

# Some Title Involving $ZH \rightarrow ll\bar{b}\bar{b}$

A DISSERTATION PRESENTED  
BY  
STEPHEN K. CHAN  
TO  
THE DEPARTMENT OF PHYSICS

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

HARVARD UNIVERSITY  
CAMBRIDGE, MASSACHUSETTS  
NOVEMBER 2017

©2017 – STEPHEN K. CHAN  
ALL RIGHTS RESERVED.

# Some Title Involving $ZH \rightarrow ll\bar{b}\bar{b}$

## ABSTRACT

The Higgs looks more Standard Model by the day. The bulk of this thesis is vomiting up what amount to book reports of the main analysis documents, a technical paper, and a RI quality theory steak haché.

For a “unique intellectual contribution,” I made three different BDT’s and went shake and bake to what I’m sure will be a set of inconclusive results of dubious actual scientific value.

If you insist, though:

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

Thesis advisor: Professor John Huth

Stephen K. Chan

# Contents

0 INTRODUCTION	I
1 THEORY	4
2 THE LARGE HADRON COLLIDER AND THE ATLAS DETECTOR	5
3 DATA AND SIMULATED SAMPLES	7
4 OBJECT DEFINITIONS AND EVENT SELECTION	9
4.1 Object Definitions . . . . .	9
4.2 Event Selection and Analysis Regions . . . . .	10
5 SIGNAL AND BACKGROUND MODELING	13
5.1 Signal Processes . . . . .	14
5.2 Background . . . . .	16
5.3 Notes . . . . .	21
6 EXPERIMENTAL SYSTEMATIC UNCERTAINTIES	28
7 MULTIVARIATE ANALYSIS	29
7.1 Training Samples and Variables . . . . .	29
7.2 Lorentz Invariants . . . . .	30
7.3 RestFrames Variables . . . . .	32
7.4 Correlations . . . . .	33
7.5 MVA Training . . . . .	34
7.6 Statistics Only BDT Performance . . . . .	37
8 STATISTICAL FIT MODEL AND VALIDATION	49
8.1 The Fit Model . . . . .	49
8.2 Fit Inputs . . . . .	51
8.3 Systematic Uncertainties . . . . .	52
8.4 Full Breakdown of Errors . . . . .	54
8.5 S/B Plot . . . . .	57
8.6 Postfit Distributions . . . . .	57
8.7 Nuisance Parameter Pulls . . . . .	57
8.8 Nuisance Parameter Correlations . . . . .	66

8.9	Summary of Results . . . . .	67
8.10	$z$ and $\geq 3$ Jet Fits . . . . .	69
8.11	Nuisance Parameter Ranking Plots and Breakdowns . . . . .	72
8.12	Postfit Distributions . . . . .	75
8.13	Nuisance Parameter Pulls . . . . .	81
8.14	Nuisance Parameter Correlations . . . . .	84
9	FIT RESULTS	86
10	CONCLUSIONS	90
APPENDIX A MICROMEGAS TRIGGER PROCESSOR SIMULATION		93
A.1	Algorithm Overview . . . . .	94
A.2	Monte Carlo Samples . . . . .	99
A.3	Nominal Performance . . . . .	99
A.4	Fit Quantities . . . . .	100
A.5	Efficiencies . . . . .	106
A.6	Incoherent Background . . . . .	109
A.7	Charge Threshold . . . . .	114
A.8	Misalignments and Corrections . . . . .	117
A.9	Simulation Correction of the Algorithm Under Nominal Conditions . . . . .	127
A.10	Translation Misalignments Along the Horizontal Strip Direction ( $\Delta s$ ) . . . . .	130
A.11	Translation Misalignments Orthogonal to the Beamline and Horizontal Strip Direction ( $\Delta z$ ) . . . . .	132
A.12	Translation Misalignments Parallel to the Beamline ( $\Delta t$ ) . . . . .	134
A.13	Chamber Tilts Towards and Away from the IP ( $\gamma_s$ Rotation) . . . . .	135
A.14	Rotation Misalignments Around the Wedge Vertical Axis ( $\beta_z$ ) . . . . .	137
A.15	Rotation Misalignments Around the Axis Parallel to the Beamline ( $\alpha_t$ ) . . . . .	138
A.16	Conclusion . . . . .	140
APPENDIX B TELESCOPING JETS		142
B.1	Monte Carlo Simulation . . . . .	143
B.2	Jet Reconstruction and Calibration . . . . .	144
B.3	Event Reconstruction and Selection . . . . .	146
B.4	Validation of Jet Calibration . . . . .	147
B.5	Truth-Level Analysis . . . . .	151
B.6	Errors on Telescoping Significances . . . . .	154
B.7	Counting . . . . .	154
B.8	Multiple Event Interpretations . . . . .	155
B.9	$t(z) = z$ . . . . .	156
B.10	$t(z) = \rho_S(z) / \rho_B(z)$ . . . . .	157
B.11	Reconstructed-Level Analysis . . . . .	162

B.12 Conclusions and Prospects . . . . .	168
REFERENCES	170
REFERENCES	173

THIS IS THE DEDICATION.

