary of error breakdowns is given in Tables 8.2 (Asimov) and 8.3
¹⁷⁹⁶ (observed) for total error, data statistics contributions, total systematic error contributions, and cat-
¹⁷⁹⁷ egories for which the total impact is ≥ 0.1 for the standard fit. As is to be expected for both the
¹⁷⁹⁸ 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.)

¹⁷⁹⁹ 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.

¹⁸⁰⁰ The contribution from systematic uncertainties, however, does vary considerably across the vari-
¹⁸⁰¹ able sets. The Asimov fits are a best case scenario in the sense that, by construction, all NP's are equal
¹⁸⁰² to their predicted values (and so no "penalty" is paid for pulls on Gaussian NP's). The systematics er-
¹⁸⁰³ ror from the LI fit is slightly higher (subpercent) than that from the standard fit, and 4.6% higher er-
¹⁸⁰⁴ ror overall due to differences in data stats. The RF Asimov fit, however, has a 6.5% lower total error

[†]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.

1805 from systematics than the standard Asimov fit (and a 2.2% lower error overall). Moreover, for both
 1806 the LI and RF sets, errors are markedly smaller for the MET and Flavor Tagging categories, with the
 1807 RF fit also featuring a smaller errors on Z +jets modeling; the only notable exception to this trend in
 1808 Asimov fits are the signal systematics.

1809 These trends are more pronounced in the observed fits. As can be seen in Table 8.3, both the LI
 1810 and RF fits have smaller errors from systematic uncertainties, both overall and in all principal cate-
 1811 gories, with the LI and RF fits having 7.5% (3.7%) and 16% (8.8%) lower systematics (total) error on
 1812 $\hat{\mu}$, respectively.

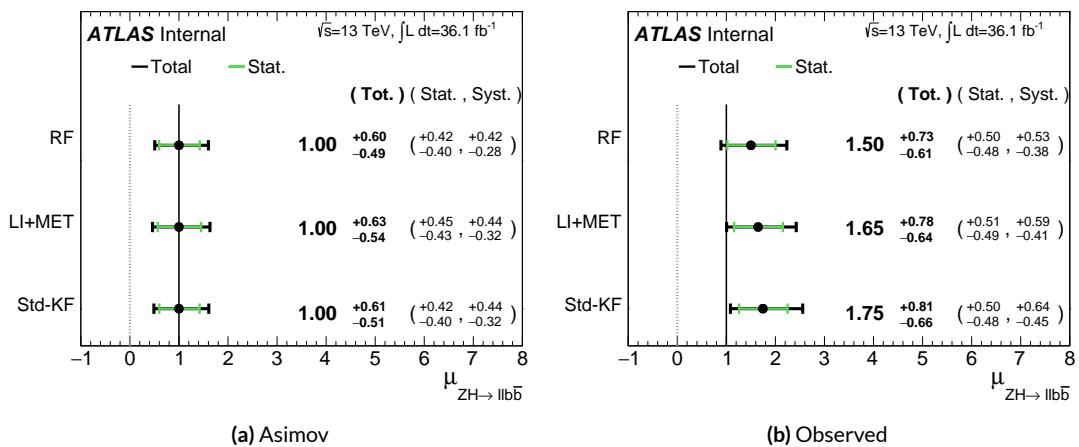


Figure 8.1: μ summary plots for the standard, LI, and RF variable sets. The Asimov case (with $\mu = 1$ by construction) is in (a), and $\hat{\mu}$ best fit values and error summary are in (b).

1813 Studying the performance of the Lorentz Invariant and RestFrames variable sets at both a data
 1814 statistics only context and with the full fit model in the $ZH \rightarrow \ell\ell b\bar{b}$ channel of the $VH(b\bar{b})$ anal-
 1815 ysis suggests that these variables may offer a potential method for better constraining systematic un-
 1816 certainties in $VH(b\bar{b})$ searches as more orthogonal bases in describing the information in collision

1817 events.

1818 The marginally worse performance of the LI and RF variables (7.9% and 6.9%, respectively) with
1819 respect to the standard variable at a stats only level illustrates that neither variable set has greater
1820 intrinsic descriptive power in the absence of systematics in this closed final state. Hence, any gains
1821 from either of these variable sets in a full fit come from improved treatment of systematic uncertain-
1822 ties.

1823 With full systematics, the LI variable set narrows the sensitivity gap somewhat, with lower signif-
1824 icances by 6.7% (1.7%) on expected fits to Asimov (data) and by 2.8% on observed significances. The
1825 RF variable set outperforms the standard set in expected fits with 3.5% (3.4%) higher significance
1826 on Asimov (data), but has an 8.6% lower observed significance, though the observed significances
1827 should be viewed in the context of observed $\hat{\mu}$ values.

1828 Moreover, the LI and RF variable sets generally perform better in the context of the error on μ .
1829 The LI fit is comparable to the standard set on Asimov data and has a 7.5% lower total systematics er-
1830 ror on $\hat{\mu}$ on observed data, while the RF fit is lower in both cases, with systematics error being 6.5%
1831 (16%) lower on Asimov (observed) data.

1832 These figures of merit suggest that both the LI and RF variables are more orthogonal than the
1833 standard variable set used in the fiducial analysis. Moreover, the RF variable set does seem to con-
1834 sistently perform better than the LI set. Furthermore, both variable sets have straightforward exten-
1835 sions to the other lepton channels in the $VH(b\bar{b})$ analysis. The magnitude of any gain from the
1836 more sophisticated treatment of E_T^{miss} in these extensions is beyond the scope of these studies, but
1837 the performance in this closed final state do suggest that there is some value to be had in these non-

¹⁸³⁸ standard descriptions independent of these considerations.

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

Isaac Newton

1839

9

1840

Measurement Combinations

1841 WHILE THE DISCUSSION thus far has focused on improvements looking towards future in just the
1842 $ZH \rightarrow \ell\ell b\bar{b}$ channel, any actual result for SM $VH(b\bar{b})$ combines all channels and all available
1843 datasets. Using additional channels at a given center of mass energy is straightforward since the fit
1844 model is designed with this combination in mind. Combining dataset results (known as “workspaces”)

1845 from different center of mass energies is not so simple an exercise since both the underlying physics
1846 (and its associated modeling) and the treatment of key experimental considerations, like flavor tag-
1847 ging, and their associated systematics change from dataset to dataset. A combined fit model must
1848 take these considerations into account, and the formulation of the fit model combining the Run
1849 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 ($\sqrt{s} = 13$
1850 TeV with) SM $VH(b\bar{b})$ results is the topic of Section 9.1. Its results, as reported in ⁴², are given in
1851 9.2.

1852 9.1 THE COMBINED FIT MODEL

1853 It is clear the signal strength parameter of interest should be fully correlated among the different
1854 datasets. Some signal modeling systematics were left unchanged from Run 1 through Run 2 and/or
1855 were designed to be explicitly correlated. Beyond these two special cases, it is not immediately clear
1856 what level of correlation should be imposed. The general methodology for settling upon a correla-
1857 tion scheme is as follows:

- 1858 1. Identify which NP categories have significant impacts on μ
- 1859 2. Of these NP's, identify which have one-to-one correspondences or established correlation
1860 schemes among \sqrt{s} values
- 1861 3. Test whether correlation has a sizeable impact on expected fit quantities

1862 The only two sizeable experimental NP categories are jet energy scale (JES) and flavor tagging sys-
1863 tematics. Correlation schemes of varying degrees of completeness exist for these categories, so ex-
1864 plicit NP correlations can be tested for these two categories. As these studies were conducted before

1865 unblinding, “sizeable impact” was judged by comparing fit results (sensitivities, pull comparisons,
1866 and breakdowns) on combined workspaces using the unblinded and public $\mu = 0.51$ result for
1867 Run 1 and Asimov data for the Run 2 result. These are treated in Sections 9.1.1 and 9.1.2. Modeling
1868 systematics require a slightly different treatment, and are explored in 9.1.3.

1869 As noted in Chapter 7 when looking at pull comparison plots for combined workspaces, the error
1870 bars in these plots are calculated using a simultaneous HESSE matrix inversion, which can fail to give
1871 sensible values for high dimensional models (the combined workspaces have well over 500 NP’s).
1872 This is not true of the nuisance parameter ranking plots, which use a MINOS based approach to test
1873 the effect of each NP individually. This is much slower but much more rigorous, which is why only
1874 ranking plots appear outside of supporting material and pull comparisons are considered “diagno-
1875 tic” plots.

1876 9.1.1 JET ENERGY SCALE SYSTEMATICS

1877 Fortunately for the case of jet energy scale systematics, the JetEtMiss group provides two recom-
1878 mended “strong” and “weak” correlation schemes between Run 1 and Run 2. These were used as
1879 a point of departure for the JES combination correlation scheme. However, the JES NP’s in both
1880 the Run 1 and Run 2 workspaces are a reduced set of NP’s, with some 56 (75) NP’s reduced to 6 (8)
1881 for Run 1 (2). In order to restore the full set of JES NP’s, the effective NP’s in each workspace are un-
1882 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:

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

- 1884 where *eff* and *unf* are for effective and unfolded NP's, respectively, the A_{ij} 's are scalar coefficients
 1885 taken from raw maps, and $|NP_{j,unf}|$ are the amplitudes of the unfolded NP's. The raw A_{ij} and scaled
 1886 maps for Run 1 and Run 2 may be found in Figure 9.1

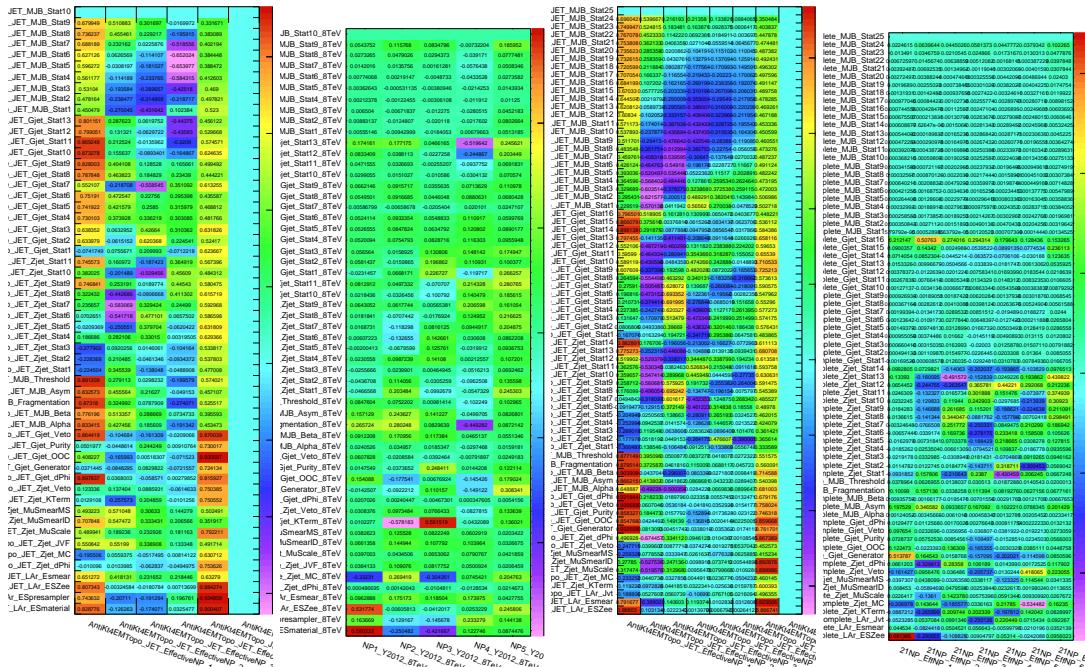


Figure 9.1: The raw and scaled coefficients for unfolding Run 1 (a and b) and Run 2 (c and d), respectively

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

seen in Tables 9.1– 9.4.

	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 ₂ Unfold	R ₂ 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.

1889 It was also found that fit sensitivities and breakdowns were similarly indifferent to the use of ei-
 1890 ther 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.

1891 Comparisons of top ranked nuisance parameters in Figures 9.2–9.4 and for the complete JES pull

1892 comparisons in Figures 9.5–9.8 also show very little difference with respect to correlation scheme
 1893 (except obviously for the number of JES NP's). Constrained pulls in pull comparisons should once
 again be taken as a shortcoming of HESSE and not the fit model.

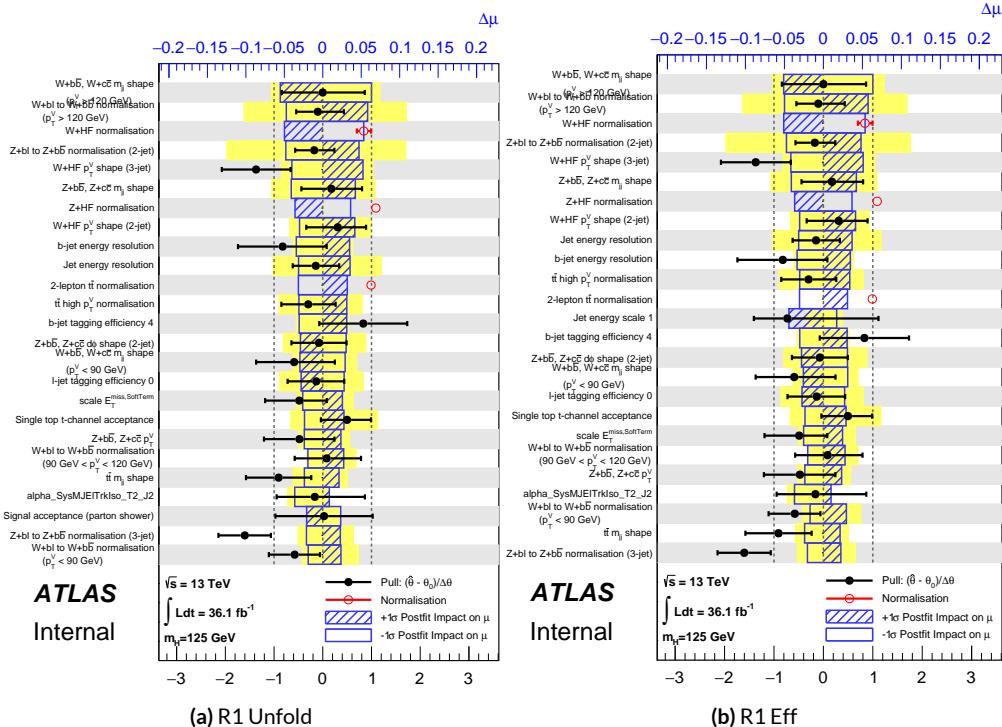


Figure 9.2: Ranks for the effective and unfolded JES NP Run1 combined workspaces.

1894 As a result of these studies, the weak JES correlation scheme with uncorrelated effective JES NP's
 1895 (i.e. just the b -jet energy scale NP) has been chosen as the treatment of JES in the Run 1 + Run 2
 1897 combined fit.

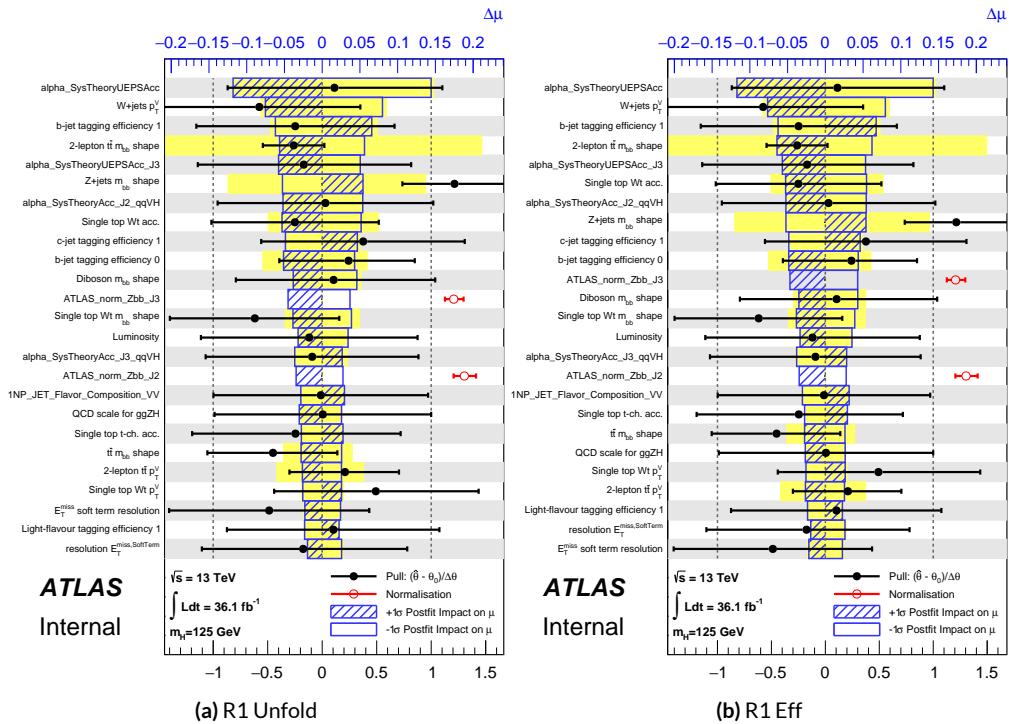


Figure 9.3: Ranks for the effective and unfolded JES NP Run2 combined workspaces.

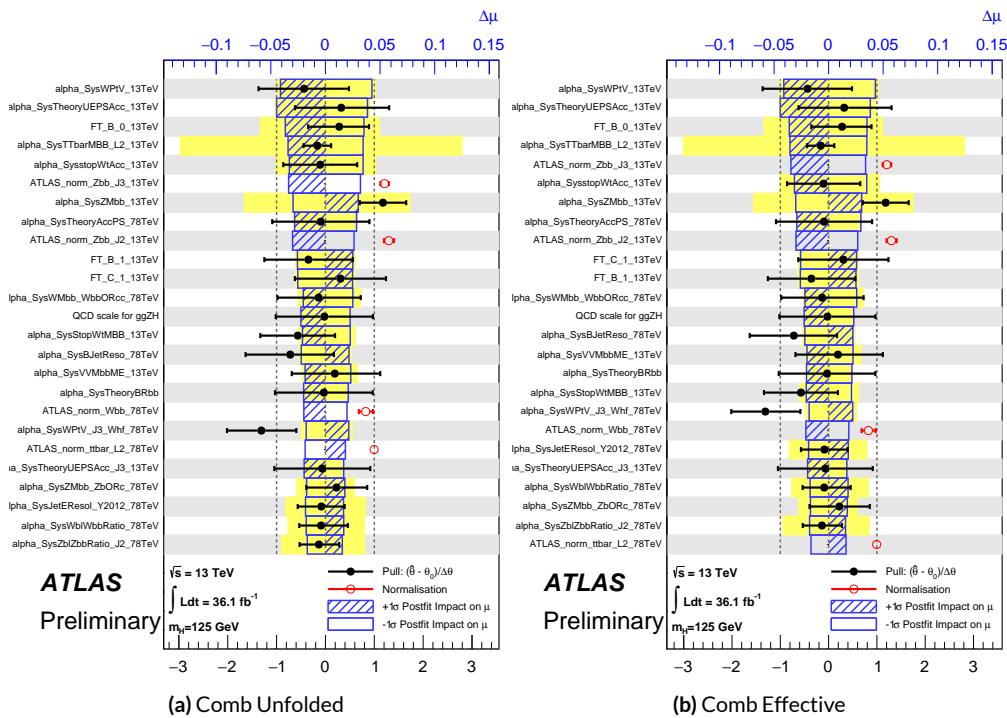


Figure 9.4: Ranks for the effective and unfolded JES NP Run1+Run2 combined workspaces.

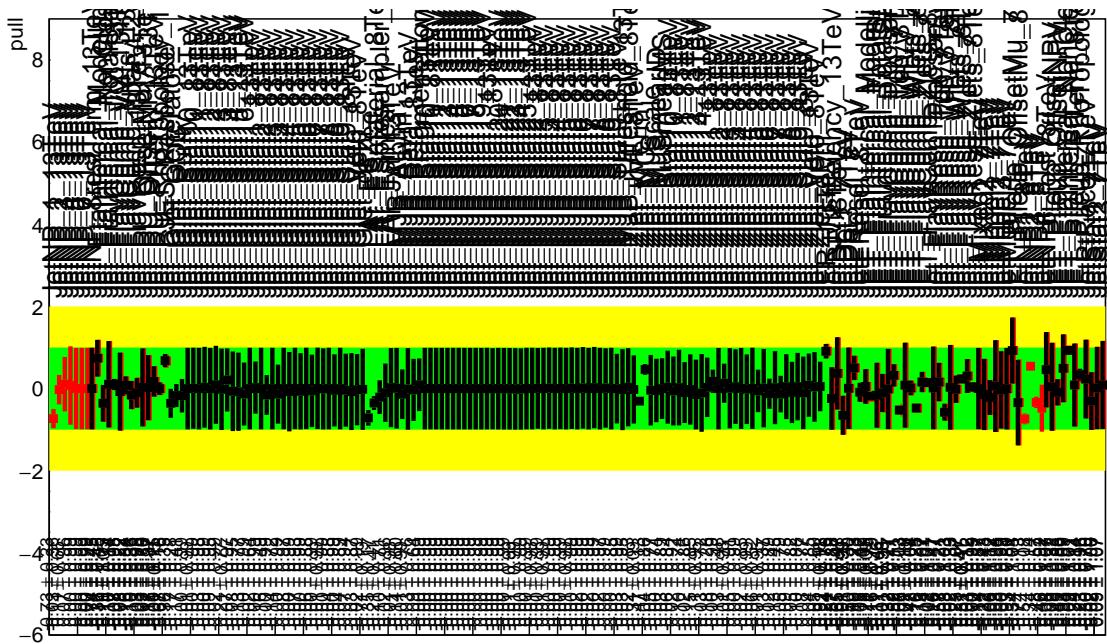


Figure 9.5: Pull Comparisons: jesu---Jet Comb Unfold, Comb Eff, Strong Unfold, Strong Eff

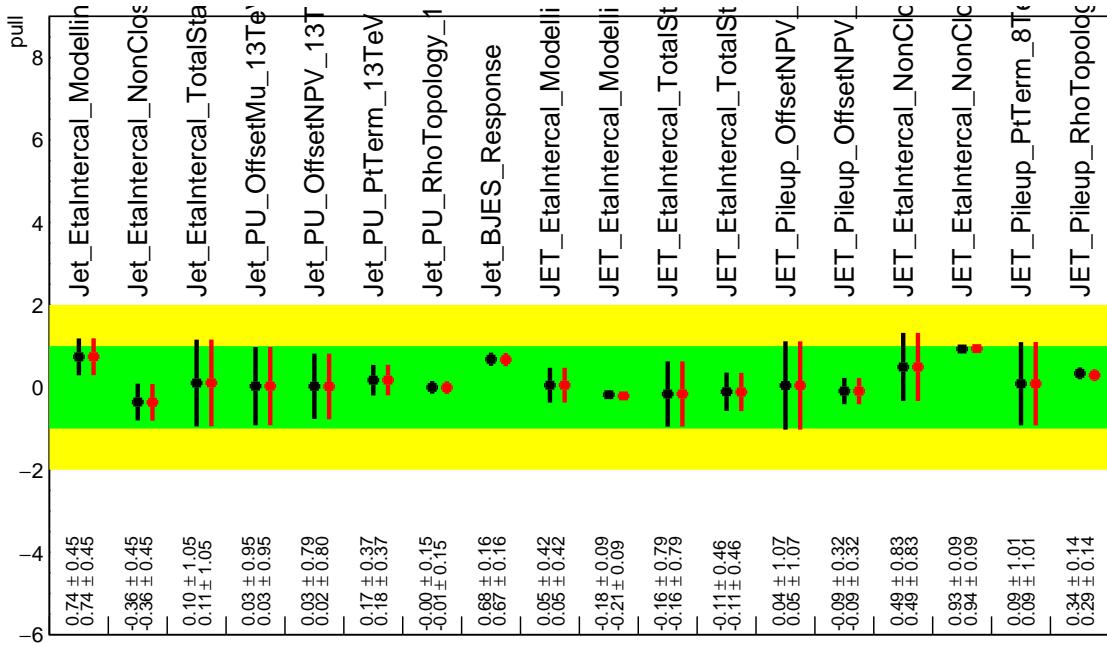


Figure 9.6: Pull Comparisons: jesu---JetMatched Comb Unfold, Comb Eff, Strong Unfold, Strong Eff

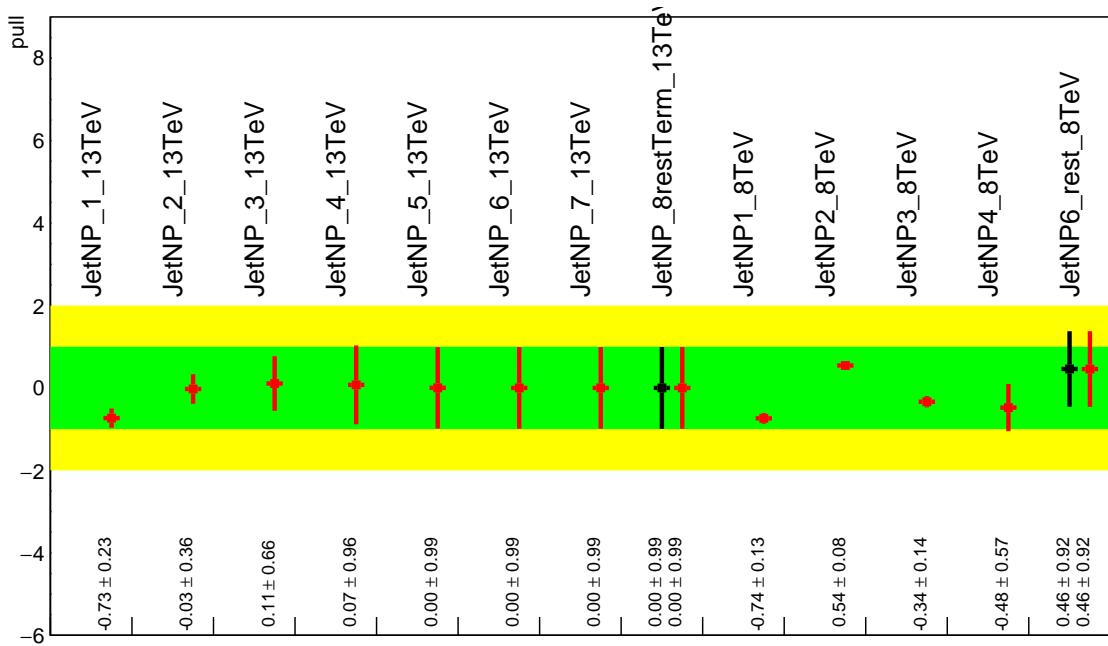


Figure 9.7: Pull Comparisons: jesu---JetEff Comb Unfold, **Comb Eff**, Strong Unfold, **Strong Eff**

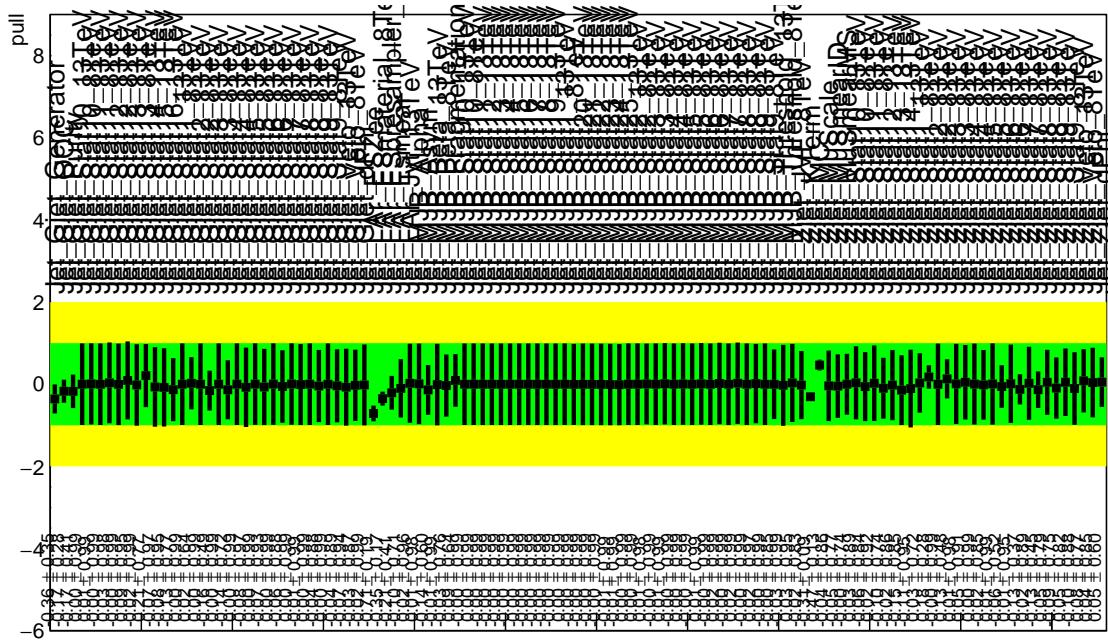


Figure 9.8: Pull Comparisons: jesu---JetUnfold Comb Unfold, **Comb Eff**, Strong Unfold, **Strong Eff**

1898 9.1.2 FLAVOR TAGGING

1899 Unfortunately, the ATLAS Flavor Tagging group did not provide any recommendations for corre-
1900 lating Run 1 and Run 2 NP's, though given the high ranking of these NP's in the Run 2, result, per-
1901 forming at least some studies was deemed crucial. Nevertheless, great improvements and changes to
1902 the treatment of flavor tagging between Run 1 and Run 2 does weaken the argument for any strong
1903 flavor tagging correlation scheme.

