

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

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

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

14                  HARVARD UNIVERSITY  
15                  CAMBRIDGE, MASSACHUSETTS  
16                  2018

<sup>17</sup> ©2017 – STEPHEN K. CHAN

<sup>18</sup> ALL RIGHTS RESERVED.

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

21           **ABSTRACT**

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

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

# Contents

34

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

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

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



# Acknowledgments

<sup>123</sup> THIS THESIS WOULD NOT HAVE BEEN POSSIBLE without large amounts of espresso.

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

Zhuangzi

# 0

124

125

## Introduction

126 SINCE THE DISCOVERY of a Standard Model (SM) like Higgs boson at the LHC in 2012<sup>21</sup>, one of  
127 the main outstanding physics goals of the LHC has been to observe the primary SM Higgs decay  
128 mode,  $H \rightarrow b\bar{b}$ , with efforts primarily targeted at searching for Higgs bosons produced in associa-  
129 tion with a leptonically decaying vector ( $W$  or  $Z$ , denoted generically as  $V$ ) boson. As the integrated

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

134 While this effort will likely require a combination of several different methods at various different  
135 stages in the analysis chain, one possible avenue forward is to revise the multivariate analysis (MVA)  
136 discriminant input variables used, as various schemes offer the promise of reducing systematic uncer-  
137 tainties through more efficient use of both actual and simulated collision data. This thesis discusses  
138 two such alternate MVA schemes, the RestFrames (RF) and Lorentz Invariants (LI) variables, in the  
139 context of the 2-lepton channel of the Run 2 analysis in<sup>34</sup> and<sup>56</sup>, henceforth referred to as the “fidu-  
140 cial analysis,” before a brief discussion of combinations across channels and datasets.

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

148 Editorial notes:

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

*Noli turbare circulos meos*

Archimedes

# 1

<sup>151</sup>

## <sup>152</sup> The Large Hadron Collider and the ATLAS

### Detector

<sup>153</sup>

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

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

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

<sup>158</sup> **I.I THE CERN ACCELERATOR COMPLEX**

<sup>159</sup> The journey of protons from hydrogen canister to high energy collisions through the CERN acceler-  
<sup>160</sup> ator complex, illustrated in Figure I.I, is also one through the history of CERN’s accelerator program.  
<sup>161</sup> After being ionized in an electric field, protons are first accelerated in a linear accelerator, LINAC 2\*,  
<sup>162</sup> to a kinetic energy of 50 MeV. From there, they are fed into the Proton Synchotron Booster†, which  
<sup>163</sup> further accelerates them to 1.4 GeV and, as its name implies, feeds them to the 628 m Proton Syn-  
<sup>164</sup> chotron (PS, 1959<sup>3</sup>) and up to 25 GeV. The penultimate stage is the 7 km Super Proton Synchotron  
<sup>165</sup> (SPS, 1976; responsible for the discovery of the  $W$  and  $Z$  bosons and the 1983 Nobel Prize<sup>5</sup>), which  
<sup>166</sup> accelerates the protons to a kinetic energy of 450 GeV. Finally, these 450 GeV protons are injected  
<sup>167</sup> into the LHC<sup>42</sup>, a proton-proton collider housed in the 27 km circumference tunnel that housed the  
<sup>168</sup> Large Electron Positron Collider (LEP) before its operations ceased in 2000.

<sup>169</sup> **I.2 THE LARGE HADRON COLLIDER**

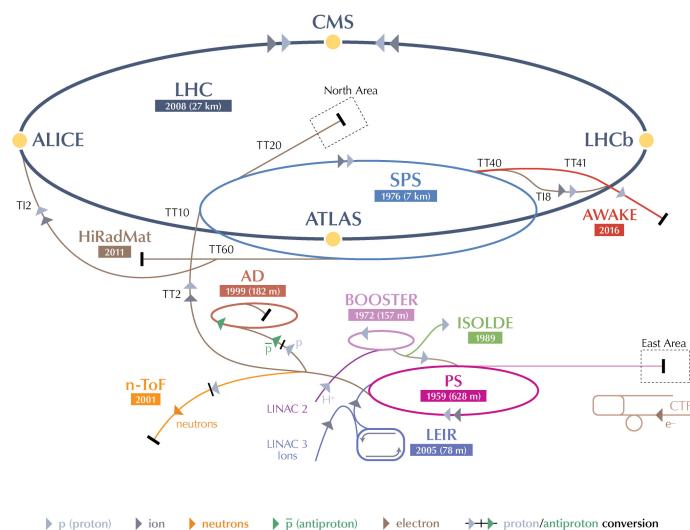
<sup>170</sup> The LHC was designed to function primarily as a proton-proton collider with a center of mass en-  
<sup>171</sup> ergy  $\sqrt{s} = 14$  TeV and an instantaneous luminosity of  $1 \times 10^{34} \text{ cm}^{-2} \cdot \text{s}^{-1}$ , though it is also capable of  
<sup>172</sup> producing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

---

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

† Protons can be directly from a LINAC into the PS, but the higher injection energy allows for approximately 100 times more protons to be used at once<sup>4</sup>, 1972.

## CERN's Accelerator Complex



LHC Large Hadron Collider SPS Super Proton Synchrotron PS Proton Synchrotron

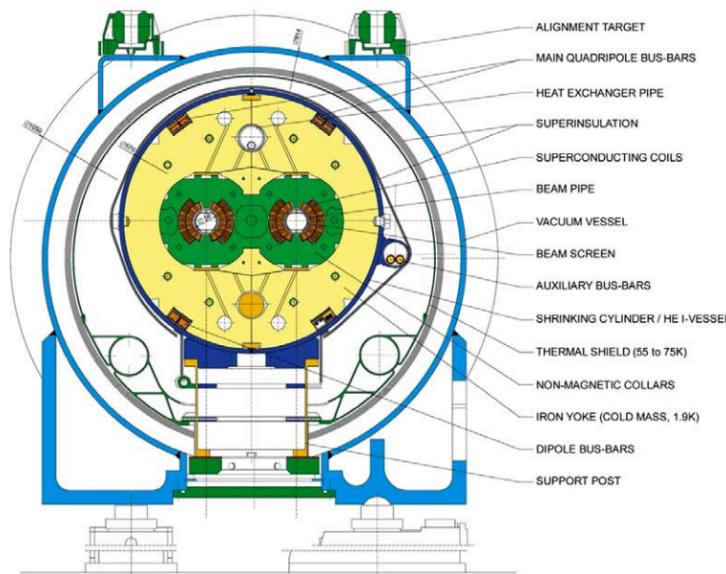
AD Antiproton Decelerator CTF3 Clic Test Facility AWAKE Advanced WAKefield Experiment ISOLDE Isotope Separator OnLine Dvice  
LEIR Low Energy Ion Ring LINAC LINear ACcelerator n-ToF Neutrons Time Of Flight HiRadMat High-Radiation to Materials

©CERN 2013

Figure 1.1: The CERN Accelerator Complex<sup>55</sup>

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

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

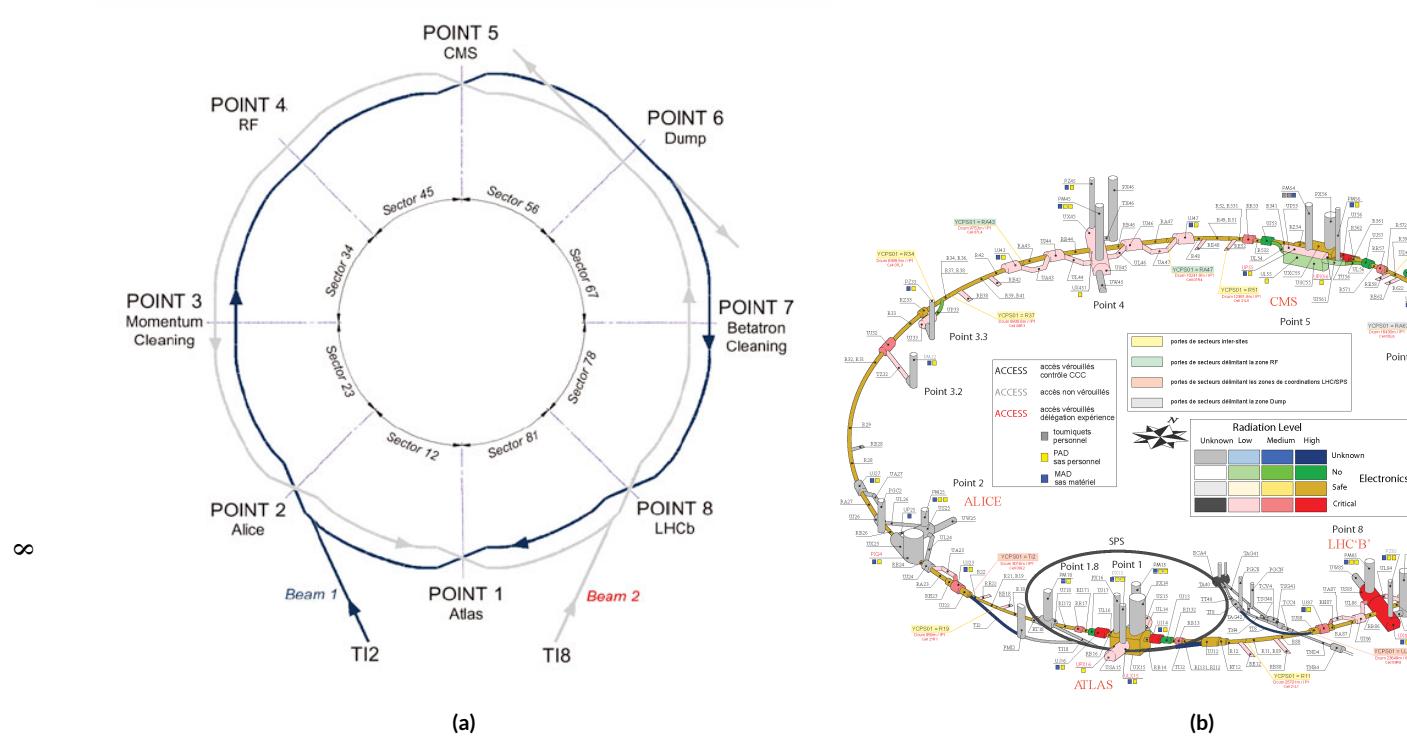


**Figure 1.2:** Schematic drawing of an LHC dipole magnet and cryogenics system.

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

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

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



**Figure 1.3:** Schematic and detailed views of the LHC ring. IC: <sup>27</sup>, <sup>63</sup>

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

210 1.3 ATLAS AT A GLANCE

211 1.3.1 COORDINATES AND DISTANCES IN THE ATLAS DETECTOR

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

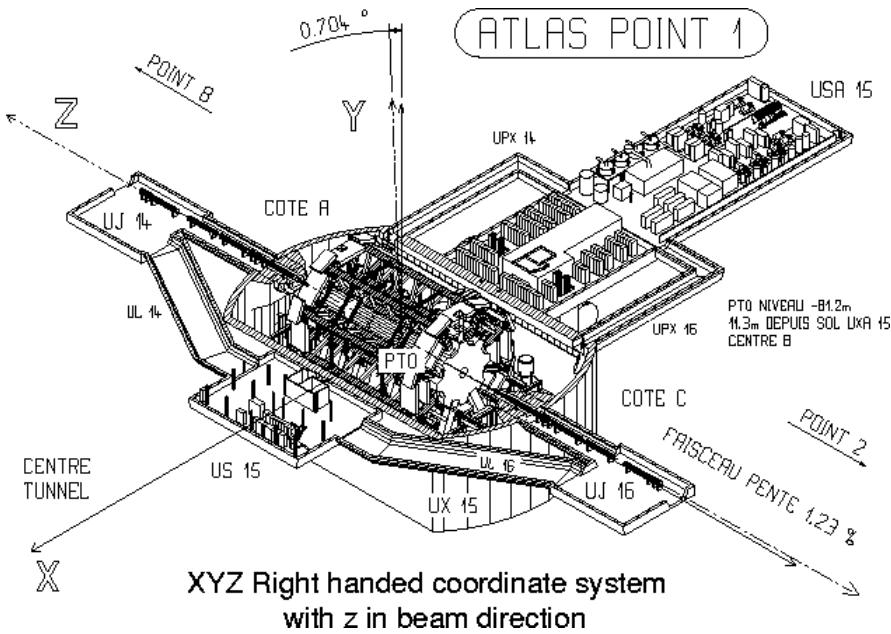
217 The ATLAS coordinate system is shown in Figure 1.4. It is a right-handed coordinate system  
218 centered at the nominal collision point, with the  $x$  axis pointing towards the center of the LHC ring,  
219 the  $z$  axis pointing up, and the  $y$  axis completing the right-handed coordinate system.

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

---

<sup>‡</sup>This is the only reason CMS can call itself “compact.”

<sup>§</sup>“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.

<sup>228</sup> In a lepton collider where total momentum is conserved, a useful coordinate is the relativistic

<sup>229</sup> rapidity of a particle:

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

<sup>230</sup> with  $E$  and  $p_z$  as the energy and longitudinal momentum of the particle, respectively. The rapidity

<sup>231</sup> is the relativistic analog of a rotation angle; boosts can be added in a manner similar to rotations<sup>¶</sup>,

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

<sup>233</sup> in the hard scatter are partons inside of the proton of unknown momentum fraction, longitudinal

<sup>234</sup> momentum is not conserved. Nevertheless, since the incident momentum is entirely longitudinal,

---

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

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

239

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

240 This last expression, denoted  $\eta$ , is known as the pseudorapidity.

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

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

---

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

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

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

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

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

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

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

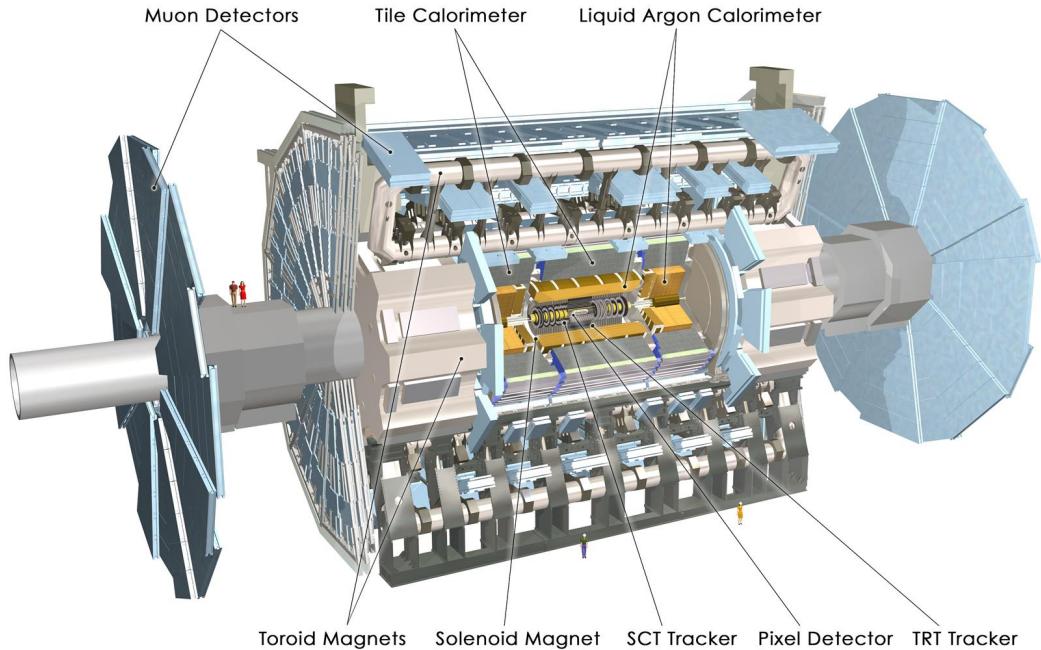
<sup>264</sup> typically pions, and is called the nuclear interaction length.

<sup>265</sup> **I.3.2 GENERAL LAYOUT OF ATLAS**

<sup>266</sup> The ATLAS detector and its main components are shown in Figure 1.5. ATLAS is designed as a  
<sup>267</sup> largely hermetic detector, offering full coverage in  $\phi$  and coverage in  $|\eta|$  up to 4.7. The multiple sub-  
<sup>268</sup> systems allow for good characterization of the decay products from collisions in the LHC. The in-  
<sup>269</sup> nermost system is the inner detector (ID); composed primarily of silicon pixels and strips immersed  
<sup>270</sup> in a magnetic field, it is designed to construct the curved trajectories of charged particles produced in  
<sup>271</sup> collisions while taking up as little material as possible.

<sup>272</sup> Surrounding the ID is the liquid argon based electromagnetic calorimeter (ECAL), which is de-  
<sup>273</sup> signed to capture all of the energy of the electromagnetic showers produced by electrons and pho-  
<sup>274</sup> tons coming from particle collisions. The ECAL is in turn encapsulated by a scintillating tile based  
<sup>275</sup> and liquid argon based hadronic calorimeter (HCAL) that captures any remaining energy from the  
<sup>276</sup> jets produced by hadronizing quarks and gluons coming from collisions.

<sup>277</sup> The outermost layer of ATLAS is the muon spectrometer (MS) with its own magnetic field pro-  
<sup>278</sup> duced by toroidal magnets. Muons are highly penetrating particles that escape the calorimeters with  
<sup>279</sup> most of their initial momentum, so the MS and its magnets are designed to curve these charged  
<sup>280</sup> particles and measure their trajectories to measure their outgoing momenta. Each of these detector  
<sup>281</sup> systems has several principal subsystems and performance characteristics, which will be described in  
<sup>282</sup> turn below.

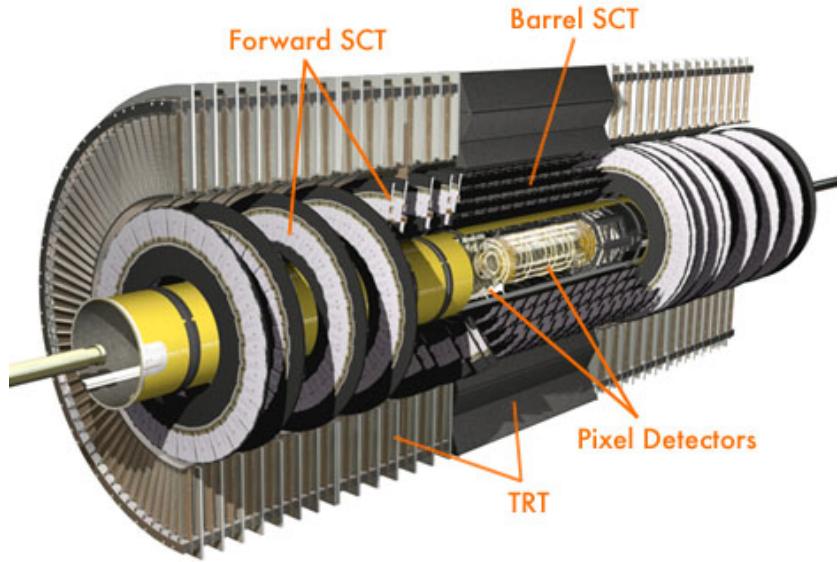


**Figure 1.5:** The ATLAS detector with principal subsystems shown.

<sup>283</sup> **I.4 THE INNER DETECTOR**

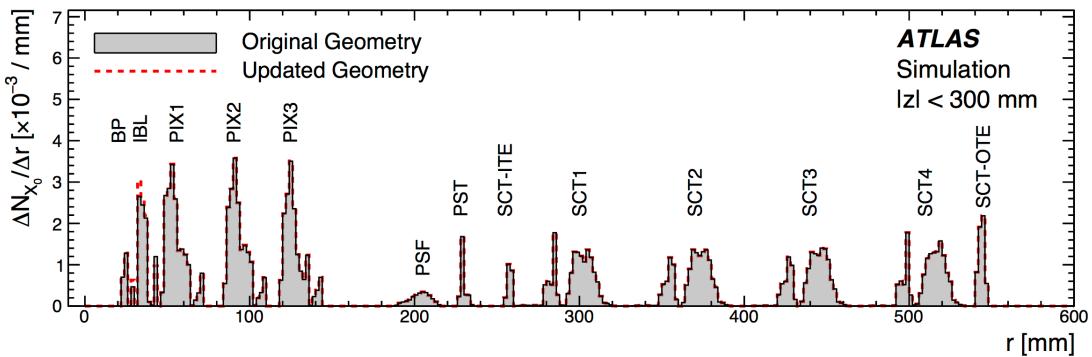
<sup>284</sup> ATLAS's inner detector (ID) is surrounded by a superconducting solenoid that is cryogenically  
<sup>285</sup> cooled to a temperature of 4.5 K, which immerses it in a 2 T magnetic field. The ID uses two sili-  
<sup>286</sup> con detector subsystems (the Pixel and SemiConductor (strip) Tracker (SCT)) to track the curved  
<sup>287</sup> trajectories of charged particles emanating from particle collisions and one subsystem composed of  
<sup>288</sup> gas straw detectors with filaments for  $e/\pi$  discrimination, as shown in Figure 1.9. The ID offers full  
<sup>289</sup> coverage in  $\phi$  and extends to an  $|\eta|$  of 2.5.

<sup>290</sup> Since the components of the ID do not provide an energy measurement, it is desirable for a track-  
<sup>291</sup> 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:<sup>36</sup>

292 may be done in the calorimeters. Generally, there are two radiation lengths in the inner detector (it  
 293 varies with  $\eta$ ); the full material budget, with the layout of the individual layers in each subsystem,  
 294 can be seen in 1.8.



**Figure 1.7:** The ID material budget. IC:<sup>35</sup>

295    **I.4.1 THE PIXEL DETECTOR**

296    The innermost part of ATLAS is the pixel detector, which, as the name suggests, is comprised of  
297    four layers of pixels in the barrel at 32, 51, 89, and 123 mm from the beam pipe, and three layers in the  
298    end caps at 495, 580, and 650 mm from the beam pipe, with over 80 million channels total. The in-  
299    nermost layer of pixels, the insertable *B* layer (IBL) was installed during the 2013–14 LHC shutdown.  
300    The pixels are cooled to a temperature of  $\sim -5$  with  $N_2$  gas and operate at 150–600 V. The pixels  
301    themselves come in two sizes  $50 \times 400(600) \times 250 \mu\text{m}$ , with the larger pixels being in the outer  
302    layers. They provide nominal resolution of  $10(115) \mu\text{m}$  resolution in  $r - \phi(z)$  direction.

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

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

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

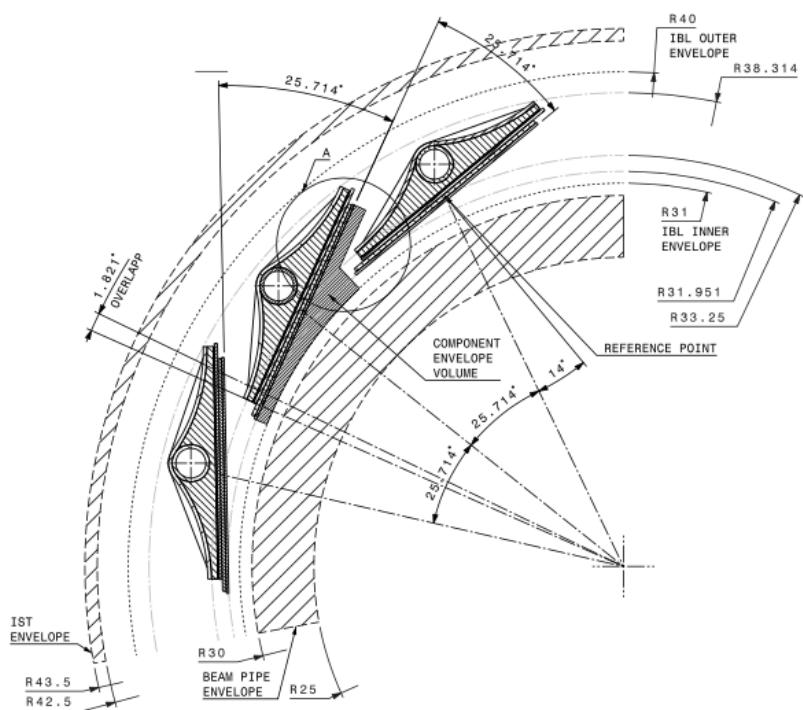


Figure 1.8: Arrangement of pixels in the barrel. IC:<sup>26</sup>

312 1.4.3 TRANSITION RADIATION TRACKER (TRT)

313 The final and outermost subsystem in the ID is the transition radiation tracker (TRT). It provides  
314 coverage for  $|\eta|$  up to 2.0 and is composed of straw detectors with a 4 mm diameter and run the  
315 length of the detector module, which provide 130  $\mu\text{m}$  resolution and filled with a Xe-CO<sub>2</sub>-O<sub>2</sub> (70-  
316 27-3) gas combination and operate at -1500 V. The filaments and foil lining inside the straws induce  
317 X-ray emission in electrons and pions passing through the TRT as they move from a dielectric to  
318 a gas; this “transition radiation” is the source of the TRT’s name. Since the energy deposited due  
319 to transition radiation is proportional to the relativistic boost  $\gamma$ , for constant momentum, this is  
320 inversely proportional to mass, so electrons will have  $\sim 130/0.5 = 260 \times$  more transition radiation  
321 than pions, in principle enabling excellent electron/pion discrimination. The TRT will be replaced  
322 by silicon strips in the Phase II upgrade.

323 1.5 THE ATLAS CALORIMETERS

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

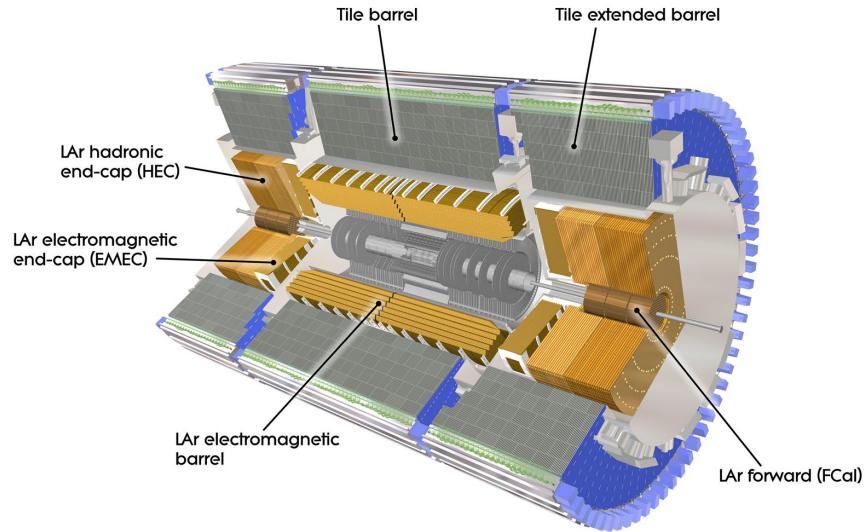


Figure 1.9: The ATLAS calorimeters.

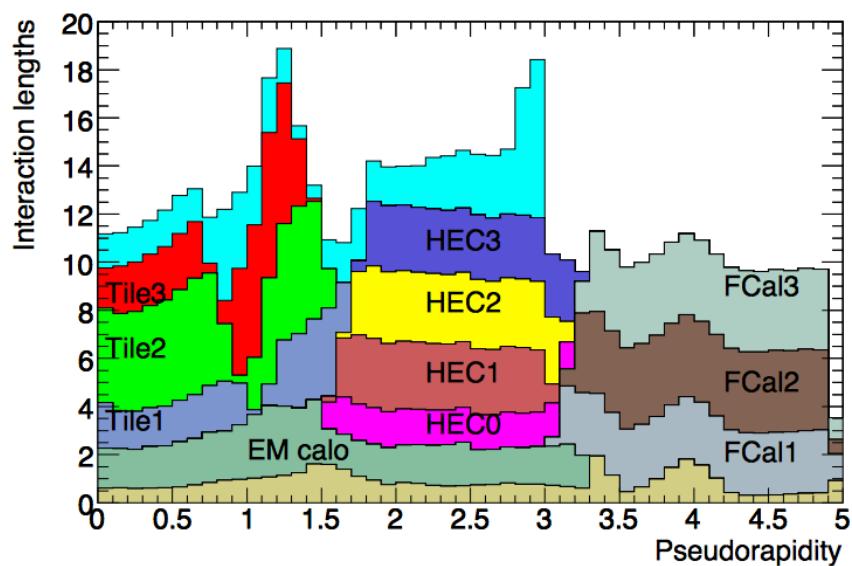


Figure 1.10: Material depth of the ATLAS calorimeters. IC;<sup>36</sup>

328 1.5.1 CALORIMETER RESOLUTION

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

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

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

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

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

### 350 1.5.2 THE ELECTROMAGNETIC CALORIMETER (ECAL)

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

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

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

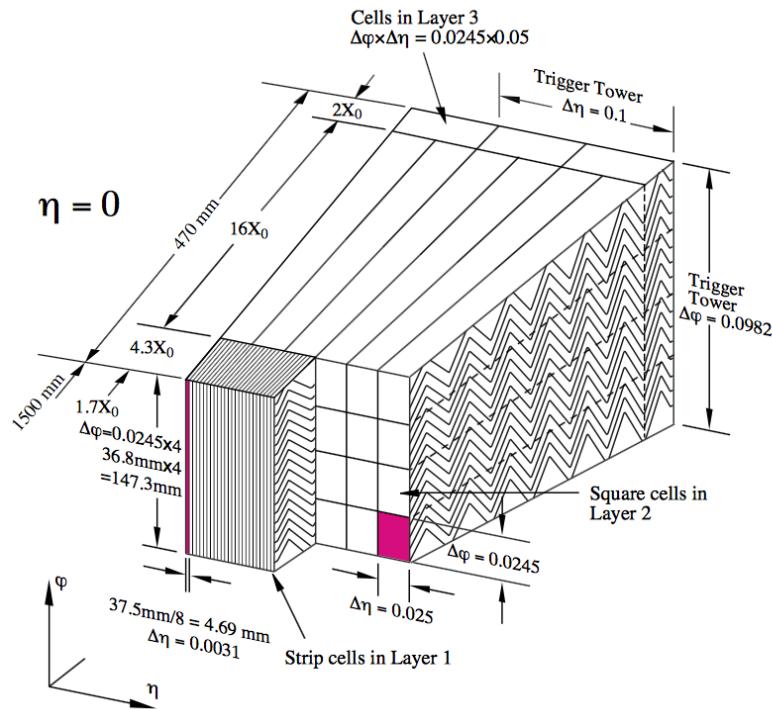
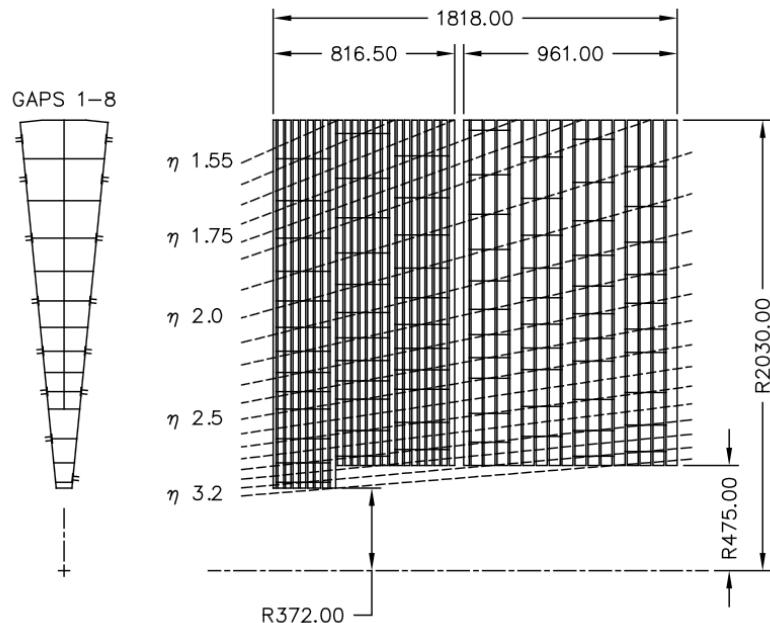


Figure 1.11: The accordion geometry of the LAr electromagnetic calorimeters is prominently shown in this illustration of an ECAL barrel module. IC:<sup>36</sup>

<sup>365</sup> cluster reconstruction in ATLAS.

<sup>366</sup> **I.5.3 HADRONIC END CAPS (HEC)**

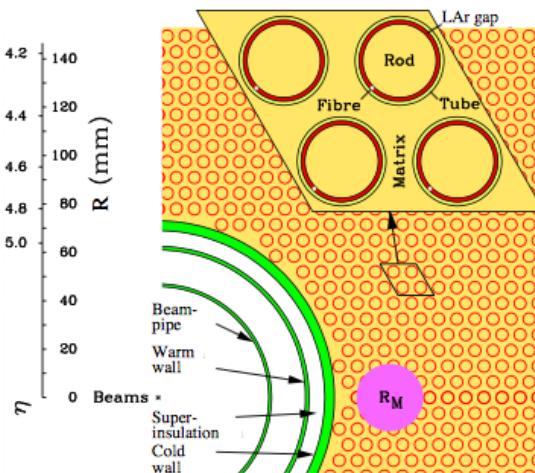
<sup>367</sup> The HEC covers an  $|\eta|$  range of 1.5 to 3.2. Like the ECAL end caps, the HEC consists of two identi-  
<sup>368</sup> cal wheels out to a distance from the beam axis of 2030 mm; its layout is shown in Figure I.12. The  
<sup>369</sup> HEC also has LAr as the active material, but instead has flat copper plates as absorbers for sampling  
<sup>370</sup> fraction of 4.4% and 2.2% in the first and second wheels, respectively. Its granularity in  $\phi$  is  $0.1 \times 0.1$   
<sup>371</sup> for  $|\eta|$  up to 2.5 and  $0.2 \times 0.2$  in the more forward regions.



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

<sup>372</sup> 1.5.4 THE FORWARD CALORIMETER (FCAL)

<sup>373</sup> The FCAL covers an  $|\eta|$  range from 3.1 to 4.9, again using LAr as the active material in gaps between  
<sup>374</sup> rods and tubes in a copper-tungsten matrix, as shown in Figure 1.13. These system has characteris-  
<sup>375</sup> tic performance corresponding to stochastic term of  $S \approx 1 \text{ GeV}^{-1/2}$ . There are three modules in  
<sup>376</sup> the FCAL: one electromagnetic and two hadronic, with the latter two featuring a higher tungsten  
<sup>377</sup> content for a larger absorption length.



**Figure 1.13:** The material layout for a typical section of the FCAL in the transverse plane. IC:<sup>36</sup>

<sup>378</sup> 1.5.5 THE HADRONIC TILE CALORIMETER

<sup>379</sup> The tile calorimeter, covering at  $|\eta|$  of up to 1.7 is made up of 64 modules in the barrel (each cover-  
<sup>380</sup> ing  $\Delta\phi$  of  $360/64 = 5.625$ , each with a layout as in Figure 1.14. It is designed to be self-supporting  
<sup>381</sup> for structural reasons, and so is the only calorimeter without LAr as a an active medium, with a stag-  
<sup>382</sup> gered matrix of active scintillating polystyrene and supporting steel. It operates at 1800 V with a 400

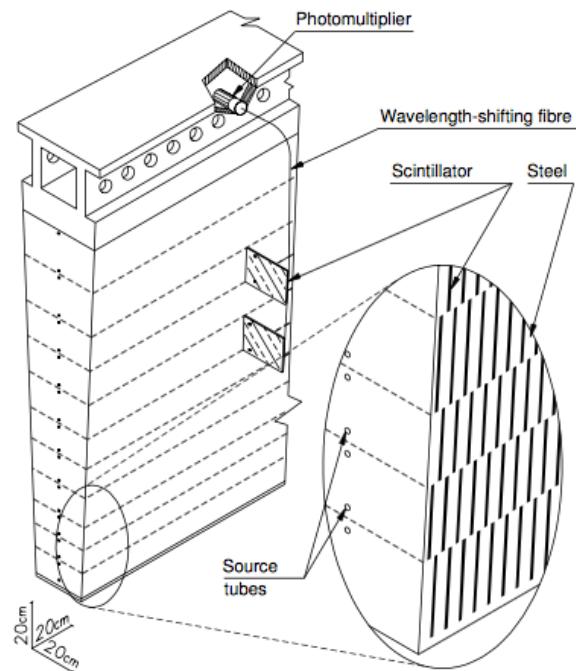


Figure 1.14: The material layout for a typical section of the hadronic tile calorimeter. IC.<sup>36</sup>

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

## 386 1.6 THE MUON SPECTROMETER

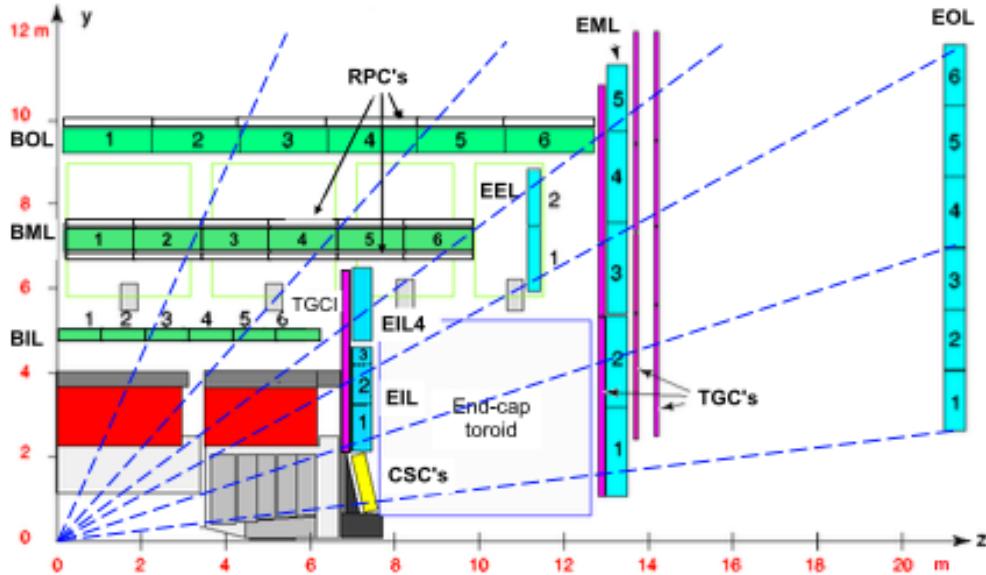
387 Since the energy of muons is not captured within the calorimeters, the stations of the ATLAS MS  
388 surrounds the entire detectors and provides tracks of outgoing muons that can be matched to tracks  
389 in the ID. The ATLAS toroids, which provide field strengths of up to 2.5 (3.5) T in the barrel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

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

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

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

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



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

401  
402     The MS can reconstruct muons with transverse momenta from 5 GeV up to 3 TeV (with 10%  
403 resolution at 1 TeV (3% at 100 GeV)). Detectors in the MS fall into two broad headings, precision  
404 detectors and trigger detectors, both described below. Nominal performance of the current detec-  
405 tor types in the MS is summarized in Figure 1.16, a table taken from <sup>36</sup>. It should be noted that  $|\eta|$   
406 ranges quoted below, where applicable, do not include the range 0-0.1, where this a gap in the MS to  
407 allow for cabling and other services to the ATLAS detector; for a discussion of compensatory mea-  
408 sures in muon reconstruction, see Chapter 5.

Type	Function	Chamber resolution (RMS) in			Measurements/track		Number of	
		$z/R$	$\phi$	time	barrel	end-cap	chambers	channels
MDT	tracking	35 $\mu\text{m}$ ( $z$ )	—	—	20	20	1088 (1150)	339k (354k)
CSC	tracking	40 $\mu\text{m}$ ( $R$ )	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:<sup>36</sup>

#### 409 I.6.1 PRECISION DETECTORS

410 The ATLAS MS has two types of precision detectors: monitored drift tubes (MDT's) and cathode  
 411 strip chambers (CSC's). An MDT is a tube with a 3 cm diameter with length depending on the sta-  
 412 tion in which the tube is located. The tube is filled with an Ar/CO<sub>2</sub> gas mixture and has a tungsten-  
 413 rhenium wire at its center that is kept at 3 000 V when operational. The MDT's provide 35  $\mu\text{m}$  res-  
 414 olution (per chamber) in their cross-sectional dimension (there is no sensitivity along the axis of the  
 415 wire). Resolution of this magnitude requires very precise knowledge of the location of the wires  
 416 within the MDT's; this is generally true for detectors in the MS (trigger as well as precision); to this  
 417 end, stations of the MS are aligned using an optical laser system. For a detailed description of how  
 418 misalignment can affect performance, see Appendix A for a detailed discussion of misalignment's  
 419 simulated effects on the performance of the proposed micromegas trigger processor in the New Small  
 420 Wheel (NSW) of the Phase I upgrade. Their 700 ns dead time, however, precludes their use as trig-  
 421 ger detectors and also in the region of the small wheel (innermost endcap) closest to the beam pipe  
 422 ( $|\eta|$  from 2.0 to 2.7), where rates are highest.

423 In this region, the precision detectors are the CSC's, which have a much lower dead time of  $\sim 40$

<sup>424</sup> ns. These are multiwire proportional chambers with cathode planes that have orthogonal sets of  
<sup>425</sup> strips, allowing for a measurement in both the principal directions. CSC detector sizes also vary  
<sup>426</sup> by station, coming in both small and large chambers. The CSC strip pitch is 5.31 (5.56) mm for the  
<sup>427</sup> large (small) chambers, with position determined from the induced charge distribution in the strips.

<sup>428</sup> This corresponds to a nominal resolution of 60 (5 000)  $\mu\text{m}$  per plane in the bending (non-bending)  
<sup>429</sup> direction. These are slated be replaced micromegas detectors in the NSW.

#### <sup>430</sup> 1.6.2 TRIGGER DETECTORS

<sup>431</sup> Trigger detectors have a fundamentally different role than the precision detectors, instead needing  
<sup>432</sup> to deliver “good enough” approximate values of muon track positions and  $p_T$  values. The MS has  
<sup>433</sup> two types of trigger detectors: resistive plate chambers (RPC’s) in the barrel and thin gap chambers  
<sup>434</sup> (TGC’s) in the end caps. They collectively cover an  $|\eta|$  range to 2.4, and their arrangement is shown  
<sup>435</sup> in Figure 1.17.

<sup>436</sup> The RPC’s are parallel plate detectors with a dead time of 5 ns and a thickness of 2 mm, kept at  
<sup>437</sup> a potential of 9 800 V; they are deployed in three layers. RPC’s, too, feature strips with orthogonal  
<sup>438</sup> arrangements on the top and bottom planes, with a strip pitch of 23–35 mm.

<sup>439</sup> The TGC’s are multiwire proportional chambers with a dead time of 25 ns. Also, featuring or-  
<sup>440</sup> thogonal strips, the TGC’s also provide a  $\phi$  measurement to compensate for the lack of MDT sensi-  
<sup>441</sup> tivity in this direction. There are four layers of TGC’s in the end cap. TGC’s will be supplanted by  
<sup>442</sup> sTGC’s (small thin gap chambers) in the NSW.

<sup>443</sup> For more details on how detector level trigger objects work in the ATLAS MS, see Appendix A

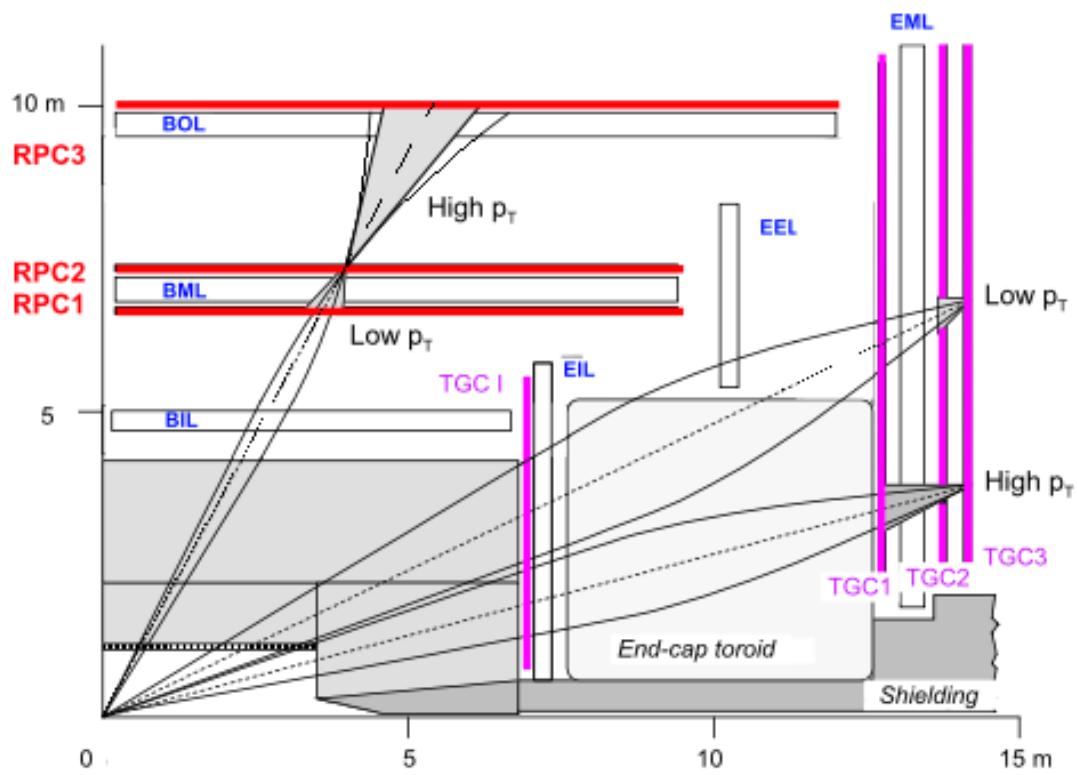


Figure 1.17: ATLAS MS trigger detector arrangement. IC:<sup>36</sup>

<sup>444</sup> 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

445

446

## The Standard Model Higgs and Colider

447

### Event Variables

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

<sup>450</sup> be addressed here, following the discussion and notation of<sup>57</sup> Chapter 11. We then move onto the  
<sup>451</sup> treatment of kinematic variables in collider events, including the two novel schemes considered in  
<sup>452</sup> this thesis, the Lorentz Invariants (LI) and RestFrames (RF) variables.

## <sup>453</sup> 2.1 THE STANDARD MODEL HIGGS BOSON

<sup>454</sup> The generic scalar Lagrangian potential (the kinetic term will be addressed later) for a scalar in the  
<sup>455</sup> SM is:

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

<sup>456</sup> where  $\Phi$  is the Higgs field, a complex scalar doublet under  $SU(2)$ . Its four degrees of freedom are  
<sup>457</sup> typically decomposed as follows:

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

<sup>458</sup>  $\phi^+$  is the complex charged component of the Higgs doublet, and  $\phi^0$  and  $\alpha^0$  are the CP-even and  
<sup>459</sup> CP-odd neutral components, respectively.

<sup>460</sup> If the sign of  $m^2 \Phi^\dagger \Phi$  is negative,  $\Phi$  acquires a *vacuum expectation value* or VEV:

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

<sup>461</sup> with this value typically denoted  $v = \sqrt{2m^2/\lambda} = \sqrt{\sqrt{2}G_F}246$  GeV (with the coupling of the

<sup>462</sup> 4-Fermi effective theory of weak interactions measured through experiments involving muon decay),

<sup>463</sup> and  $\phi^0$  is rewritten as  $\phi^0 = H + v$ .

<sup>464</sup> It is this non-zero VEV that induces spontaneous symmetry breaking in the SM's gauge (local)

<sup>465</sup> symmetry group of  $SU(3)_C \times SU(2)_L \times U(1)_Y$  since the VEV does not respect the  $SU(2)_L \times U(1)_Y$

<sup>466</sup> symmetry of the Lagrangian (i.e.  $\langle \Phi \rangle$  is not invariant under a gauge transformation of this group).

<sup>467</sup> Three of the four generators of this subgroup are spontaneously broken, which implies the existence

<sup>468</sup> of three massless Goldstone bosons, which are in turn “eaten” by linear combinations of the  $W^a$

<sup>469</sup> and  $B$  bosons to form the longitudinal components of the familiar  $W^\pm$  and  $Z$  bosons, with the last

<sup>470</sup> generator giving rise to the usual, unbroken  $U(1)_{EM}$  symmetry and its massless photon,  $A$ , as well

<sup>471</sup> as the scalar Higgs boson  $H$ . To see this one starts with the full Higgs SM Lagrangian (kinetic minus

<sup>472</sup> potential only)

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

<sup>473</sup> One simply plugs in the reparametrized  $\Phi$  with  $\phi^0 = H + v$ , collects the terms involving  $v$  together

<sup>474</sup> with the appropriate  $W$  and  $B$  kinetic terms to extract:

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

<sup>475</sup> This is left as an exercise for the reader; this exercise also makes it manifest that the Higgs couplings

<sup>476</sup> with the  $W^\pm$  and  $Z$  scale quadratically with their masses. Since the Higgs field also respects the

<sup>477</sup>  $SU(3)_C$  color symmetry, the eight gluons are also left massless, and the  $H$  is left interacting with  
<sup>478</sup> photons and gluons primarily through heavy quark loops.

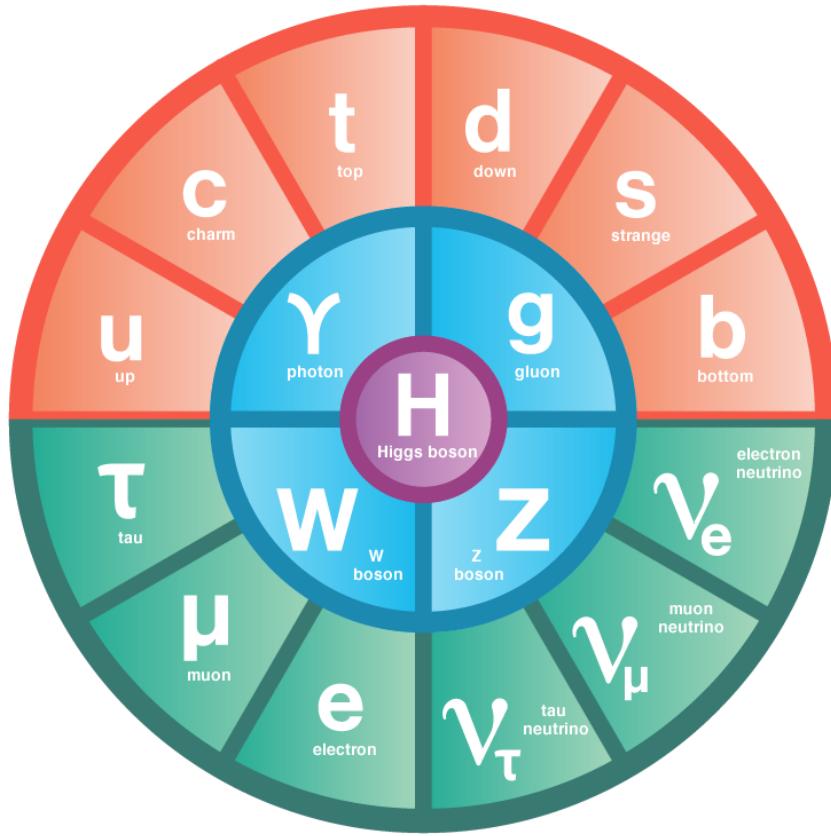


Figure 2.1: The fundamental particles of the Standard Model. IC:<sup>65</sup>

<sup>479</sup> The Higgs is often introduced to the public at large as the mechanism through which fundamen-  
<sup>480</sup> tal fermions (enumerated in Figure 2.1) acquire mass—this is through the Yukawa interactions of the  
<sup>481</sup> Higgs:

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

<sup>482</sup> 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

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

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

## 492 2.2 HIGGS BOSON PRODUCTION AND DECAY AT THE LARGE HADRON COLLIDER

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

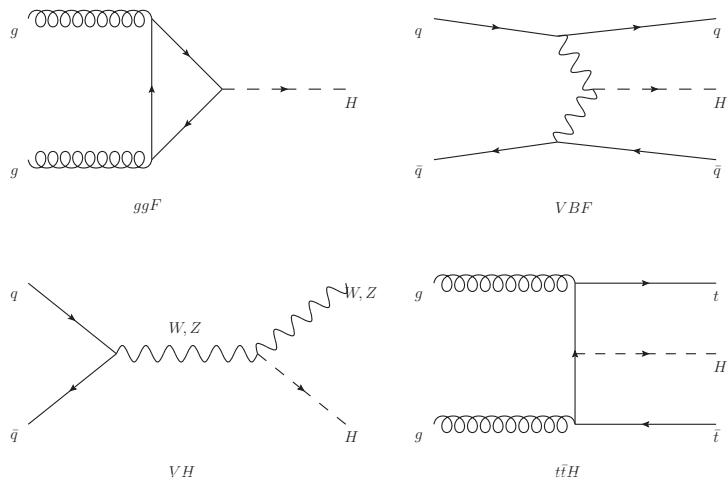


Figure 2.2: Dominant Higgs production modes.

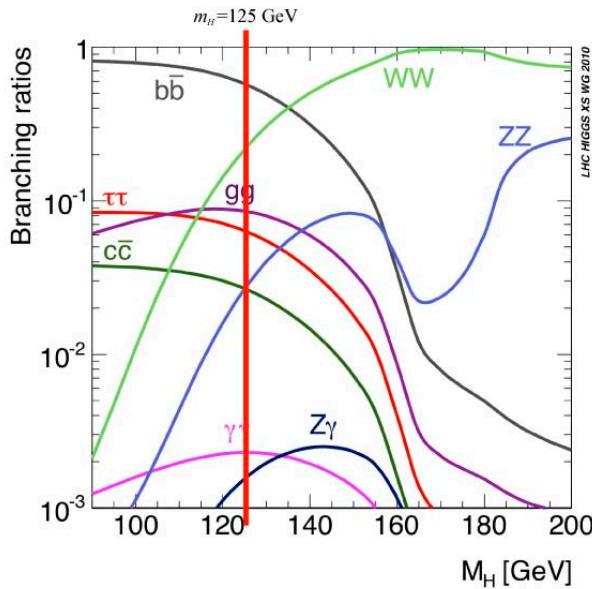
<sup>494</sup> some 87% of Higgs production at the nominal LHC center of mass energy of 14 TeV, is gluon-gluon  
<sup>495</sup> fusion (ggF), shown in Figure 2.2 (a). At high center of mass energies, most of a proton's momen-  
<sup>496</sup> tum is predominantly carried by sea gluons (as opposed to the constituent valence quarks associated  
<sup>497</sup> with the hadron's identity). This, along with the difficulties associated with high luminosity antipro-  
<sup>498</sup> ton beam production, is why the LHC was designed as a proton-proton collider instead of a proton-  
<sup>499</sup> antiproton collider (like the Tevatron). As mentioned above, the Higgs does not couple directly to  
<sup>500</sup> gluons but must instead be produced through the fermion loop shown in the figure. The heaviest  
<sup>501</sup> fundamental fermion by far is the top quark, with  $m_t = 173$  GeV, so top loops dominate this pro-  
<sup>502</sup> cess. While not particularly relevant for this thesis, about 14% of events in the 2-lepton channel of  
<sup>503</sup> the  $H \rightarrow b\bar{b}$  analysis are ggF initiated.

<sup>504</sup> The next most prevalent process is vector boson fusion (VBF), where vector bosons ( $W$  or  $Z$ ,  
<sup>505</sup> denoted generically as  $V$ ) from quarks in the colliding quarks “fuse” to form a Higgs. These quarks  
<sup>506</sup> typically form jets in the forward region, which provides a unique signature for this process. This  
<sup>507</sup> process is not relevant for this thesis.

<sup>508</sup> The third leading process is “Higgsstrahlung” or Higgs production in association with a vector  
<sup>509</sup> boson, often simply  $VH$  production. In this process, a quark-antiquark pair in the colliding protons  
<sup>510</sup> forms an energetic vector boson, which then radiates a Higgs (this is similar to photon emission of  
<sup>511</sup> accelerating electrons, called “bremsstrahlung,” hence the name). Some fraction of the time (about  
<sup>512</sup> 21% of the time for  $WH$  and 6.7% of the time for  $ZH$ ), the energetic  $V$  will decay leptonically (i.e.  
<sup>513</sup> into a decay involving an electron or a muon), which provides a unique and triggerable signature for  
<sup>514</sup> this process. Another 20% of the time for  $ZH$  production, the  $Z$  will decay to neutrinos, which are

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

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

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

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

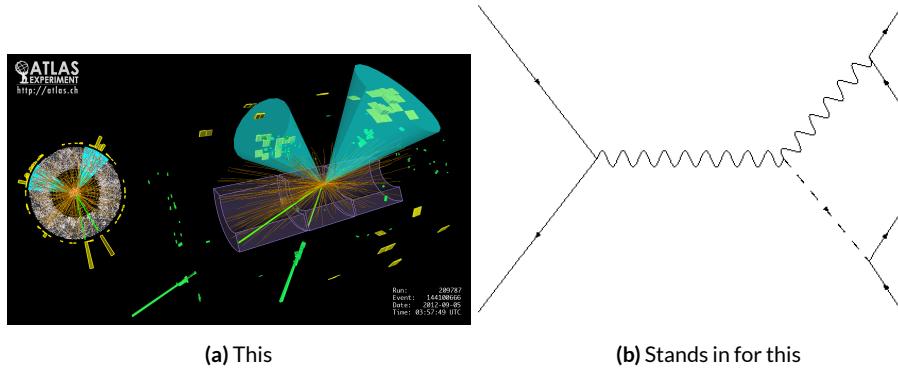
$\sqrt{s}$ (TeV)	ZH	WH	ggF	total $\sigma$	$N_{V \rightarrow \ell^+ \nu} H$
7	$0.34^{+4\%}_{-4\%}$	$0.58^{+3\%}_{-3\%}$	$15.3^{+10\%}_{-10\%}$	17.5	$4.7 \text{ fb}^{-1} \rightarrow 589$
8	$0.42^{+5\%}_{-5\%}$	$0.70^{+3\%}_{-3\%}$	$19.5^{+10\%}_{-10\%}$	22.3	$20.3 \text{ fb}^{-1} \rightarrow 3100$
13	$0.88^{+5\%}_{-5\%}$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

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

### 531 2.3 COLLIDER EVENTS AND EVENT LEVEL VARIABLES

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

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



**Figure 2.4:** Reconstruction in a nutshell

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

---

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

551 2.4 CHARACTERIZATION WITH EVENT-LEVEL VARIABLES

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

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

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

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

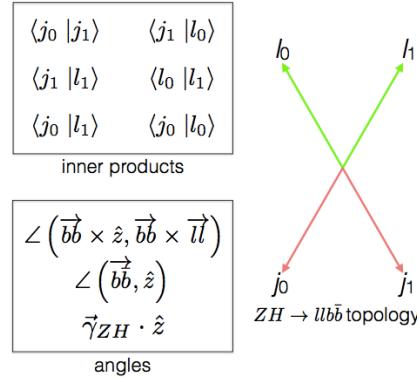
592 used in the VHbb search in Chapter 7.

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

## 603 2.5 LORENTZ INVARIANTS

604 The LI variables, first put forth by S. Hagebeck and others<sup>45</sup>, are based upon the fact that the four-  
605 vectors of an event are determined, all of the information in an event are encoded into their inner  
606 products (Lorentz invariant quantities, hence the name) and the angles between them. This makes  
607 for 16 quantities in all: the ten inner products of the four vectors, the three Euler angles, and the  
608 three parameters specifying the boost of the  $ZH$  system. The masses of the four final state objects  
609 are not considered very useful and so can be removed to leave six meaningful inner products (the  
610  ${}_4C_2$  combinations between distinct final state four vectors). Since these inner products can have an  
611 ill-defined physical interpretation and in order to help MVA training, each inner product is scaled

by:



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

612

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

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

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

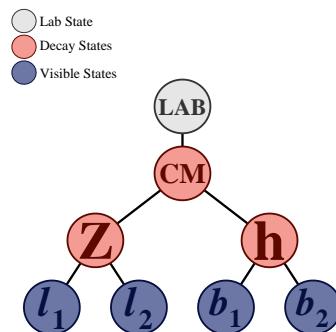
622 These variables do contain a lot of information similar to the usual set: there are mass equivalents

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

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

## 631 2.6 RESTFRAMES VARIABLES

632 The RestFrames variables<sup>47</sup>, calculated using the software package of the same name, is based upon  
 633 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.

634  
 635 erally, one does not typically have enough information to determine exactly each of the intermediate  
 636 rest frames or the boosts between the frames, but in a completely closed final state like  $ZH \rightarrow \ell\ell b\bar{b}$ ,

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

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

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

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

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

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

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

## 2.7 EXTENSIONS TO THE 1 AND 0 LEPTON CHANNELS

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

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

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

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

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

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

676 As alluded to in the 2-lepton discussion, when there is missing information in the final state from  
677 invisible particles and/or combinatoric ambiguities, recursive jigsaw reconstruction (RJR) offers a  
678 standard toolkit for deriving estimated boosts between rest frames by analytically minimizing on  
679 unknown quantities. While in more exotic final states with multiple invisible particles and com-  
680 binatoric ambiguities the choice of jigsaw rule can be subjective, the case of  $W$  is well-studied and  
681 outlined in detail in Section V.A. of<sup>47</sup>. It reproduces the usual transverse mass of the  $W$  in place of  
682  $m_Z$  in the 2-lepton case.

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

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

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

Conventional Wisdom

# 3

688

689

## Data and Simulated Samples

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

694        Details about MC samples may be found in<sup>38</sup>. The  $ZH \rightarrow \ell\ell b\bar{b}$  process is considered for both  
695        multivariate analysis (MVA) optimization and the final statistical analysis, while  $WH \rightarrow \ell\nu b\bar{b}$   
696        and  $ZH \rightarrow \nu\nu b\bar{b}$  production are included in the final statistical analysis only. Signal MC samples  
697        were generated separately for  $qq$  and  $gg$  initiated  $VH$  processes.  $qqVH$  samples were generated with  
698        POWHEG MiNLO + PYTHIA8 with the AZNLO tune set and NNPDF3.0 PDF. Nominal  $ggZH$   
699        samples were generated using POWHEG for the matrix element (ME) and PYTHIA8 for the parton  
700        shower (PS), underlying event (UE), and multiple parton interactions (MPI), again applying the  
701        AZNLO tune and NNPDF3.0 PDF set.<sup>31</sup>

702        The background processes considered in these studies are  $Z+jets$ ,  $t\bar{t}$ , and diboson production  
703        for both MVA optimization and the final statistical analysis and single top production and  $W+jets$   
704        only considered in the final statistical analysis.  $V+jets$  samples are generated using SHERPA 2.2.1<sup>32</sup>  
705        for both the ME and PS. These samples are generated in different groups, according to the identity  
706        of the  $V$ , the max ( $H_T, p_T^V$ ) of events, and also further subdivided according to the flavor of the  
707        two leading jets in an event,  $b$ ,  $c$ , or  $l$ , for a total of six categories..  $t\bar{t}$  samples are generated using  
708        POWHEG with the NNPDF3.0 PDF set interfaced with PYTHIA8 using the NNPDF2.3 PDF's and  
709        the A14 tune<sup>38</sup>. Single top samples use POWHEG with the CT10 PDF's interfaced with PYTHIA6  
710        using the CTEQ6L1 PDF's<sup>14,48</sup>. Diboson samples are generated with SHERPA 2.2.1 interfaced with  
711        the NNPDF3.0 NNLO PDF set normalized to NLO cross sections<sup>25</sup>.

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

Conventional Wisdom

# 4

712

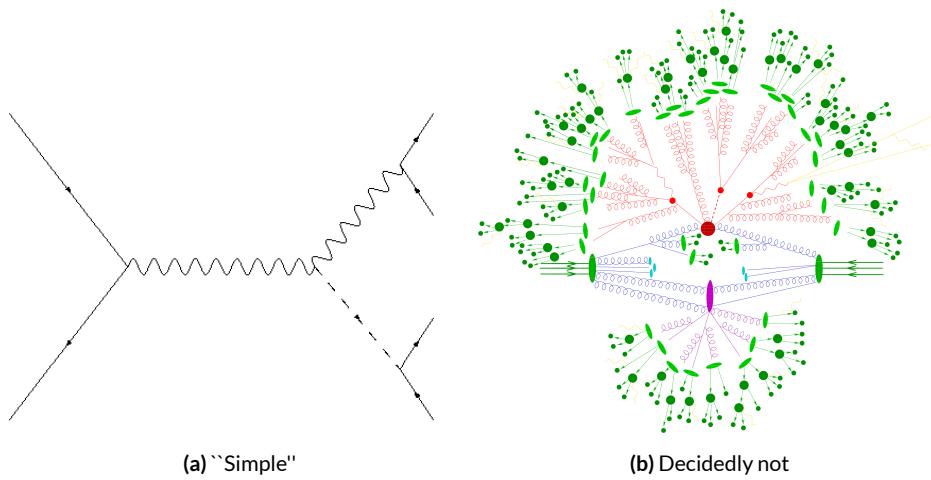
## Signal and Bacground Modeling

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

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

720 **4.1 EVENT GENERATION IN A NUTSHELL**

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



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

727 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.