# Acknowledgments

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

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

Conventional Wisdom

# 0

## Introduction

MUCH HAS BEEN SAID Since the discovery of a Standard Model (SM) like Higgs boson at the LHC in 2012<sup>?</sup>, one of the main outstanding physics goals of the LHC has been to observe the primary SM Higgs decay mode,  $H \rightarrow b\bar{b}$ , with efforts primarily targeted at searching for Higgs bosons produced in association with a leptonically decaying vector ( $W$  or  $Z$ , denoted generically as  $V$ ) boson. As the

integrated luminosity of data collected at the LHC increases,  $H \rightarrow b\bar{b}$  searches will increasingly become limited by the ability to constrain systematic uncertainties, with the latest result from ATLAS at  $\sqrt{s} = 13$  TeV using  $36.1 \text{ fb}^{-1}$  of  $pp$  collision data already approaching this regime, having a  $VH(b\bar{b})$  signal strength of  $1.20^{+0.24}_{-0.23}(\text{stat.})^{+0.34}_{-0.28}(\text{syst.})$  at  $m_H = 125 \text{ GeV}$ <sup>31</sup>.

While this effort will likely require a combination of several different methods at various different stages in the analysis chain, one possible avenue forward is to revise the multivariate analysis (MVA) discriminant input variables used. Novel variable sets often promise to increase performance in two ways. The first is by having higher descriptive power, often through some sophisticated treatment of the missing transverse energy in an event,  $E_T^{\text{miss}}$ . The second is through using a more orthogonal basis of description, which allows one to more efficiently use data and simulation samples.<sup>\*</sup> This set of studies will seek to address the latter issue.

In order to largely factor out the first issue, gains from better treatments of  $E_T^{\text{miss}}$ , a closed final state, the 2-lepton  $ZH \rightarrow \ell\ell b\bar{b}$  channel, will be studied here in an analysis that very closely mirrors the approach in<sup>31</sup> (henceforth referred to as the “fiducial analysis”). In addition to the standard variable set considered there, two additional variable sets, the “Lorentz Invariant” (LI)<sup>?</sup> and “Rest-Frames inspired” (RF) variable<sup>?</sup> are also studied.

Data and simulation samples used are described in Section ??, and event reconstruction definitions and event selection requirements are outlined in Section ???. The multivariate analysis, includ-

---

<sup>\*</sup>Heuristically, the more orthogonal one’s basis, the less overlapping information the variables contain, and the more efficiently something like a numerical minimization can proceed. Hence, even if the physical likelihood that something like an MVA is approximating has the same discriminating power for two variable sets in a stats only context, a more orthogonal basis can allow for a more efficient exploration of the extra dimensionality added through systematic uncertainty terms in a typical analysis, mitigating the usual broadening and smearing of the likelihood from systematics and reducing errors on fit quantities of interest.

ing a description of the LI and RF variable sets and a summary of performance in the absence of systematic uncertainties, is described in Section ???. The statistical fit model and systematic uncertainties are described in Section ??, and the fit results may be found in Section ???. Finally, conclusions and discussion are presented in Section A.16.

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

Conventional Wisdom

# 1

## Theory

MUCH HAS BEEN SAID about the so-called Standard Model of particle physics

*Noli turbare circulos meos*

Archimedes

# 2

## The Large Hadron Collider and the ATLAS Detector

Look at<sup>26, 19</sup>

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

nuity and

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

Conventional Wisdom

# 3

## Data and Simulated Samples

### MUCH HAS BEEN SAID

The data and Monte Carlo simulation (MC) samples are the same as in the fiducial analysis. The data corresponds to  $36.1 \text{ fb}^{-1}$  of  $p\bar{p}$  collision data collected in 2015+16 at the ATLAS detector at  $\sqrt{s} = 13 \text{ TeV}$ . Only events recorded with all systems in ATLAS in good working order and passing certain

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

Details about MC samples may be found in<sup>33</sup>. The  $ZH \rightarrow \ell\ell b\bar{b}$  process is considered for both multivariate analysis (MVA) optimization and the final statistical analysis, while  $WH \rightarrow \ell\nu b\bar{b}$  and  $ZH \rightarrow \nu\nu b\bar{b}$  production are included in the final statistical analysis only. Signal MC samples were generated separately for  $qq$  and  $gg$  initiated  $VH$  processes.  $qqVH$  samples were generated with POWHEG MiNLO + PYTHIA8 with the AZNLO tune set and NNPDF3.0 PDF. Nominal  $ggZH$  samples were generated using POWHEG for the matrix element (ME) and PYTHIA8 for the parton shower (PS), underlying event (UE), and multiple parton interactions (MPI), again applying the AZNLO tune and NNPDF3.0 PDF set.<sup>3</sup>