1904 Given that c -tagging changed significantly between Run 1 and Run 2 and that light tagging NP's
1905 are very lowly ranked, these sets of NP's are left uncorrelated. Moreover, the change in the physical
1906 meaning of the effective b -tagging NP's means a full correlation of such NP's (insomuch as they exist
1907 in each result) is one of limited utility. Hence, it was decided to leave flavor tagging NP's uncorre-
1908 lated. However, since the meaning of the leading b -tagging NP's is approximately constant across
1909 years and since Run 2 b -tagging NP's are very highly ranked in both the Run 2 only and combined
1910 fits, tests correlating these NP's were conducted, the results of which can be seen below. It should be
1911 noted that the leading B NP at 8 TeV, `SysBTagB0Effic_Y2012_8TeV`, has an opposite effect on $t\bar{t}$
1912 normalization than the 7 and 13 TeV NP's, and so must be flipped using a similar strategy as for JES
1913 unfolding. Initial studies of flavor tagging correlations did not flip this NP, and so results for this
1914 scheme (labeled "Bo 8TeV Not Flipped") have also been included for comparison.

1915 It is clear from these results that correlating the leading effective Eigen NP associated with b 's can
1916 have a noticeable effect on final fit results and that the 8 TeV Bo NP is the most important compo-
1917 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}_{ll} for a combination featuring the weak JES scheme and a combination with the weak JES scheme + leading b NP's correlated.

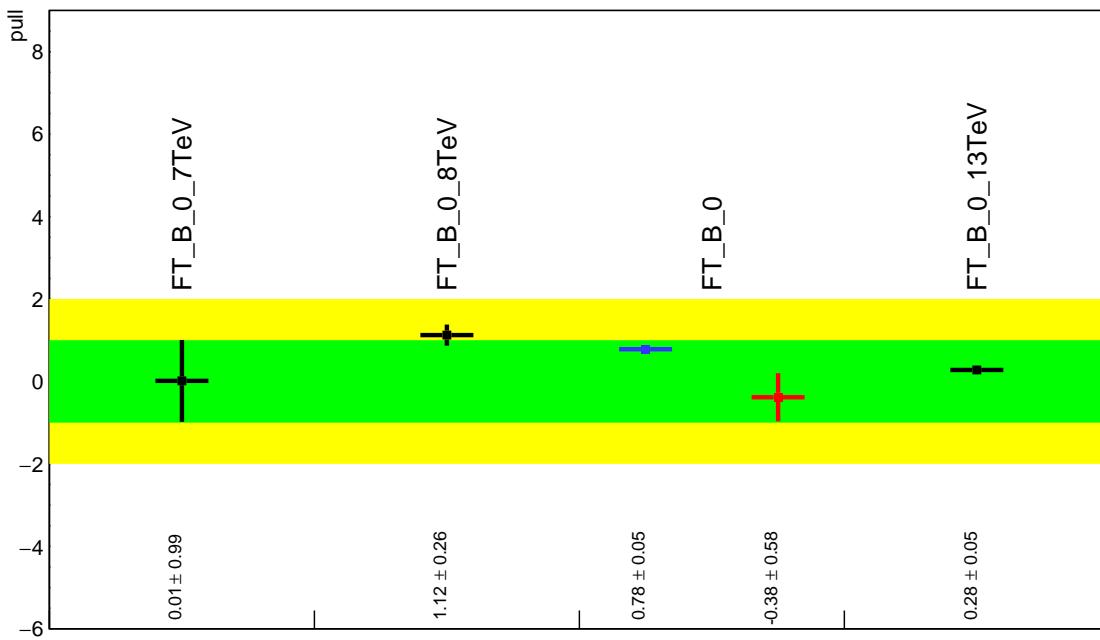


Figure 9.9: Pull Comparisons: btag-b---BTagBO Comb Eff, BTag BO

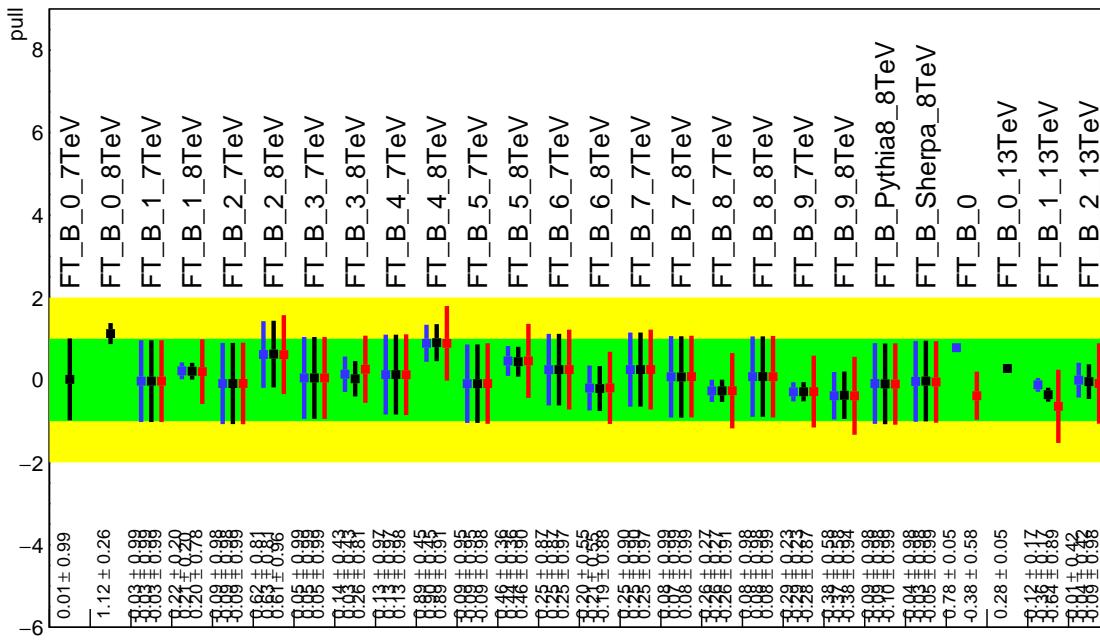


Figure 9.10: Pull Comparisons: btag-b---BTagB Comb Eff, BTag BO

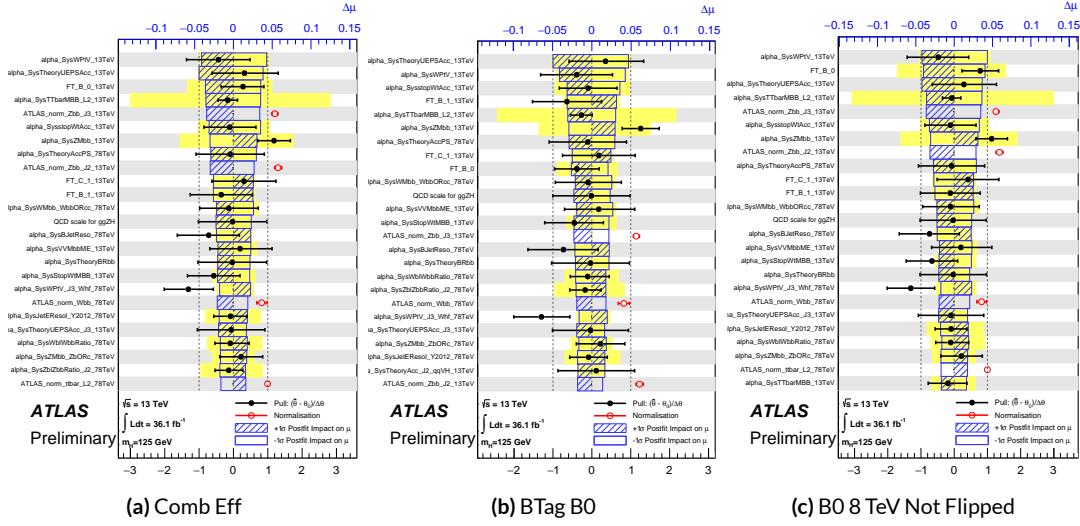


Figure 9.11: NP rankings for a combination featuring the weak JES scheme and a combination with the weak JES scheme + leading b NP's correlated.

1918 bined nuisance parameter since it is the only result to make use of both pseudocontinuous tagging-
 1919 based and $1 b$ -tag regions into the final fit, implicitly yielding much more information about b 's. The
 1920 13 TeV fit has neither of these regions. What is less clear is whether there are sufficient grounds for
 1921 implementing this correlation (i.e. does the correspondence of these NP's across years warrant a full
 1922 correlation). While there are no current plans to do so, this matter warrants careful scrutiny if Run 1
 1923 is to be combined with future results.

1924 9.1.3 MODELING SYSTEMATICS

1925 Another principal systematic category is modeling uncertainties. The effect of correlating groups
 1926 of systematics was estimated using the same strategy employed by the ATLAS/CMS SM $VH(b\bar{b})$
 1927 combination for Run 1. This extrapolation can be used to estimate the impact of correlations on
 1928 the estimated signal strength, the total error on the signal strength, and the χ^2 of the result. The

¹⁹²⁹ impact of such correlations is no more than a few percent effect, as the following tables demonstrate,
¹⁹³⁰ beginning with the category with the greatest shift, W+jets modeling, in Table 9.9.

	$ \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

¹⁹³¹ 9.1.4 FINAL CORRELATION SCHEME

¹⁹³² The final Run 1 + Run 2 correlation scheme is shown in Table 9.10. As detailed above, neither JES
¹⁹³³ nor modeling systematics had any demonstrable effect on combined fit results. Hence, only signal
¹⁹³⁴ NP's and the b -jet energy scale are correlated (the weak JES scheme without unfolding). While the
¹⁹³⁵ effect of flavor tagging correlations is less clear, the result physical arguments for correlation are less
¹⁹³⁶ strong; the size of effect was discovered rather late in the analysis process; and no nuisance parameter
¹⁹³⁷ unfolding maps exist for flavor tagging as they do for JES, so it was decided to leave these uncorre-
¹⁹³⁸ lated 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.

¹⁹³⁹ **9.2 COMBINED FIT RESULTS**

¹⁹⁴⁰ **9.2.1 COMBINED FIT MODEL VALIDATION**

¹⁹⁴¹ Before moving onto the final results, we present the rest of the validation for the Run 1 + Run 2
¹⁹⁴² combined fits, beginning with impacts of ranked individual nuisance parameters in Figure 9.12 and
¹⁹⁴³ for all nuisance parameter categories in Table 9.11. Both of these sets of results point to the most im-
¹⁹⁴⁴ portant nuisance parameters being signal systematics, b -tagging, and V +jets modeling systematics,
¹⁹⁴⁵ with Run 2 NP's generally being higher ranked. That some NP's are strongly pulled is not unusual
¹⁹⁴⁶ as the fit model has so many NP's; V +jets modeling in particular has been historically difficult.

¹⁹⁴⁷ In addition to looking at the behaviors of nuisance parameters to gauge fit model performance
¹⁹⁴⁸ and stability, fits are conducted using multiple parameters of interest. Typical divisions are Run 1
¹⁹⁴⁹ vs. Run 2, lepton channels, and WH vs ZH . As mentioned in Chapter 7, the profile likelihood test
¹⁹⁵⁰ statistic given in Equation 7.2 is, in the limit of large sample statistics, a χ^2 distribution with degrees
¹⁹⁵¹ of freedom equal to the number of parameters of interest plus number of nuisance parameters.
¹⁹⁵² Thus, changing the number of interest parameters and leaving the rest of the fit model unchanged

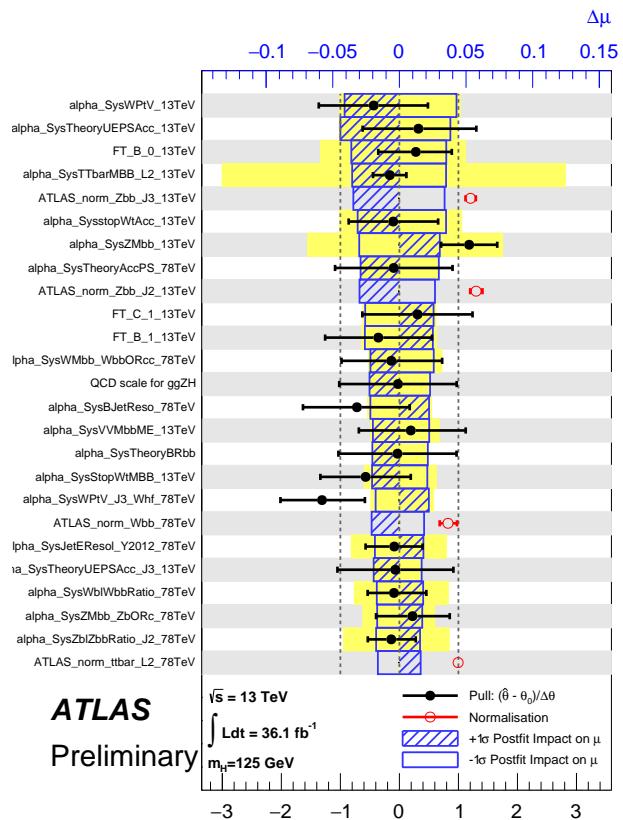


Figure 9.12: Ranked nuisance parameters for the Run1+Run2 combination.

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.

¹⁹⁵³ means that the difference between the nominal fit and a fit with more parameters of interest ought
¹⁹⁵⁴ to also be distributed as a χ^2 distribution with degrees of freedom equivalent to the number of extra
¹⁹⁵⁵ parameters of interest. This difference can then be interpreted as a compatibility between the two
¹⁹⁵⁶ results using the standard tables for this distribution, giving another gauge of fit performance. These
¹⁹⁵⁷ are shown 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.

¹⁹⁵⁸ The low compatabilities associated with treating the lepton channels as separate parameters of
¹⁹⁵⁹ interest are a symptom of the low signal strengths associated with the Run 1 0-lepton channel, in par-
¹⁹⁶⁰ ticular the 7 TeV result. Given the relatively small amount of data associated with the 7 TeV result,
¹⁹⁶¹ this should not be a cause for alarm. Signal strength summary plots for the fits treating Run 1 and
¹⁹⁶² Run 2 separately are shown in Figures 9.13-9.15, where the effect of the Run 1 parameters can be seen
¹⁹⁶³ graphically.

¹⁹⁶⁴ 9.2.2 FINAL RESULTS

¹⁹⁶⁵ The combined results yields an observed (expected) significance of 3.57 (4.00) and an observed (ex-
¹⁹⁶⁶ pected) limit of 1.37 ($0.510^{+0.200}_{-0.143}$), with a signal strength of $\hat{\mu} = 0.898^{+0.278}_{-0.261}$.

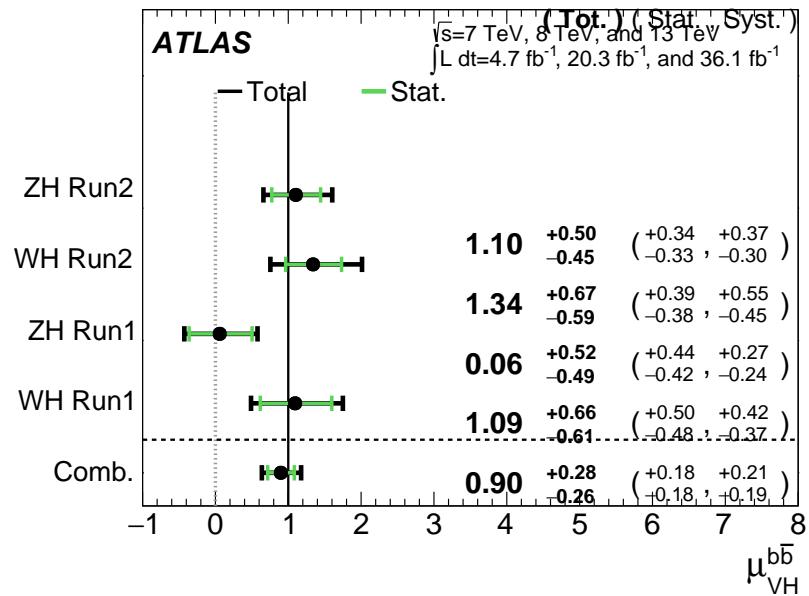


Figure 9.13: $\hat{\mu}$ summary plot for a four parameter of interest fit.

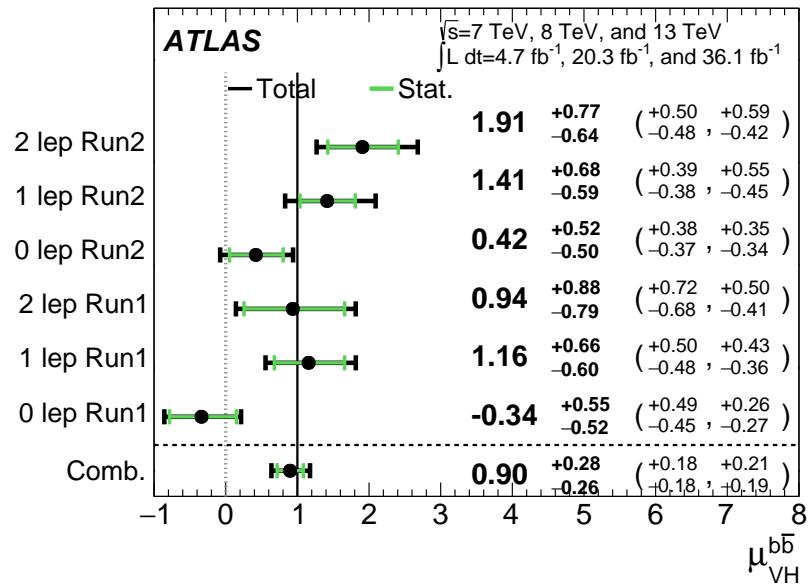


Figure 9.14: $\hat{\mu}$ summary plot for a six parameter of interest fit.

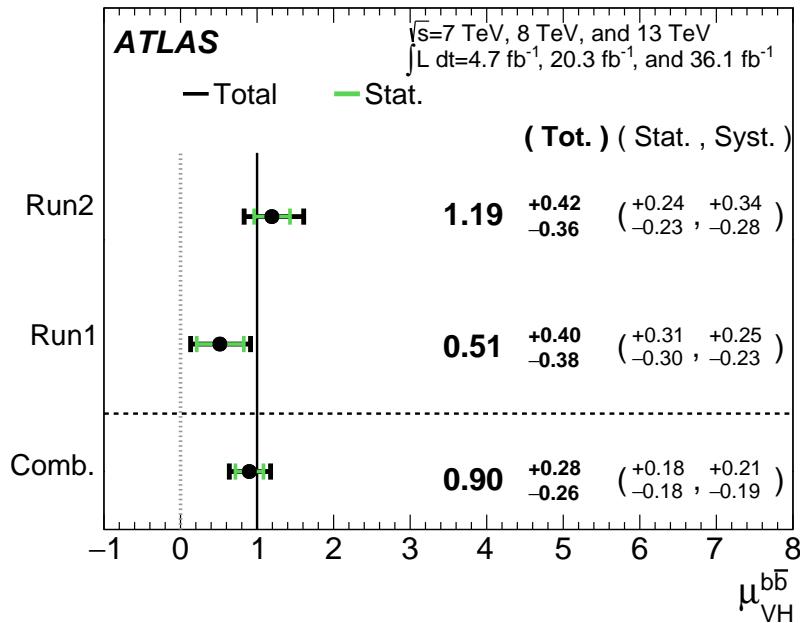


Figure 9.15: $\hat{\mu}$ summary plot for a two parameter of interest (Run 1 and Run 2) values.

1967 The two and three parameter of interest fit signal strength summary plots, as well as a summary
 1968 of the historical values of the 7, 8, and 13 TeV results may be found in Figures 9.16-9.18. The main
 1969 results for Run 1, Run 2, and the combination may be found in Table 9.13. These results were collec-
 1970 tively noted as the first ever experimental evidence for SM $VH(b\bar{b})$ in ⁴².

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

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

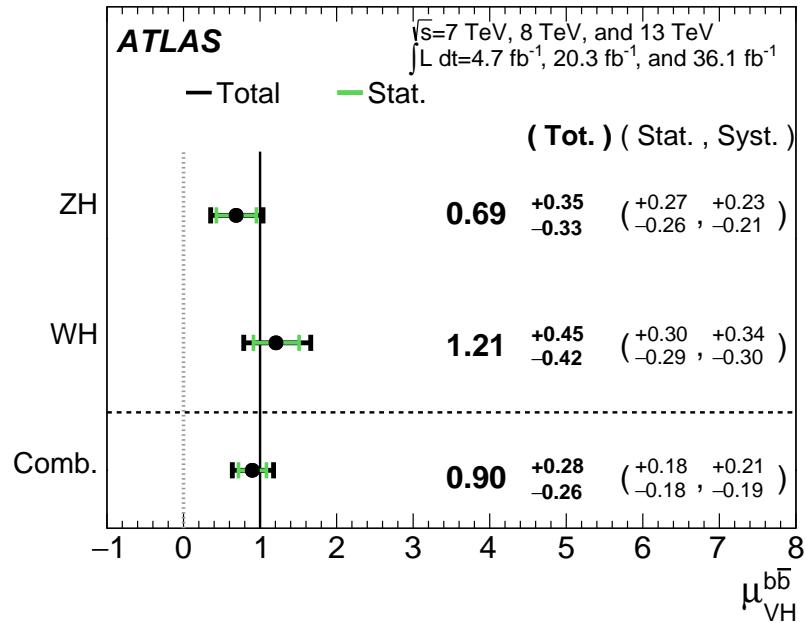


Figure 9.16: $\hat{\mu}$ summary plot for a two parameter of interest fit.

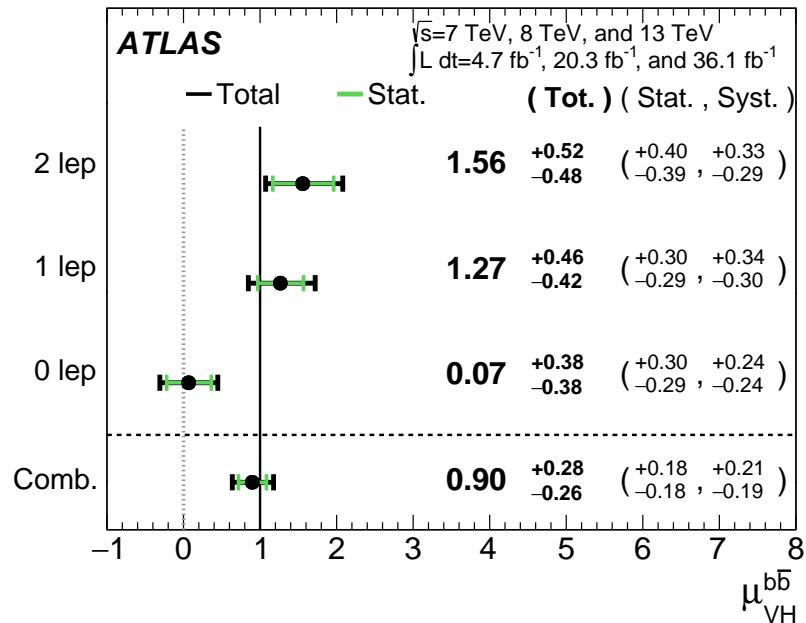


Figure 9.17: $\hat{\mu}$ summary plot for a three parameter of interest fit.

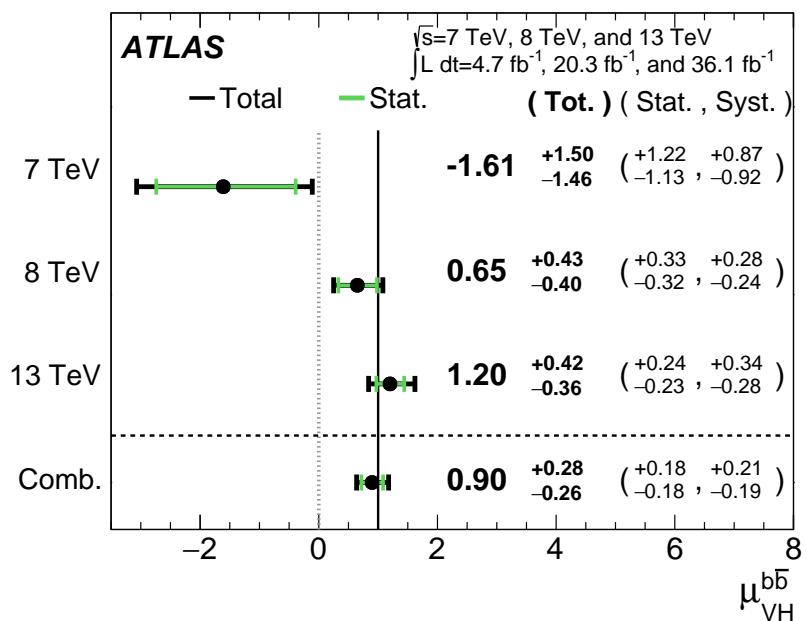


Figure 9.18: $\hat{\mu}$ summary plot for different \sqrt{s} values.

Vanitas vanitatum, omnis vanitas

Ecclesiastes 1:2

10

1971

1972

Closing Thoughts

1973 SINCE BOTH THE LHC and ATLAS are performing very well, it is only a matter of time before the

1974 evidence for SM $VH(b\bar{b})$ passes the 5 Gaussian standard deviation threshold necessary for discovery.

1975 Official discovery may come less than a year after reports of first evidence and may not even require

1976 a combination with the Run 1 result, depending on the latter two years of ATLAS Run 2 data (2017

₁₉₇₇ and 2018).