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

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

745 Typically, it takes several steps and tools to accomplish this. The hard scatter itself is often done  
746 with a dedicated event generator like PowHEG<sup>59</sup> or MadGraph<sup>16</sup>, with events generated being in-  
747 terfaced with a tool like Pythia<sup>62</sup> for the PS, UE, and MPI, though there are exceptions (Sherpa<sup>43</sup>,  
748 for example, can do both for some processes).

749 4.2 DESCRIPTION OF MODELING UNCERTAINTY CATEGORIES

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

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

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

---

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

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

769       **4.2.1 PHYSICS CONSIDERATIONS**

770       As described in the preceding section, many choices must be made when generating events for anal-  
771       ysis studies. In general, evaluating the uncertainties arising from these choices entails generating  
772       alternate samples of events, which practically means tuning parameters in the various software pack-  
773       ages and/or using alternate packages/libraries to make new samples. Once these samples have been  
774       created, they are compared at truth-level (particle level) using a package called Rivet<sup>23</sup> instead of us-  
775       ing the full ATLAS detector reconstruction for computational considerations. Given the nature of  
776       the problem and the tools, there are generally three main categories of physics issues, each described  
777       below.

778       **UNDERLYING EVENT AND PARTON SHOWER**

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

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

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

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

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

## 796 QCD SCALE

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

803 PARTON DISTRIBUTION FUNCTIONS AND  $\alpha_s$

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

809 4.2.2 MODELING SYSTEMATIC TYPES

810 ACCEPTANCE/NORMALIZATION

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

816 As an example, the uncertainty on the theoretical prediction of the  $H \rightarrow b\bar{b}$  branching ratio,  
817 denoted ATLAS\_BR\_bb (it is an ATLAS-wide systematic), is expressed as a normalization systematic  
818 with a value of 1.7%, affecting all  $VH$  processes. Now imagine we have an event in a  $VH$  sample with  
819 weight 1.0. The nominal histograms for this region gets filled with this event's relevant information  
820 with weight 1.0, while the ATLAS\_BR\_bb\_\_1up (\_\_1do) histograms get filled with weight 1.017  
821 (0.983).

822 SHAPE SYSTEMATICS

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

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

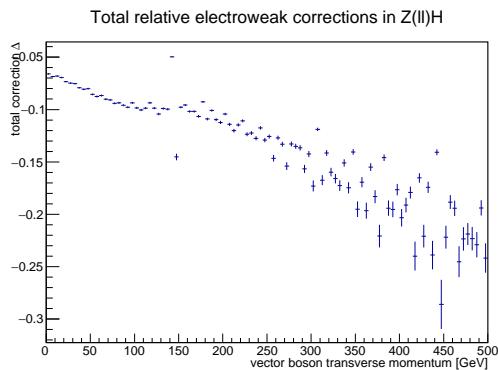
826 where  $w_{event}$  is the simulated event’s weight,  $A_{region}$  is the overall normalization (in principle in-  
827 cluding any systematics), and  $f_{region}(event)$  is some function of event-level variables, usually a single  
828 variable, like  $p_T^V$  or  $m_{bb}$ . The purpose of these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

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

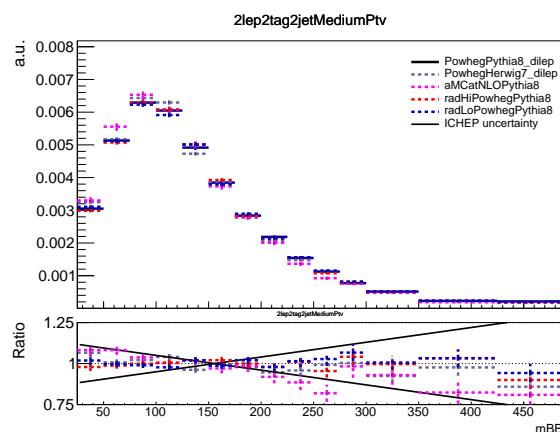
835 An example of the latter case is the systematic associated with the  $m_{bb}$  dependence of the the  
836  $t\bar{t}$  normalization for 2 jet,  $p_T^V \in [75, 150]$  GeV, 2 lepton events. In this case, a variety of effects are  
837 studied (ME, PS, UE), but the treatment of the ME calculation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

---

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



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



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

840 DIVIDING MODELING UNCERTAINTIES: ACCEPTANCE RATIOS

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

847 THEORETICAL ASIDE: STEWART-TACKMANN A way to calculate uncertainties on processes in re-  
848 gions with different numbers of jets was developed by Stewart and Tackmann and is implicitly used  
849 for most  $n_{jet}$  ratio systematics<sup>64</sup>. The problem is how to calculate the cross section and associated  
850 uncertainty for a process with exclusively  $N$  jets in the final state. Generically:

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

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

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

---

systematic covered the largest effect, so others were unnecessary.

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

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

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

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

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

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

872 Of particular importance to ratio systematics is the so-called “double ratio” that is often take as  
873 the scale of variation (plus one). The ATLAS\_UEPS\_VH\_hbb systematic mentioned above, for exam-  
874 ple, has associated with it, ATLAS\_UEPS\_VH\_hbb\_32JR. This systematic is evaluated by dividing the  
875 3 jet to 2 jet ratio in the nominal setup by the same ratio in an alternate setup. Such a ratio generically  
876 looks like:

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

### 877 4.3 PROCESS SPECIFIC SYSTEMATIC SUMMARIES

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

#### 885 4.3.1 SIGNAL PROCESSES

886 Nominal signal  $qqVH$  samples are generated using PowHEG with the MiNLO (multiscale improved  
887 NLO)<sup>53</sup> procedure applied interfaced with PyTHIA8 using the AZNLO tune<sup>12</sup> and NNPDF3.0

Process	Systematics
Signal	$H \rightarrow bb$ decay, QCD scale, PDF+ $\alpha_s$ 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.

888 PDF set<sup>21</sup>. For the 2-lepton case, gluon fusion initiated Higgs production is also considered (ac-  
 889 counting for  $\sim 14\%$  of the total cross section in this channel), with samples generated with Powheg interfaced  
 890 with Pythia8 using the AZNLO tune; the NNPDF2.3 set<sup>20</sup> is used for both the ME and UE+PS.

891 Alternate samples  $qqVH$  samples are generated using MadGraph5\_aMC@NLO<sup>15</sup> for the ME  
 892 and Pythia8 for the UE+PS, hadronizaiton and MPI. The NNPDF2.3\_5f FFN PDF sets and the  
 893 Al14 tune<sup>6</sup>; the latter has variations included. Powheg+Minlo+Herwig<sup>7</sup> were samples were also  
 894 used for systematics.

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

899 Table 4.2, reproduced from<sup>58</sup>, summarizes the signal cross section systematics, which are applied  
 900 uniformly across the analysis channels (as applicable).

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

Sys Name	source	Norm. effect	applied to
ATLAS_BR_bb	$H \rightarrow bb$ decay uncertainties (HO effects, $m_b$ , $\alpha_s$ )	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+ $\alpha_s$ uncertainty	1.9% 1.6%	$qq \rightarrow WH$ $qq \rightarrow ZH$
ATLAS_pdf_Higgs_ggZH	PDF+ $\alpha_s$ uncertainty	5.0%	$gg \rightarrow ZH$

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

samples mentioned above; QCD scale uncertainties were derived by varying scales by  $1/3$  and  $3$ ; and PDF uncertainties were derived by comparing the nominal set with the PDF4LHC15\_30 PDF set<sup>22</sup>.

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

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

905 4.3.2  $V+JETS$

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

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

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

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

---

<sup>¶</sup>SHERPA 2.1 is used for some variations not available in SHERPA 2.2.1.

jet (3+ jet for 2-lepton) and then o-lepton versus 2-lepton (1-lepton) for Z+hf (W+hf<sup>¶</sup>). The final uncertainty contains the sum in quadrature of four effects:

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

933     Summaries of the  $Z$ +jets uncertainties are provided here; the reader is referred to<sup>58</sup> for the  $W$ +jets  
 934     systematics, as these events are virtually non-existent in the 2-lepton case with which this thesis is al-  
 935     most exclusively concerned. In Table 4.4, from<sup>58</sup> 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 Normalisation		
Z+hf	SysZbbNorm_L2_J3	-	-	-	30%	30%	-
	SysZbbNorm_J3	-	-	-	-	-	17%
	SysZbbNorm_0L	-	-	7%	-	-	7%
	SysZbbPTV	effect on each region obtained from shape rw					

**Table 4.4:** Effect of modelling systematics on  $Z$ +jets normalisation in the 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.

936     The flavor composition ratio systematics are in Table 4.5, also from<sup>58</sup>.

<sup>¶</sup>The W+hf CR versus the SR is also considered for W+hf

\*\* cf.<sup>59</sup>, 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 $\geq 3$ 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 $\geq 3$ 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 $\geq 3$ jet)

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

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

#### 4.3.3 TOP-PAIR PRODUCTION

Nominal  $t\bar{t}$  samples are produced with PowHEG at NLO for the ME calculation using the NNPDF3.0 PDF set interfaced with Pythia8.210 using the A14 tune and the NNPDF2.3 PDF set at LO. The parameters `hdamp` (nominal value  $1.5 m_{top}$ , a resummation damping factor for ME/PS matching that can heuristically thought of as tuning high  $p_T$  radiation) in PowHEG and `pThard` (nominal value 0) and `pTdef` (nominal value 2) in Pythia (both control merging with PowHEG) are varied to evaluate certain systematics. Alternative  $t\bar{t}$  samples use PowHEG+Herwig7, MadGraph55\_aMC@NLO+Pythia8.2,

<sup>††</sup>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.

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

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

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

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

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

**Table 4.7:** Effect of modelling systematics on normalisation in the 2lepton regions. The SysTTbarPTV\_L2\_L2 systematic is implemented as a shape systematic over the full  $VpT > 75$  GeV range, and as a result has different acceptance effects in the low and high  $VpT$  regions.

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

---

<sup>#</sup>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.

954    4.3.4 DIBOSON PRODUCTION

955    Three diboson production processes (collectively denoted  $VV$ ) are important for these analyses:  $ZZ$ ,  
 956     $WZ$ , and  $WW$ . Nominal samples are created using **SHERPA 2.2.1** using the NNPDF3.0 PDF set. Al-  
 957    ternative samples use **PowHEG+Pythia8** and **PowHEG+Herwig++**. The methodology here is  
 958    similar to that of the  $t\bar{t}$  systematics, with both overall acceptance and lepton channel specific uncer-  
 959    tainties, with the exception that UE+PS and QCD scale are treated separately (PDF+ $\alpha_S$  was found  
 960    to be negligible).  $p_T^V$  shape systematics are described using linear fits, while  $m_{b\bar{b}}$  shape systematics  
 961    are described using hyperbolic tangents (third degree polynomials) in the 2 jet (3 jet) regions. Once  
 962    again, summary tables from<sup>58</sup> 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.

### 963 4.3.5 SINGLE TOP PRODUCTION

964 Single top sample are generated separately for the different production channels ( $s$ ,  $t$ , and  $Wt$ ) us-  
 965 ing PowHEG with the CT10 NLO PDF's interfaced with PYTHIA6 using the PERUGIA2012 PS  
 966 tune and the corresponding CTEQ6l LO PDF's and PHOTOS (TAUOLA) for QED final state ( $\tau$ )  
 967 decays. Just as with  $t\bar{t}$  samples, PowHEG and PYTHIA settings are varied for certain systematics. Al-  
 968 ternative samples use PowHEG+PYTHIA6 with  $Wt$  diagram subtraction (DS) (instead of “diagram  
 969 removal” for the ME calculation) and MADGRAPH5\_aMC@NLO+HERWIG++. Systematics are  
 970 derived separately in each channel, and are well described in the sumamry Table 4.11 taken from<sup>58</sup>.

---

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

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

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

*“...what would you do first?”*

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

Confucius, *The Analects*

# 5

971

972

## Object and Event Reconstruction and

973

## Selection

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

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

<sup>976</sup> of these descriptions will illuminate each other. Summary tables are almost exclusively taken from <sup>24</sup>  
<sup>977</sup> or <sup>56</sup>. This analysis, like most typical analyses in ATLAS use central object definitions from collabora-  
<sup>978</sup> tion combined performance (CP) groups using standard analysis tools and recommendations from  
<sup>979</sup> these groups for the various objects and their accompanying systematic uncertainties.

<sup>980</sup> Before proceeding to the objects used in this analysis, we begin with a few remarks on uncertain-  
<sup>981</sup> ties associated with object reconstruction. Event-level variables and selections will be discussed more  
<sup>982</sup> in depth in following chapters. As described in Section , systematics quantify the uncertainty asso-  
<sup>983</sup> ciated with certain effects, and are generally treated in an analysis by saving histograms of discrimi-  
<sup>984</sup> nating distributions corresponding to the nominal analysis except with the systematic in question  
<sup>985</sup> varied by plus and minus one standard deviation each (one histogram each). While for modeling  
<sup>986</sup> systematics this only corresponds to different event weight, for experimental systematics like those  
<sup>987</sup> described in this chapter (with the exception of flavor tagging and certain trigger systematics), this is  
<sup>988</sup> done by varying the parameter in question and re-running reconstruction with the systematic varied  
<sup>989</sup> before recomputing all event level quantities and then saving discriminant values in their appropri-  
<sup>990</sup> ate distributions. This is, in general, a much more computationally intensive process in the analysis,  
<sup>991</sup> which is why an entire software framework, the `CxAODFramework`, was created for this analysis (see  
<sup>992</sup> Section 3 of <sup>24</sup> for more details).

## <sup>993</sup> 5.1 TRIGGERS

<sup>994</sup> Tables of the triggers used with the 2015 and 2016 datasets are given in Tables 5.1 and 5.2.

<sup>995</sup> The 0-lepton channel uses a  $\vec{E}_T^{\text{miss}}$  trigger, while 2-lepton channels use single lepton triggers, with

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

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

period	o lep	1 lep	2 lep
A	HLT_xe90_mht_L1XE50	HLT_xe90_mht_L1XE50 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0	HLT_mu24_ilosse(data) HLT_mu24_ilosse_L1MU15(MC) OR HLT_mu40 HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e60_medium OR HLT_e140_lhloose_nod0
B-D <sub>3</sub>	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 <sub>4</sub> -E <sub>3</sub>	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.

<sup>996</sup> the 1-lepton analysis using both. Since the 0- and 1-lepton channels are largely beyond the scope of  
<sup>997</sup> this thesis, the discussion here will be limited to the single lepton triggers; the interested reader is  
<sup>998</sup> directed towards <sup>24</sup> and its cited sources for an in-depth discussion of the use of the  $\vec{E}_T^{\text{miss}}$  trigger.

<sup>999</sup> The efficiency of triggers is in general different on simulated datasets than in actual data collected  
<sup>1000</sup> in ATLAS, so a scale factor to correct for this difference in efficiency must be applied to simulation  
<sup>1001</sup> events to correct for this difference. This scale factor is given by the muon CP group for muons for  
<sup>1002</sup> both the 1- and 2-lepton cases and from the electron CP group for the 1-lepton case. For the two  
<sup>1003</sup> electron case, this was calculated by the analysis team as (details in the <sup>24</sup>):

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

<sup>1004</sup> There are also systematic uncertainties associated with these trigger efficiencies. The single elec-  
<sup>1005</sup> tron trigger efficiency systematic uncertainty is encapsulated in a single systematic, `EL_EFF_Trigger_-`  
<sup>1006</sup> `Total_1NPCOR_PLUS_UNCOR`, while the single muon trigger efficiency has two components, one  
<sup>1007</sup> each for the sample statistics, `MUON_EFF_TrigStatUncertainty`, and systematic uncertainties  
<sup>1008</sup> `MUON_EFF_TrigSystUncertainty` associated with that efficiency's measurement.

<sup>1009</sup> While the momentum associated with the lowest un-prescaled single lepton triggers changes  
<sup>1010</sup> depending on data-taking conditions (the numbers associated with the triggers in the tables can  
<sup>1011</sup> be thought of as nominal  $p_T$  values for trigger level objects), the lowest typical value is  $\sim 25$  GeV.  
<sup>1012</sup> In order to maintain this triggering capability on low  $p_T$  muons in the higher luminosity environ-  
<sup>1013</sup> ment of the Run 3 LHC and beyond, trigger-capable detectors will be installed in upgraded New

1014 Small Wheels (NSW) of the ATLAS muon detector during the Phase I upgrade. Detailed studies in  
1015 simulation of the trigger algorithm performance under nominal and misaligned conditions for the  
1016 Micromegas detectors to be installed in the NSW may be found in Appendix A.

1017 **5.2 ELECTRONS**

1018 Electrons in ATLAS are reconstructed using a combination of the ATLAS electromagnetic calorime-  
1019 ter (ECAL) and inner detector (ID). Reconstruction begins by searching for so-called “seed clusters”  
1020 in the ECAL. The ECAL is divided into a  $200 \times 256$  tower grid in the  $\eta - \phi$  plane, with each tower  
1021 having a size of  $0.025$  square in  $\eta$  and  $\phi$ , corresponding to the granularity of the ECAL in its middle  
1022 layer, with all energy in a tower summed longitudinally. A “sliding window” of  $3 \times 5$  cells in the  
1023 plane is then used to identify EM clusters associated with electrons based on criteria detailed in <sup>41</sup>.  
1024 This comparatively simple algorithm (in contrast to jet reconstruction detailed below) is effective  
1025 since electromagnetic showers have a well defined behavior and shape.

1026 Once seed clusters have been formed, they are associated with tracks in the inner detector. Com-  
1027 bined cluster-tracks pairs form electron candidates. In order for a electron candidate to be consid-  
1028 ered a suitable electron for analysis, it must pass certain quality requirements, based on a cut on the  
1029 value of a likelihood-based (LH) discriminant (cf. <sup>7</sup> for details). This discriminant is given by:

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

1030 where the  $s$  and  $S$  ( $b$  and  $B$ ) subscripts refer to distributions in fiducial signal (background) distribu-

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

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

**Table 5.3:** Electron selection requirements.

### 5.2.1 ELECTRON SYSTEMATICS

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

1047 electron energy) and energy resolution (the width associated with this). This is in practice a very  
1048 complicated procedure, with over 60 systematics associated, but this analysis is not at all sensitive to  
1049 these effects and so a simplified model of two systematics, EG\_RESOLUTION\_ALL and EG\_SCALE-  
1050 \_ALL, is used.

1051 **5.3 MUONS**

1052 This analysis uses the standard CP muon collection in an event, though these muons in ATLAS  
1053 are constructed in a variety of ways; for full details see<sup>18</sup> and<sup>19</sup>. Most muons are constructed using  
1054 tracks in the chambers of the muon spectrometer (MS), with a variety of algorithms available. MS  
1055 tracks are sufficient to reconstruct a muon (a fit on these tracks can be used to point back to an in-  
1056 teraction point for vertex matching, for example) and, in the  $|\eta| \in (2.5, 2.7)$  interval where there  
1057 is no tracking, these standalone (SA) muons are the default. The most common and robust form  
1058 of muon reconstruction combines tracks in the MS with tracks in the ID (more precisely, a global  
1059 refit with hits from both subsystems is typically done) to form combined (CB) muons. CB and SA  
1060 muons automatically pass the loose reconstruction requirements for the Loose muons used in this  
1061 analysis. Additionally, since there is a gap in the  $|\eta| < 0.1$  range in the MS to make room for cabling  
1062 and other detector services, there are two further muon types used in this range: the segment tagged  
1063 (ST) muons that match ID tracks to segments in the MDT or CSC chambers and the calorimeter  
1064 tagged (CT) muons that match ID tracks to calorimeter clusters consistent with minimum ionizing  
1065 particles (which muons in ATLAS generally are).

1066 Further quality requirements are imposed on Loose muons for the different muon categories

1067 used in this analysis. Isolation requirements similar to the electrons in corresponding categories are  
 1068 imposed, and impact parameter requirements are also imposed. The  $ZH$  signal muons also have a  
 1069  $p_T$  cut at 27 GeV and a requirement that the muon fall within the  $|\eta|$  range of the ID.

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

Table 5.4: Muon selection requirements.

### 1070 5.3.1 MUON SYSTEMATICS

1071 Similar to the treatment of systematic uncertainties associated with the electrons, muons have CP de-  
 1072 fined systematics. The muon momentum scale and resolution systematics are divided into three cat-  
 1073 egories associated one for uncertainties related to ID tracks (MUONS\_ID), one for MS tracks (MUONS-  
 1074 \_MS), one for the overall scale (MUONS\_SCALE), and two for charge dependent momentum scales  
 1075 (MUON\_SAGITTA\_RHO and MUON\_SAGITTA\_RESBIAS). The remaining systematics have a STAT  
 1076 and SYS component corresponding to the sample statistics and systematic uncertainties for their  
 1077 individual components. For Efficiency scale factors use different standard candles in different  $p_T$   
 1078 ranges ( $J/\psi$ 's ( $Z$ 's) below (above) 15 GeV), and so these systematics are broken up into two categories  
 1079 (MUON\_EFF\_STAT and MUON\_EFF\_SYS; MUON\_EFF\_STAT\_LOWPT and MUON\_EFF\_SYS\_LOWPT).  
 1080 There are also isolation systematics (MUON\_ISO\_STAT, MUON\_ISO\_SYS) and track to vertex associa-  
 1081 tion systematics (MUON\_TTVA\_STAT, MUON\_TTVA\_SYS).

1082    5.4 MISSING TRANSVERSE ENERGY

1083    High precision performance of  $\vec{E}_T^{\text{miss}}$  is not so crucial to the 2-lepton analysis (though it is very im-  
1084    portant to the other channels), so the interested reader is referred to<sup>19</sup> and its references for an in-  
1085    depth discussion.  $\vec{E}_T^{\text{miss}}$  in ATLAS is the negative vectoral sum of physics objects (in this analysis just  
1086    jets and leptons, though in principle also including  $\tau$ 's and  $\gamma$ 's) and a so-called track based soft term  
1087    (TST). The TST is comprised of valid ID tracks not associated with any physics objects in an event.  
1088    These tracks must be associated to an event's primary vertex, have a  $p_T > 0, 4 \text{ GeV}$ , and pass other  
1089    quality requirements.

1090    The  $\vec{E}_T^{\text{miss}}$  systematic uncertainties relevant to this analysis are related to track based energy scale  
1091    and resolutions in both the soft term and in the jets and are: MET\_SoftTrk\_ResoPara, MET\_Soft-  
1092    Trk\_ResoPerp, MET\_SoftTrk\_ScaleDown, MET\_SoftTrk\_ScaleUp, MET\_JetTrk\_Scale-  
1093    Down, and MET\_JetTrk\_ScaleUp.

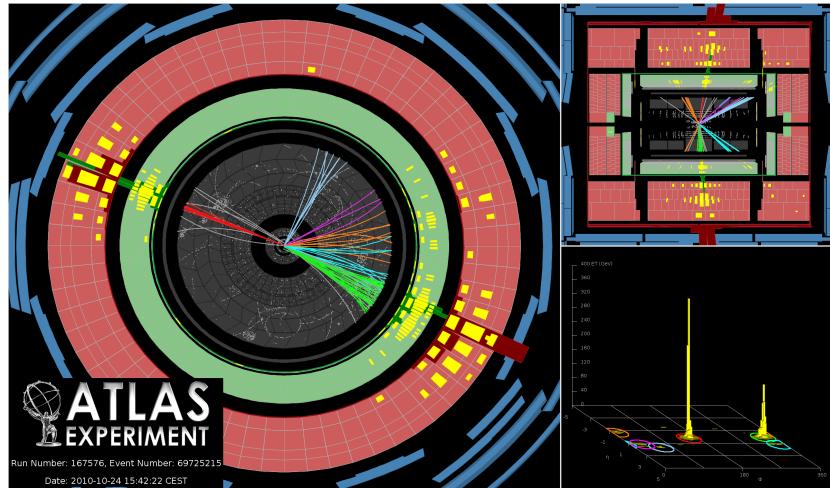
1094    5.5 JETS

1095    5.5.1 JET ALGORITHMS

1096    The hadronic nature of jets makes jet reconstruction a lot more complicated than electron or photon  
1097    reconstruction, where a regular shower shape and track matching (or lack thereof in the case of the  
1098    chargeless photon) provide a fairly straightforward and robust approach. The interested reader is  
1099    referred to<sup>61</sup> for an excellent survey, from which this discussion is greatly abbreviated.

1100    Looking at an event like the one in Figure 5.1, the jets are particularly easy to identify, more or less

1101 popping out of the plane plot, but this is not always the case.



**Figure 5.1:** A clean ATLAS dijet event.

1102 Two general methods of turning particles/calorimeter towers into jets exist: cone-based and se-  
1103 quential recombination. The general theme of the former is to find a hard (energetic) particle and  
1104 draw a circle around it in the plane in an intelligent manner, while the theme of the latter is to find  
1105 some metric of distance between particles and then to cluster pairs based on this distance into jets in  
1106 an intelligent way. Cone-based algorithms are simple (and therefore generally quite fast) but gener-  
1107 ally lack some nice properties of the sequentially recombined jets (though there are notable excep-  
1108 tions like SISCone). These jets are important for trigger level objects in ATLAS, though since no jet  
1109 triggers are used in this analysis, they will not be discussed any further here.

1110 The general drawback of cone-based algorithms is that they are not infrared and collinear (IRC)  
1111 safe. That is, the emission of a soft (IR) quark or gluon during hadronization should not change  
1112 the final jet collection in an event, and neither should the collinear splitting of hard particles during

1113 hadronization. These are fairly common edge cases and can lead to certain pathologies in QCD cal-  
1114 culations. Infrared and collinear safety are diagrammed schematically in Figures 5.2 (a) and (b), taken  
1115 from <sup>61</sup>.

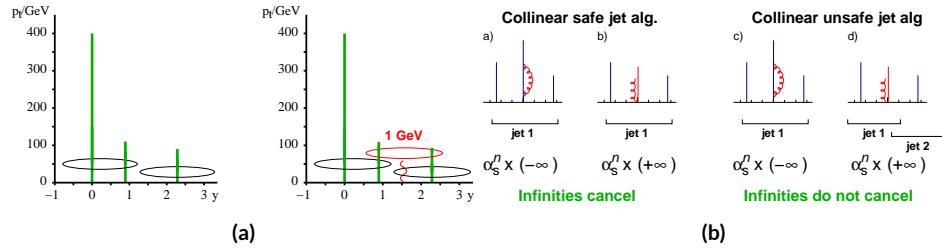


Figure 5.2: Infrared (a) and collinear (b) safety.

1116 The sequential recombination algorithms are generally safe from these effects, as these edge cases  
1117 are very “close” to each other by construction. A sequential recombination algorithm proceeds as  
1118 follows

- 1119 1. Evaluate the set of distances  $d_{ij}$  (for pairs of objects) and  $d_{iB}$  (the “beam distance” for each  
1120 individual object)

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

- 1121 2. Find the minimum distance
- 1122 3. If the minimum distance is:
- 1123 • A  $d_{ij}$ : cluster these objects together, and go to step 1
  - 1124 • A  $d_{iB}$ : call the  $i^{\text{th}}$  object a jet, remove it from the set of objects to be clustered, and go to  
1125 step 1
- 1126 4. Repeat until all objects are clustered into jets

1127 The choices one must make in sequential recombination are the size parameter  $R$ , akin to a cone

1128 radius in cone-based algorithms, and the momentum power  $p$ . Common choices and their trade-offs

1129 are:

- 1130 • +1: the  $k_t$  algorithm; favors the softer particles in an event, so the cluster sequence gives a  
1131 history of hadronization, but jet shapes are irregular
- 1132 • 0: the Cambridge-Aachen algorithm: a pure distance metric; less substructure but jets are less  
1133 “weird” looking
- 1134 • -1: the anti- $k_t$  algorithm: clustering begins with hardest particles in an event; regular, localized  
1135 jet shapes, but virtually no substructure in clustering history

1136 As shown in Figure 5.3

1137 All three algorithms have uses for different applications in ATLAS, with anti- $k_t R = 0.4$  jets

1138 being the default jet collection. These are the jets used in this analysis.

1139 If the choice of jet algorithm seems a little arbitrary, it is. There is no one-size-fits-all jet collection

1140 perfect for every application, and analyzers have to make these choices for themselves. One interest-

1141 ing choice is the jet size parameter,  $R$ . A large  $R$  jet will contain more of the radiation coming from a

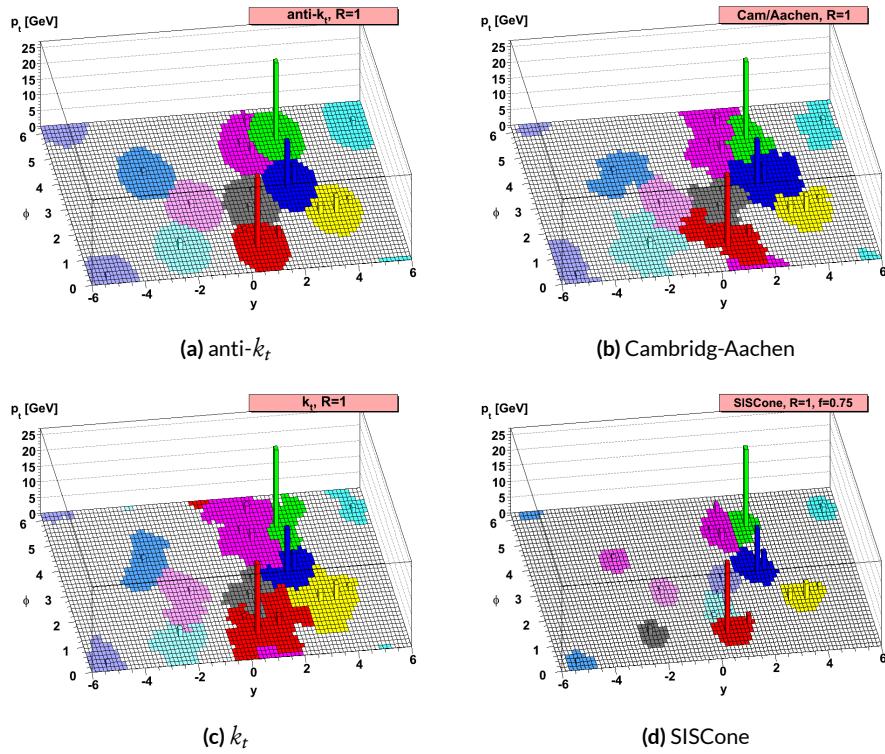
1142 final state object, but its large size makes it susceptible to contamination from the underlying event

1143 and pileup, with small  $R$  jets having the opposite features.  $R = 0.4$  is a fairly middle-of-the-road

1144 choice. A natural question to ask is whether there needs to be just one jet collection in an analysis.

1145 Might there not be more information to be gained from looking at more jet sizes or clusterings? Pre-

1146 liminary studies point to this answer being yes and are addressed in Appendix B.



**Figure 5.3:** Different jet algorithms used on the same event. IC:<sup>61</sup>

1147 5.5.2 STANDARD ATLAS HBB JETS

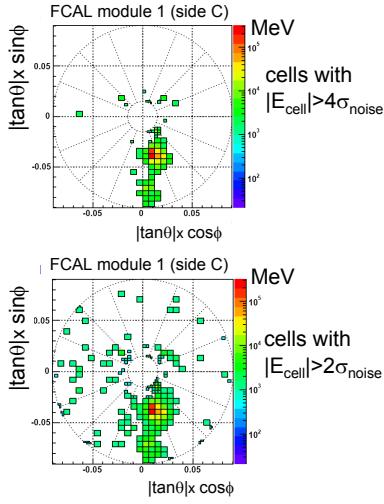
1148 There are a few considerations that arise with jets in physical detectors. The first is what type of  
1149 object to use as the stand in for “particles” since ATLAS measures energy deposits, not pions. The  
1150 approach ATLAS has settled upon are calorimeter topological clusters (or CaloTopoClusters for  
1151 short)<sup>49</sup>. Unlike the sliding window algorithm used for electron clusters, CaloTopoClusters use  
1152 a noise significance based approach in the “4-2-1” algorithm. Each cell in the electromagnetic and  
1153 hadronic calorimeters (ECAL and HCAL, respectively) has associated with it a characteristic noise  
1154 level, with this noise level in each channel, it is possible to construct a “significance” for the regis-  
1155 tered energy deposit in the event by dividing the measured value by its characteristic noise. Groups  
1156 of cells having a significance of 4 are taken as the centers of clusters in the plane. The second layer  
1157 in a cluster includes all neighboring cells to the central layer with significance of 2, and the final layer  
1158 includes all the nearest neighbors to the second layer. This is described in Figure 5.4 from<sup>52</sup>.

1159 Once CaloTopoClusters have been formed and clustered into jets, they are calibrated using the  
1160 electromagnetic (EM) scale (the scale for clusters coming form EM showers). Further details may be  
1161 found in<sup>9</sup>.

1162 Jets in this analysis fall into two categories, “signal” and “forward” jets, and are required to pass  
1163 certain quality requirements, described in Table 5.5. All jets must pass a series of jet cleaning require-  
1164 ments using calorimeter level variables to eliminate jets coming from problematic calorimeter cells  
1165 and certain backgrounds. Some signal jet candidates also make use of a Jet Vertex Tagger (JVT) that  
1166 uses primary vertex and jet and track  $p_T$  information to decide whether certain soft jets are likely

## Local Hadronic Calibration: Clusters

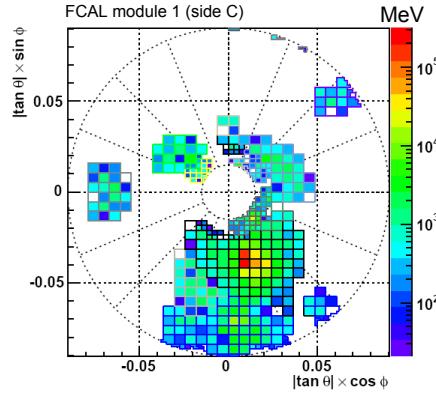
Sven Menke



□ Topological clustering

❖ 4,2,0 clusters in FCal

➢ jets with  $p_T > 50 \text{ GeV}$



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

M. Lefebvre, P. Loch

33

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

- <sup>1167</sup> to have come from the primary (hard scatter) vertex in an event or are to be considered pileup.  
<sup>1168</sup> Further details on JVT may be found in <sup>17</sup>. Jets are further corrected using standard CP tools and a  
<sup>1169</sup> dedicated PtReco correction, all outlined in Section 7.3 of <sup>24</sup>.

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.

- <sup>1170</sup> Overlap removal in this analysis is done according to the following precedence, taken from <sup>24</sup>

1171 with further steps only taken into account if an object survives previous steps:

1172 • tau-electron: If  $(\tau, e) < 0.2$ , the  $\tau$  lepton is removed.

1173 • tau-muon: If  $(\tau, \mu) < 0.2$ , the  $\tau$  lepton is removed, with the exception that if the  $\tau$  lepton has  
1174  $> 50$  GeV and the muon is not a combined muon, then the  $\tau$  lepton is not removed.

1175 • electron-muon: If a combined muon shares an ID track with an electron, the electron is re-  
1176 moved.

1177 If a calo-tagged muon shares an ID track with an electron, the muon is removed.

1178 • electron-jet: If  $(jet, e) < 0.2$  the jet is removed.

1179 For any surviving jets, if  $(jet, e) < \min(0.4, 0.04 + 10 \text{ GeV}/e)$ , the electron is removed.

1180 • muon-jet If  $(jet, \mu) < 0.2$  or the muon ID track is ghost associated to the jet, then the jet is re-  
1181 moved if the jet has less than three associated tracks with  $> 500$  MeV ( $\text{NumTrkPt}500\text{PV}^{\text{jet}} < 3$ )  
1182 or both of the following conditions are met: the ratio of the muon and jet is larger than  $0.5$  ( $\mu/\text{jet} >$   
1183  $0.5$ ) and the ratio of the muon to the sum of tracks with  $> 500$  MeV associated to the jet is larger  
1184 than  $0.7$  ( ${}^{\text{muon}}/\text{SumPtTrkPt}500\text{PV}^{\text{jet}} > 0.7$ ).

1185 For any surviving jets, if  $(jet, \mu) < \min(0.4, 0.04 + 10 \text{ GeV}/\mu)$ , the muon is removed.

1186 • tau-jet: If  $(\tau, jet) < 0.2$ , the jet is removed.

1187 • electron-fat jet: If  $(e, \text{fat jet}) < 1.2$ , the fat jet is removed.

1188 Jets are corrected using a muon-in-jet correction and then a kinematic fitter (Appendix D of <sup>56</sup>)

1189 for the 2-lepton case (PtReco correction for the 0- and 1- lepton case). The muon-in-jet correction

1190 is designed for  $b$ -jets. Since the decay of a  $b$ -quark to a  $c$ -quark and finally to a light quark (these are  
1191 the multiple vertices for which JetFitter searches) involves two weak decays, there are two  $W$ -bosons

<sup>1192</sup> involved in the decay. Some of these will decay semileptonically, and, while electron and  $\tau$  energy  
<sup>1193</sup> will be captured by the calorimeters, semileptonic  $\mu$ 's will only be registered in the MS, which occurs  
<sup>1194</sup> in some 44% of all decays from a theoretical standpoint, which amounts to about 12% in practice  
<sup>1195</sup> (due to track isolation requirements for the leptons). This value is about 1–2% for electrons, which  
<sup>1196</sup> deposit their energy in the calorimeter and so require no correction; any jet with a valid lepton associ-  
<sup>1197</sup> ated to it is deemed semileptonic (all others are called hadronic). Any jet with muons associated with  
<sup>1198</sup> it has the closest muon's 4-vector (in the plane) added to it.

<sup>1199</sup> The PtReco correction is a scale factor on the muon-in-jet corrected jet's 4-vector based on the  
<sup>1200</sup> jet's  $p_T$  and whether the jet is hadronic or semileptonic. This correction factor is based on particle  
<sup>1201</sup> level studies done on a TruthWZ sample. As the 0- and 1-lepton cases are not the focus of this thesis,  
<sup>1202</sup> the interested reader is directed to Section 7.3 of<sup>24</sup>.

<sup>1203</sup> The kinematic fitter used in 2-lepton events with two or three jets takes as its input 12 fit parame-  
<sup>1204</sup> ters,

- <sup>1205</sup> • energies of 2 electron or of 2 muons
- <sup>1206</sup> • energies of 2  $b$ -jets
- <sup>1207</sup> •  $\eta, \phi$  of 2 leptons and 2 jets
- <sup>1208</sup> •  $p_X$  and  $p_Y$  of  $\ell\ell b\bar{b}$  system.
- <sup>1209</sup> •  $m_{\ell\ell}$

<sup>1210</sup> and 3 constraints for the variation of these parameters,

- <sup>1211</sup> • parameters : Gaussian (b-jet energy : Transfer Functions (TF); these are denoted  $L$ , with an  
<sup>1212</sup>  $L_{truth}$  as a prior) (the  $\phi$  parameters)

<sup>1213</sup> •  $p_X$  and  $p_Y$  of  $\ell\ell b\bar{b}$  system : zero with a width of 9 GeV obtained from  $ZH$  signal MC.

<sup>1214</sup> •  $m_{ll}$  : Breit-Wigner (BW) distribution of  $Z$  boson (final term, leptons denoted  $\Omega$ )

<sup>1215</sup> which leads to test statistic from the usual likelihood formalism to be minimized in each event:

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

### <sup>1217</sup> 5.5.3 JET SYSTEMATICS

<sup>1218</sup> As with the electron systematics, jet energy scale (JES) and resolution (JER) are the two principal

<sup>1219</sup> considerations for systematic uncertainties, with even more standard. JER, as with the electron en-

<sup>1220</sup> ergy resolution, is a single systematic uncertainty, `JET_JER_SINGLE_NP`. There is also a single JVT

<sup>1221</sup> efficiency `JET_JvtEfficiency` systematic uncertainty. There are 88 nominal JES systematics, and

<sup>1222</sup> this analysis is sufficiently sensitive to these variations that a single systematic is grossly insufficient.

<sup>1223</sup> Nevertheless, some simplification is possible, with the 75 of these nuisance parameters (mostly sta-

<sup>1224</sup> tistical uncertainties related to the  $Z$ +jet and  $\gamma$ +jet calibrations) being reduced to 8, and 13 explicitly

<sup>1225</sup> named nuisance parameter. These remaining 13 are related to the  $\eta$  intercalibration used to extrap-

<sup>1226</sup> olate standard calibrations to other jet  $\eta$  regions, uncertainties related to the flavor composition of

<sup>1227</sup> principle background samples ( $W/Z$ +jets, top, and diboson), a single systematic for the  $b$ -jet energy

1228 scale, a high  $p_T$  jet energy scale systematic, and one for jets that punch through the HCAL to leave  
1229 energy deposits in the MS. These are listed explicitly in Table 5.7.

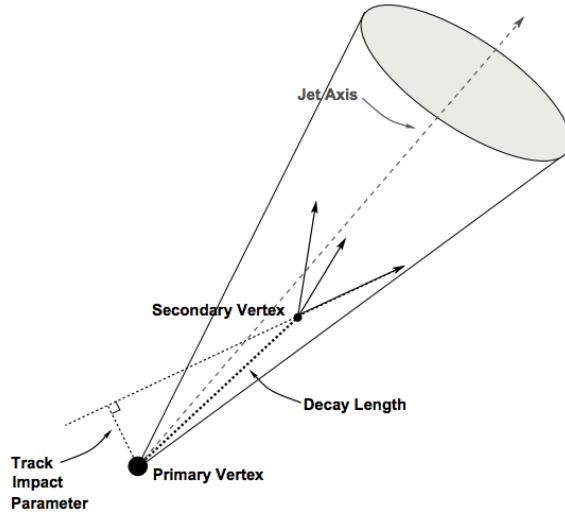
1230 5.5.4 FLAVOR TAGGING

1231 Given that the final state in this analysis involves pairs of jets originating from  $b$ -quarks, deploying  
1232 effective flavor tagging algorithms is imperative. While flavor tagging in general can be used to isolate  
1233 any flavor ( $b$ ,  $c$ , or light ( $u$ ,  $d$ ,  $s$ , or gluon-initiated jets)), this analysis exclusively looks for  $b$ -jets, so  
1234 this discussion will focus on  $b$ 's. At truth-level in simulation this is fairly straightforward: one need  
1235 only look at the particles contained within a jet and seeing if any include a  $b$ -quark (sometimes a  $B$   
1236 hadron) in the decay chain.

1237 DIRECT TAGGING

1238 One of the most distinctive features of  $b$ -jets is the presence of secondary vertices, as illustrated in  
1239 Figure 5.5. While most partons created in particle collisions will hadronize promptly,  $b$ -quarks will  
1240 first hadronize into  $B$ -hadrons, which have lifetimes of about a picosecond. This small but finite  
1241 lifetime means that these particles will travel about half a millimeter or so before decaying into a jet  
1242 in much the usual way, and the tracks from this decay will point back to this displaced, secondary  
1243 vertex.

1244 There are various secondary vertex algorithms used as inputs to the nominal  $b$ -tagging algo-  
1245 rithm<sup>8</sup>, with three main types of algorithms used as inputs  
1246 1. Track impact parameter based algorithms:  $I_2PD$  (signed transverse only; more pileup robust),  
1247  $I_3PD$  (signed transverse and longitudinal)



**Figure 5.5:** An illustration of a secondary vertex in a  $b$ -jet. Image credit:<sup>46</sup>

- 1248     2. Inclusive secondary vertex reconstruction: SV1 (start with two track vertex pairs and con-
- 1249       struct a secondary vertex)
- 1250     3. Multiple vertex reconstruction (decay chain): JetFitter ( $PV \rightarrow b \rightarrow c$  decay chain using Kalman
- 1251       filter)

1252       All of these are combined into a boosted decision tree (BDT) and trained on five million  $t\bar{t}$  events  
 1253       with an 90%/10%  $c$ /light jet background to form the MV<sub>2c10</sub> algorithm, with 10 referring to the  
 1254       percentage of charm samples in the training background. The 10% charm ratio was found to be a  
 1255       good balance between increased charm rejection capability (as opposed to MV<sub>2c00</sub>, which has no  
 1256       charm in the background training) and loss in light jet rejection (compared to MV<sub>2c20</sub>, which has

1257       20

1258       ANALYSIS SPECIFIC CONCERNS AND SYSTEMATIC UNCERTAINTIES   In addition to specifying  
 1259       the tagging algorithm, the working point efficiency must be specified. As with selection algorithms

1260 in general, there is a trade off between efficiency (identifying all the  $b$ -jets, minimizing type II error)  
 1261 and purity (making sure all jets positively identified are in fact  $b$ -jets, minimizing type I error). Nom-  
 1262 inal efficiency working points have been calibrated by the flavor tagging CP group and are outlined  
 1263 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).

1264 These values are aggregate figures, as both the jet's  $p_T$  and  $\eta$  are inputs to the MV2c10 discrim-  
 1265 inant. The working point chosen for this analysis is the 70% FixedCutBEff\_70 working point,  
 1266 with “fixed cut” referring to the fact that this particular usage of the MV2c10 BDT value is a simple  
 1267 cut value.

1268 Just as with the trigger and lepton identification efficiencies, efficiencies differ from their nom-  
 1269 inal values somewhat depending on what simulation or data sample is being used. To account for  
 1270 this difference, just as in the other cases, scale factors are applied to simulation event weights. It is  
 1271 through these event weights, as with the modeling systematics, that the flavor tagging systematic  
 1272 uncertainties are applied. Given that there are 24 input variables to MV2c10 and that flavor tagging  
 1273 is in general a very difficult problem, it is not surprising that, as with the JES, there are very many  
 1274 systematic uncertainties associated with flavor tagging. However, as with JES, the CP group has com-  
 1275 pacted the full systematic set into a reduced set of 13 systematic uncertainties: 3 each associated with

<sub>1276</sub>  $c$  and light jets,  $\varsigma$  for  $b$ -jets (with the naming convention `FT_EFF_Eigen_(B|C|Light)N`), one  
<sub>1277</sub> for the extrapolation of scale factors to different jet  $p_T$  regimes (`FT_EFF_Eigen_extrapolation`),  
<sub>1278</sub> and one for the charm to bottom extrapolation (`FT_EFF_Eigen_extrapolation_from_charm`)<sup>67</sup>.  
<sub>1279</sub> This schematic is a middle-of-the-road “Medium” set of systematics.

## <sub>1280</sub> TRUTH TAGGING

<sub>1281</sub> Since imposing a  $z$   $b$ -tag requirement overwhelmingly rejects events dominated by  $c$ - and light jets,  
<sub>1282</sub> statistics in such MC samples are very low. In order to circumvent this problem and restore full MC  
<sub>1283</sub> statistics, the tag rate function, or “truth-tagging” procedure (in contrast to the standard or “direct  
<sub>1284</sub> tagging” procedure) is applied, in which all events are kept but given a weight that preserves the  
<sub>1285</sub> overall shape and normalization of underlying distributions. Intuitively, this is done by giving events  
<sub>1286</sub> with real  $b$ -jets in MC a much higher weight than events having only  $c$ - or light jets. Truth-tagging is  
<sub>1287</sub> applied to all samples when conducting MVA training in order to maximize statistics and reduce the  
<sub>1288</sub> risk of overtraining. Truth-tagging is also used for data-MC comparison plots in  $z$ -tag regions and  
<sub>1289</sub> for  $V + cc$ ,  $V + c\ell$ ,  $V + \ell$ , and  $WW$  samples used in the final likelihood fit. A detailed description of  
<sub>1290</sub> the truth-tagging process is provided below.

<sub>1291</sub> Each jet in a given event has associated with it a  $b$ -tagging efficiency, denoted  $\varepsilon$ , that is a function  
<sub>1292</sub> of its  $p_T$ ,  $\eta$ , and real flavor ( $b$ ,  $c$ , or light) from truth-level information in MC. Intuitively, this effi-  
<sub>1293</sub> ciency can be thought of as the likelihood that a given jet will be  $b$ -tagged. Hence,  $b$ -jets have a much  
<sub>1294</sub> higher  $b$ -tagging efficiency than  $c$ -jets, which in turn have a higher  $b$ -tagging efficiency than light jets.  
<sub>1295</sub> We define a truth-tag weight for a given combination of tagged and untagged jets as the product of

1296 the efficiencies of the tagged jets times the product of the complement of the efficiencies of the un-  
 1297 tagged jets. For example, for an event with three jets, labeled 1, 2, and 3, if jets 1 and 2 are tagged, and  
 1298 jet 3 is untagged, the truth-tag weight associated with this combination is

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

1299 In order to obtain a truth-tag weight for an event, one takes the sum of the weights for each pos-  
 1300 sible tag combination. The current analysis requires that all events have exactly two  $b$ -tagged jets, so  
 1301 the truth-tag weight is the sum of all the weights of all possible pairs of tagged jets (events with fewer  
 1302 than two jets are discarded). Going back to the three jet example, one has the possible combinations:  
 1303 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  
 1304 tagged and jet 1 as untagged, which yields a total event weight of

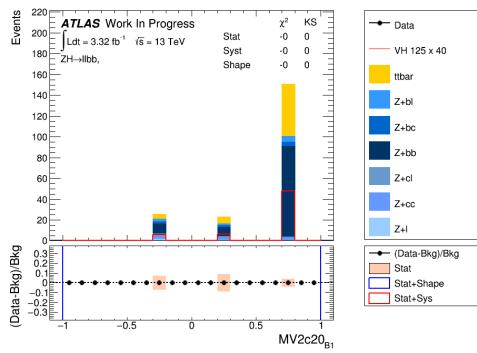
$$w_{tot} = \varepsilon_1 \varepsilon_2 (1 - \varepsilon_3) + \varepsilon_1 \varepsilon_3 (1 - \varepsilon_2) + \varepsilon_2 \varepsilon_3 (1 - \varepsilon_1) \quad (5.6)$$

1305 For some applications (e.g. in order to use variables like pTB1, the  $p_T$  of the harder  $b$ -tagged jet  
 1306 in an event, in MVA training), it is necessary to choose a combination of jets in an event as “tagged.”  
 1307 This combination is chosen randomly, with the probability for a given combination to be chosen  
 1308 being proportional to its truth-tag weight. In the three jet example, the probability of tagging jets  
 1309 1+2 is:

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

1310     Though not used in the current analysis, functionality exists for generic truth-tagging require-  
 1311     ments (i.e. an arbitrary number of tags on an arbitrary number of jets) through the logical combina-  
 1312     toric extension and for so-called “pseudo-continuous tagging,” where a  $b$ -tag score is generated for  
 1313     each jet in a given event. Since a random combination of jets is set by hand to pass the  $b$ -tagging cuts  
 1314     regardless of its  $b$ -tag score, a new score must be generated if this information is to be used in further  
 1315     analysis. Under current settings, jets that are tagged are assigned a random  $b$ -tag score that is sampled  
 1316     from the MV2c10 cumulative distribution above the 70% efficiency working point cut. All other  
 1317     jets in the event are assigned a random  $b$ -tag score below the 70% working point cut. Since these dis-  
 1318     tributions are discrete, the scores are not truly continuous (cf. example distributions in Figure ??),  
 hence the “pseudo-continuous” nomenclature.

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



1319  
 1320     A number of closure tests were performed on both the nominal and several systematics cases. In  
 1321     the plots that follow, truth (solid) and direct (dashed) tagging distributions for  $m_{bb}$  and  $\Delta R(b_1, b_2)$   
 1322     in different  $p_T^V$  regimes for 2 lepton, 2 jet events. Agreement between the truth and direct tagging  
 1323     cases is generally very good, an example of which can be seen in Figure ?? for a signal qqZllH125 sam-

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

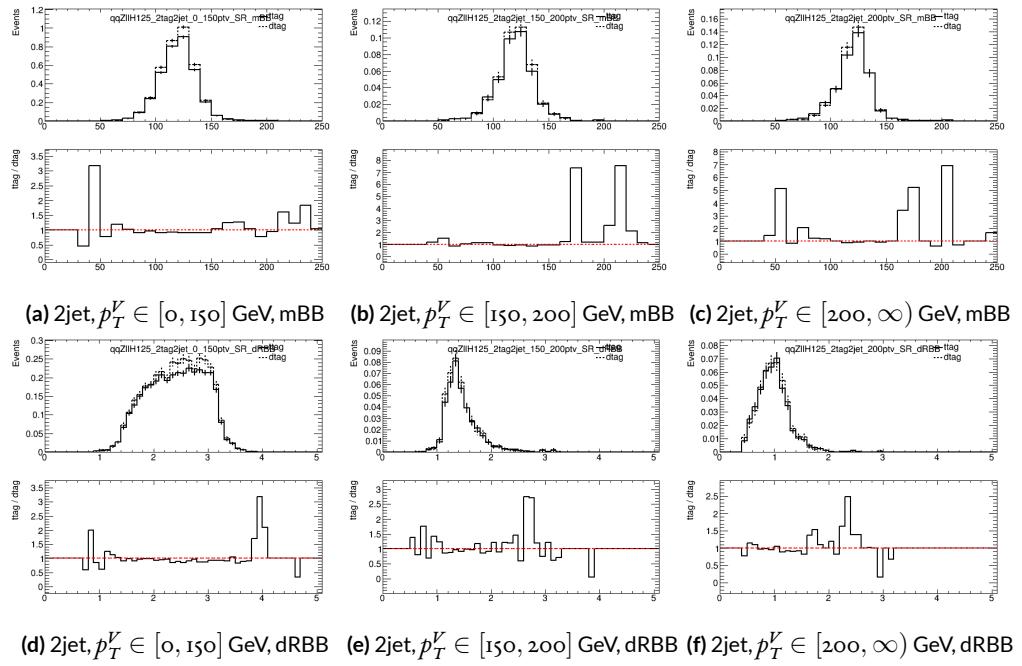
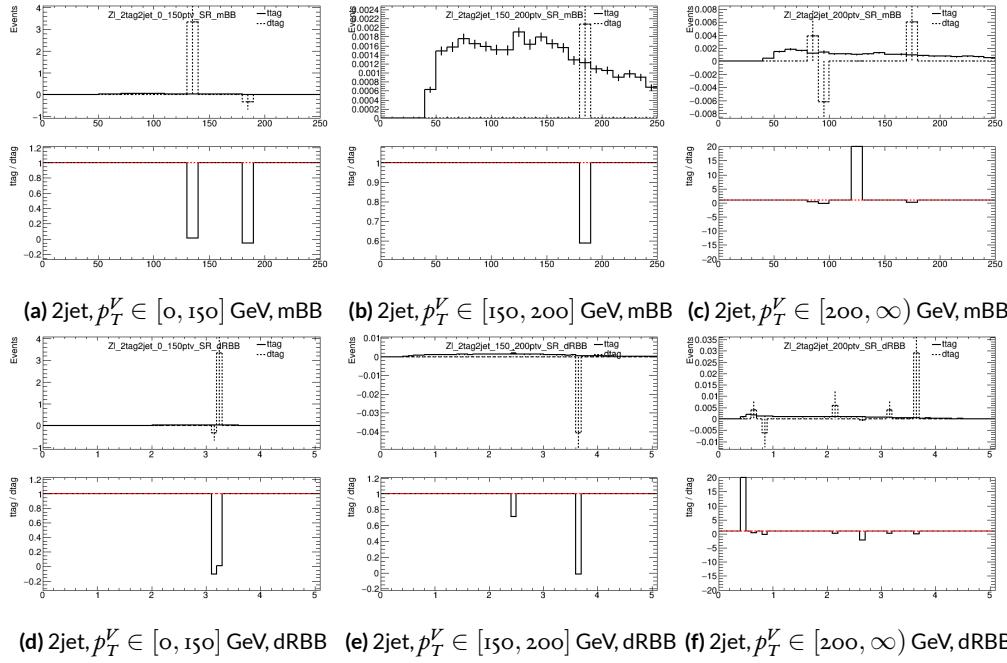
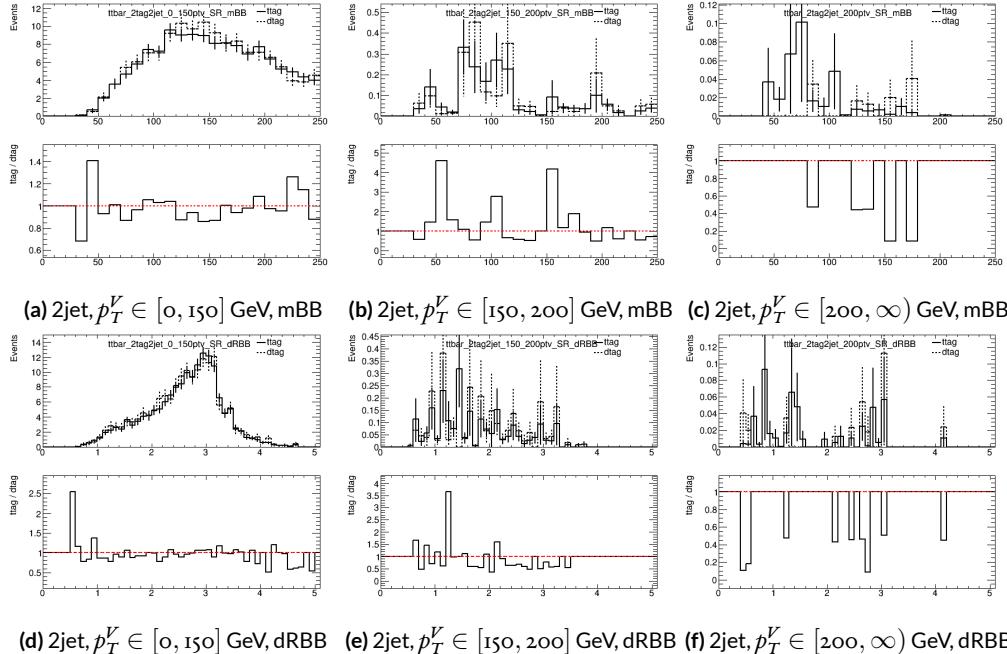


Figure 5.7: Truth-tagging closure tests for 2 lepton, 2 jet qqZIIH125 samples in three different  $p_T^V$  regions.

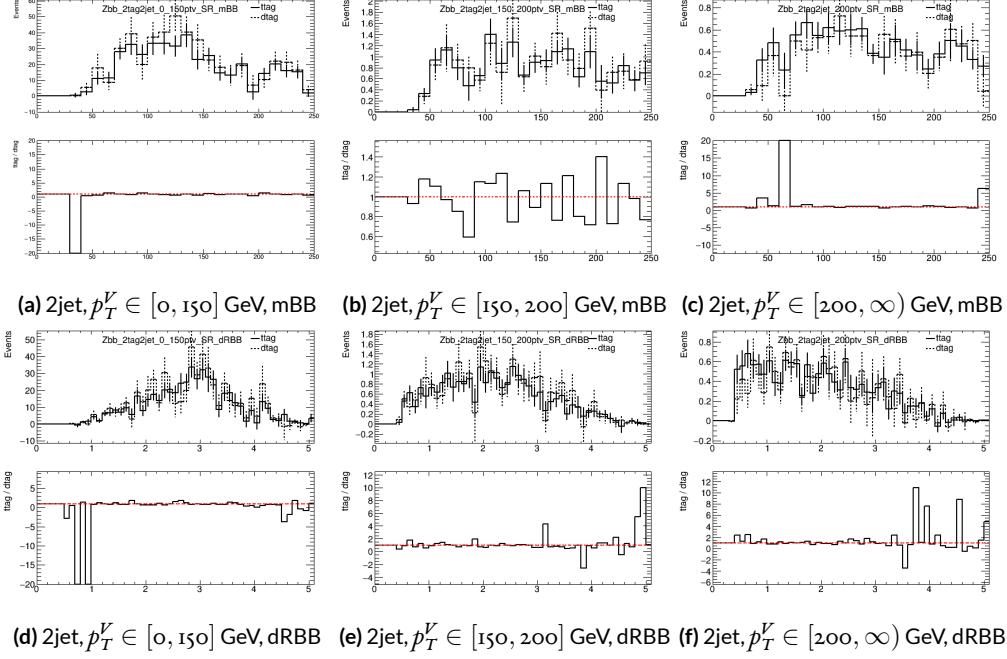
1330



**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.

1331 5.6 MISCELLANIA AND SYSTEMATICS SUMMARY

1332 A summary of all experimental systematics, taken from<sup>56</sup>, may be found below. In addition to the  
1333 systematics discussed above, there are also two further systematics, on the total integrated luminosity  
1334 and on the event reweighting factor used to account for pileup, both included in Table 5.7

1335 5.7 EVENT SELECTION AND ANALYSIS REGIONS

1336 With object and event reconstruction described, it is now time to address which events are actually  
1337 selected for use in analysis. This analysis focuses specifically on the 2-lepton channel of the fiducial  
1338 analysis, with the event selection and analysis region definitions being identical. Common to all  
1339 lepton channels in the fiducial analysis is the set of requirements on the jets in a given event. There  
1340 must be at least two central jets and exactly two signal jets that have been “*b*-tagged” according to the  
1341 MV2c10 algorithm<sup>7</sup>, with at least one of these *b*-jets having  $p_T > 45$  GeV. For MVA training and  
1342 certain background samples, a process known as “truth-tagging” is applied instead of the standard *b*-  
1343 tagging to boost sample statistics and stabilize training/fits (cf.<sup>56</sup> Section 4.2 for details). After event  
1344 selection, the *muon-in-jet* and *PtReco* corrections, described in<sup>24</sup> 6.3.3-4, are applied to the *b*-jets.