The background processes considered in these studies are  $Z+jets$ ,  $t\bar{t}$ , and diboson production for both MVA optimization and the final statistical analysis and single top production and  $W+jets$  only considered in the final statistical analysis.  $V+jets$  samples are generated using SHERPA 2.2.1<sup>3</sup> for both the ME and PS. These samples are generated in different groups, according to the identity of the  $V$ , the max ( $H_T, p_T^V$ ) of events, and also further subdivided according to the flavor of the two leading jets in an event,  $b$ ,  $c$ , or  $l$ , for a total of six categories..  $t\bar{t}$  samples are generated using POWHEG with the NNPDF3.0 PDF set interfaced with PYTHIA8 using the NNPDF2.3 PDF's and the A14 tune<sup>3</sup>. Single top samples use POWHEG with the CT10 PDF's interfaced with PYTHIA6 using the CTEQ6L1 PDF's<sup>33</sup>. Diboson samples are generated with SHERPA 2.2.1 interfaced with the NNPDF3.0 NNLO PDF set normalized to NLO cross sections<sup>3</sup>.

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

Conventional Wisdom

# 4

## Object Definitions and Event Selection

MUCH HAS BEEN SAID

### 4.1 OBJECT DEFINITIONS

Detailed descriptions of object definitions may be found in<sup>21</sup>. Tables 4.2, 4.3, and 4.4 are reproduced from<sup>31</sup> and summarize electron, muon, and jet definitions, respectively.

## 4.2 EVENT SELECTION AND ANALYSIS REGIONS

This analysis focuses specifically on the 2-lepton channel of the fiducial analysis, with the event selection and analysis region definitions being identical. Common to all lepton channels in the fiducial analysis is the set of requirements on the jets in a given event. There must be at least two central jets and exactly two signal jets that have been “*b*-tagged” according to the MV2c10 algorithm<sup>2</sup>, with at least one of these *b*-jets having  $p_T > 45$  GeV. For MVA training and certain background samples, a process known as “truth-tagging” is applied instead of the standard *b*-tagging to boost sample statistics and stabilize training/fits (cf.<sup>3</sup> Section 4.2 for details). After event selection, the *muon-in-jet* and *PtReco* corrections, described in<sup>2</sup> 6.3.3-4, are applied to the *b*-jets.

In addition to the common selections, there are 2-lepton specific selections. All events are required to pass an un-prescaled single lepton trigger, a full list of which may be found in Tables 5 and 6 of<sup>2</sup> with the requirement that one of the two selected leptons in the event must have fired the trigger. There must be 2 VH-loose leptons, and at least one of these must be a ZH-signal lepton (cf. Tables 4.2 and 4.3 for definitions). This lepton pair must have an invariant mass between 81 and 101 GeV. In addition to the jet corrections described above, a kinematic fitter is applied to the leptons and two leading corrected jets in an event with three or fewer jets<sup>\*</sup> to take advantage of the fact that the 2-lepton final state is closed (cf.<sup>19</sup>); these objects are only used for MVA training/fit inputs.

In order to increase analysis sensitivity, the analysis is split into orthogonal regions based on the number of jets and the transverse momentum of the  $Z$  candidate (the vectoral sum of the lepton

---

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

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 regions where the leptons are required to be the same flavor ( $e$  or  $\mu$ ), there are top  $e - \mu$  control regions used to constrain the top backgrounds.

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 4.1:** Event selection requirements

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

**Table 4.2:** Electron selection requirements.

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

**Table 4.3:** Muon selection requirements.

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

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

Conventional Wisdom

# 5

## Signal and Bacgkround Modeling

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.

ATL-COM-PHYS-2016-1724 section 4 and Main reference: ATL-COM-PHYS-2016-1747.pdf<sup>33</sup>

(we use the latter as a scaffold for this, basically plucking out stuff section by section)

Most of these studies (unless noted) are truth-level studies (particle level) done in Rivet<sup>20</sup> (F.4 for Rivet/reco comparison....not there)

### 5.1 SIGNAL PROCESSES

The dominant process is Higgsstrahlung; ggF is  $\sim 14\%$

$q\bar{q}$  Powheg with MiNLO (multiscale improved NLO generator Ref 1) applied as generator, Pythia 8 + AZNLO tune (Ref 3) + NNPDF<sub>3.0</sub> PDF (Ref 4) set;  $gg \rightarrow ZH$  Powheg + Pythia 8 (Ref 2)