₁₉₇₈ It is entirely natural to ask, then, how essential the techniques and results described in this thesis
₁₉₇₉ will prove to be moving forward. Neither the LI/RF multivariate techniques nor combination with
₁₉₈₀ Run 1 datasets and their accompanying low signal strength values are necessary for discovery, and
₁₉₈₁ the latter may not even be essential to timely^{*} discovery of SM $VH(b\bar{b})$. Nevertheless, both sets of
₁₉₈₂ results hold great potential as key parts of a concerted ensemble of efforts towards precision Higgs
₁₉₈₃ physics.

₁₉₈₄ With the perhaps final major center of mass energy increase at the energy frontier ever complete,
₁₉₈₅ analyses must rely on increased integrated luminosity. Hence, it is becoming increasingly likely that
₁₉₈₆ any new fundamental physics at colliders will require the use of results of systematics limited analy-
₁₉₈₇ ses. This is the regime where the techniques described in this thesis will be most useful.

₁₉₈₈ As the LHC and its experiments undergo successive stages of upgrades and operate in evermore
₁₉₈₉ extreme environments, the statistical fit models used to describe LHC data will continue to evolve in
₁₉₉₀ complexity and diverge from their predecessors. The techniques described in Chapter 9 will become
₁₉₉₁ increasingly more vital to producing the best physics results possible. The improvement in precision
₁₉₉₂ from $\hat{\mu}_{VH} = 1.20^{+0.24}_{-0.23}(\text{stat.})^{+0.34}_{-0.28}(\text{syst.})$ to $\hat{\mu}_{VH} = 0.90^{+0.18}_{-0.18}(\text{stat.})^{+0.21}_{-0.19}(\text{syst.})$ is just the begin-
₁₉₉₃ ning.

₁₉₉₄ The best methods for reduction of systematic uncertainties will naturally depend in part on the
₁₉₉₅ state of the art for both fundamental physics process and detector modeling, but techniques that
₁₉₉₆ can reduce systematic uncertainties independent of fit model, dataset, and physics process provide

^{*}i.e. before or coincident with CMS

1997 a promising avenue forward. The improvements in systematic uncertainties using the Lorentz In-
 1998 variant and RestFrames variable techniques in the $ZH \rightarrow \ell\ell b\bar{b}$ analysis, summarized in Table 10.1,
 1999 show that a smarter and more orthogonal decomposition of information in a collision event pro-
 2000 vides benefits independent of any clever treatment of \vec{E}_T^{miss} (which both schemes also provide). Both
 2001 techniques are readily extendible to other analysis channels, with the RestFrames concept demon-
 2002 strating stronger performance and greater flexibility to nearly completely generic final states.

	Standard	LI	RF
$\hat{\mu}$	$1.75^{+0.50, 0.64}_{-0.48, 0.45})$	$1.65^{+0.51, 0.59}_{-0.49, 0.41}$	$1.50^{+0.50, 0.53}_{-0.48, 0.36}$
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. Uncertainties on $\hat{\mu}$ are quoted stat., syst. 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.

2003 Critical work remains to be done refining and extending the treatment of both the LI and RF

2004 techniques in $VH(b\bar{b})$ analyses and their fit models, and completely independent techniques, like
 2005 the use of multiple event interpretations addressed in Appendix B promise further improvements
 2006 still.

2007 No one can say for certain what the future of the energy frontier of experimental particle physics
 2008 may hold, but more nuanced treatments of the information in collision events born of meaningful
 2009 physical insight are sure to light the way.

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

Conventional Wisdom

A

2010

2011 Micromegas Trigger Processor Simulation

2012 IN ORDER TO PRESERVE key physics functionality by maintaining the ability to trigger on low p_T
2013 muons, the Phase I Upgrade to ATLAS includes a New Small Wheel (NSW) that will supply muon
2014 track segments to the Level 1 trigger. These NSW trigger segments will combine segments from the
2015 sTGC and Micromegas (MM) trigger processors (TP). This note will focus in particular on the algo-

2016 rithm for the MMTP, described in detail with initial studies in ⁴¹. The goal of this note is to describe
2017 the MMTP algorithm performance under a variety of algorithm settings with both nominal and
2018 misaligned chamber positions, as well as addressing a number of performance issues.

2019 This note is organized as follows: the algorithm and its outputs are briefly described in Section
2020 A.1; Monte Carlo samples used are in Section B.1; nominal algorithm performance and certain quan-
2021 tities of interest are described in Section A.3; algorithm performance under misalignment, misalign-
2022 ment corrections, and corrected performance are shown in Section A.9; and conclusions are pre-
2023 sented in Section A.24.

2024 **A.1 ALGORITHM OVERVIEW**

2025 The MMTP algorithm is shown schematically in Figure A.1, taken from ⁴¹, where a more detailed
2026 description may be found. The algorithm begins by reading in hits, which are converted to slopes.
2027 These slopes are calculated under the assumption that the hit originates from the IP; slopes calcu-
2028 lated under this assumption are denoted by a superscript g for global in order to distinguish them
2029 from local slopes calculated using only hits in the wedge. In the algorithm simulation, events are
2030 screened at truth level to make sure they pass certain requirements. The track's truth-level coor-
2031 dinates must place it with the wedge since some generated tracks do not reach the wedge. These
2032 hits are stored in a buffer two bunch crossings (BCs) in time deep that separates the wedge into so-
2033 called "slope-roads." If any given slope-road has sufficient hits to pass what is known as a coinci-
2034 dence threshold, a fit proceeds. A coincidence threshold is a requirement for an event expressed as
2035 $aX+bUV$, which means that an slope-road must have at least a hits in horizontal (X) planes and at

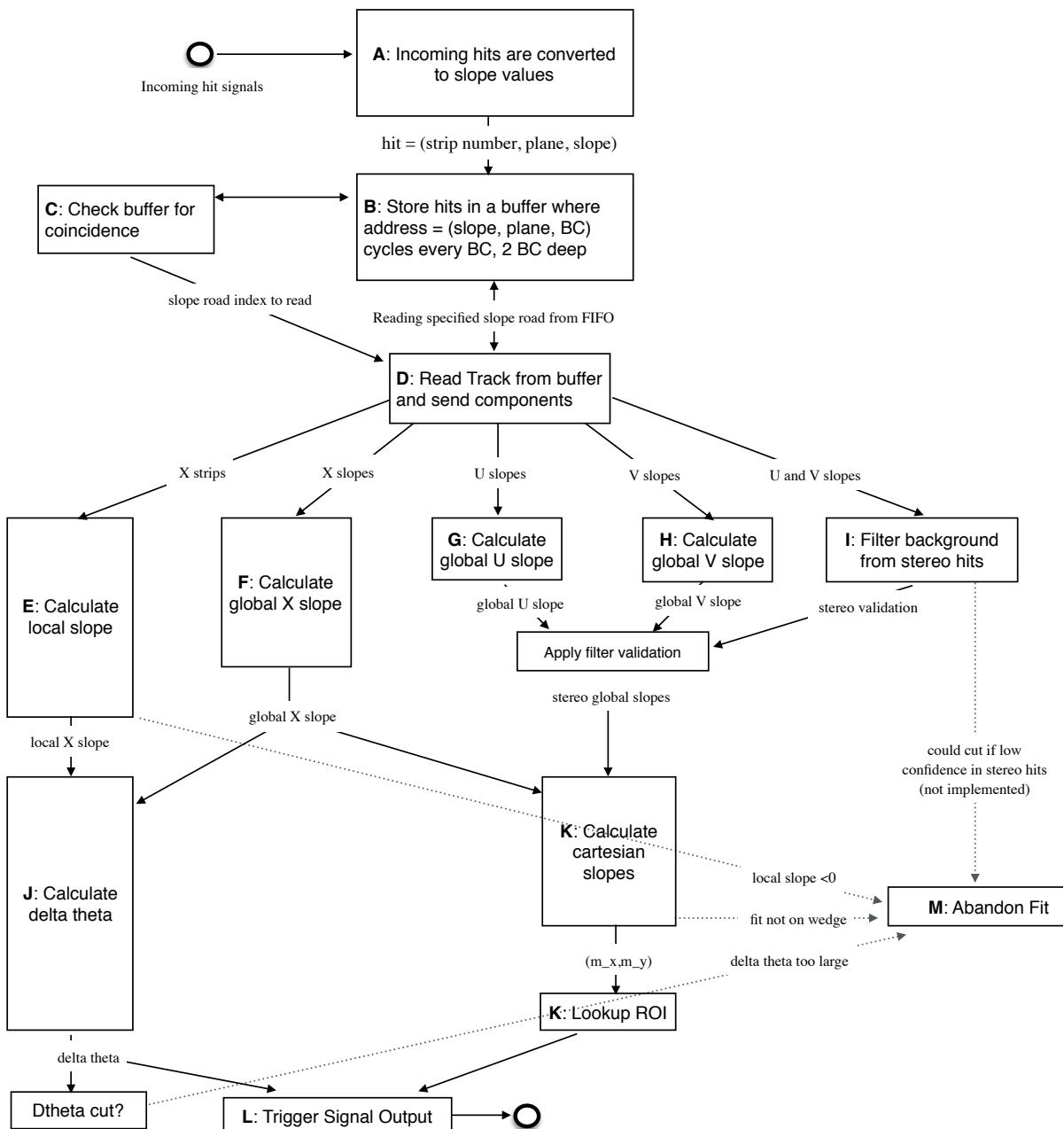


Figure A.1: A flow chart describing the algorithm steps, taken from⁴¹.

2036 least b hits in stereo (U or V (corresponding to positive and negative stereo rotations)) planes. For
 2037 coincidence thresholds with a $2X$ hit requirement there is the extra requirement that, in the case of
 2038 only $2X$ hits, one be on each quadruplet in order to ensure an adequate lever arm for the $\Delta\theta$ calcu-
 2039 lation. Note that less stringent (lower hit) coincidence thresholds are inclusive; i.e. a slope-road pass-
 2040 ing a $4X+4UV$ cut automatically passes $2X+1UV$. The coincidence threshold, size of the slope-roads
 2041 (denoted b), and the number of slope-roads into which each horizontal and stereo hits get written
 2042 centered upon their nominal value are configurable parameters of the algorithm.

2043 An individual hit's slope is calculated as shown in Equation A.1, where y_{base} is the local y coordi-
 2044 nate (orthogonal to the beamline and direction of the horizontal strips) of a station's base, w_{str} is the
 2045 strip pitch, n_{str} is the hit's strip number, and z_{plane} is the location of the hit's plane along the beam-
 2046 line.

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

2047 In the fit, individual hit slopes in a slope-road are used to calculate global slopes associated with each
 2048 plane type, which are averages (e.g. M_X^{ℓ} for the average slope of horizontal planes). These in turn are
 2049 used to calculate the three composite slopes: slopes associated with the horizontal (m_x) and vertical
 2050 coordinates (m_y) and the local slope of hits in the horizontal planes (M_X^l), all of which are shown in
 2051 Equation A.4. Note that the expression for M_X^l differs but is equivalent to the expression given in ⁴¹.
 2052 This is due to a procedural change in the algorithm. The local X slope is expressed in ⁴¹ as:

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

2053 Procedurally, this entails doing the sums over y_i and $y_i z_i$, multiplying the sums by A_k , B_k , and then
 2054 subtracting both of these numbers, $\mathcal{O}(10^3)$, to get local slopes, $\mathcal{O}(10^{-1})$, while requiring preci-
 2055 sion on these numbers on the order of $\mathcal{O}(10^{-3})$. This requires precision in the sums $\mathcal{O}(10^{-7})$,
 2056 and with 32 bit fixed point numbers, there are deviations with respect to the floating point calcula-
 2057 tions at the level of $\mathcal{O}(10^{-5})$, which is enough to introduce a significant bias in the $\Delta\theta$ calculation.

2058 In order to prevent these errors, we do the subtraction first

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

2059 Thus, we change the order of operations and store $1/\bar{z}$ instead of A_k in addition to B_k . We also
 2060 change the units of y_i and z_i in the calculation by dividing the millimeter lengths by 8192.* With
 2061 these changes, a 32 bit fixed point based algorithm has essentially identical performance to that of an
 2062 algorithm based on the usual C++ 32 floating point numbers. Future work includes converting the
 2063 32 bit fixed point arithmetic to 16 bit where possible in the algorithm. While introducing 16 bit num-
 2064 bers uniformly might seem preferable, since simple 16-bit operations in the firmware can be done in
 2065 a single clock tick, and a larger number of bits increases the algorithm latency, some numbers in the
 2066 algorithm will require a larger number of bits, in particular in the local slope calculation, which is
 2067 the single calculation in the algorithm requiring the largest numeric range.

2068 In Equation A.4, θ_{st} is the stereo angle of 1.5 degrees; the sums are over relevant planes; \bar{z} is the
 2069 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

2070 and z coordinates of hits in X planes.

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

2071 From these composite slopes, the familiar expressions for the fit quantities θ (the zenith), ϕ (the az-
2072 imuth[†]), and $\Delta\theta$ (the difference in θ between the direction of the segment extrapolated back to the
2073 interaction point and its direction when entering the detector region; the following is an approxima-
2074 tion) may be calculated, as noted in ⁴¹:

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

2075 Looking at Equations A.4 and A.5, the dependence of fit quantities on input hit information be-
2076 comes clear. $\Delta\theta$ relies exclusively on information from the horizontal (X) planes. Both θ and ϕ rely
2077 on both horizontal and stereo slope information. However, the sum in quadrature of m_x and m_y in
2078 the arctangent for θ means that θ is less sensitive to errors in stereo hit information than ϕ . Given
2079 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
2080 is simply an average over slopes. This means that while errors in horizontal hit information will af-
2081 fect all three fit quantities, comparable errors in stereo hits will have a proportionately larger effect
2082 on θ and particularly on ϕ . The $\Delta\theta$ cut after step J in Figure A.1 has been implemented, requiring
2083 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 $\phi = 0$

2084 algorithm efficiency.

2085 A.2 MONTE CARLO SAMPLES

2086 The Monte Carlo (MC) samples used for these studies were generated in Athena release 20.1.0.2 us-
2087 ing simulation layout ATLAS-R2-2015-01-01-00 with muon GeoModel override version MuonSpectrometer-
2088 R.07.00-NSW and modifications to have two modules per multiplet and xxuvuvxx geometry with a
2089 stereo angle of 1.5 degrees. Muons of a single p_T were generated around the nominal IP with a smear-
2090 ing of 50 mm along the beam line and 0.015 mm orthogonal to it; these muons were pointed toward
2091 a single, large sector of the NSW. Each event consists of one muon fired towards the single NSW
2092 wedge separated by effectively infinite time from other events.

2093 A.3 NOMINAL PERFORMANCE

2094 In order to evaluate algorithm performance, a number of quantities are evaluated, including the fit
2095 quantities θ , ϕ , and $\Delta\theta$ as well as algorithm efficiency. Unless otherwise stated, that algorithm is
2096 run with a 4X+4UV coincidence threshold, slope-road size of 0.0009, an X tolerance of two slope-
2097 roads (i.e. hits in horizontal planes are written into the two slope-roads closest to the hits' value),
2098 a UV tolerance of four slope-roads[†], and a charge threshold requirement on hits of 1 (measured in
2099 units of electron charge) for a sample of 30 000 events with a muon p_T of 100 GeV. Samples were
2100 also generated for p_T values of 10 GeV, 20 GeV, 30 GeV, 50 GeV, and 200 GeV, which were used in

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

2101 some of the following studies.

2102 **A.4 FIT QUANTITIES**

2103 In order to evaluate the performance of the algorithm’s fit quantities θ , ϕ , and $\Delta\theta$, fit values are com-
2104 pared to truth-level MC values. The residual of the three fit quantities, $\theta_{fit} - \theta_{tru}$, $\phi_{fit} - \phi_{tru}$, and
2105 $\Delta\theta_{fit} - \Delta\theta_{tru}$, are recorded for every fitted track. The distributions of these quantities, in particular
2106 their biases and standard deviations, are then used to evaluate performance. In most cases, follow-
2107 ing⁴¹, the mean and standard deviation of a 3σ Gaussian fit are quoted, as they capture the main
2108 features of the algorithm and generally behave like the raw mean and rms. Nevertheless, discussion
2109 of the raw quantities will be included when their behavior deviates markedly from that of the 3σ fit
2110 quantities.

2111 The truth-level quantities used in residual distribution are taken from information in the MC.

2112 These come directly from the MC for θ , ϕ , and $\Delta\theta$. These quantities, along with the geometry of
2113 the (large) wedge, are then in turn used to calculate truth-level values for any intermediate quantities
2114 used in the algorithm. $m_{x,tru}$, for instance, is given by $\tan \theta_{tru} \sin \phi_{tru}$.

2115 Residual distributions for fit quantities under the previously described default settings of the al-
2116 gorithm are shown in Figure A.2. Both the $\theta_{fit} - \theta_{tru}$ and $\Delta\theta_{fit} - \Delta\theta_{tru}$ distributions feature a
2117 mostly Gaussian shape with more pronounced tails. The mean bias for these distributions is negligi-
2118 ble at under one tenth of a milliradian, and the fitted (raw) rms values are 0.349 (0.614) mrad for θ
2119 and 1.03 (2.55) mrad for $\Delta\theta$. The case of the $\phi_{fit} - \phi_{tru}$ distribution is less straightforward, with both
2120 the shape and bias arising from the xxuvuvxx geometry and relatively large extent of one of the two

²¹²¹ η -stations, as explained in Appendix B of³⁸. The fitted (raw) rms for the ϕ distribution is 8.67 (16.6)
²¹²² mrad.

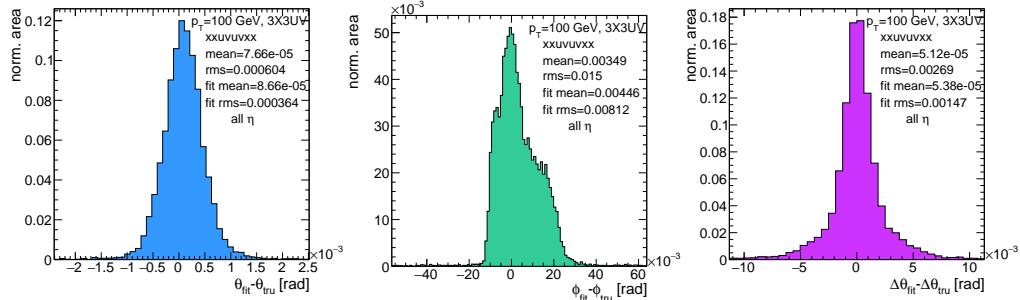


Figure A.2: Nominal residual plots; $\theta, \phi, \Delta\theta$ for $p_T = 100$ GeV muons

²¹²³ Both increasing muon p_T and increasing muon η for a fixed p_T imply increasing muon energy. As
²¹²⁴ muons become more energetic, two effects compete in affecting the quality of fit. On the one hand,
²¹²⁵ higher energy muons are deflected less by the ATLAS magnetic field, which should tend to improve
²¹²⁶ the quality of the fit, since the fitted θ (upon which $\Delta\theta$ also relies) and ϕ values are calculated under
²¹²⁷ the infinite momentum muon (straight track) assumption. However, as muon energy increases, the
²¹²⁸ likelihood that the muon will create additional secondaries increases, which creates extra hits that
²¹²⁹ degrade the quality of the fit. While the geometry of the multiplet is such that there is very good res-
²¹³⁰ olution in the direction orthogonal to the horizontal strip direction, the shallow stereo angle of 1.5
²¹³¹ degrees means that early hits caused by secondaries can have an outsize impact on m_x . $\Delta\theta$, which
²¹³² does not rely upon stereo information should feel the effect of secondaries the least and benefit from
²¹³³ straighter tracks the most and hence benefit from higher muon energies; ϕ , relying upon stereo in-
²¹³⁴ formation the most, would be most susceptible to secondaries and benefit the least from straighter

2135 tracks and hence least likely to benefit from higher muon energy; θ relies upon both horizontal and
 2136 vertical slope information, though small errors are less likely to seriously affect the calculation, so the
 2137 two effects are most likely to be in conflict for this fit quantity.

2138 The interplay of these effects on the residual standard deviations can be seen in their dependen-
 2139 cies on η (Figure A.3; note that the final point in each of these plots is the rms of the distribution
 2140 overall η) and p_T (Figure A.4). For $p_T = 100$ GeV muons, $\Delta\theta$ performance increases with η (en-
 2141 ergy), and ϕ performance decreases, as expected;[§] for θ , the two effects appear to compete, with per-
 2142 formance first increasing with η until the effects of secondaries begins to dominate. Integrated over
 2143 all η , the effects are less clearly delineated. Both $\Delta\theta$ and θ performance increases with increasing p_T ,
 2144 suggesting straighter tracks with increasing energy are the dominant effect for these quantities, while
 2145 ϕ performance appears to improve and then deteriorate (the slight improvement at high p_T is due to
 2146 the addition of the $\Delta\theta$ cut into the algorithm, which filters out very poor quality fits).

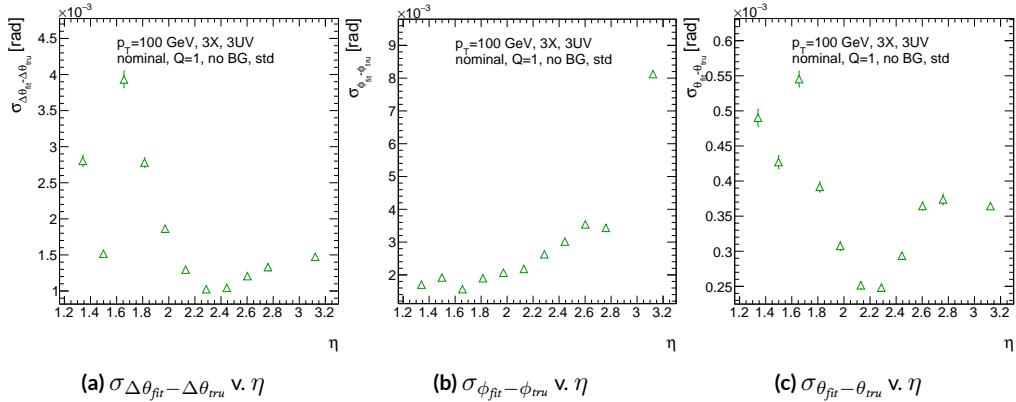


Figure A.3: The rms distributions of $\Delta\theta$, ϕ , and θ as a function of η for $p_T = 100$ GeV; the final point in each plot is the rms obtained from a fit to the full distribution including all η bins.

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

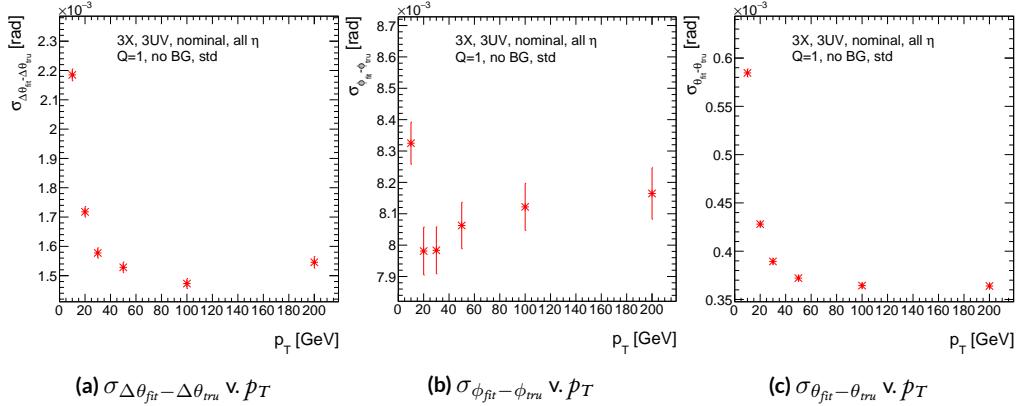


Figure A.4: The rms distributions of $\Delta\theta$, ϕ , and θ as a function of p_T .

2147 The rms of the three benchmark quantities as a function of algorithm set (i.e. slope-road) coinci-
 2148 dence threshold are shown in Figure A.5 using Gaussian fits and in Figure A.6 for the raw quantities.
 2149 The fitted σ 's for θ and ϕ are fairly stable across coincidence threshold. $\Delta\theta$, on the other hand, per-
 2150 forms better particularly for the most stringent coincidence threshold; this is a result of the fact that
 2151 additional information for more hits greatly improves the quality of the local slope fit calculation.
 2152 The raw rms is a different story. Naïvely, one would expect the performance to get better with more
 2153 stringent coincidence threshold, but this is not the case in Figure A.6. As the coincidence thresh-
 2154 old gets more stringent, fewer and fewer tracks are allowed to be fit. When moving from 2X hits to
 2155 3X hits, the tracks that get vetoed populate the tails of the distribution outside the 3σ fit range but
 2156 are not in the very extremes of the distribution. While tracks with 2X hits are of lower quality than
 2157 those with 3 and 4 X hits, tracks with the very worst fit values pass even the most stringent coinci-
 2158 dence threshold requirements (e.g. as a result of many hits arising from a shower of secondaries).
 2159 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.

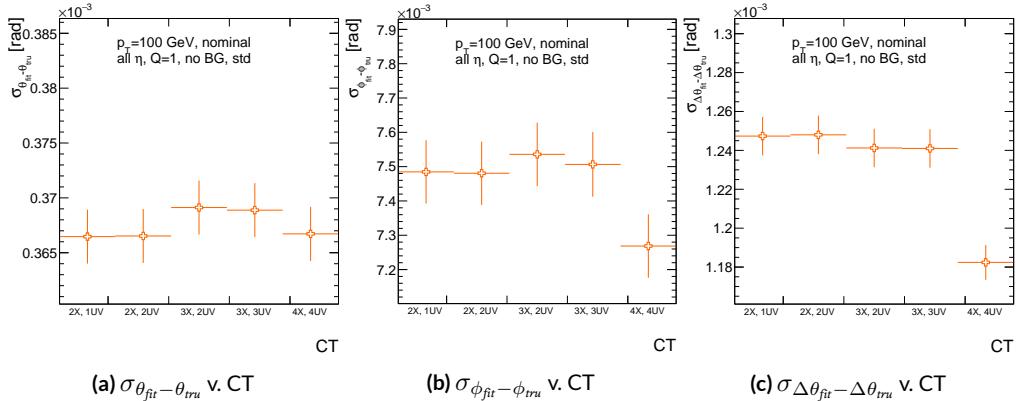


Figure A.5: The fitted rms of residual distributions for θ , ϕ , and $\Delta\theta$ as a function of coincidence threshold for $p_T = 100$ GeV.

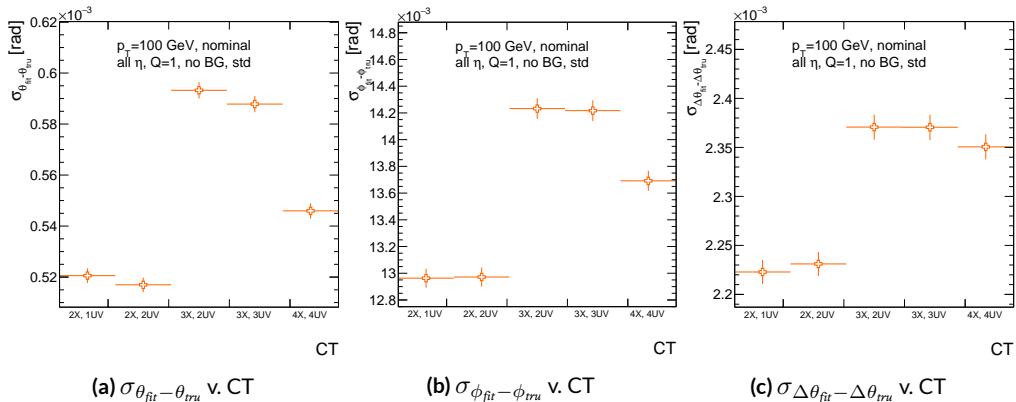


Figure A.6: The raw rms of residual distributions for θ , ϕ , and $\Delta\theta$ as a function of coincidence threshold for $p_T = 100$ GeV.