1345 In addition to the common selections, there are 2-lepton specific selections. All events are re-  
1346 quired to pass an un-prescaled single lepton trigger, a full list of which may be found in Tables 5  
1347 and 6 of<sup>24</sup> with the requirement that one of the two selected leptons in the event must have fired  
1348 the trigger. There must be 2 VH-loose leptons, and at least one of these must be a ZH-signal lepton  
1349 (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. <sup>24</sup>
Pileup Reweighting	uncertainty on pileup reweighting	Section 11.1 in Ref. <sup>24</sup>
Electrons		
EL_EFF_Trigger_Total_iNPCOR_PLUS_UNCOR	trigger efficiency uncertainty	Section 11.2.2. in Ref. <sup>24</sup>
EL_EFF_Reco_Total_iNPCOR_PLUS_UNCOR	reconstruction efficiency uncertainty	Section 11.3.1. in Ref. <sup>24</sup>
EL_EFF_ID_Total_iNPCOR_PLUS_UNCOR	ID efficiency uncertainty	Section 11.3.1. in Ref. <sup>24</sup>
EL_EFF_Iso_Total_iNPCOR_PLUS_UNCOR	isolation efficiency uncertainty	Section 11.3.1. in Ref. <sup>24</sup>
EG_SCALE_ALL	energy scale uncertainty	Section 11.3.2. in Ref. <sup>24</sup>
EG_RESOLUTION_ALL	energy resolution uncertainty	Section 11.3.2. in Ref. <sup>24</sup>
Muons		
MUON_EFF_TrigStatUncertainty	trigger efficiency uncertainty	Section 11.2.2. in Ref. <sup>24</sup>
MUON_EFF_TrigSystUncertainty	reconstruction and ID efficiency uncertainty for muons with $> 15$ GeV	Section 11.4.1. in Ref. <sup>24</sup>
MUON_EFF_STAT	reconstruction and ID efficiency uncertainty for muons with $< 15$ GeV	Section 11.4.1. in Ref. <sup>24</sup>
MUON_EFF_SYS	isolation efficiency uncertainty	Section 11.4.1. in Ref. <sup>24</sup>
MUON_EFF_STAT_LOWPT	track-to-vertex association efficiency uncertainty	Section 11.4.1. in Ref. <sup>24</sup>
MUON_EFF_SYST_LOWPT	momentum resolution uncertainty from inner detector	Section 11.4.2. in Ref. <sup>24</sup>
MUON_ISO_STAT	momentum resolution uncertainty from muon system	Section 11.4.2. in Ref. <sup>24</sup>
MUON_ISO_SYS	momentum scale uncertainty	Section 11.4.2. in Ref. <sup>24</sup>
MUON_TTVA_STAT	charge dependent momentum scale uncertainty	Section 11.4.2 in Ref. <sup>24</sup>
MUON_TTVA_SYS		
MUON_ID		
MUON_MS		
MUON_SCALE		
MUON_SAGITTA_RHO		
MUON_SAGITTA_RESBIAS		
Jets		
JET_2iNP_JET_EffectiveNP_1	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EffectiveNP_2	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EffectiveNP_3	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EffectiveNP_4	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EffectiveNP_5	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EffectiveNP_6	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EffectiveNP_7	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EffectiveNP_8restTerm	energy scale uncertainty from the in situ analyses splits into 8 components	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EtaIntercalibration_Modeling	energy scale uncertainty on eta-intercalibration (modeling)	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EtaIntercalibration_TotalStat	energy scale uncertainty on eta-intercalibrations (statistics/method)	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_EtaIntercalibration_NonClosure	energy scale uncertainty on eta-intercalibrations (non-closure)	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Pileup_OffsetMu	energy scale uncertainty on pile-up (mu dependent)	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Pileup_OffsetNPV	energy scale uncertainty on pile-up (NPV dependent)	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Pileup_PtTerm	energy scale uncertainty on pile-up (pt term)	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Pileup_RhoTopology	energy scale uncertainty on pile-up (density $\rho$ )	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Flavor_Composition_Zjets	energy scale uncertainty on $Z+jets$ sample's flavour composition	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Flavor_Composition_Wjets	energy scale uncertainty on $W+jets$ sample's flavour composition	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Flavor_Composition_top	energy scale uncertainty on top sample's flavour composition	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Flavor_Composition	energy scale uncertainty on $VV$ and $VH$ sample's flavour composition	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_Flavor_Response	energy scale uncertainty on samples' flavour response	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_BJES_Response	energy scale uncertainty on b-jets	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_PunchThrough_MC15	energy scale uncertainty for punch-through jets	Section 11.5.1. in Ref. <sup>24</sup>
JET_2iNP_JET_SingleParticle_HighPt	energy scale uncertainty from the behaviour of high-pT jets	Section 11.5.1. in Ref. <sup>24</sup>
JET_JER_SINGLE_NP	energy resolution uncertainty	Section 11.5.1. in Ref. <sup>24</sup>
JET_JvtEfficiency	JVT efficiency uncertainty	Section 11.5.1 in Ref. <sup>24</sup>
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. <sup>24</sup>
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. <sup>24</sup>
FT_EFF_Eigen_extrapolation_from_charm	$b$ -tagging efficiency uncertainty on tau jets	Section 11.7. in Ref. <sup>24</sup>
MET		
METTrigStat	trigger efficiency uncertainty	Section 11.2.1. in Ref. <sup>24</sup>
METTrigTop/Z	track-based soft term related longitudinal resolution uncertainty	Section 11.6. in Ref. <sup>24</sup>
MET_SoftTrk_ResoPara	track-based soft term related transverse resolution uncertainty	Section 11.6. in Ref. <sup>24</sup>
MET_SoftTrk_ResoPerp	track-based soft term related longitudinal scale uncertainty	Section 11.6. in Ref. <sup>24</sup>
MET_SoftTrk_Scale	track MET scale uncertainty due to tracks in jets	Section 11.6. in Ref. <sup>24</sup>

**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.<sup>24</sup>.

<sub>1350</sub> 101 GeV. In addition to the jet corrections described above, a kinematic fitter is applied to the leptons  
<sub>1351</sub> and two leading corrected jets in an event with three or fewer jets<sup>\*</sup> to take advantage of the fact that  
<sub>1352</sub> the 2-lepton final state is closed (cf.<sup>132</sup>); these objects are only used for MVA training/fit inputs.

<sub>1353</sub> In order to increase analysis sensitivity, the analysis is split into orthogonal regions based on the  
<sub>1354</sub> number of jets and the transverse momentum of the  $Z$  candidate (the vectoral sum of the lepton  
<sub>1355</sub> pair; this  $p_T$  is denoted  $p_T^V$ ): 2 and  $\geq 3$  jets;  $p_T^V$  in  $[75, 150], [150, \infty)$  GeV. In addition to the signal  
<sub>1356</sub> regions where the leptons are required to be the same flavor ( $e$  or  $\mu$ ), there are top  $e - \mu$  control  
<sub>1357</sub> regions used to constrain the top backgrounds.

<sub>1358</sub> All of these requirements are summarized in ??.

Category	Requirement
Trigger	un-prescaled, single lepton
Jets	$\geq 2$ central jets; 2 $b$ -tagged signal jets, harder jet with $p_T > 45$ GeV
Leptons	2 VH-loose leptons ( $\geq 1$ 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

<sub>1359</sub> It should be noted that the use of  $\geq 3$  jet events is a 2-lepton specific selection. These regions are  
<sub>1360</sub> exclusive 3 jet regions in the 0- and 1-lepton channels, but the fiducial 2-lepton analysis was found to  
<sub>1361</sub> see a  $\sim 4\%$  gain in sensitivity in studies by including  $\geq 4$  jet events<sup>56</sup>.

---

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

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

Conventional Wisdom

# 6

1362

1363

## Multivariate Analysis Configuration

1364 IN ORDER TO fully leverage the descriptive power of the 13 TeV dataset, this analysis makes use of a  
1365 multivariate (MVA) discriminant. Where traditionally event counts or single discriminating vari-  
1366 ables per region of phase space have been fed to fits, MVA discriminants seek to integrate additional  
1367 information not captured in the conventional phase space cuts plus dijet invariant mass distribu-

1368 tions. Formulating the MVA discriminant is an exercise in supervised learning to construct a binary  
1369 classifier, where one uses labeled “signal” and “background” MC events to train optimize the param-  
1370 eters of a statistical model—in this case a boosted decision tree (BDT) with some set of physically  
1371 motivated variables (or “factors”). The interested reader is directed to the standard references on  
1372 machine learning for further details. Sample and variable selection, including the RestFrames and  
1373 Lorentz Invariants variable concepts, are discussed in Section 6.1; MVA training is treated in Sec-  
1374 tion 6.2; and the data statistics only (no systematics) performance of the three MVA discriminants is  
1375 explored in Section .

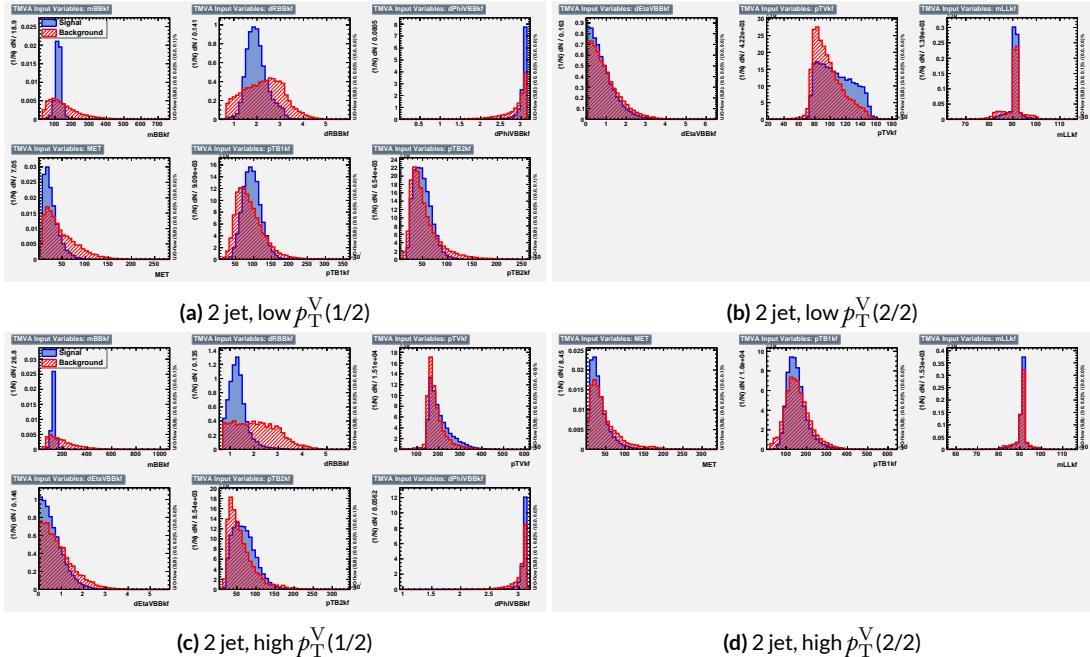
## 1376 6.1 TRAINING SAMPLES AND VARIABLE SELECTION

1377 A subset of samples described in Section 3 was used for multivariate analysis training, with  $qqZH \rightarrow$   
1378  $\ell\ell b\bar{b}$  and  $ggZH \rightarrow \ell\ell b\bar{b}$  used as signal samples and  $Z+jets$ ,  $t\bar{t}$ , and  $VV$  used as background samples.  
1379 Truth-tagging is used on all samples in MVA training to improve training statistics and stability. All  
1380 figures quoted in this section scale distributions to a luminosity of  $36.1 \text{ fb}^{-1}$ .

### 1381 6.1.1 STANDARD VARIABLES

1382 The standard set of variables taken as a baseline is the same as used in the fiducial analysis. The vari-  
1383 ables fall into several main categories: energy/momenta scales of composite objects ( $m_{bb}$ ,  $m_{bbj}$ ,  
1384  $p_T^V$ ,  $m_{\ell\ell}$ ), angles ( $\Delta R(b_1, b_2)$ ,  $\Delta\phi(V, H)$ ,  $\Delta\eta(V, H)$ ), transverse momenta of the jets in the event  
1385 ( $p_T^{b_1}$ ,  $p_T^{b_2}$ ,  $p_T^{j_3}$ ), and  $E_T^{miss}$ . Input distributions for these variables in all the 2 ( $\geq 3$  jet) analysis signal  
1386 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



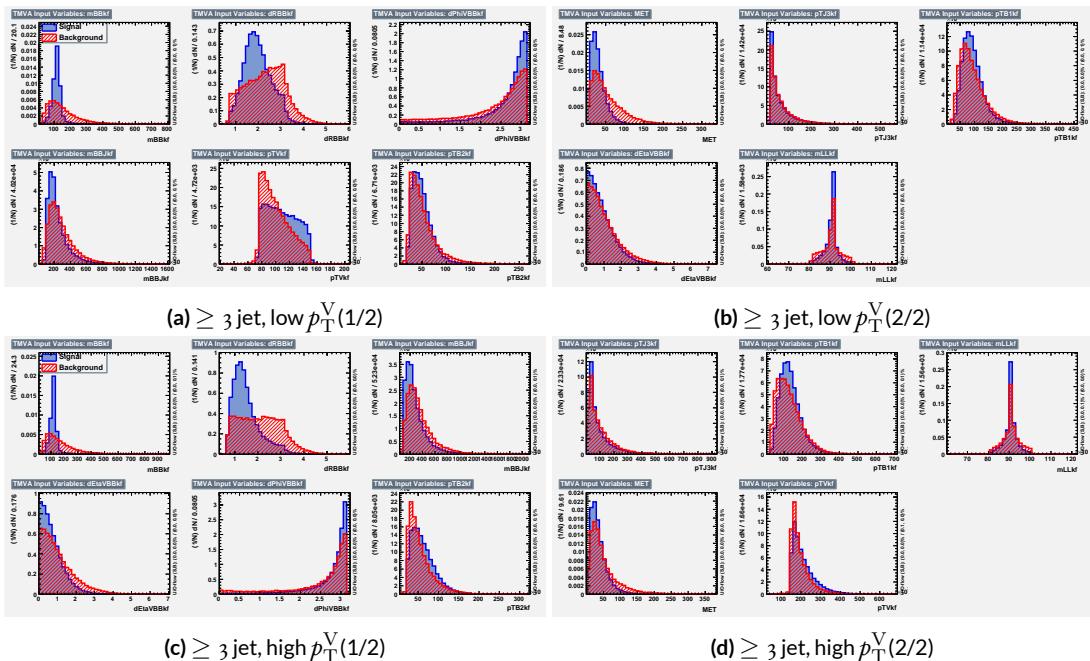
**Figure 6.1:** Input variables in 2 jet signal regions for the standard variable set. Signal distributions are in red, and background distributions are in blue.

1387

1388 are those used as inputs for one of the two k-folded final discriminants, and the order of the distri-  
 1389 butions is the hyperparameter optimized order for feeding into the BDT; what precisely this means  
 1390 will be discussed in following sections. While variables in the analysis regions are generally similar,  
 1391 there are some notable exceptions.  $p_T^V$  and the correlated  $\Delta R(b_1, b_2)$  have different shapes, by con-  
 1392 struction for the former and by correlation for the latter, at low and high  $p_T^V$ . \* The  $\geq 3$  jet regions  
 1393 also have variables that are not applicable to the 2 jet regions; the inclusion of  $m_{BBJ}$  (the invariant  
 1394 mass of the two  $b$ -jets and leading untagged jet) in particular is of note and suggests an avenue for

---

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



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

1395 refinements of the non-standard variables.

1396 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

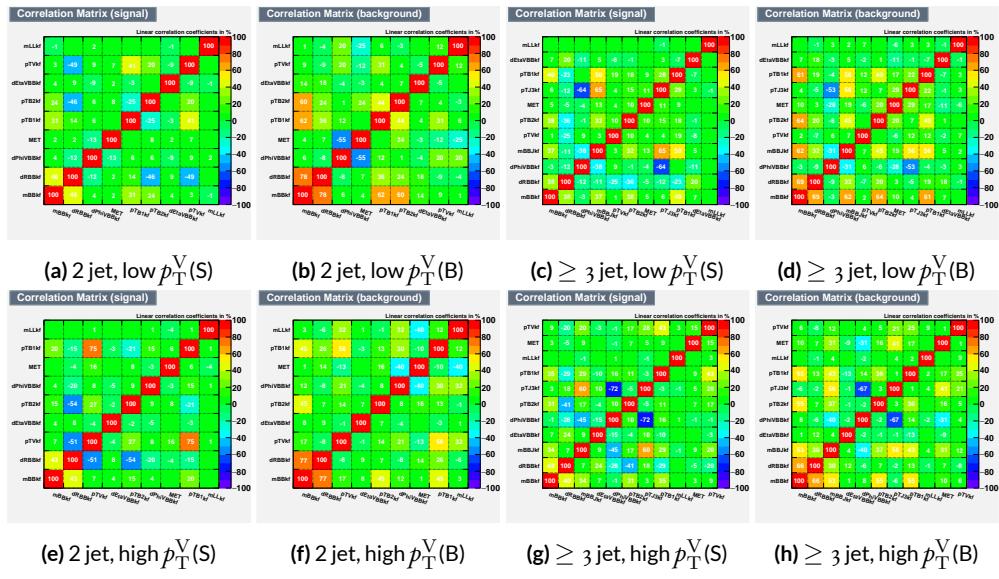
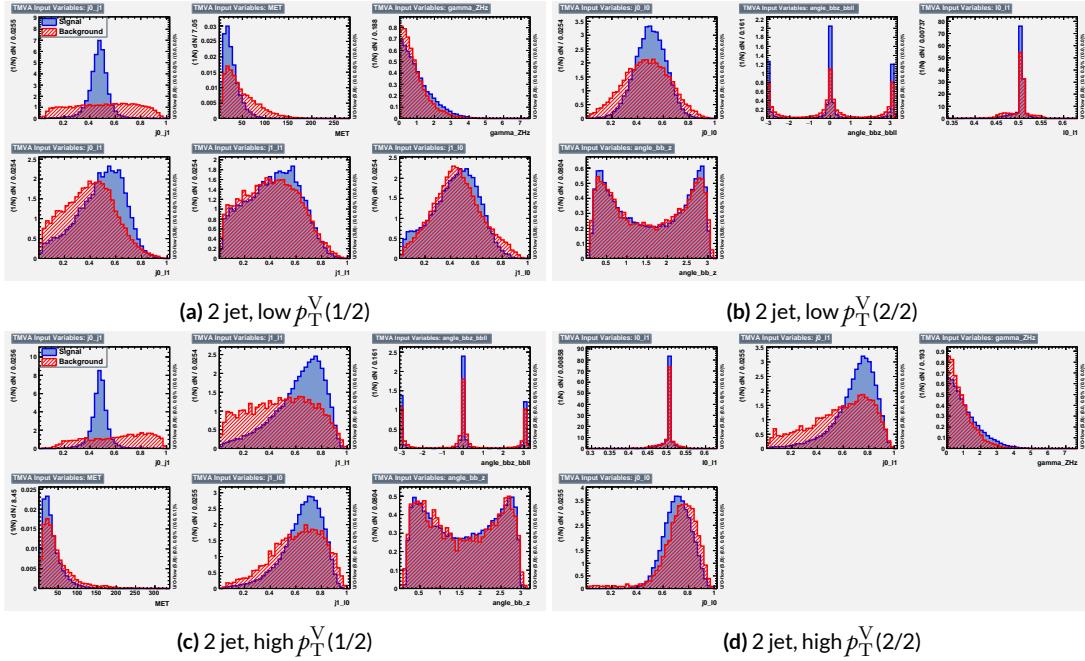


Figure 6.3: Signal and background variable correlations for the standard variable set.

1397

1398 6.1.2 LORENTZ INVARIANTS

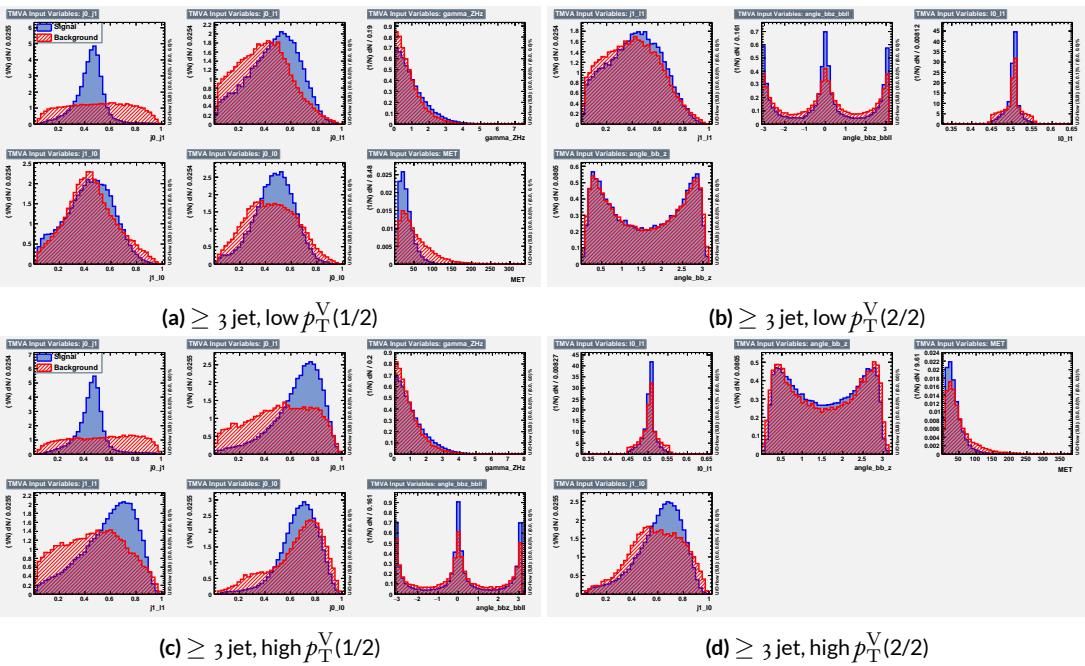
1399 In choosing the set of variables used for a set of Lorentz Invariants based discriminants, we decided  
 1400 to use S. Hagebeck's set from<sup>45</sup> and related studies. Distributions of these variables in the same ar-  
 rangement as with the standard variables may be seen in Figures 6.4 and 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.

1401

1402 about the variable set chosen here is that  $\vec{E}_T^{\text{miss}}$  has been added to the standard LI set. Since the LI  
 1403 construction assumes that this quantity is zero, there is no obvious way to include it. Nevertheless,  
 1404 as the correlation matrices for the LI variables show in Figure 6.6, there is actually very little corre-  
 1405 lation between  $\vec{E}_T^{\text{miss}}$  and the other variables (with this being slightly less the case for the background  
 1406 correlations, as to be expected since  $t\bar{t}$ , a principal background, is  $\vec{E}_T^{\text{miss}}$ -rich). Hence, if including



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

1407  $\vec{E}_T^{\text{miss}}$  violates the spirit somewhat of the LI variables, it does not break terribly much with the aim  
of having a more orthogonal set.

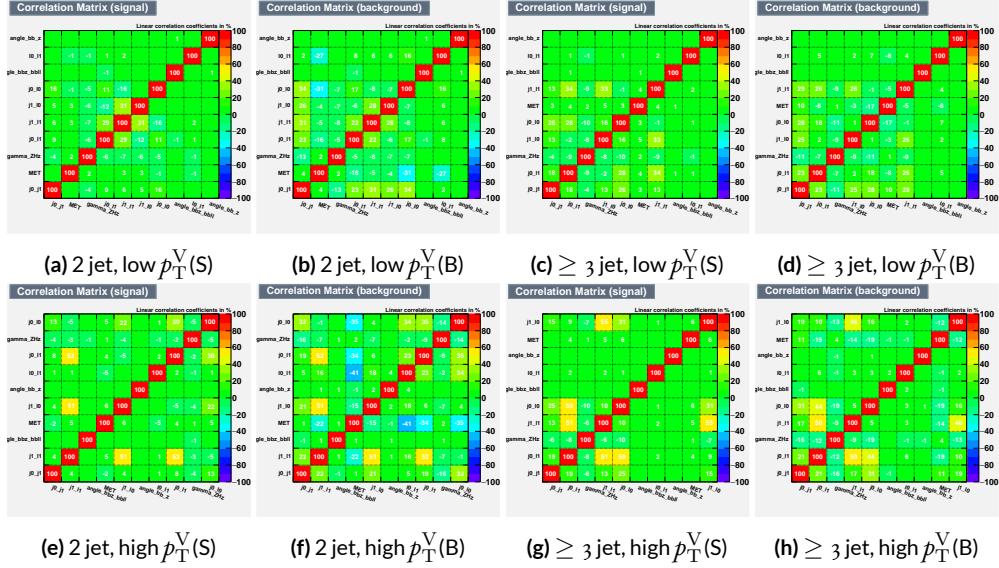


Figure 6.6: Signal and background variable correlations for the LI variable set.

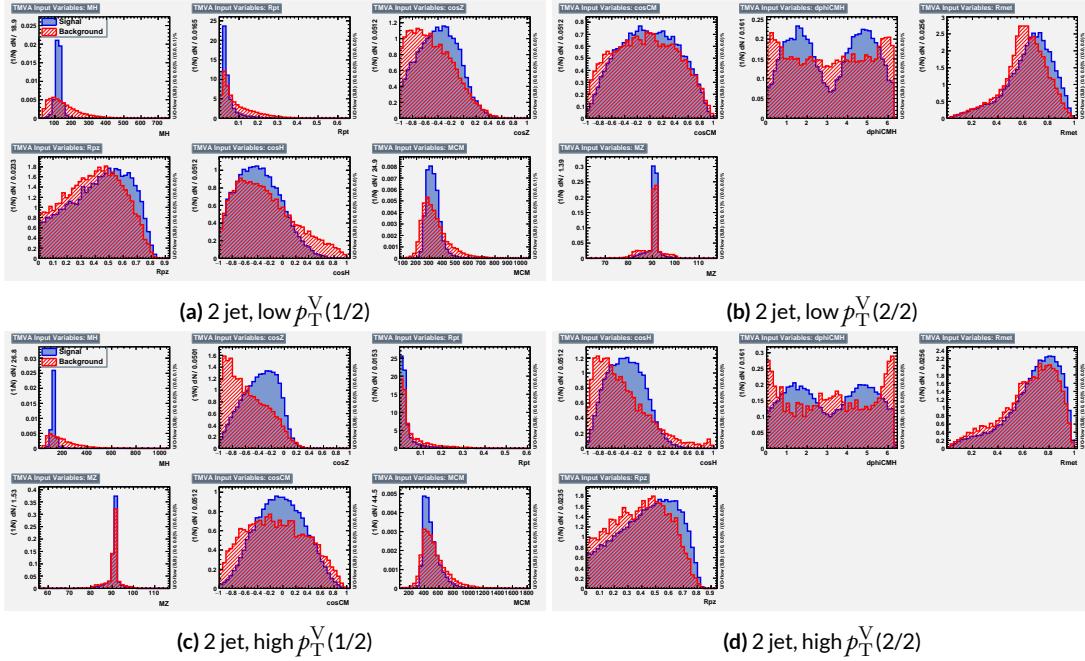
1409 6.1.3 RESTFRAMES VARIABLES

1410 There is no precedent for using the RestFrames variables in the  $ZH \rightarrow \ell\ell b\bar{b}$  analysis, so a subset  
1411 of possible RF variables had to be selected as the basis of a discriminant. The masses and cosines of  
1412 boost angles from parent frames for the CM,  $Z$ , and  $H$  frames gives six variables, and it was decided  
1413 that it would be good to match the LI in terms of variable number and treatment (i.e. no special  
1414 treatment of the third jet), which leaves four more variables. In addition to the cosines, there are also  
1415 the  $\Delta\phi$  angles. Furthermore, there are the event-by-event scaled momentum ratios, both longitudi-  
1416 nal and transverse. There is also both a  $\Delta\phi$  and an event ratio for the  $\vec{E}_T^{\text{miss}}$ . All of these variables  
1417 were included in a ranking using slightly different training settings as the main hyperparameter  
1418 optimization variable rankign described below. The goal of this study was not to develop a discir-  
1419 iminant, as the number of variables is too high, but rather to see which ones are generally useful.  
1420 Table 6.1 shows the results of this study. Percent gains (losses) at each step by adding the variable  
1421 with biggest gain (smallest loss) are shown in green (red). The final row shows an aggregate ranking,  
1422 calculated simply by adding up a variables ranks in all bins and ordering the variables smallest to  
1423 greatest. This simple aggregation does not take into account which regions are potentially more sen-  
1424 sitive and so where taken simply to give an idea of how variables generally performed. With this in  
1425 mind, the RF variables were chosen to be the masses  $M_{CM}$ ,  $M_H$ , and  $M_Z$ , the angles  $\cos_{CM}$ ,  $\cos_H$ ,  $\cos_Z$ ,  
1426  $d\phi_{CMH}$ , and the ratios  $R_{pt}$ ,  $R_{pz}$ , and  $R_{met}$ . Their distributions may be seen in Figures ?? and ??.

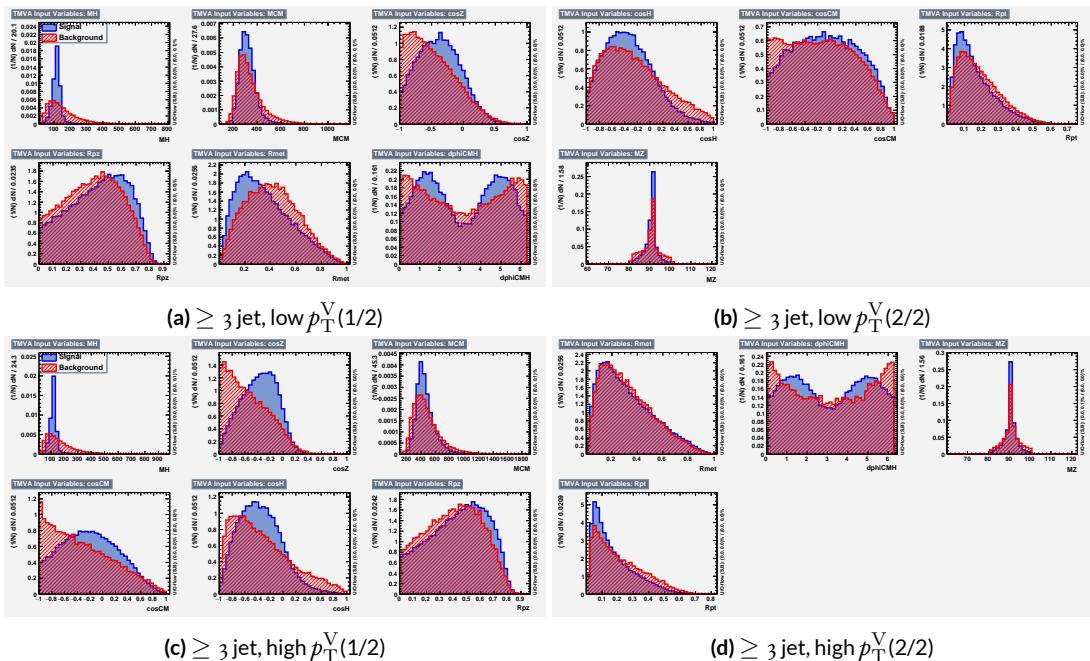
1427 Correlations for the chosen RF variables are shown in Figure 6.9. These correlations are much  
1428 lower than for the standard case but still slightly higher than for the LI case. Given the generally bet-

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

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

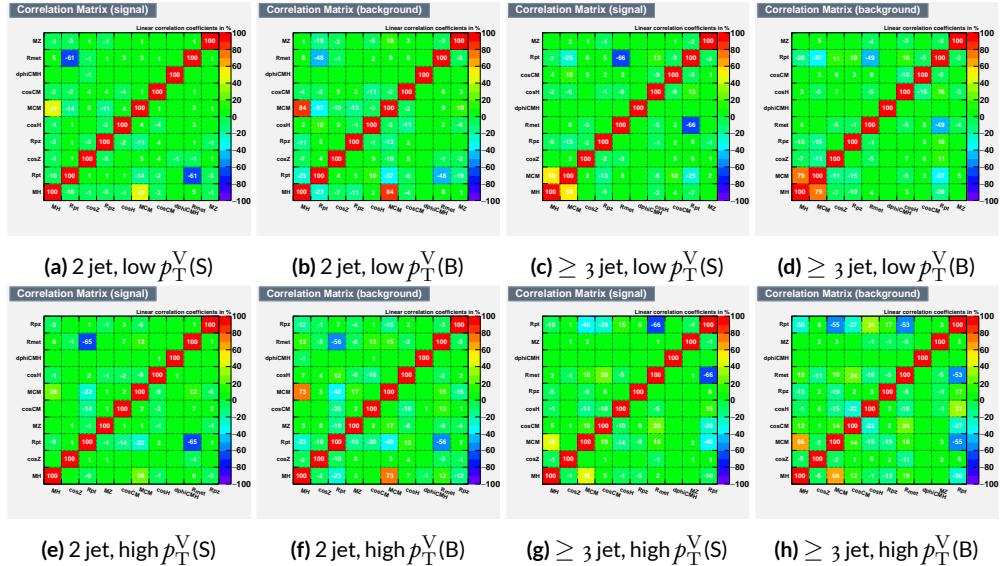


**Figure 6.7:** Input variables in 2 jet signal regions for the RF variable set. Signal distributions are in red, and background distributions are in blue.



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

<sup>1429</sup> ter performance of the RF sets, as we shall see in following sections and chapters, this slight tradeoff  
<sup>1430</sup> is likely an aesthetic one, with the main benefits of a more orthogonal basis likely realized at this level  
<sup>1431</sup> of correlation.



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

1432 A summary of the variables used in the three cases is given in ??.

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

**Table 6.2:** Variables used in MVA training. Variables in parentheses are only used in the  $\geq 3$  jet regions.

1433 6.2 MVA TRAINING

1434 With variables chosen the MVA discriminants must be trained and optimized. MVA training and  
1435 hyperparameter optimization (in this case, just the order in which variables are fed into the MVA) is  
1436 conducted using the “holdout” method. In this scheme, events are divided into three equal portions  
1437 (in this case using `EventNumber%3`), with the first third (the “training” set) being used for the initial  
1438 training, the second third (the “validation” set) being used for hyperparameter optimizaiton, and  
1439 the final third (the “testing” set) used to evaluate the performanace of the final discriminants in each  
1440 analysis region.

1441 The MVA discriminant used is a boosted descision tree (BDT). Training is done in TMVA using  
1442 the training settings of the fiducial analysis<sup>36†</sup>. For the purposes of hyperparameterization and test-  
1443 ing, transformation D with  $z_s = z_b = 10$  is applied to the BDT distributions, and the cumulative  
1444 sum of the significance  $S/\sqrt{S + B}$  in each bin is calculated for each pair of distributions.

1445 Transformation D is a histogram transformation, developed during the Run 1 VHbb search, de-  
1446 signed to reduce the number of bins in final BDT distributions (to reduce the effect of statistical  
1447 fluctuations in data) while also maintaining sensitivity. Such an arbitrary transformation may be  
1448 expressed as:

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

1449 where

---

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

- 1450     •  $I[k, l]$  is an interval of the histograms, containing the bins between bin  $k$  and bin  $l$ ;
- 1451     •  $N_s$  is the total number of signal events in the histogram;
- 1452     •  $N_b$  is the total number of background events in the histogram;
- 1453     •  $n_s(I[k, l])$  is the total number of signal events in the interval  $I[k, l]$ ;
- 1454     •  $n_b(I[k, l])$  is the total number of background events in the interval  $I[k, l]$ ;
- 1455     •  $z_s$  and  $z_b$  are parameters used to tune the algorithm.

1456     Transformation D uses:

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

1457     Rebinning occurs as follow:

- 1458       1. Begin with the highest valued bin in the original pair of distributions. Call this the “last” bin  
1459           and use it as  $l$ , and have  $k$  be this bin as well.
- 1460       2. Calculate  $Z(I[k, l])$
- 1461       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  
1462            $k - 1$  the new “last” bin  $l$ .
- 1463       4. Continue until all bins have been iterated through; if  $Z \leq 1$  for any remaining  $n$  of the  
1464           lowest-valued bins (as is often the case), simply rebin these as a single bin.

1465     Variable ranking is done iteratively in each analysis region. In each set, the validation significance  
1466     of a BDT using an initial subset of variables is calculated (dRBB and mBB for the standard set; j0\_j1  
1467     for the LI set; and MH for the RF set). Each of the remaining unranked variables are then added sep-  
1468     arately, one at a time, to the BDT. The variable yielding the highest validation significance is then  
1469     added to the set list of ranked variables and removed from the list of unranked variables. This pro-  
1470     cess is repeated until no variables remain. These rankings are shown in Figures 6.10–6.12. Rankings  
1471     tend to be fairly stable.

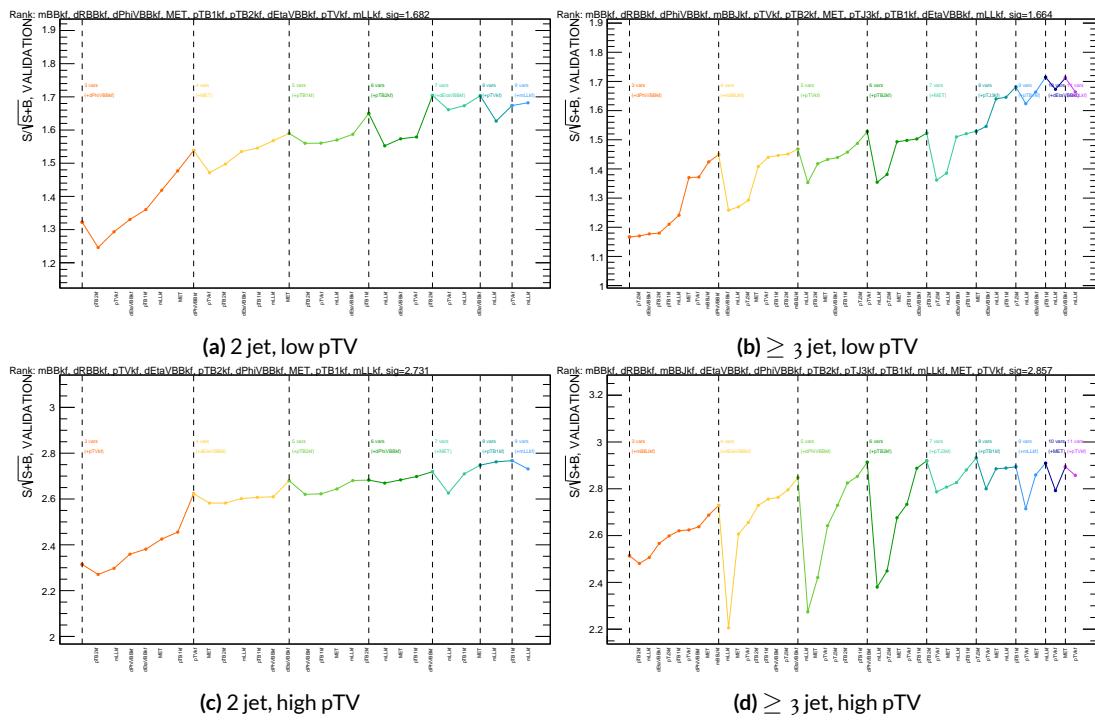


Figure 6.10: Rankings for the standard variable set.

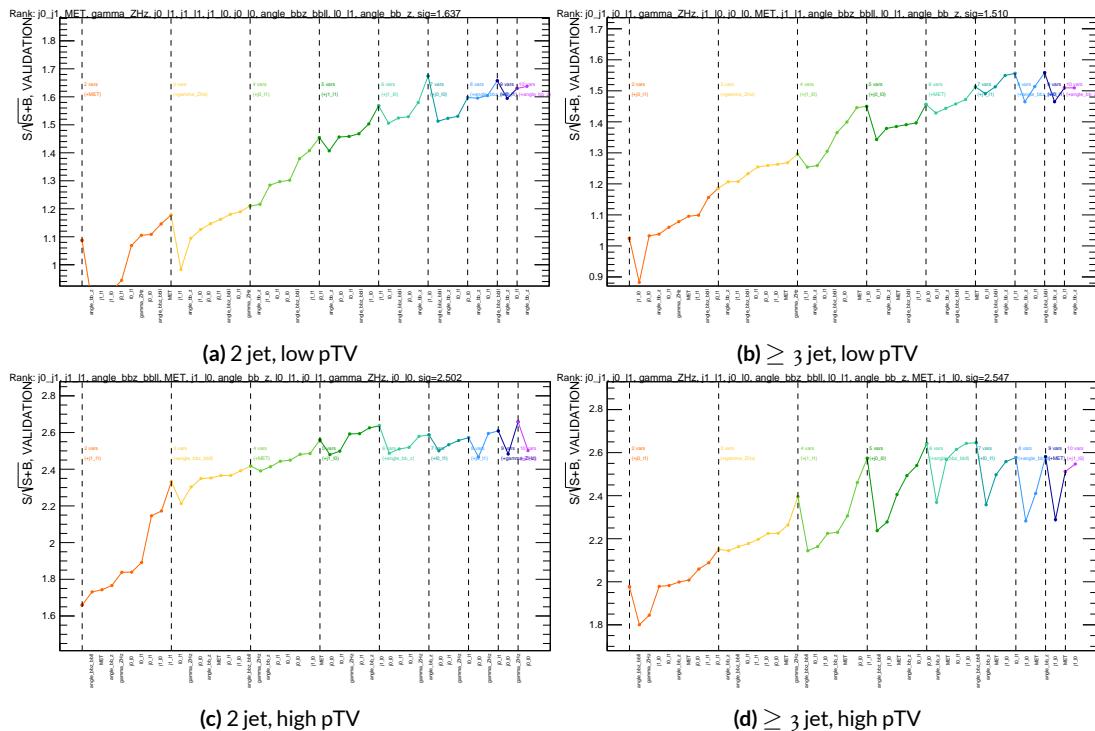


Figure 6.11: Rankings for the LL variable set.

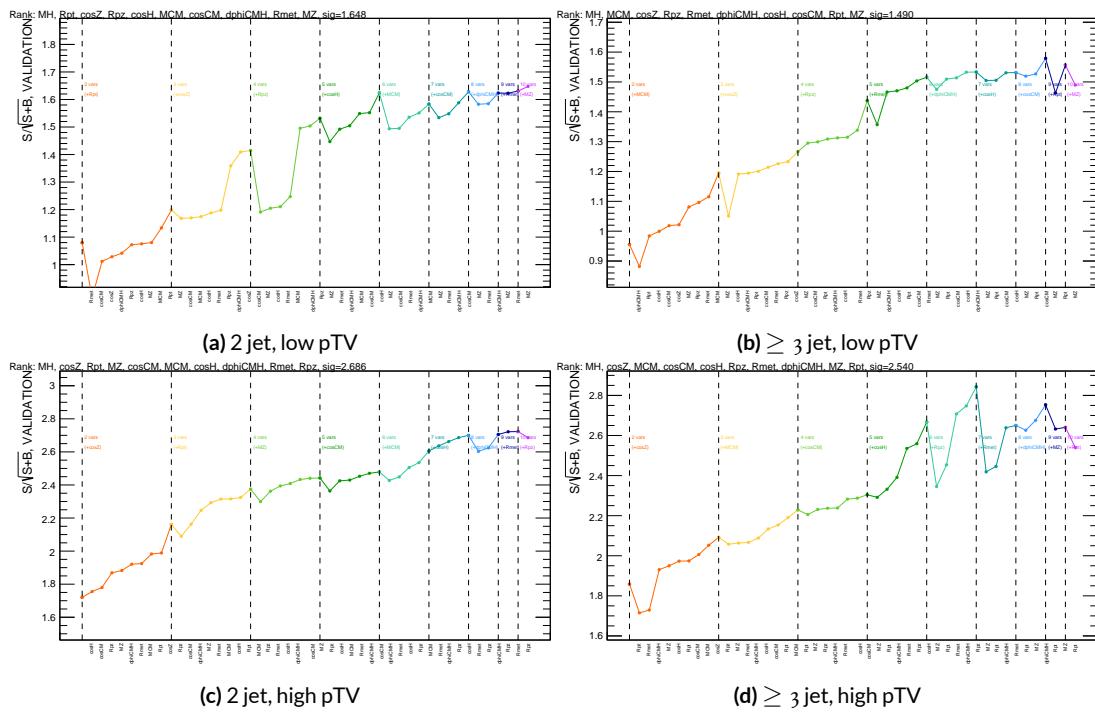


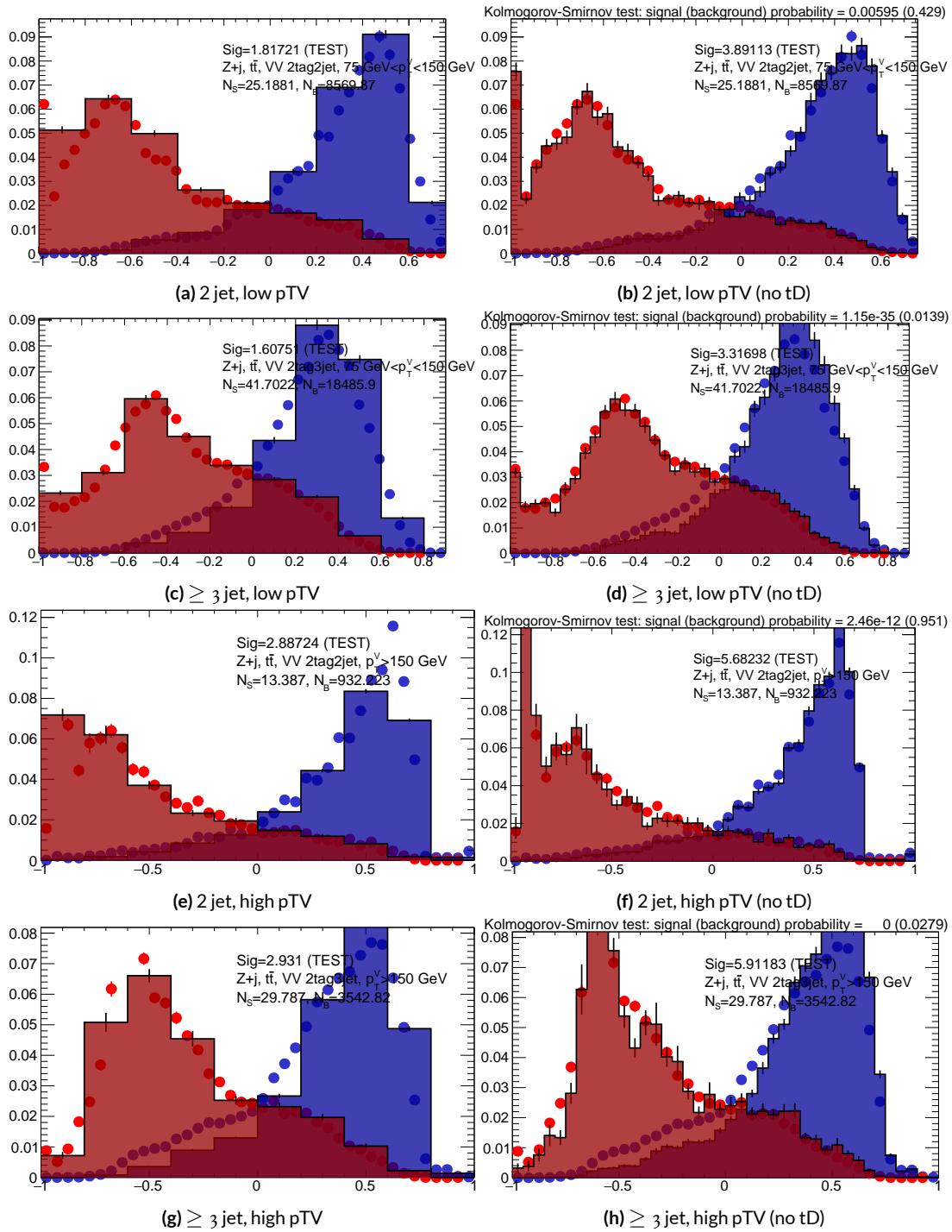
Figure 6.12: Rankings for the RF variable set.

<sup>1472</sup> Once variables have been ranked, the BDT may be used both to evaluate performance in a simpli-  
<sup>1473</sup> fied analysis scenario in the absence of systematic uncertainties (described below in Section 6.3) and  
<sup>1474</sup> to create xml files for the production of fit inputs for an analysis including systematics. Following  
<sup>1475</sup> the approach taken in the fiducial analysis, BDT discriminants using two “k-folds” are produced to  
<sup>1476</sup> prevent overtraining, since the samples used for training are the same as those used to produce in-  
<sup>1477</sup> puts for the full profile likelihood fit. In this scheme, a BDT trained on events with an even (odd)  
<sup>1478</sup> `EventNumber` are used to evaluate events with an odd (even) `EventNumber`.

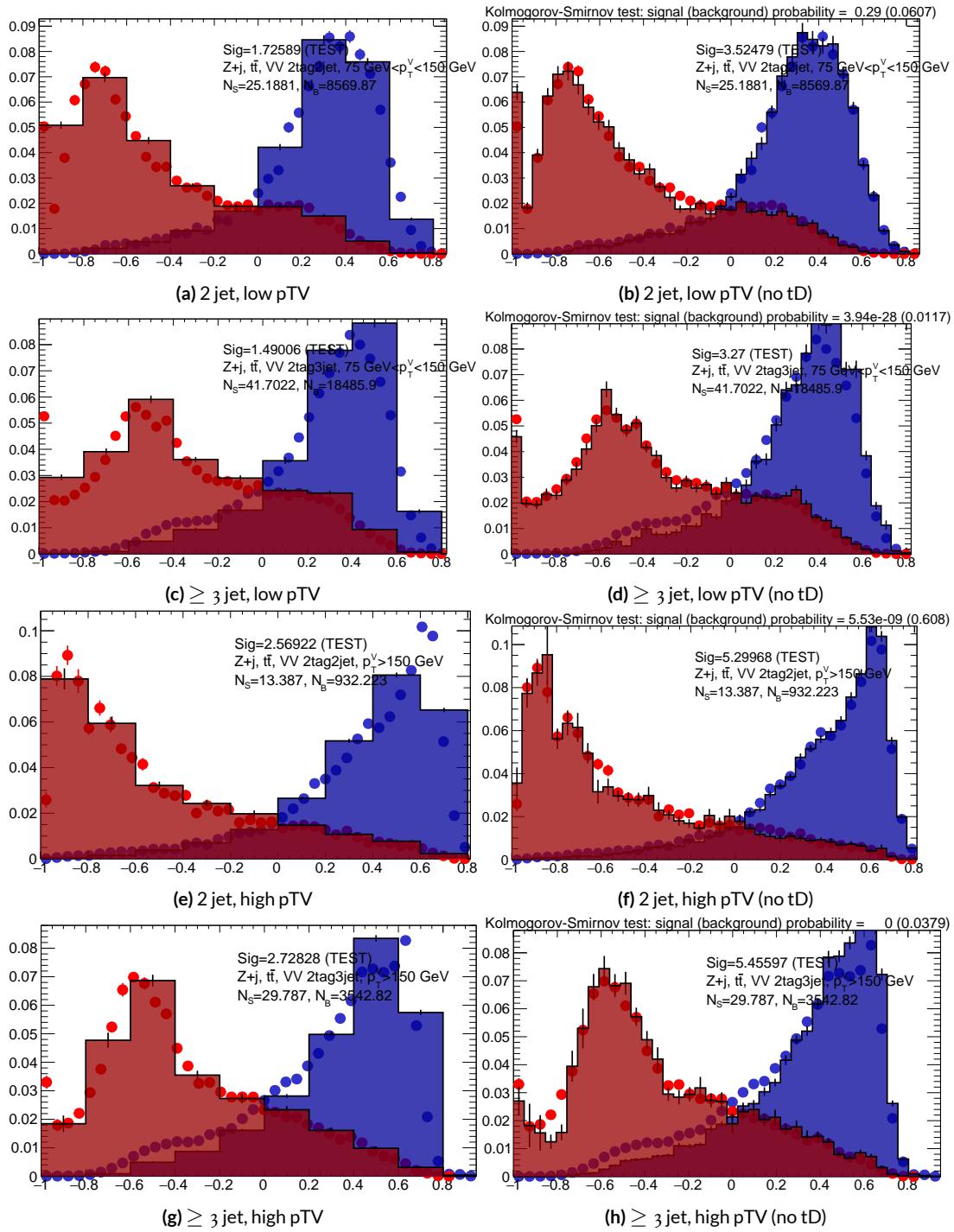
### <sup>1479</sup> 6.3 STATISTICS ONLY BDT PERFORMANCE

<sup>1480</sup> As described above, cumulative significances can be extracted from pairs of signal and background  
<sup>1481</sup> BDT output distributions in a given region. In order to evaluate performance of variable sets in  
<sup>1482</sup> the absence of systematic uncertainties, such pairs can be constructed by evaluating BDT score on  
<sup>1483</sup> the testing set of events using the optimal variable rankings in each region. We show two versions  
<sup>1484</sup> of each testing distribution for each variable set in each signal region in Figures ??–???. The training  
<sup>1485</sup> distribution is always shown as points. The plots with block histograms with numbers of bins that  
<sup>1486</sup> match (do not match) the training distribution do not (do) have transformation D applied. Trans-  
<sup>1487</sup> formation D histograms are included to show the distributions actually used for significance evalu-  
<sup>1488</sup> ation, while the untransformed histograms are included to illustrate that the level of overtraining is  
<sup>1489</sup> not too terrible. For better comparison of the distributions, all histograms have been scaled to have  
<sup>1490</sup> the same normalization.

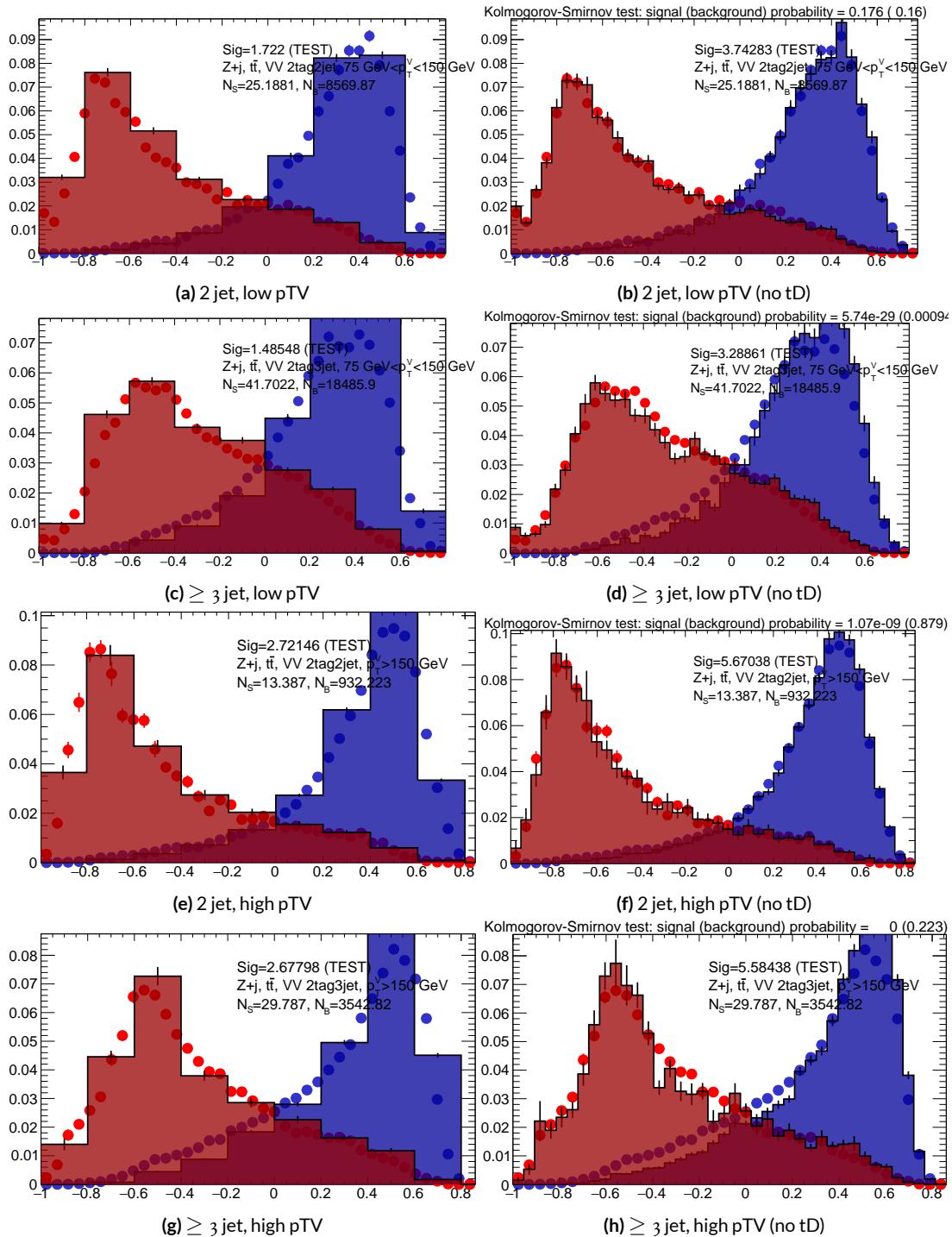
<sup>1491</sup> As can be seen in the summary of cumulative significances for each of these analysis regions and



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

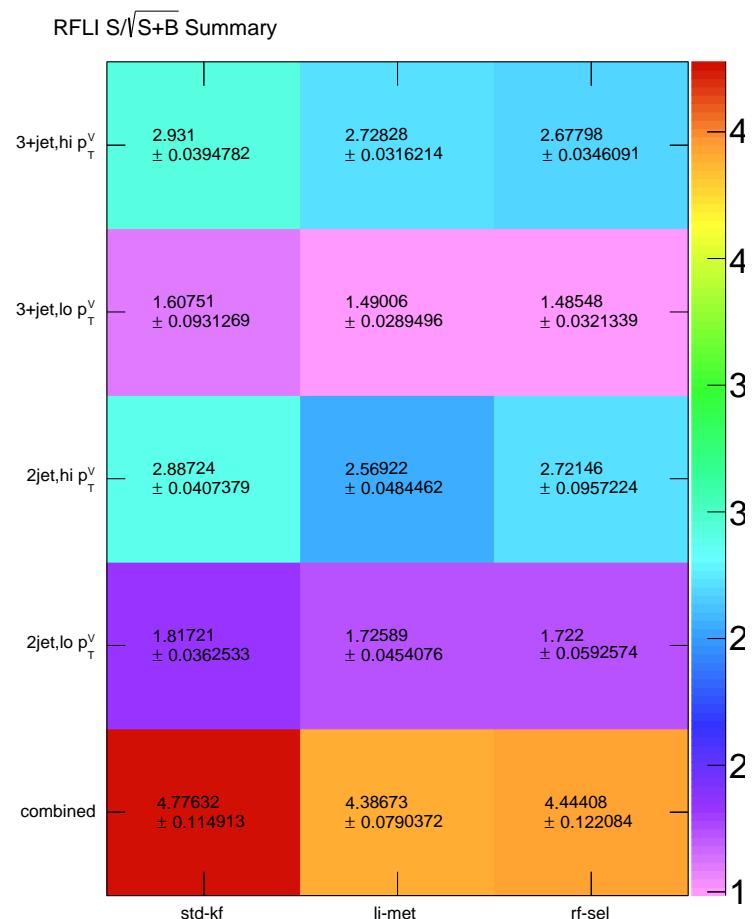


**Figure 6.14:** Training (points) and testing (block histogram) MVA distributions used for stat only testing for the LI variable set.



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

<sup>1492</sup> variable sets in Figure 6.16, the performance of each of the variable sets is quite similar. The standard  
<sup>1493</sup> set performs best, with the LI (RF) set having a cumulative significance that is 7.9% (6.9%) lower.  
<sup>1494</sup> This suggests that the LI and RF variables, in the  $ZH \rightarrow \ell\ell b\bar{b}$  closed final state, have no more in-  
<sup>1495</sup> trinsic descriptive power than the standard set. That these figures are all relatively high ( $\sim 4.5$ ) is  
<sup>1496</sup> due largely to the absence of systematics and possibly in part due to the fact that many of the most  
<sup>1497</sup> significant bins occur at high values of the BDT output, which, as can be seen in any of the testing  
<sup>1498</sup> distributions, contain a small fraction of background events. An interesting feature to note in Fig-  
<sup>1499</sup> ure 6.16 is that while the standard set does perform better in all regions, the gap is larger in the  $\geq 3$   
<sup>1500</sup> jet regions, suggesting that further optimization in the  $\geq 3$  jet case could be useful. Moreover, as  
<sup>1501</sup> discussed at the end of Chapter 5, the choice of  $\geq 3$  jet and not exclusive 3 jet regions is a 2-lepton  
<sup>1502</sup> specific choice and may not be justified for the non-standard variable sets.



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

*Multivac picked you as most representative this year.*

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

Isaac Asimov, “Franchise”

1503

# 7

1504

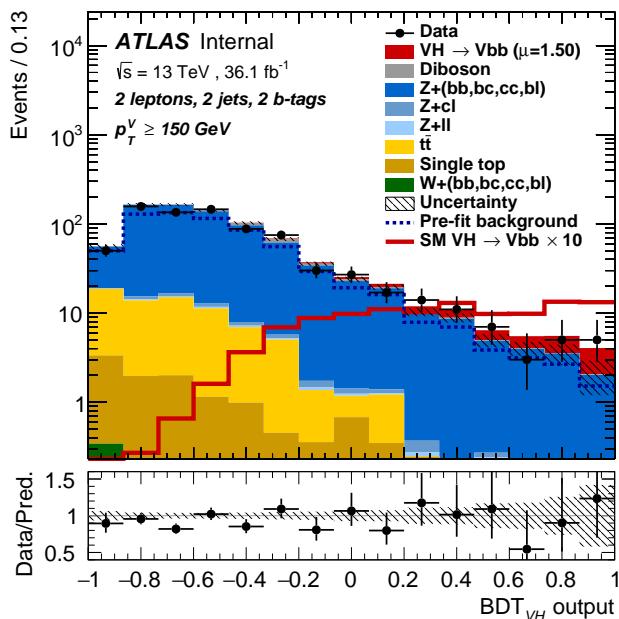
## Statistical Fit Model and Validation

1505 THE ULTIMATE GOAL of an analysis like the search for SM VHbb decay is to say with as much preci-  
1506 sion as possible with the ATLAS collision data whether or not the SM-like Higgs observed in other  
1507 decay modes also decays to  $b$ -quarks and, if so, whether this rate matches the SM prediction. In the  
1508 limit of perfect modeling of both background processes and detector/reconstruction, the only free

1509 parameter is this production rate, referred to typically as a “signal strength,” denoted  $\mu$ , with  $\mu = 1$   
1510 corresponding to the SM prediction.

1511 To get a better sense of what this would look like, take a look at the example discriminant distri-  
1512 bution in Figure 7.1. The black points are data (with statistical error bars), and the colored block  
1513 histograms have size corresponding to the number of predicted events for each process in each bin  
1514 of the final BDT. In the limit of perfect understanding, a fit would correspond to a constant scale  
1515 factor on the red, signal histogram, where one would choose a best fit  $\mu$  value, denoted  $\hat{\mu}$ , that  
1516 would minimize the sum in quadrature of differences between the number of observed data events  
1517 and  $\mu s_i + b_i$ , where  $s_i$  and  $b_i$  are the predicted number of signal and background events in each bin.

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



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

1518

1519 perfect understanding,  $\mu$  could be fitted to arbitrary precision. This, of course, is not the case since  
 1520 there are very many sources of systematic uncertainty, discussed in previous chapters. This chapter  
 1521 will first describe how systematic uncertainties are integrated into the statistical fit of this analysis  
 1522 before describing two sets of cross checks on both a validation  $VZ$  fit and on the fit for the  $VH$  fit of  
 1523 interest.

## 1524 7.1 THE FIT MODEL

1525 In order to derive the strength of the signal process  $ZH \rightarrow \ell\ell b\bar{b}$  and other quantities of interest  
 1526 while taking into account systematic uncertainties or nuisance parameters (NP's, collectively de-  
 1527 noted  $\theta$ ), a binned likelihood function is constructed as the product over bins of Poisson distribu-  
 1528 tions:

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

1529 where  $n$  is the total number of events observed,  $s_i$  and  $b_i$  are the number of expected signal and back-  
 1530 ground events in each bin, and  $S$  and  $B$  are the total expected signal and background events. The  
 1531 signal and background expectations generally are functions of the NP's  $\theta$ . NP's related to the nor-  
 1532 malization of signal and background processes fall into two categories. The first set is left to float  
 1533 freely like  $\mu$  while the second set are parametrized as log-normally distributed to prevent negative  
 1534 predicted values. All other NP's are parametrized with Gaussian priors. This results in a "penalty"  
 1535 on the NLL discussed below of  $(\hat{\alpha} - \mu_\alpha)^2 / \sigma_\alpha^2$ , for NP  $\alpha$ , normally parametrized with mean  $\mu_\alpha$   
 1536 and variance  $\sigma_\alpha^2$  for an MLE of  $\hat{\alpha}$ .

1537 One can maximize<sup>\*</sup> the likelihood in Equation 7.1 for a fixed value of  $\mu$  to derive estimators for  
 1538 the NP's  $\theta$ ; values of  $\theta$  so derived are denoted  $\hat{\theta}_\mu$  to emphasize that these are likelihood maximizing  
 1539 for a given  $\theta$ . The profile likelihood technique finds the likelihood function's maximum by com-  
 1540 paring the values of the likelihood over all possible values of  $\mu$  using these "profiles" and picking  
 1541 the one with the greatest  $\mathcal{L}(\mu, \hat{\theta}_\mu)$  value; these values of  $\mu$  and  $\theta$  are denoted  $\hat{\mu}$  and  $\hat{\theta}$ . The profile  
 1542 likelihood can further be used to construct a test statistic<sup>†</sup>

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

1543 This statistic can be used to derive the usual significance ( $p$  value), by setting  $\mu = 0$  to find the  
 1544 compatibility with the background-only hypothesis<sup>37</sup>. If there is insufficient evidence for the signal  
 1545 hypothesis, the  $CL_s$  method can be used to set limits<sup>11</sup>.