Alternate samples: MadGraph 5(ref 26)\_aMC@NLO+Pythia 8

Cross section: this is done NNLO in QCD and NLO in EW except for ggZH NLO+NNL (QCD). WH normalized to values in the table; ZH: total as 0.88, ggZH as 0.12, and then qqZH as total - ggZH (refs 15-18)

Process	$\sigma(\text{pb})$
$WH$	$1.37 \pm 0.04$
$W^+ H$	0.84
$W^- H$	0.53
$ZH$	$0.88^{+0.04}_{-0.03}$
$gg \rightarrow ZH$	0.12
$qq \rightarrow ZH$	0.76

**Table 5.1:** Summary of inclusive cross sections for signal processes.

NLO EWK correction: same as Run 1; they use HAWK to calculate a differential cross section as a function of pTV (take their Figure 3) for a correction factor of  $k_{EW}^{NLO}(p_T^V) = 1 + \delta_{EW}$ ; qqVH only.

N(N)LO EWK systematic:  $\Delta_{EW} = \max\{1\%, \delta_{EW}^2, \Delta \}$ ,  $\delta_{EW}$  from above correction,  $\Delta$  is  $\gamma$

induced cross section uncertainty to the total [WZ]H xsec

Overall signal acceptance uncertainties: cross section and branching ratio (LHC Higgs WG;<sup>28, 32</sup>)

- ATLAS\_BR\_bb (1.7%)
- ATLAS\_QCDscale\_(VH|ggZH) (0.7%, 27%) vary  $\mu_{R,F}$  for renorm/factorization scale by 1/3 to 3 of original value
  - to get ggZH; assume QCD scale  $\sigma$  same for qq[WZ]H; assume ref 20 inclusive ZH production and take diff in quadrature of inc. and qqZH
- ATLAS\_pdf\_Higgs\_(V[WZ]H|ggZH) (1.9%, 1.6%, 5.0%) (also  $\alpha_s$ , 68%CL on PDF4LHC15\_nnlo\_mc PDF set)
  - qqWH is bigger here than ZH; get ggZH from 19, qqZH from 20 assuming ggZH small, so overall ZH is qqVH

Analysis specific: analysis category acceptances; pTV, mBB shape

- PS/UE (Table 4)
  - MadGraph vs. A14 varied (tunes); nominal Powheg/Minlo/Pythia8 vs Powheg+minlo+Herwig7 (PS)
  - Now vary up and down in each nLep x nJet bin and save as a ratio wrt nominal
  - $\sum_{tunes} \max_{tune} (|R_{up} - R_{down}|) \oplus \sigma_{PS} (\text{ATLAS\_UEPS\_VH\_hbb})$
  - Now add a 2/3 jet ratio systematic (i.e. (2/3 acceptane ratio nominal)/(2/3 ratio alternative)) (ATLAS\_UEPS\_VH\_hbb\_32JR); combine in same way
  - pTV (mBB) shape: linear (quadratic); fit up and down for each variation; 2/3jet separate for mBB; use histogram with largest deviation as shape (ATLAS\_UEPS\_VH\_hbb\_(VPT|MBB))
    - \* shape only, except for L2 pTV (shape+norm)

- ggZH same as qqZH and correlated
- Scale variations (Table 5)
  - Vary  $\mu_R$ ,  $\mu_F$  (probably the 1/2 to 2 scheme in steps with no more than blah blah)
    - \* Stewart Tackmann for nJet bins (QCDscale\_VH\_ANA\_hbb\_J[23]; both for 2jet)
    - \* JVeto for L[01] (3 jet exclusive)
  - Same pTV, mBB NP scheme for nLep/nJet as for UEPS
  - ggZH same as qqZH and de-correlated (Run 1 says they're difference)
- PDF+ $\alpha_s$ 
  - Powheg/Minlo/Pythia8 v. PDF4LHC15\_30 PDF set; reco-level distributions (all others use Rivet, which doesn't like a lot of weight variations)
  - PDF: quad sum of variations of PDF uncertainties (go through the set? no probably the same way as above)
  - $\alpha_s$ : average of variations from altering  $\alpha_s$
  - pdf\_HIGGS\_VH\_ANA\_hbb, pdf\_VH\_ANA\_hbb\_(VPT|MBB)
  - Same pTV, mBB NP scheme for nLep/nJet as for UEPS

## 5.2 BACKGROUND

Main backgrounds are V+jet, ttbar, VV, single top (, and multijet in 1lep)