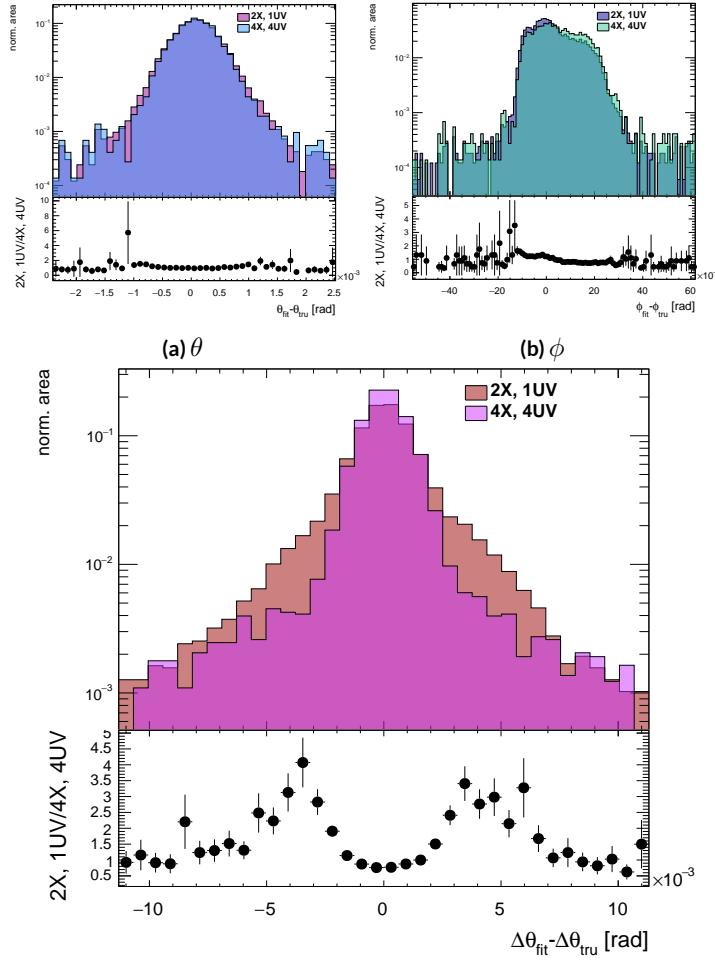


Figure A.7: Nominal $\Delta\theta$ residual distribution for $p_T = 100$ GeV muons with coincidence thresholds 2X+1UV and 4X+4UV normalized to the same area and plotted together (top) as well as the ratio of the 2X+1UV distribution and the 4X+4UV per bin.

2164 A.5 EFFICIENCIES

2165 Two general efficiencies have been formulated to study the performance of the MMTP algorithm.

2166 The first, denoted ε_{alg} , is the fraction of tracks that pass some (slope-road) coincidence threshold

2167 configuration that are successfully fit. An event that passes a slope-road coincidence but does not fit

2168 fails because some of the hits included are of sufficiently poor quality to throw off the fit. This effi-

2169 ciency answers the question of how often the algorithm performs fits when technically possible, giv-

2170 ing a measure of overall algorithm performance for a given configuration. For example, $\varepsilon = 95\%$ for

2171 $3X+2UV$ means that 95% of tracks that produce at least $3X$ hits and $2UV$ hits in at least one slope-

2172 road will be successfully fitted 95% of the time. The performance of this efficiency as a function of

2173 coincidence threshold, η (with the final point once again being the efficiency integrated over all η),

2174 and p_T is shown in Figure A.8. ε_{alg} is fairly constant in η and decreases with increased p_T , which can

2175 be attributed to the increased likelihood of secondaries introducing lower quality hits that cause the

2176 fit to fail.

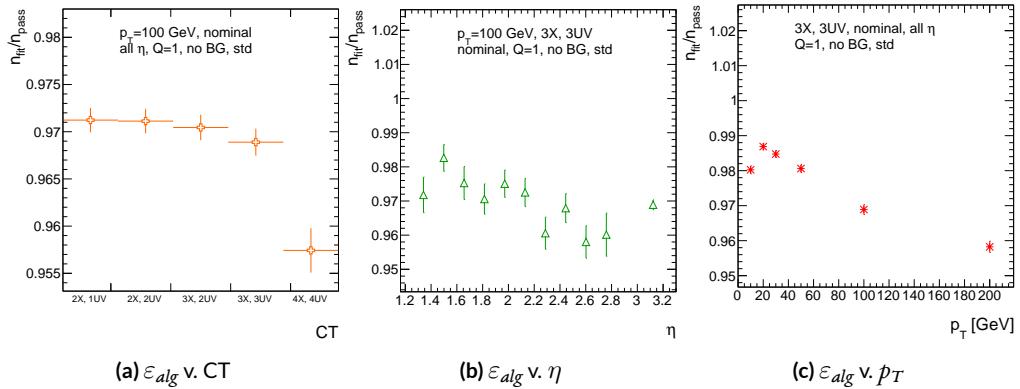


Figure A.8: ε_{alg} and as a function of coincidence threshold, η (final point is ε_{alg} integrated over all η), and p_T .

2177 The second efficiency type, denoted ε_{fit} , is the fraction of tracks that enter the wedge whose fits
 2178 (if any) satisfy a given coincidence threshold. This efficiency can be used to help establish an optimal
 2179 coincidence threshold setting in the algorithm, balancing the improved overall fit quality of higher
 2180 thresholds with the greater number of fits for lower thresholds. Hence, an ε_{fit} of 95% at 3X+2UV
 2181 means that 95% of tracks entering the wedge are fit and that these fits include at least 3X and 2UV
 2182 hits. ε_{fit} as a function of coincidence threshold is shown in Figure A.9 (a), which shows that the ma-
 2183 jority of fits having at most 3X+3UV hits. That there is a marked drop to 4X+4UV is not surpris-
 2184 ing, as there is a substantial population outside the 4X+4UV bin in Figure A.10. The behavior of
 2185 ε_{fit} with η in Figure A.9 (b) (with the final point once again being the efficiency integrated over all
 2186 η) is much more varied, with geometric effects of detector acceptance coming into play. The per-
 2187 formance of ε_{fit} as a function of p_T , shown in Figure A.9 (c), is similar to that of ε_{alg} coincidence
 2188 threshold, again consistent with the effects of secondaries at higher energies.

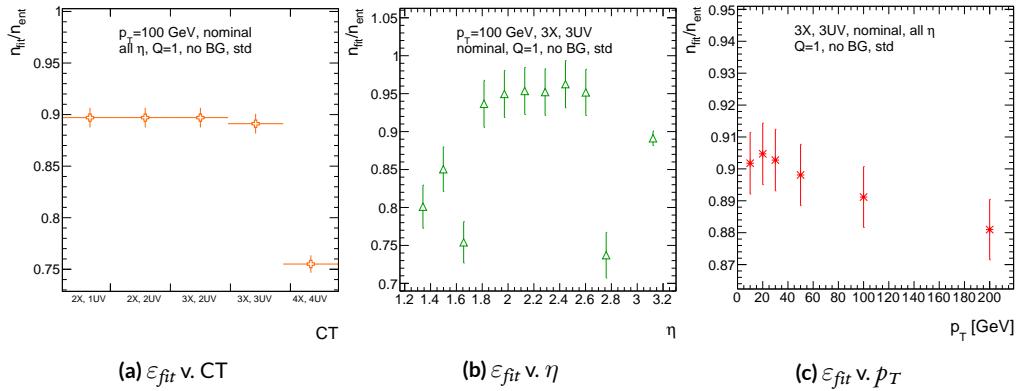


Figure A.9: ε_{fit} and as a function of coincidence threshold, η (final point is ε_{fit} integrated over all η), and p_T .

2189 In order to better understand efficiency behavior with coincidence threshold, the distribution of

2190 highest slope-road coincidence thresholds in events is shown in Figure A.10, with the o,o bin con-
 2191 taining events that did not meet requirements for the minimum $2X+1UV$ coincidence threshold for
 2192 a fit to occur. That the efficiency is lower at higher coincidence threshold suggests that most of the
 2193 fits that fail have high hit multiplicity (i.e. a similar number fails in each of the coincidence thresh-
 2194 old bins in Figure A.8 (a)), which is consistent with the interpretation that the primary source of fit
 2195 failures is bad hits originating from secondaries created by higher energy muons.

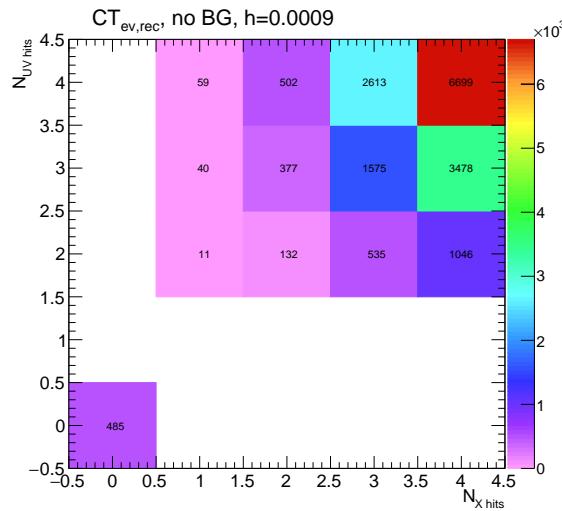


Figure A.10: The distribution of highest slope-road coincidence thresholds in events; the 0,0 bin is the number of events passing selection requirements that fail to form the minimum $2X+1UV$ coincidence threshold necessary for a fit.

2196 A.6 INCOHERENT BACKGROUND

2197 The default slope-road size and tolerances associated with horizontal and stereo hits used in the
2198 above studies were configured to optimize algorithm performance, similar to studies in ⁴¹. In order
2199 to evaluate algorithm performance under conditions with more limited resources, as might be ex-
2200 pected at run-time, additional studies were conducted with the slope-road size and hit tolerances set
2201 equivalent to the sensitive area of a single VMM chip[¶] both with and without generation of incoher-
2202 ent background.

2203 Incoherent background is generated based on the assumption that the intensity only varies as a
2204 function of the distance from a point to the beamline, r . The number of hits per unit area per unit
2205 time as a function of r is given in Equation A.6 and taken from ⁴¹.

$$I = I_0 (r/r_0)^{-2.125} \quad (\text{A.6})$$

2206 where $r_0 = 1000$ mm and $I_0 = 0.141$ kHz/mm²

2207 Background generation happens per event as follows:

- 2208 1. Determine the total number of hits to be generated in this event according to a Poisson distri-
2209 bution
- 2210 2. Assign a time to hits uniformly in $[t_{\text{start}} - t_{\text{VMM}}, t_{\text{end}}]$ where start and end are for the event
2211 clock and t_{VMM} is the VMM chip deadtime (100 ns)
- 2212 3. Assign a plane to hits uniformly
- 2213 4. Assign a ϕ value to hits uniformly

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

2214 5. Assign an r to hits according to Equation A.6

2215 6. Calculate hit information according to these values.

2216 The expectation value for the Poisson distribution is determined by integrating Equation A.6

2217 over the surface area of the wedge to get the total hit rate for the wedge, Γ , and then multiplying this

2218 by the length of the time window over which hits may be generated. With $H = 982$ mm, $b_1 =$

2219 3665 mm, and $\theta_w = 33\pi/180$, we find^{||}:

$$\Gamma = 2I_0 r_0^{2.125} \int_0^{\theta_w/2} d\phi \int_{H \sec \phi}^{(H+b_1) \sec \phi} r dr r^{-2.125} = 98.6657 \text{ MHz} \quad (\text{A.7})$$

2220 In this case, we have taken the nominal values of the MM sector geometry for H (wedge base), b_1

2221 (the wedge height), and θ_w (the wedge opening angle).

2222 The effects of incoherent background and larger slope road size are summarized in Figure A.11 for

2223 efficiencies and in Figure A.13 and Table A.1 for residual of fit quantities.

2224 Figure A.11 show the effect of both wider slope-roads and the introduction of background on ef-

2225 ficiencies. The introduction of wider slope-roads increases the chance that an early errant hit (either

2226 from secondaries/ionization or background) will be introduced into the fit, and the presence of in-

2227 coherent background greatly increases the number of such errant hits. Both wider slope-roads and

2228 background drive down the number of fits (numerator) in both efficiencies, and background can

2229 artificially inflate the denominator of ε_{alg} , a reco-level, slope-road coincidence threshold. The shape

2230 of the ε_{fit} versus coincidence threshold distributions remains fairly constant with each complicat-

^{||}Using Mathematica and the extra factor of r from the volume element

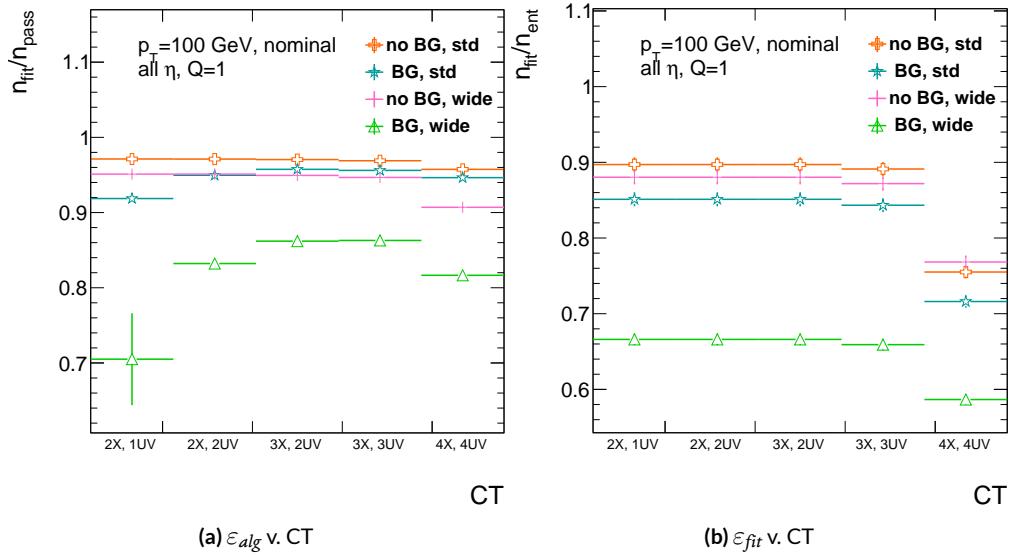


Figure A.11: The algorithm and total efficiencies as a function of coincidence threshold for different background settings and slope-road sizes (standard and wide (one slope road as 1 VMM chip)).

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

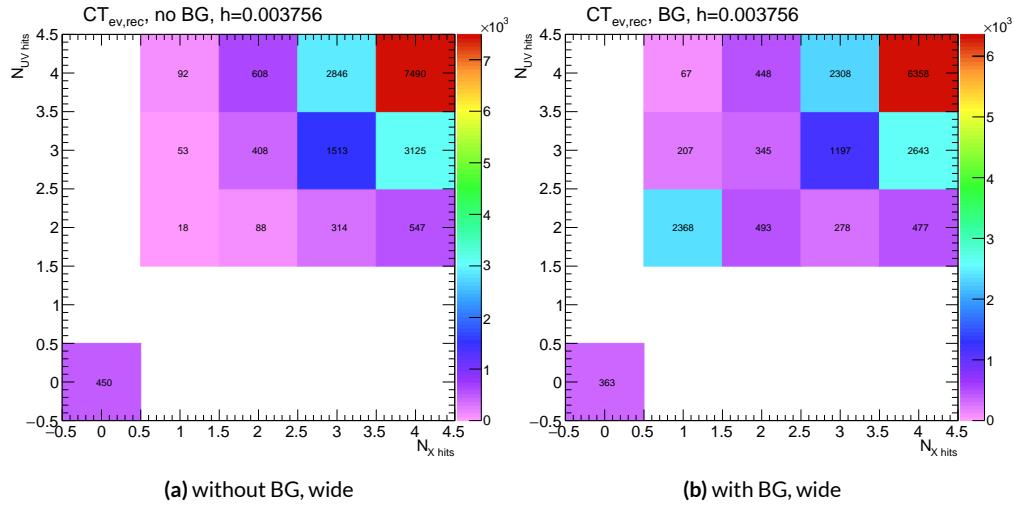


Figure A.12: The distribution of highest slope-road coincidence thresholds in events for the algorithm with wide slope-roads (width of 1 VMM) both without (a) and with (b) incoherent background; the 0,0 bin is the number of events passing selection requirements that fail to form the minimum $2X+1UV$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

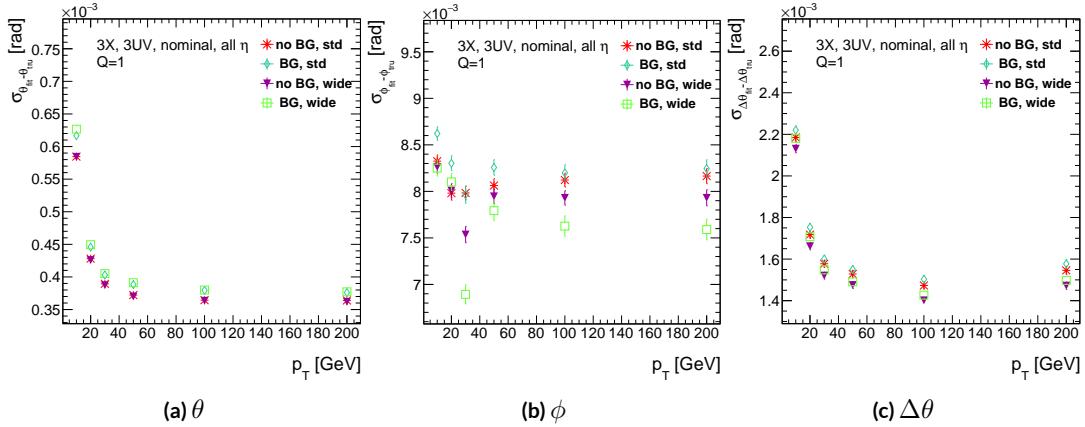


Figure A.13: The three fit quantity residual rms values as a function of p_T for different background settings and slope-road sizes (standard and wide (one slope road as 1 VMM chip)).

	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.

2252 As Table A.1 shows, rms values appear to be robust to an increase in slope-road size. Neverthe-
2253 less, though the fitted σ residual values are also fairly robust to the introduction of background, the
2254 raw rms values are not. While the raw $\Delta\theta$ rms stays stable, both θ and ϕ suffer noticeable degra-
2255 dation, which suggests that the introduction of background has a detrimental effect on horizontal
2256 slope residual (i.e. on stereo strips in particular). This level of degradation is likely acceptable for θ ,
2257 though further steps may need to be taken to address ϕ .

2258 A.7 BCID

2259 A fitted track's BCID is determined by the most common BCID associated with its hits. Concerns
2260 were raised that this might cause incorrect BCID association for fitted tracks. In order to address
2261 this, the rate of successful BCID association for fitted tracks was recorded. Figure A.14 shows the
2262 dependence of this success rate as a function of p_T and coincidence threshold in the different back-
2263 ground and resource conditions used in the previous section. The successful BCID identification
2264 rate is always over 99.5%, demonstrating that this issue is not a concern with the state-of-the-art de-
2265 tector simulation.

2266 A.8 CHARGE THRESHOLD

2267 The MMTP uses the first hits registered passing a charge threshold requirement given in units of
2268 electron charge. In principle, it would be beneficial to be able to use any hits that are registered re-
2269 gardless of deposited charge, but in the high rate environment envisioned for the NSW, this require-
2270 ment might need to be raised. Nominal algorithm settings have this charge threshold requirement

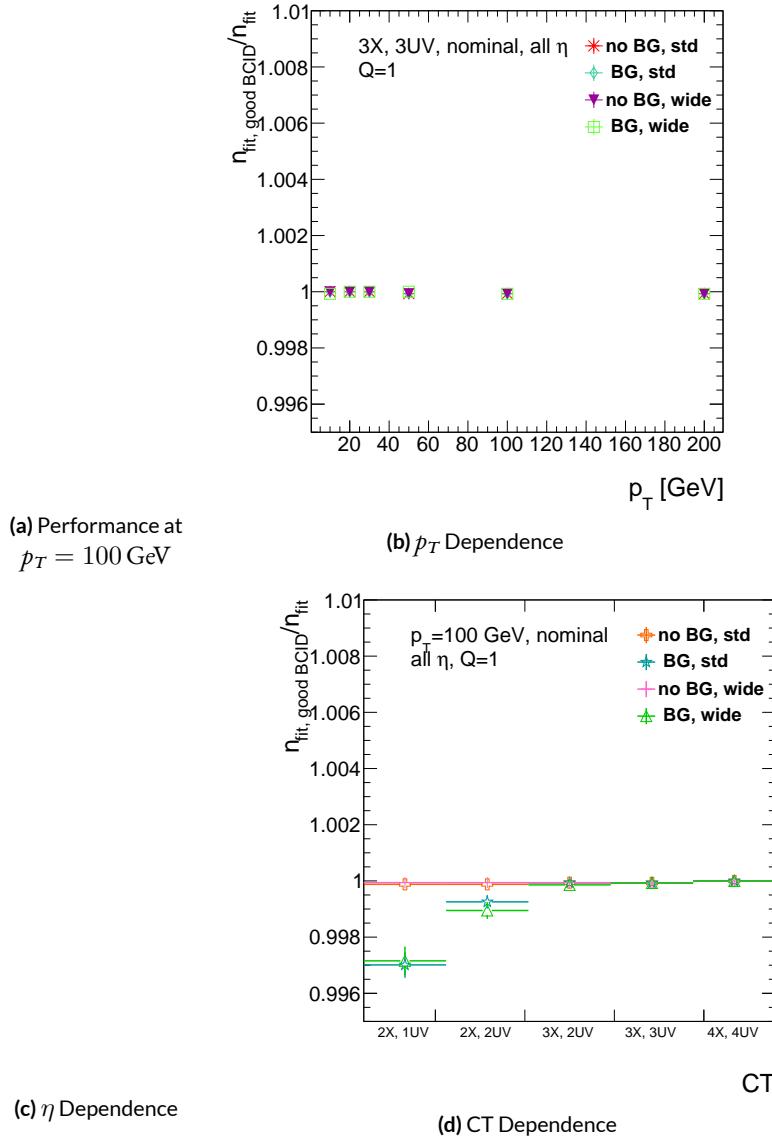


Figure A.14: The rate of good BCID association based majority hit BCID as a function of p_T and coincidence threshold.

2271 set to 1, and studies were conducted on algorithm performance for charge threshold values of 0, 1,
 2272 and 2. Efficiencies as a function of coincidence threshold for different charge thresholds are shown
 2273 in Figure A.15. Increasing the charge threshold lowers both efficiencies, particularly at high coinci-
 2274 dence threshold, which suggests that energetic muons with secondaries create both very many hits
 2275 and hits with higher charge. While the shapes of the fit quantity distributions as a function of p_T in
 2276 Figure A.16 are fairly constant across charge threshold, performance is not. θ and $\Delta\theta$ show some im-
 2277 provement with higher charge threshold, particularly at low p_T , suggesting that resolution improves
 2278 in the vertical direction, but ϕ shows degradation at higher charge threshold, which is a symptom
 2279 of more highly charged particles experiencing greater bending in the ATLAS magnetic field in the ϕ
 2280 direction.

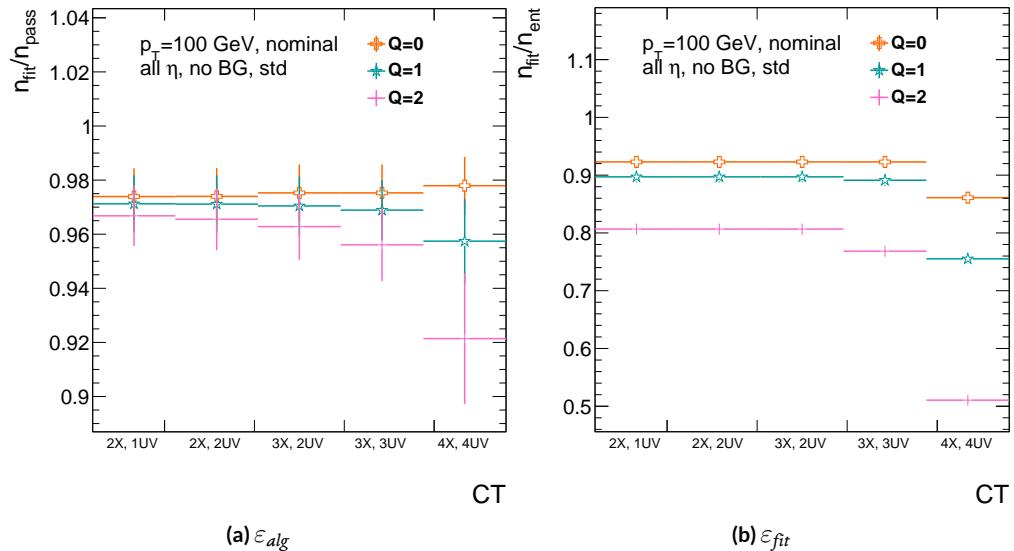


Figure A.15: The efficiencies as a function of coincidence threshold for charge thresholds of 0, 1, and 2.

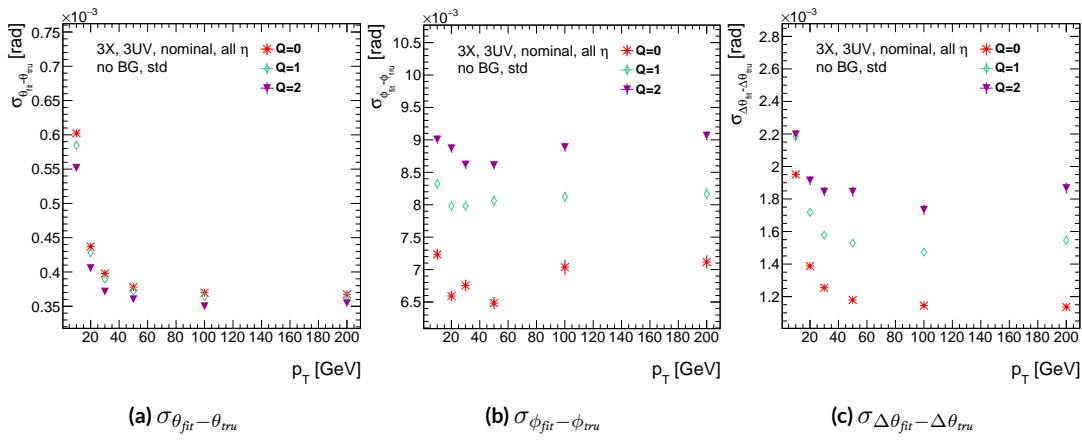


Figure A.16: The fit quantity residual rms values as a function of p_T for charge thresholds of 0, 1, and 2.