1546 In order to both validate the fit model and study the behavior of fits independent of a given  
 1547 dataset, a so-called "Asimov" dataset can be constructed for a given fit model; this dataset has each  
 1548 bin equal to its expectation value for assumed values of the NP's and a given  $\mu$  value (in this case,  
 1549  $\mu = 0$ , the SM prediction).

---

<sup>\*</sup>Maximization is mathematically identical to finding the minimum of the negative logarithm of the likelihood, which is numerically an easier problem. This is what is done.

<sup>†</sup>The factor of -2 is added so that this statistic gives, in the asymptotic limit of large  $N$ , a  $\chi^2$  distribution.

1550 7.2 FIT INPUTS

1551 Inputs to the binned likelihood are distributions of the BDT outputs described in Chapter 6 for  
1552 the signal regions and of  $m_{bb}$  for the top  $e - \mu$  control regions. These regions split events accord-  
1553 ing to their  $p_T^V$  and number of jets. All events are required to have two  $b$ -tags, as well as pass the  
1554 other event selection requirements summarized in Table ??; the only difference between the signal  
1555 and control region selections is that the same flavor requirement (i.e. leptons both be electrons or  
1556 muons) is flipped so that events in the control region have exactly one electron and one muon. The  
1557 BDT outputs are binned using transformation D, while the  $m_{bb}$  distributions have 50 GeV bins,  
1558 with the exception of the 2 jet, high  $p_T^V$  region, where a single bin is used due to low statistics.

1559 Input distributions in MC are further divided according to their physics process. The signal pro-  
1560 cesses are divided based on both the identity of associated  $V$  and the number of leptons in the final  
1561 state;  $ZH \rightarrow \ell\ell b\bar{b}$  events are further separated into distributions for  $qq$  and  $gg$  initiated processes.  
1562  $V+jets$  events are split according to  $V$  identity and into the jet flavor bins described in Chapter 3.  
1563 Due to the effectiveness of the 2  $b$ -tag requirement suppressing the presence of both  $c$  and  $l$  jets,  
1564 truth-tagging is used to boost MC statistics in the  $cc$ ,  $cl$ , and  $ll$  distributions.<sup>‡</sup> For top backgrounds,  
1565 single top production is split according to production mode ( $s$ ,  $t$ , and  $Wt$ ), with  $t\bar{t}$  as single category.  
1566 Diboson background distributions are also split according to the identity of the  $V$ 's ( $ZZ$ ,  $WZ$ , and  
1567  $WW$ ). Fit input segmentation is summarized in Table 7.1.

---

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

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

**Table 7.1:** Fit input segmentation.

### 1568 7.3 SYSTEMATIC UNCERTAINTIES REVIEW

1569 Tables 7.2 and 7.3 summarize modeling (Chapter 4) and experimental (Chapter 5) systematic uncer-  
 1570 tainties considered in this analysis, respectively. In addition to these, simulation statistics uncertain-  
 1571 ties (“MC stat errors”) are also included in the fit model. There is one distribution per systematic  
 1572 (one each for up and down) per sample per region. The  $\pm 1\sigma$  variation for a systematic is calculated  
 1573 as the difference in the integrals between the nominal and up/down varied distributions.

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

**Table 7.2:** Summary of modeling systematic uncertainties.

1574 The systematics distributions undergo processes known as “smoothing” and “pruning” before  
 1575 being combined into the final likelihood used in minimization.  
 1576 The difference between systematics varied distributions and nominal distributions approaches

Process	Systematics
Jets	21 NP scheme for JES, JER as single NP
$E_T^{\text{miss}}$	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.

1577 some stable value in the limit of large simulation statistics, but if the fluctuations due to simulation  
 1578 statistics in a distribution are large compared to the actual physical effect (whether this is because  
 1579 the actual effect is small or if the actual distribution is derived from a small number of simulation  
 1580 events), then systematic uncertainty will be overestimated by, in effect, counting the MC stat error  
 1581 multiple times. Smoothing is designed to mitigate these effects by merging adjacent bins in some  
 1582 input distributions. Smoothing happens in two steps (the full details of smoothing algorithms may  
 1583 be found in <sup>56</sup> and in the `WSMaker` code):  
 1584 Merge bins iteratively where bin differences are smallest in input distributions until no local extrema  
 1585 remain (obviously, a single peak or valley is allowed to remain)  
 1586 Sequentially merge bins (highest to lowest, like transformation D) until the statistical uncertainty in  
 1587 a given bin is smaller than 5% of merged bin content  
 1588 Not all systematic uncertainties defined are included in the final fit. Pruning Systematics are  
 1589 subject “pruning” (individually in each region/sample: there are two histograms per systematic

<sub>1590</sub> (up/down) per region per sample, so pruning just consists of removing the histograms from the  
<sub>1591</sub> set of distributions included in the likelihood) if they are do not have a significant impact, defined as  
<sub>1592</sub> follows.

- <sub>1593</sub> • Normalization/acceptance systematics are pruned away if either:

<sub>1594</sub>     – The variation is less than 0.5%

<sub>1595</sub>     – Both up and down variations have the same sign

- <sub>1596</sub> • Shape systematics pruned away if either:

<sub>1597</sub>     – Not one single bin has a deviation over 0.5% after the overall normalisation is removed

<sub>1598</sub>     – If only the up or the down variation is non-zero

- <sub>1599</sub> • Shape+Normalisation systematics are pruned away if the associated sample is less than 2% of  
<sub>1600</sub> the total background and either:

<sub>1601</sub>     – If the predicted signal is < 2% of the total background in all bins and the shape and  
<sub>1602</sub> normalisation error are each < 0.5% of the total background

<sub>1603</sub>     – If instead at least one bin has a signal contribution > 2% of the total background, and  
<sub>1604</sub> only in each of these bins, the shape and normalisation error are each < 2% of the  
<sub>1605</sub> signal yield

1606 7.4 THE VZ VALIDATION FIT

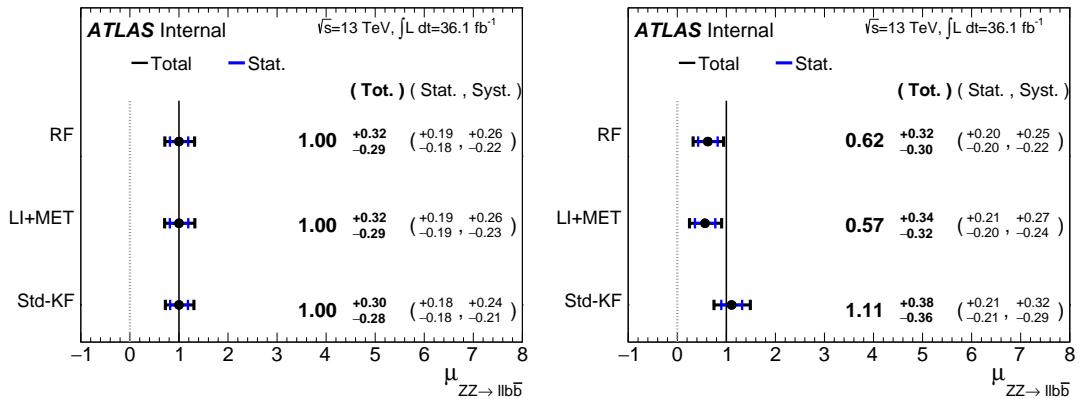
1607 One of the primary validation cross-checks for the fiducial analysis was a  $VZ$  fit—that is, conducting  
1608 the entire analysis but looking for  $Z \rightarrow b\bar{b}$  decays instead of the Higgs. The idea here is that the  $Z$  is  
1609 very well understood and so “rediscovering” the  $Z$  is taken as a benchmark of analysis reliability since  
1610 the complexity of the fit model precludes the use of orthogonal control regions for validation as is  
1611 done in other analyses (generally, if there is a good control region, one prefers to use it to constrain  
1612 backgrounds and improve the fit model). To do this, a new MVA discriminant is made by keeping  
1613 all hyperparameter configurations the same (e.g. variable ranking) but using diboson samples as  
1614 signal. For the 2-lepton case, this means using  $ZZ \rightarrow \ell\ell b\bar{b}$  as the signal sample. This new MVA is  
1615 used to make the inputs described in Section 7.2, and the fit is then run as for the  $VH$  fit (again, with  
1616  $ZZ$  as signal).

1617 The  $VZ$  fit sensitivities for the standard, LI, and RF fits are summarized in Table 7.4. The ex-  
1618 pected significances are all fairly comparable and about what was the case in the fiducial analysis.  
1619 The observed significance for the standard set matches fairly well with the expected value on data,  
1620 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.

1621 These values, however, are consistent with the observed signal strength values, which can be seen  
 1622 in Figure 7.2 (b), with both the LI and RF fits showing a deficit of signal events with respect to the  
 1623 SM expectation, though not by much more than one standard deviation (a possible explanation is  
 1624 explored in the following section). Just as in the  $VH$  fits, errors arising systematic uncertainties are  
 1625 lower in the fits to the observed dataset. That the effect is not noticeable in Asimov fits is not too  
 1626 surprising, since this analysis (and these variable configurations in particular), is not optimized for  
 1627  $VZ$ .



**Figure 7.2:**  $\mu$  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).

#### 1628 7.4.I 2 AND $\geq 3$ JET FITS

1629 While the treatment of simply ignoring any additional jets in the event seems adequate for the  $VH$   
 1630 analysis (discussed below), the potential shortcoming of this treatment appears in the  $VZ$  analysis  
 1631 when the  $2$  and  $\geq 3$  jet cases are fit separately<sup>§</sup>, as can be seen in Figure 7.3. Compared to the stan-  
 1632 dard fit, the LI and RF fits have lower  $\hat{\mu}_{\geq 3 \text{ jet}}$  values, consistent with the interpretation that the addi-

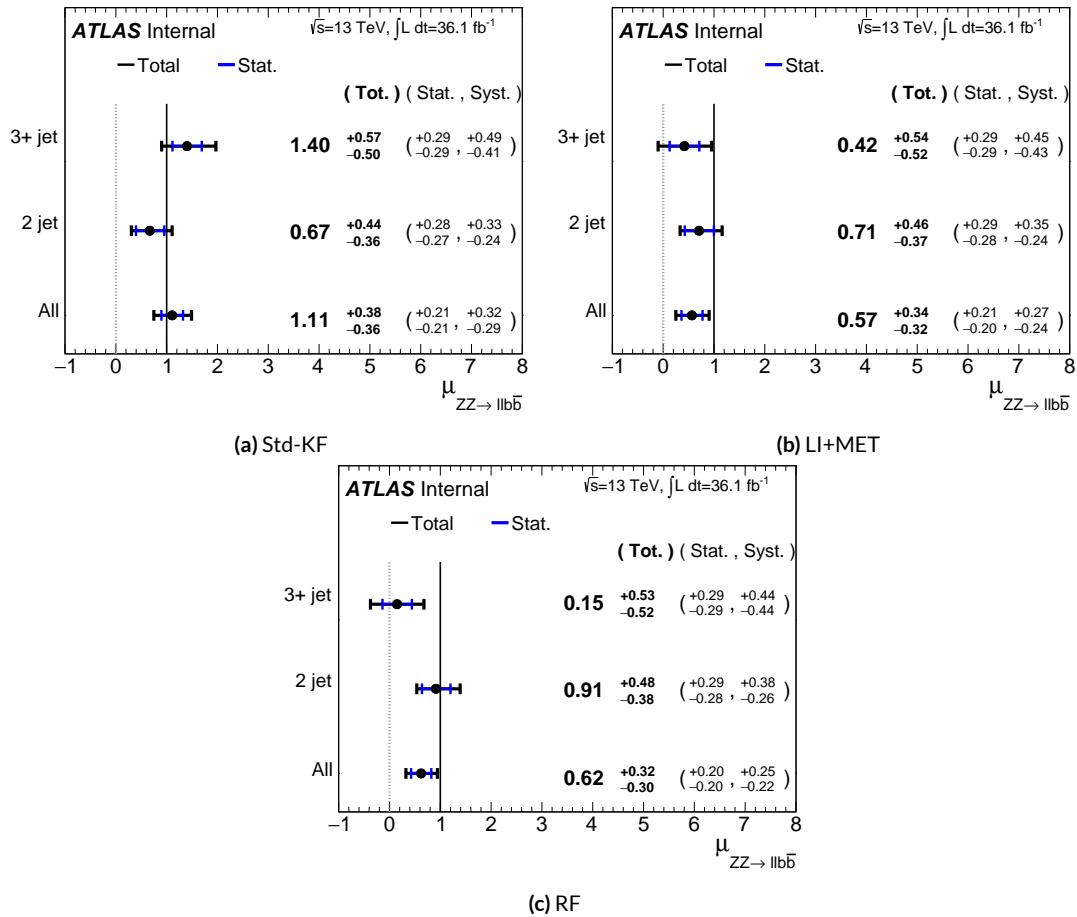
<sup>§</sup>standalone fits, with half the regions each, not 2 POI fits

1633 tional information in the  $\geq 3$  jet regions for the standard case is important for characterizing events  
1634 in these regions for  $VZ$  fits.

1635 A natural question to ask is why this would be an issue for the  $VZ$  but not the  $VH$  case. One  
1636 potential answer is that at high transverse boosts, there is a greater probability for final state  
1637 radiation in the hadronically decaying  $Z$ , so there are more events where the third jet should be in-  
1638 cluded in the calculation of variables like  $m_{b\bar{b}}$  or for angles involving the  $b\bar{b}$  system (e.g.  $\cosh$  in the  
1639 RF case). While the absolute scale at which the low and high  $p_T^V$  regions are separated remains the  
1640 same does not change from the  $VH$  to the  $VZ$  analysis, 150 GeV, the implicit cutoff on the transverse  
1641 boost of the hadronically decaying boson does. For the Higgs, with a mass of 125 GeV, the  $p_T^V$  cutoff  
1642 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  
1643 23–37% higher.

1644 If either the LI or RF schemes were to be used in a mainstream analysis, these validation fits sug-  
1645 gest that the third jet ought to be included in variable schemes (e.g. by adding the third jet to the  
1646 Higgs in the high  $p_T^V$  case). On the issue of whether or not  $\geq 4$  jet events should be included, the  
1647 RF set shows very little sensitivity to this change (a 2 jet and 3 jet only fit moves  $\hat{\mu}$  to 0.64, while  
1648 doing so for the LI set moves it to 0.40, so this, like the addition of the third jet into the variable  
1649 sets, would have to be addressed individually. Nevertheless, this optimization is beyond the scope of  
1650 this thesis, which aim to preserve as much of the fiducial analysis as possible for as straightforward a  
1651 comparison as possible.

1652 For completeness, we include the full set of fit validation results for the  $VZ$  fit, explaining them in  
1653 turn.



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

1654 7.5 NUISANCE PARAMETER PULLS

1655 The first set of plots statistical fit experts will want to look at are the “pulls” and “pull comparisons.”

1656 In these plots, the best fit (nominal) values and one standard deviation error bars are shown for ob-  
1657 served (Asimov) pull plots, with the green and yellow bands corresponding to  $\pm 1, 2\sigma$ , respectively.

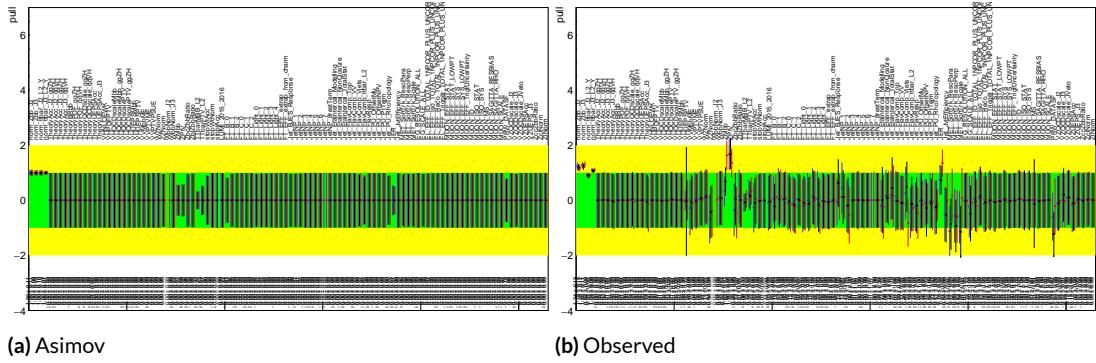
1658 These plots are divided by NP category for readability. ¶ In pull comparisons, these pulls are over-  
1659 layed and color-coded. Pull comparisons here have the following color code: black is the standard  
1660 variable set, red is the LI set, and blue is the RF set.

1661 A well-behaved fit has pulls close to nominal values (“closeness” should be interpreted in the  
1662 context of pull value divided by pull error). As can be seen in Figures 7.4–7.8, the fits for the three  
1663 different variable sets are fairly similar from a NP pull perspective, though the  $Z+jets$   $m_{bb}$  and  $p_T^V$   
1664 NP’s and the jet energy resolution NP are heavily pulled (a handful of poorly behaved pulls is not  
1665 uncommon, though typically warrants further investigation). As a general note, these pull plots  
1666 calculate pulls using a simultaneous HESSE matrix inversion, which is fine for relatively small fits,  
1667 but the more reliable MINOS result, which calculates the impact of each NP on its own, should be  
1668 cross-checked for significant pulls. The ranking plots below do this.

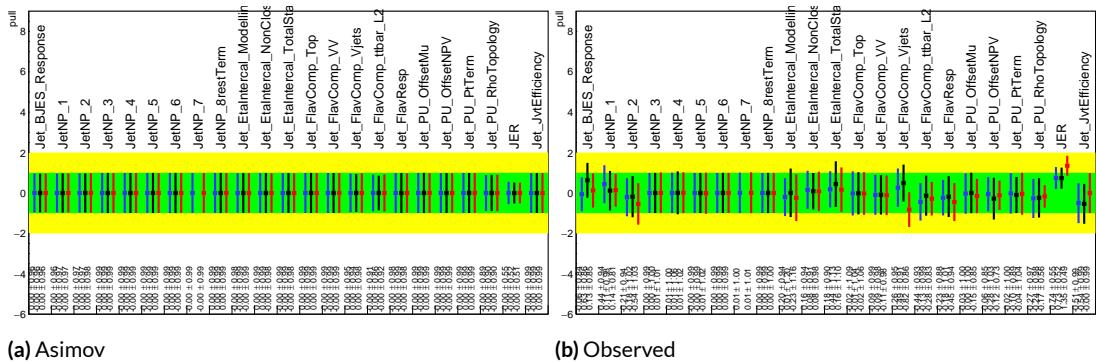
1669 Nuisance parameter correlation matrices (for correlations with magnitude at least 0.25) for all  
1670 three variable set fits can be found in Figures 7.10–7.12. These are useful for seeing which NP’s move  
1671 together (if there is no physical argument for them to do so, this is a potential indicator that further  
1672 investigation is warranted).

---

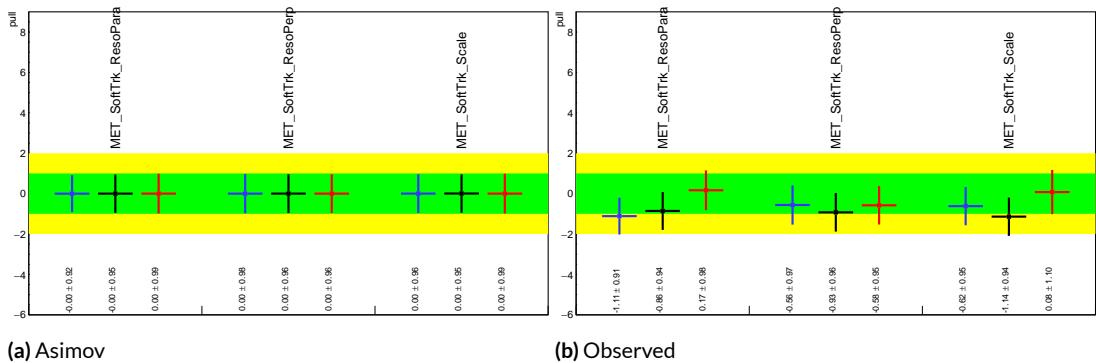
¶ Over 100 non-MC stat NP’s survive pruning in these 2-lepton only Run 2 fits; well over 500 survive in the Run 1+Run 2 combined fit.



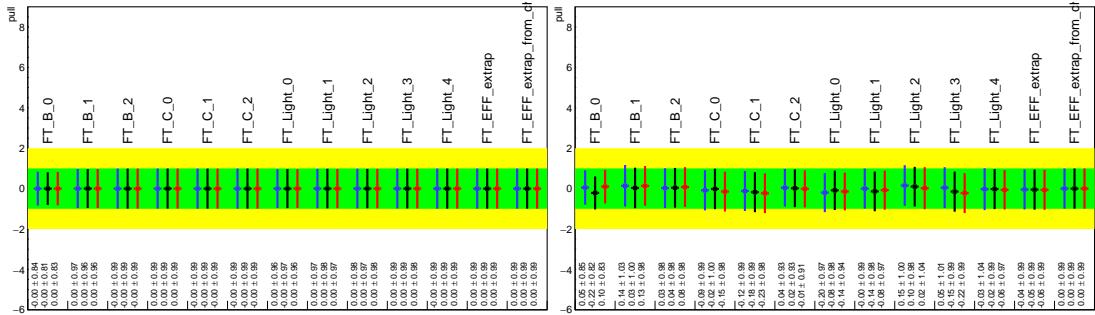
**Figure 7.4:** Pull comparison for all NP's but MC stats.



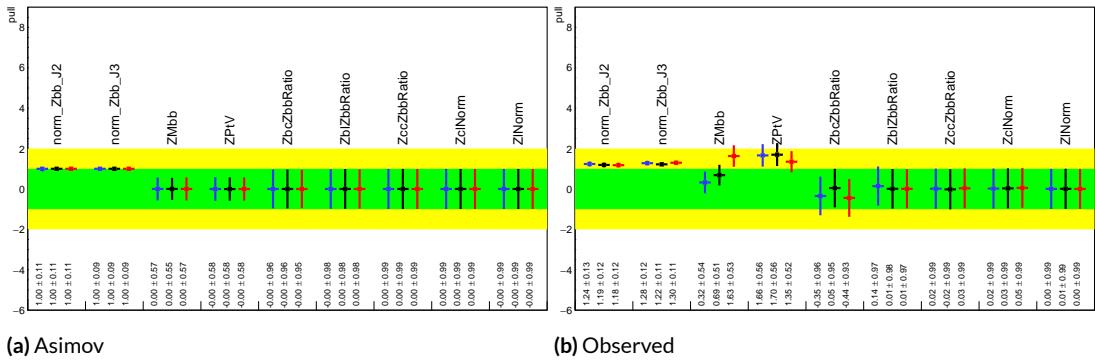
**Figure 7.5:** Pull comparison for jet NP's.



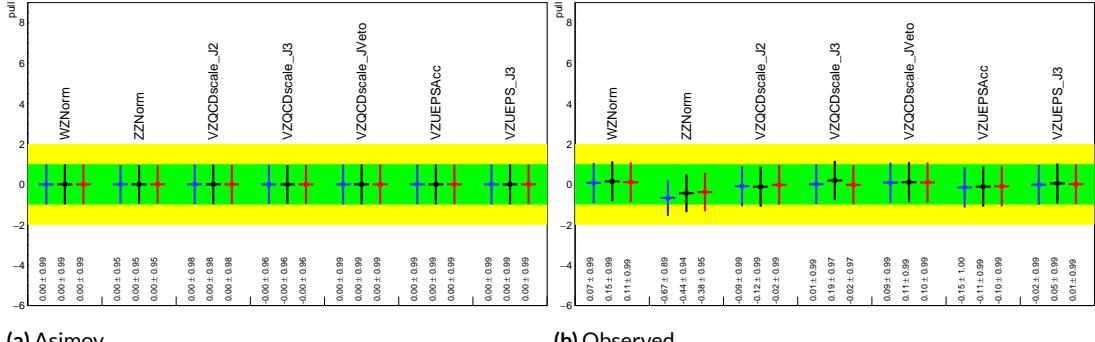
**Figure 7.6:** Pull comparison for MET NP's.



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



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



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

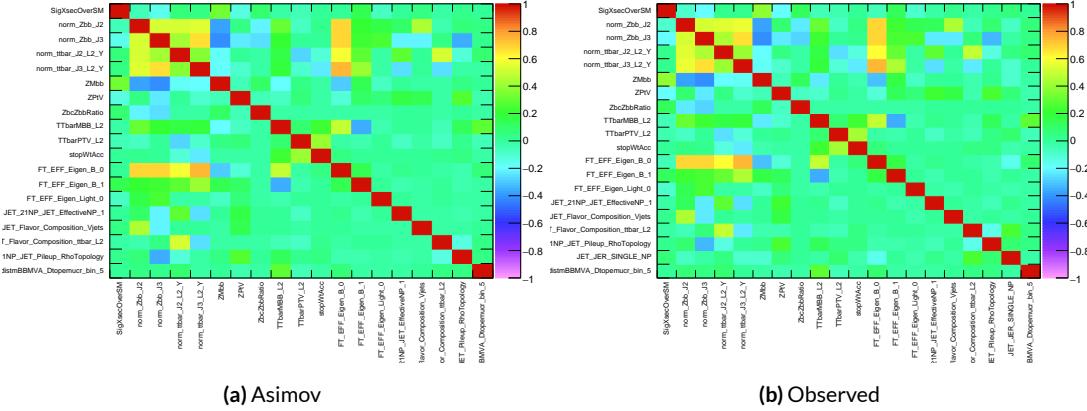


Figure 7.10: NP correlations for standard variable fits.

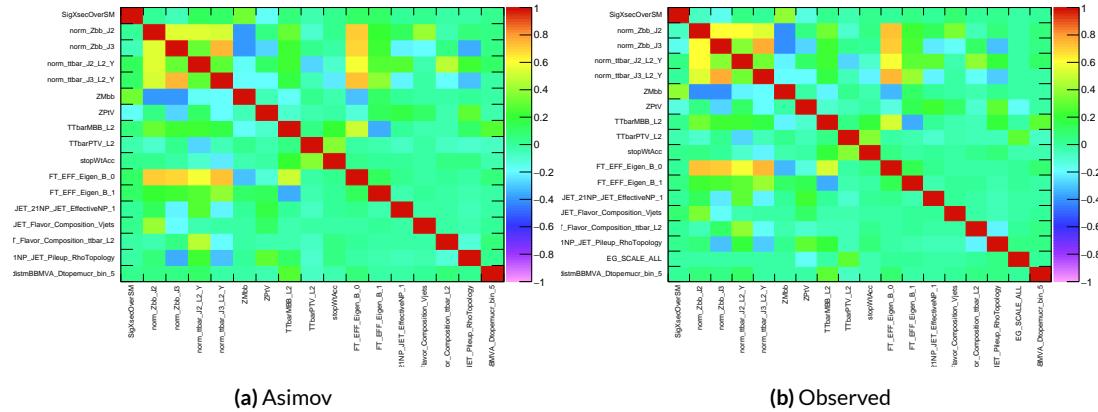


Figure 7.11: NP correlations for L1 variable fits.

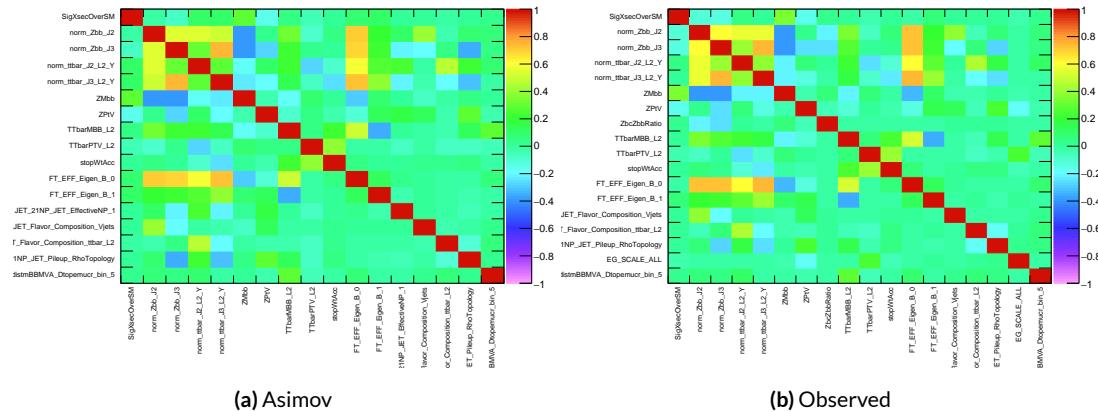


Figure 7.12: NP correlations for RF variable fits.

## 1673 7.5.1 NUISANCE PARAMETER RANKING PLOTS AND BREAKDOWNS

1674 The next set of fit results that is used to diagnose the quality of a fit is the impact of different nui-  
 1675 sance parameters on the total error on  $\mu$ , both individually and as categories. Figure 7.13 shows the  
 1676 top 25 nuisance parameters ranked by their postfit impact on  $\hat{\mu}$ ; these plots use the aforementioned  
 1677 more reliable MINOS approach. This set of rankings is fairly similar, with  $Z+jets$  systematics being  
 1678 particularly prominent. The advantage of seeing individual nuisance parameter rankings, as op-  
 1679 posed to impacts of categories in aggregate, is that particularly pathological NP's are easier to see; in  
 1680 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  $\mu$ .

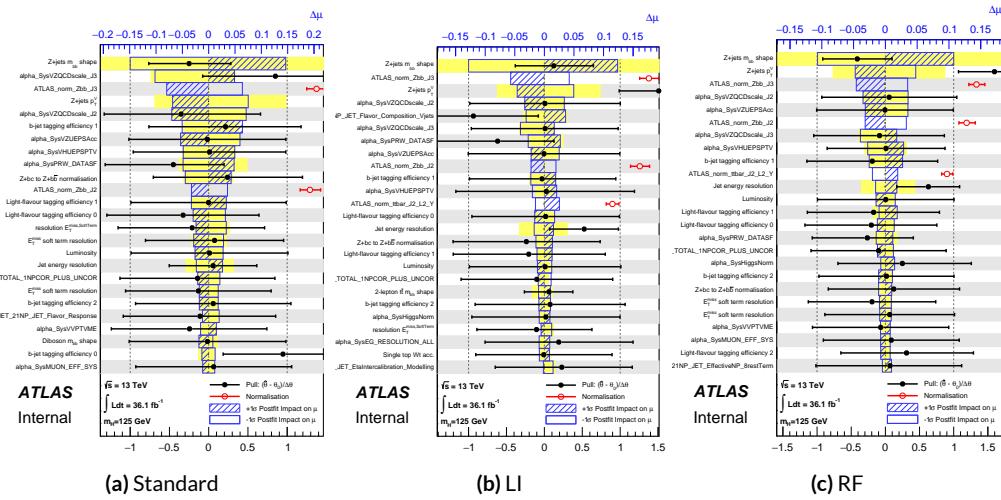


Figure 7.13: Plots for the top 25 nuisance parameters according to their postfit impact on  $\hat{\mu}$  for the standard (a), LI (b), and RF (c) variable sets.

1681

1682 This is consistent with the picture of NP's taken in aggregate categories in Tables 7.5 and 7.6,  
 1683 known as “breakdowns,” with  $Z+jets$  in particular featuring prominently. Of particular interest is

<sup>1684</sup> 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  $\mu$  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.

1685 7.6 POSTFIT DISTRIBUTIONS

1686 Finally, postfit distributions for the MVA discriminant ( $m_{bb}$ ) distribution in the signal (top  $e - \mu$   
1687 control) region for the standard, Lorentz Invariant, and RestFrames variable sets are shown. It is  
1688 generally considered good practice to check the actual postfit distributions of discriminating quan-  
1689 tities used to make sure there is good agreement.<sup>¶</sup> It should be noted that agreement is not always  
1690 great when “eyeballing” a distribution, as fits are messy and  $V+hf$  modeling is notoriously hard to  
1691 get correct. This is particularly true in the  $VZ$  fit since normalizations for  $Z+hf$  in particular are de-  
1692 rived using  $VH$  optimized sidebands. This is also why a lot of these plots are presented as log plots  
1693 (which hide disagreement better; the general argument goes that one has the ratio plots on the bot-  
1694 tom).

1695 7.7 VH FIT MODEL VALIDATION

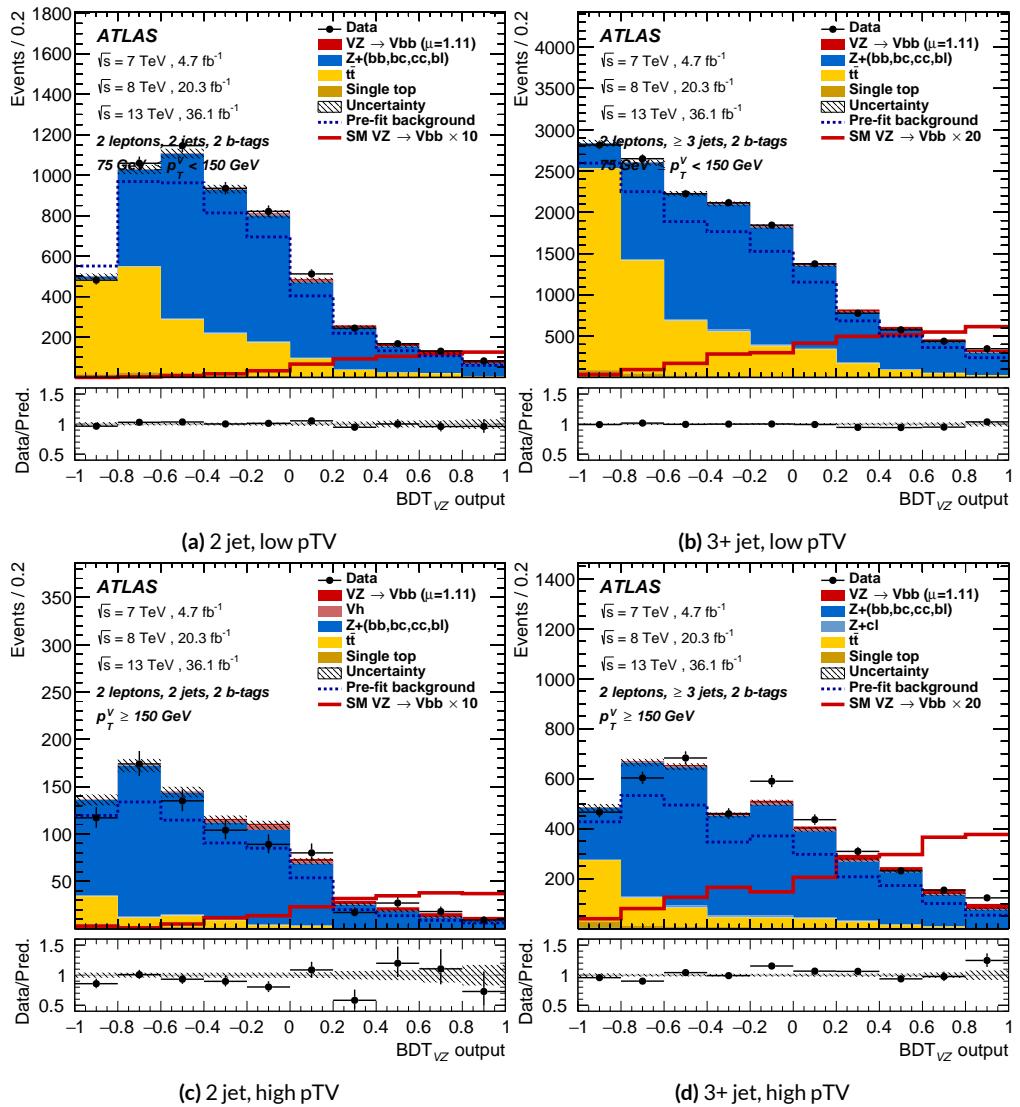
1696 We now move onto the fit validation distributions and numbers for the  $VH$  fit of interest.

1697 7.7.1 NUISANCE PARAMETER PULLS

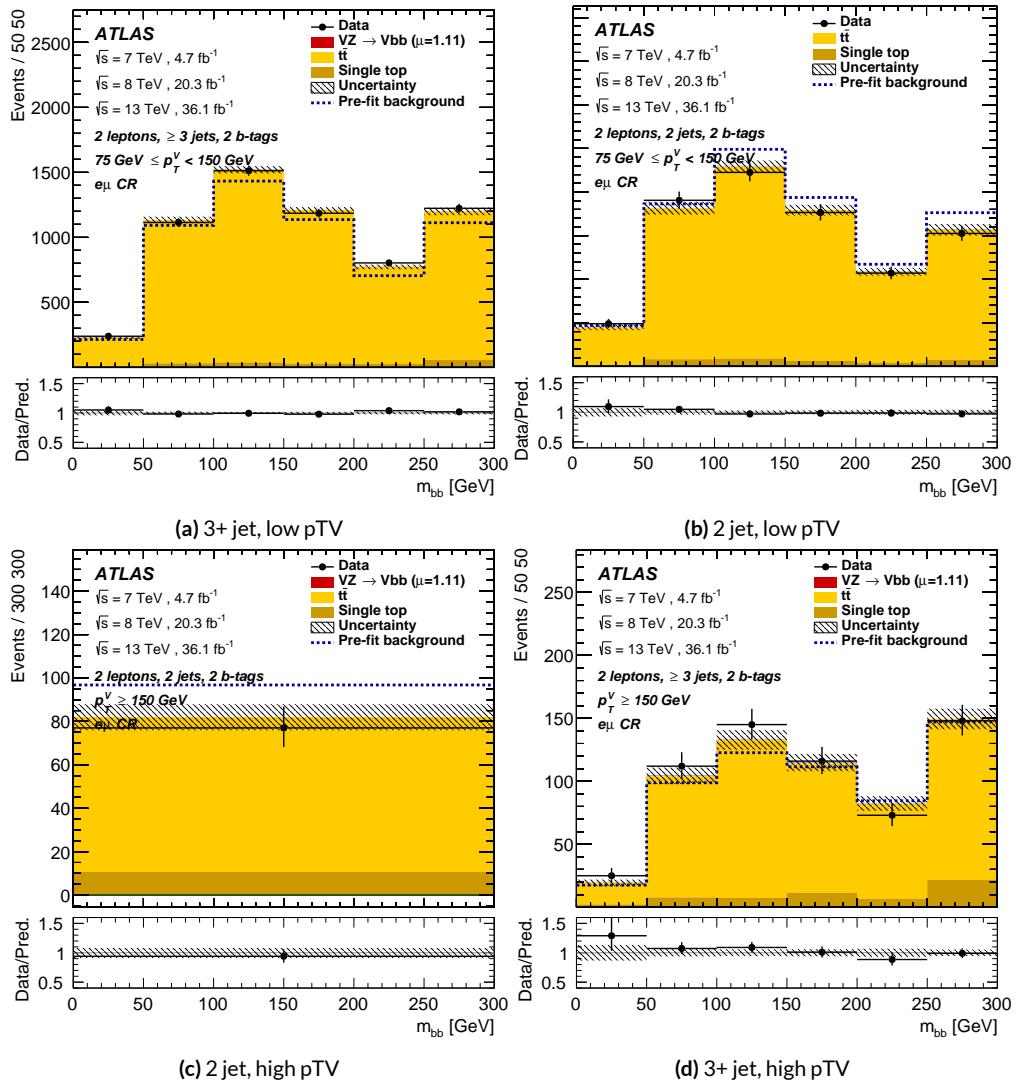
1698 As can be seen in Figures 7.20–7.24, the fits for the three different variable sets are fairly similar from  
1699 a NP pull perspective. Again, black is the standard variable set, red is the LI set, and blue is the RF  
1700 set. The possible exception is the signal UE+PS  $p_T^V$  systematic, which looks very different for all  
1701 three cases (underconstrained for the standard, but overconstrained for the novel variable cases),

---

1701 <sup>¶</sup>Sometimes distributions of input variables (MC histograms scaled by their postfit normalizations) are also used.



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



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

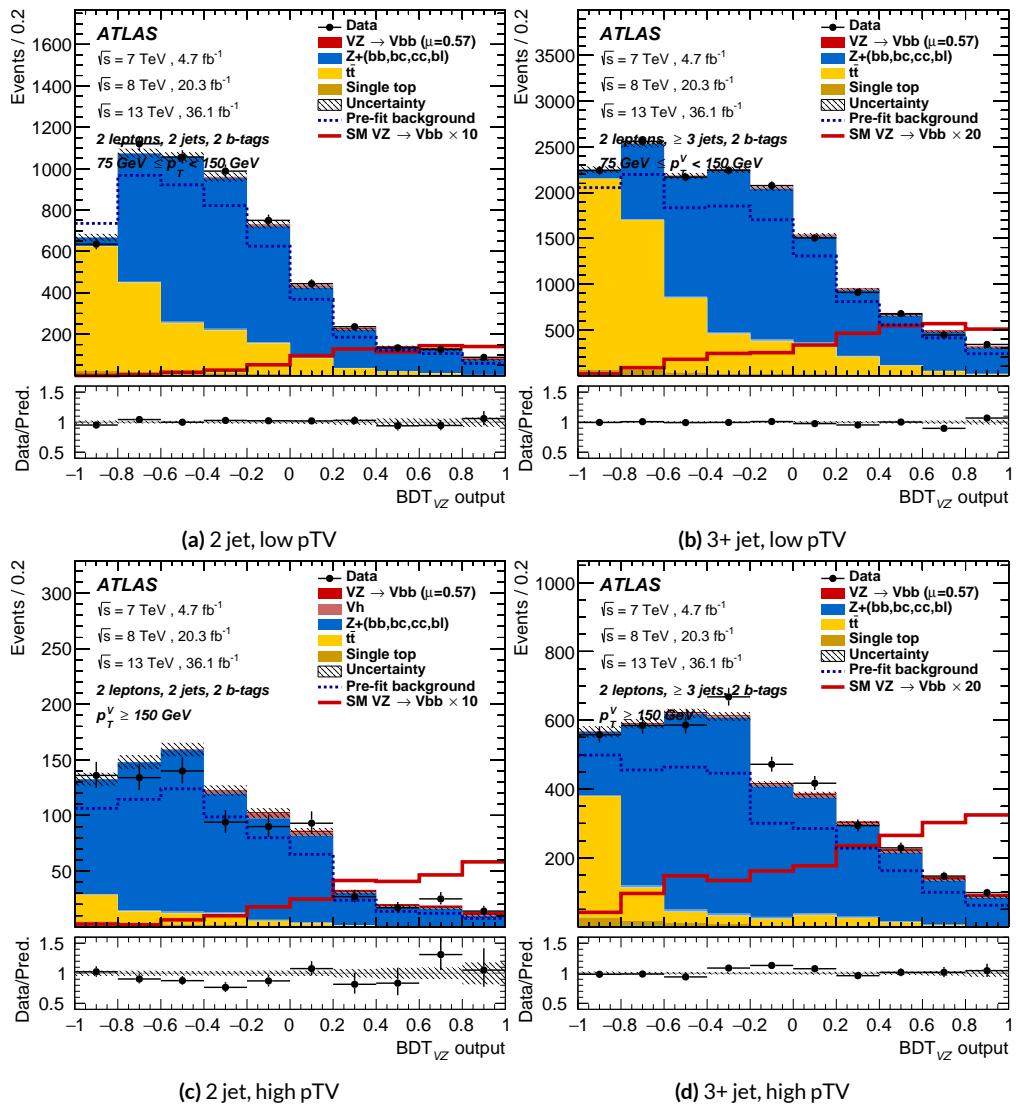


Figure 7.16: Postfit  $BDT_{VZ}$  plots in the signal region for the LI variable set.

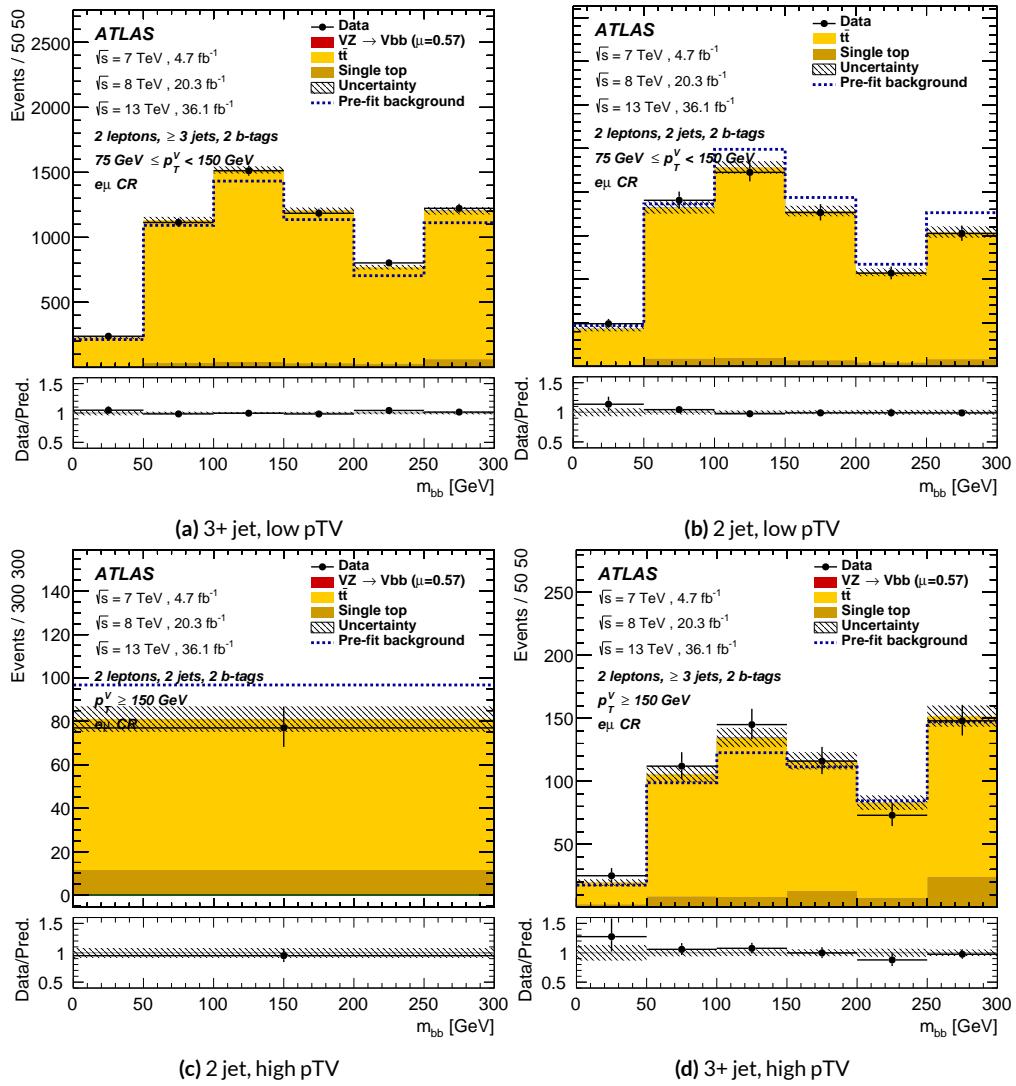
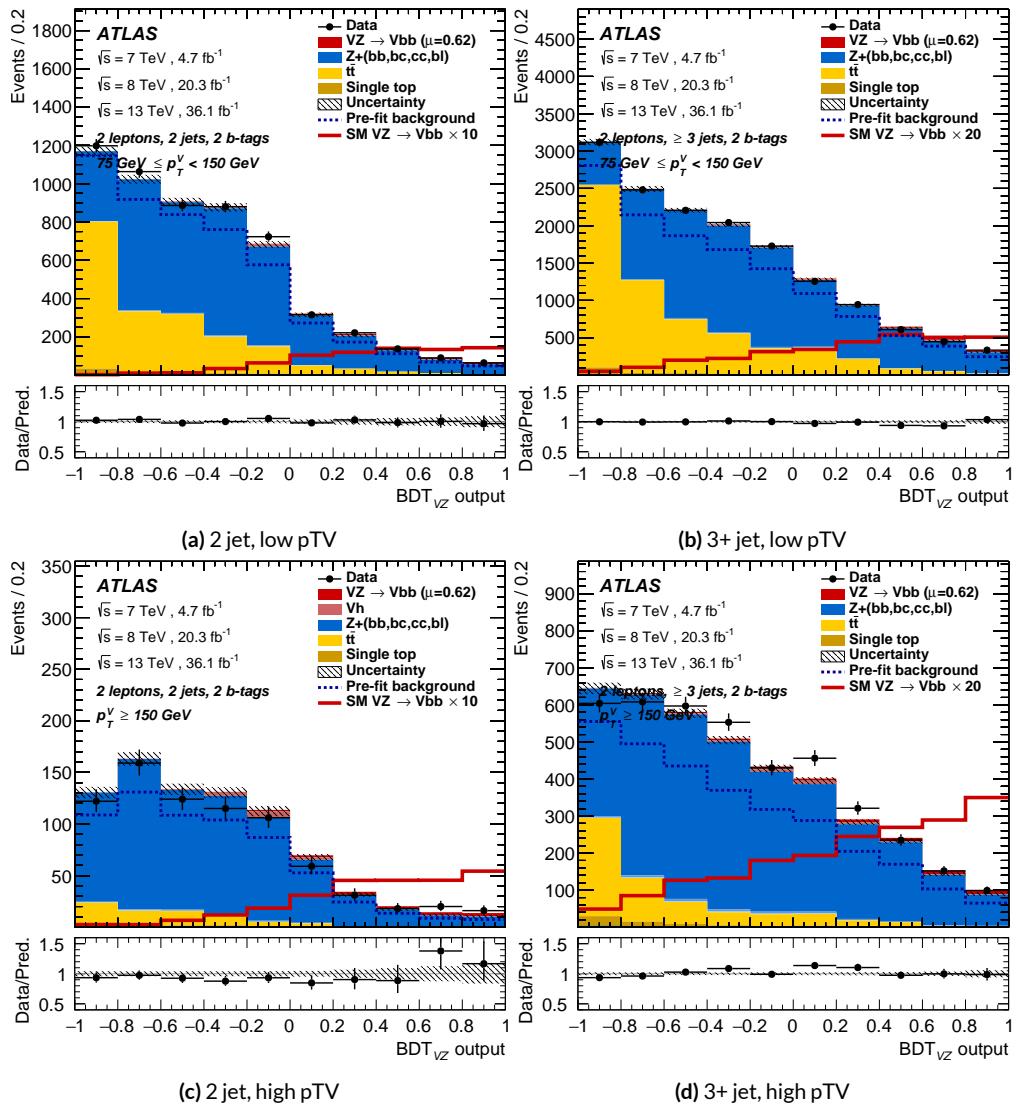
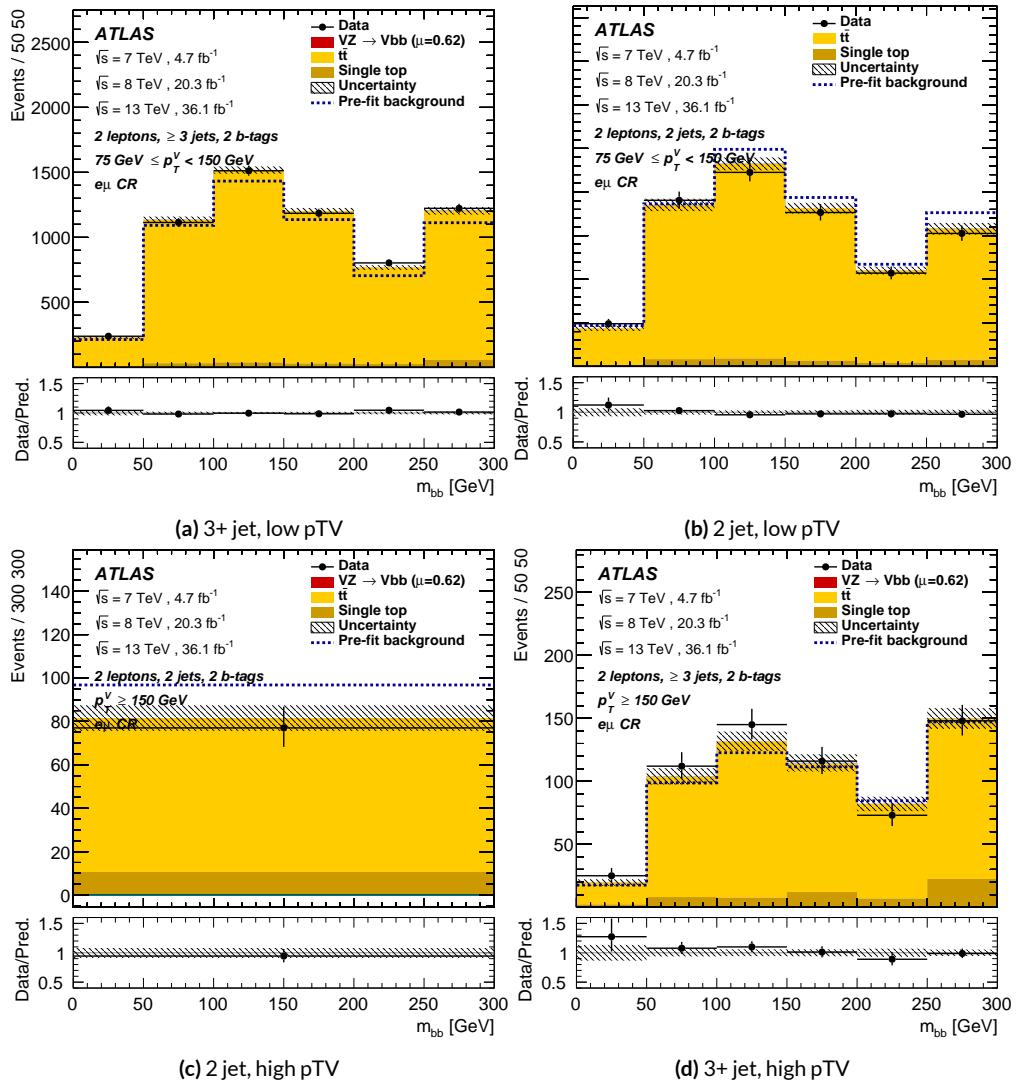


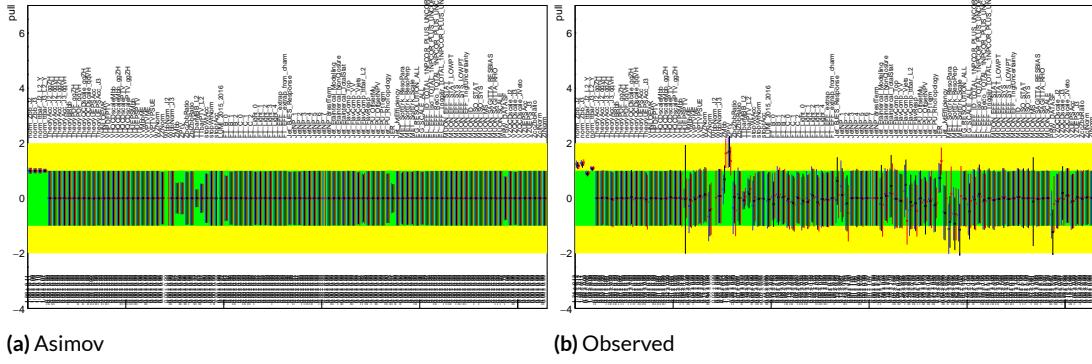
Figure 7.17: Postfit  $m_{bb}$  plots in the top  $e - \mu$  CR for the LI variable set.



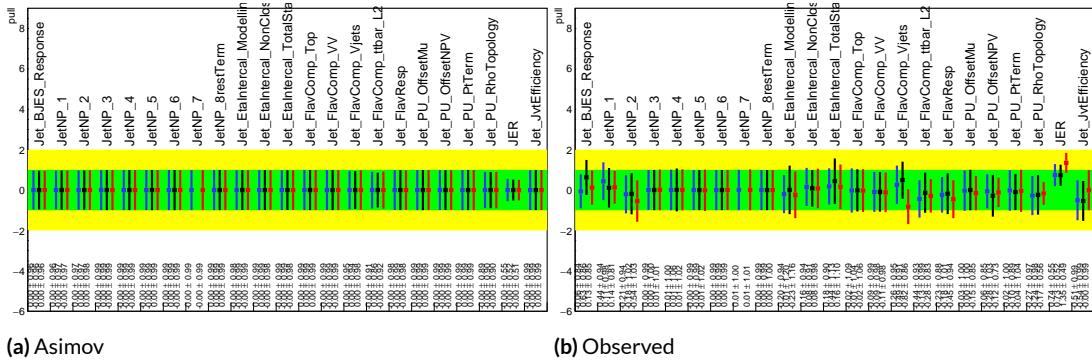
**Figure 7.18:** Postfit  $BDT_{VZ}$  plots in the signal region for the RF variable set.



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



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



**Figure 7.21:** Pull comparison for jet NP's.

1702 though this difference goes away in the ranking plot.

1703 Nuisance parameter correlation matrices (for correlations with magnitude at least 0.25) for all

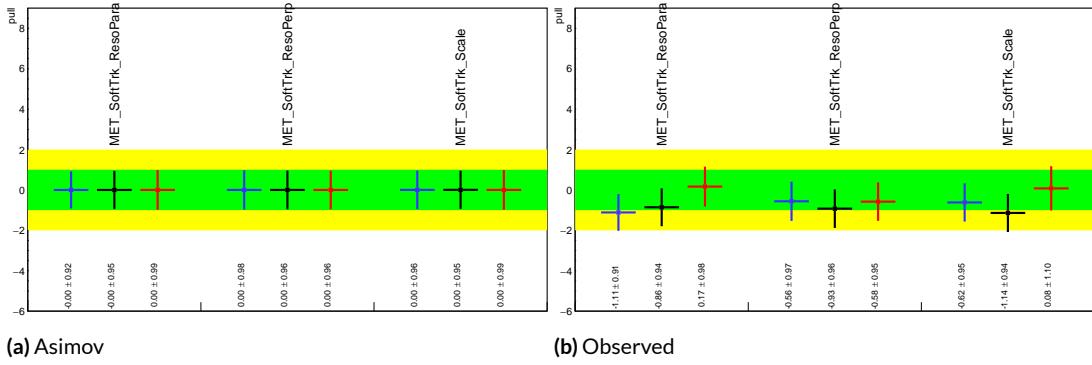
1704 three variable set fits can be found in Figures 7.26–7.28.

### 1705 7.7.2 FULL BREAKDOWN OF ERRORS

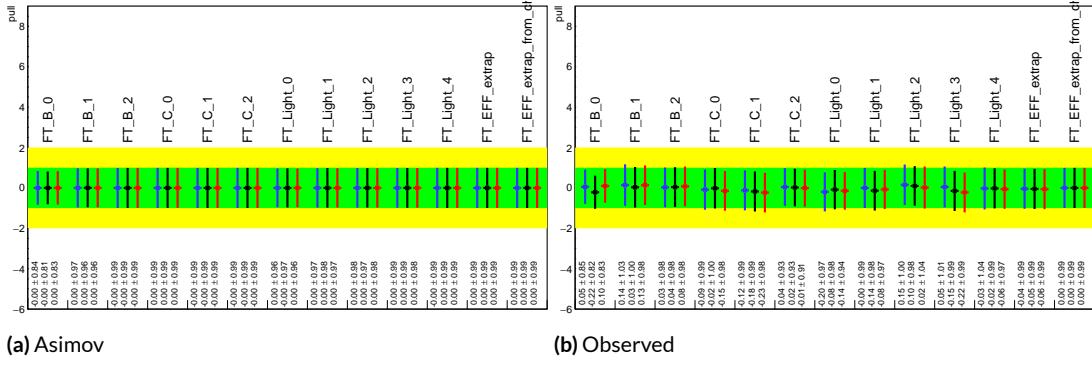
1706 A postfit ranking of nuisance parameters according to their impact on  $\hat{\mu}$  for the different variable

1707 sets may be found in Figure 7.29, with rankings being fairly similar. In particular, the signal UE+PS

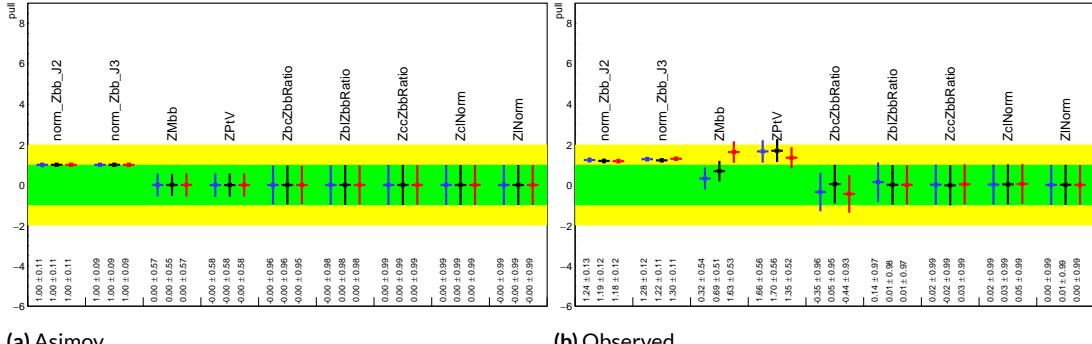
1708  $p_T^V$  systematic is top-ranked for all three variable sets and also looks very similar, unlike in the pull



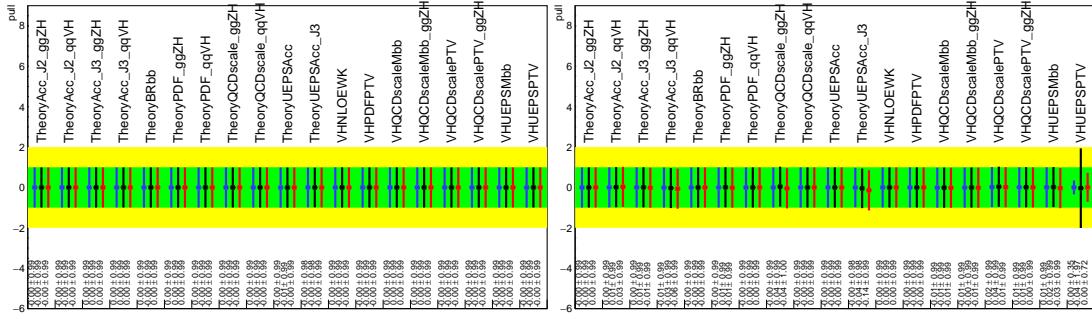
**Figure 7.22:** Pull comparison for MET NP's.



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



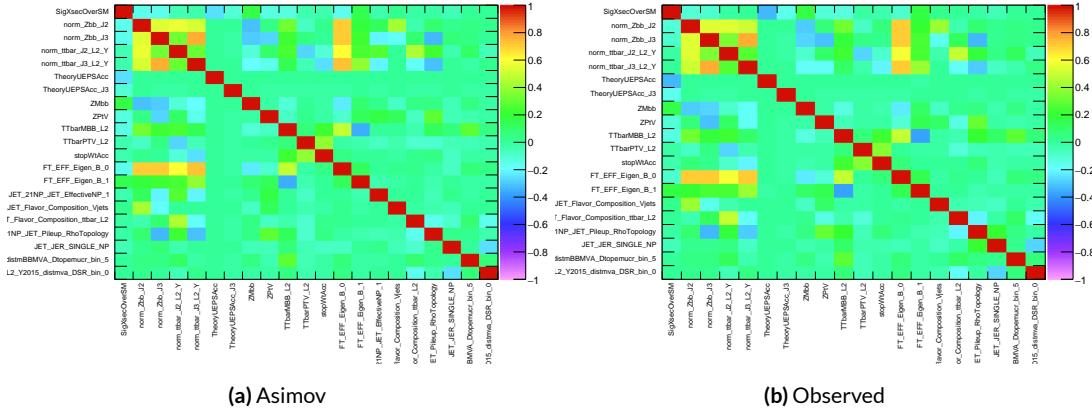
**Figure 7.24:** Pull comparison for  $Z$ +jets NP's.