### 5.2.1 V+JET

cf.<sup>18</sup> for details of MC generation

- Sherpa 2.2.1@NLO<sup>27</sup> for matrix element (ME) and PS tuning (Tables 7–10)
  - ME’s for up to 2 (3–4) partons at NLO (LO); for more, use showering (Sherpa’s own UEPS)
  - “The merging of different parton multiplicities is achieved through a matching scheme based on the CKKW-L [24] [25] merging technique using a merging scale of  $\underline{Q_{\text{cut}}} = 20 \text{ GeV}$ <sup>30 29</sup>
  - 5 quark flavors mass(less) quarks in the shower (ME)
  - $\max(H_T, P_T^V)$  slices:  $[0-70, 70-140, 140-280, 280-500, 500-1000, >1000] \text{ GeV}$
  - Slices in [CB](Veto|Filter) for flavors
    - \* BFilter: at least 1 b-hadron with  $|\eta| < 4, p_T > 0 \text{ GeV}$
    - \* CFILTERBVeto: at least 1 c-hadron with  $|\eta| < 3, p_T > 4 \text{ GeV}$ ; veto events which pass the BFilter
    - \* CVetoBVeto veto events which pass the BFilter or the CFILTERBVeto
  - Variations of  $\mu_{R,F}$  at 0.5, 2; PDF variation for MMHT2014nnlo68cl and CT14nnlo
  - Sherpa 2.1 for resummation scale at 0.5, 2; CKKW 15, 20 GeV
- Alternate samples use MadGraph5+Pythia8 (UEPS)
  - LO QCD ME’s, merging parton multiplicities up to 4 (for more, use PS), NNPDF2.3 LO PDFs; A14 tune (ATLAS)
  - CKKW-L scheme with a merging scale of  $\underline{Q_{\text{cut}}} = 30 \text{ GeV}$ .
  - 5 flavor scheme

- Cross section  $k$ -factors: our generators are NLO, but V production is known to NNLO—add factors to rescale
  - Take total events, average over lepton flavors for filter efficiencies, and compare to NNLO (ref 27)
  - For  $L_2$ , there's a 40 GeV generator mLL cut, but the NNLO calcu is done in (66,116) GeV, so another scale
  - For Lo, take  $L_2$  (since NNLO not calc) and correct for  $BR(Z \rightarrow \nu\nu) / BR(Z \rightarrow \ell\ell)$ , consider with no mass cuts, remove “ $Z/\gamma^*$  interference”
  - Differences between nominal and alternative MC's can be explained to higher order BR's and EW schemes w.r.t. PDG recommendations

Anyway, V+jet is broken up into V+hf (V+b\*, V+cc), V+cl, V+l(ight)

- Relative acceptance between regions
  - Understand correlation between/among regions (you can float these normalizaitons in the fit to fix your understanding of things using more ifnrmation)
  - 2jet vs 3(p)jet for  $L[01](2)$ , Lo vs.  $L_2$  ( $Z+hf$ ), Lo vs.  $L_1$  ( $W+hf$ ), WCR vs. SR for  $L_1$  ( $W+hf$ )
  - These norm's are RooGaussian's with priors from MC studies (Rivet, Appendix A <sup>33</sup>)
  - Their uncertainties are double ratios between regions and then MC's with components...
    - \* Envelope of varying  $\mu_R$ ,  $\mu_F$  in Sherpa
    - \*  $0.5 \sum_{\oplus}$  (up-down on CKKW, merging scale variation; weird because done with Sherpa 2.1, so no central value comparison)
    - \* max variation between nominal/alt PDF reweighting
    - \* diff btw Sherpa/MadGraph