2281 A.9 MISALIGNMENTS AND CORRECTIONS

2282 The performance of the trigger algorithm under misalignment has been studied for each of the six
2283 alignment quantities (three translations and three rotations all along the principal axes) described
2284 in[?] and[?], whose convention we will follow here. For the simulated wedge studied here the local co-
2285 ordinates described in[?] are taken to be centered at the center of the base of the wedge^{**}, the local t
2286 axis corresponds to the axis of the beam line, the local z axis corresponds to the direction orthogo-
2287 nal to both the beam line and the horizontal strips, and the local s axis completes the right-handed
2288 coordinate system. The rotation angles α , β , and γ correspond to rotations around the local t , z ,
2289 and s axes, respectively. Note that the local s , z , and $-t$, axes correspond to the usual global x , y , and
2290 z axes. Misalignments were studied in twenty evenly spaced increments from nominal positions
2291 to misalignments of 1.5 mrad for the rotations (-1.5 mrad to +1.5 mrad for the γ case), and of 5 mm
2292 (a roughly corresponding linear shift) for the translations. In all cases, the front quadruplet is mis-
2293 aligned while the rear quadruplet remains in its nominal position. While only the front quadruplet
2294 of a single wedge is misaligned, the framework for misalignment presented below could be used to
2295 study generic local and global misalignments. The six misalignments are schematically represented
2296 in Figure A.17.

2297 Chamber misalignments manifest themselves as altered strips in algorithm input. In order to sim-
2298 ulate the effects of misalignment, the change in the local y coordinate—the distance from the bot-

**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.

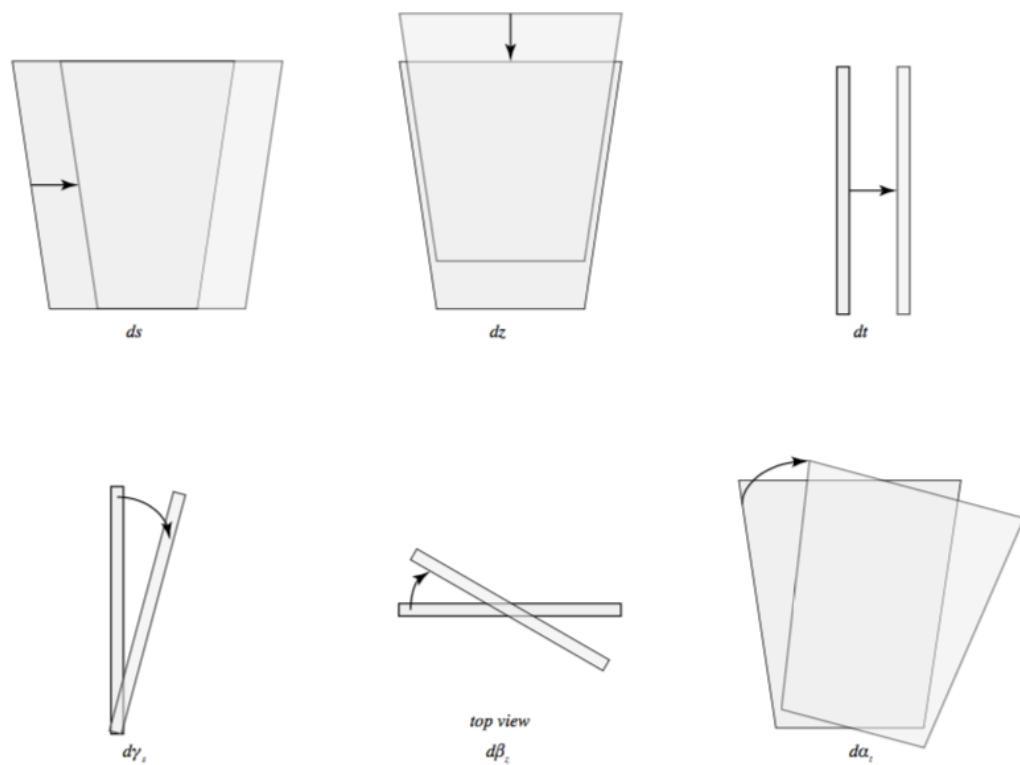


Figure A.17: The different misalignment cases as defined in the AMDB manual.

2299 tom wedge center in the direction perpendicular to both the beamline and the strip direction—is
 2300 calculated for a track coming straight from the interaction point defined by the truth-level θ and ϕ
 2301 angles for generic misalignment. This displacement in y is then added to input hit information and
 2302 the algorithm is then run normally.

2303 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; $(0, y_{base}, z_{pl})$ ($(0, 0, 0)$) for the nominal case in global (local) coordinates
$\vec{\zeta}$	$\vec{s} - \vec{\mathcal{O}}_{base}$
primed quant.	quantities after misalignment

2304 Generically speaking, a hit is the intersection of a line (the muon track) with a plane (the individual
 2305 plane in the multiplet). We assume the muon moves in a straight line defined by the origin and
 2306 the truth-level θ_{pos} and ϕ_{pos} (i.e. the infinite momentum limit) and that the MM plane is rigid and
 2307 defined by a point, which we take to be the center of the bottom edge of the plane, and a normal
 2308 vector, which we take to be the z axis in the nominal case.

2309 The coordinate axes x, y, z axes used here correspond to the usual AMDB $s, z, -t$ axes. Since the
 2310 direction does not really matter when studying misalignment or corrections thereof, the major dif-

²³¹¹ ference is the choice of origin.

²³¹² The muon track we denote^{††} \vec{s} , the bottom point of the plane $\vec{\mathcal{O}}_{base}$, and the normal vector \hat{n} .

²³¹³ The muon track will always be given as (the wedge gets moved, not the muon):

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

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

$$\vec{s} = k\hat{s} = \frac{z_{pl}}{\cos \theta_{pos}} \hat{s} = z_{pl} (\tan \theta \sin \phi \hat{x} + \tan \theta \cos \phi \hat{y} + 1) \quad (\text{A.10})$$

²³¹⁴ where $k \in \mathbb{R}$, along with the unit vector \hat{s} , defines the point where the track intersects the wedge.

²³¹⁵ Rotations are done before translations, according to the order prescribed in the AMDB guide for

²³¹⁶ chamber alignment, so the axes the principal axes of the plane are rotated according to the following

²³¹⁷ matrix (where s , c , and t are the obvious trigonometric substitutions)

$$\begin{aligned} & \begin{pmatrix} 1 & 0 & 0 \\ 0 & c\gamma & -s\gamma \\ 0 & s\gamma & c\gamma \end{pmatrix} \begin{pmatrix} c\beta & 0 & s\beta \\ 0 & 1 & 0 \\ -s\beta & 0 & c\beta \end{pmatrix} \begin{pmatrix} c\alpha & -s\alpha & 0 \\ s\alpha & c\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c\gamma & -s\gamma \\ 0 & s\gamma & c\gamma \end{pmatrix} \begin{pmatrix} c\alpha c\beta & -s\alpha c\beta & s\beta \\ s\alpha & c\alpha & 0 \\ -c\alpha s\beta & s\alpha s\beta & c\beta \end{pmatrix} \\ & = \boxed{\begin{pmatrix} c\alpha c\beta & -s\alpha c\beta & s\beta \\ s\alpha c\gamma + c\alpha s\beta s\gamma & c\alpha c\gamma - s\alpha s\beta s\gamma & -c\beta s\gamma \\ s\alpha s\gamma - c\alpha s\beta c\gamma & c\alpha s\gamma + s\alpha s\beta c\gamma & c\beta c\gamma \end{pmatrix}} = \mathcal{A} \end{aligned} \quad (\text{A.11})$$

^{††}Recall ϕ_{pos} is defined with respect to the y axis instead of the x axis, as might otherwise be typical.

2318 The thing that matters is what the new strip hit is—i.e. what the new y value is since this, along
 2319 with a plane number, is all that is fed into the algorithm. To find this, we must solve for the new
 2320 point of intersection with the rotated plane and then apply the effects of translations. The path con-
 2321 necting the base of the wedge with the intersection of the muon track will always be orthogonal to
 2322 the normal vector of the plane. Our quantities after misalignment, denoted by primed quantities,
 2323 will look like

$$\mathcal{O}_{base} \rightarrow \mathcal{O}_{base} + ds\hat{x} + dz\hat{y} + dt\hat{z} = \mathcal{O}'_{base}, \quad \hat{n} \rightarrow A\hat{n} = A\hat{z} = \hat{z}', \quad \vec{s} \rightarrow k'\hat{s} + \mathcal{O}_{IP} = \vec{s}' \quad (\text{A.12})$$

2324 so, moving to explicit, global coordinates in the last line so we can do the computation (relying on
 2325 the fact that any vector in the wedge, namely $\vec{\zeta} = \vec{s} - \mathcal{O}$ the local coordinates of the interaction
 2326 point, is necessarily orthogonal to \hat{n}):

$$0 = \hat{n} \cdot (\vec{\mathcal{O}}_{base} - \vec{s}) \rightarrow 0 = A\hat{z}' \cdot (\vec{\mathcal{O}}'_{base} - (k'\hat{s} + \vec{\mathcal{O}}_{IP})) \quad (\text{A.13})$$

$$\rightarrow k' = \frac{s\beta\vec{\mathcal{O}}'_{base-IP,x} - c\beta s\gamma\vec{\mathcal{O}}'_{base-IP,y} + c\beta c\gamma\vec{\mathcal{O}}'_{base-IP,z}}{\hat{s} \cdot \hat{z}'} \quad (\text{A.14})$$

$$= \frac{s\beta ds - c\beta s\gamma(y_{base} + dz) + c\beta c\gamma(z_{pl} + dt)}{s\beta s\theta s\phi - c\beta s\gamma s\theta c\phi + c\beta c\gamma c\theta} \quad (\text{A.15})$$

2327 To find our new y coordinate, we need to evaluate $s'_y = \hat{y}' \cdot k'\vec{s}$ to find the final correction of:

$$\Delta y = \vec{\zeta}' \cdot \hat{y}' - \vec{\zeta} \cdot \hat{y} = (k'\hat{s} - \vec{\mathcal{O}}'_{base}) \cdot \hat{y}' - (s_y - y_{base}) \quad (\text{A.16})$$

²³²⁸ The correction will be plane dependent since (denoting the stereo angle ω):

$$\hat{y}_x = \hat{y} \rightarrow \hat{y}'_x = -s\alpha c\beta \hat{x} + (\alpha c\gamma - s\alpha s\beta s\gamma) \hat{y} + (\alpha s\gamma + s\alpha s\beta c\gamma) \hat{z} \quad (\text{A.17})$$

²³²⁹ and

$$\begin{aligned} \hat{y}_{U,V} = & \pm s\omega \hat{x}' + \omega \hat{y}'_{U,V} = [\pm \alpha c\beta s\omega - s\alpha c\beta \omega] \hat{x} + [\pm (s\alpha c\gamma + s\alpha s\beta s\gamma) s\omega \\ & + (\alpha c\gamma - s\alpha s\beta s\gamma) \omega] \hat{y} + [\pm (s\alpha s\gamma - c\alpha s\beta c\gamma) s\omega + (\alpha s\gamma + s\alpha s\beta c\gamma) \omega] \hat{z} \end{aligned}$$

²³³⁰

²³³¹ A.10 INDIVIDUAL CASES

²³³² Currently we only study the cases where one misalignment parameter is not zero. We examine these
²³³³ in detail below, calculating the most pertinent quantities in the misalignment calculation, k'/k and
²³³⁴ the new horizontal and stereo y axes. Before setting out, we simplify the expressions for the trans-
²³³⁵ formed \hat{y}' 's, removing any terms with the product of two sines of misalignment angles, which will be
²³³⁶ zero.^{††}

$$\hat{y}'_x = -s\alpha c\beta \hat{x} + \alpha c\gamma \hat{y} + \alpha s\gamma \hat{z} \quad (\text{A.19})$$

²³³⁷

$$\hat{y}'_{U,V} = [\pm \alpha c\beta s\omega - s\alpha c\beta \omega] \hat{x} + [\pm s\alpha c\gamma s\omega + c\alpha c\gamma \omega] \hat{y} + [\mp c\alpha s\beta c\gamma s\omega + c\alpha s\gamma \omega] \hat{z} \quad (\text{A.20})$$

^{††}If only one misalignment parameter is non-zero, then two or more sines will contain at least one term will contain $\sin 0 = 0$.

²³³⁸ If the translations are zero,

$$k' = \frac{-c\beta s\gamma y_{base} + c\beta c\gamma z_{pl}}{s\beta s\theta s\phi - c\beta s\gamma s\theta c\phi + c\beta c\gamma c\theta}, \quad k'/k = \frac{-c\beta s\gamma y_{base}/z_{pl} + c\beta c\gamma}{s\beta t\theta s\phi - c\beta s\gamma t\theta c\phi + c\beta c\gamma} \quad (\text{A.21})$$

²³³⁹ A.II $ds \neq 0$

²³⁴⁰ $k'/k = 1$ (the point of intersection does not move closer or further from the IP), and only the stereo
²³⁴¹ planes are affected. Note that only relevant term in Equation A.16, for the stereo strip \hat{y} for $\vec{\mathcal{O}}'_{base} =$
²³⁴² $ds\hat{x}$ is:

$$\pm \sin \omega ds \approx \pm 0.0261 ds \quad (\text{A.22})$$

²³⁴³ meaning that a displacement in x of 17 mm, more than three times the range of misalignments studied,
²³⁴⁴ would be necessary for a shift in the stereo planes corresponding to one strip width.

²³⁴⁵ A.12 $dz \neq 0$

²³⁴⁶ $k'/k = 1$ (the point of intersection does not move closer or further from the IP). This case is the
²³⁴⁷ trivial one (cf. Equation A.16 with $\vec{\mathcal{O}}'_{base} = dz\hat{y}$). y just gets moved in the opposite direction as the
²³⁴⁸ wedge. Correction is an additive constant.

²³⁴⁹ A.13 $dt \neq 0$

²³⁵⁰ $k'/k = (z_{pl} + dt) / z_{pl}$. y gets modified by a simple scale factor. Correct by storing changing definitions
²³⁵¹ of plane positions in algorithm to match the misaligned values.

²³⁵² A.14 $\alpha \neq 0$

²³⁵³ $k'/k = 1$ and

$$\hat{y}'_x = -s\alpha\hat{x} + c\alpha\hat{y} \quad (\text{A.23})$$

$$\hat{y}'_{U,V} = [\pm c\alpha s\omega - s\alpha c\omega] \hat{x} + [\pm s\alpha s\omega + c\omega] \hat{y} \quad (\text{A.24})$$

²³⁵⁴ A.15 $\beta \neq 0$

²³⁵⁵ We have $k'/k = (1 + \tan \beta \tan \theta \sin \phi)^{-1}$, and

$$\hat{y}'_x = \hat{y} \quad (\text{A.25})$$

$$\hat{y}'_{U,V} = \hat{y} \pm (c\beta\hat{x} - s\beta\hat{z}) s\omega \quad (\text{A.26})$$

²³⁵⁶ A.16 $\gamma \neq 0$

$$k'/k = \frac{1 - \tan \gamma \frac{y_{base}}{z_{pl}}}{1 - \tan \gamma \tan \theta \cos \phi} \quad (\text{A.27})$$

$$\hat{y}'_x = c\gamma\hat{y} + s\gamma\hat{z} \quad (\text{A.28})$$

$$\hat{y}'_{U,V} = \pm s\omega\hat{x} + c\omega\hat{y} - s\gamma c\omega\hat{z} \quad (\text{A.29})$$

²³⁵⁷ In order to evaluate algorithm performance under misalignment and corrections for misalign-

²³⁵⁸ ment, the absolute means and relative resolutions of the fit quantities θ , ϕ , and $\Delta\theta$ are measured as

2359 a function of misalignment. In the following, results will only be shown for which the effects of mis-
2360 alignment are significant. “Significant,” for misalignments of 1 mm (0.3 mrad) for translations (ro-
2361 tations) means more than a 5% degradation in rms and/or bias shifts in θ , ϕ , and $\Delta\theta$ of 0.01 mrad, 1
2362 mrad, and 0.1 mrad, respectively.

2363 While corrections are typically done on a case-by-base basis, they fall under two general cate-
2364 gories, analytic and simulation based. Analytic corrections rely upon specific knowledge of the mis-
2365 alignment, with each case being handled separately; as such, the additional resources required, both
2366 extra constants and operations, if any, vary accordingly. Simulation based corrections are all done in
2367 the same manner. The algorithm is run over a training MC sample (same setup but with $p_T = 200$
2368 GeV instead of the normal 100 GeV sample so as not to overtrain the corrections), and the mean bi-
2369 ases for θ , ϕ , and $\Delta\theta$ are saved for different, equally spaced regions in the $\eta - \phi$ plane over the wedge
2370 based on the fitted θ and ϕ values. Currently, these values are saved for 10 η and 10 ϕ bins (100 η, ϕ
2371 bins total), with the number of bins in each direction being a configurable parameter. When the al-
2372 gorithm runs with simulation based correction, this table of constant corrections is saved in a LUT
2373 before runtime, and corrections are added to final fit quantities based on the (uncorrected) θ and
2374 ϕ fit values. With the settings mentioned, this is 300 extra constants ($10\eta\text{-bins} \times 10\phi\text{-bins} \times 3$ fit
2375 quantities) and two extra operations (a lookup and addition for each quantity done in parallel). The
2376 simulation correction can, in principle, also be applied to the algorithm in nominal conditions with
2377 non-trivial improvements, as detailed below in Section A.17. Depending on the misalignment case in
2378 question, different approaches work better. A summary of correction methods, including resources
2379 necessary for the individual analytic cases, is shown in Table A.3.

	Δ_s	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

2380 A.17 SIMULATION CORRECTION OF THE ALGORITHM UNDER NOMINAL CONDITIONS

2381 In addition to using simulation based correction to counter the effects of several classes of misalign-
 2382 ment, the correction can be applied at to the algorithm under nominal conditions. The main effect
 2383 of this correction is to mitigate the effects of the bias in stereo strips. As such, the correction has a
 2384 larger effect on quantities that rely on the aggregate slope m_y , as can be seen in in Figure A.18, im-
 2385 proving $\sigma_{\theta_{fit} - \theta_{true}}$ resolution by about 25%, and reducing $\sigma_{\phi_{fit} - \phi_{true}}$ by over 50% and restoring a largely
 2386 Gaussian shape. The slight, apparent degradation in $\Delta\theta$ is due to a more mild version of the effect
 2387 seen in Figure A.7.

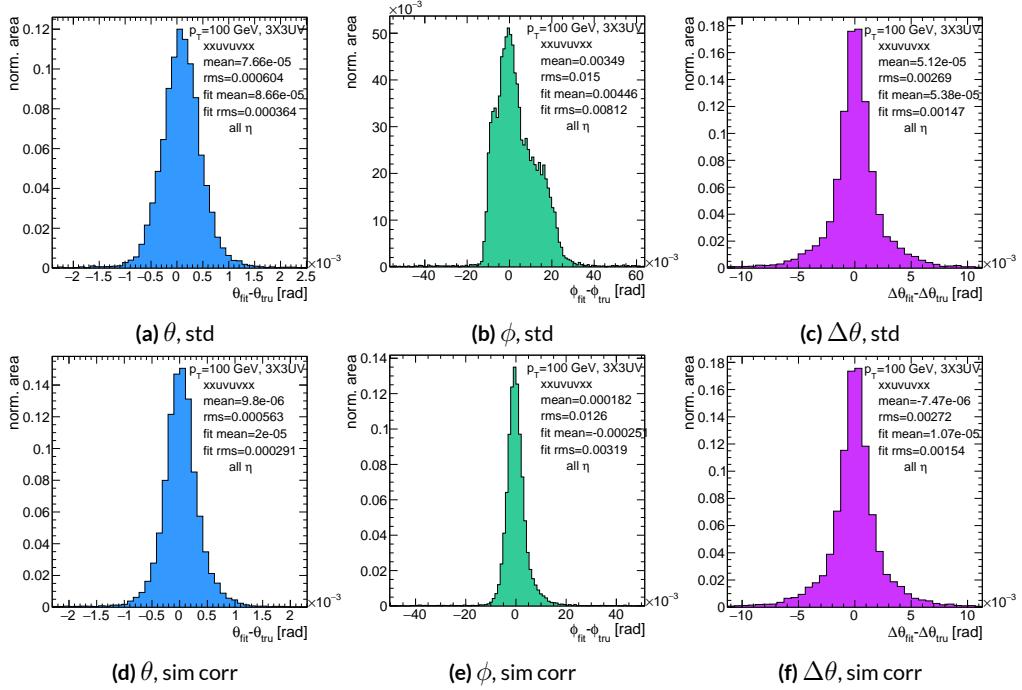


Figure A.18: Nominal residual plots for both uncorrected and simulation corrected cases; $\theta, \phi, \Delta\theta$ for $p_T = 100$ GeV muons

2388 As can be seen in Figure A.19, the simulation based correction also removes the η dependence to
 2389 fit quantity resolution distributions, as expected. One consequence of this is that simulation-based
 2390 corrections applied to the misalignment cases below will restore performance to the “sim” and not
 2391 the “std” distributions of Figure A.18. Hence, when making comparisons between simulation cor-
 2392 rected curves and the nominal performance point, simulation-corrected distributions of benchmark
 quantities versus misalignment will often look generally better.

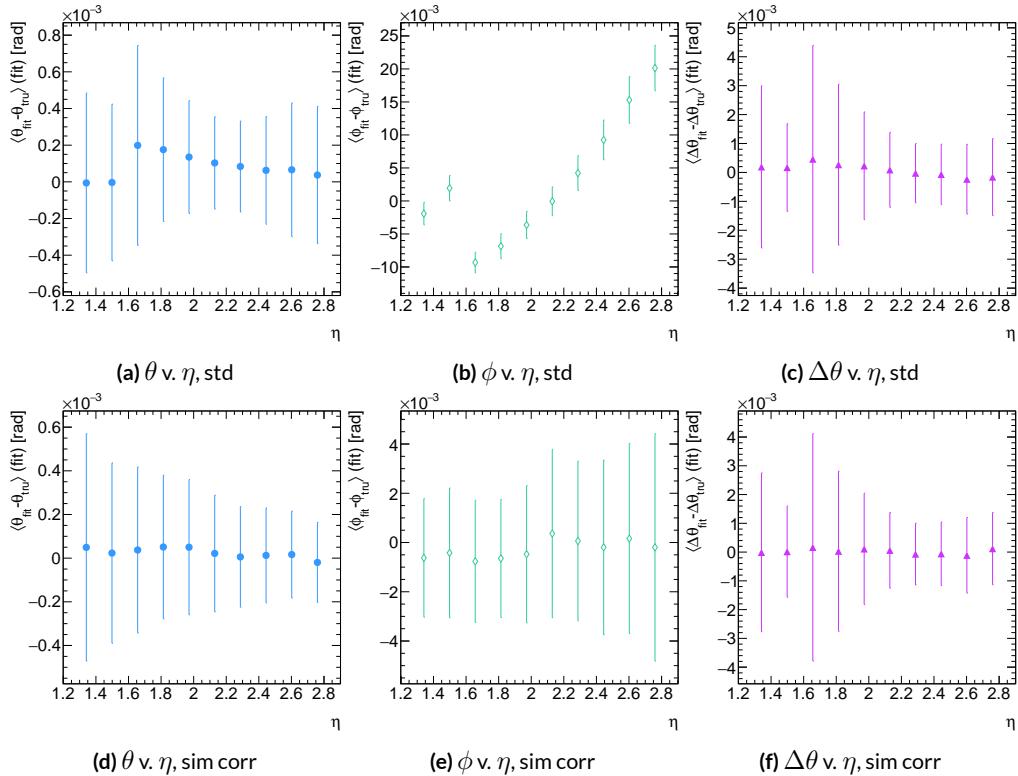


Figure A.19: Nominal residual plots as a function of η with points as means and error bars as rms values in each η bin for the angles $\theta, \phi, \Delta\theta$ for $p_T = 100$ GeV muons in the uncorrected and simulation corrected cases.

2393

2394 That the improvements from a simulation-based correction improve performance of the algo-

2395 rithm in nominal conditions most for the quantities that depend most on stereo information (ϕ and
 2396 θ) and remove the η dependence of fit quantity resolutions suggests that there could, in principle, be
 2397 analytic corrections that could be applied to the nominal algorithm. One possible solution is to in-
 2398 troduce an additional set of constants, having the y_{base} depend on the strip number, similar to the γ_s
 2399 correction for z_{plane} described in Section A.21, which would add a lookup per hit and $8 \times n_{bins,y}$ extra
 2400 constants that would be optimized as the γ_s correction was.

$$M_{hit} = \frac{\gamma}{z} = \frac{y_{base}}{z_{plane}} (n_{str}) + \frac{w_{str}}{z_{plane}} n_{str} \quad (\text{A.30})$$

2401 The simulation correction residual rms values suggest a limit on the quality of such correction
 2402 and could perhaps be implemented generically on their own regardless of misalignment for rms val-
 2403 ues on fit quantities of 0.291 mrad for θ , 3.19 mrad for ϕ , and 1.54 for $\Delta\theta$, which represent a 20%
 2404 improvement for θ , a 62% improvement for ϕ , and a slight degradation in $\Delta\theta$ of 4.7%, again owing
 2405 to an effect similar to the one in A.7.

2406 A.18 TRANSLATION MISALIGNMENTS ALONG THE HORIZONTAL STRIP DIRECTION (Δs)

2407 A translation in s (i.e. along the direction of a horizontal strip) only affects the stereo strips, and,
2408 since the stereo angle is small, a very large misalignment is necessary for effects to be noticeable (a
2409 misalignment of roughly 17 mm corresponds to one strip's misalignment in the stereo planes). The
2410 only quantity to show any meaningful deviation with misalignments with translations in s is the ϕ
2411 residual bias (a change of 0.4 mrad at $\Delta s = 1$ mm), as can be seen in the uncorrected curve of Figure
2412 A.20.

2413 A translation in s induces a constant shift in the calculated horizontal slope, m_x in Equation A.4.
2414 This constant shift should only depend on which stereo planes included in a fit are misaligned and
2415 how misaligned they are. Hence, the correction to m_x , for a sum over misaligned stereo planes i ,
2416 with their individual misalignments in s and plane positions in z is:

$$\Delta m_x = \frac{1}{N_{\text{stereo}}} \sum_{i, \text{misal stereo}} \frac{\Delta s_i}{z_{i, \text{plane}}} \quad (\text{A.31})$$

2417 Given prior knowledge of misalignment, these corrections to m_x can be performed ahead of time
2418 and saved in a lookup table (LUT), similar to the LUT used for constants in the X local slope (M_x^l)
2419 calculation. The added overhead of this analytic correction is hence eleven constants in memory, a
2420 lookup, and one addition. The correction perfectly corrects the effects of misalignment, as can be
2421 seen in Figure A.20. The simulation based correction described above can also be used to correct
2422 for Δs misalignments, with the results of that correction also shown in Figure A.20. The apparent

2423 discrepancy between the simulated and analytic correction is a natural consequence of the fact that
 2424 the simulation correction, as previously mentioned, restores the ϕ residual distribution to an overall
 2425 more Gaussian shape.

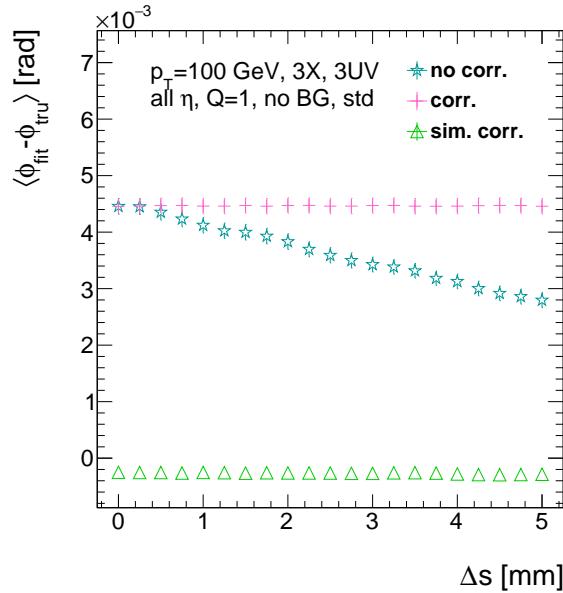


Figure A.20: The mean of the ϕ residual as a function of misalignment for the uncorrected case and the analytic and simulation correction cases.