(a) Asimov

(b) Observed

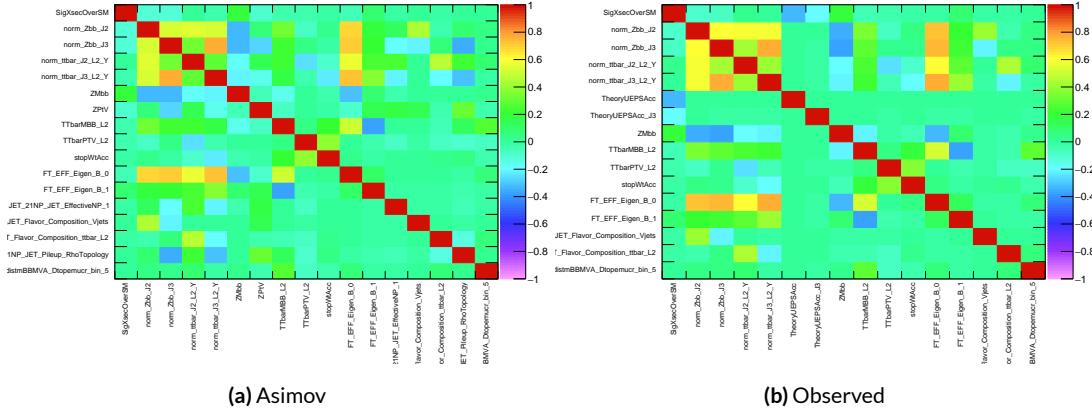
Figure 7.25: Pull comparison for signal process modeling NP's.



(a) Asimov

(b) Observed

Figure 7.26: NP correlations for standard variable fits.



(a) Asimov

(b) Observed

Figure 7.27: NP correlations for L1 variable fits.

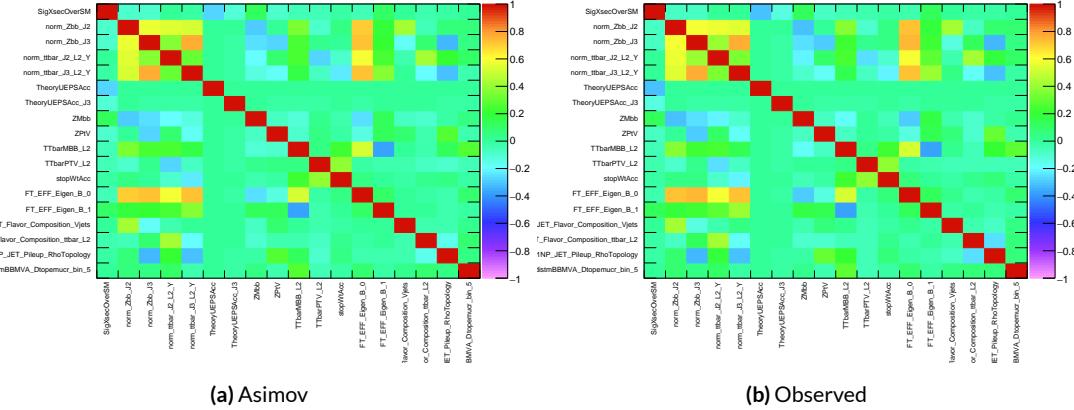


Figure 7.28: NP correlations for RF variable fits.

1709 comparison plot, reiterating the importance of evaluating individually the impact of highly ranked  
 1710 NP's. The  $Z+jets p_T^V$  is highly pulled in all three cases, though this is less severe for the non-standard  
 1711 set (it is off the scale for the standard). The RF discriminant mitigates the effect of poorly modeled  
 1712 jet energy resolution better than the other sets.

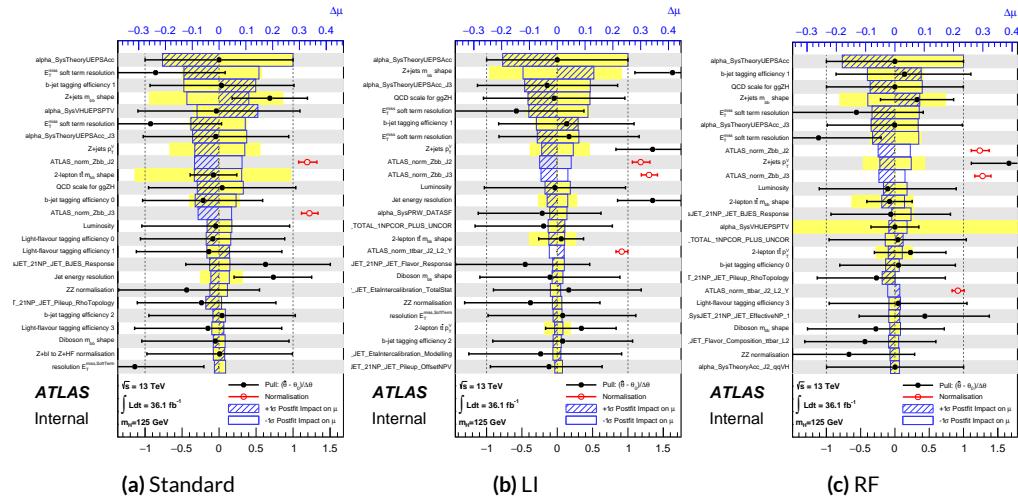


Figure 7.29: Plots for the top 25 nuisance parameters according to their postfit impact on  $\hat{\mu}$  for the standard (a), LI (b), and RF (c) variable sets.

<sup>1713</sup> The Asimov (Table 7.7) and observed (Table 7.8) breakdowns both consistently suggest that the  
<sup>1714</sup> LI variable set does a better job of constraining systematic uncertainties than the standard set and  
<sup>1715</sup> that the RF set does better still. It is also not surprising that the gain is more substantial in the ob-  
<sup>1716</sup> served fit than in the Asimov fits, as in the latter there are the “penalty” terms from pulls in addition  
<sup>1717</sup> 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

1718 7.7.3 POSTFIT DISTRIBUTIONS AND S/B PLOTS

1719 Postfit distributions for the MVA discriminant ( $m_{bb}$ ) distribution in the signal (top  $e - \mu$  control) re-  
1720 gion for the standard, Lorentz Invariant, and RestFrames variable sets. Here, as in the  $VZ$  fit, agree-  
1721 ment is reasonable. In a combined fit with all three channels,  $Z + hf$  normalizations in particular  
1722 would be correlated across the 0- and 2-lepton channels, which might help to better constrain this  
1723 mismodeling (and perhaps as a result some of the  $Z + jets$  systematics as well).

1724 One final type of plot presented as a result is the binned  $\log_{10} (S/B)$  in signal regions distribu-  
1725 tions may be found in Figure 7.36. For these plots, one fills a histogram with the  $\log_{10} (S/B)$  ratio  
1726 in each postfit distribution bin weighted by the total number of events. In this case, a log plot is  
1727 helpful because otherwise the highest bins would be invisible on a linear plot. These distributions  
1728 are allegedly useful for seeing where most of one's sensitivity lies. Practically, it is problematic if the  
1729 pull (from the null hypothesis) is higher at lower  $S/B$  values, which may indicate a poorly optimized  
1730 discriminant.

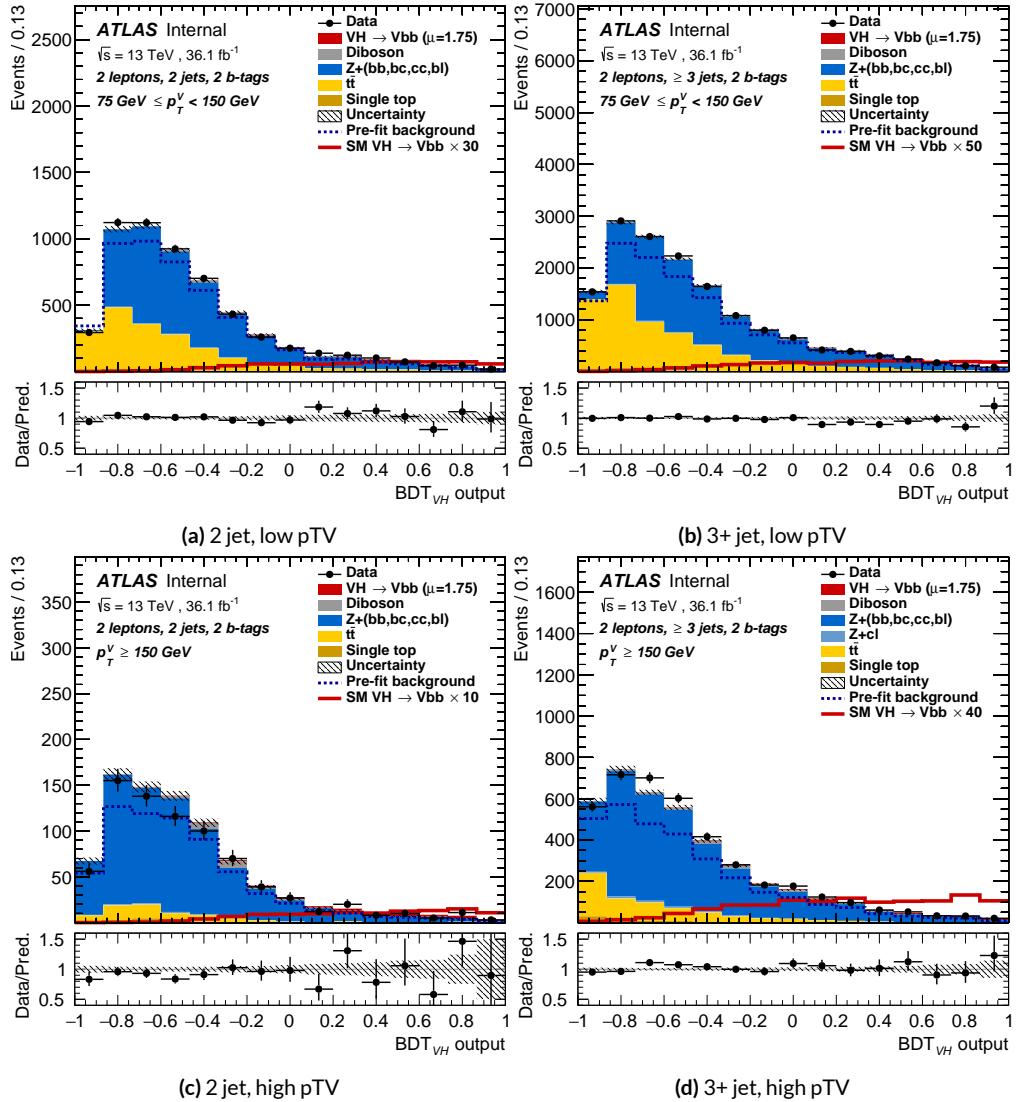
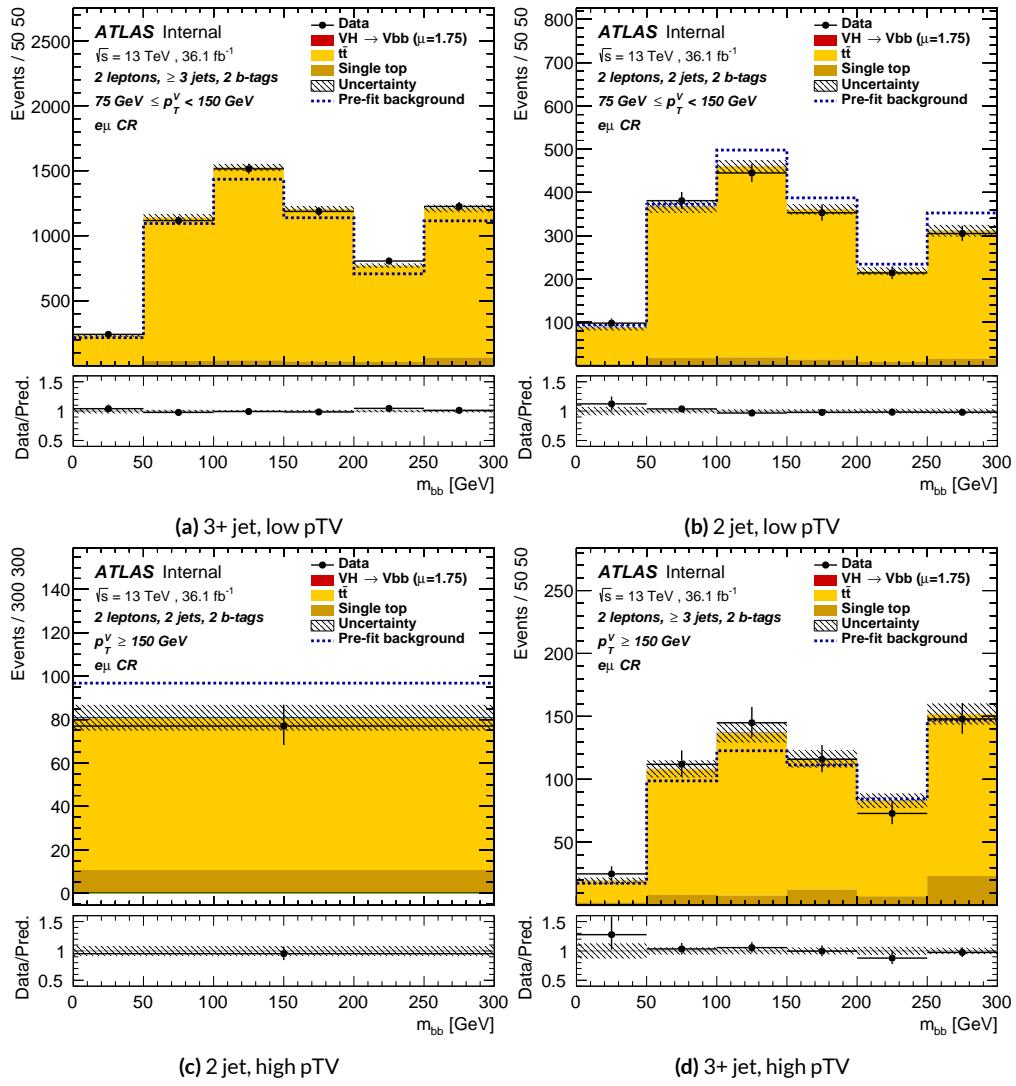


Figure 7.30: Postfit  $BDT_{VH}$  plots in the signal region for the standard variable set.



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

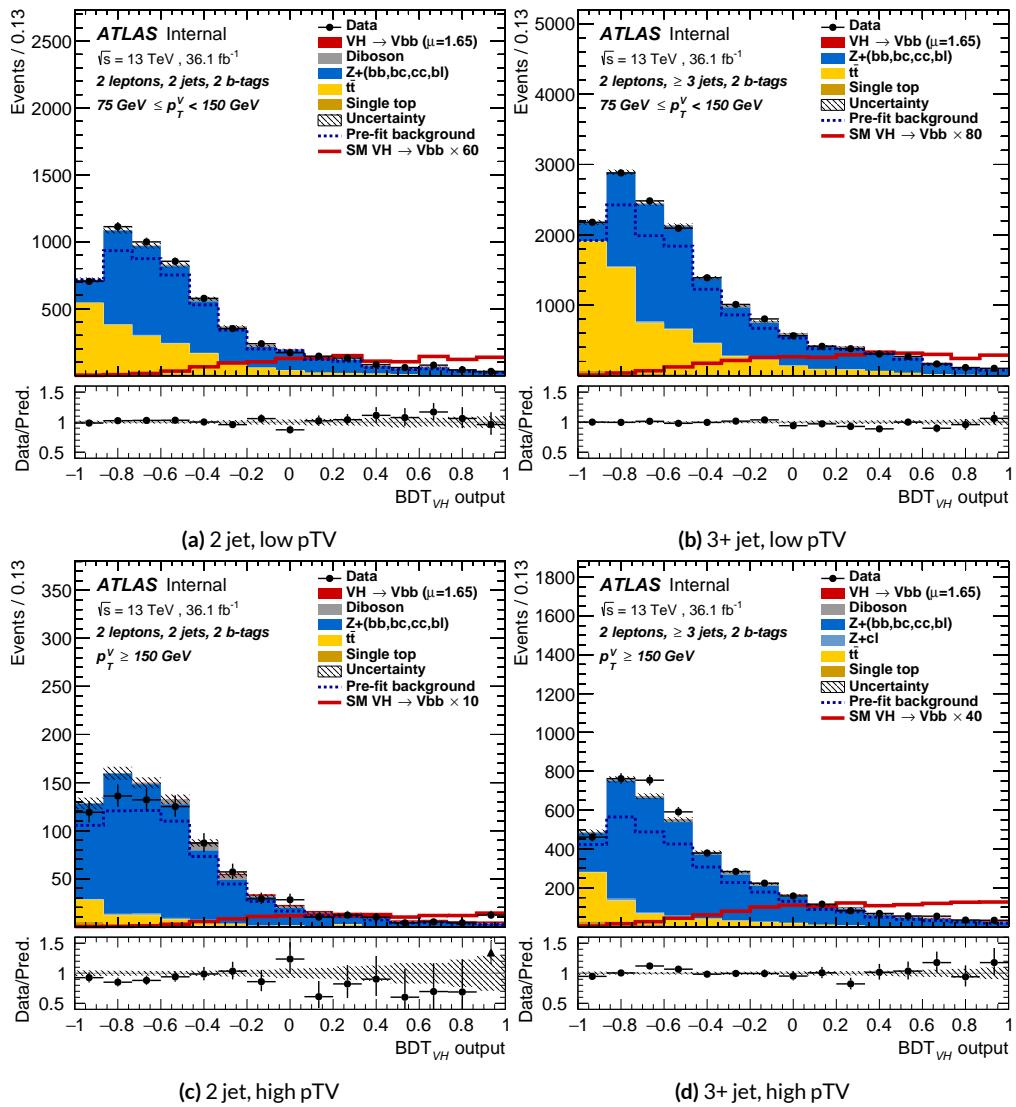


Figure 7.32: Postfit  $BDT_{VH}$  plots in the signal region for the LI variable set.

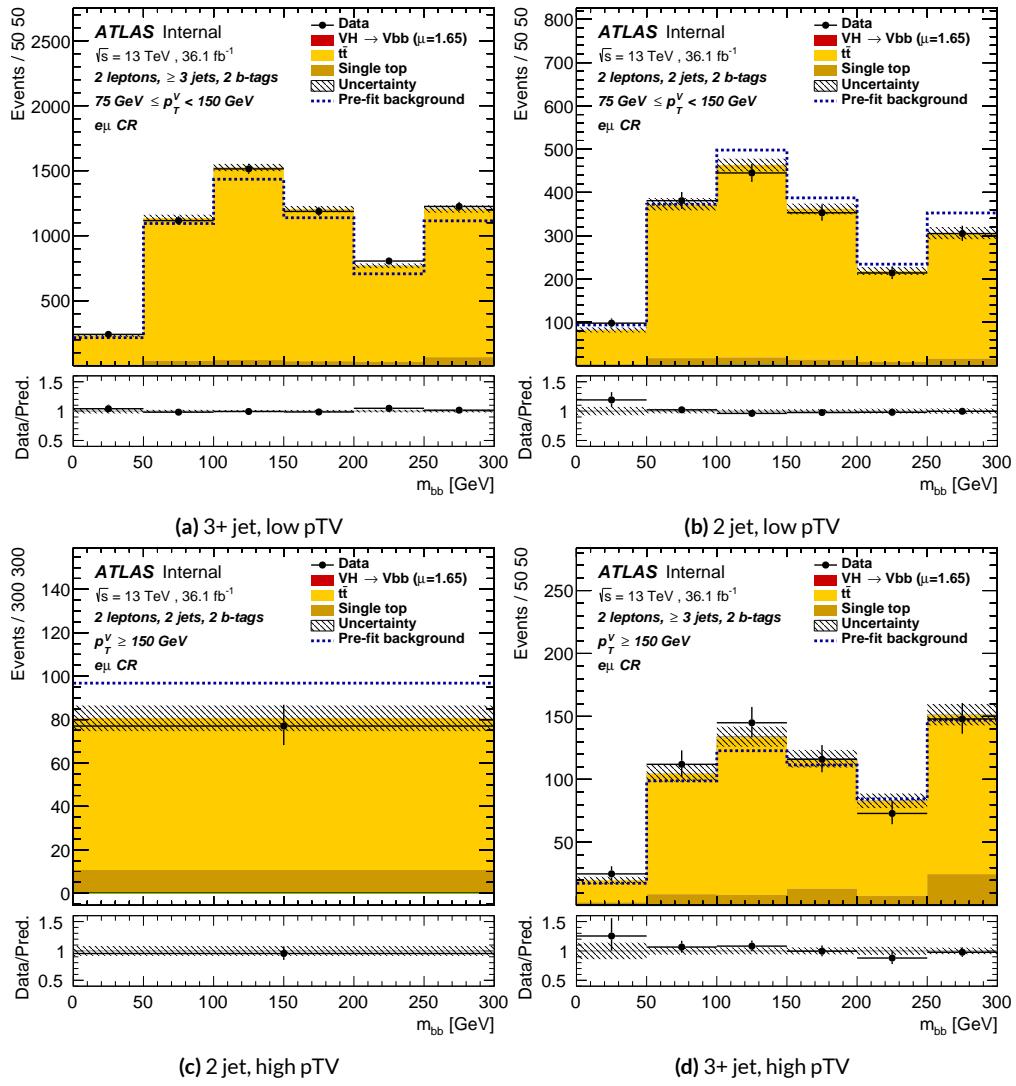
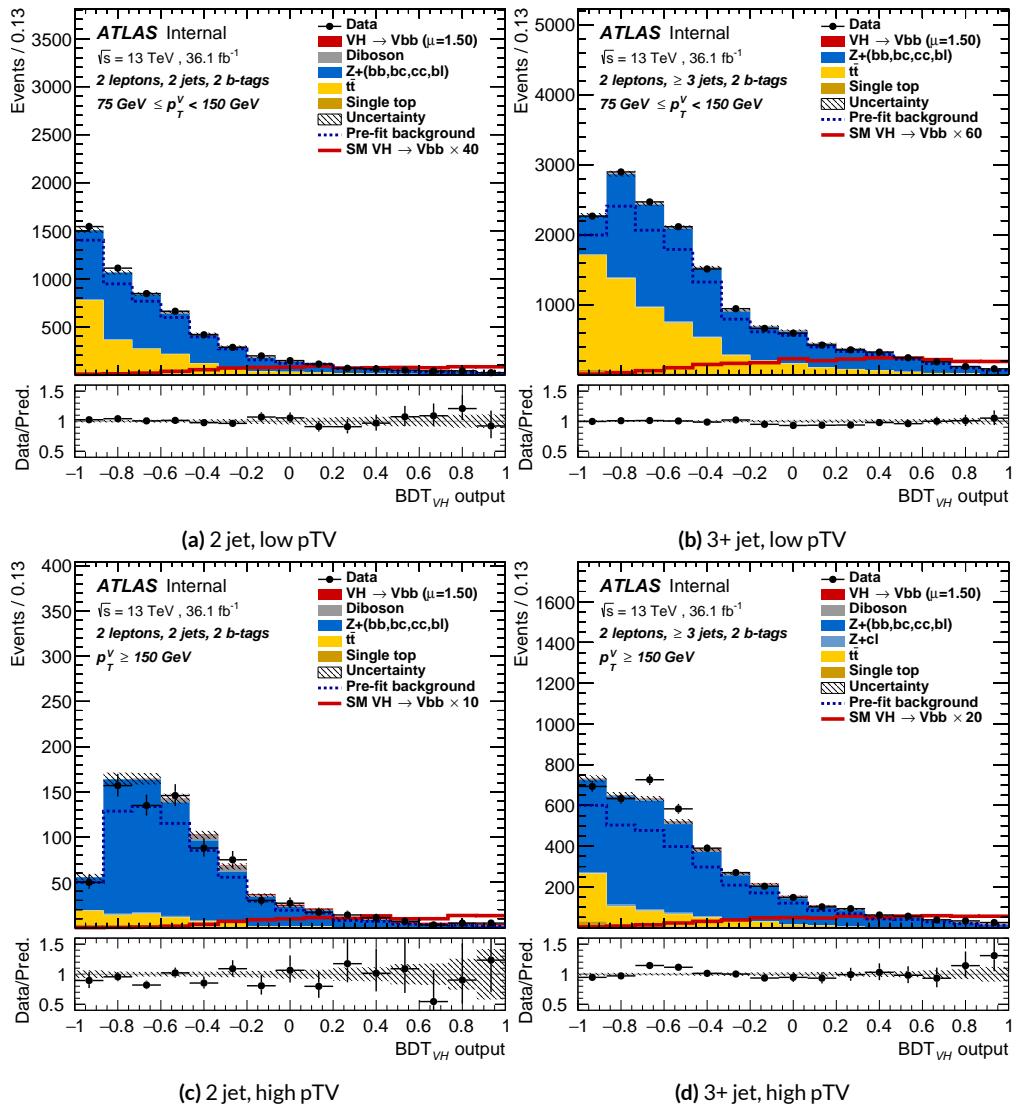
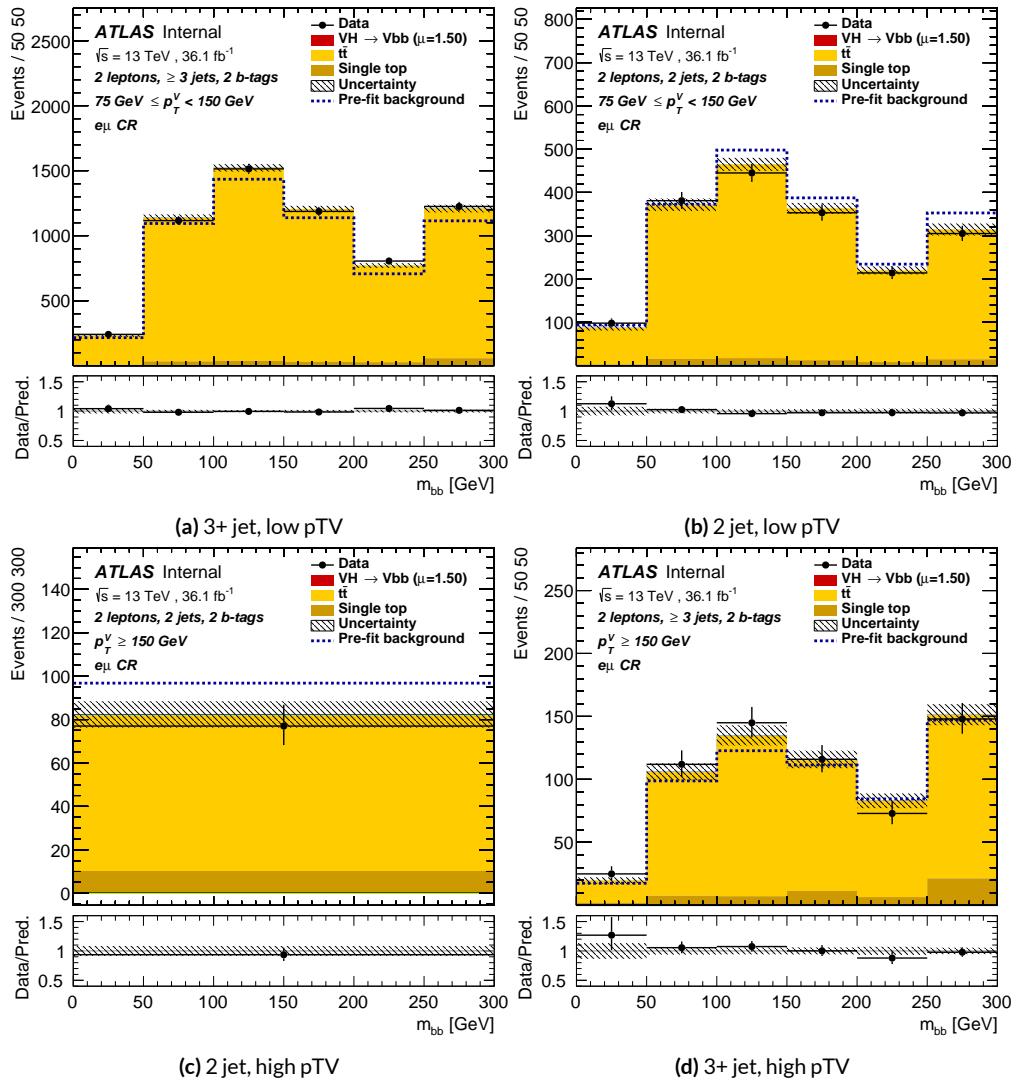


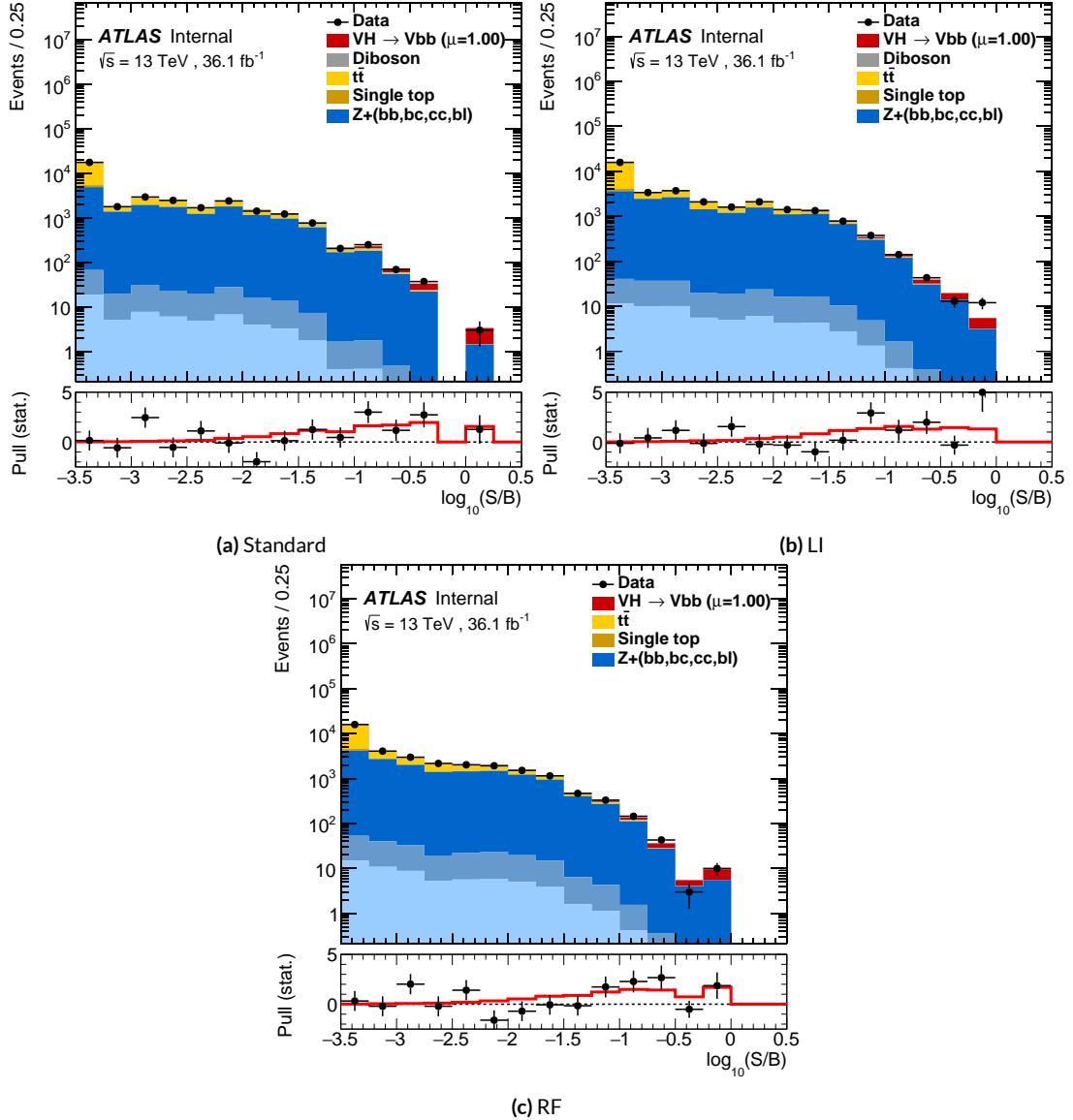
Figure 7.33: Postfit  $m_{bb}$  plots in the top  $e - \mu$  CR for the LI variable set.



**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

1731

1732

## Fit Results

1733 THE RESULTS IN THIS CHAPTER were first reported in <sup>29</sup> and describes how the three different fit models detailed and validated Chapter 7, corresponding to the standard, RF, and LI variable sets described in Chapter 6 perform on actual  $VH$  fits. In particular sensitivities, nuisance parameter impacts, and signal strengths on expected fits to Asimov datasets and both expected and observed

<sub>1737</sub> fits on the actual dataset are compared.

<sub>1738</sub> Expected and observed sensitivities for the different variable sets may be found in Table 8.1. The  
<sub>1739</sub> RF fits feature the highest expected sensitivities, outperforming the standard set by 3.5% and 3.4%  
<sub>1740</sub> for fits to Asimov and observed datasets, respectively. The LI variable has a lower significance than  
<sub>1741</sub> both for expected fits to both Asimov and data with a 6.7% (1.7%) significance than the standard set  
<sub>1742</sub> for the Asimov (observed) dataset. While the fit using standard variables does have a higher observed  
<sub>1743</sub> significance than both the LI and RF fits, by 2.8% and 8.6%, respectively, these numbers should be  
<sub>1744</sub> viewed in the context of the best fit  $\hat{\mu}$  values, discussed below. That is, the standard set may yield the  
<sub>1745</sub> highest sensitivity for this particular dataset, but this is not necessarily (and likely is not) the case for  
<sub>1746</sub> 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.

<sub>1747</sub> A summary of fitted signal strengths and errors for both the Asimov (a) and observed (b) datasets  
<sub>1748</sub> are shown in Figure 8.1.\* A summary of error breakdowns is given in Tables 8.2 (Asimov) and 8.3  
<sub>1749</sub> (observed) for total error, data statistics contributions, total systematic error contributions, and  
<sub>1750</sub> categories for which the total impact is  $\geq 0.1$  for the standard fit. As is to be expected for both the  
<sub>1751</sub> Asimov and observed dataset fits, the contribution to the total error on  $\mu$  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.)

<sup>1752</sup> is nearly identical, since each set of fits uses the same selections and data.<sup>†</sup>

	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  $\mu$  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.

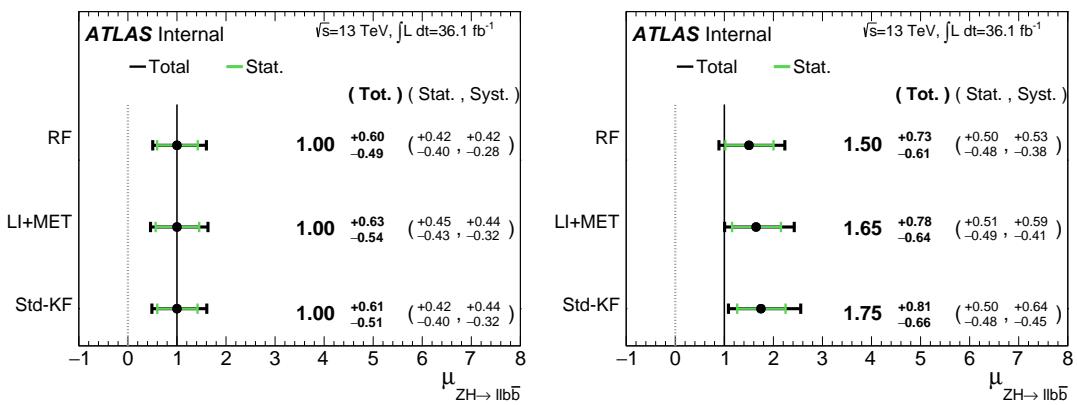
<sup>1753</sup> The contribution from systematic uncertainties, however, does vary considerably across the vari-  
<sup>1754</sup> able sets. The Asimov fits are a best case scenario in the sense that, by construction, all NP's are equal  
<sup>1755</sup> to their predicted values (and so no "penalty" is paid for pulls on Gaussian NP's). The systematics  
<sup>1756</sup> error from the LI fit is slightly higher (subpercent) than that from the standard fit, and 4.6% higher  
<sup>1757</sup> error overall due to differences in data stats. The RF Asimov fit, however, has a 6.5% lower total er-

---

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

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

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



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

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

<sub>1771</sub>      The marginally worse performance of the LI and RF variables (7.9% and 6.9%, respectively) with  
<sub>1772</sub> respect to the standard variable at a stats only level illustrates that neither variable set has greater  
<sub>1773</sub> intrinsic descriptive power in the absence of systematics in this closed final state. Hence, any gains  
<sub>1774</sub> from either of these variable sets in a full fit come from improved treatment of systematic uncertain-  
<sub>1775</sub> ties.

<sub>1776</sub>      With full systematics, the LI variable set narrows the sensitivity gap somewhat, with lower signif-  
<sub>1777</sub> icances by 6.7% (1.7%) on expected fits to Asimov (data) and by 2.8% on observed significances. The  
<sub>1778</sub> RF variable set outperforms the standard set in expected fits with 3.5% (3.4%) higher significance  
<sub>1779</sub> on Asimov (data), but has an 8.6% lower observed significance, though the observed significances  
<sub>1780</sub> should be viewed in the context of observed  $\hat{\mu}$  values.

<sub>1781</sub>      Moreover, the LI and RF variable sets generally perform better in the context of the error on  $\mu$ .  
<sub>1782</sub>      The LI fit is comparable to the standard set on Asimov data and has a 7.5% lower total systematics  
<sub>1783</sub> error on  $\hat{\mu}$  on observed data, while the RF fit is lower in both cases, with systematics error being  
<sub>1784</sub> 6.5% (16%) lower on Asimov (observed) data.

<sub>1785</sub>      These figures of merit suggest that both the LI and RF variables are more orthogonal than the  
<sub>1786</sub> standard variable set used in the fiducial analysis. Moreover, the RF variable set does seem to con-  
<sub>1787</sub> sistently perform better than the LI set. Furthermore, both variable sets have straightforward exten-  
<sub>1788</sub> sions to the one lepton channel in the  $VH(b\bar{b})$  analysis, and the RF set has a straightforward ex-  
<sub>1789</sub> tension to the zero lepton channel as well. The magnitude of any gain from the more sophisticated  
<sub>1790</sub> treatment of  $E_T^{miss}$  in these extensions is beyond the scope of these studies, but the performance in  
<sub>1791</sub> this closed final state do suggest that there is some value to be had in these non-standard descriptions

<sup>1792</sup> independent of these considerations.

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

Isaac Newton

# 9

1793

1794

## Measurement Combinations

1795 WHILE THE DISCUSSION thus far has focused on improvements looking towards future in just  
1796 the  $ZH \rightarrow \ell\ell b\bar{b}$  channel, any actual result for SM  $VH(b\bar{b})$  combines all channels and all avail-  
1797 able datasets. Using additional channels at a given center of mass energy is straightforward since  
1798 the fit model is designed with this combination in mind. Combining dataset results (known as

1799 “workspaces”) from different center of mass energies is not so simple an exercise since both the un-  
1800 derlying physics (and its associated modeling) and the treatment of key experimental considerations,  
1801 like flavor tagging, and their associated systematics change from dataset to dataset. A combined fit  
1802 model must take these considerations into account, and the formulation of the fit model combin-  
1803 ing the Run 1 ( $\sqrt{s} = 7$  TeV with  $4.7 \text{ fb}^{-1}$  of data, and  $\sqrt{s} = 8$  TeV with  $20.3 \text{ fb}^{-1}$  of data) and Run 2  
1804 ( $\sqrt{s} = 13$  TeV with ) SM  $VH(b\bar{b})$  results is the topic of Section 9.1. Its results, as reported in <sup>34</sup>, are  
1805 given in 9.2.

1806 9.1 THE COMBINED FIT MODEL

1807 It is clear the signal strength parameter of interest should be fully correlated among the different  
1808 datasets. Some signal modeling systematics were left unchanged from Run 1 through Run 2 and/or  
1809 were designed to be explicitly correlated. Beyond these two special cases, it is not immediately clear  
1810 what level of correlation should be imposed. The general methodology for settling upon a correla-  
1811 tion scheme is as follows:

- 1812 1. Identify which NP categories have significant impacts on  $\mu$
- 1813 2. Of these NP’s, identify which have one-to-one correspondences or established correlation  
1814 schemes among  $\sqrt{s}$  values
- 1815 3. Test whether correlation has a sizeable impact on expected fit quantities

1816 The only two sizeable experimental NP categories are jet energy scale (JES) and flavor tagging  
1817 systematics. Correlation schemes of varying degrees of completeness exist for these categories, so ex-  
1818 plicit NP correlations can be tested for these two categories. As these studies were conducted before

1819 unblinding, “sizeable impact” was judged by comparing fit results (sensitivities, pull comparisons,  
1820 and breakdowns) on combined workspaces using the unblinded and public Run 1  $\mu = 0.51$  result  
1821 and Asimov data for the Run 2 result. These are treated in Sections 9.1.1 and 9.1.2.

1822 Modeling systematics require a slightly different treatment, and are explored in 9.1.3.

1823 As a general note when looking at pull comparison plots for combined workspaces, the error bars  
1824 in these plots are calculated using a simultaneous HESSE matrix inversion, which can fail to give  
1825 sensible values for high dimensional models (the combined workspaces have well over 500 NP’s).

1826 This is not true of the nuisance parameter ranking plots, which use a MINOS based approach to test  
1827 the effect of each NP individually. This is much slower but much more rigorous, which is why only  
1828 ranking plots appear outside of supporting material and pull comparisons are considered “diagnos-  
1829 tic” plots.

### 1830 9.1.1 JET ENERGY SCALE SYSTEMATICS

1831 Fortunately for the case of jet energy scale systematics, the JetEtMiss group provides two recom-  
1832 mended “strong” and “weak” correlation schemes between Run 1 and Run 2. These were used as a  
1833 point of departure for the JES combination correlation scheme. However, the JES NP’s in both the  
1834 Run 1 and Run 2 workspaces are a reduced set of NP’s, with some 56 (75) NP’s reduced to 6 (8) for  
1835 Run 1 (2). In order to restore the full set of JES NP’s, the effective NP’s in each workspace are un-  
1836 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:

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

1838 where *eff* and *unf* are for effective and unfolded NP's, respectively, the  $A_{ij}$ 's are scalar coefficients

1839 taken from raw maps, and  $|NP_{j,unf}|$  are the amplitudes of the unfolded NP's. The raw  $A_{ij}$  and  
1840 scaled maps for Run 1 and Run 2 may be found in Figure 9.1.

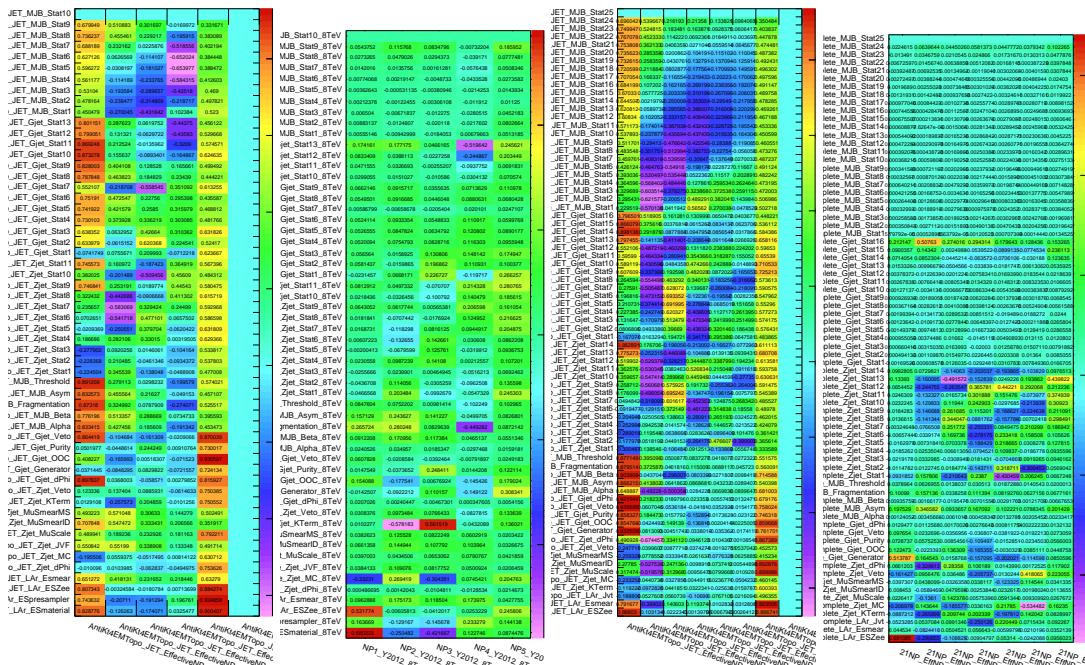


Figure 9.1: The raw and scaled coefficients for unfolding Run 1 (a and b) and Run 2 (c and d), respectively

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

1842 seen in Tables ?? and ??.

	R <sub>1</sub> Unfold	R <sub>1</sub> Eff	R <sub>2</sub> Unfold	R <sub>2</sub> 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 <sup>+0.296</sup> <sub>-0.211</sub>	0.755 <sup>+0.296</sup> <sub>-0.211</sub>	0.732 <sup>+0.287</sup> <sub>-0.205</sub>	0.732 <sup>+0.287</sup> <sub>-0.205</sub>	0.512 <sup>+0.201</sup> <sub>-0.143</sub>	0.51 <sup>+0.2</sup> <sub>-0.143</sub>
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 <sub>1</sub> Unfold	R <sub>1</sub> 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 <sub>1</sub> Unfold	R <sub>1</sub> Eff
R <sub>2</sub> Unfold	R <sub>2</sub> Eff	Comb Unfold
Comb Eff		
$ \Delta\hat{\mu} $	0.0	
$\hat{\mu}$	1.2051	1.2052
Total	+0.421 / -0.366	+0.421 / -0.366
DataStat	+0.239 / -0.234	+0.239 / -0.234
FullSyst	+0.346 / -0.282	+0.346 / -0.282
Jets	+0.066 / -0.047	+0.066 / -0.047
BTag	+0.119 / -0.106	+0.119 / -0.106

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

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

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

1843 It was also found that fit sensitivities and breakdowns were similarly indifferent to the use of  
1844 either the strong or weak JES correlation schemes, as shown in Tables 9.5 and 9.6.

	JES Weak Unfold	JES Weak Eff	JES Strong Unfold	JES Strong Eff
Exp. Sig.	3.57	3.57	3.59	3.59
Exp. Limit	0.493 <sup>+0.193</sup> <sub>-0.138</sub>	0.494 <sup>+0.193</sup> <sub>-0.138</sub>	0.493 <sup>+0.193</sup> <sub>-0.138</sub>	0.493 <sup>+0.193</sup> <sub>-0.138</sub>

**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.

1845 Comparisons of top ranked nuisance parameters in Figures 9.2–9.4 and for the complete JES pull

comparisons in Figures 9.5–9.8

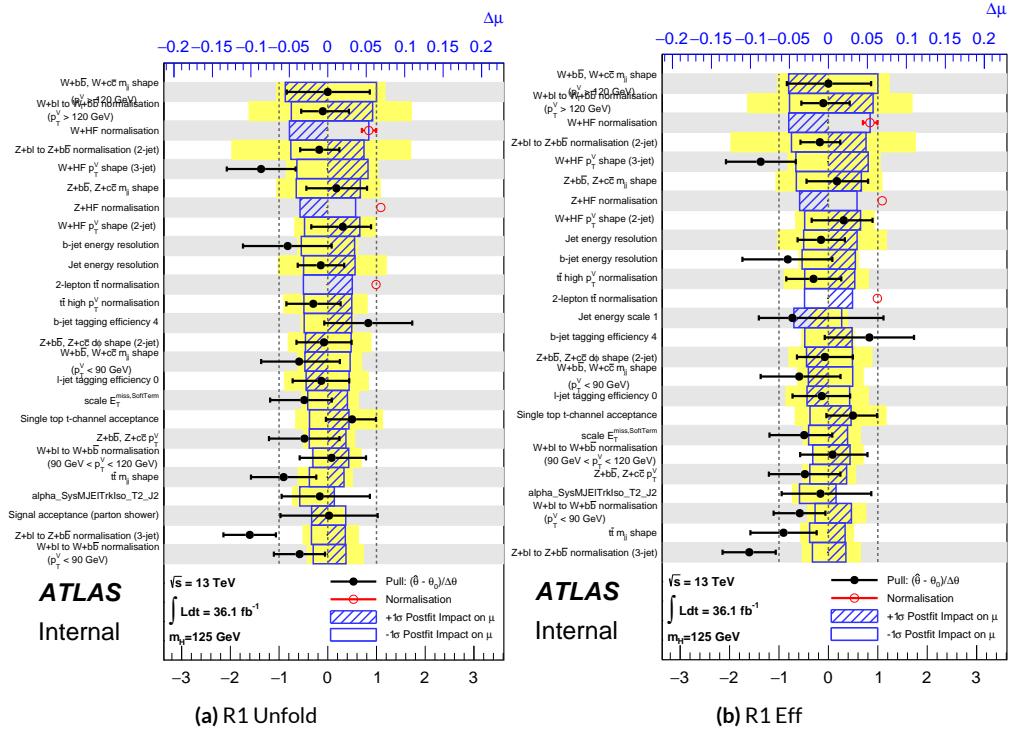


Figure 9.2: Ranks for the effective and unfolded JES NP Run1 combined workspaces.

1846

1847 As a result of these studies, the weak JES correlation scheme with uncorrelated effective JES NP's  
 1848 (i.e. just the  $b$ -jet energy scale NP) has been chosen as the treatment of JES in the Run 1 + Run 2  
 1849 combined fit.

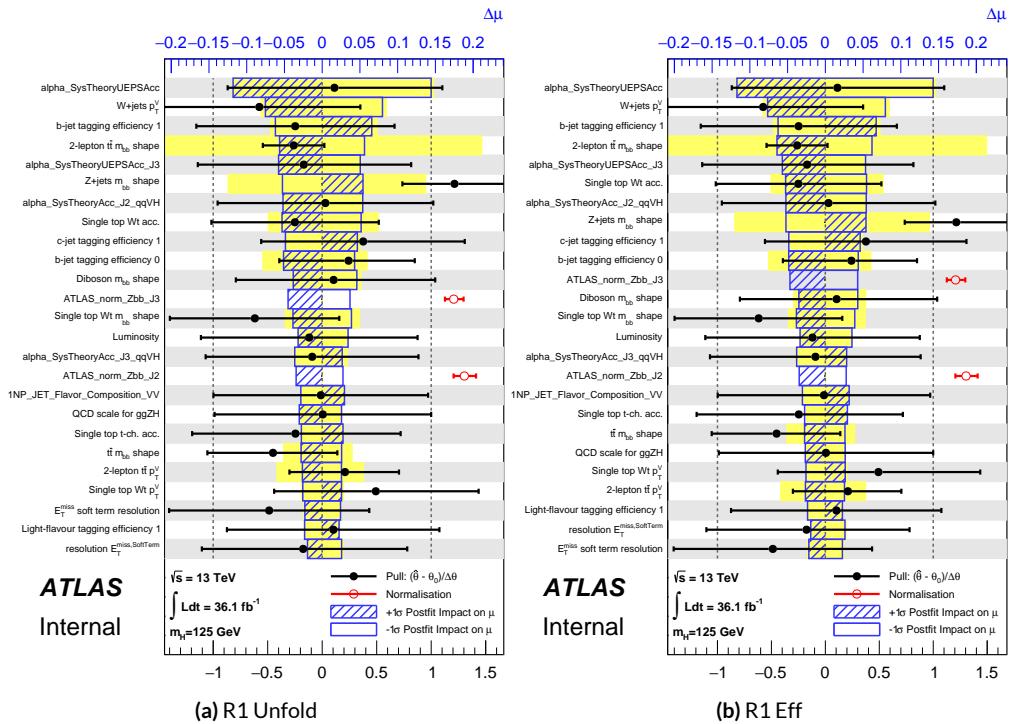


Figure 9.3: Ranks for the effective and unfolded JES NP Run2 combined workspaces.

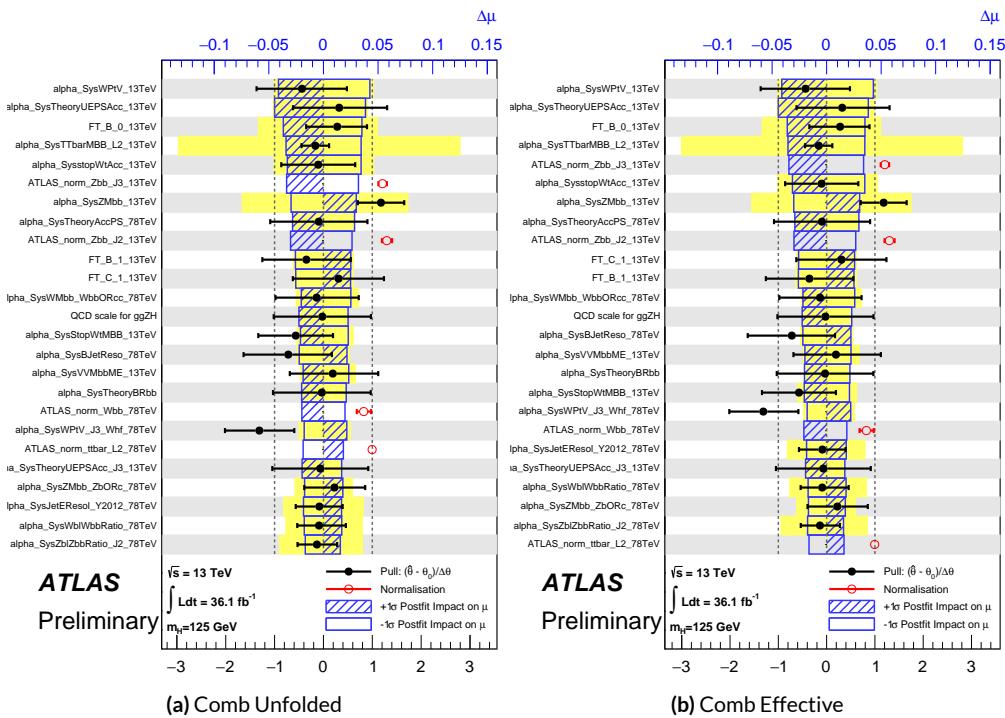


Figure 9.4: Ranks for the effective and unfolded JES NP Run1+Run2 combined workspaces.

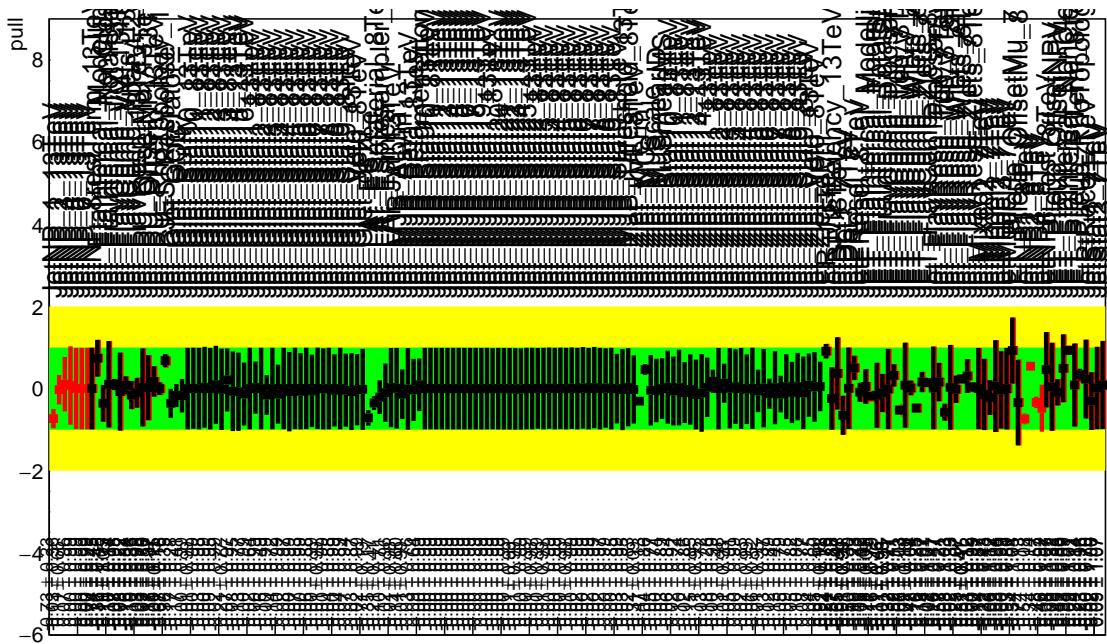


Figure 9.5: Pull Comparisons: jesu---Jet Comb Unfold, Comb Eff, Strong Unfold, Strong Eff

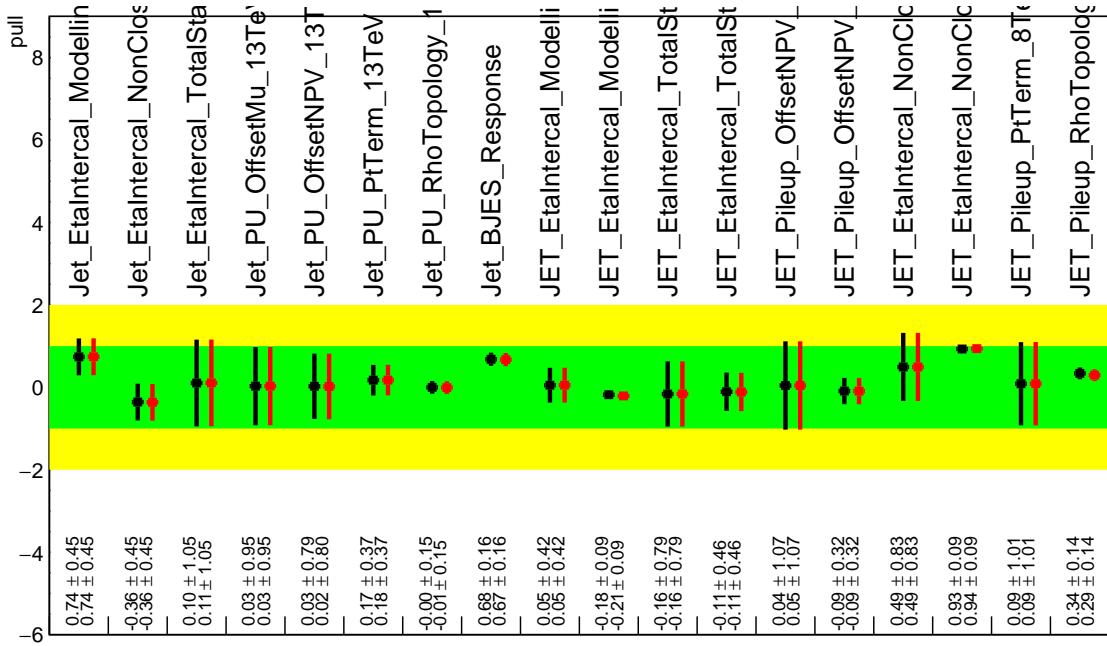


Figure 9.6: Pull Comparisons: jesu---JetMatched Comb Unfold, Comb Eff, Strong Unfold, Strong Eff

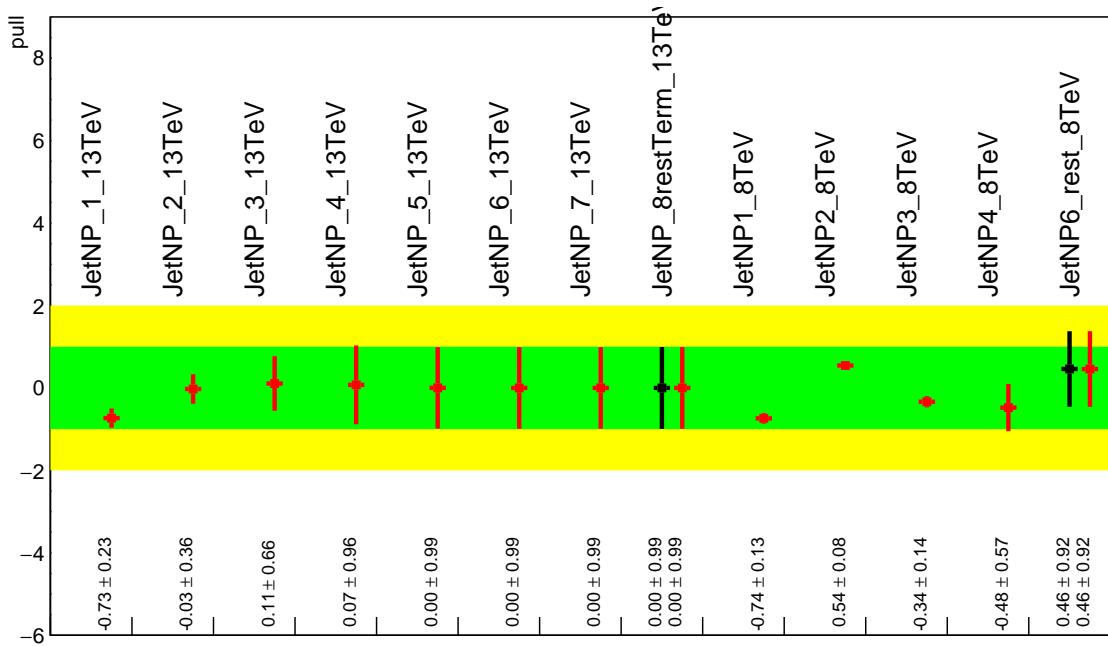


Figure 9.7: Pull Comparisons: jesu---JetEff Comb Unfold, **Comb Eff**, Strong Unfold, **Strong Eff**

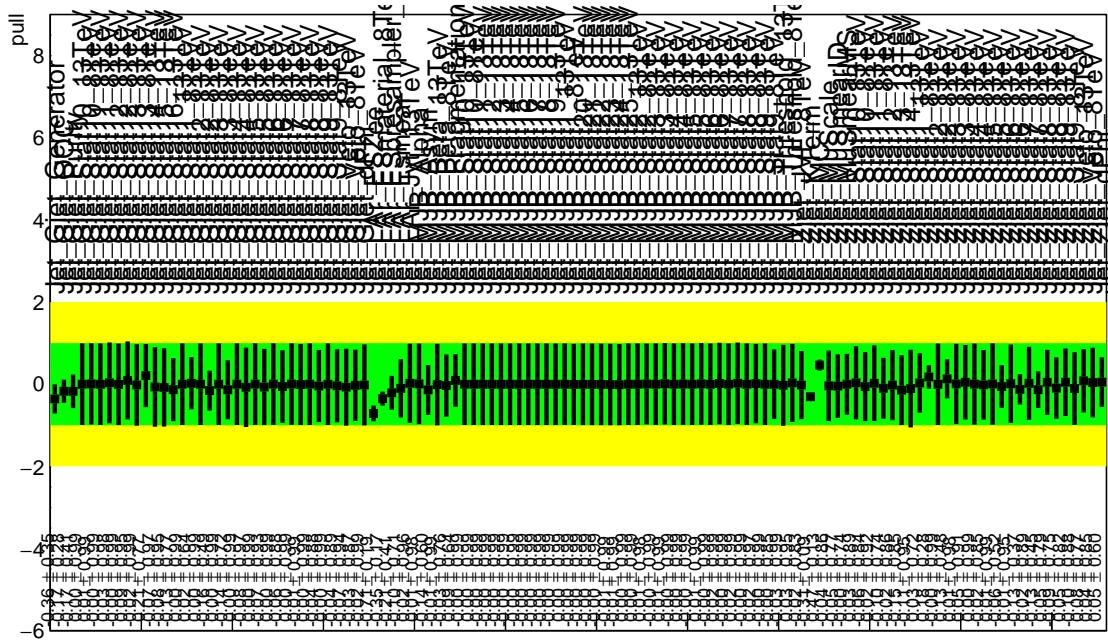


Figure 9.8: Pull Comparisons: jesu---JetUnfold Comb Unfold, **Comb Eff**, Strong Unfold, **Strong Eff**

1850 9.1.2 FLAVOR TAGGING

1851 Unfortunately, the ATLAS Flavor Tagging group did not provide any recommendations for cor-  
1852 relating Run 1 and Run 2 NP's, though given the high ranking of these NP's in the Run 2, result,  
1853 some studies at least were crucial. Nevertheless, great improvements and changes to the treatment  
1854 of flavor tagging between Run 1 and Run 2 does weaken the argument for any strong flavor tagging  
1855 correlation scheme.