- pTV, mBB shape uncertainties: data driven and MC techniques—you normalize distributions to the same area, compare, then do functional fits, then pick the biggest one and symmetrize
- W+jets
  - Normalization/acceptance systs (Table 13):  $\text{Sys}(\text{Wcl}|\text{WL})\text{Norm}$  (one for all regions is fine since  $b$ -tagging suppresses), a floating  $\text{norm\_Wbb}$ ,  $\text{SysWbbNorm}_{-}(\text{J3}|\text{DWhfCR\_L1}|\text{L0})$ ;  $\text{J}_3$  is 3-to-2 jet;  $\text{DWhfCR\_L1}$  is CR-SR;  $\text{Lo}$  is  $\text{Lo-L1}$
  - Flavor composition (Tables 14, 15): W+hf breakdown;  $\text{Sys}(\text{Wbc}|\text{Wbl}|\text{Wcc})\text{WbbRatio}$
  - pTV: a linear  $\text{SysWPtV}$ , which happens to be Serpa 2.2.1 v. MadGraph in all regions (largest variation)
  - mBB: a linear  $\text{SysWMbb}$ , which happens to be Serpa 2.2.1 v. MadGraph in all regions (largest variation) (not a typo; it's the same as pTV)
- Z+jets: L[02] SR only (topemucr is pretty pure; not really in L1)
  - Normalization/acceptance (Table 16):  $\text{Sys}(\text{Zcl}|\text{Zl})\text{Norm}$  (one for all regions is fine since  $b$ -tagging suppresses; less than 1% here), a floating  $\text{norm\_Zbb}$ ,  $\text{SysWbbNorm}_{-}(\text{L2\_J3}|\text{J3}|\text{0L})$ ;  $\text{L2\_J3}$ ,  $\text{J}_3$  is 3-to-2 jet ( $\text{L2}$  correlates  $\text{lo}/\text{hi}$  pTV;  $\text{Lo}$  is separate because of selection differences);  $\text{oL}$  is 0 to 2 lepton ( $\text{hi}$  pTV only)
  - Flavor composition (Tables 17): Z+hf breakdown;  $\text{Sys}(\text{Zbc}|\text{Zbl}|\text{Zcc})\text{ZbbRatio}$ —norm uncertainties with diff priors in  $\text{Lo}$ ,  $\text{L2-2jet}$ ,  $\text{L2-3pjet}$ ; Sherpa 2.2.1 v MG main diff
  - $\text{L2}$  CR:  $\text{METHT} < 3.5$ , [012]-tag, 2 and 3pjet, no mJJ in (110,140) GeV for 2tag, pTV regions; subtract off non Z+jet and then scale MC to data
  - pTV: shape+norm, fit to data in  $\text{L2}$  CR;  $\pm 0.2 \log_{10}(p_T^V/500 \text{ GeV})$
  - mBB: shape only, fit to data in  $\text{L2}$  CR;  $\pm 0.0005 \log_{10}(m_{jj} - 100 \text{ GeV})$

### 5.2.2 TOP-PAIR PRODUCTION

MC production— $b_{damp}$  is transverse momentum scale at which Sudakov resummation becomes unimportant: smaller damp means higher suppression (cf. Table 20)

- Powheg+Pythia8
  - Powheg: NNPDF3.0 (NLO) for ME (Powheg);  $b_{damp} = 1.5m_{top}$  (resummation damping factor for ME/PS matching; controls high pT rad)
  - Pythia: PS,UE,had; v 8.210, A14 PDF set, NNPDF2.3 LO for PS; pTdef=2, pThard=0 control Powheg/Pythia8 merging thorough shower vetoing
  - $\sigma_{t\bar{t}}(m_{top} = 172.5 \text{ GeV}) = 831.76^{+40}_{-46} \text{ pb}$ : NNLO QCD; NNLL soft gluon terms;
    - \* QCD scale variations:  $^{+19.77}_{-29.20} \text{ pb}$ ; PDF:  $\pm 35.06 \text{ pb}$ : “The PDF and  $\alpha_s$  uncertainties were calculated using the PDF4LHC prescription [8] with the MSTW2008 68
    - \* 3.3 times higher than 8 TeV
- Powheg+Herwig7: different PS, UE, had, MPI; H7UE tune
- MadGraph 5\_aMC@NLO+Pythia 8.2: different hard scatter (i.e. ME)
- Powheg+Pythia8 low radiation sample (double  $\mu_{R,F}$ ;  $b_{damp}$ , pTdef, pThard same; A14 tune Var3c Down variation used)
- Powheg+Pythia8 high radiation sample (halve  $\mu_{R,F}$ ; pTdef, pThard same;  $b_{damp} = 3m_{top}$  (doubled) A14 tune Var3c Up variation used)

Systematics—Rivet

- Powheg+Pythia8

—

### 5.3 NOTES

Notes from Kevin's thesis: Signal:

- $pTV NLOEWK$  The signal processes have some pTV dependence at next to leading order (NLO) due to electroweak corrections
- $TheoryQCDScale$ ,  $TheoryPDF$  for renormalization/scale uncertainties, PDF uncertainties
- $TheoryAcc_J[23]$  Stewart-Tackmann stuff
- $TheoryAccPDF$  do acceptance calculations with different PDF's
- $TheoryVPtQCD$  this is one of those functional things—probably different in Run2; linear of pTV

Background

- $ZDPhi \Delta\phi(b_1, b_2)$  mismodeling; shape—another linear of dphi; a correction and the correction is a systematic for each event
- $ZPtV \Delta\phi(b_1, b_2)$  mismodeling; const+log and half the correction is a systematic for each event
- $Z+jet Normalizations$  broken down by flavor region; both Norm's and Ratio between regions
- $ZMbb$  const(mbb e-3 -c const); systematic
- $ttbar$  pT, (2/3 jet ratio across generators), mBB
- $VV$  NLO xsec, s/PDF's, mJJ