2426 A.19 TRANSLATION MISALIGNMENTS ORTHOGONAL TO THE BEAMLINE AND HORIZON-
2427 TAL STRIP DIRECTION (Δz)

2428 A translation in AMDB z , the direction orthogonal to both the beamline and the horizontal strip
2429 direction, corresponds to a translation in the y of Equation A.1, affecting all slope calculations. This
2430 has a large impact on the θ residual bias and both the bias and rms of $\Delta\theta$ residual, as can be seen in
2431 Figures A.21 (a)–(c). The marked degradation and non-linear behavior in performance at very high
2432 levels of misalignments is a result of low statistics; there are fewer fits at high level of misalignments
2433 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 $\Delta z =$
2434 1 mm, and $\Delta\theta$ shifts by about 5 mrad for the same level of misalignment. While the fitted rms of the
2435 $\Delta\theta$ residual remains fairly stable for $\Delta z < 1$ mm or so, between $\Delta z = 2$ mm and $\Delta z = 3$ mm, the
2436 rms increases by 15% before the $\Delta\theta$ cut issue mentioned above intervenes.

2437 Fortunately, these misalignments are straightforward to correct with knowledge of the misalign-
2438 ment. The only modification necessary for this correction is to change the definitions of y_{base} in
2439 Equation A.1 for the individual hit slope addressing. This is done before runtime and adds no over-
2440 head to the algorithm, and the correction quality is only limited by knowledge of the misalignment.
2441 The results of this correction are also shown in Figures A.21 (a)–(c) and restore nominal perfor-
2442 mance.

Since $\Delta\theta = \frac{M_x^l - M_x^e}{1 + M_x^l M_x^e}$ 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

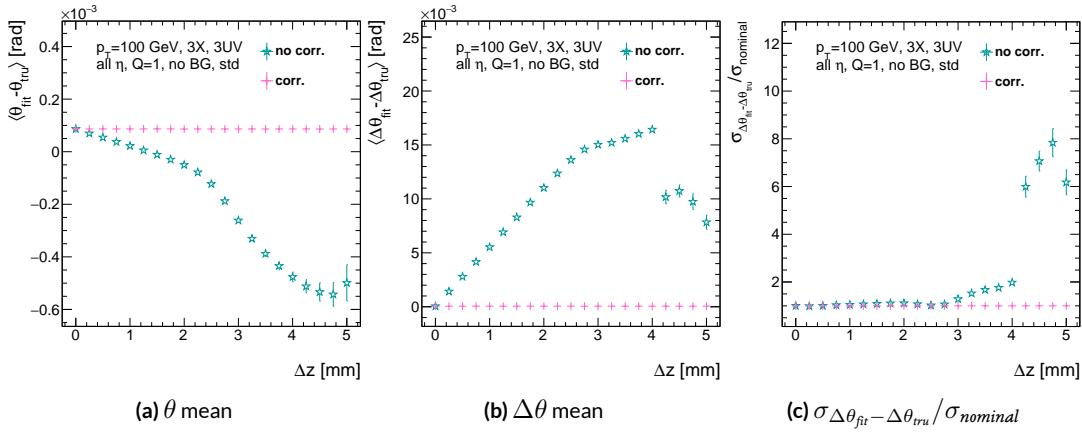


Figure A.21: The affected quantities of Δz misalignments: θ bias, $\Delta\theta$ bias, and $\sigma_{\Delta\theta_{\text{fit}} - \Delta\theta_{\text{tru}}} / \sigma_{\text{nominal}}$ for both the misaligned and corrected cases.

2443 A.20 TRANSLATION MISALIGNMENTS PARALLEL TO THE BEAMLINE (Δt)

2444 The effects of misalignment due to translations in t are very similar to those due to translations in
 2445 z without the complication of the $\Delta\theta$ cut, affecting the z instead of the y coordinate that enters
 2446 into hit slope calculations. Again, θ bias, $\Delta\theta$ bias, and $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}$ are the primarily affected quan-
 2447 tities. For $\Delta t = 1$ mm, θ bias shifts by about 0.02 mrad, $\Delta\theta$ bias shifts by just under 2 mrad, and
 2448 $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}$ degrades by about 20%. The correction for this misalignment once again costs no over-
 2449 head and consists of changing stored constants in the algorithm, in this case the positions along the
 2450 beamline of the misaligned planes, with results similarly limited by knowledge of the misalignment.
 2451 The slight improvement with correction to $\Delta\theta$ rms is due to the real effect of a larger lever arm.
 2452 Both the misaligned and corrected distributions of affected quantities of interest are shown in Fig-

ure A.22.

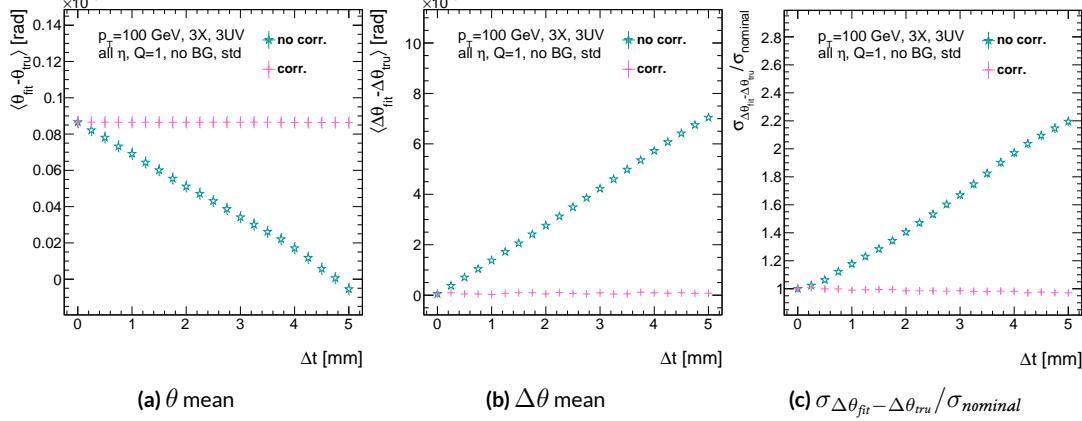


Figure A.22: The affected quantities of Δt misalignments: θ bias, $\Delta\theta$ bias, and $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}} / \sigma_{nominal}$ for both the misaligned and corrected cases.

2453

2454 A.21 CHAMBER TILTS TOWARDS AND AWAY FROM THE IP (γ_s ROTATION)

2455 Chamber misalignment due to rotations around the s axis act effectively like a translation in t that
2456 depends on strip number. These rotations tilt misaligned chambers away from (towards) the IP for
2457 positive (negative) values of γ_s . Since, unlike for the other two rotation cases that will be studied,
2458 positive and negative rotation values are not symmetric, this misalignment is studied for both posi-
2459 tive and negative γ_s values. The divergent effect at the tails is a result of a large population of fits not
2460 having fit quantities within the cores, and so not appearing in the fit rms. Once again, affected quan-
2461 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
2462 A.23 (a)–(c). The relationship between biases and γ_s is roughly linear with $\Delta\gamma_s = 0.3$ mrad (the an-
2463 gular scale corresponding to linear shifts of ~ 1 mm) corresponding to 0.005 mrad (0.12 mrad) for θ
2464 ($\Delta\theta$). For $\sigma_{\Delta\theta_{fit} - \Delta\theta_{tru}}$, degradation is not symmetric. For negative (positive) γ_s , with the quadruplet
2465 tilted towards (away from) the IP, slope-roads are artificially expanded (shrunk), decreasing (increas-
2466 ing) the granularity of the trigger, explaining the asymmetry in Figure A.23 (c), with the degradation
2467 being a 10% (25%) effect for γ_s of $+(-)0.3$ mrad.

2468 Corrections are less simple in this case. In principle, corrections of the same accuracy of the trans-
2469 lations could be calculated per strip, but the overhead of one correction per strip (many thousands
2470 of constants) is prohibitive. Instead, each plane was divided into eight equal segments with a t value
2471 (z in the slope calculation) assigned to strips in each region to correct for the misalignment. This
2472 amounts to 56 extra constants and a 2D instead of a 1D LUT for z positions while the algorithm
2473 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.

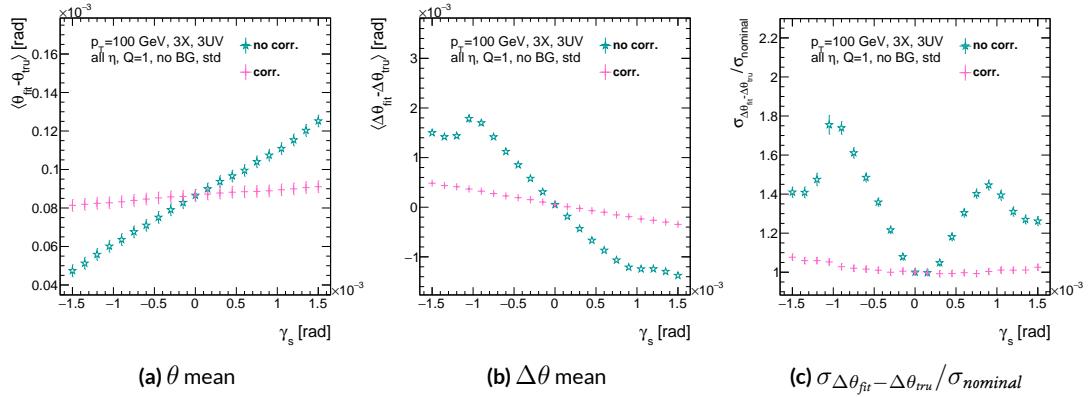


Figure A.23: The noticeable effects of rotations in the s axis and the behavior of these quantities (θ and $\Delta\theta$ bias shifts and $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}/\sigma_{nominal}$) with and without misalignment correction.

2476

2477 A.22 ROTATION MISALIGNMENTS AROUND THE WEDGE VERTICAL AXIS (β_z)

2478 While misalignments coming from rotations around the z axis (the direction orthogonal to both
2479 the beamline and the horizontal strip direction) foreshorten the strips as seen from the IP and add
2480 a deviation in t , the long lever arm largely washes out any effects of this misalignment. Only the
2481 $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$ is noticeably affected, though only at severe misalignments, with only about a 1% degra-
2482 dation in performance at $\beta_z = 0.3$ mrad (corresponding to a linear shift of ~ 1 mm). A simulation
2483 based correction works well to cancel out the effects of this misalignment, and the $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$ as a
2484 function of misalignment with and without corrections are shown in Figure A.24. The apparent
2485 2% effect in the simulation corrected curve is a result of a more mild version of the effect shown in
2486 Figure A.7.

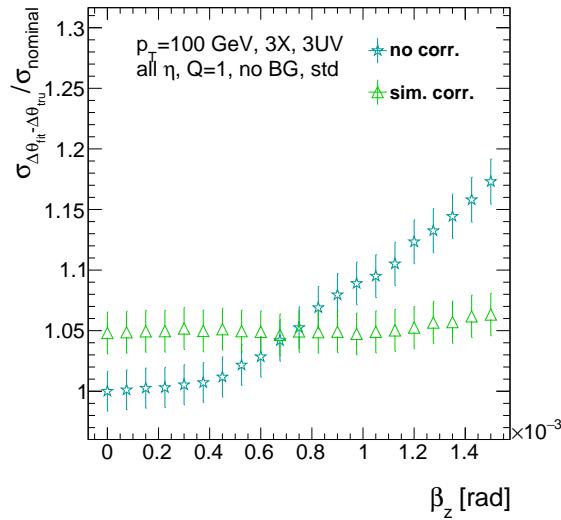


Figure A.24: The effects of rotations in the z axis on $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}} / \sigma_{nominal}$ a function of β_z both with and without misalignment corrections.

2487 A.23 ROTATION MISALIGNMENTS AROUND THE AXIS PARALLEL TO THE BEAMLINE (α_t)

2488 Misalignments arising from rotations around the t axis (parallel to the beamline at the center of
2489 the base of the wedge) are essentially rotations in the ϕ direction. The quantities of interest most
2490 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-
2491 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
2492 (corresponding to a linear shift of ~ 1 mm). The raw instead of fitted mean ϕ biases is used in Fig-
2493 ure A.25 (a) to better illustrate the effect of misalignment.

2494 Since the effect of misalignment is dependent on horizontal (along the strip direction, \hat{s}) in addi-
2495 tion to vertical information, corrections cannot be applied before a fit takes place. The ϕ bias shift is
2496 uniform over the entire wedge, so a constant additive correction to ϕ based on the level of misalign-
2497 ment can be applied to all fits depending on how many misaligned stereo planes enter in the fit. $\Delta\theta$
2498 is less straightforward, but corrections to the y and z information used in the local slope calculation
2499 in Equation A.4 can be applied once θ_{fit} and ϕ_{fit} are known. These corrections are calculated ahead
2500 of time in bins of uniform η and ϕ as with the simulation corrections using the same framework
2501 as the misalignment calculation. The results of both types of correction can be seen in Figure A.22.
2502 The apparent discrepancy between the simulation and analytic corrections in the ϕ bias happens for
2503 the same reason as in the Δs misalignment correction cases, as simulation correction restores a more
2504 Gaussian shape to the ϕ residual distribution opposed to the uncorrected nominal case, as discussed
2505 in Section A.17.

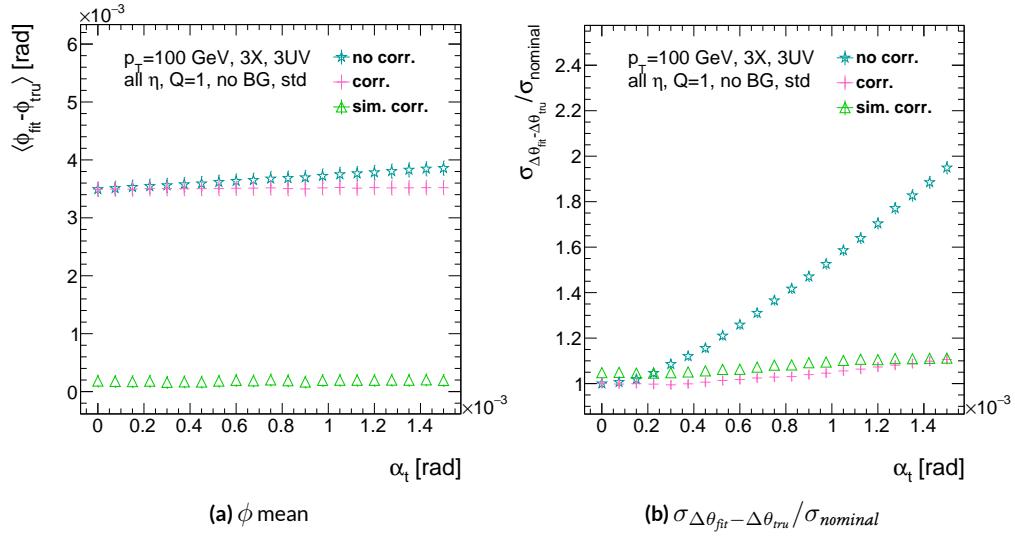


Figure A.25: The effects of rotation misalignments around the t axis for ϕ bias and $\sigma_{\Delta\theta_{\text{fit}} - \Delta\theta_{\text{true}}} / \sigma_{\text{nominal}}$ as a function of misalignment. The uncorrected and both the analytic and simulation correction cases are shown.

2506 A.24 CONCLUSION

2507 The algorithm for Micromegas detectors in the NSW Trigger Processor performs well in a variety of
2508 conditions and has proven robust to a number of effects to deliver measurements on muon tracks
2509 of the three angles θ , ϕ , $\Delta\theta$. Under nominal conditions, the rms values for the residuals of these
2510 quantities are 0.364 mrad for θ , 8.12 mrad for ϕ , and 1.47 mrad for $\Delta\theta$. Algorithm performance was
2511 found to be largely independent of the charge threshold setting, and a hit majority BCID associa-
2512 tion was found to provide proper timing information over 99.7% even in the most relaxed settings
2513 (2X+1UV coincidence threshold requirement+wide slope-road+background). The introduction of
2514 wide slope-roads to better mimic potentially limited algorithm resources at run time and the intro-
2515 duction of incoherent background was found to have a manageable effect on fit quantity residual
2516 rms values and on total algorithm efficiency for sufficiently stringent coincidence threshold. The ef-
2517 ffects of the three translation and three rotation misalignments specified by AMDB convention were
2518 studied, and correction methods for each of the six cases was developed. Simulation-based correc-
2519 tions were found to improve nominal algorithm performance to residual rms value of 0.291 mrad for
2520 θ , 3.19 mrad for ϕ , and 1.54 for $\Delta\theta$, which represent improvements of 20%, 62%, and -4.7%, respec-
2521 tively. Misalignment corrections were found to restore nominal performance for all but the rotation
2522 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.

Tod-Not-Brot

Old German Proverb

2523

B

2524

Telescoping Jets

2525 ANOTHER APPROACH TO IMPROVING $ZH \rightarrow \ell\ell b\bar{b}$ is the use of telescoping jets³⁶, which har-
2526 nesses the power of multiple event interpretations. The use of multiple event interpretations was
2527 originally developed with non-deterministic jet algorithms like the Q-jets (“quantum” jets) algo-
2528 rithm⁶⁹. When a traditional or “classical” algorithm, such as the Cambridge-Aachen⁸⁰ and anti- k_t ⁶³

256

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. ⁴⁷. 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.

B.I MONTE CARLO SIMULATION

The MC simulated samples used in this study are the same as in Ref.[?]. The signal sample used is generated in PYTHIA8⁷⁶ with the CTEQ6L1 parton distributions functions (PDFs) and AU2 tune^{55,2,3} for the ZH process with $m_H = 125$ GeV (henceforth, $ZH125$). 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 generator⁷⁸.

2549 B.2 JET RECONSTRUCTION AND CALIBRATION

2550 In order to construct telescoping jets, jet axes must first be found around which to “telescope.” In
2551 the reconstructed-level analysis, the anti- k_t algorithm with $R = 0.4$ is used to reconstruct jets from
2552 topological clusters in the calorimeters. The four vectors of these anti- k_t algorithm with $R = 0.4$ jets
2553 are calibrated to match truth information obtained from simulation and validated in data. To take
2554 into account the effect of pile-up interactions, jet energies are corrected using a jet-area based tech-
2555 nique³², and each jet with $p_T < 50$ GeV and $|\eta| < 2.4$ is subject to a requirement that at least 50% of
2556 the scalar sum of the p_T of tracks matched to this jet be composed of tracks also associated with the
2557 primary vertex. Jet energies are also calibrated using p_T and η -dependent correction factors¹¹. Fur-
2558 thermore, at least two jets must have $|\eta| < 2.5$ in order to be b -tagged. The MV1 algorithm^{42???}
2559 is used for b -tagging. Once jets are reconstructed and b -tag weights have been calculated, the two
2560 hardest, b -tagged jets are used as the telescoping jet axes. Additional details can be found in Ref.¹².

2561 After the telescoping jet axes have been established, telescoping jets are constructed using topolog-
2562 ical clusters in the calorimeters at a variety of jet cone sizes. Including the original anti- k_t jets used for
2563 the $R = 0.4$ case, twelve total sets of jets of cone sizes ranging from $R = 0.4\text{--}1.5$ are constructed,
2564 with each successive size having a radius 0.1 larger than the preceding set. For each jet axis, telescop-
2565 ing jets consist of any topological cluster lying within R of the axis. In the event of overlap, clusters
2566 are assigned to the closer jet axis. If a given cluster is equidistant from the two jet axes, the cluster
2567 is assigned to whichever jet axis is associated with the anti- k_t jet with higher p_T . Calibration for the
2568 telescoping jets is conducted using corrections for anti- k_t calorimeter topological cluster jets; the

2569 $R = 0.4$ corrections are used for telescoping $R = 0.5$, and the $R = 0.6$ corrections are used
2570 for telescoping $R \geq 0.6$ (cf. Sec. B.4). The telescoping cone jets ($R \geq 0.5$) at reconstructed level
2571 are trimmed using Cambridge-Aachen jets with $R = 0.3$ and $f_{cut} = 0.05$ with respect to the
2572 untrimmed jet p_T^{48} . Since these jets are trimmed, the active area correction is not applied. In the
2573 event a Z candidate electron falls within R of the axis of a telescoping jet, its 4-momentum is sub-
2574 tracted from that of the jet vectorially.

2575 A similar process is used to construct telescoping jets in the truth-level analysis below. Instead of
2576 the two hardest b -tagged anti- k_t with $R = 0.4$ jets reconstructed with calorimeter topological clus-
2577 ters, the two hardest truth b -jets in an event are used. Instead of making a cut on b -tagging weight
2578 to b -tag, truth jets are examined to see whether a b -hadron with $p_T > 5$ GeV is contained within
2579 $\Delta R < 0.4$ of the jet axis; the presence of a b -hadron is used to b -tag truth-level jets. These two jets
2580 again provide the jets for the $R = 0.4$ case and the axes around which telescoping takes place. Stable
2581 truth particles, not including muons and neutrinos, are used in place of calorimeter topological clus-
2582 ters. Z candidate electron-telescoping jet overlap removal is performed at truth level, too. Missing E_T
2583 is calculated using the vector sum of the four momenta of stable truth-level neutrinos. Since there
2584 are no pileup particles stored at truth level, truth-level telescoping jets are not trimmed.

2585 B.3 EVENT RECONSTRUCTION AND SELECTION

2586 Events are selected on the basis of a combination of leptonic, jet, and missing E_T requirements,
2587 which are outlined in Table B.1. Leptons are categorized by three sets of increasingly stringent qual-
2588 ity requirements, which include lower limits on E_T , upper limits on $|\eta|$, impact-parameter require-

2589 ments, and track-based isolation criteria. The requirements differ for electrons⁵ and muons¹. Events
2590 are selected with a combination of single lepton, dielectron, and dimuon requirements. Each event
2591 must contain at least one lepton passing medium requirements and at least one other lepton pass-
2592 ing loose requirements. These leptons are used to create a dilepton invariant mass cut to ensure the
2593 presence of a Z boson and suppress multijet backgrounds.

2594 Event selection requirements are also imposed on the anti- k_t with $R = 0.4$ jets. There must be at
2595 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
2596 the second b -tagged jet must have p_T of at least 20 GeV. There are further topological cuts on the
2597 separation of the two jets $\Delta R(b, \bar{b})$, the distance between the two jets in the (η, ϕ) plane, according
2598 to the transverse momentum of the Z boson, p_T^Z . These are shown in Table B.2.

2599 The truth-level analysis has the same missing E_T , jet p_T , m_{ll} , and additional topological selection
2600 criteria, but the use of truth-level information simplifies the other requirements. Instead of lepton
2601 quality requirements, Z boson candidate leptons' statuses and MC record barcodes are checked to
2602 ensure the leptons are stable.

2603 In the jet calibration validation, the reconstructed level analysis lepton and m_{ll} requirements are
2604 imposed, but neither the missing E_T nor the jet selection requirements are applied so as not to bias
2605 the validation.

2606 B.4 VALIDATION OF JET CALIBRATION

2607 In order to validate the jet energy scale and resolution of jets constructed with this telescoping-jets
2608 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)		$> 45 \text{ GeV} (> 20) \text{ GeV}$	—
Missing E_T		$E_T^{\text{miss}} < 60 \text{ GeV}$	—
Z boson		$83 < m_{ll} < 99 \text{ GeV}$	

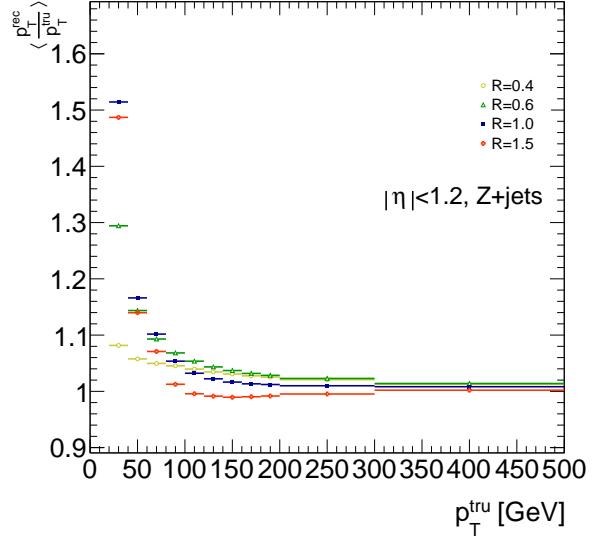
Table B.2: Topological requirements of the event selection.

$p_T^Z [\text{GeV}]$	$\Delta R(b, b)$
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

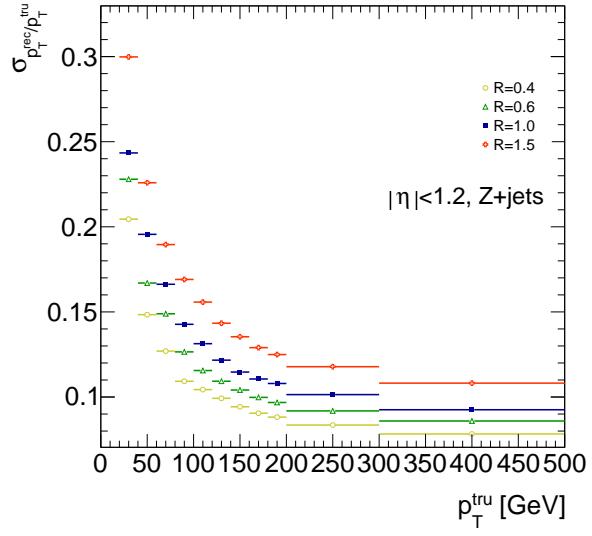
2609 event, all jets, not just the two hardest b -tagged jets, are telescoped. These jets are constructed in the
2610 same way as in the reconstructed and truth-level analyses; reconstructed-level jets are made from
2611 calorimeter topological clusters within R of the anti- k_t with $R = 0.4$ jet axes and then trimmed,
2612 and truth-level jets are made from stable truth particles within R of the anti- k_t with $R = 0.4$ jet axes.
2613 The reconstructed and truth-level telescoping jet ensembles are matched according to the separation
2614 in the (η, ϕ) plane of their corresponding anti- k_t with $R = 0.4$ jets used as seeds. Only jets with
2615 $|\eta| < 1.2$ are examined here, and the results of studies on the $ZH125$, ZZ , and $t\bar{t}$ samples, as well as
2616 over other $|\eta|$ ranges, are outlined in³⁶. Any reconstructed jets not within $\Delta R = 0.3$ of a truth jet
2617 are discarded. In the event that multiple reconstructed jets are the same distance away from a given
2618 truth jet, the reconstructed jet with the highest p_T gets matched. Matching is retained for all R values
2619 (i.e. telescoping seeds are matched, and telescoping jets are assumed to match if the anti- k_t jets from
2620 which their seeds are derived match).