1856 Given that  $c$ -tagging changed significantly between Run 1 and Run 2 and that light tagging NP's  
1857 are very lowly ranked, these sets of NP's are left uncorrelated. Moreover, the change in the physical  
1858 meaning of the effective  $b$ -tagging NP's means a full correlation of such NP's (insomuch as they exist  
1859 in each result) is one of limited utility. Hence, it was decided to leave flavor tagging NP's uncorre-  
1860 lated. However, since the meaning of the leading  $b$ -tagging NP's is approximately constant across  
1861 years and since Run 2  $b$ -tagging NP's are very highly ranked in both the Run 2 only and combined  
1862 fits, tests correlating these NP's were conducted, the results of which can be seen below. It should be  
1863 noted that the leading B NP at 8 TeV, `SysBTagB0Effic_Y2012_8TeV`, has an opposite effect on  $t\bar{t}$   
1864 normalization than the 7 and 13 TeV NP's, and so must be flipped using a similar strategy as for JES  
1865 unfolding. Initial studies of flavor tagging correlations did not flip this NP, and so results for this  
1866 scheme (labeled "Bo 8TeV Not Flipped") have also been included for comparison.

1867 It is clear from these results that correlating the leading effective Eigen NP associated with  $b$ 's can  
1868 have a noticeable effect on final fit results and that the 8 TeV Bo NP is the most important compo-  
1869 nent of a combined Bo NP. It is also not so surprising that the 8 TeV result should drive the com-

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

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

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

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

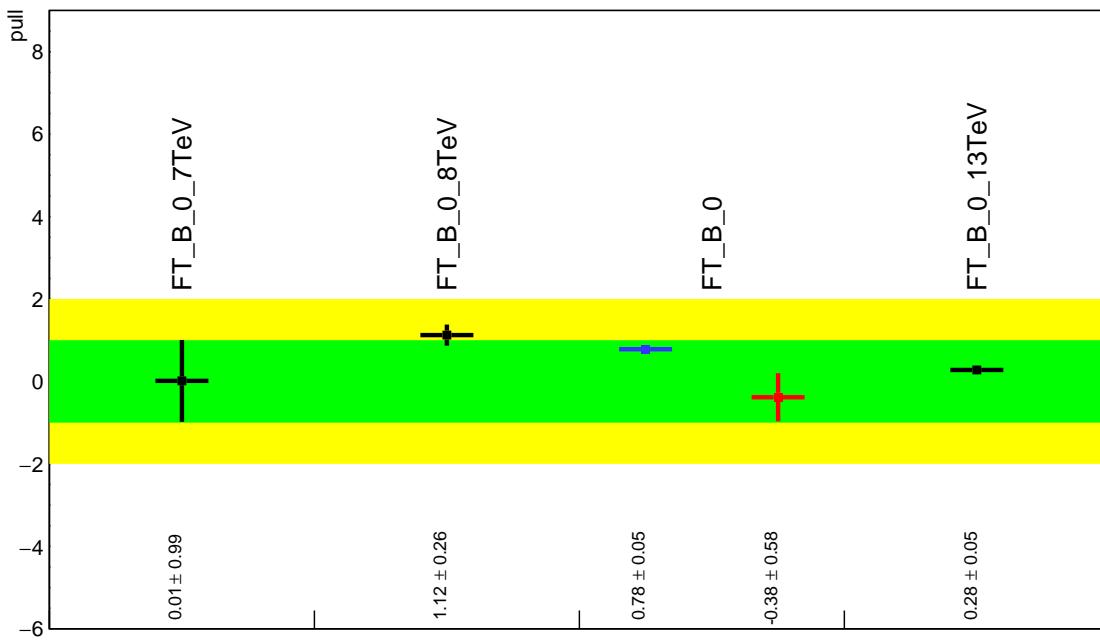


Figure 9.9: Pull Comparisons: btag-b---BTagBO Comb Eff, BTag BO

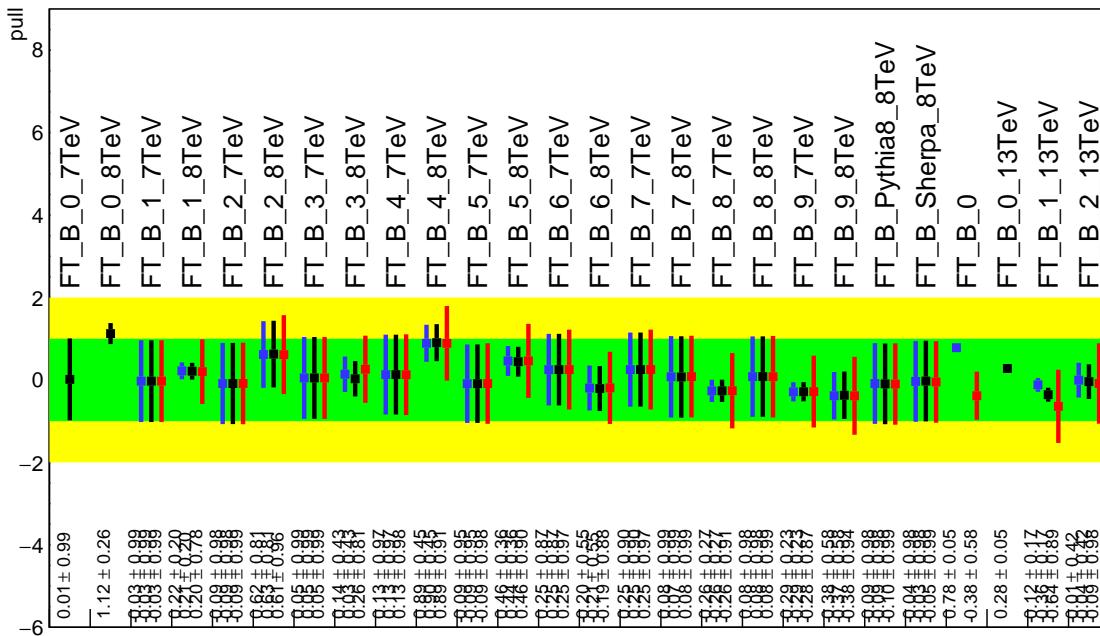
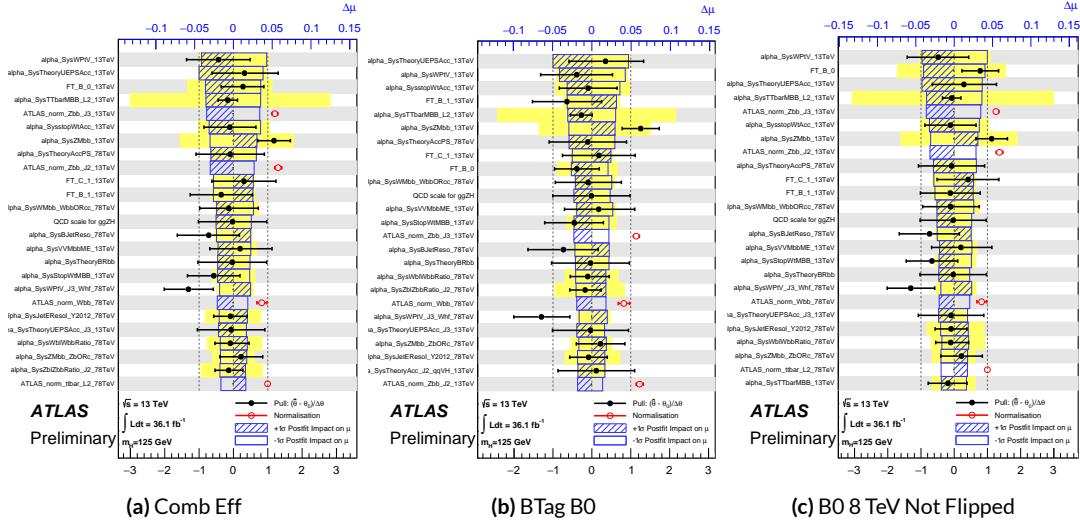


Figure 9.10: Pull Comparisons: btag-b---BTagB Comb Eff, BTag BO



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

1870 bined nuisance parameter since it is the only result to make use of both pseudocontinuous tagging-  
 1871 based and 1  $b$ -tag regions into the final fit, implicitly yielding much more information about  $b$ 's. The  
 1872 13 TeV fit has neither of these regions. What is less clear is whether there are sufficient grounds for  
 1873 doing so (i.e. does the correspondence of these NP's across years warrant a full correlation). While  
 1874 there are no current plans to do so, this matter warrants careful scrutiny if Run 1 is to be included in  
 1875 future results.

### 1876 9.1.3 MODELING SYSTEMATICS

1877 Another principal systematic category is modeling uncertainties. The effect of correlating groups  
 1878 of systematics was estimated using the same strategy employed by the ATLAS/CMS SM  $VH$  ( $b\bar{b}$ )  
 1879 combination for Run 1. This extrapolation can be used to estimate the impact of correlations on  
 1880 the estimated signal strength, the total error on the signal strength, and the  $\chi^2$  of the result. The

1881 impact of such correlations is no more than a few percent effect, as the following tables demonstrate,  
1882 beginning with the category with the greatest shift, W+jets modeling, in Table ??.

	$ \Delta\mu $	$\sigma$	$ \Delta\sigma $	$\chi^2$
$\rho=-1$	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

#### 1883 9.1.4 FINAL CORRELATION SCHEME

1884 The final Run 1 + Run 2 correlation scheme is shown in Table 9.10. As detailed above, neither JES  
1885 nor modeling systematics had any demonstrable effect on combined fit results. Hence, only signal  
1886 NP's and the  $b$ -jet energy scale are correlated. While the effect of flavor tagging correlations is less  
1887 clear, the result physical arguments for correlation are less strong; the size of effect was discovered  
1888 rather late in the analysis process; and has no nuisance parameter unfolding maps exist for flavor  
1889 tagging as they do for JES, so it was decided to leave these uncorrelated as well.

$7\text{ TeV NP}$	$8\text{ TeV NP}$	$13\text{ TeV NP}$
	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.

1890 **9.2 COMBINED FIT RESULTS**

1891 **9.2.1 COMBINED FIT MODEL VALIDATION**

1892 Before moving onto the final results, we present the rest of the validation for the Run 1 + Run 2  
1893 combined fits, beginning with impacts of ranked individual nuisance parameters in Figure 9.12 and  
1894 for all nuisance parameter categories in Table ???. Both of these sets of results point to the most im-  
1895 portant nuisance parameters being signal systematics,  $b$ -tagging, and  $V$ +jets modeling systematics,  
1896 with Run 2 NP's generally being higher ranked. That some NP's are strongly pulled is not unusual  
1897 as the fit model has so many NP's;  $V$ +jets modeling in particular has been historically difficult.

1898 In addition to looking at the behaviors of nuisance parameters to gauge fit model performance  
1899 and stability, fits are conducted using multiple parameters of interest. Typical divisions are Run 1  
1900 vs. Run 2, lepton channels, and  $WH$  vs  $ZH$ . As mentioned in Chapter 7, the profile likelihood test  
1901 statistic given in Equation 7.2 is, in the limit of large sample statistics, a  $\chi^2$  distribution with degrees  
1902 of freedom equal to the number of nuisance parameters plus nuisance parameters. Thus, changing  
1903 the number of interest parameters and leaving the rest of the fit model unchanged means that the

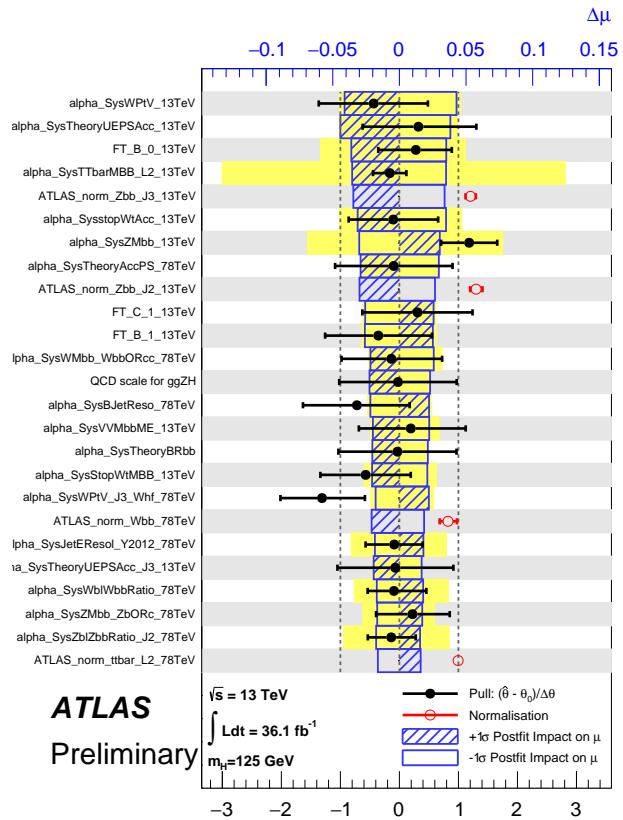


Figure 9.12: Ranked nuisance parameters for the Run1+Run2 combination.

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

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

1904 difference between the nominal fit and a fit with more parameters of interest ought to also be dis-  
 1905 tributed as a  $\chi^2$  distribution with degrees of freedom equivalent to the number of extra parameters  
 1906 of interest. This difference can then be interpreted as a compatibility between the two results using  
 1907 the standard tables for this distribution, giving another gauge of fit performance. These are shown  
 1908 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.

1909 The low compatabilities associated with treating the lepton channels as separate parameters of  
 1910 interest are a symptom of the low signal strengths associated with the Run 1 0-lepton channel, in par-  
 1911 ticular the 7 TeV result. Given the relatively small amount of data associated with the 7 TeV result,  
 1912 this should not be a cause for alarm. Signal strength summary plots for the fits treating Run 1 and  
 1913 Run 2 separately are shown in Figures 9.13-9.15, where the effect of the Run 1 parameters can be seen  
 1914 graphically.

### 1915 9.2.2 FINAL RESULTS

1916 The combined results yields an observed (expected) significance of 3.57 (4.00) and an observed (ex-  
 1917 pected) limit of 1.37 ( $0.510^{+0.200}_{-0.143}$ ), with a signal strength of  $\hat{\mu} = 0.898^{+0.278}_{-0.261}$ .

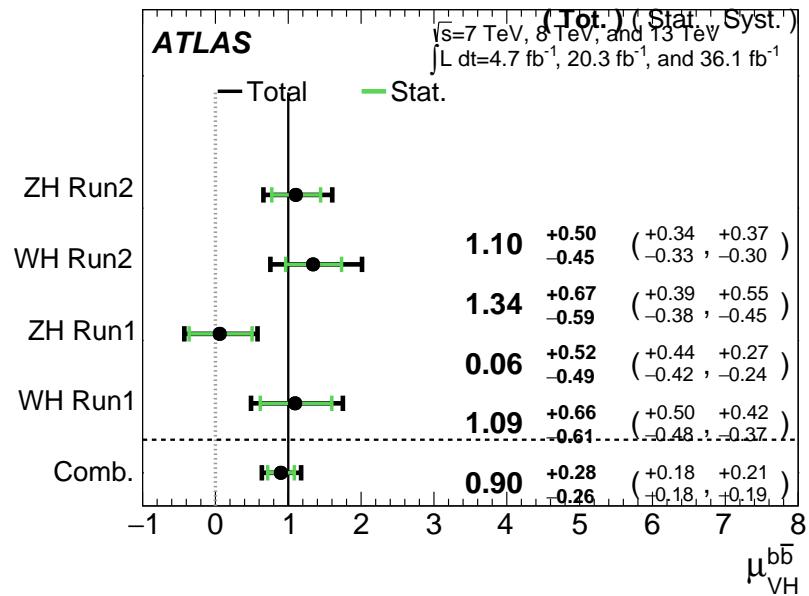


Figure 9.13:  $\hat{\mu}$  summary plot for a four parameter of interest fit.

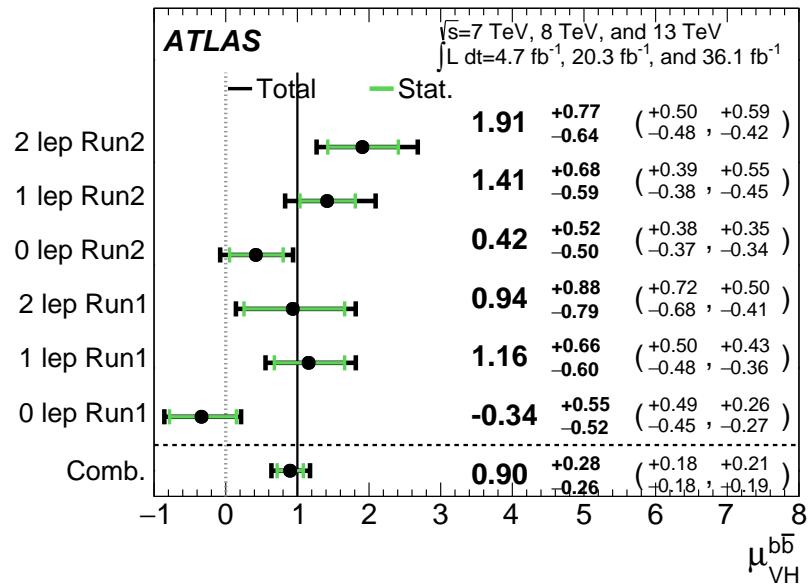


Figure 9.14:  $\hat{\mu}$  summary plot for a six parameter of interest fit.

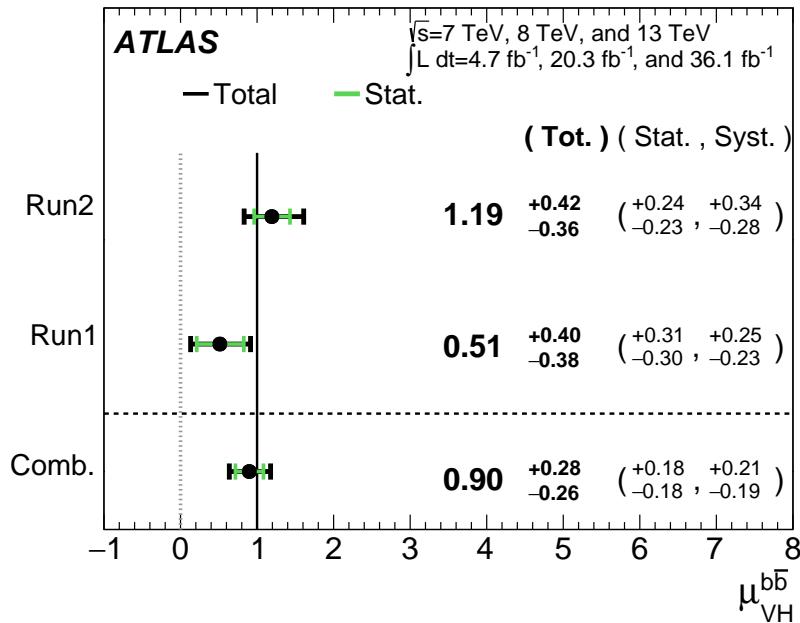


Figure 9.15:  $\hat{\mu}$  summary plot for a two parameter of interest (Run 1 and Run 2) values.

1918      The two and three parameter of interest fit signal strength summary plots, as well as a summary  
 1919      of the historical values of the 7, 8, and 13 TeV results may be found in Figures 9.16-9.18. The main  
 1920      results for Run 1, Run 2, and the combination may be found in Table 9.13. These results were collec-  
 1921      tively noted as the first ever experimental evidence for SM  $VH(b\bar{b})$  in<sup>34</sup>.

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

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

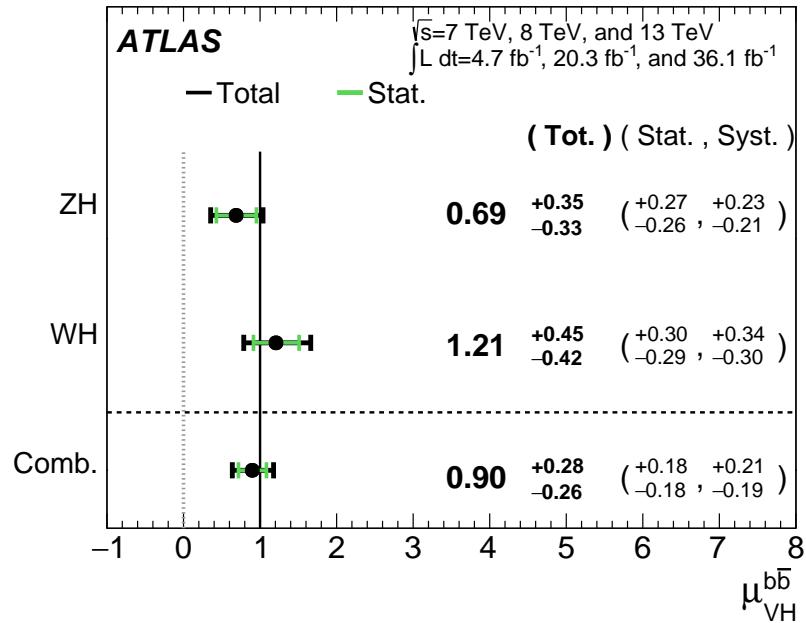


Figure 9.16:  $\hat{\mu}$  summary plot for a two parameter of interest fit.

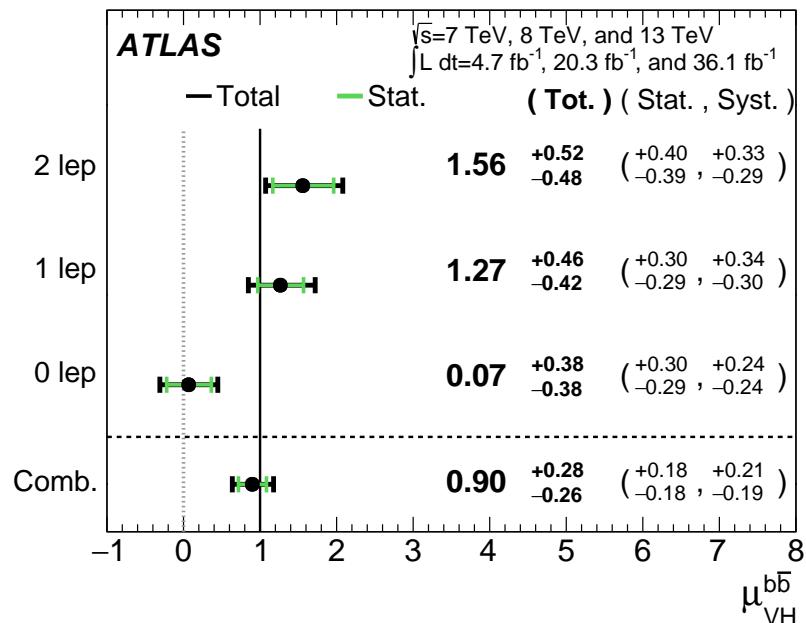


Figure 9.17:  $\hat{\mu}$  summary plot for a three parameter of interest fit.

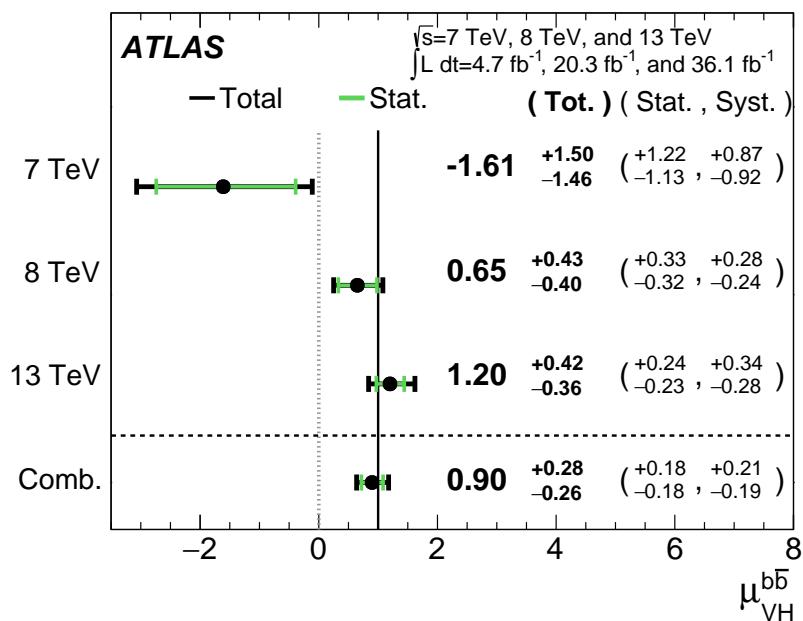


Figure 9.18:  $\hat{\mu}$  summary plot for different  $\sqrt{s}$  values.

*Vanitas vanitatum, omnis vanitas*

Ecclesiastes 1:2

# 10

1922

1923

## Closing Thoughts

- 1924 WITH BOTH THE LHC and ATLAS performing very well, it is only a matter of time before the  
1925 evidence for SM  $VH(b\bar{b})$  passes the 5 Gaussian standard deviation threshold necessary for discovery.  
1926 Depending the latter two years of ATLAS data from Run 2 of the LHC (2017 and 2018), this may  
1927 come less than a year after reports of first evidence and may not even require a combination with

<sup>1928</sup> Run 1 data.

<sup>1929</sup> It is entirely natural to ask, then, how essential the techniques and results described in this disser-  
<sup>1930</sup> tation will prove to be moving forward. Neither the LI/RF multivariate techniques nor combina-  
<sup>1931</sup> tion with Run 1 datasets and their accompanying low signal strength values are necessary for discov-  
<sup>1932</sup> ery, and the latter may not even be essential to timely<sup>\*</sup> discovery of SM  $VH(b\bar{b})$ . Nevertheless, both  
<sup>1933</sup> sets of results hold great potential as key parts of a concerted ensemble of efforts towards precision  
<sup>1934</sup> Higgs physics.

<sup>1935</sup> With the final major center of mass energy increase at the energy frontier for the foreseeable fu-  
<sup>1936</sup> ture behind us, it is becoming increasingly likely that any new fundamental physics at colliders will  
<sup>1937</sup> require the use of results of systematics limited analyses. This is the regime where the techniques  
<sup>1938</sup> described in this dissertation will be most useful.

<sup>1939</sup> As the LHC and its experiments undergo successive stages of upgrades and operate in evermore  
<sup>1940</sup> extreme environments, the statistical fit models used to describe LHC data will continue to evolve in  
<sup>1941</sup> complexity and diverge from their predecessors. The techniques described in Chapter 9 will become  
<sup>1942</sup> increasingly more vital to producing the best physics results possible. The improvement in precision  
<sup>1943</sup> from  $1.20^{+0.24}_{-0.23}$  (stat.) $^{+0.34}_{-0.28}$  (syst.) to  $0.90^{+0.18}_{-0.18}$  (stat.) $^{+0.21}_{-0.19}$  (syst.) is just the beginning.

<sup>1944</sup> The best methods for reduction of systematic uncertainties will naturally depend in part on the  
<sup>1945</sup> state of the art for both fundamental physics process and detector modeling, but techniques that  
<sup>1946</sup> can reduce systematic uncertainties independent of fit model, dataset, and physics process provide a  
<sup>1947</sup> promising avenue forward. The improvements in systematic uncertainties using the Lorentz Invari-

---

\*i.e. before or coincident with CMS

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

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

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

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

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

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

Conventional Wisdom

A

1960

## 1961 Micromegas Trigger Processor Simulation

1962 IN ORDER TO PRESERVE key physics functionality by maintaining the ability to trigger on low  $p_T$   
1963 muons, the Phase I Upgrade to ATLAS includes a New Small Wheel (NSW) that will supply muon  
1964 track segments to the Level 1 trigger. These NSW trigger segments will combine segments from the  
1965 sTGC and Micromegas (MM) trigger processors (TP). This note will focus in particular on the algo-

<sup>1966</sup> rithm for the MMTP, described in detail with initial studies in<sup>33</sup>. The goal of this note is to describe  
<sup>1967</sup> the MMTP algorithm performance under a variety of algorithm settings with both nominal and  
<sup>1968</sup> misaligned chamber positions, as well as addressing a number of performance issues.

<sup>1969</sup> This note is organized as follows: the algorithm and its outputs are briefly described in Section  
<sup>1970</sup> A.1; Monte Carlo samples used are in Section B.1; nominal algorithm performance and certain quan-  
<sup>1971</sup> tities of interest are described in Section A.3; algorithm performance under misalignment, misalign-  
<sup>1972</sup> ment corrections, and corrected performance are shown in Section A.9; and conclusions are pre-  
<sup>1973</sup> sented in Section A.24.

## <sup>1974</sup> A.1 ALGORITHM OVERVIEW

<sup>1975</sup> The MMTP algorithm is shown schematically in Figure A.1, taken from<sup>33</sup>, where a more detailed  
<sup>1976</sup> description may be found. The algorithm begins by reading in hits, which are converted to slopes.  
<sup>1977</sup> These slopes are calculated under the assumption that the hit originates from the IP; slopes calcu-  
<sup>1978</sup> lated under this assumption are denoted by a superscript  $g$  for global in order to distinguish them  
<sup>1979</sup> from local slopes calculated using only hits in the wedge. In the algorithm simulation, events are  
<sup>1980</sup> screened at truth level to make sure they pass certain requirements. The track's truth-level coor-  
<sup>1981</sup> dinates must place it with the wedge since some generated tracks do not reach the wedge. These  
<sup>1982</sup> hits are stored in a buffer two bunch crossings (BCs) in time deep that separates the wedge into  
<sup>1983</sup> so-called "slope-roads." If any given slope-road has sufficient hits to pass what is known as a coin-  
<sup>1984</sup> cidence threshold, a fit proceeds. A coincidence threshold is a requirement for an event expressed as  
<sup>1985</sup>  $aX+bUV$ , which means that an slope-road must have at least  $a$  hits in horizontal (X) planes and at

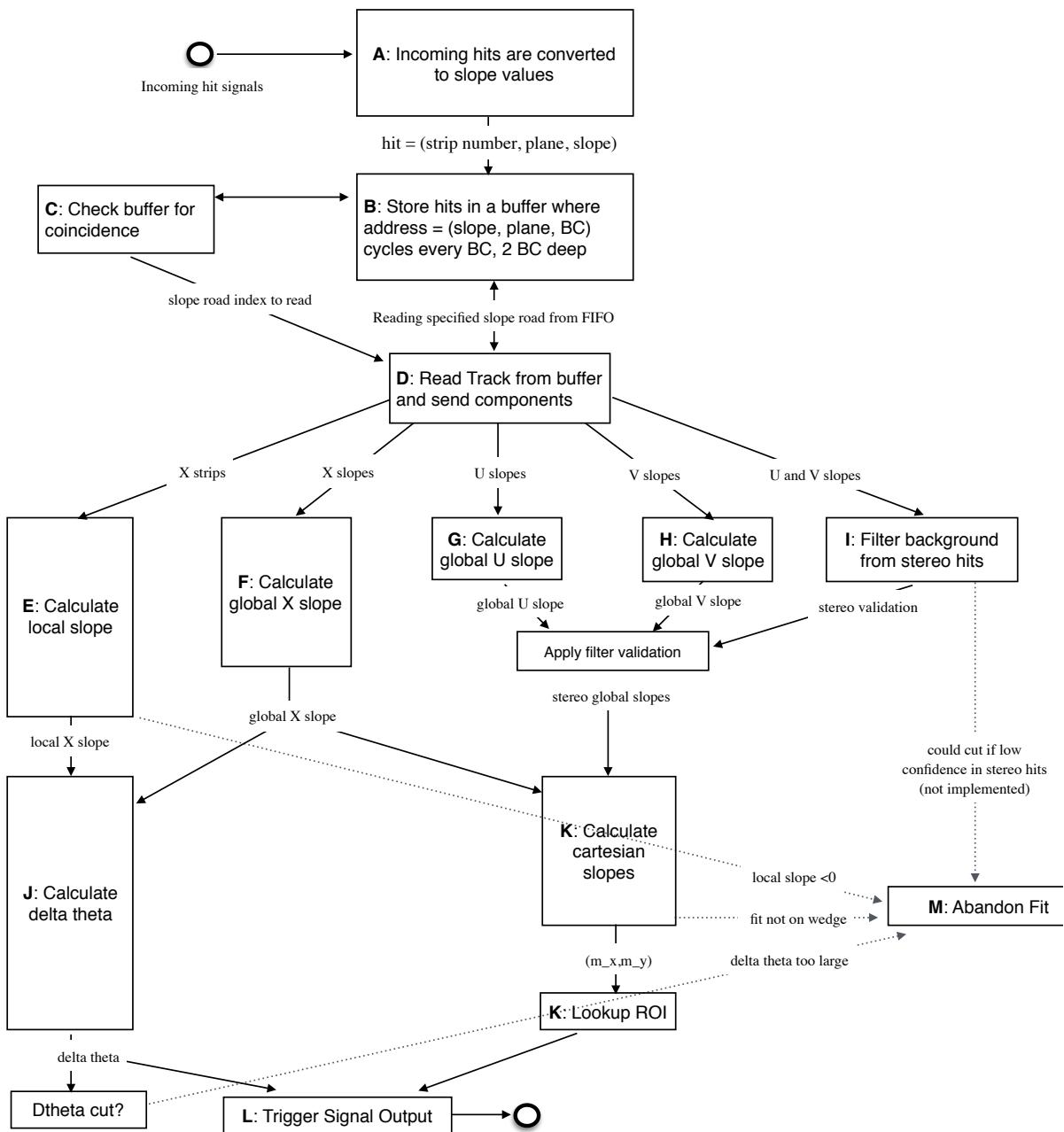


Figure A.1: A flow chart describing the algorithm steps, taken from <sup>33</sup>.

1986 least  $b$  hits in stereo (U or V (corresponding to positive and negative stereo rotations)) planes. For  
 1987 coincidence thresholds with a  $2X$  hit requirement there is the extra requirement that, in the case of  
 1988 only  $2X$  hits, one be on each quadruplet in order to ensure an adequate lever arm for the  $\Delta\theta$  calcu-  
 1989 lation. Note that less stringent (lower hit) coincidence thresholds are inclusive; i.e. a slope-road pass-  
 1990 ing a  $4X+4UV$  cut automatically passes  $2X+1UV$ . The coincidence threshold, size of the slope-roads  
 1991 (denoted  $b$ ), and the number of slope-roads into which each horizontal and stereo hits get written  
 1992 centered upon their nominal value are configurable parameters of the algorithm.

1993 An individual hit's slope is calculated as shown in Equation A.1, where  $y_{base}$  is the local  $y$  coordi-  
 1994 nate (orthogonal to the beamline and direction of the horizontal strips) of a station's base,  $w_{str}$  is the  
 1995 strip pitch,  $n_{str}$  is the hit's strip number, and  $z_{plane}$  is the location of the hit's plane along the beam-  
 1996 line.

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

1997 In the fit, individual hit slopes in a slope-road are used to calculate global slopes associated with each  
 1998 plane type, which are averages (e.g.  $M_X^{\ell}$  for the average slope of horizontal planes). These in turn are  
 1999 used to calculate the three composite slopes: slopes associated with the horizontal ( $m_x$ ) and vertical  
 2000 coordinates ( $m_y$ ) and the local slope of hits in the horizontal planes ( $M_X^l$ ), all of which are shown in  
 2001 Equation A.4. Note that the expression for  $M_X^l$  differs but is equivalent to the expression given in <sup>33</sup>.  
 2002 This is due to a procedural change in the algorithm. The local X slope is expressed in <sup>33</sup> as:

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

2003 Procedurally, this entails doing the sums over  $y_i$  and  $y_i z_i$ , multiplying the sums by  $A_k$ ,  $B_k$ , and then  
 2004 subtracting both of these numbers,  $\mathcal{O}(10^3)$ , to get local slopes,  $\mathcal{O}(10^{-1})$ , while requiring precision  
 2005 on these numbers on the order of  $\mathcal{O}(10^{-3})$ . This requires precision in the sums  $\mathcal{O}(10^{-7})$ , and with  
 2006 32 bit fixed point numbers, there are deviations with respect to the floating point calculations at the  
 2007 level of  $\mathcal{O}(10^{-5})$ , which is enough to introduce a significant bias in the  $\Delta\theta$  calculation.

2008 In order to prevent these errors, we do the subtraction first

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

2009 Thus, we change the order of operations and store  $1/\bar{z}$  instead of  $A_k$  in addition to  $B_k$ . We also  
 2010 change the units of  $y_i$  and  $z_i$  in the calculation by dividing the millimeter lengths by 8192.\* With  
 2011 these changes, a 32 bit fixed point based algorithm has essentially identical performance to that of an  
 2012 algorithm based on the usual C++ 32 floating point numbers. Future work includes converting the  
 2013 32 bit fixed point arithmetic to 16 bit where possible in the algorithm. While introducing 16 bit num-  
 2014 bers uniformly might seem preferable, since simple 16-bit operations in the firmware can be done in  
 2015 a single clock tick, and a larger number of bits increases the algorithm latency, some numbers in the  
 2016 algorithm will require a larger number of bits, in particular in the local slope calculation, which is  
 2017 the single calculation in the algorithm requiring the largest numeric range.

2018 In Equation A.4,  $\theta_{st}$  is the stereo angle of 1.5 degrees; the sums are over relevant planes;  $\bar{z}$  is the  
 2019 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

2020 and  $z$  coordinates of hits in X planes.

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

2021 From these composite slopes, the familiar expressions for the fit quantities  $\theta$  (the zenith),  $\phi$  (the az-  
2022 imuth<sup>†</sup>), and  $\Delta\theta$  (the difference in  $\theta$  between the direction of the segment extrapolated back to the  
2023 interaction point and its direction when entering the detector region; the following is an approxima-  
2024 tion) may be calculated, as noted in<sup>33</sup>:

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

2025 Looking at Equations A.4 and A.5, the dependence of fit quantities on input hit information be-  
2026 comes clear.  $\Delta\theta$  relies exclusively on information from the horizontal (X) planes. Both  $\theta$  and  $\phi$  rely  
2027 on both horizontal and stereo slope information. However, the sum in quadrature of  $m_x$  and  $m_y$  in  
2028 the arctangent for  $\theta$  means that  $\theta$  is less sensitive to errors in stereo hit information than  $\phi$ . Given  
2029 that  $\theta_{st}$  is small,  $\cot \theta_{st}$  is large ( $\sim 38$ ), so  $m_x$  multiplies small differences in  $\mathcal{M}_U$  and  $\mathcal{M}_V$ , where  $m_y$   
2030 is simply an average over slopes. This means that while errors in horizontal hit information will af-  
2031 fect all three fit quantities, comparable errors in stereo hits will have a proportionately larger effect  
2032 on  $\theta$  and particularly on  $\phi$ . The  $\Delta\theta$  cut after step J in Figure A.1 has been implemented, requiring  
2033 all fits to have  $|\Delta\theta| < 16$  mrad. This requirement ensures good quality fits but also slightly reduces

---

<sup>†</sup>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  $\phi = 0$

2034 algorithm efficiency.

2035 A.2 MONTE CARLO SAMPLES

2036 The Monte Carlo (MC) samples used for these studies were generated in Athena release 20.1.0.2 us-  
2037 ing simulation layout ATLAS-R2-2015-01-01-00 with muon GeoModel override version MuonSpectrometer-  
2038 R.07.00-NSW and modifications to have two modules per multiplet and xxuvuvxx geometry with a  
2039 stereo angle of 1.5 degrees. Muons of a single  $p_T$  were generated around the nominal IP with a smear-  
2040 ing of 50 mm along the beam line and 0.015 mm orthogonal to it; these muons were pointed toward  
2041 a single, large sector of the NSW. Each event consists of one muon fired towards the single NSW  
2042 wedge separated by effectively infinite time from other events.

2043 A.3 NOMINAL PERFORMANCE

2044 In order to evaluate algorithm performance, a number of quantities are evaluated, including the fit  
2045 quantities  $\theta$ ,  $\phi$ , and  $\Delta\theta$  as well as algorithm efficiency. Unless otherwise stated, that algorithm is run  
2046 with a 4X+4UV coincidence threshold, slope-road size of 0.0009, an X tolerance of two slope-roads  
2047 (i.e. hits in horizontal planes are written into the two slope-roads closest to the hits' value), a UV  
2048 tolerance of four slope-roads<sup>‡</sup>, and a charge threshold requirement on hits of 1 (measured in units  
2049 of electron charge) for a sample of 30 000 events with a muon  $p_T$  of 100 GeV. Samples were also  
2050 generated for  $p_T$  values of 10 GeV, 20 GeV, 30 GeV, 50 GeV, and 200 GeV, which were used in some

---

<sup>‡</sup>The larger tolerance on stereo hits takes into account the particulars of the  $m_x$  calculation mentioned in Section A.1.

2051 of the following studies.

2052 A.4 FIT QUANTITIES

2053 In order to evaluate the performance of the algorithm’s fit quantities  $\theta$ ,  $\phi$ , and  $\Delta\theta$ , fit values are com-  
2054 pared to truth-level MC values. The residual of the three fit quantities,  $\theta_{fit} - \theta_{tru}$ ,  $\phi_{fit} - \phi_{tru}$ , and  
2055  $\Delta\theta_{fit} - \Delta\theta_{tru}$ , are recorded for every fitted track. The distributions of these quantities, in particular  
2056 their biases and standard deviations, are then used to evaluate performance. In most cases, follow-  
2057 ing<sup>33</sup>, the mean and standard deviation of a  $3\sigma$  Gaussian fit are quoted, as they capture the main  
2058 features of the algorithm and generally behave like the raw mean and rms. Nevertheless, discussion  
2059 of the raw quantities will be included when their behavior deviates markedly from that of the  $3\sigma$  fit  
2060 quantities.

2061 The truth-level quantities used in residual distribution are taken from information in the MC.

2062 These come directly from the MC for  $\theta$ ,  $\phi$ , and  $\Delta\theta$ . These quantities, along with the geometry of  
2063 the (large) wedge, are then in turn used to calculate truth-level values for any intermediate quantities  
2064 used in the algorithm.  $m_{x,tru}$ , for instance, is given by  $\tan \theta_{tru} \sin \phi_{tru}$ .

2065 Residual distributions for fit quantities under the previously described default settings of the al-  
2066 gorithm are shown in Figure A.2. Both the  $\theta_{fit} - \theta_{tru}$  and  $\Delta\theta_{fit} - \Delta\theta_{tru}$  distributions feature a  
2067 mostly Gaussian shape with more pronounced tails. The mean bias for these distributions is negligi-  
2068 ble at under one tenth of a milliradian, and the fitted (raw) rms values are 0.349 (0.614) mrad for  $\theta$   
2069 and 1.03 (2.55) mrad for  $\Delta\theta$ . The case of the  $\phi_{fit} - \phi_{tru}$  distribution is less straightforward, with both  
2070 the shape and bias arising from the xxuvuvxx geometry and relatively large extent of one of the two

2071  $\eta$ -stations, as explained in Appendix B of<sup>30</sup>. The fitted (raw) rms for the  $\phi$  distribution is 8.67 (16.6)  
 2072 mrad.

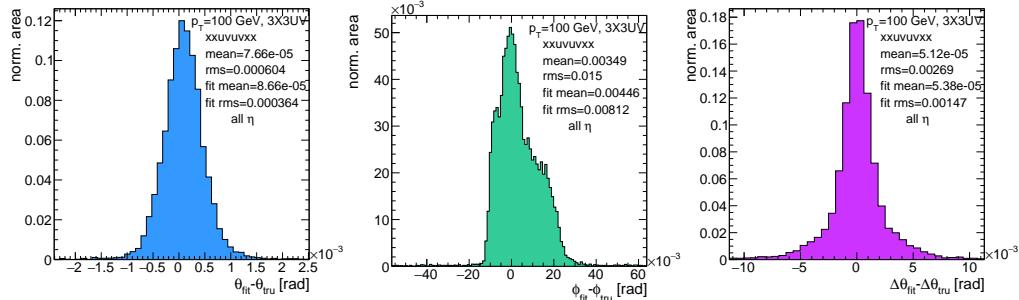
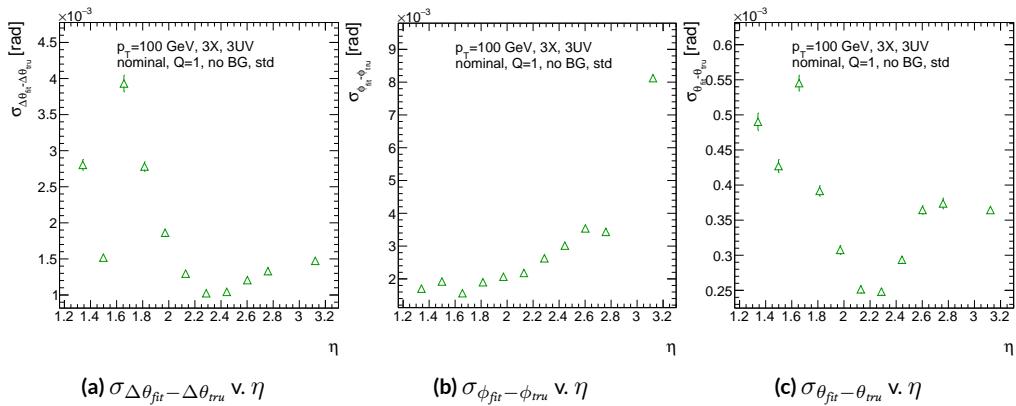


Figure A.2: Nominal residual plots;  $\theta, \phi, \Delta\theta$  for  $p_T = 100 \text{ GeV}$  muons

2073 Both increasing muon  $p_T$  and increasing muon  $\eta$  for a fixed  $p_T$  imply increasing muon energy. As  
 2074 muons become more energetic, two effects compete in affecting the quality of fit. On the one hand,  
 2075 higher energy muons are deflected less by the ATLAS magnetic field, which should tend to improve  
 2076 the quality of the fit, since the fitted  $\theta$  (upon which  $\Delta\theta$  also relies) and  $\phi$  values are calculated under  
 2077 the infinite momentum muon (straight track) assumption. However, as muon energy increases, the  
 2078 likelihood that the muon will create additional secondaries increases, which creates extra hits that  
 2079 degrade the quality of the fit. While the geometry of the multiplet is such that there is very good res-  
 2080 olution in the direction orthogonal to the horizontal strip direction, the shallow stereo angle of 1.5  
 2081 degrees means that early hits caused by secondaries can have an outsize impact on  $m_x$ .  $\Delta\theta$ , which  
 2082 does not rely upon stereo information should feel the effect of secondaries the least and benefit from  
 2083 straighter tracks the most and hence benefit from higher muon energies;  $\phi$ , relying upon stereo in-  
 2084 formation the most, would be most susceptible to secondaries and benefit the least from straighter

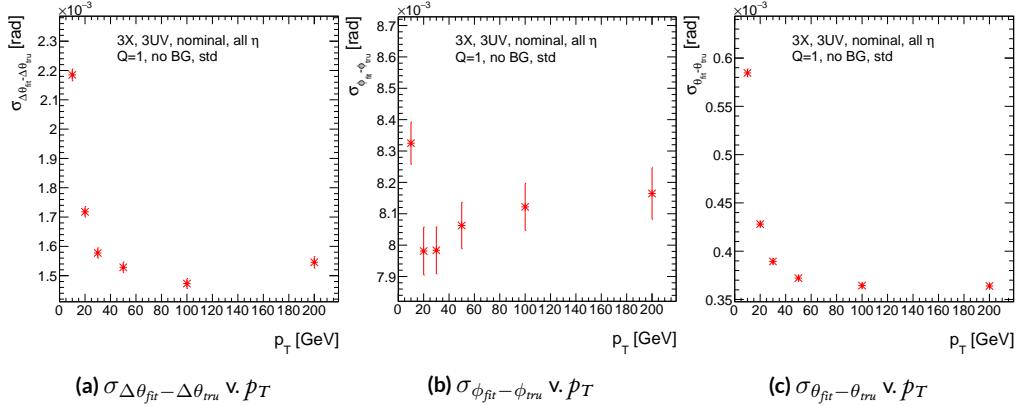
2085 tracks and hence least likely to benefit from higher muon energy;  $\theta$  relies upon both horizontal and  
 2086 vertical slope information, though small errors are less likely to seriously affect the calculation, so the  
 2087 two effects are most likely to be in conflict for this fit quantity.

2088 The interplay of these effects on the residual standard deviations can be seen in their dependen-  
 2089 cies on  $\eta$  (Figure A.3; note that the final point in each of these plots is the rms of the distribution  
 2090 overall  $\eta$ ) and  $p_T$  (Figure A.4). For  $p_T = 100$  GeV muons,  $\Delta\theta$  performance increases with  $\eta$  (energy),  
 2091 and  $\phi$  performance decreases, as expected;<sup>§</sup> for  $\theta$ , the two effects appear to compete, with perfor-  
 2092 mance first increasing with  $\eta$  until the effects of secondaries begins to dominate. Integrated over all  
 2093  $\eta$ , the effects are less clearly delineated. Both  $\Delta\theta$  and  $\theta$  performance increases with increasing  $p_T$ ,  
 2094 suggesting straighter tracks with increasing energy are the dominant effect for these quantities, while  
 2095  $\phi$  performance appears to improve and then deteriorate (the slight improvement at high  $p_T$  is due to  
 2096 the addition of the  $\Delta\theta$  cut into the algorithm, which filters out very poor quality fits).



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

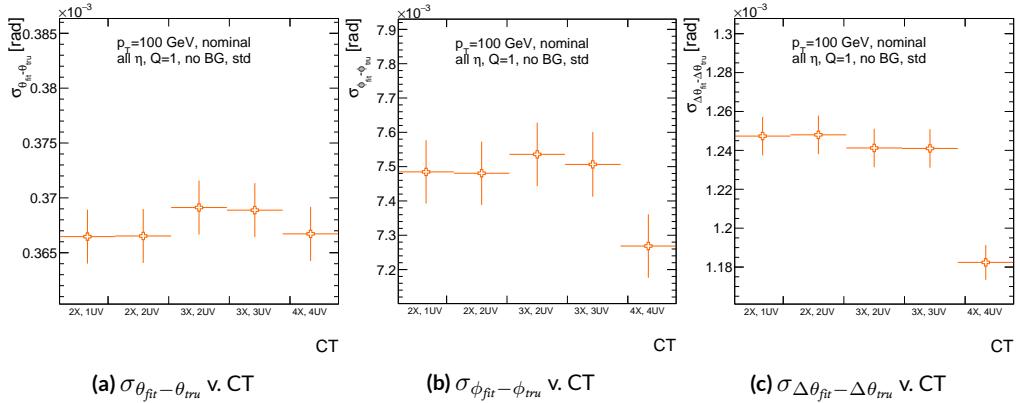
§The much worse overall performance for  $\phi$  is due to the  $\eta$  dependent bias and other effects



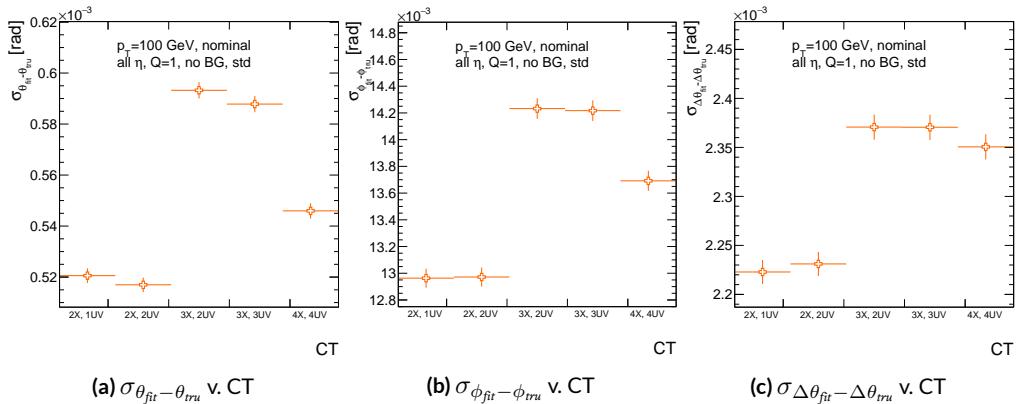
**Figure A.4:** The rms distributions of  $\Delta\theta$ ,  $\phi$ , and  $\theta$  as a function of  $p_T$ .

2097 The rms of the three benchmark quantities as a function of algorithm set (i.e. slope-road) coinci-  
 2098     dence threshold are shown in Figure A.5 using Gaussian fits and in Figure A.6 for the raw quantities.  
 2099     The fitted  $\sigma$ 's for  $\theta$  and  $\phi$  are fairly stable across coincidence threshold.  $\Delta\theta$ , on the other hand, per-  
 2100     forms better particularly for the most stringent coincidence threshold; this is a result of the fact that  
 2101     additional information for more hits greatly improves the quality of the local slope fit calculation.  
 2102     The raw rms is a different story. Naïvely, one would expect the performance to get better with more  
 2103     stringent coincidence threshold, but this is not the case in Figure A.6. As the coincidence thresh-  
 2104     old gets more stringent, fewer and fewer tracks are allowed to be fit. When moving from 2X hits to  
 2105     3X hits, the tracks that get vetoed populate the tails of the distribution outside the  $3\sigma$  fit range but  
 2106     are not in the very extremes of the distribution. While tracks with 2X hits are of lower quality than  
 2107     those with 3 and 4 X hits, tracks with the very worst fit values pass even the most stringent coinci-  
 2108     dence threshold requirements (e.g. as a result of many hits arising from a shower of secondaries).  
 2109     This is best illustrated when comparing the 2X+1UV  $\Delta\theta$  residual distribution with the 4X+4UV

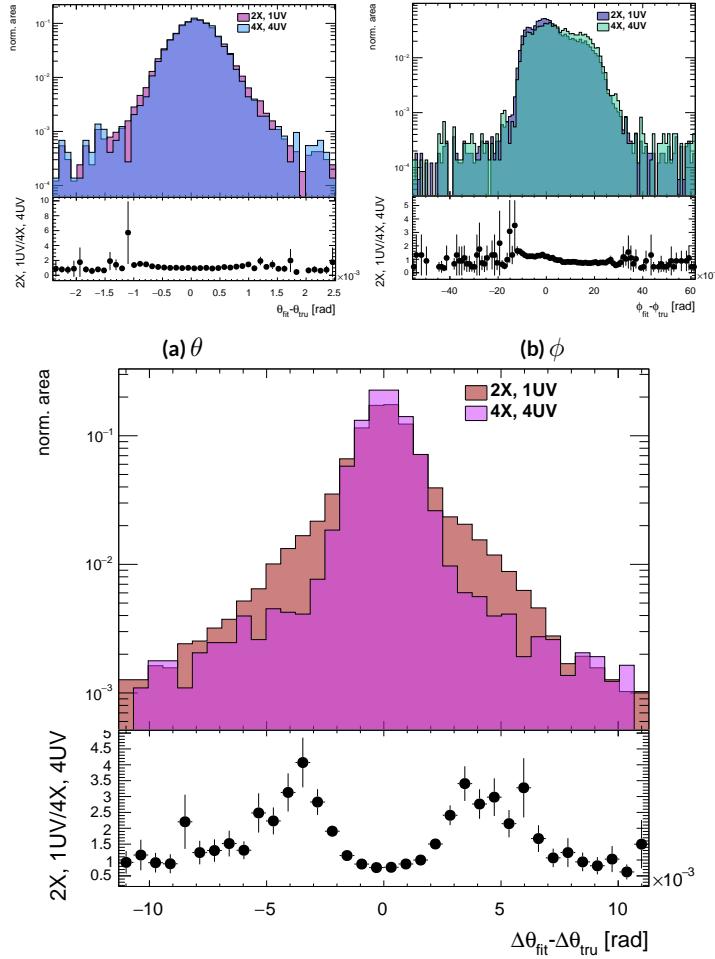
distribution in Figure A.7. As both the overlayed normalized curves and ratio distribution show,  
 while the most central regions are fairly similar, the  $\omega X + 1$  UV distribution is much more prominent  
 in the tails but not the extreme tails, which means that, though the overall  $\omega X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



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



**Figure A.6:** The raw rms of residual distributions for  $\theta$ ,  $\phi$ , 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.

2114 A.5 EFFICIENCIES

2115 Two general efficiencies have been formulated to study the performance of the MMTP algorithm.

2116 The first, denoted  $\varepsilon_{alg}$ , is the fraction of tracks that pass some (slope-road) coincidence threshold

2117 configuration that are successfully fit. An event that passes a slope-road coincidence but does not fit

2118 fails because some of the hits included are of sufficiently poor quality to throw off the fit. This effi-

2119 ciency answers the question of how often the algorithm performs fits when technically possible, giv-

2120 ing a measure of overall algorithm performance for a given configuration. For example,  $\varepsilon = 95\%$  for

2121  $3X+2UV$  means that 95% of tracks that produce at least  $3X$  hits and  $2UV$  hits in at least one slope-

2122 road will be successfully fitted 95% of the time. The performance of this efficiency as a function of

2123 coincidence threshold,  $\eta$  (with the final point once again being the efficiency integrated over all  $\eta$ ),

2124 and  $p_T$  is shown in Figure A.8.  $\varepsilon_{alg}$  is fairly constant in  $\eta$  and decreases with increased  $p_T$ , which can

2125 be attributed to the increased likelihood of secondaries introducing lower quality hits that cause the

2126 fit to fail.

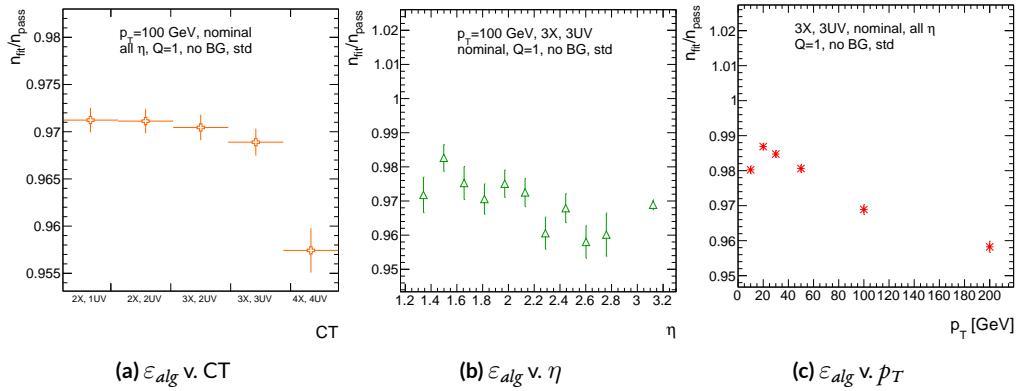
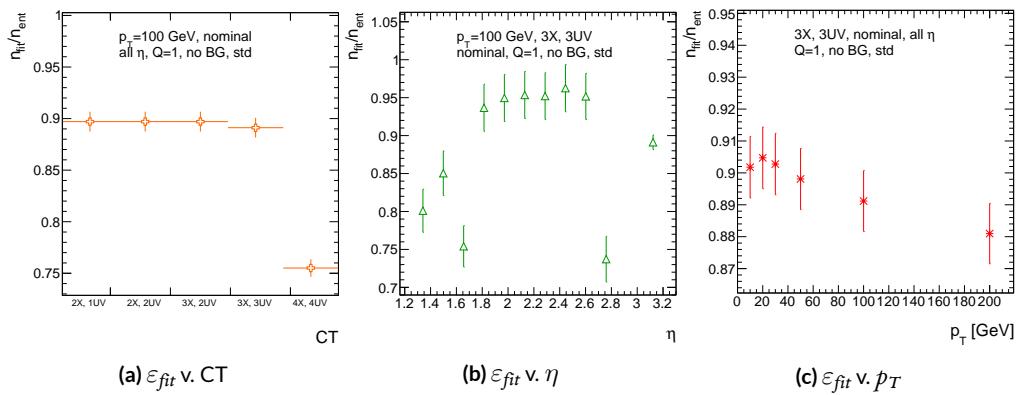


Figure A.8:  $\varepsilon_{alg}$  and as a function of coincidence threshold,  $\eta$  (final point is  $\varepsilon_{alg}$  integrated over all  $\eta$ ), and  $p_T$ .

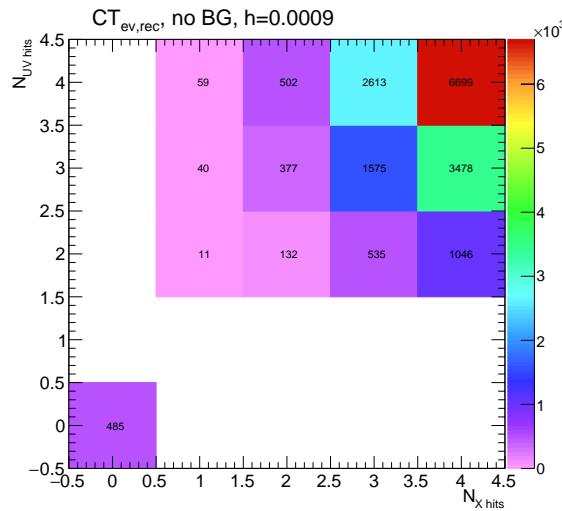
2127        The second efficiency type, denoted  $\varepsilon_{fit}$ , is the fraction of tracks that enter the wedge whose fits  
 2128        (if any) satisfy a given coincidence threshold. This efficiency can be used to help establish an optimal  
 2129        coincidence threshold setting in the algorithm, balancing the improved overall fit quality of higher  
 2130        thresholds with the greater number of fits for lower thresholds. Hence, an  $\varepsilon_{fit}$  of 95% at 3X+2UV  
 2131        means that 95% of tracks entering the wedge are fit and that these fits include at least 3X and 2UV  
 2132        hits.  $\varepsilon_{fit}$  as a function of coincidence threshold is shown in Figure A.9 (a), which shows that the  
 2133        majority of fits having at most 3X+3UV hits. That there is a marked drop to 4X+4UV is not sur-  
 2134        prising, as there is a substantial population outside the 4X+4UV bin in Figure A.10. The behavior  
 2135        of  $\varepsilon_{fit}$  with  $\eta$  in Figure A.9 (b) (with the final point once again being the efficiency integrated over  
 2136        all  $\eta$ ) is much more varied, with geometric effects of detector acceptance coming into play. The per-  
 2137        formance of  $\varepsilon_{fit}$  as a function of  $p_T$ , shown in Figure A.9 (c), is similar to that of  $\varepsilon_{alg}$  coincidence  
 2138        threshold, again consistent with the effects of secondaries at higher energies.



**Figure A.9:**  $\varepsilon_{fit}$  and as a function of coincidence threshold,  $\eta$  (final point is  $\varepsilon_{fit}$  integrated over all  $\eta$ ), and  $p_T$ .

2139        In order to better understand efficiency behavior with coincidence threshold, the distribution

2140 of highest slope-road coincidence thresholds in events is shown in Figure A.10, with the 0,0 bin  
 2141 containing events that did not meet requirements for the minimum  $2X+1UV$  coincidence threshold  
 2142 for a fit to occur. That the efficiency is lower at higher coincidence threshold suggests that most  
 2143 of the fits that fail have high hit multiplicity (i.e. a similar number fails in each of the coincidence  
 2144 threshold bins in Figure A.8 (a)), which is consistent with the interpretation that the primary source  
 2145 of fit failures is bad hits originating from secondaries created by higher energy muons.



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

2146 A.6 INCOHERENT BACKGROUND

2147 The default slope-road size and tolerances associated with horizontal and stereo hits used in the  
2148 above studies were configured to optimize algorithm performance, similar to studies in<sup>33</sup>. In order to  
2149 evaluate algorithm performance under conditions with more limited resources, as might be expected  
2150 at run-time, additional studies were conducted with the slope-road size and hit tolerances set equiv-  
2151 alent to the sensitive area of a single VMM chip<sup>¶</sup> both with and without generation of incoherent  
2152 background.

2153 Incoherent background is generated based on the assumption that the intensity only varies as a  
2154 function of the distance from a point to the beamline,  $r$ . The number of hits per unit area per unit  
2155 time as a function of  $r$  is given in Equation A.6 and taken from<sup>33</sup>.

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

2156 where  $r_o = 1000$  mm and  $I_o = 0.141$  kHz/mm<sup>2</sup>

2157 Background generation happens per event as follows:

- 2158 1. Determine the total number of hits to be generated in this event according to a Poisson distri-  
2159 bution
- 2160 2. Assign a time to hits uniformly in  $[t_{start} - t_{VMM}, t_{end}]$  where start and end are for the event  
2161 clock and  $t_{VMM}$  is the VMM chip deadtime (100 ns)
- 2162 3. Assign a plane to hits uniformly
- 2163 4. Assign a  $\phi$  value to hits uniformly

---

<sup>¶</sup>One VMM is assumed to cover 64 MM strips at 0.445 mm each.