### 5.3.1 STEWART-TACKMANN

A way to calculate uncertainties on processes in different nJet bins<sup>34</sup>. Generically:

$$\sigma_{\geq N} = \sigma_N + \sigma_{\geq N+1} \quad (5.1)$$

There's some quantity that you make a cutoff in an integral that defines the border between jet regions.

$$\sigma_{\geq N} = \int_0^{p_{cut}} \frac{d\sigma_N}{dp} + \int_{p_{cut}} \frac{d\sigma_{\geq N+1}}{dp} \quad (5.2)$$

So for some fucking reason, inclusive cross sections are easier to calculate, so you can just vary  $\alpha_s$  in the usual way for those and treat the two inclusive cross sections. Anywho, we assume the inclusive uncertainties are uncorrelated, for a covariance matrix for  $\{\sigma_{\geq N}, \sigma_N, \sigma_{\geq N+1}\}$  of:

$$\Sigma = \begin{pmatrix} \Delta_{\geq N}^2 & \Delta_{\geq N}^2 & 0 \\ \Delta_{\geq N}^2 & \Delta_{\geq N}^2 + \Delta_{\geq N+1}^2 & -\Delta_{\geq N+1}^2 \\ 0 & -\Delta_{\geq N+1}^2 & \Delta_{\geq N+1}^2 \end{pmatrix} \quad (5.3)$$

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

“In the limit  $\alpha_s L^2 \approx 1$ , the fixed-order perturbative expansion breaks down and the logarithmic

terms must be resummed to all orders in  $\alpha_s$  to obtain a meaningful result. For typical experimental values of pcut fixed-order perturbation theory can still be considered, but the logarithms cause large corrections at each order and dominate the series. This means varying the scale in  $\alpha_s$  in Eq. (9) directly tracks the size of the large logarithms and therefore allows one to get some estimate of the size of missing higher-order terms caused by  $p_{cut}$ , that correspond to  $\Delta_{cut}$ . Therefore, we can approximate  $\Delta_{cut} = \Delta_{\geq 1}$ , where  $\Delta_{\geq 1}$  is obtained from the scale variation for  $\sigma_{\geq 1}$ .”

They use the example of ggF Higgs production with  $\{\sigma_{total}, \sigma_o, \sigma_{\geq 1}\}$  and say this works to all  $N$  for all processes, provided one picks  $\mu \approx Q$  so you can use perturbative expansions.

Anyway, the upshot is this: we've got 2 and 3 jet bins. For 2 jet TheoryAcc\_J2 and TheoryAcc\_J3; 3 jet has TheoryAcc\_J3, which is anti-correlated with the 2 jet J3 term

### 5.3.2 CKKW-L

When you're looking to generate MC events, there are two main event generators. There are the parton shower event generators (PSEG), like Pythia, and the matrix element generators (MEG) like MadGraph or Powheg, both of which have nice and not-so-nice features. If we follow<sup>29</sup>, section 2, we get a nice illustration. Sherpa does both and stitches things together for you.

So PSEG's have the nice feature that you don't get nasty infinities. You start with some primary hard scatter (say  $e^+ e^- \rightarrow q\bar{q}$ ) and then let your incoming and outgoing partons cascade via iterative 1 → 2 branching. You order the emissions by some “evolution scale  $\xi$ ,” starting at  $\xi_0$  and decreasing until you reach some pre-determined cutoff  $\xi_c$  (usually to match some model) to generate 0, 1 . . .  $n$  extra partons, there are exclusive cross sections involving well-ordered, intermediate scales  $\xi_i$ ,

some phase space variables (like momentum fractions  $z_i$ ) denoted  $\Omega_i$ , probabilities of non-emission between scales in the form of Sudakov form factors  $\Delta_{S_i}(\xi_i, \xi_{i+1})$ , coefficients  $c_{nn}^{PS}$  associated with splitting functions that depend on  $\xi_i, \Omega_i$  and sum over flavors, blah blah.

The  $\Delta$ 's look like:

$$\Delta_S(\xi_i, \xi_{i+1}) = \exp \left( - \int_{i+1}^i \frac{d\xi}{\xi} \alpha_s(\xi) \int dz P(z) \right) \quad (5.4)$$

and these can be written as a perturbative series in  $\alpha_s$  ("duh")

$$\sigma_{+o} = \sigma_o \Delta_{S_o}(\xi_o, \xi_c) \quad (5.5)$$

$$\sigma_{+n} = \sigma_o c_{nn}^{PS} \Delta_{S_n}(\xi_o, \xi_c) \prod_{i=1}^n \alpha_s(\xi_i) \Delta_{S_{i-1}}(\xi_{i-1}, \xi_i) d\xi_i d\Omega_i$$