2621 Once anti- k_t with $R = 0.4$ reconstructed and truth jets are matched, response functions are cre-
2622 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
2623 bin for 200–300 GeV, and one bin for 300–500 GeV for each R , with bins chosen for purposes of
2624 statistics. Ensembles with $p_T^{tru} < 20$ GeV are ignored since no calibration exists for jets with trans-
2625 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
2626 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,
2627 and the error on the mean is taken from the error of this parameter in the fit. The resolutions are the
2628 values of the square root of the variance on this fit. As the total response distributions in Figure B.1
2629 show, performance is best for low R values and high values of p_T^{tru} . Figure B.1 shows the $R = 0.4$

2630 (anti k_t) case to show a baseline for performance, $R = 0.6$ to show the deviations with “correct”
2631 calibrations, and $R = 1.0, 1.5$ to show how big those deviations get with larger R jets. The resolu-
2632 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
2633 consistent over various R values. Resolution, as might naïvely be expected, is worse for increasingly
2634 larger values of R . For $p_T^{tru} < 60$ GeV, resolution degrades, and response degrades in particular for
2635 increasing R ; this is likely a result from residual pileup effects.



(a)



(b)

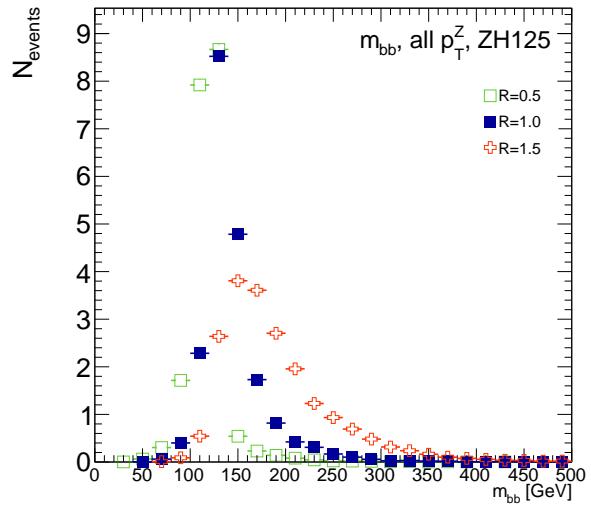
Figure B.1: The mean and resolution of p_T^{rec}/p_T^{tru} for the background $Z+jets$ sample for $|\eta| < 1.2$ and for $R = 0.4, 0.6, 1.0$, and 1.5 in 20 GeV bins of p_T^{tru} for $20\text{--}200\text{ GeV}$, one bin for $200\text{--}300\text{ GeV}$, and one bin for $300\text{--}500\text{ GeV}$, with bins chosen for purposes of statistics.

2636 B.5 TRUTH-LEVEL ANALYSIS

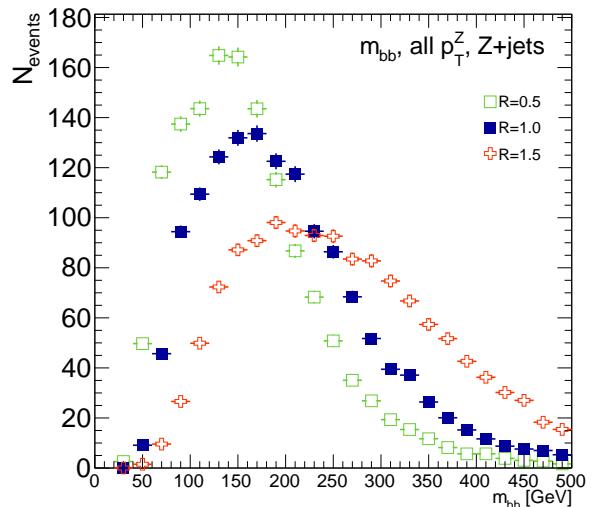
2637 To understand the limits and sources of any potential improvements, a truth-level analysis was con-
2638 ducted on MC samples with a ZH_{125} signal sample and a $Z+jets$ background sample. Distributions
2639 for the dijet invariant mass, m_{bb} , were made for each telescoping radius. Both signal and background
2640 samples develop more pronounced tails in the high m_{bb} region as R increases, as shown in Figure
2641 B.2. N_{events} is normalized to expected values in data.

2642 One way to take advantage of this information is to make a cut on m_{bb} for two different radii.
2643 This is graphically depicted in Figure B.3 for the optimized combination of $m_{bb,R=0.9}$ (telescoping
2644 cone jets constructed as outlined in Sec. B.2) vs. $m_{bb,R=0.4}$ (anti- k_t jets). At truth-level, the majority
2645 of events in the signal ZH_{125} sample are concentrated in relatively narrow region of parameter space,
2646 where this is certainly not the case for the more diffuse $Z+jets$ background sample.

2647 Another way to take advantage of multiple event interpretations is to make use of an event's cut-
2648 weight, denoted z and defined as the fraction of interpretations in a given event that pass a certain set
2649 of cuts (in this note, a cut on m_{bb}). The distribution of cut-weights for a sample of events is denoted
2650 $\rho(z)$. To enhance the significance of a cut-based analysis, events can be weighted by the cut-weight
2651 or any function $t(z)$ of the cut-weight. Weighting events by $t(z)$ modifies the usual $S/\delta B$ formula
2652 used to calculate significances. In this note, δB is based on Poissonian statistics and is taken as $0.5 +$
2653 $\sqrt{0.25 + N_B}$, where N_B is the number of background events.

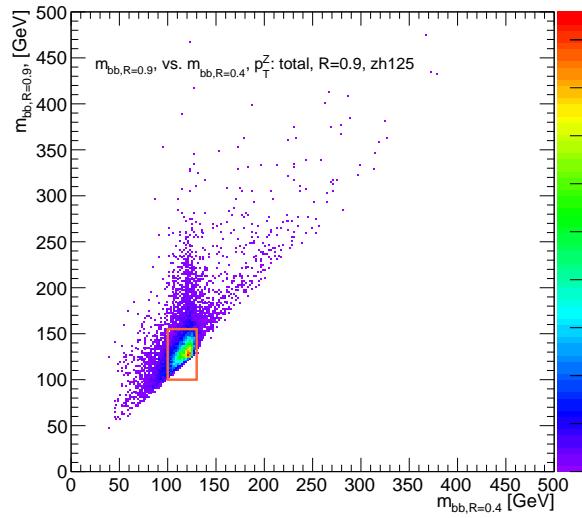


(a)

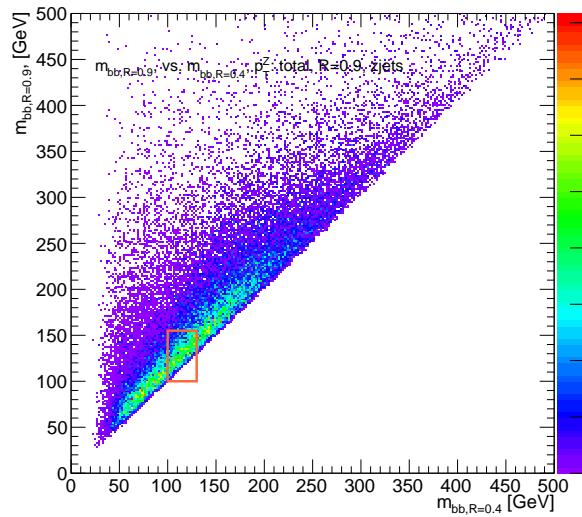


(b)

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.

2654 B.6 ERRORS ON TELESCOPING SIGNIFICANCES

2655 Significances of measurements are quoted in units of expected background fluctuations, schemati-
2656 cally, $S/\delta B$. For counting experiments with high numbers of events, we can use Gaussian statistics
2657 and express this as S/\sqrt{B} , which we here denote as \mathcal{S} . However, with lower statistics, it becomes
2658 more appropriate to use Poissonian statistics, and

$$\mathcal{S}_{gaus} = \frac{S}{\sqrt{B}} \rightarrow \mathcal{S}_{pois} = \frac{S}{0.5 + \sqrt{0.25 + B}} \quad (\text{B.1})$$

2659 where $0.5 + \sqrt{0.25 + B}$ is the characteristic upward fluctuation expected in a Poissonian data set
2660 using the Pearson chi-square test⁷⁸.

2661 B.7 COUNTING

2662 The significance is given as above, where $S = N_S$ and $B = N_B$. That is, the signal and background
2663 are just the number of events in signal and background that pass some cuts. The error for the Guas-
2664 sian case is the standard:

$$\Delta \mathcal{S}_{gaus} = \frac{1}{\sqrt{B}} \Delta S \oplus \frac{S}{2B^{3/2}} \Delta B \quad (\text{B.2})$$

2665 The error for the Poissonian case is:

$$\Delta \mathcal{S}_{pois} = \frac{1}{0.5 + \sqrt{0.25 + B}} \Delta S \oplus \frac{S}{2(0.5 + \sqrt{0.25 + B})^2 \sqrt{0.25 + B}} \Delta B \quad (\text{B.3})$$

2666 where \oplus denotes addition in quadrature, and $\Delta S(B)$ is the error on signal (background).

2667 **B.8 MULTIPLE EVENT INTERPRETATIONS**

2668 Using multiple event interpretations changes the formulae used in with simple counting. That is, S

2669 is not necessarily merely N_S , the number of events passing some signal cuts, and similarly for B and

2670 N_B . Using an event weighting by some function of the cut-weight, z , denoted $t(z)$, $S = N_S \langle t \rangle_{\rho_S}$

2671 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 telescoping 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,

$$N = \sum_{i=0}^{n_{tel}} \rho_i, \quad \int_0^1 dz t(z) \rho_S(z) = \sum_{i=0}^{n_{tel}} t_i \rho_{S,i}, \quad \int_0^1 dz t^2(z) \rho_B(z) = \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}$$

2672 For the calculations that follow, let $\xi = \sum_{i=0}^{n_{tel}} t_i \rho_{S,i}$, $\psi = 0.5 + \sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}$,

2673 $\partial_S = \frac{\partial}{\partial \rho_{S,i}}$ (and similarly for B), so $\mathcal{S}_t = N_S \xi / \psi$

²⁶⁷⁴ Some partial derivatives:

$$\begin{aligned}
 \partial_S N_S &= 1, & \partial_{B,i} N_B &= 1 \\
 \partial_S \xi &= t_i + (\partial_S t_i) \rho_{S,i}, & \partial_B \xi &= (\partial_B t_i) \rho_{B,i} \\
 \partial_S \psi &= \frac{N_B t_i (\partial_S t_i) \rho_{B,i}}{\sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}}, & \partial_B \psi &= \frac{\sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i} + N_B (t_i^2 + 2t_i (\partial_B t_i) \rho_{B,i})}{2\sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}} \\
 \partial_S \mathcal{S}_t &= \frac{\xi}{\psi} + \frac{N_S}{\psi} \partial_S \xi - \frac{N_S \xi}{\psi^2} \partial_S \psi, & \partial_B \mathcal{S}_t &= N_S \left(\frac{1}{\psi} \partial_B \xi - \frac{\xi}{\psi^2} \partial_B \psi \right)
 \end{aligned}$$

²⁶⁷⁵ Thus,

$$\Delta \mathcal{S}_{t,i} = \left[\frac{\xi}{\psi} + \frac{N_S}{\psi} \partial_S \xi - \frac{N_S \xi}{\psi^2} \partial_S \psi \right] \Delta \rho_{S,i} \oplus N_S \left[\frac{1}{\psi} \partial_B \xi - \frac{\xi}{\psi^2} \partial_B \psi \right] \Delta \rho_{B,i} \quad (\text{B.5})$$

²⁶⁷⁶ and the total error is given by the sum in quadrature over all bins i of $\Delta \mathcal{S}_{t,i}$.

²⁶⁷⁷ B.9 $t(z) = z$

²⁶⁷⁸ With $t(z) = z$, $t_i = i/n_{tel}$, so $\partial_S t_i = \partial_B t_i = 0$. So:

$$\begin{aligned}
 \partial_S \psi &= \partial_B \xi = 0 \\
 \partial_S \xi &= \frac{i}{n_{tel}} \\
 \partial_B \psi &= \frac{\sum_i i^2 \rho_{B,i} + N_B t^2}{n_{tel} \sqrt{n_{tel}^2 + N_B \sum_i i^2 \rho_{B,i}}}
 \end{aligned}$$

²⁶⁷⁹ so $\Delta\mathcal{S}_{z,i}$ reduces to

$$\Delta\mathcal{S}_{t,i} = \left[\frac{\xi + N_S t_i}{\psi} \right] \Delta\rho_{S,i} \oplus \left[\frac{N_S \xi}{\psi^2} \partial_B \psi \right] \Delta\rho_{B,i} \quad (\text{B.6})$$

²⁶⁸⁰ B.10 $t(z) = \rho_S(z) / \rho_B(z)$

²⁶⁸¹ 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 $\partial_B t_i =$

²⁶⁸² $-\rho_{S,i} / \rho_{B,i}^2$. So:

$$\begin{aligned} \partial_S \xi &= 2 \frac{\rho_{S,i}}{\rho_{B,i}} = 2t_i \\ \partial_B \xi &= -\frac{\rho_{S,i}}{\rho_{B,i}} = -t_i \\ \partial_S \psi &= \frac{N_B t_i}{\sqrt{0.25 + N_B \sum_i \rho_{S,i}^2 / \rho_{B,i}}} \\ \partial_B \psi &= \frac{\sum_i \rho_{S,i}^2 / \rho_{B,i} - N_B (\rho_{S,i} / \rho_{B,i})^2}{\sqrt{1 + 4N_B \sum_i \rho_{S,i}^2 / \rho_{B,i}}} \end{aligned}$$

²⁶⁸³ simplifying somewhat the terms in the per bin error in Equation B.6.

²⁶⁸⁴ The new significance figure using multiple event interpretations becomes, with ρ_S and ρ_B denot-
²⁶⁸⁵ 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

2686 Of particular interest is the likelihood optimized $t(z)$,[†] $t^*(z) = \rho_S(z)/\rho_B(z)$. m_{bb} windows are
 2687 chosen separately for each scheme studied to maximize total significances and are summarized in
 2688 Table B.3.

$$\left(\frac{S}{\delta B}\right)_z = \frac{N_S \epsilon_S}{0.5 + \sqrt{0.25 + N_B (\epsilon_B^2 + \sigma_B^2)}} \quad (B.8)$$

2689

$$\left(\frac{S}{\delta B}\right)_{t^*(z)} = \frac{N_S \int_0^1 dz \frac{\rho_S^2(z)}{\rho_B(z)}}{0.5 + \sqrt{0.25 + N_B \int_0^1 dz \frac{\rho_S^2(z)}{\rho_B(z)}}} \quad (B.9)$$

2690 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
 2691 Refs.^{36,47} 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	$\text{anti-}k_t R = 0.4$ $t(z) = z$ $t(z) = \rho_S(z)/\rho_B(z)$ $\text{anti-}k_t R = 0.4$, telescoping $R = 0.6$	$90\text{--}140 \text{ GeV}$ $110\text{--}155 \text{ GeV}$ $110\text{--}155 \text{ GeV}$ $95\text{--}140 \text{ GeV} (R = 0.4), 105\text{--}160 \text{ GeV} (R = 0.6)$
Truth	$\text{anti-}k_t R = 0.4$ $t(z) = z$ $t(z) = \rho_S(z)/\rho_B(z)$ $\text{anti-}k_t R = 0.4$, telescoping $R = 0.9$	$100\text{--}130 \text{ GeV}$ $115\text{--}140 \text{ GeV}$ $120\text{--}135 \text{ GeV}$ $100\text{--}130 \text{ GeV} (R = 0.4), 100\text{--}155 \text{ GeV} (R = 0.9)$

2692 The truth-level distributions $\rho_S(z)$, $\rho_B(z)$, and $\rho_S(z)/\rho_B(z)$ are shown for the m_{bb} window
 2693 that optimizes $(S/\delta B)_{t^*(z)}$ in Figure B.4, and significance improvements as a function of p_T^Z are
 2694 summarized in Figure B.5. Uncertainties in Figures B.5 and B.9 are statistical uncertainties. JES sys-

[†]Derived under the assumption of Gaussian statistics in Ref⁴⁷

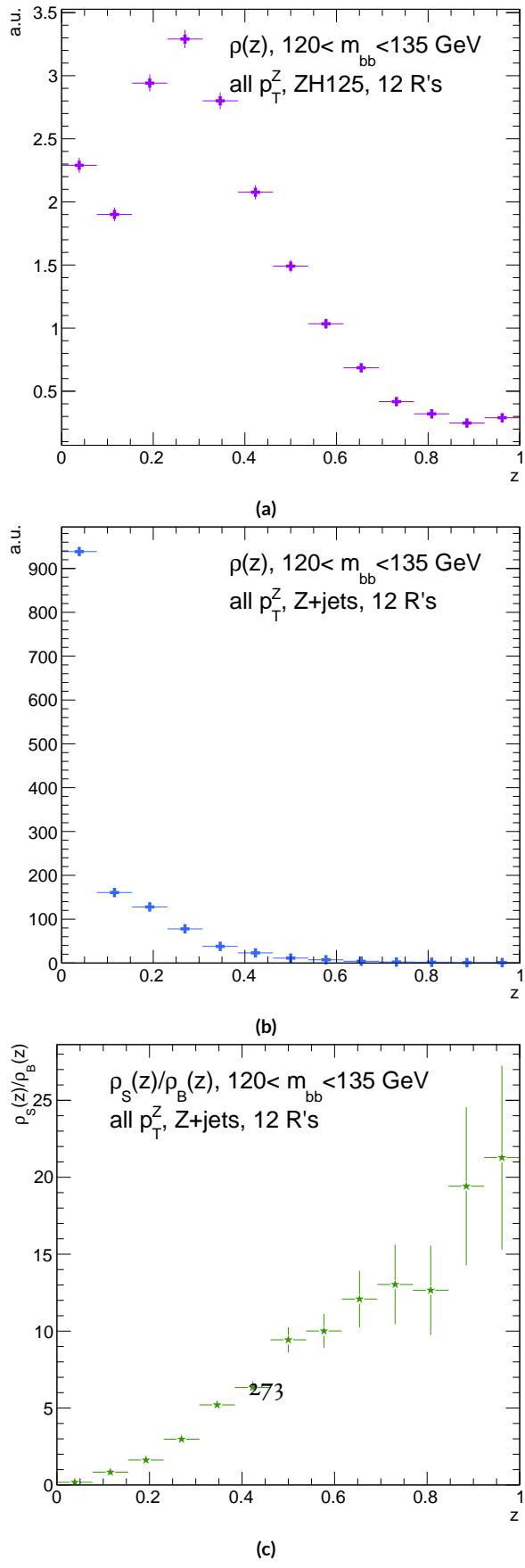


Figure B.4: Truth-level $\rho(z)$ distributions for the m_{bb} window optimizing $(S/\delta B)_{t^*(z)}$. $\rho_S(z)$ for the signal $ZH125$ sample is shown in (a), and $\rho_B(z)$ for the background $Z+jets$ sample is shown in (b). The distribution of $\rho_S(z) / \rho_B(z)$

tematics 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 increase 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.³⁹.

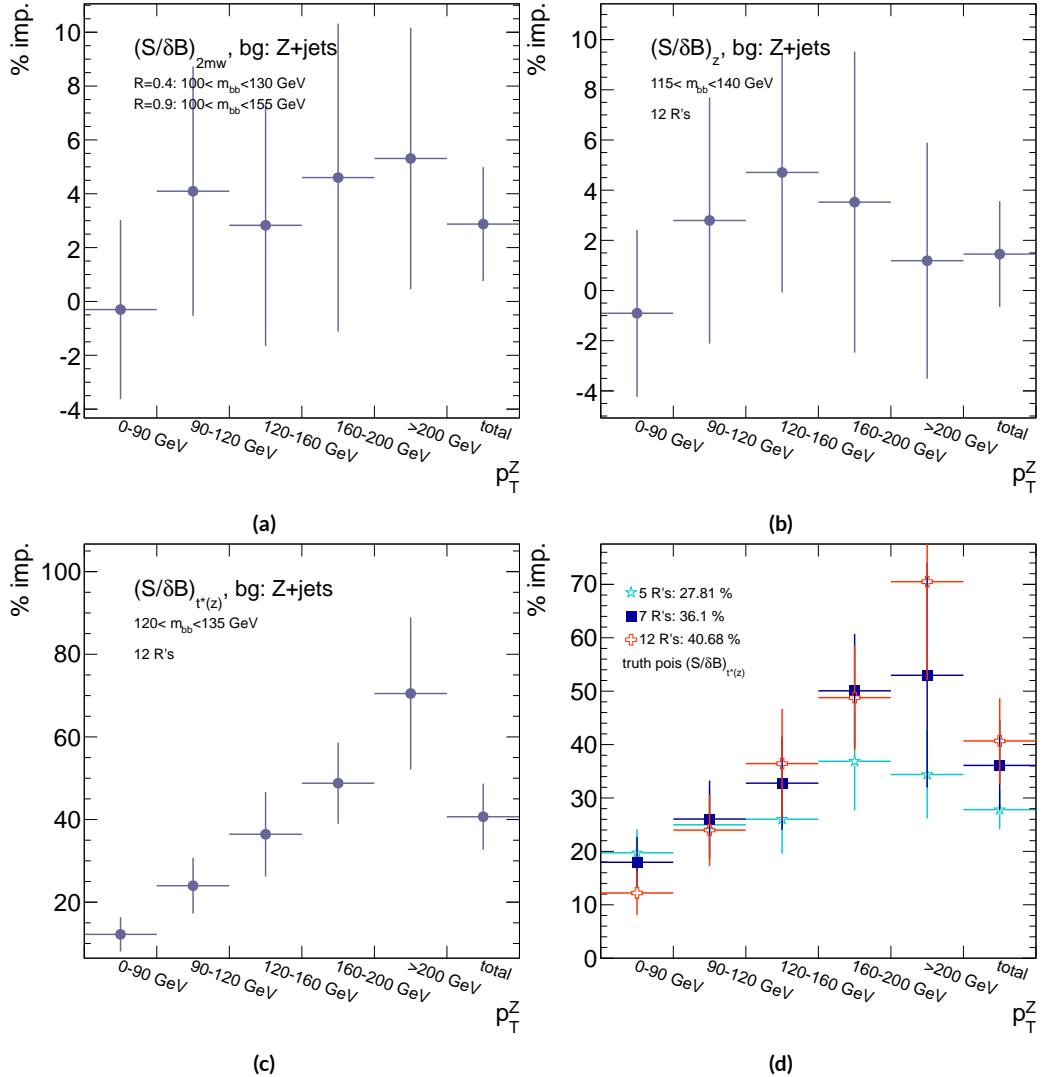


Figure B.5: A summary of the improvements for different truth-level telescoping jet cuts and weights is shown in bins of p_T^Z . The final bin is the total improvement over all p_T^Z . Shown are improvements for the 2D m_{bb} cut (a), $t(z) = z$ (b), $t(z) = t^*(z)$ with 12 radii (c), and $t(z) = t^*(z)$ for various radii (d).

2710 B.II RECONSTRUCTED-LEVEL ANALYSIS

2711 At reconstructed level, the same overall effect of introducing a high tail in m_{bb} distributions with
2712 increasing R is evident in comparing Figures B.2 and B.6. The optimal m_{bb} windows, however, grow
2713 larger, due to the lack of truth-level information.

2714 Total significance gains at reconstructed level for the two dimensional m_{bb} cut and the $t(z) = z$
2715 case are similar, at 2.87% and 1.45%, respectively. The optimal two-dimensional m_{bb} cut at recon-
2716 structed level is $95 < m_{bb,R=0.4} < 140 \text{ GeV}$, $105 < m_{bb,R=0.6} < 160 \text{ GeV}$. Just as at truth level,
2717 the $R = 0.4$ m_{bb} cut is comparable to the optimal single $R = 0.4$ m_{bb} cut, and the second m_{bb} cut is
2718 at similar values (cf. Table B.3 and Figures B.3 and B.7). However, the optimal second telescoping ra-
2719 dius is markedly smaller at $R = 0.6$ versus the optimal truth-level second radius of $R = 0.9$, which
2720 suggests that effects like pileup at reconstructed level obscure correlations between the $R = 0.4$
2721 interpretations and limit the usefulness of larger R interpretations in this particular scheme. The
2722 $t(z) = z$ case has a wider optimal window and yields about half the improvement it does at truth
2723 level.

2724 The optimal m_{bb} window for the $t^*(z)$ case is also markedly wider at reconstructed level, at $110 <$
2725 $m_{bb} < 155 \text{ GeV}$ in comparison to the truth-level optimal $120 < m_{bb} < 135 \text{ GeV}$. The $\rho(z)$ dis-
2726 tributions for the signal $ZH125$ and background $Z+\text{jets}$ as well as the $\rho_S(z) / \rho_B(z)$ in this window
2727 are shown in Figure B.8. Compared with the truth-level distributions in Figure B.4, both the sig-
2728 nal and background optimal $\rho(z)$ distributions have higher values at higher z . The peak in $\rho_S(z)$ at
2729 $z = 1$ suggests that at reconstructed level, maximizing the number of more “signal-like” events is

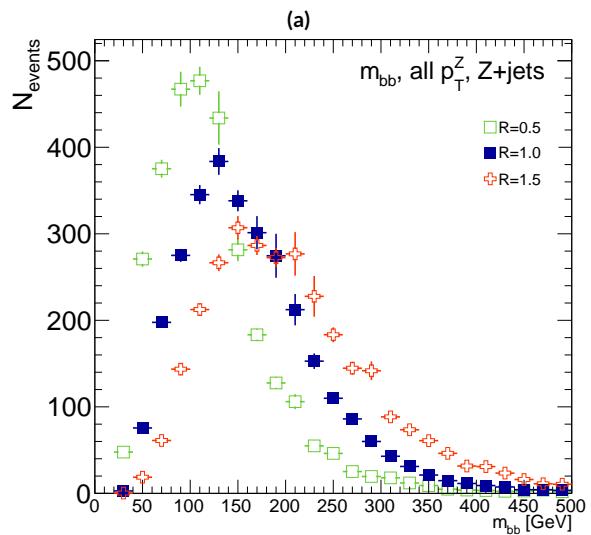
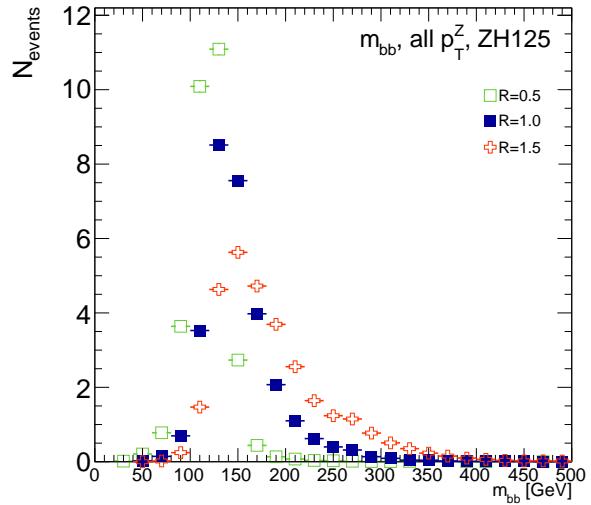


Figure B.6: The m_{bb} distribution for the telescoping jets with $R = 0.5$, $R = 1.0$, and $R = 1.5$ reconstructed-level jets is shown for the signal and background samples in (a) and (b), respectively.