2164 5. Assign an  $r$  to hits according to Equation A.6

2165 6. Calculate hit information according to these values.

2166 The expectation value for the Poisson distribution is determined by integrating Equation A.6

2167 over the surface area of the wedge to get the total hit rate for the wedge,  $\Gamma$ , and then multiplying this

2168 by the length of the time window over which hits may be generated. With  $H = 982$  mm,  $b_t = 3665$

2169 mm, and  $\theta_w = 33\pi/180$ , we find<sup>11</sup>:

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

2170 In this case, we have taken the nominal values of the MM sector geometry for  $H$  (wedge base),  $b_t$

2171 (the wedge height), and  $\theta_w$  (the wedge opening angle).

2172 The effects of incoherent background and larger slope road size are summarized in Figure A.11 for

2173 efficiencies and in Figure A.13 and Table A.1 for residual of fit quantities.

2174 Figure A.11 show the effect of both wider slope-roads and the introduction of background on ef-

2175 ficiencies. The introduction of wider slope-roads increases the chance that an early errant hit (either

2176 from secondaries/ionization or background) will be introduced into the fit, and the presence of in-

2177 coherent background greatly increases the number of such errant hits. Both wider slope-roads and

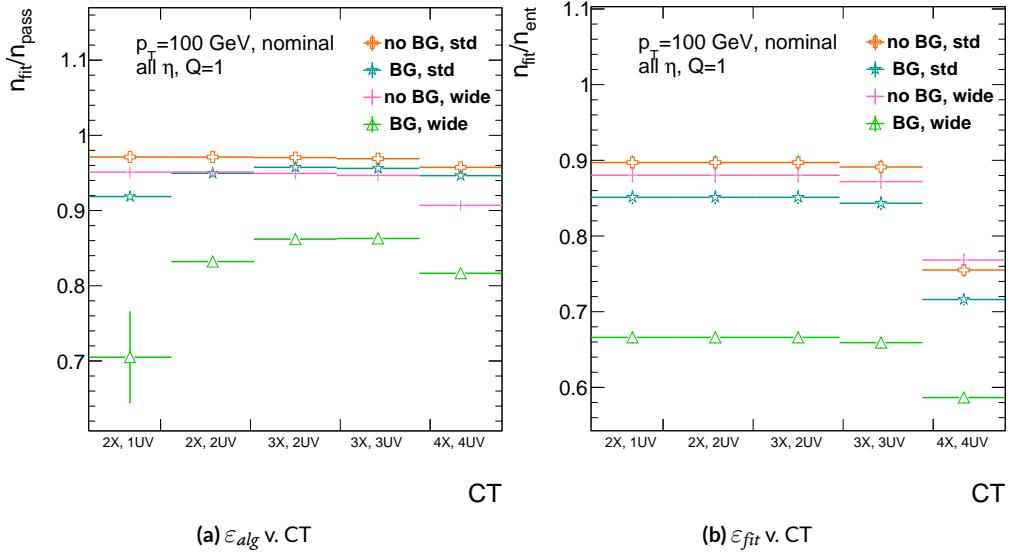
2178 background drive down the number of fits (numerator) in both efficiencies, and background can

2179 artificially inflate the denominator of  $\varepsilon_{alg}$ , a reco-level, slope-road coincidence threshold. The shape

2180 of the  $\varepsilon_{fit}$  versus coincidence threshold distributions remains fairly constant with each complicating

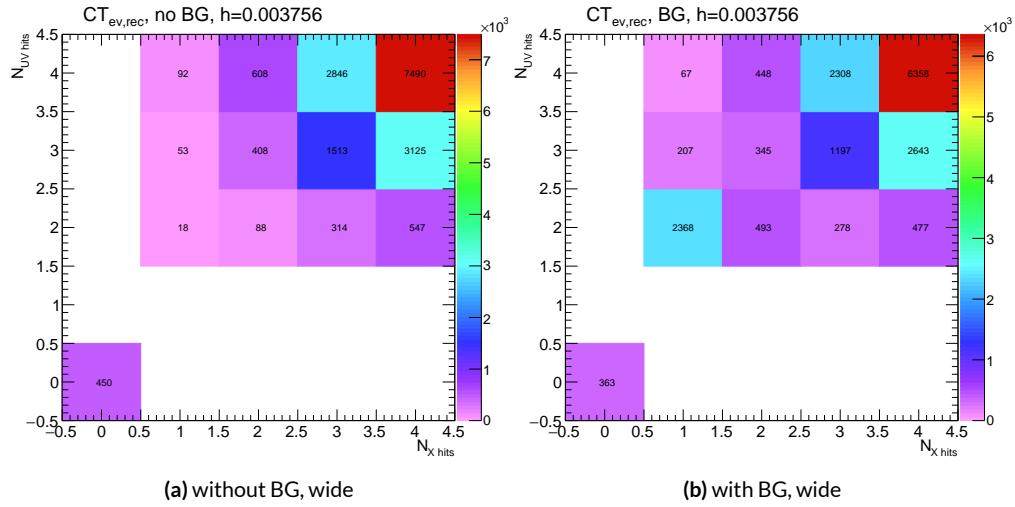
---

<sup>11</sup>Using Mathematica and the extra factor of  $r$  from the volume element



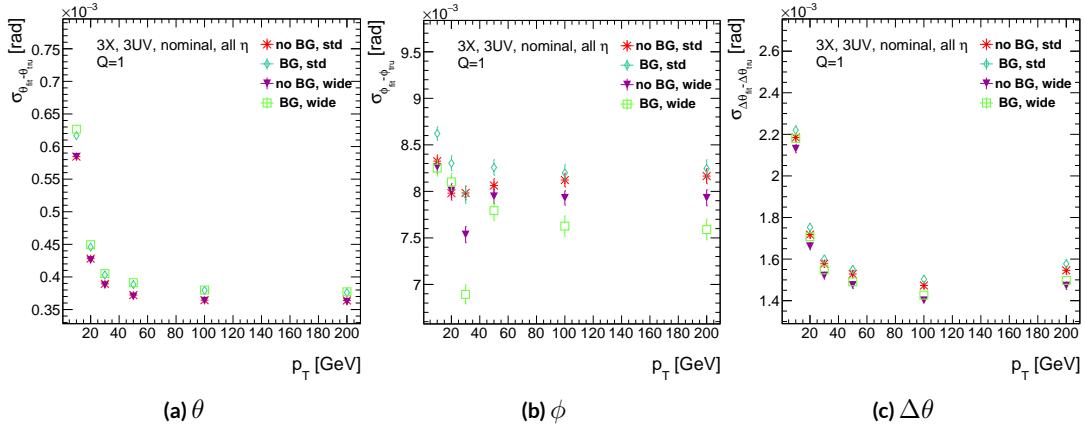
**Figure A.11:** The algorithm and total efficiencies as a function of coincidence threshold for different background settings and slope-road sizes (standard and wide (one slope road as 1 VMM chip)).

2181 factor (standard, wider slope-roads, background, both wider slope-roads and background), suggest-  
 2182 ing many muons will simply not be fit with any number of hits;  $\varepsilon_{\text{fit}}$  does not take into account the  
 2183 coincidence threshold of tracks that are not fit, so the effect appears uniform across coincidence  
 2184 threshold. The effects seen for  $\varepsilon_{\text{alg}}$ , which are not uniform across coincidence threshold can be bet-  
 2185 ter understood when examining the distribution of event highest coincidence thresholds, shown for  
 2186 wide slope-roads both without and with background in Figure A.12. Take, for example the 2X+1UV  
 2187 case. The 2X+1UV bin in particular has a marked increase when background is introduced. No  
 2188 new, good tracks are introduced between the no backgrond and background cases, so the increase is  
 2189 entirely due to bad, background hits; hence, these events do not (and should not) fit, causing the  
 2190 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),  $\theta$  shows some degradation in performance, and the  $\phi$  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
$\theta$	0.364 (0.604)	0.363 (0.542)	0.379 (0.886)	0.380 (1.07)
$\phi$	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)  $\sigma$  of fit quantity residuals in mrad under different algorithm settings.

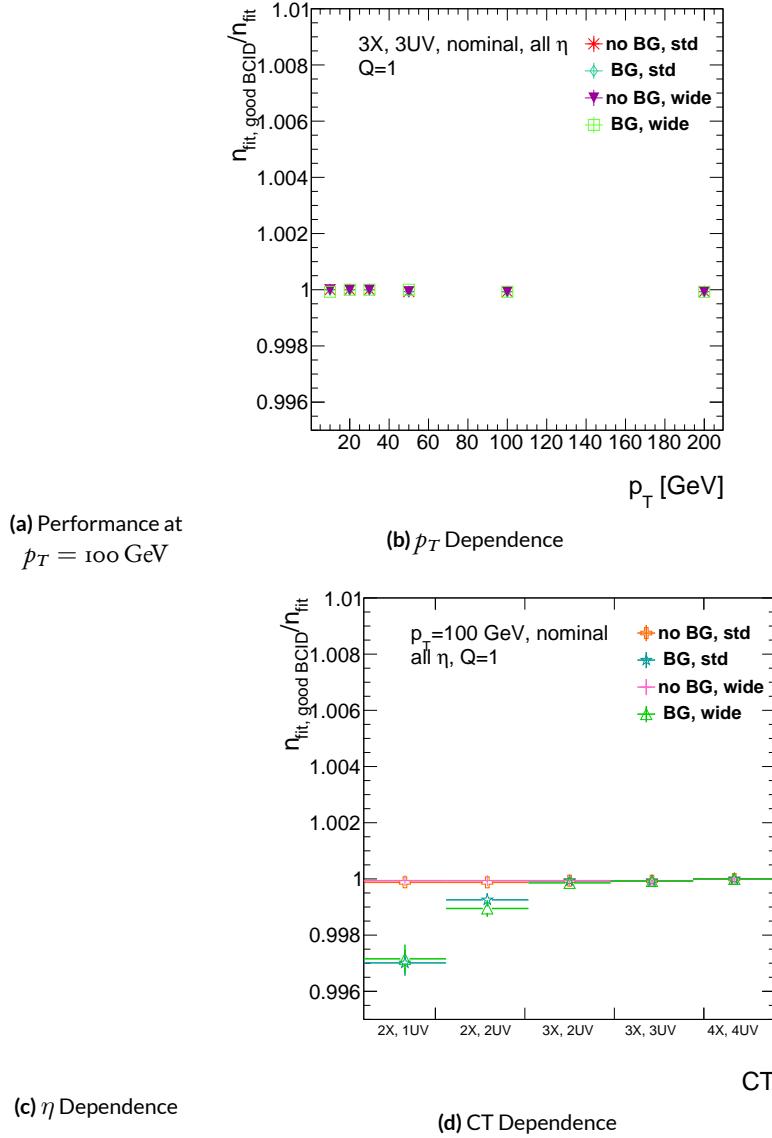
2202 As Table A.1 shows, rms values appear to be robust to an increase in slope-road size. Neverthe-  
2203 less, though the fitted  $\sigma$  residual values are also fairly robust to the introduction of background, the  
2204 raw rms values are not. While the raw  $\Delta\theta$  rms stays stable, both  $\theta$  and  $\phi$  suffer noticeable degra-  
2205 dation, which suggests that the introduction of background has a detrimental effect on horizontal  
2206 slope residual (i.e. on stereo strips in particular). This level of degradation is likely acceptable for  $\theta$ ,  
2207 though further steps may need to be taken to address  $\phi$ .

2208 **A.7 BCID**

2209 A fitted track's BCID is determined by the most common BCID associated with its hits. Concerns  
2210 were raised that this might cause incorrect BCID association for fitted tracks. In order to address this,  
2211 the rate of successful BCID association for fitted tracks was recorded. Figure A.14 shows the depen-  
2212 dence of this success rate as a function of  $p_T$  and coincidence threshold in the different background  
2213 and resource conditions used in the previous section. The successful BCID identification rate is al-  
2214 ways over 99.5%, demonstrating that this issue is not a concern with the state-of-the-art detector  
2215 simulation.

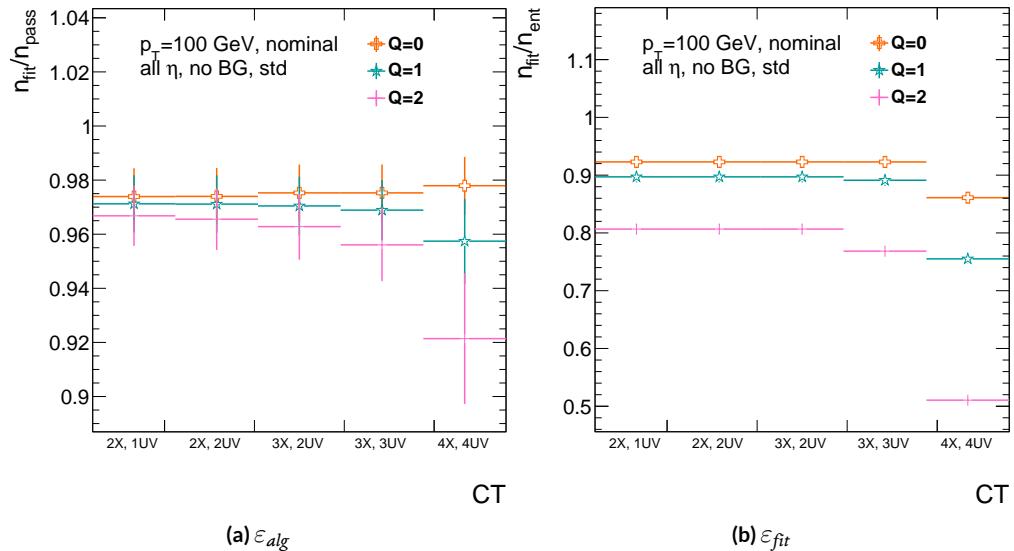
2216 **A.8 CHARGE THRESHOLD**

2217 The MMTP uses the first hits registered passing a charge threshold requirement given in units of  
2218 electron charge. In principle, it would be beneficial to be able to use any hits that are registered re-  
2219 gardless of deposited charge, but in the high rate environment envisioned for the NSW, this require-  
2220 ment might need to be raised. Nominal algorithm settings have this charge threshold requirement

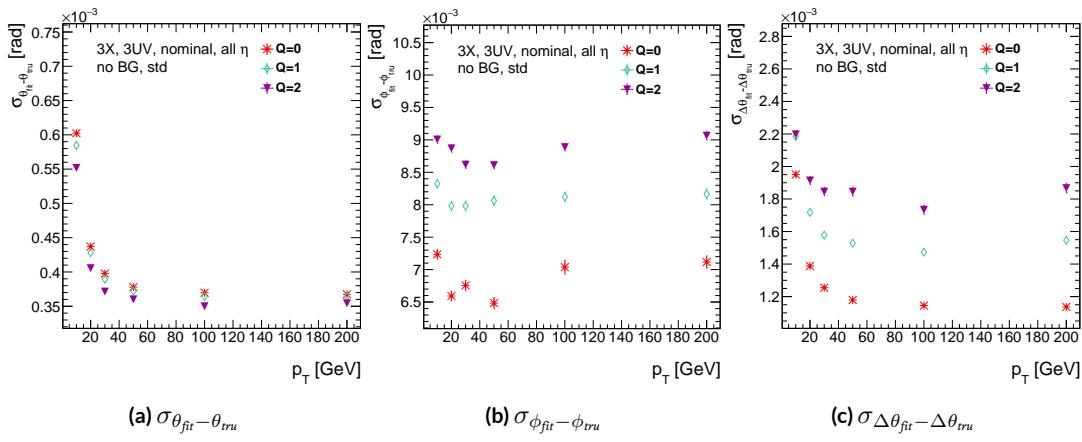


**Figure A.14:** The rate of good BCID association based majority hit BCID as a function of  $p_T$  and coincidence threshold.

2221 set to 1, and studies were conducted on algorithm performance for charge threshold values of 0, 1,  
 2222 and 2. Efficiencies as a function of coincidence threshold for different charge thresholds are shown  
 2223 in Figure A.15. Increasing the charge threshold lowers both efficiencies, particularly at high coinci-  
 2224 dence threshold, which suggests that energetic muons with secondaries create both very many hits  
 2225 and hits with higher charge. While the shapes of the fit quantity distributions as a function of  $p_T$  in  
 2226 Figure A.16 are fairly constant across charge threshold, performance is not.  $\theta$  and  $\Delta\theta$  show some im-  
 2227 provement with higher charge threshold, particularly at low  $p_T$ , suggesting that resolution improves  
 2228 in the vertical direction, but  $\phi$  shows degradation at higher charge threshold, which is a symptom  
 2229 of more highly charged particles experiencing greater bending in the ATLAS magnetic field in the  $\phi$   
 2230 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.

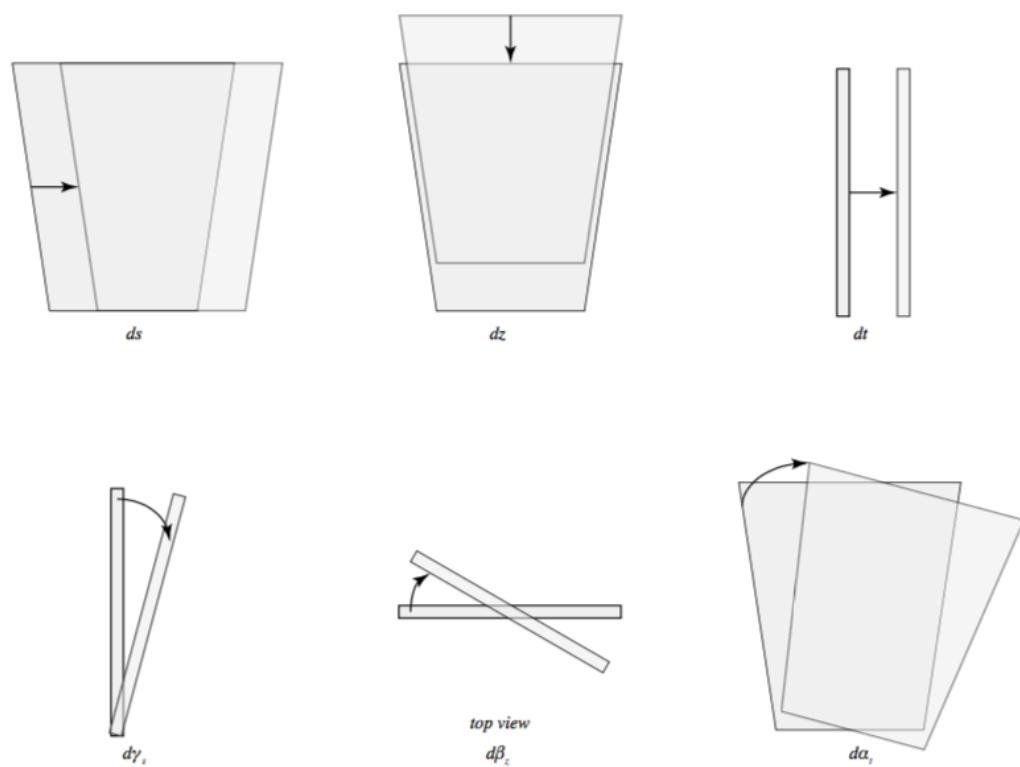
2231 A.9 MISALIGNMENTS AND CORRECTIONS

2232 The performance of the trigger algorithm under misalignment has been studied for each of the six  
2233 alignment quantities (three translations and three rotations all along the principal axes) described  
2234 in<sup>?</sup> and<sup>?</sup>, whose convention we will follow here. For the simulated wedge studied here the local co-  
2235 ordinates described in<sup>?</sup> are taken to be centered at the center of the base of the wedge<sup>\*\*</sup>, the local  $t$   
2236 axis corresponds to the axis of the beam line, the local  $z$  axis corresponds to the direction orthogo-  
2237 nal to both the beam line and the horizontal strips, and the local  $s$  axis completes the right-handed  
2238 coordinate system. The rotation angles  $\alpha$ ,  $\beta$ , and  $\gamma$  correspond to rotations around the local  $t$ ,  $z$ ,  
2239 and  $s$  axes, respectively. Note that the local  $s$ ,  $z$ , and  $-t$ , axes correspond to the usual global  $x$ ,  $y$ , and  
2240  $z$  axes. Misalignments were studied in twenty evenly spaced increments from nominal positions  
2241 to misalignments of 1.5 mrad for the rotations (-1.5 mrad to +1.5 mrad for the  $\gamma$  case), and of 5 mm  
2242 (a roughly corresponding linear shift) for the translations. In all cases, the front quadruplet is mis-  
2243 aligned while the rear quadruplet remains in its nominal position. While only the front quadruplet  
2244 of a single wedge is misaligned, the framework for misalignment presented below could be used to  
2245 study generic local and global misalignments. The six misalignments are schematically represented  
2246 in Figure A.17.

2247 Chamber misalignments manifest themselves as altered strips in algorithm input. In order to  
2248 simulate the effects of misalignment, the change in the local  $y$  coordinate—the distance from the

---

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



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

2249 bottom wedge center in the direction perpendicular to both the beamline and the strip direction—is  
 2250 calculated for a track coming straight from the interaction point defined by the truth-level  $\theta$  and  $\phi$   
 2251 angles for generic misalignment. This displacement in  $y$  is then added to input hit information and  
 2252 the algorithm is then run normally.

2253 To understand how this displacement is calculated, some notation first needs to be described.

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

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

2254 Generically speaking, a hit is the intersection of a line (the muon track) with a plane (the individual plane in the multiplet). We assume the muon moves in a straight line defined by the origin and  
 2255 the truth-level  $\theta_{pos}$  and  $\phi_{pos}$  (i.e. the infinite momentum limit) and that the MM plane is rigid and  
 2256 defined by a point, which we take to be the center of the bottom edge of the plane, and a normal  
 2257 vector, which we take to be the  $z$  axis in the nominal case.

2259 The coordinate axes  $x, y, z$  axes used here correspond to the usual AMDB  $s, z, -t$  axes. Since  
 2260 the direction does not really matter when studying misalignment or corrections thereof, the major

2261 difference is the choice of origin.

2262 The muon track we denote<sup>††</sup>  $\vec{s}$ , the bottom point of the plane  $\vec{\mathcal{O}}_{base}$ , and the normal vector  $\hat{n}$ .

2263 The muon track will always be given as (the wedge gets moved, not the muon):

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

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

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

2264 where  $k \in \mathbb{R}$ , along with the unit vector  $\hat{s}$ , defines the point where the track intersects the wedge.

2265 Rotations are done before translations, according to the order prescribed in the AMDB guide for

2266 chamber alignment, so the axes the principal axes of the plane are rotated according to the following

2267 matrix (where  $s$ ,  $c$ , and  $t$  are the obvious trigonometric substitutions)

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

---

<sup>††</sup>Recall  $\phi_{pos}$  is defined with respect to the  $y$  axis instead of the  $x$  axis, as might otherwise be typical.

2268      The thing that matters is what the new strip hit is—i.e. what the new  $y$  value is since this, along  
 2269      with a plane number, is all that is fed into the algorithm. To find this, we must solve for the new  
 2270      point of intersection with the rotated plane and then apply the effects of translations. The path  
 2271      connecting the base of the wedge with the intersection of the muon track will always be orthogonal  
 2272      to the normal vector of the plane. Our quantities after misalignment, denoted by primed quantities,  
 2273      will look like

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

2274      so, moving to explicit, global coordinates in the last line so we can do the computation (relying on  
 2275      the fact that any vector in the wedge, namely  $\vec{\zeta} = \vec{s} - \mathcal{O}$  the local coordinates of the interaction  
 2276      point, is necessarily orthogonal to  $\hat{n}$ ):

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

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

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

2277      To find our new  $y$  coordinate, we need to evaluate  $s'_y = \hat{y}' \cdot k'\vec{s}$  to find the final correction of:

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

<sup>2278</sup> The correction will be plane dependent since (denoting the stereo angle  $\omega$ ):

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

<sup>2279</sup> and

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

<sup>2280</sup>

## <sup>2281</sup> A.10 INDIVIDUAL CASES

<sup>2282</sup> Currently we only study the cases where one misalignment parameter is not zero. We examine these  
<sup>2283</sup> in detail below, calculating the most pertinent quantities in the misalignment calculation,  $k'/k$  and  
<sup>2284</sup> the new horizontal and stereo  $y$  axes. Before setting out, we simplify the expressions for the trans-  
<sup>2285</sup> formed  $\hat{y}'$ 's, removing any terms with the product of two sines of misalignment angles, which will be  
<sup>2286</sup> zero.<sup>††</sup>

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

<sup>2287</sup>

---


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

<sup>††</sup>If only one misalignment parameter is non-zero, then two or more sines will contain at least one term will contain  $\sin \theta = 0$ .

<sup>2288</sup> If the translations are zero,

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

<sup>2289</sup> A.II  $ds \neq 0$

<sup>2290</sup>  $k'/k = 1$  (the point of intersection does not move closer or further from the IP), and only the stereo planes are affected. Note that only relevant term in Equation A.16, for the stereo strip  $\hat{y}$  for  $\vec{\mathcal{O}}'_{base} =$   
<sup>2291</sup>  $ds\hat{x}$  is:

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

<sup>2293</sup> meaning that a displacement in  $x$  of 17 mm, more than three times the range of misalignments studied, would be necessary for a shift in the stereo planes corresponding to one strip width.  
<sup>2294</sup>

<sup>2295</sup> A.12  $dz \neq 0$

<sup>2296</sup>  $k'/k = 1$  (the point of intersection does not move closer or further from the IP). This case is the trivial one (cf. Equation A.16 with  $\vec{\mathcal{O}}'_{base} = dz\hat{y}$ ).  $y$  just gets moved in the opposite direction as the wedge. Correction is an additive constant.  
<sup>2297</sup>  
<sup>2298</sup>

<sup>2299</sup> A.13  $dt \neq 0$

<sup>2300</sup>  $k'/k = (z_{pl} + dt) / z_{pl}$ .  $y$  gets modified by a simple scale factor. Correct by storing changing definitions of plane positions in algorithm to match the misaligned values.  
<sup>2301</sup>

2302 A.14  $\alpha \neq 0$

2303  $k'/k = 1$  and

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

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

2304 A.15  $\beta \neq 0$

2305 We have  $k'/k = (1 + \tan \beta \tan \theta \sin \phi)^{-1}$ , and

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

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

2306 A.16  $\gamma \neq 0$

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

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

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

2307 In order to evaluate algorithm performance under misalignment and corrections for misalignment, the absolute means and relative resolutions of the fit quantities  $\theta$ ,  $\phi$ , and  $\Delta\theta$  are measured

2309 as a function of misalignment. In the following, results will only be shown for which the effects of  
2310 misalignment are significant. “Significant,” for misalignments of 1 mm (0.3 mrad) for translations  
2311 (rotations) means more than a 5% degradation in rms and/or bias shifts in  $\theta$ ,  $\phi$ , and  $\Delta\theta$  of 0.01 mrad,  
2312 1 mrad, and 0.1 mrad, respectively.

2313 While corrections are typically done on a case-by-base basis, they fall under two general categories,  
2314 analytic and simulation based. Analytic corrections rely upon specific knowledge of the misalign-  
2315 ment, with each case being handled separately; as such, the additional resources required, both extra  
2316 constants and operations, if any, vary accordingly. Simulation based corrections are all done in the  
2317 same manner. The algorithm is run over a training MC sample (same setup but with  $p_T = 200$   
2318 GeV instead of the normal 100 GeV sample so as not to overtrain the corrections), and the mean bi-  
2319 ases for  $\theta$ ,  $\phi$ , and  $\Delta\theta$  are saved for different, equally spaced regions in the  $\eta - \phi$  plane over the wedge  
2320 based on the fitted  $\theta$  and  $\phi$  values. Currently, these values are saved for 10  $\eta$  and 10  $\phi$  bins (100  $\eta, \phi$   
2321 bins total), with the number of bins in each direction being a configurable parameter. When the al-  
2322 gorithm runs with simulation based correction, this table of constant corrections is saved in a LUT  
2323 before runtime, and corrections are added to final fit quantities based on the (uncorrected)  $\theta$  and  
2324  $\phi$  fit values. With the settings mentioned, this is 300 extra constants ( $10\eta\text{-bins} \times 10\phi\text{-bins} \times 3$  fit  
2325 quantities) and two extra operations (a lookup and addition for each quantity done in parallel). The  
2326 simulation correction can, in principle, also be applied to the algorithm in nominal conditions with  
2327 non-trivial improvements, as detailed below in Section A.17. Depending on the misalignment case in  
2328 question, different approaches work better. A summary of correction methods, including resources  
2329 necessary for the individual analytic cases, is shown in Table A.3.

	$\Delta s$	$\Delta z$	$\Delta t$	$\gamma_s$	$\beta_z$	$\alpha_t$
Analytic Resources	yes+ 11c/2op	yes+ oc/oop	yes+ oc/oop	yes 56c/1op	no —	yes 400c/2n <sub>X</sub> op, 32c/12n <sub>X</sub> 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

2330 A.17 SIMULATION CORRECTION OF THE ALGORITHM UNDER NOMINAL CONDITIONS

2331 In addition to using simulation based correction to counter the effects of several classes of misalign-  
 2332 ment, the correction can be applied at to the algorithm under nominal conditions. The main effect  
 2333 of this correction is to mitigate the effects of the bias in stereo strips. As such, the correction has a  
 2334 larger effect on quantities that rely on the aggregate slope  $m_y$ , as can be seen in in Figure A.18, im-  
 2335 proving  $\sigma_{\theta_{fit} - \theta_{true}}$  resolution by about 25%, and reducing  $\sigma_{\phi_{fit} - \phi_{true}}$  by over 50% and restoring a largely  
 2336 Gaussian shape. The slight, apparent degradation in  $\Delta\theta$  is due to a more mild version of the effect  
 2337 seen in Figure A.7.

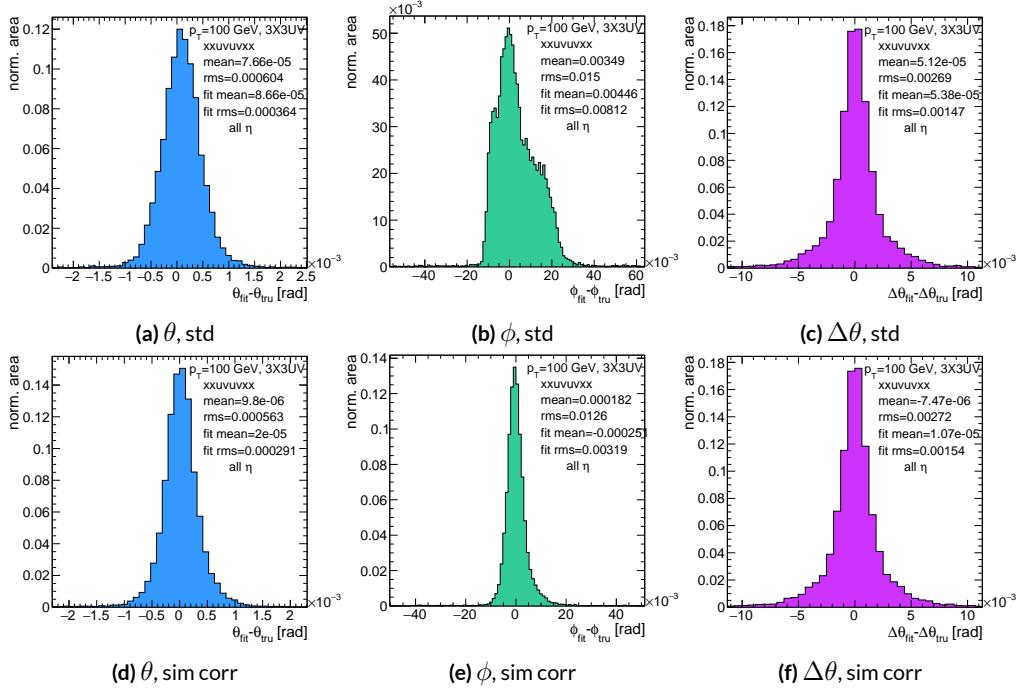
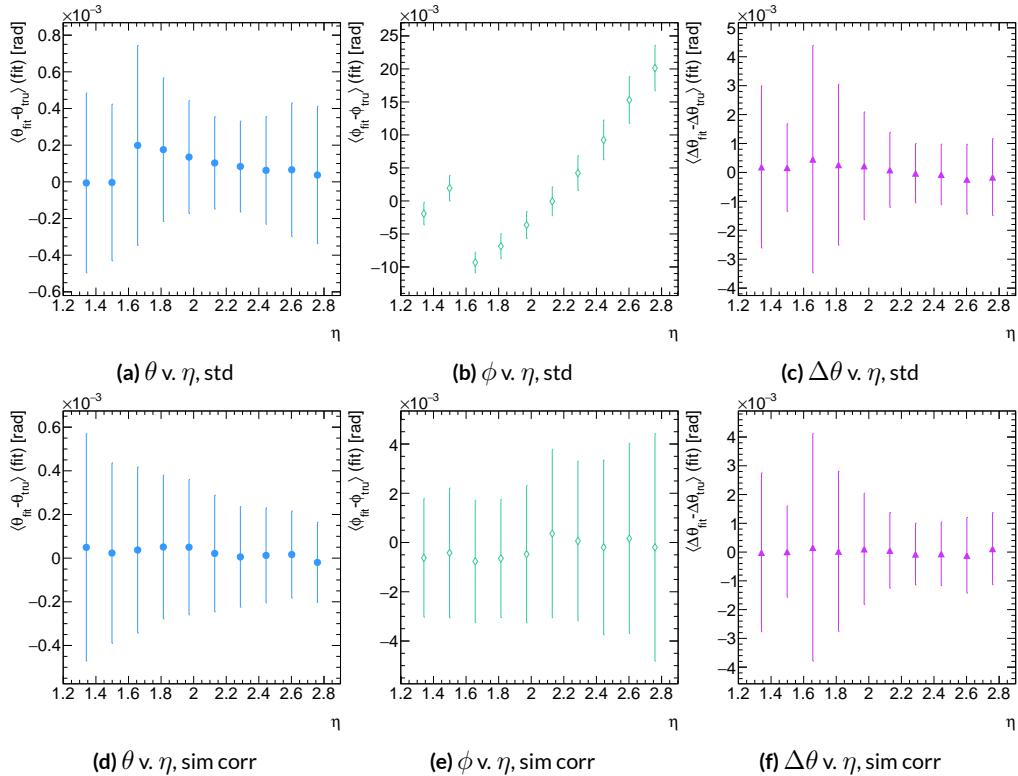


Figure A.18: Nominal residual plots for both uncorrected and simulation corrected cases;  $\theta, \phi, \Delta\theta$  for  $p_T = 100$  GeV muons

2338 As can be seen in Figure A.19, the simulation based correction also removes the  $\eta$  dependence to  
 2339 fit quantity resolution distributions, as expected. One consequence of this is that simulation-based  
 2340 corrections applied to the misalignment cases below will restore performance to the “sim” and not  
 2341 the “std” distributions of Figure A.18. Hence, when making comparisons between simulation cor-  
 2342 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  $\eta$  with points as means and error bars as rms values in each  $\eta$  bin for the angles  $\theta, \phi, \Delta\theta$  for  $p_T = 100$  GeV muons in the uncorrected and simulation corrected cases.

2343

2344 That the improvements from a simulation-based correction improve performance of the algo-

2345 rithm in nominal conditions most for the quantities that depend most on stereo information ( $\phi$  and  
 2346  $\theta$ ) and remove the  $\eta$  dependence of fit quantity resolutions suggests that there could, in principle, be  
 2347 analytic corrections that could be applied to the nominal algorithm. One possible solution is to in-  
 2348 troduce an additional set of constants, having the  $y_{base}$  depend on the strip number, similar to the  $\gamma_s$   
 2349 correction for  $z_{plane}$  described in Section A.21, which would add a lookup per hit and  $8 \times n_{bins,y}$  extra  
 2350 constants that would be optimized as the  $\gamma_s$  correction was.

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

2351 The simulation correction residual rms values suggest a limit on the quality of such correction  
 2352 and could perhaps be implemented generically on their own regardless of misalignment for rms  
 2353 values on fit quantities of 0.291 mrad for  $\theta$ , 3.19 mrad for  $\phi$ , and 1.54 for  $\Delta\theta$ , which represent a 20%  
 2354 improvement for  $\theta$ , a 62% improvement for  $\phi$ , and a slight degradation in  $\Delta\theta$  of 4.7%, again owing  
 2355 to an effect similar to the one in A.7.

2356 A.18 TRANSLATION MISALIGNMENTS ALONG THE HORIZONTAL STRIP DIRECTION ( $\Delta s$ )

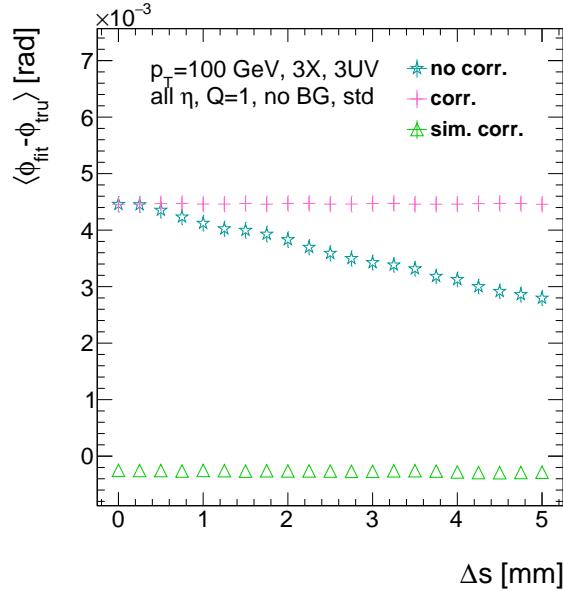
2357 A translation in  $s$  (i.e. along the direction of a horizontal strip) only affects the stereo strips, and,  
2358 since the stereo angle is small, a very large misalignment is necessary for effects to be noticeable (a  
2359 misalignment of roughly 17 mm corresponds to one strip's misalignment in the stereo planes). The  
2360 only quantity to show any meaningful deviation with misalignments with translations in  $s$  is the  $\phi$   
2361 residual bias (a change of 0.4 mrad at  $\Delta s = 1$  mm), as can be seen in the uncorrected curve of Figure  
2362 A.20.

2363 A translation in  $s$  induces a constant shift in the calculated horizontal slope,  $m_x$  in Equation A.4.  
2364 This constant shift should only depend on which stereo planes included in a fit are misaligned and  
2365 how misaligned they are. Hence, the correction to  $m_x$ , for a sum over misaligned stereo planes  $i$ ,  
2366 with their individual misalignments in  $s$  and plane positions in  $z$  is:

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

2367 Given prior knowledge of misalignment, these corrections to  $m_x$  can be performed ahead of time  
2368 and saved in a lookup table (LUT), similar to the LUT used for constants in the X local slope ( $M_x^l$ )  
2369 calculation. The added overhead of this analytic correction is hence eleven constants in memory, a  
2370 lookup, and one addition. The correction perfectly corrects the effects of misalignment, as can be  
2371 seen in Figure A.20. The simulation based correction described above can also be used to correct  
2372 for  $\Delta s$  misalignments, with the results of that correction also shown in Figure A.20. The apparent

2373 discrepancy between the simulated and analytic correction is a natural consequence of the fact that  
 2374 the simulation correction, as previously mentioned, restores the  $\phi$  residual distribution to an overall  
 2375 more Gaussian shape.



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

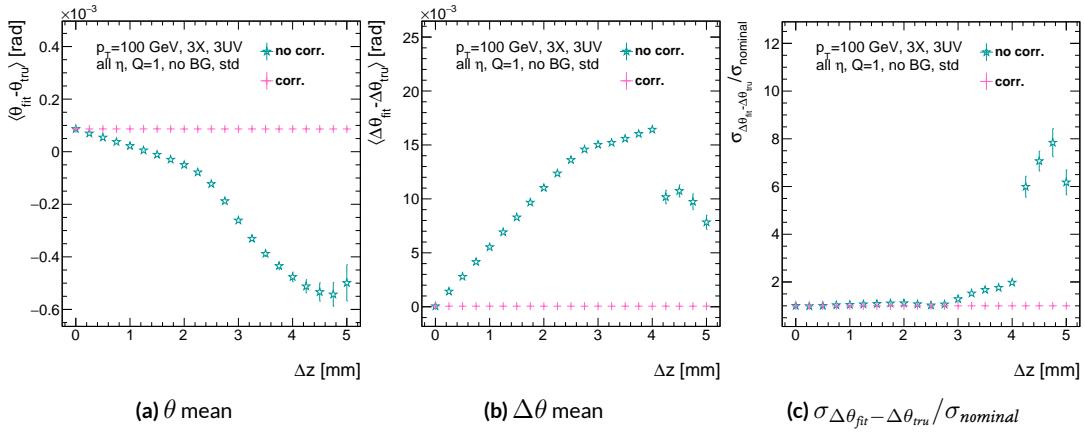
2376 A.19 TRANSLATION MISALIGNMENTS ORTHOGONAL TO THE BEAMLINE AND HORIZON-  
2377 TAL STRIP DIRECTION ( $\Delta z$ )

2378 A translation in AMDB  $z$ , the direction orthogonal to both the beamline and the horizontal strip  
2379 direction, corresponds to a translation in the  $y$  of Equation A.1, affecting all slope calculations. This  
2380 has a large impact on the  $\theta$  residual bias and both the bias and rms of  $\Delta\theta$  residual, as can be seen in  
2381 Figures A.21 (a)–(c). The marked degradation and non-linear behavior in performance at very high  
2382 levels of misalignments is a result of low statistics; there are fewer fits at high level of misalignments  
2383 since for  $\Delta z \gtrsim 3$  mm, most fits will fail the  $\Delta\theta$  cut. The  $\theta$  bias shifts by about 0.075 mrad at  $\Delta z = 1$   
2384 mm, and  $\Delta\theta$  shifts by about 5 mrad for the same level of misalignment. While the fitted rms of the  
2385  $\Delta\theta$  residual remains fairly stable for  $\Delta z < 1$  mm or so, between  $\Delta z = 2$  mm and  $\Delta z = 3$  mm, the  
2386 rms increases by 15% before the  $\Delta\theta$  cut issue mentioned above intervenes.

2387 Fortunately, these misalignments are straightforward to correct with knowledge of the misalign-  
2388 ment. The only modification necessary for this correction is to change the definitions of  $y_{base}$  in  
2389 Equation A.1 for the individual hit slope addressing. This is done before runtime and adds no over-  
2390 head to the algorithm, and the correction quality is only limited by knowledge of the misalignment.  
2391 The results of this correction are also shown in Figures A.21 (a)–(c) and restore nominal perfor-  
2392 mance.

---

Since  $\Delta\theta = \frac{M'_X - M''_X}{1 + M'_X M''_X}$  and  $M'_X = B_k \sum y_i (z/\bar{z} - 1)$ , a shift  $\Delta y$  translates (with typical slope values of  $\sim 0.3$ ) 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

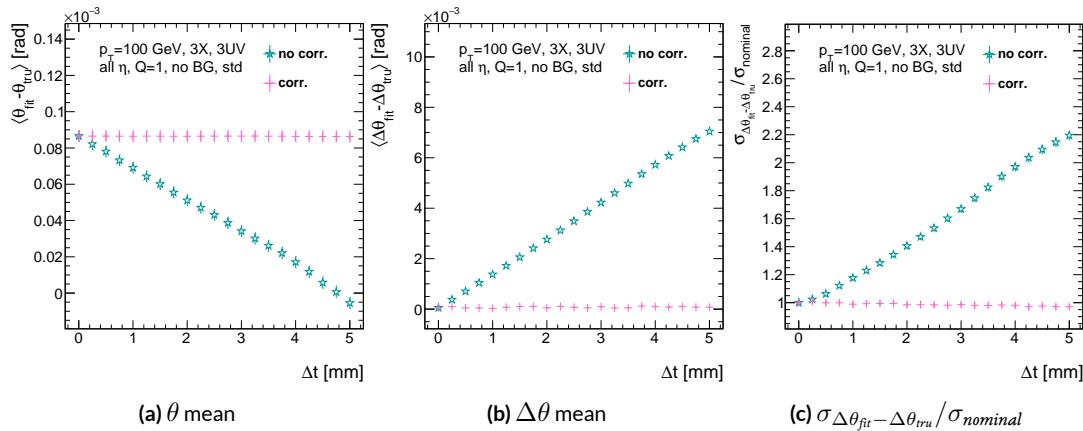


**Figure A.21:** The affected quantities of  $\Delta z$  misalignments:  $\theta$  bias,  $\Delta\theta$  bias, and  $\sigma_{\Delta\theta_{\text{fit}} - \Delta\theta_{\text{tru}}} / \sigma_{\text{nominal}}$  for both the misaligned and corrected cases.

2393 A.20 TRANSLATION MISALIGNMENTS PARALLEL TO THE BEAMLINE ( $\Delta t$ )

2394 The effects of misalignment due to translations in  $t$  are very similar to those due to translations in  
 2395  $z$  without the complication of the  $\Delta\theta$  cut, affecting the  $z$  instead of the  $y$  coordinate that enters  
 2396 into hit slope calculations. Again,  $\theta$  bias,  $\Delta\theta$  bias, and  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}$  are the primarily affected quan-  
 2397 tities. For  $\Delta t = 1$  mm,  $\theta$  bias shifts by about 0.02 mrad,  $\Delta\theta$  bias shifts by just under 2 mrad, and  
 2398  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}$  degrades by about 20%. The correction for this misalignment once again costs no over-  
 2399 head and consists of changing stored constants in the algorithm, in this case the positions along  
 2400 the beamline of the misaligned planes, with results similarly limited by knowledge of the misalign-  
 2401 ment. The slight improvement with correction to  $\Delta\theta$  rms is due to the real effect of a larger lever  
 2402 arm. Both the misaligned and corrected distributions of affected quantities of interest are shown in

Figure A.22.



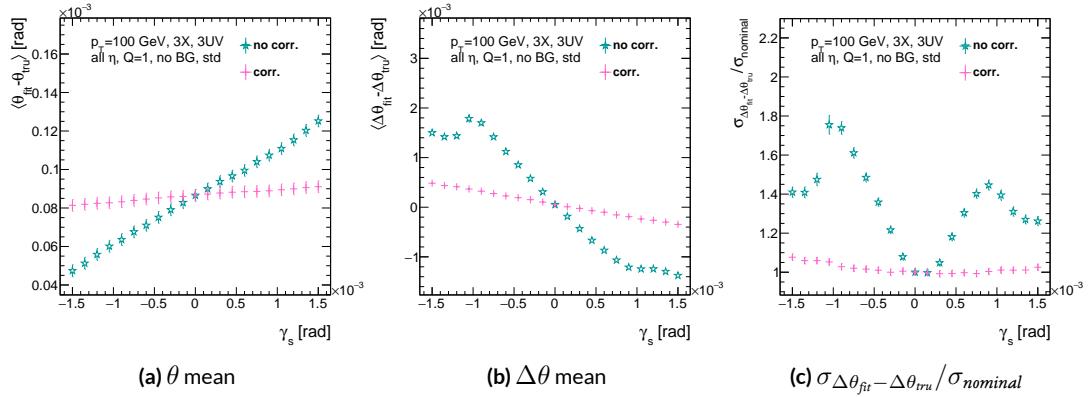
2403 **Figure A.22:** The affected quantities of  $\Delta t$  misalignments:  $\theta$  bias,  $\Delta\theta$  bias, and  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}} / \sigma_{nominal}$  for both the misaligned and corrected cases.

2404 A.21 CHAMBER TILTS TOWARDS AND AWAY FROM THE IP ( $\gamma_s$  ROTATION)

2405 Chamber misalignment due to rotations around the  $s$  axis act effectively like a translation in  $t$  that  
2406 depends on strip number. These rotations tilt misaligned chambers away from (towards) the IP for  
2407 positive (negative) values of  $\gamma_s$ . Since, unlike for the other two rotation cases that will be studied,  
2408 positive and negative rotation values are not symmetric, this misalignment is studied for both posi-  
2409 tive and negative  $\gamma_s$  values. The divergent effect at the tails is a result of a large population of fits not  
2410 having fit quantities within the cores, and so not appearing in the fit rms. Once again, affected quan-  
2411 tities of interest  $\theta$  bias,  $\Delta\theta$  bias, and  $\sigma_{\Delta\theta_{fit} - \Delta\theta_{tru}}$ . The effects of misalignment can be seen in Figures  
2412 A.23 (a)–(c). The relationship between biases and  $\gamma_s$  is roughly linear with  $\Delta\gamma_s = 0.3$  mrad (the an-  
2413 gular scale corresponding to linear shifts of  $\sim 1$  mm) corresponding to  $0.005$  mrad ( $0.12$  mrad) for  $\theta$   
2414 ( $\Delta\theta$ ). For  $\sigma_{\Delta\theta_{fit} - \Delta\theta_{tru}}$ , degradation is not symmetric. For negative (positive)  $\gamma_s$ , with the quadruplet  
2415 tilted towards (away from) the IP, slope-roads are artificially expanded (shrunk), decreasing (increas-  
2416 ing) the granularity of the trigger, explaining the asymmetry in Figure A.23 (c), with the degradation  
2417 being a  $10\%$  ( $25\%$ ) effect for  $\gamma_s$  of  $+(-)0.3$  mrad.

2418 Corrections are less simple in this case. In principle, corrections of the same accuracy of the trans-  
2419 lations could be calculated per strip, but the overhead of one correction per strip (many thousands  
2420 of constants) is prohibitive. Instead, each plane was divided into eight equal segments with a  $t$  value  
2421 ( $z$  in the slope calculation) assigned to strips in each region to correct for the misalignment. This  
2422 amounts to 56 extra constants and a 2D instead of a 1D LUT for  $z$  positions while the algorithm  
2423 runs. The corrected distributions can also be seen in Figures A.23 (a)–(c). The corrections, while not

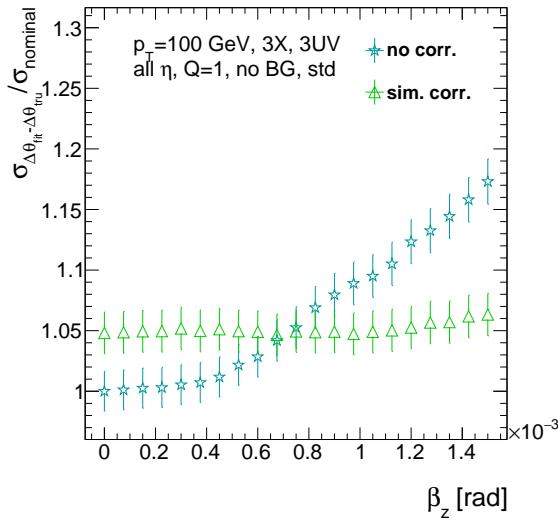
as effective as for the simple translation cases, are still very effective with the quoted misalignment values for bias shifts down to 0.001 mrad (0.25 mrad) for  $\theta$  ( $\Delta\theta$ ) and no more than a 2% degradation in  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{true}}$  for  $|\gamma_s| = 0.3$  mrad.



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

<sup>2427</sup> A.22 ROTATION MISALIGNMENTS AROUND THE WEDGE VERTICAL AXIS ( $\beta_z$ )

<sup>2428</sup> While misalignments coming from rotations around the  $z$  axis (the direction orthogonal to both  
<sup>2429</sup> the beamline and the horizontal strip direction) foreshorten the strips as seen from the IP and add  
<sup>2430</sup> a deviation in  $t$ , the long lever arm largely washes out any effects of this misalignment. Only the  
<sup>2431</sup>  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$  is noticeably affected, though only at severe misalignments, with only about a 1% degra-  
<sup>2432</sup> dation in performance at  $\beta_z = 0.3$  mrad (corresponding to a linear shift of  $\sim 1$  mm). A simulation  
<sup>2433</sup> based correction works well to cancel out the effects of this misalignment, and the  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$  as a  
<sup>2434</sup> function of misalignment with and without corrections are shown in Figure A.24. The apparent  
<sup>2435</sup> 2% effect in the simulation corrected curve is a result of a more mild version of the effect shown in  
<sup>2436</sup> Figure A.7.

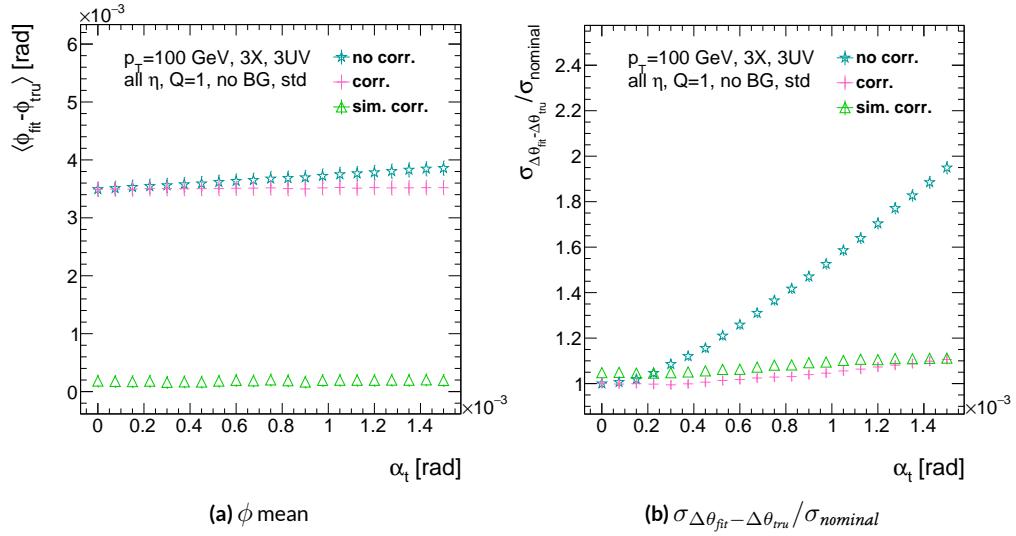


**Figure A.24:** The effects of rotations in the  $z$  axis on  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}} / \sigma_{nominal}$  a function of  $\beta_z$  both with and without misalignment corrections.

<sup>2437</sup> A.23 ROTATION MISALIGNMENTS AROUND THE AXIS PARALLEL TO THE BEAMLINE ( $\alpha_t$ )

<sup>2438</sup> Misalignments arising from rotations around the  $t$  axis (parallel to the beamline at the center of  
<sup>2439</sup> the base of the wedge) are essentially rotations in the  $\phi$  direction. The quantities of interest most  
<sup>2440</sup> affected are the  $\phi$  bias and  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$ , as shown in Figures A.25 (a) and (b), respectively, and cor-  
<sup>2441</sup> respond to a shift in  $\phi$  bias of 0.2 mrad and a 10% degradation in  $\sigma_{\Delta\theta_{fit}-\Delta\theta_{tru}}$  for  $\alpha_t = 0.3$  mrad  
<sup>2442</sup> (corresponding to a linear shift of  $\sim 1$  mm). The raw instead of fitted mean  $\phi$  biases is used in Figure  
<sup>2443</sup> A.25 (a) to better illustrate the effect of misalignment.

<sup>2444</sup> Since the effect of misalignment is dependent on horizontal (along the strip direction,  $\hat{s}$ ) in addi-  
<sup>2445</sup> tion to vertical information, corrections cannot be applied before a fit takes place. The  $\phi$  bias shift is  
<sup>2446</sup> uniform over the entire wedge, so a constant additive correction to  $\phi$  based on the level of misalign-  
<sup>2447</sup> ment can be applied to all fits depending on how many misaligned stereo planes enter in the fit.  $\Delta\theta$   
<sup>2448</sup> is less straightforward, but corrections to the  $y$  and  $z$  information used in the local slope calculation  
<sup>2449</sup> in Equation A.4 can be applied once  $\theta_{fit}$  and  $\phi_{fit}$  are known. These corrections are calculated ahead  
<sup>2450</sup> of time in bins of uniform  $\eta$  and  $\phi$  as with the simulation corrections using the same framework  
<sup>2451</sup> as the misalignment calculation. The results of both types of correction can be seen in Figure A.22.  
<sup>2452</sup> The apparent discrepancy between the simulation and analytic corrections in the  $\phi$  bias happens for  
<sup>2453</sup> the same reason as in the  $\Delta s$  misalignment correction cases, as simulation correction restores a more  
<sup>2454</sup> Gaussian shape to the  $\phi$  residual distribution opposed to the uncorrected nominal case, as discussed  
<sup>2455</sup> in Section A.17.



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

2456 A.24 CONCLUSION

2457 The algorithm for Micromegas detectors in the NSW Trigger Processor performs well in a variety of  
2458 conditions and has proven robust to a number of effects to deliver measurements on muon tracks  
2459 of the three angles  $\theta$ ,  $\phi$ ,  $\Delta\theta$ . Under nominal conditions, the rms values for the residuals of these  
2460 quantities are 0.364 mrad for  $\theta$ , 8.12 mrad for  $\phi$ , and 1.47 mrad for  $\Delta\theta$ . Algorithm performance was  
2461 found to be largely independent of the charge threshold setting, and a hit majority BCID associa-  
2462 tion was found to provide proper timing information over 99.7% even in the most relaxed settings  
2463 (2X+1UV coincidence threshold requirement+wide slope-road+background). The introduction of  
2464 wide slope-roads to better mimic potentially limited algorithm resources at run time and the intro-  
2465 duction of incoherent background was found to have a manageable effect on fit quantity residual  
2466 rms values and on total algorithm efficiency for sufficiently stringent coincidence threshold. The ef-  
2467 ffects of the three translation and three rotation misalignments specified by AMDB convention were  
2468 studied, and correction methods for each of the six cases was developed. Simulation-based correc-  
2469 tions were found to improve nominal algorithm performance to residual rms value of 0.291 mrad for  
2470  $\theta$ , 3.19 mrad for  $\phi$ , and 1.54 for  $\Delta\theta$ , which represent improvements of 20%, 62%, and -4.7%, respec-  
2471 tively. Misalignment corrections were found to restore nominal performance for all but the rotation  
2472 around the  $s$  axis, and a summary of tolerances may be found in Table A.4.

	No Correction	Correction
$\Delta s$	4 mm ( $\phi$ bias)	$> 5$ mm
$\Delta z$	0.25 mm ( $\Delta\theta$ )	$> 5$ mm
$\Delta t$	0.25 mm ( $\Delta\theta$ )	$> 5$ mm
$\gamma_s$	0.15 mrad ( $\Delta\theta$ bias)	0.75 mrad
$\beta_z$	0.9 mrad ( $\Delta\theta$ rms)	$> 1.5$ mrad
$\alpha_t$	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  $\theta$ , 1 mrad for  $\phi$ , and 0.25 mrad for  $\Delta\theta$  for both the uncorrected and corrected cases;  $> 5$  mm and  $> 1.5$  mrad mean that such a degradation does not occur for the range of misalignment studied. Most affected quantity in parentheses.

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

Conventional Wisdom

# B

2473

2474

## Telescoping Jets

2475 ANOTHER APPROACH TO IMPROVING  $ZH \rightarrow \ell\ell b\bar{b}$  is the use of telescoping jets<sup>28</sup>, which har-  
2476 nesses the power of multiple event interpretations. The use of multiple event interpretations was  
2477 originally developed with non-deterministic jet algorithms like the Q-jets (“quantum” jets) algo-  
2478 rithm<sup>60</sup>. When a traditional or “classical” algorithm, such as the Cambridge-Aachen<sup>68</sup> and anti- $k_t$ <sup>54</sup>

249

algorithms, is applied to an event, it produces one set of jets for that event, i.e. a single interpretation of that event. With multiple event interpretations, each event is instead given an ensemble of interpretations. In the case of Q-jets, this ensemble is created through a non-deterministic clustering process for an anti- $k_t$  jet algorithm. With telescoping jets, multiple jet cone radii (the characteristic size parameter,  $R$ ) around a set of points in the pseudorapidity-azimuth ( $\eta - \phi$ ) plane are used to generate a series of jet collections. Instead of an event passing or not-passing a given set of cuts, a fraction (called the cut-weight,  $z$ ) of interpretations will pass these cuts. This cut-weight allows for enhanced background suppression and increased significance of observed quantities for a given data set, as detailed in Ref.<sup>39</sup>. 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.<sup>?</sup>. The signal sample used is generated in PYTHIA8<sup>?</sup> with the CTEQ6L1 parton distributions functions (PDFs) and AU2 tune<sup>???</sup> for the  $ZH$  process with  $m_H = 125$  GeV (henceforth,  $ZH_{125}$ ). The primary background processes examined in this study were  $Z +$  jets with massive  $b$  and  $c$  quarks. These samples are generated with version 1.4.1 of the SHERPA generator<sup>?</sup>.

2499 B.2 JET RECONSTRUCTION AND CALIBRATION

2500 In order to construct telescoping jets, jet axes must first be found around which to “telescope.” In  
2501 the reconstructed-level analysis, the anti- $k_t$  algorithm with  $R = 0.4$  is used to reconstruct jets from  
2502 topological clusters in the calorimeters. The four vectors of these anti- $k_t$  algorithm with  $R = 0.4$  jets  
2503 are calibrated to match truth information obtained from simulation and validated in data. To take  
2504 into account the effect of pile-up interactions, jet energies are corrected using a jet-area based tech-  
2505 nique<sup>3</sup>, and each jet with  $p_T < 50$  GeV and  $|\eta| < 2.4$  is subject to a requirement that at least 50% of  
2506 the scalar sum of the  $p_T$  of tracks matched to this jet be composed of tracks also associated with the  
2507 primary vertex. Jet energies are also calibrated using  $p_T$  and  $\eta$ -dependent correction factors<sup>3</sup>. Fur-  
2508 thermore, at least two jets must have  $|\eta| < 2.5$  in order to be  $b$ -tagged. The MV1 algorithm<sup>?????</sup>  
2509 is used for  $b$ -tagging. Once jets are reconstructed and  $b$ -tag weights have been calculated, the two  
2510 hardest,  $b$ -tagged jets are used as the telescoping jet axes. Additional details can be found in Ref.<sup>3</sup>.

2511 After the telescoping jet axes have been established, telescoping jets are constructed using topolog-  
2512 ical clusters in the calorimeters at a variety of jet cone sizes. Including the original anti- $k_t$  jets used for  
2513 the  $R = 0.4$  case, twelve total sets of jets of cone sizes ranging from  $R = 0.4\text{--}1.5$  are constructed,  
2514 with each successive size having a radius 0.1 larger than the preceding set. For each jet axis, telescop-  
2515 ing jets consist of any topological cluster lying within  $R$  of the axis. In the event of overlap, clusters  
2516 are assigned to the closer jet axis. If a given cluster is equidistant from the two jet axes, the cluster  
2517 is assigned to whichever jet axis is associated with the anti- $k_t$  jet with higher  $p_T$ . Calibration for the  
2518 telescoping jets is conducted using corrections for anti- $k_t$  calorimeter topological cluster jets; the

2519     $R = 0.4$  corrections are used for telescoping  $R = 0.5$ , and the  $R = 0.6$  corrections are used for  
2520    telescoping  $R \geq 0.6$  (cf. Sec. B.4). The telescoping cone jets ( $R \geq 0.5$ ) at reconstructed level are  
2521    trimmed using Cambridge-Aachen jets with  $R = 0.3$  and  $f_{cut} = 0.05$  with respect to the untrimmed  
2522    jet  $p_T^{40}$ . Since these jets are trimmed, the active area correction is not applied. In the event a  $Z$  can-  
2523    didate electron falls within  $R$  of the axis of a telescoping jet, its 4-momentum is subtracted from that  
2524    of the jet vectorially.

2525    A similar process is used to construct telescoping jets in the truth-level analysis below. Instead of  
2526    the two hardest  $b$ -tagged anti- $k_t$  with  $R = 0.4$  jets reconstructed with calorimeter topological clus-  
2527    ters, the two hardest truth  $b$ -jets in an event are used. Instead of making a cut on  $b$ -tagging weight  
2528    to  $b$ -tag, truth jets are examined to see whether a  $b$ -hadron with  $p_T > 5$  GeV is contained within  
2529     $\Delta R < 0.4$  of the jet axis; the presence of a  $b$ -hadron is used to  $b$ -tag truth-level jets. These two jets  
2530    again provide the jets for the  $R = 0.4$  case and the axes around which telescoping takes place. Stable  
2531    truth particles, not including muons and neutrinos, are used in place of calorimeter topological clus-  
2532    ters.  $Z$  candidate electron-telescoping jet overlap removal is performed at truth level, too. Missing  $E_T$   
2533    is calculated using the vector sum of the four momenta of stable truth-level neutrinos. Since there  
2534    are no pileup particles stored at truth level, truth-level telescoping jets are not trimmed.