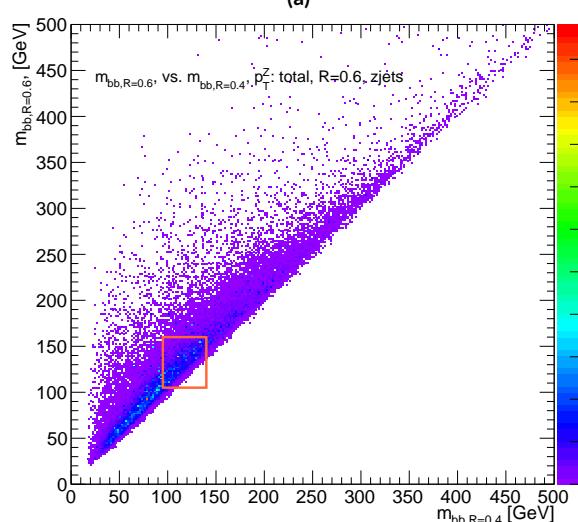
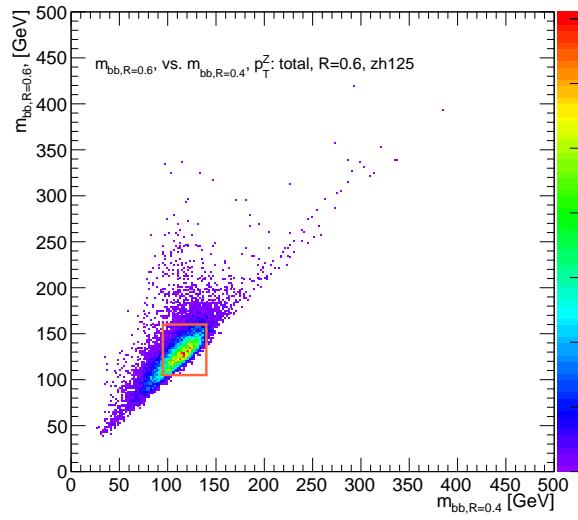


Figure B.7: The 2D distribution of $m_{bb,R=0.8}$ vs. $m_{bb,R=0.4}$ is shown for signal and background reconstructed-level samples in (a) and (b), respectively. The region chosen for the double m_{bb} cut is outlined in orange.

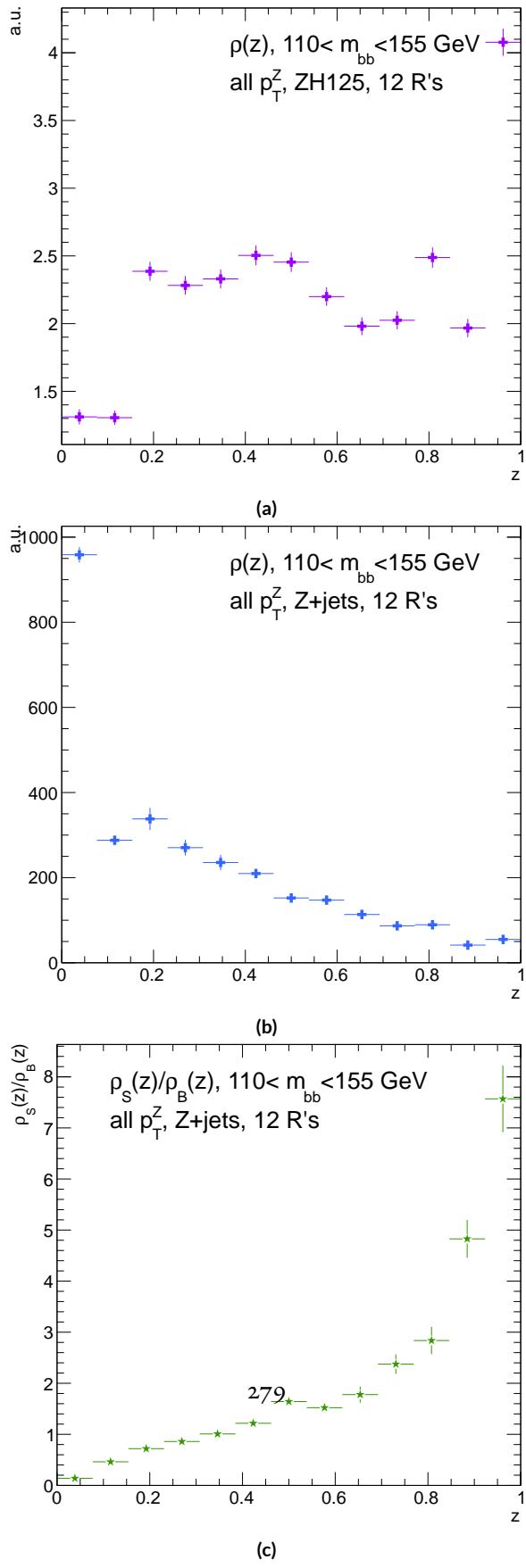


Figure B.8: Reconstructed-level $\rho(z)$ distributions for the m_{bb} window optimizing $(S/\delta B)_{t^*(z)} \cdot \rho_S(z)$ for the signal ZH125 sample is shown in (a), and $\rho_B(z)$ for the background Z+jets sample is shown in (b). The distribution of $\rho_s(z)/\rho_B(z)$ for these samples is shown in (c).

2730 the key to optimizing significances, as opposed to the optimal, background suppressing $\rho(z)$ distri-
 2731 butions at truth level. The use of a greater number of interpretations per event (telescoping radii)
 2732 does appear to result in overall greater improvement as at truth level, as twelve radii performed bet-
 2733 ter than five, but this is less clear at reconstructed level, as shown in Figure B.9 (d). The improve-
 2734 ment at reconstructed level using an event weight of $t^*(z)$ is 20.5%, just over half the improvement
 2735 at truth level but still quite significant. Summaries of improvements as a function of p_T^Z for all three
 2736 cases studied and for the $t^*(z)$ case for different numbers of telescoping radii are shown in Figure
 2737 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

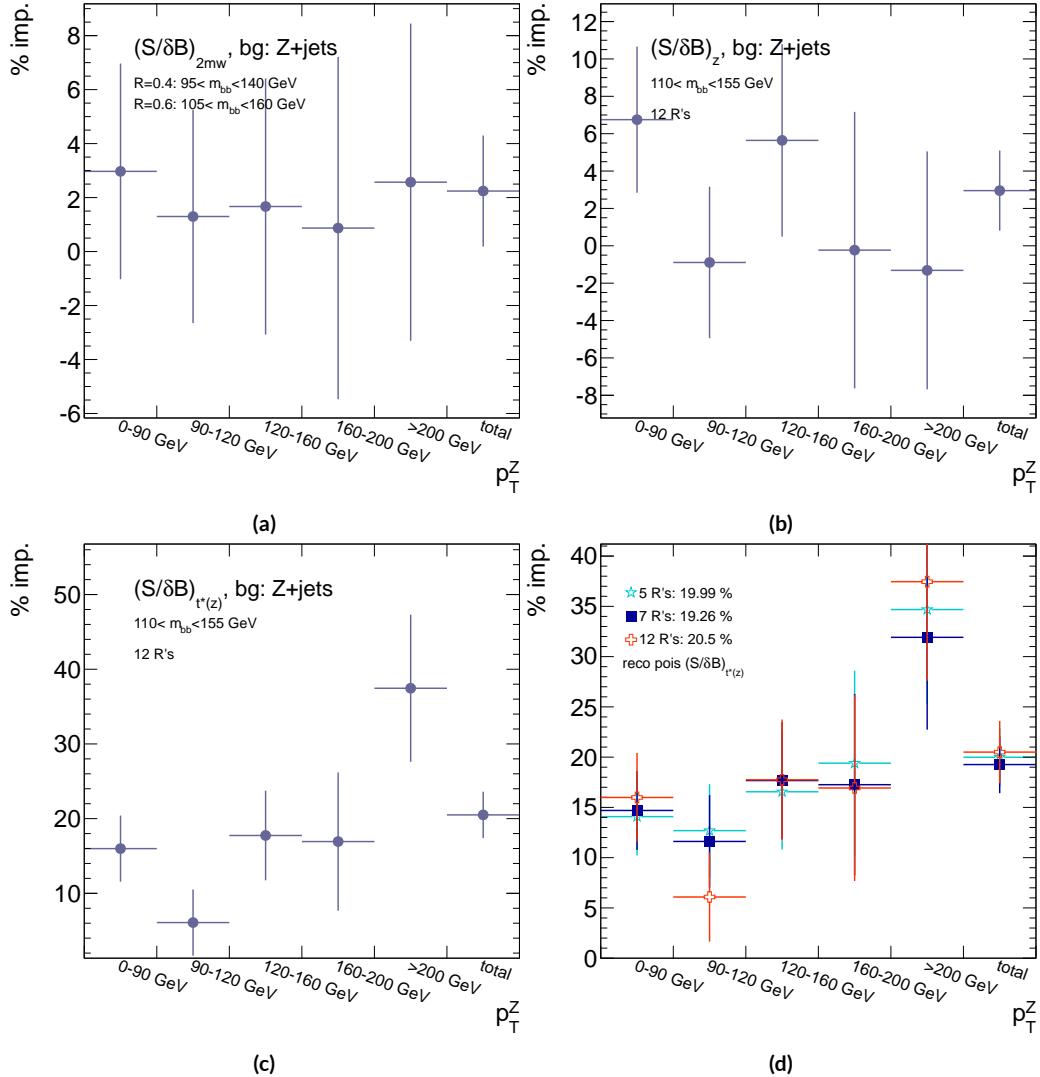


Figure B.9: A summary of the improvements for different reconstructed-level telescoping jet cuts and weights is shown in bins of p_T^Z . The final bin is the total improvement over all p_T^Z . Shown are improvements for the 2D m_{bb} cut (a), $t(z) = z$ (b), $t(z) = t^*(z)$ with 12 radii (c), and $t(z) = t^*(z)$ for various radii (d).

2738 B.12 CONCLUSIONS AND PROSPECTS

2739 The use of telescoping jets to provide multiple event interpretations shows promise as an avenue to
2740 increase significances in the $H \rightarrow b\bar{b}$ search in ATLAS and make an observation in the systematics-
2741 limited environment of early Run 2. A preliminary study using the telescoping jets algorithm with
2742 12 telescoping radii to build 12 event interpretations on 2012 Monte Carlo based on the full, cut-
2743 based Run 1 analysis yielded a 20.5% improvement in $S/\delta B$ over using anti- k_t with $R = 0.4$ alone
2744 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
2745 resolution in the large- R interpretations. The algorithm, in particular, showed discriminating power
2746 at high p_T^Z , so better performance can be expected in Run 2 with a higher \sqrt{s} and higher numbers of
2747 events with large p_T^Z . Additionally, the many simplifying assumptions regarding jet calibration and
2748 the relatively basic use of information[§] from multiple invariant masses in this note suggest that even
2749 further improvements than those quoted are possible. While this note did not explore the corre-
2750 lations between multiple event interpretations and the variables used in the BDT of the latest multi-
2751 variate version of the $H \rightarrow b\bar{b}$ analysis²⁰, new phenomenological studies suggest that such corre-
2752 lations are not strong³⁹. The corresponding reconstructed-level study, using a BDT, is left for future
2753 work. Also left for future work are better understanding the effects of jet trimming and which inter-
2754 pretations are the most useful.

[§]For examples of more sophisticated treatments compared to the treatment in this note, see Ref³⁹.

References

- [1] (2010). Measurement of the $w \rightarrow l\nu$ and $z/\gamma^* \rightarrow ll$ production cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the atlas detector. *JHEP*, 1012.
- [2] (2011a). *ATLAS tunes of PYTHIA 6 and Pythia 8 for MCII*. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva.
- [3] (2011b). *New ATLAS event generator tunes to 2010 data*. Technical Report ATL-PHYS-PUB-2011-008, CERN, Geneva.
- [4] (2012a). *b-jet tagging calibration on c-jets containing D^{*+} mesons*. Technical Report ATLAS-CONF-2012-039, CERN, Geneva.
- [5] (2012b). Electron performance measurements with the atlas detector using the 2010 lhc proton-proton collision data. *Eur. Phys. J.*, C72.
- [6] (2012a). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716, 30–61.
- [7] (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.
- [8] (2012). The Proton Synchrotron.
- [9] (2012). The Proton Synchrotron Booster.
- [10] (2012). The Super Proton Synchrotron.
- [11] (2013a). Jet energy measurement with the atlas detector in proton-proton collisions at $\sqrt{s} = 7$ TeV *Eur. Phys. J.*, C73.
- [12] (2013b). *Search for the bb decay of the Standard Model Higgs boson in associated W/ZH production with the ATLAS detector*. Technical Report ATLAS-CONF-2013-079, CERN, Geneva.

- 2779 [13] (2014). *ATLAS Run 1 Pythia8 tunes*. Technical Report ATL-PHYS-PUB-2014-021, CERN,
 2780 Geneva.
- 2781 [14] (2014). *Electron efficiency measurements with the ATLAS detector using the 2012 LHC*
 2782 *proton-proton collision data*. Technical Report ATLAS-CONF-2014-032, CERN, Geneva.
- 2783 [15] (2015). *Expected performance of the ATLAS b-tagging algorithms in Run-2*. Technical Report
 2784 ATL-PHYS-PUB-2015-022, CERN, Geneva.
- 2785 [16] (2015). *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS*
 2786 *Detector at $\sqrt{s} = 13 \text{ TeV}$* . Technical Report ATL-PHYS-PUB-2015-015, CERN, Geneva.
- 2787 [17] (2015). *Muon reconstruction performance in early $\sqrt{s}=13 \text{ TeV}$ data*. Technical Report ATL-
 2788 PHYS-PUB-2015-037, CERN, Geneva.
- 2789 [18] A. L. Read (2002). Presentation of search results: the CL_s technique. *J. Phys. G*, 28, 2693–
 2790 2704.
- 2791 [19] Aad, G. et al. (2014). Measurement of the Z/γ^* boson transverse momentum distribution in
 2792 pp collisions at $\sqrt{s} = 7 \text{ TeV}$ with the ATLAS detector. *JHEP*, 09, 145.
- 2793 [20] Ahmadov, F., Alio, L., Allbrooke, B., Bristow, T., Buescher, D., Buzatu, A., Coadou, Y.,
 2794 Debenedetti, C., Enari, Y., Facini, G., Fisher, W., Francavilla, P., Gaycken, G., Gentil, J.,
 2795 Goncalo, R., Gonzalez Parra, G., Grivaz, J., Gwilliam, C., Hageboeck, S., Halladjian, G., Jack-
 2796 son, M., Jamin, D., Jansky, R., Kiuchi, K., Kostyukhin, V., Lohwasser, K., & Lopez Mateos,
 2797 D, e. a. (2014). *Supporting Document for the Search for the bb decay of the Standard Model*
 2798 *Higgs boson in associated (W/Z)H production with the ATLAS detector*. Technical Report
 2799 ATL-COM-PHYS-2014-051, CERN, Geneva.
- 2800 [21] Aliev, M., Lacker, H., Langenfeld, U., Moch, S., Uwer, P., et al. (2011). HATHOR: Hadronic
 2801 Top and Heavy quarks cross section calculator. *Comput.Phys.Commun.*, 182, 1034–1046.
- 2802 [22] Alwall, J., Frederix, R., Frixione, S., Hirschi, V., Maltoni, F., Mattelaer, O., Shao, H. S.,
 2803 Stelzer, T., Torrielli, P., & Zaro, M. (2014). The automated computation of tree-level and
 2804 next-to-leading order differential cross sections, and their matching to parton shower simula-
 2805 tions. *JHEP*, 07, 079.
- 2806 [23] Alwall, J., Herquet, M., Maltoni, F., Mattelaer, O., & Stelzer, T. (2011). Madgraph 5 : Going
 2807 beyond.

- 2808 [24] ATLAS Collaboration (2014). *Tagging and suppression of pileup jets with the ATLAS detector*. Technical Report ATLAS-CONF-2014-018, CERN, Geneva.
- 2809
2810 [25] ATLAS Collaboration (2015a). MCPAnalysisGuidelinesMC15.
2811 <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesMC15>.
- 2812 [26] ATLAS Collaboration (2015b). *Performance of missing transverse momentum reconstruction*
2813 *for the ATLAS detector in the first proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$* . Technical Re-
2814 port ATL-PHYS-PUB-2015-027, CERN, Geneva.
- 2815 [27] Ball, R. D. et al. (2013). Parton distributions with LHC data. *Nucl. Phys.*, B867, 244–289.
- 2816 [28] Ball, R. D. et al. (2015). Parton distributions for the LHC Run II. *JHEP*, 04, 040.
- 2817 [29] Botje, M. et al. (2011). The PDF4LHC Working Group Interim Recommendations.
- 2818 [30] Buckley, A., Butterworth, J., Grellscheid, D., Hoeth, H., Lonnblad, L., Monk, J., Schulz, H.,
2819 & Siegert, F. (2010). Rivet user manual.
- 2820 [31] Buzatu, A. & Wang, W. (2016). *Object selections for SM Higgs boson produced in associa-
2821 tion with a vector boson in which $H \rightarrow b\bar{b}$ and V decays leptonically with Run-2 data: Object
2822 support note for $VH(bb)$ 2015+2016 dataset publication*. Technical Report ATL-COM-PHYS-
2823 2016-1674, CERN, Geneva. This is a support note for the $VH(bb)$ SM publication using the
2824 2015+2016 datasets.
- 2825 [32] Cacciari, M., G. P. Salam, M. C., & Salam, G. P. (2008). Pileup subtraction using jet areas.
2826 *Phys. Lett.*, B659.
- 2827 [33] Campbell, J. M. & Ellis, R. K. (2010). MCFM for the Tevatron and the LHC. *Nucl. Phys.
2828 Proc. Suppl.*, 205-206, 10–15.
- 2829 [34] Capeans, M., Darbo, G., Einsweiller, K., Elsing, M., Flick, T., Garcia-Sciveres, M., Gemme,
2830 C., Pernegger, H., Rohne, O., & Vuillermet, R. (2010). *ATLAS Insertable B-Layer Technical
2831 Design Report*. Technical Report CERN-LHCC-2010-013, ATLAS-TDR-19.
- 2832 [35] CERN (2008). LHC first beam: a day to remember.
- 2833 [36] Chan, S., Huth, J., Lopez Mateos, D., & Mercurio, K. (2015a). *ZH → ll $b\bar{b}$ Analysis with
2834 Telescoping Jets*. Technical Report ATL-PHYS-INT-2015-002, CERN, Geneva.

- 2835 [37] Chan, S. K.-w. & Huth, J. (2017). *Variations on MVA Variables in the SM ZH → ℓℓ $b\bar{b}$ Search*. Technical Report ATL-COM-PHYS-2017-1318, CERN, Geneva.
- 2836
- 2837 [38] Chan, S. K.-w., Lopez Mateos, D., & Huth, J. (2015b). *Micromegas Trigger Processor Algo-*
2838 *rithm Performance in Nominal, Misaligned, and Corrected Misalignment Conditions*. Tech-
2839 nical Report ATL-COM-UPGRADE-2015-033, CERN, Geneva.
- 2840 [39] Chien, Y.-T., Farhi, D., Krohn, D., Marantan, A., Mateos, D. L., & Schwartz, M. (2014).
2841 Quantifying the power of multiple event interpretations.
- 2842 [40] Ciccolini, M., Dittmaier, S., & Kramer, M. (2003). Electroweak radiative corrections to asso-
2843 ciated WH and ZH production at hadron colliders. *Phys. Rev.*, D68, 073003.
- 2844 [41] Clark, B., Lopez Mateos, D., Felt, N., Huth, J., & Oliver, J. (2014). *An Algorithm for Mi-*
2845 *cromegas Segment Reconstruction in the Level-1 Trigger of the New Small Wheel*. Technical
2846 Report ATL-UPGRADE-INT-2014-001, CERN, Geneva.
- 2847 [42] Collaboration, A. (2017a). Evidence for the $h \rightarrow b\bar{b}$ decay with the atlas detector.
- 2848 [43] Collaboration, A. (2017b). Study of the material of the atlas inner detector for run 2 of the
2849 lhc.
- 2850 [44] Collaboration, T. A., Aad, G., Abat, E., Abdallah, J., & A A Abdelalim, e. a. (2008). The atlas
2851 experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08), S08003.
- 2852 [45] Cowan, G., Cranmer, K., Gross, E., & Vitells, O. (2011). Asymptotic formulae for likelihood-
2853 based tests of new physics. *Eur. Phys. J. C*, 71, 1554.
- 2854 [46] Czakon, M., Fiedler, P., & Mitov, A. (2013). Total Top-Quark Pair-Production Cross Section
2855 at Hadron Colliders Through $\alpha(\frac{4}{5})$. *Phys. Rev. Lett.*, 110, 252004.
- 2856 [47] D. Kahawala, D. Kahawala, D. K. & Schwartz, M. D. (2013). Jet sampling: Improving event
2857 reconstruction through multiple interpretations.
- 2858 [48] D. Krohn, D. Krohn, J. T. L. W. (2009). Jet trimming.
- 2859 [49] Delmastro, M., Gleyzer, S., Hengler, C., Jimenez, M., Koffas, T., Kuna, M., Liu, K., Liu, Y.,
2860 Marchiori, G., Petit, E., Pitt, M., Soldatov, E., & Tackmann, K. (2014). *Photon identification*
2861 *efficiency measurements with the ATLAS detector using LHC Run 1 data*. Technical Report
2862 ATL-COM-PHYS-2014-949, CERN, Geneva.

- 2863 [50] Evans, L. & Bryant, P. (2008). Lhc machine. *Journal of Instrumentation*, 3(08), S08001.
- 2864 [51] Gleisberg, T. et al. (2009a). Event generation with SHERPA 1.1. *JHEP*, 02, 007.
- 2865 [52] Gleisberg, T., Höche, S., Krauss, F., Schönherr, M., Schumann, S., Siegert, F., & Winter, J. (2009b). Event generation with sherpa 1.1. *Journal of High Energy Physics*, 2009(02), 007.
- 2866 [53] Hagebock, S. (CERN, Geneva, 2017). Lorentz Invariant Observables for Measurements of Hbb Decays with ATLAS.
- 2867 [54] Heinemann, B., Hirsch, F., & Strandberg, S. (2010). *Performance of the ATLAS Secondary Vertex b-tagging Algorithm in 7 TeV Collision Data*. Technical Report ATLAS-COM-CONF-2010-042, CERN, Geneva. (Was originally 'ATL-COM-PHYS-2010-274').
- 2868 [55] J. Pumplin, J. Pumplin, D. S. J. H. H. L. P. M. N. e. a. (2002). New generation of parton distributions with unvertainties from global qcd analysis. *JEGP*, 0207.
- 2869 [56] Jackson, P. & Rogan, C. (2017). Recursive jigsaw reconstruction: Hep event analysis in the presence of kinematic and combinatoric ambiguities.
- 2870 [57] Kant, P., Kind, O., Kintscher, T., Lohse, T., Martini, T., Molbitz, S., Rieck, P., & Uwer, P. (2015). Hathor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions. *Computer Physics Communications*, 191, 74 – 89.
- 2871 [58] Lampl, W., Laplace, S., Lelas, D., Loch, P., Ma, H., Menke, S., Rajagopalan, S., Rousseau, D., Snyder, S., & Unal, G. (2008). *Calorimeter Clustering Algorithms: Description and Performance*. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva.
- 2872 [59] Lavesson, N. & Lonnblad, L. (2005). W+jets matrix elements and the dipole cascade.
- 2873 [60] LHC Higgs Cross Section Working Group, Dittmaier, S., Mariotti, C., Passarino, G., & Tanaka (Eds.), R. (CERN, Geneva, 2011). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*.
- 2874 [61] Loch, Peter and Lefebve, Michel (2007). Introduction to Hadronic Calibration in ATLAS.

- 2889 [62] Luisoni, G., Nason, P., Oleari, C., & Tramontano, F. (2013). H_w ±/hz + o and 1 jet at nlo
 2890 with the powheg box interfaced to gosam and their merging within minlo. *Journal of High*
 2891 *Energy Physics*, 2013(10), 83.
- 2892 [63] M. Cacciari, M. Cacciari, G. P. S. & Soyez, G. (2008). The anti- k_t jet clustering algorithm.
 2893 *JHEP*, 0804.
- 2894 [64] Marcastel, F. (2013). CERN's Accelerator Complex. La chaîne des accélérateurs du CERN.
 2895 General Photo.
- 2896 [65] Masubuchi, T., Benitez, J., Bell, A. S., Argyropoulos, S., Arnold, H., Amaral Coutinho, Y.,
 2897 Sanchez Pineda, A. R., Buzatu, A., Calderini, G., & Chan, Stephen Kam-wah, e. a. (2016).
 2898 *Search for a Standard Model Higgs boson produced in association with a vector boson and de-*
 2899 *caying to a pair of b-quarks*. Technical Report ATL-COM-PHYS-2016-1724, CERN, Geneva.
- 2900 [66] Patrignani, C. et al. (2016). Review of Particle Physics. *Chin. Phys.*, C40(10), 100001.
- 2901 [67] Robson, A., Piacquadio, G., & Schopf, E. (2016). *Signal and Background Modelling Stud-*
 2902 *ies for the Standard Model VH, H → b̄b and Related Searches: Modelling support note*
 2903 *for VH(b̄b) 2015+2016 dataset publication*. Technical Report ATL-COM-PHYS-2016-1747,
 2904 CERN, Geneva. This is a support note for the VH(b̄b) SM publication using the 2015+2016
 2905 datasets.
- 2906 [68] S. Alioli et al. (2009). NLO Higgs boson production via gluon fusion matched with shower
 2907 in POWHEG. *JHEP*, 0904, 002.
- 2908 [69] S. D. Ellis, S. D. Ellis, A. H. T. S. R. D. K. & Schwartz, M. D. (2012). Qjets: A non-
 2909 deterministic approach to tree-based jet substructure. *Phys. Rev. Lett.*, 108.
- 2910 [70] Salam, G. P. (2009). Towards jetography.
- 2911 [71] Sjostrand, T., Mrenna, S., & Skands, P. Z. (2008). A Brief Introduction to PYTHIA 8.1.
 2912 *Comput.Phys.Commun.*, 178, 852–867.
- 2913 [72] Stancari, G., Previtali, V., Valishev, A., Bruce, R., Redaelli, S., Rossi, A., & Salvachua Fer-
 2914 rando, B. (2014). *Conceptual design of hollow electron lenses for beam halo control in the*
 2915 *Large Hadron Collider*. Technical Report FERMILAB-TM-2572-APC. FERMILAB-TM-
 2916 2572-APC, CERN, Geneva. Comments: 23 pages, 1 table, 10 figures.

- 2917 [73] Stewart, I. W. & Tackmann, F. J. (2011). Theory uncertainties for higgs and other searches
2918 using jet bins.
- 2919 [74] Symmetry Magazine (2015). The Standard Model of Particle Physics.
- 2920 [75] T. Gleisberg et al., T. G. e. a. (2009). Event generation with sherpa 1.1. *JHEP*, 02.
- 2921 [76] T. Sjöstrand, T. Sjöstrand, S. M. & Sands, P. Z. (2008). A brief introduction to pythia 8.1.
2922 *Comput. Phys. Commun.*, 178.
- 2923 [77] Tully, C. C. (2011). *Elementary particle physics in a nutshell*. Princeton, NJ: Princeton Univ.
2924 Press.
- 2925 [78] Verkerke, W. & Kirkby, D. (2003). The RooFit toolkit for data modeling. In *2003 Computing*
2926 *in High Energy and Nuclear Physics, CHEP03*.
- 2927 [79] Watts, G., Filthaut, F., & Piacquadio, G. (2015). *Extrapolating Errors for b-tagging*. Technical
2928 Report ATL-COM-PHYS-2015-711, CERN, Geneva. This is for internal information only, no
2929 approval to ever be seen outside of ATLAS.
- 2930 [80] Y. L. Dokshitzer, Y. L. Dokshitzer, G. D. L. S. M. & Webber, B. R. (1997). Better jet clustering
2931 algorithms. *JHEP*, 9708.