### 2535    B.3    EVENT RECONSTRUCTION AND SELECTION

2536    Events are selected on the basis of a combination of leptonic, jet, and missing  $E_T$  requirements,  
2537    which are outlined in Table B.1. Leptons are categorized by three sets of increasingly stringent qual-  
2538    ity requirements, which include lower limits on  $E_T$ , upper limits on  $|\eta|$ , impact-parameter require-

2539 ments, and track-based isolation criteria. The requirements differ for electrons<sup>3</sup> and muons<sup>3</sup>. Events  
2540 are selected with a combination of single lepton, dielectron, and dimuon requirements. Each event  
2541 must contain at least one lepton passing medium requirements and at least one other lepton pass-  
2542 ing loose requirements. These leptons are used to create a dilepton invariant mass cut to ensure the  
2543 presence of a  $Z$  boson and suppress multijet backgrounds.

2544 Event selection requirements are also imposed on the anti- $k_t$  with  $R = 0.4$  jets. There must be at  
2545 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  
2546 the second  $b$ -tagged jet must have  $p_T$  of at least 20 GeV. There are further topological cuts on the  
2547 separation of the two jets  $\Delta R(b, \bar{b})$ , the distance between the two jets in the  $(\eta, \phi)$  plane, according  
2548 to the transverse momentum of the  $Z$  boson,  $p_T^Z$ . These are shown in Table B.2.

2549 The truth-level analysis has the same missing  $E_T$ , jet  $p_T$ ,  $m_{ll}$ , and additional topological selection  
2550 criteria, but the use of truth-level information simplifies the other requirements. Instead of lepton  
2551 quality requirements,  $Z$  boson candidate leptons' statuses and MC record barcodes are checked to  
2552 ensure the leptons are stable.

2553 In the jet calibration validation, the reconstructed level analysis lepton and  $m_{ll}$  requirements are  
2554 imposed, but neither the missing  $E_T$  nor the jet selection requirements are applied so as not to bias  
2555 the validation.

#### 2556 B.4 VALIDATION OF JET CALIBRATION

2557 In order to validate the jet energy scale and resolution of jets constructed with this telescoping-jets  
2558 algorithm, values of  $p_T^{rec}/p_T^{tru}$  are studied for each value of  $R$  for the  $Z+jets$  MC sample. In a given

**Table B.1:** A summary of basic event selection requirements. Truth-level  $b$ -tagging is done with truth-level information.

Requirement	Reconstructed	Truth	Validation
Leptons	1 medium + 1 loose lepton	2 produced by $Z$ boson	1 medium + 1 loose lepton
$b$ -jet	2 $b$ -tags	2 $b$ -jets	—
$p_T$ jet 1 (jet 2)	$> 45$ GeV ( $> 20$ ) GeV		—
Missing $E_T$		$E_T^{\text{miss}} < 60$ GeV	—
$Z$ boson		$83 < m_{ll} < 99$ GeV	

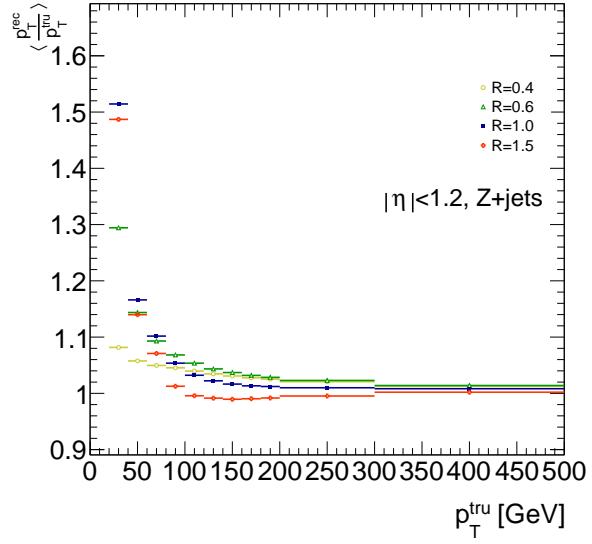
**Table B.2:** Topological requirements of the event selection.

$p_T^Z$ [GeV]	$\Delta R(b, \bar{b})$
0–90	0.7–3.4
90–120	0.7–3.0
120–160	0.7–2.3
160–200	0.7–1.8
$> 200$	$< 1.4$

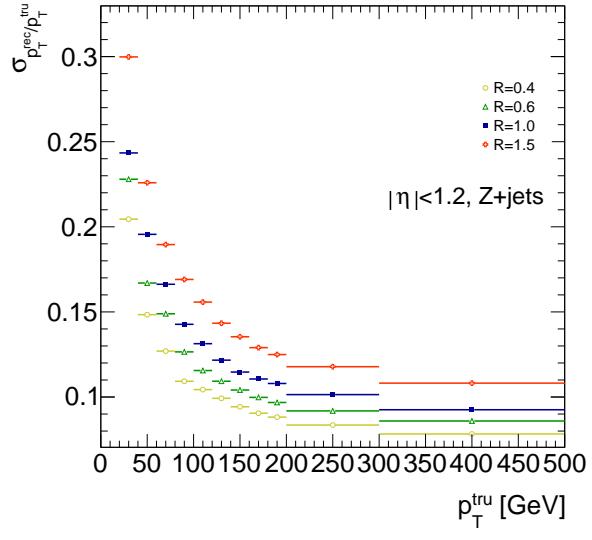
2559 event, all jets, not just the two hardest  $b$ -tagged jets, are telescoped. These jets are constructed in the  
2560 same way as in the reconstructed and truth-level analyses; reconstructed-level jets are made from  
2561 calorimeter topological clusters within  $R$  of the anti- $k_t$  with  $R = 0.4$  jet axes and then trimmed, and  
2562 truth-level jets are made from stable truth particles within  $R$  of the anti- $k_t$  with  $R = 0.4$  jet axes.  
2563 The reconstructed and truth-level telescoping jet ensembles are matched according to the separation  
2564 in the  $(\eta, \phi)$  plane of their corresponding anti- $k_t$  with  $R = 0.4$  jets used as seeds. Only jets with  
2565  $|\eta| < 1.2$  are examined here, and the results of studies on the  $ZH_{125}$ ,  $ZZ$ , and  $t\bar{t}$  samples, as well as  
2566 over other  $|\eta|$  ranges, are outlined in<sup>28</sup>. Any reconstructed jets not within  $\Delta R = 0.3$  of a truth jet  
2567 are discarded. In the event that multiple reconstructed jets are the same distance away from a given  
2568 truth jet, the reconstructed jet with the highest  $p_T$  gets matched. Matching is retained for all  $R$  values  
2569 (i.e. telescoping seeds are matched, and telescoping jets are assumed to match if the anti- $k_t$  jets from  
2570 which their seeds are derived match).

2571 Once anti- $k_t$  with  $R = 0.4$  reconstructed and truth jets are matched, response functions are cre-  
2572 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  
2573 bin for 200–300 GeV, and one bin for 300–500 GeV for each  $R$ , with bins chosen for purposes of  
2574 statistics. Ensembles with  $p_T^{tru} < 20$  GeV are ignored since no calibration exists for jets with trans-  
2575 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  
2576 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,  
2577 and the error on the mean is taken from the error of this parameter in the fit. The resolutions are the  
2578 values of the square root of the variance on this fit. As the total response distributions in Figure B.1  
2579 show, performance is best for low  $R$  values and high values of  $p_T^{tru}$ . Figure B.1 shows the  $R = 0.4$

2580 (anti $k_t$ ) case to show a baseline for performance,  $R = 0.6$  to show the deviations with “correct”  
2581 calibrations, and  $R = 1.0, 1.5$  to show how big those deviations get with larger  $R$  jets. The resolu-  
2582 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  
2583 consistent over various  $R$  values. Resolution, as might naïvely be expected, is worse for increasingly  
2584 larger values of  $R$ . For  $p_T^{tru} < 60$  GeV, resolution degrades, and response degrades in particular for  
2585 increasing  $R$ ; this is likely a result from residual pileup effects.



(a)



(b)

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

2586    B.5    TRUTH-LEVEL ANALYSIS

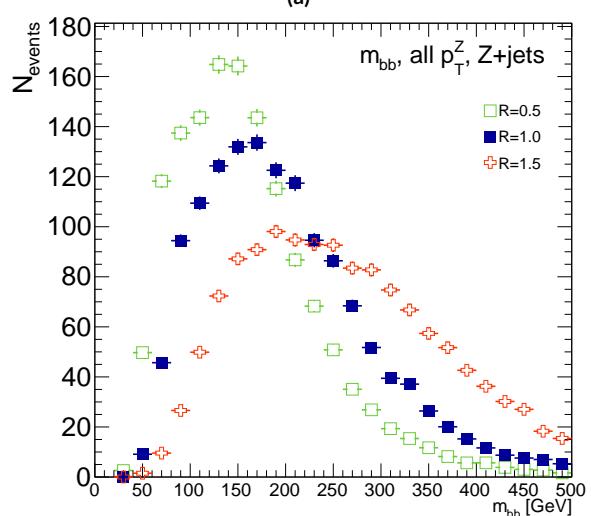
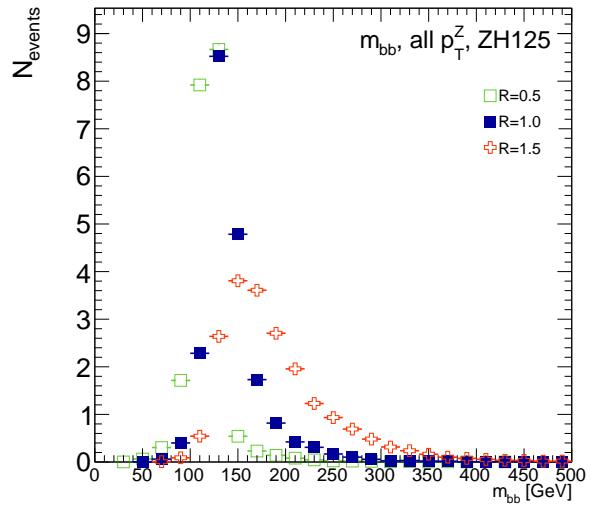
2587    To understand the limits and sources of any potential improvements, a truth-level analysis was con-  
2588    ducted on MC samples with a  $ZH_{125}$  signal sample and a  $Z+jets$  background sample. Distribu-  
2589    tions for the dijet invariant mass,  $m_{bb}$ , were made for each telescoping radius.\* Both signal and back-  
2590    ground samples develop more pronounced tails in the high  $m_{bb}$  region as  $R$  increases, as shown in  
2591    Figure B.2.  $N_{events}$  is normalized to expected values in data.

2592       One way to take advantage of this information is to make a cut on  $m_{bb}$  for two different radii.  
2593       This is graphically depicted in Figure B.3 for the optimized combination of  $m_{bb,R=0.9}$  (telescoping  
2594       cone jets constructed as outlined in Sec. B.2) vs.  $m_{bb,R=0.4}$  (anti- $k_t$  jets). At truth-level, the majority  
2595       of events in the signal  $ZH_{125}$  sample are concentrated in relatively narrow region of parameter space,  
2596       where this is certainly not the case for the more diffuse  $Z+jets$  background sample.

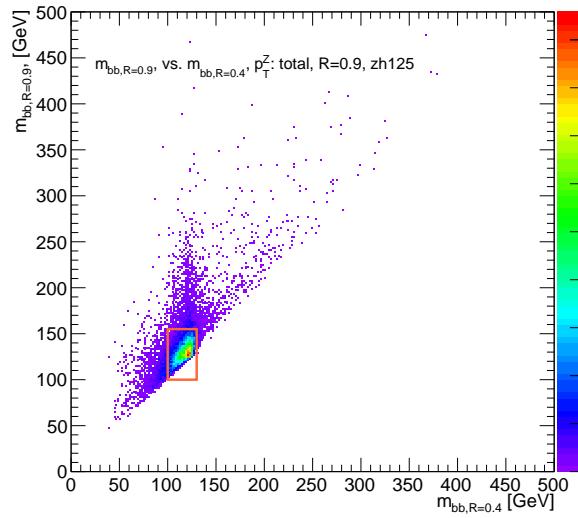
2597       Another way to take advantage of multiple event interpretations is to make use of an event's cut-  
2598       weight, denoted  $z$  and defined as the fraction of interpretations in a given event that pass a certain set  
2599       of cuts (in this note, a cut on  $m_{bb}$ ). The distribution of cut-weights for a sample of events is denoted  
2600        $\rho(z)$ . To enhance the significance of a cut-based analysis, events can be weighted by the cut-weight  
2601       or any function  $t(z)$  of the cut-weight. Weighting events by  $t(z)$  modifies the usual  $S/\delta B$  formula  
2602       used to calculate significances. In this note,  $\delta B$  is based on Poissonian statistics and is taken as  $0.5 +$   
2603        $\sqrt{0.25 + N_B}$ , where  $N_B$  is the number of background events.

---

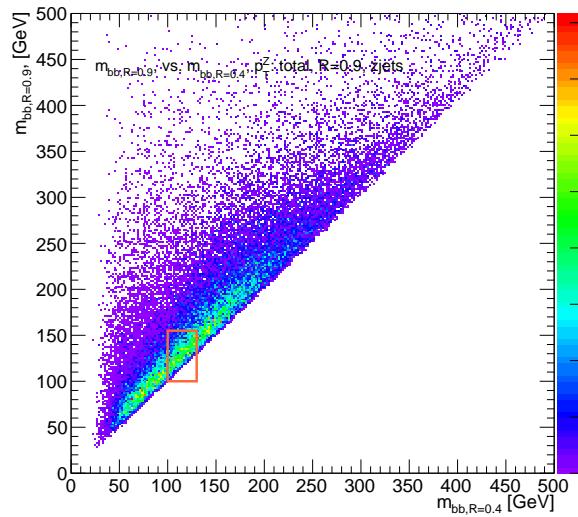
\*Distributions for  $m_{bb}$  at truth and reconstructed level for all telescoping radii studied may be found in Appendix ??



**Figure B.2:** The  $m_{bb}$  distribution for the telescoping jets with  $R = 0.5, 1.0$ , and  $1.5$  truth-level jets is shown for the signal and background samples in (a) and (b), respectively.



(a)



(b)

**Figure B.3:** The 2D distribution of  $m_{bb,R=0.9}$  vs.  $m_{bb,R=0.4}$  is shown for signal and background truth-level samples in (a) and (b), respectively. The region chosen for the double  $m_{bb}$  cut is outlined in orange.

2604 B.6 ERRORS ON TELESCOPING SIGNIFICANCES

2605 Significances of measurements are quoted in units of expected background fluctuations, schemati-  
2606 cally,  $S/\delta B$ . For counting experiments with high numbers of events, we can use Gaussian statistics  
2607 and express this as  $S/\sqrt{B}$ , which we here denote as  $\mathcal{S}$ . However, with lower statistics, it becomes  
2608 more appropriate to use Poissonian statistics, and

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

2609 where  $0.5 + \sqrt{0.25 + B}$  is the characteristic upward fluctuation expected in a Poissonian data set  
2610 using the Pearson chi-square test<sup>66</sup>.

2611 B.7 COUNTING

2612 The significance is given as above, where  $S = N_S$  and  $B = N_B$ . That is, the signal and background  
2613 are just the number of events in signal and background that pass some cuts. The error for the Guas-  
2614 sian case is the standard:

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

2615 The error for the Poissonian case is:

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

2616 where  $\oplus$  denotes addition in quadrature, and  $\Delta S(B)$  is the error on signal (background).

2617 B.8 MULTIPLE EVENT INTERPRETATIONS

2618 Using multiple event interpretations changes the formulae used in with simple counting. That is,  $S$

2619 is not necessarily merely  $N_S$ , the number of events passing some signal cuts, and similarly for  $B$  and

2620  $N_B$ . Using an event weighting by some function of the cut-weight,  $z$ , denoted  $t(z)$ ,  $S = N_S \langle t \rangle_{\rho_S}$

2621 and  $B = N_B \langle t^2 \rangle_{\rho_B}$ . So

$$\mathcal{S}_{t,gaus} = \frac{N_S \langle t \rangle_{\rho_S}}{\sqrt{N_B \langle t^2 \rangle_{\rho_B}}} \rightarrow \mathcal{S}_{t,pois} = \frac{N_S \langle t \rangle_{\rho_S}}{0.5 + \sqrt{0.25 + N_B \langle t^2 \rangle_{\rho_B}}} = \frac{N_S \int_0^1 dz t(z) \rho_S(z)}{0.5 + \sqrt{0.25 + N_B \int_0^1 dz t^2(z) \rho_B(z)}} \quad (B.4)$$

For histograms, everything is done bin-wise. The notation used below is as follows:  $\rho_i$  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,

$$N = \sum_{i=0}^{n_{tel}} \rho_i, \quad \int_0^1 dz t(z) \rho_S(z) = \sum_{i=0}^{n_{tel}} t_i \rho_{S,i}, \quad \int_0^1 dz t^2(z) \rho_B(z) = \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}$$

2622 For the calculations that follow, let  $\xi = \sum_{i=0}^{n_{tel}} t_i \rho_{S,i}$ ,  $\psi = 0.5 + \sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}$ ,

2623  $\partial_S = \frac{\partial}{\partial \rho_{S,i}}$  (and similarly for  $B$ ), so  $\mathcal{S}_t = N_S \xi / \psi$

<sup>2624</sup> Some partial derivatives:

$$\partial_S N_S = 1, \quad \partial_{B,i} N_B = 1$$

$$\partial_S \xi = t_i + (\partial_S t_i) \rho_{S,i}, \quad \partial_B \xi = (\partial_B t_i) \rho_{B,i}$$

$$\begin{aligned}\partial_S \psi &= \frac{N_B t_i (\partial_S t_i) \rho_{B,i}}{\sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}}, & \partial_B \psi &= \frac{\sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i} + N_B (t_i^2 + 2t_i (\partial_B t_i) \rho_{B,i})}{2 \sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \rho_{B,i}}} \\ \partial_S \mathcal{S}_t &= \frac{\xi}{\psi} + \frac{N_S}{\psi} \partial_S \xi - \frac{N_S \xi}{\psi^2} \partial_S \psi, & \partial_B \mathcal{S}_t &= N_S \left( \frac{1}{\psi} \partial_B \xi - \frac{\xi}{\psi^2} \partial_B \psi \right)\end{aligned}$$

<sup>2625</sup> Thus,

$$\Delta \mathcal{S}_{t,i} = \left[ \frac{\xi}{\psi} + \frac{N_S}{\psi} \partial_S \xi - \frac{N_S \xi}{\psi^2} \partial_S \psi \right] \Delta \rho_{S,i} \oplus N_S \left[ \frac{1}{\psi} \partial_B \xi - \frac{\xi}{\psi^2} \partial_B \psi \right] \Delta \rho_{B,i} \quad (\text{B.5})$$

<sup>2626</sup> and the total error is given by the sum in quadrature over all bins  $i$  of  $\Delta \mathcal{S}_{t,i}$ .

<sup>2627</sup> B.9  $t(z) = z$

<sup>2628</sup> With  $t(z) = z$ ,  $t_i = i/n_{tel}$ , so  $\partial_S t_i = \partial_B t_i = 0$ . So:

$$\begin{aligned}\partial_S \psi &= \partial_B \xi = 0 \\ \partial_S \xi &= \frac{i}{n_{tel}} \\ \partial_B \psi &= \frac{\sum_i i^2 \rho_{B,i} + N_B i^2}{n_{tel} \sqrt{n_{tel}^2 + N_B \sum_i i^2 \rho_{B,i}}}\end{aligned}$$

<sup>2629</sup> so  $\Delta\mathcal{S}_{z,i}$  reduces to

$$\Delta\mathcal{S}_{t,i} = \left[ \frac{\xi + N_S t_i}{\psi} \right] \Delta\rho_{S,i} \oplus \left[ \frac{N_S \xi}{\psi^2} \partial_B \psi \right] \Delta\rho_{B,i} \quad (\text{B.6})$$

<sup>2630</sup> B.10  $t(z) = \rho_S(z) / \rho_B(z)$

<sup>2631</sup> With the likelihood optimized<sup>†</sup>  $t^*(z) = \rho_S(z) / \rho_B(z)$ ,  $t_i = \rho_{S,i} / \rho_{B,i}$ , so  $\partial_S t_i = 1 / \rho_{B,i}$  and

<sup>2632</sup>  $\partial_B t_i = -\rho_{S,i} / \rho_{B,i}^2$ . So:

$$\begin{aligned} \partial_S \xi &= 2 \frac{\rho_{S,i}}{\rho_{B,i}} = 2t_i \\ \partial_B \xi &= -\frac{\rho_{S,i}}{\rho_{B,i}} = -t_i \\ \partial_S \psi &= \frac{N_B t_i}{\sqrt{0.25 + N_B \sum_i \rho_{S,i}^2 / \rho_{B,i}}} \\ \partial_B \psi &= \frac{\sum_i \rho_{S,i}^2 / \rho_{B,i} - N_B (\rho_{S,i} / \rho_{B,i})^2}{\sqrt{1 + 4N_B \sum_i \rho_{S,i}^2 / \rho_{B,i}}} \end{aligned}$$

<sup>2633</sup> simplifying somewhat the terms in the per bin error in Equation B.6.

<sup>2634</sup> The new significance figure using multiple event interpretations becomes, with  $\rho_S$  and  $\rho_B$  denot-  
<sup>2635</sup> ing the cut-weight distributions in signal and background, respectively

$$\frac{S}{\delta B} = \frac{N_S \langle t \rangle_{\rho_S}}{0.5 + \sqrt{0.25 + N_B \langle t^2 \rangle_{\rho_B}}} \quad (\text{B.7})$$

---

<sup>†</sup>for the Gaussian statistics case

2636 Of particular interest is the likelihood optimized  $t(z)$ ,<sup>‡</sup>  $t^*(z) = \rho_S(z) / \rho_B(z)$ .  $m_{bb}$  windows are  
 2637 chosen separately for each scheme studied to maximize total significances and are summarized in  
 2638 Table B.3.

$$\left( \frac{S}{\delta B} \right)_z = \frac{N_S \epsilon_S}{0.5 + \sqrt{0.25 + N_B (\epsilon_B^2 + \sigma_B^2)}} \quad (B.8)$$

2639

$$\left( \frac{S}{\delta B} \right)_{t^*(z)} = \frac{N_S \int_0^1 dz \frac{\rho_S^2(z)}{\rho_B(z)}}{0.5 + \sqrt{0.25 + N_B \int_0^1 dz \frac{\rho_S^2(z)}{\rho_B(z)}}} \quad (B.9)$$

2640 where  $\epsilon_{S,B}$  are the means of  $\rho_{S,B}(z)$  and  $\sigma_B^2$  is the variance of  $\rho_B(z)$ . Further details can be found in  
 2641 Refs.<sup>28,39</sup> and Appendix B.6.

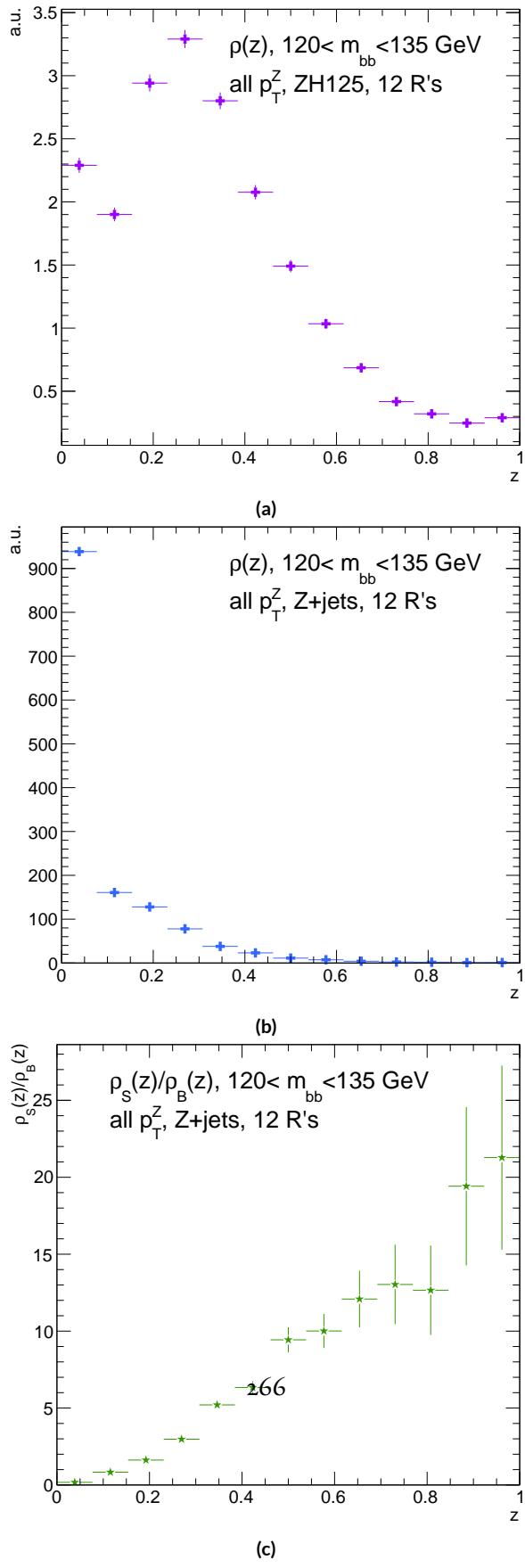
**Table B.3:**  $m_{bb}$  windows studied. These windows were chosen to optimize significances over all  $p_T^Z$ .

Analysis Type	$S/\delta B$ Type	Optimal $m_{bb}$ Window
Reconstructed	$anti-k_t R = 0.4$ $t(z) = z$ $t(z) = \rho_S(z) / \rho_B(z)$ $anti-k_t R = 0.4$ , telescoping $R = 0.6$	$90-140$ GeV $110-155$ GeV $110-155$ GeV $95-140$ GeV ( $R = 0.4$ ), $105-160$ GeV ( $R = 0.6$ )
Truth	$anti-k_t R = 0.4$ $t(z) = z$ $t(z) = \rho_S(z) / \rho_B(z)$ $anti-k_t R = 0.4$ , telescoping $R = 0.9$	$100-130$ GeV $115-140$ GeV $120-135$ GeV $100-130$ GeV ( $R = 0.4$ ), $100-155$ GeV ( $R = 0.9$ )

2642 The truth-level distributions  $\rho_S(z)$ ,  $\rho_B(z)$ , and  $\rho_S(z) / \rho_B(z)$  are shown for the  $m_{bb}$  window that  
 2643 optimizes  $(S/\delta B)_{t^*(z)}$  in Figure B.4, and significance improvements as a function of  $p_T^Z$  are summa-  
 2644 rized in Figure B.5. Uncertainties in Figures B.5 and B.9 are statistical uncertainties. JES systematics

---

<sup>‡</sup>Derived under the assumption of Gaussian statistics in Ref<sup>39</sup>



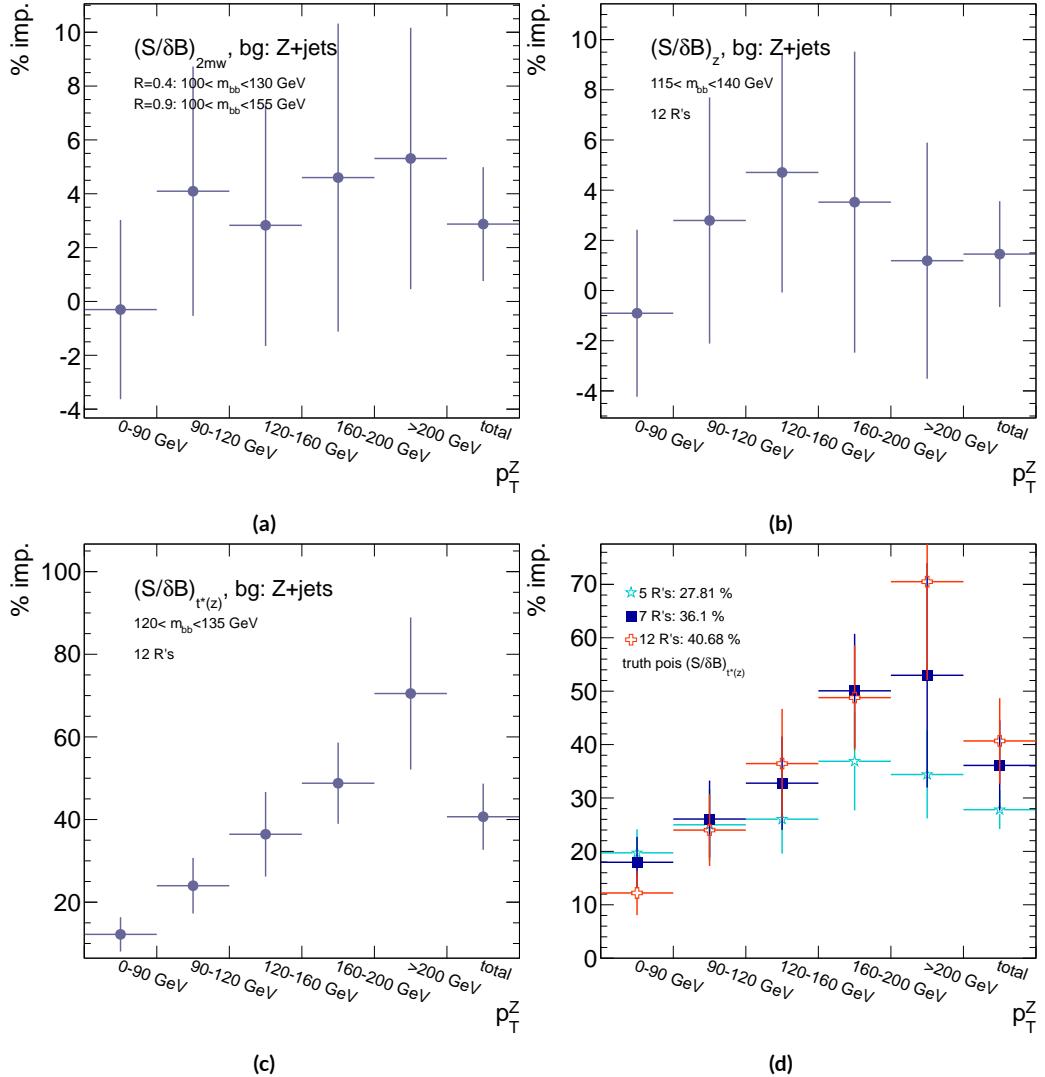
**Figure B.4:** Truth-level  $\rho(z)$  distributions for the  $m_{bb}$  window optimizing  $(S/\delta B)_{t^*(z)} \cdot \rho_S(z)$  for the signal  $ZH125$  sample is shown in (a), and  $\rho_B(z)$  for the background  $Z+jets$  sample is shown in (b). The distribution of  $\rho_S(z) / \rho_B(z)$

will need to be evaluated for different  $R$ 's, as modeling uncertainties is an outstanding issue, but  
 these systematics will likely be strongly correlated for the different  $R$ 's and are not anticipated to be  
 a very large contribution to total uncertainties. While the two dimensional  $m_{bb}$  cut and  $t(z) = z$   
 schemes only showed marginal improvement at truth level at 2.87%<sup>§</sup> and 1.45%, respectively, the  
 likelihood optimized  $t^*(z)$  showed a more substantial 40.7% improvement overall, with a steady in-  
 crease in improvement with increasing  $p_T^Z$ . Figure B.5 (d) summarizes the improvements with respect  
 to  $p_T^Z$  for the  $t^*(z)$  event weight for five, seven, and twelve telescoping radii (interpretations) per  
 event. Improvements increase with a greater number of interpretations and are more pronounced at  
 higher  $p_T^Z$  for this scheme.<sup>¶</sup> 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  $\rho_S/\rho_B$  is the relevant  
 quantity (as shown in Figure B.4 (c)).

---

<sup>§</sup>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.<sup>31</sup>.

<sup>¶</sup>This is not the case for the  $z$  event weight, which is more thoroughly treated in Section ??.



**Figure B.5:** A summary of the improvements for different truth-level telescoping jet cuts and weights is shown in bins of  $p_T^Z$ . The final bin is the total improvement over all  $p_T^Z$ . Shown are improvements for the 2D  $m_{bb}$  cut (a),  $t(z) = z$  (b),  $t(z) = t^*(z)$  with 12 radii (c), and  $t(z) = t^*(z)$  for various radii (d).

2660 B.II RECONSTRUCTED-LEVEL ANALYSIS

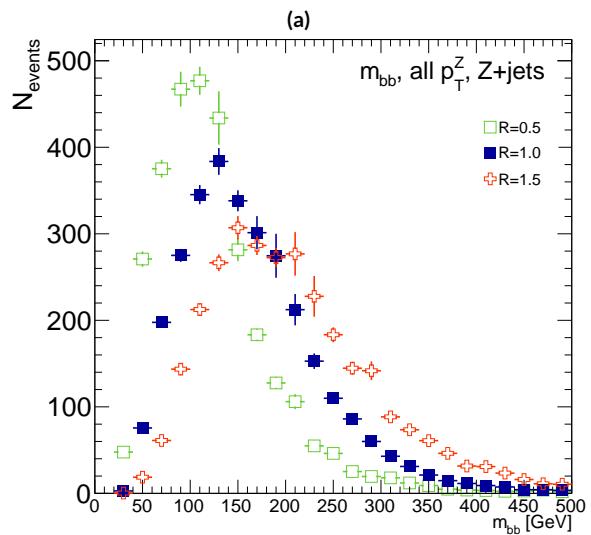
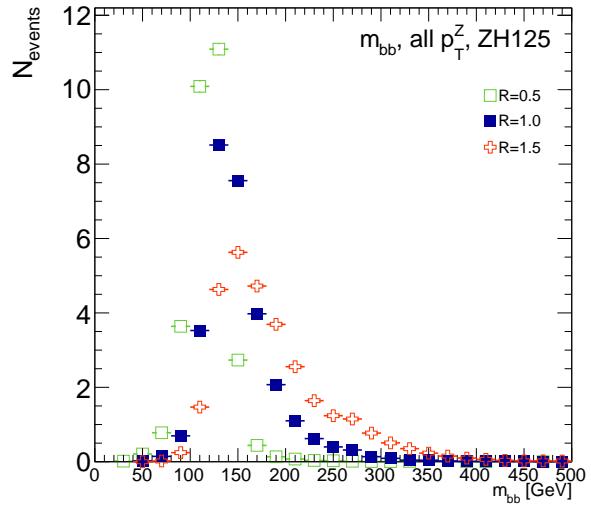
2661 At reconstructed level, the same overall effect of introducing a high tail in  $m_{bb}$  distributions with  
2662 increasing  $R$  is evident in comparing Figures B.2 and B.6. The optimal  $m_{bb}$  windows, however, grow  
2663 larger, due to the lack of truth-level information.

2664 Total significance gains at reconstructed level for the two dimensional  $m_{bb}$  cut and the  $t(z) = z$   
2665 case are similar, at 2.87% and 1.45%, respectively. The optimal two-dimensional  $m_{bb}$  cut at recon-  
2666 structed level is  $95 < m_{bb,R=0.4} < 140$  GeV,  $105 < m_{bb,R=0.6} < 160$  GeV. Just as at truth level, the  
2667  $R = 0.4$   $m_{bb}$  cut is comparable to the optimal single  $R = 0.4$   $m_{bb}$  cut, and the second  $m_{bb}$  cut is at  
2668 similar values (cf. Table B.3 and Figures B.3 and B.7). However, the optimal second telescoping ra-  
2669 dius is markedly smaller at  $R = 0.6$  versus the optimal truth-level second radius of  $R = 0.9$ , which  
2670 suggests that effects like pileup at reconstructed level obscure correlations between the  $R = 0.4$   
2671 interpretations and limit the usefulness of larger  $R$  interpretations in this particular scheme. The  
2672  $t(z) = z$  case has a wider optimal window and yields about half the improvement it does at truth  
2673 level.<sup>||</sup>

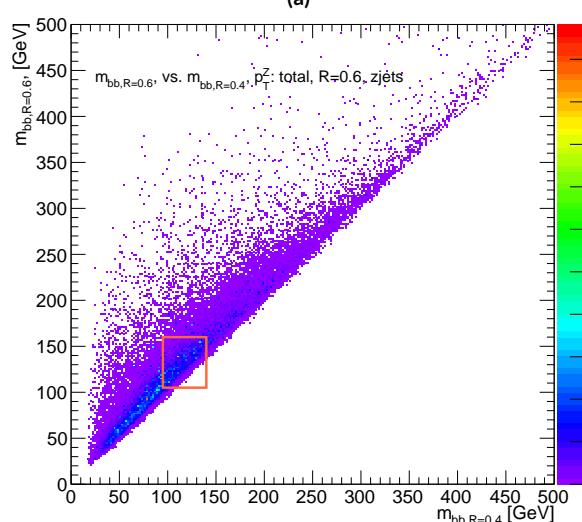
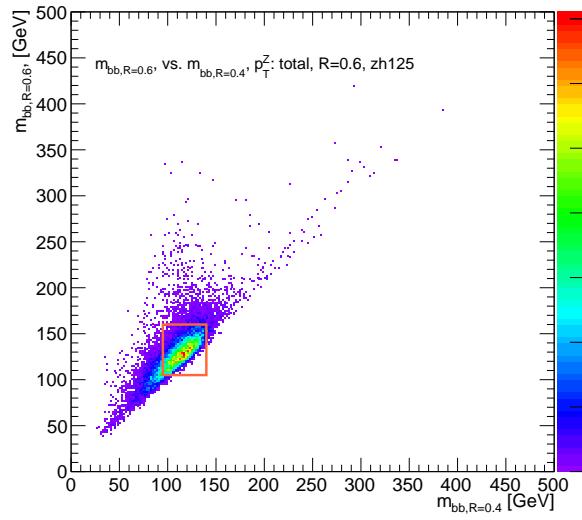
2674 The optimal  $m_{bb}$  window for the  $t^*(z)$  case is also markedly wider at reconstructed level, at  
2675  $110 < m_{bb} < 155$  GeV in comparison to the truth-level optimal  $120 < m_{bb} < 135$  GeV. The  $\rho(z)$  dis-  
2676 tributions for the signal  $ZH_{125}$  and background  $Z + \text{jets}$  as well as the  $\rho_S(z) / \rho_B(z)$  in this window  
2677 are shown in Figure B.8. Compared with the truth-level distributions in Figure B.4, both the signal  
2678 and background optimal  $\rho(z)$  distributions have higher values at higher  $z$ . The peak in  $\rho_S(z)$  at

---

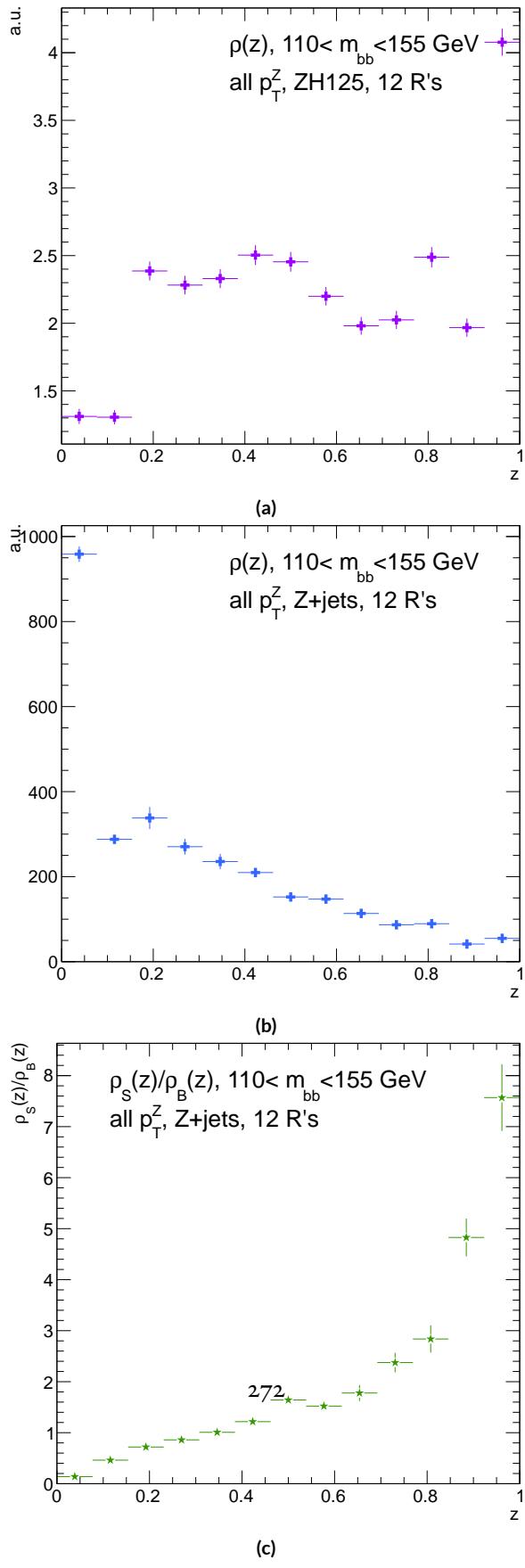
<sup>||</sup>A fuller treatment of this scheme is given in Section ??.



**Figure B.6:** The  $m_{bb}$  distribution for the telescoping jets with  $R = 0.5$ ,  $R = 1.0$ , and  $R = 1.5$  reconstructed-level jets is shown for the signal and background samples in (a) and (b), respectively.



**Figure B.7:** The 2D distribution of  $m_{bb,R=0.8}$  vs.  $m_{bb,R=0.4}$  is shown for signal and background reconstructed-level samples in (a) and (b), respectively. The region chosen for the double  $m_{bb}$  cut is outlined in orange.

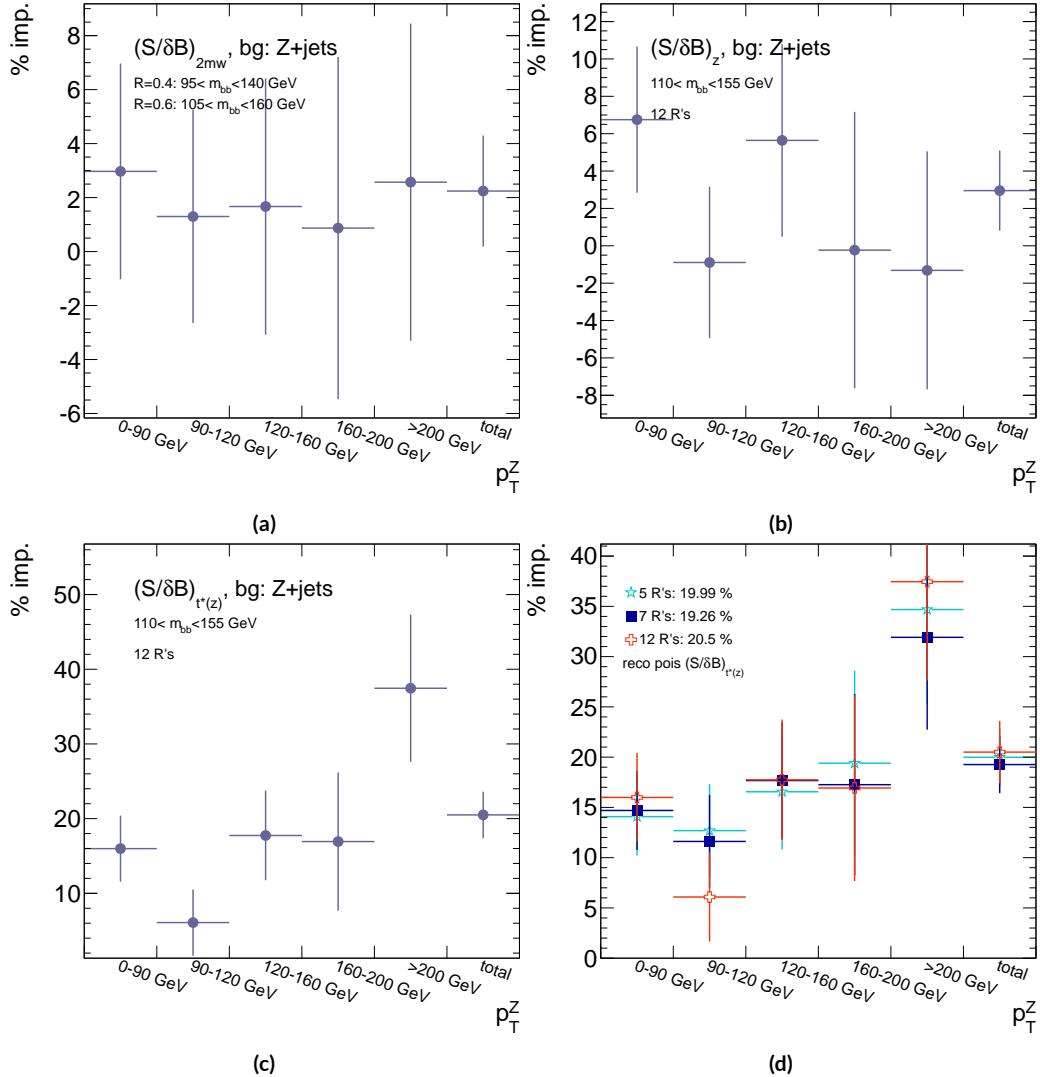


**Figure B.8:** Reconstructed-level  $\rho(z)$  distributions for the  $m_{bb}$  window optimizing  $(S/\delta B)_{t^*(z)} \cdot \rho_S(z)$  for the signal  $ZH125$  sample is shown in (a), and  $\rho_B(z)$  for the background  $Z+\text{jets}$  sample is shown in (b). The distribution of  $\rho_s(z)/\rho_B(z)$  for these samples is shown in (c).

2679  $z = 1$  suggests that at reconstructed level, maximizing the number of more “signal-like” events is the  
 2680 key to optimizing significances, as opposed to the optimal, background suppressing  $\rho(z)$  distribu-  
 2681 tions at truth level. The use of a greater number of interpretations per event (telescoping radii) does  
 2682 appear to result in overall greater improvement as at truth level, as twelve radii performed better  
 2683 than five, but this is less clear at reconstructed level, as shown in Figure B.9 (d). The improvement at  
 2684 reconstructed level using an event weight of  $t^*(z)$  is 20.5%, just over half the improvement at truth  
 2685 level but still quite significant. Summaries of improvements as a function of  $p_T^Z$  for all three cases  
 2686 studied and for the  $t^*(z)$  case for different numbers of telescoping radii are shown in Figure B.9.

**Table B.4:** A summary of significances for different weighting schemes and cuts and for reconstructed and truth jets for a luminosity of  $20.3 \text{ fb}^{-1}$ .

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).

2687 **B.12 CONCLUSIONS AND PROSPECTS**

2688 The use of telescoping jets to provide multiple event interpretations shows promise as an avenue to  
2689 increase significances in the  $H \rightarrow b\bar{b}$  search in ATLAS and make an observation in the systematics-  
2690 limited environment of early Run 2. A preliminary study using the telescoping jets algorithm with  
2691 12 telescoping radii to build 12 event interpretations on 2012 Monte Carlo based on the full, cut-  
2692 based Run 1 analysis yielded a 20.5% improvement in  $S/\delta B$  over using anti- $k_t$  with  $R = 0.4$  alone  
2693 at reconstructed level using a likelihood maximized event weighting to study the  $ZH \rightarrow llb\bar{b}$  pro-  
2694 cess. The jets used in this note at reconstructed level were trimmed in order to guarantee reasonable  
2695 resolution in the large- $R$  interpretations. The algorithm, in particular, showed discriminating power  
2696 at high  $p_T^Z$ , so better performance can be expected in Run 2 with a higher  $\sqrt{s}$  and higher numbers  
2697 of events with large  $p_T^Z$ . Additionally, the many simplifying assumptions regarding jet calibration  
2698 and the relatively basic use of information<sup>\*\*</sup> from multiple invariant masses in this note suggest  
2699 that even further improvements than those quoted are possible. While this note did not explore  
2700 the correlations between multiple event interpretations and the variables used in the BDT of the  
2701 latest multivariate version of the  $H \rightarrow b\bar{b}$  analysis<sup>13</sup>, new phenomenological studies suggest that  
2702 such correlations are not strong<sup>31</sup>. The corresponding reconstructed-level study, using a BDT, is left  
2703 for future work. Also left for future work are better understanding the effects of jet trimming and  
2704 which interpretations are the most useful.

---

2704 \*\*For examples of more sophisticated treatments compared to the treatment in this note, see Ref<sup>31</sup>.

# References

2705

- 2706 [1] (2012a). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the  
2707 LHC. *Phys.Lett.*, B716, 30–61.
- 2708 [2] (2012b). Observation of a new particle in the search for the Standard Model Higgs boson  
2709 with the ATLAS detector at the LHC. *Phys.Lett.*, B716, 1–29.
- 2710 [3] (2012). The Proton Synchrotron.
- 2711 [4] (2012). The Proton Synchrotron Booster.
- 2712 [5] (2012). The Super Proton Synchrotron.
- 2713 [6] (2014). *ATLAS Run 1 Pythia8 tunes*. Technical Report ATL-PHYS-PUB-2014-021, CERN,  
2714 Geneva.
- 2715 [7] (2014). *Electron efficiency measurements with the ATLAS detector using the 2012 LHC*  
2716 *proton-proton collision data*. Technical Report ATLAS-CONF-2014-032, CERN, Geneva.
- 2717 [8] (2015). *Expected performance of the ATLAS b-tagging algorithms in Run-2*. Technical Report  
2718 ATL-PHYS-PUB-2015-022, CERN, Geneva.
- 2719 [9] (2015). *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS*  
2720 *Detector at  $\sqrt{s} = 13 \text{ TeV}$* . Technical Report ATL-PHYS-PUB-2015-015, CERN, Geneva.
- 2721 [10] (2015). *Muon reconstruction performance in early  $\sqrt{s}=13 \text{ TeV}$  data*. Technical Report ATL-  
2722 PHYS-PUB-2015-037, CERN, Geneva.
- 2723 [11] A. L. Read (2002). Presentation of search results: the  $CL_s$  technique. *J. Phys. G*, 28, 2693–  
2724 2704.
- 2725 [12] Aad, G. et al. (2014). Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in  
2726  $pp$  collisions at  $\sqrt{s} = 7 \text{ TeV}$  with the ATLAS detector. *JHEP*, 09, 145.

- 2727 [13] Ahmadov, F., Alio, L., Allbrooke, B., Bristow, T., Buescher, D., Buzatu, A., Coadou, Y.,  
 2728 Debenedetti, C., Enari, Y., Facini, G., Fisher, W., Francavilla, P., Gaycken, G., Gentil, J.,  
 2729 Goncalo, R., Gonzalez Parra, G., Grivaz, J., Gwilliam, C., Hageboeck, S., Halladjian, G., Jack-  
 2730 son, M., Jamin, D., Jansky, R., Kiuchi, K., Kostyukhin, V., Lohwasser, K., & Lopez Mateos,  
 2731 D, e. a. (2014). *Supporting Document for the Search for the bb decay of the Standard Model*  
 2732 *Higgs boson in associated (W/Z)H production with the ATLAS detector*. Technical Report  
 2733 ATL-COM-PHYS-2014-051, CERN, Geneva.
- 2734 [14] Aliev, M., Lacker, H., Langenfeld, U., Moch, S., Uwer, P., et al. (2011). HATHOR: Hadronic  
 2735 Top and Heavy quarks cross section calculator. *Comput.Phys.Commun.*, 182, 1034–1046.
- 2736 [15] Alwall, J., Frederix, R., Frixione, S., Hirschi, V., Maltoni, F., Mattelaer, O., Shao, H. S.,  
 2737 Stelzer, T., Torrielli, P., & Zaro, M. (2014). The automated computation of tree-level and  
 2738 next-to-leading order differential cross sections, and their matching to parton shower simula-  
 2739 tions. *JHEP*, 07, 079.
- 2740 [16] Alwall, J., Herquet, M., Maltoni, F., Mattelaer, O., & Stelzer, T. (2011). Madgraph 5 : Going  
 2741 beyond.
- 2742 [17] ATLAS Collaboration (2014). *Tagging and suppression of pileup jets with the ATLAS detec-*  
 2743 *tor*. Technical Report ATLAS-CONF-2014-018, CERN, Geneva.
- 2744 [18] ATLAS Collaboration (2015a). MCPAnalysisGuidelinesMC15.  
 2745 <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCPAnalysisGuidelinesMC15>.
- 2746 [19] ATLAS Collaboration (2015b). *Performance of missing transverse momentum reconstruction*  
 2747 *for the ATLAS detector in the first proton-proton collisions at at  $\sqrt{s} = 13$  TeV*. Technical  
 2748 Report ATL-PHYS-PUB-2015-027, CERN, Geneva.
- 2749 [20] Ball, R. D. et al. (2013). Parton distributions with LHC data. *Nucl. Phys.*, B867, 244–289.
- 2750 [21] Ball, R. D. et al. (2015). Parton distributions for the LHC Run II. *JHEP*, 04, 040.
- 2751 [22] Botje, M. et al. (2011). The PDF4LHC Working Group Interim Recommendations.
- 2752 [23] Buckley, A., Butterworth, J., Grellscheid, D., Hoeth, H., Lonnblad, L., Monk, J., Schulz, H.,  
 2753 & Siegert, F. (2010). Rivet user manual.

- 2754 [24] Buzatu, A. & Wang, W. (2016). *Object selections for SM Higgs boson produced in associa-*  
 2755 *tion with a vector boson in which  $H \rightarrow b\bar{b}$  and  $V$  decays leptonically with Run-2 data: Object*  
 2756 *support note for  $VH(b\bar{b})$  2015+2016 dataset publication.* Technical Report ATL-COM-PHYS-  
 2757 2016-1674, CERN, Geneva. This is a support note for the  $VH(b\bar{b})$  SM publication using the  
 2758 2015+2016 datasets.
- 2759 [25] Campbell, J. M. & Ellis, R. K. (2010). MCFM for the Tevatron and the LHC. *Nucl. Phys.*  
 2760 *Proc. Suppl.*, 205-206, 10–15.
- 2761 [26] Capeans, M., Darbo, G., Einsweiller, K., Elsing, M., Flick, T., Garcia-Sciveres, M., Gemme,  
 2762 C., Pernegger, H., Rohne, O., & Vuillermet, R. (2010). *ATLAS Insertable B-Layer Technical*  
 2763 *Design Report.* Technical Report CERN-LHCC-2010-013. ATLAS-TDR-19.
- 2764 [27] CERN (2008). LHC first beam: a day to remember.
- 2765 [28] Chan, S., Huth, J., Lopez Mateos, D., & Mercurio, K. (2015a). *ZH → llb̄ Analysis with*  
 2766 *Telescoping Jets.* Technical Report ATL-PHYS-INT-2015-002, CERN, Geneva.
- 2767 [29] Chan, S. K.-w. & Huth, J. (2017). *Variations on MVA Variables in the SM ZH → llb̄*  
 2768 *Search.* Technical Report ATL-COM-PHYS-2017-1318, CERN, Geneva.
- 2769 [30] Chan, S. K.-w., Lopez Mateos, D., & Huth, J. (2015b). *Micromegas Trigger Processor Al-*  
 2770 *gorithm Performance in Nominal, Misaligned, and Corrected Misalignment Conditions.*  
 2771 Technical Report ATL-COM-UPGRADE-2015-033, CERN, Geneva.
- 2772 [31] Chien, Y.-T., Farhi, D., Krohn, D., Marantan, A., Mateos, D. L., & Schwartz, M. (2014).  
 2773 Quantifying the power of multiple event interpretations.
- 2774 [32] Ciccolini, M., Dittmaier, S., & Kramer, M. (2003). Electroweak radiative corrections to  
 2775 associated WH and ZH production at hadron colliders. *Phys. Rev.*, D68, 073003.
- 2776 [33] Clark, B., Lopez Mateos, D., Felt, N., Huth, J., & Oliver, J. (2014). *An Algorithm for Mi-*  
 2777 *cromegas Segment Reconstruction in the Level-1 Trigger of the New Small Wheel.* Technical  
 2778 Report ATL-UPGRADE-INT-2014-001, CERN, Geneva.
- 2779 [34] Collaboration, A. (2017a). Evidence for the  $h \rightarrow b\bar{b}$  decay with the atlas detector.
- 2780 [35] Collaboration, A. (2017b). Study of the material of the atlas inner detector for run 2 of the  
 2781 lhc.

- 2782 [36] Collaboration, T. A., Aad, G., Abat, E., Abdallah, J., & A A Abdelalim, e. a. (2008). The atlas  
2783 experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08), S08003.
- 2784 [37] Cowan, G., Cranmer, K., Gross, E., & Vitells, O. (2011). Asymptotic formulae for likelihood-  
2785 based tests of new physics. *Eur. Phys. J. C*, 71, 1554.
- 2786 [38] Czakon, M., Fiedler, P., & Mitov, A. (2013). Total Top-Quark Pair-Production Cross Section  
2787 at Hadron Colliders Through  $\alpha(\frac{4}{S})$ . *Phys. Rev. Lett.*, 110, 252004.
- 2788 [39] D. Kahawala, D. Kahawala, D. K. & Schwartz, M. D. (2013). Jet sampling: Improving event  
2789 reconstruction through multiple interpretations.
- 2790 [40] D. Krohn, D. Krohn, J. T. L. W. (2009). Jet trimming.
- 2791 [41] Delmastro, M., Gleyzer, S., Hengler, C., Jimenez, M., Koffas, T., Kuna, M., Liu, K., Liu, Y.,  
2792 Marchiori, G., Petit, E., Pitt, M., Soldatov, E., & Tackmann, K. (2014). *Photon identification*  
2793 *efficiency measurements with the ATLAS detector using LHC Run 1 data*. Technical Report  
2794 ATL-COM-PHYS-2014-949, CERN, Geneva.
- 2795 [42] Evans, L. & Bryant, P. (2008). Lhc machine. *Journal of Instrumentation*, 3(08), S08001.
- 2796 [43] Gleisberg, T. et al. (2009a). Event generation with SHERPA 1.1. *JHEP*, 02, 007.
- 2797 [44] Gleisberg, T., Höche, S., Krauss, F., Schönherr, M., Schumann, S., Siegert, F., & Winter, J.  
2798 (2009b). Event generation with sherpa 1.1. *Journal of High Energy Physics*, 2009(02), 007.
- 2799 [45] Hagebock, S. (CERN, Geneva, 2017). Lorentz Invariant Observables for Measurements of  
2800 Hbb Decays with ATLS.
- 2801 [46] Heinemann, B., Hirsch, F., & Strandberg, S. (2010). *Performance of the ATLAS Secondary*  
2802 *Vertex b-tagging Algorithm in 7 TeV Collision Data*. Technical Report ATLAS-COM-CONF-  
2803 2010-042, CERN, Geneva. (Was originally 'ATL-COM-PHYS-2010-274').
- 2804 [47] Jackson, P. & Rogan, C. (2017). Recursive jigsaw reconstruction: Hep event analysis in the  
2805 presence of kinematic and combinatoric ambiguities.
- 2806 [48] Kant, P., Kind, O., Kintscher, T., Lohse, T., Martini, T., Molbitz, S., Rieck, P., & Uwer,  
2807 P. (2015). Hathor for single top-quark production: Updated predictions and uncertainty  
2808 estimates for single top-quark production in hadronic collisions. *Computer Physics Commu-*  
2809 *nifications*, 191, 74 – 89.

- 2810 [49] Lampl, W., Laplace, S., Lelas, D., Loch, P., Ma, H., Menke, S., Rajagopalan, S., Rousseau,  
 2811 D., Snyder, S., & Unal, G. (2008). *Calorimeter Clustering Algorithms: Description and*  
 2812 *Performance*. Technical Report ATL-LARG-PUB-2008-002, ATL-COM-LARG-2008-003,  
 2813 CERN, Geneva.
- 2814 [50] Lavesson, N. & Lonnblad, L. (2005). W+jets matrix elements and the dipole cascade.
- 2815 [51] LHC Higgs Cross Section Working Group, Dittmaier, S., Mariotti, C., Passarino, G., &  
 2816 Tanaka (Eds.), R. (CERN, Geneva, 2011). Handbook of LHC Higgs Cross Sections: 1. In-  
 2817 clusiv Observables. *CERN-2011-002*.
- 2818 [52] Loch, Peter and Lefebvre, Michel (2007). Introduction to Hadronic Calibration in ATLAS.
- 2819 [53] Luisoni, G., Nason, P., Oleari, C., & Tramontano, F. (2013).  $H_W \pm/hz + o$  and 1 jet at nlo  
 2820 with the powheg box interfaced to gosam and their merging within minlo. *Journal of High*  
 2821 *Energy Physics*, 2013(10), 83.
- 2822 [54] M. Cacciari, M. Cacciari, G. P. S. & Soyez, G. (2008). The anti- $k_t$  jet clustering algorithm.  
 2823 *JHEP*, 0804.
- 2824 [55] Marcastel, F. (2013). CERN's Accelerator Complex. La chaîne des accélérateurs du CERN.  
 2825 General Photo.
- 2826 [56] Masubuchi, T., Benitez, J., Bell, A. S., Argyropoulos, S., Arnold, H., Amaral Coutinho, Y.,  
 2827 Sanchez Pineda, A. R., Buzatu, A., Calderini, G., & Chan, Stephen Kam-wah, e. a. (2016).  
 2828 *Search for a Standard Model Higgs boson produced in association with a vector boson and de-*  
 2829 *caying to a pair of b-quarks*. Technical Report ATL-COM-PHYS-2016-1724, CERN, Geneva.
- 2830 [57] Patrignani, C. et al. (2016). Review of Particle Physics. *Chin. Phys.*, C40(10), 100001.
- 2831 [58] Robson, A., Piacquadio, G., & Schopf, E. (2016). *Signal and Background Modelling Stud-*  
 2832 *ies for the Standard Model VH,  $H \rightarrow b\bar{b}$  and Related Searches: Modelling support note*  
 2833 *for VH(bb) 2015+2016 dataset publication*. Technical Report ATL-COM-PHYS-2016-1747,  
 2834 CERN, Geneva. This is a support note for the VH(bb) SM publication using the 2015+2016  
 2835 datasets.
- 2836 [59] S. Alioli et al. (2009). NLO Higgs boson production via gluon fusion matched with shower  
 2837 in POWHEG. *JHEP*, 0904, 002.

- 2838 [60] S. D. Ellis, S. D. Ellis, A. H. T. S. R. D. K. & Schwartz, M. D. (2012). Qjets: A non-  
2839 deterministic approach to tree-based jet substructure. *Phys. Rev. Lett.*, 108.
- 2840 [61] Salam, G. P. (2009). Towards jetography.
- 2841 [62] Sjostrand, T., Mrenna, S., & Skands, P. Z. (2008). A Brief Introduction to PYTHIA 8.1.  
2842 *Comput.Phys.Commun.*, 178, 852–867.
- 2843 [63] Stancari, G., Previtali, V., Valishev, A., Bruce, R., Redaelli, S., Rossi, A., & Salvachua Fer-  
2844 rando, B. (2014). *Conceptual design of hollow electron lenses for beam halo control in the*  
2845 *Large Hadron Collider*. Technical Report FERMILAB-TM-2572-APC. FERMILAB-TM-  
2846 2572-APC, CERN, Geneva. Comments: 23 pages, 1 table, 10 figures.
- 2847 [64] Stewart, I. W. & Tackmann, F. J. (2011). Theory uncertainties for higgs and other searches  
2848 using jet bins.
- 2849 [65] Symmetry Magazine (2015). The Standard Model of Particle Physics.
- 2850 [66] Verkerke, W. & Kirkby, D. (2003). The RooFit toolkit for data modeling. In *2003 Computing*  
2851 *in High Energy and Nuclear Physics, CHEP03*.
- 2852 [67] Watts, G., Filthaut, F., & Piacquadio, G. (2015). *Extrapolating Errors for b-tagging*. Technical  
2853 Report ATL-COM-PHYS-2015-711, CERN, Geneva. This is for internal information only, no  
2854 approval to ever be seen outside of ATLAS.
- 2855 [68] Y. L. Dokshitzer, Y. L. Dokshitzer, G. D. L. S. M. & Webber, B. R. (1997). Better jet cluster-  
2856 ing algorithms. *JHEP*, 9708.