(5.7)

$$\sigma_{+n} = \sigma_o c_{nn}^{PS} (1 + c_{n,n+1}^{PS} \alpha_s + c_{n,n+2}^{PS} \alpha_s^2 + \dots) \prod_{i=1}^n d\xi_i d\Omega_i \quad (5.8)$$

Now, these  $c_{ij}^{PS}$  blow up in the soft/collinear limit of  $\xi_c \rightarrow 0$ , but a resummation in all order for the  $\Delta$ 's gives a finite result for each cross section. Moreover,  $\sum_o^\infty \sigma_{+i} = \sigma_o$ . *The problem is that for several hard partons, this description only makes sense for strict ordering (the intermediate states) of hard partons because of the splitting function dependent coefficients.*

For MEG's, the picture is simpler because we use tree-level matrix elements for each parton final state. However, the cross-sections are *inclusive* (so each of these is at least  $n$  jets), and these all blow up in the soft/collinear regime, where the resummation gets nasty. The authors note that you can make PSEG's look like the MEG for the first emission ( $c_{\text{II}}^{\text{PS}} \rightarrow c_{\text{II}}^{\text{ME}}$ ).

$$\sigma_{+\circ} = \sigma_{\circ} \quad (5.9)$$

$$\sigma_{+n} = \sigma_{\circ} \alpha_s^n c_{nn}^{\text{ME}} \prod_{i=1}^n d\Omega_i \quad (5.10)$$

So what to do? "...the solution should be obvious." Just use the MEG to generate your partons over some  $Q_{\text{cut}}$ , reweight the generated states with the Sudakov form factors, and use the PSEG to make parton showers for these final state objects so that the showers make everything under  $Q_{\text{cut}}$ . But those Sudakov scales need an ordered set of emission scales since all the diagrams are added together.

How does one set up an ordered set of scales? You can use the  $k_{\perp}$ -algorithm (takes pairs based on something like  $p_T$  (??)); use those scales as arguments to  $\alpha_s$ ; use  $k_{\perp}$ -algorithm resolution as a cut-off. This approach is good to NLL but has some discontinuities. Anyway,  $k_{\perp}$  is basically the same thing as  $k_t$  clustering for jets<sup>22</sup>. Actually, it *is* the same exact thing for lepton colliders, so they use the angle between particles times a minimum square energy instead of  $\Delta R$  and define beam jets...they also don't have the minimum distance built in, so there's a  $d_{\text{cut}}$ , which can be the square energy or some other thing; you can define it by the resolution in  $y$  you want by  $y_{\text{cut}} = Q_{\circ}/d_{\text{cut}}$ ). Remember,  $k_t$  starts with your softest stuff and clusters upwards from there. For the resolution variable, remem-

ber that you have some characteristic distance after which things. Blah blah, so you pre-cluster (their topocluster type stuff for hadronic deposits based on the min  $E_{T,i}^2, E_{T,j}^2, \vartheta_{ij}^2$  metric) until all distances remaining are bigger than  $d_{cut}$ . Now define  $y_{cut} = Q_o^2/d_{cut}$  and use  $y_{kl} = d_{kl}/d_{cut}$  and cluster until all bigger than  $y_{cut}$ . The important thing from this mess is just that  $y_{cut}$  is the resolution mentioned above; you don't have this mess with our usual algorithms because distances come in with  $\Delta R^2/R_{alg}^2$ , so if distances remaining are bigger, the plain "beam distance" keeps things unclustered.

## DIPOLE CASCADE MODEL

You can also use the dipole cascade model ( $2 \rightarrow 3$  where the 2 partons are a color dipole). The

dipoles mean you don't have to do an angular ordering for the partons? Your  $\xi$  is  $p_\perp^2 = \frac{s_{12}s_{23}}{s_{123}}$

where  $s$ 's are invariant masses of the combinations. There's also a rapidity associated with the  $p_\perp$ 's:

$y = \frac{1}{2} \ln \left( \frac{s_{12}}{s_{23}} \right)$ . The emission probability depends on splitting functions, which in turn depend on

the parton pair type (parton 2 is the emitted one in this convention), where  $x_i = 2E_i/\sqrt{s_{123}}$ :

$$D_{q\bar{q}}(p_\perp^2, y) = \frac{2}{3\pi} \frac{x_1^2 + x_3^2}{(1-x_1)(1-x_3)} \quad (5.11)$$

$$D_{qg}(p_\perp^2, y) = \frac{3}{4\pi} \frac{x_1^2 + x_3^3}{(1-x_1)(1-x_3)} \quad (5.12)$$

$$D_{gg}(p_\perp^2, y) = \frac{3}{4\pi} \frac{x_1^3 + x_3^3}{(1-x_1)(1-x_3)} \quad (5.13)$$

$$(5.14)$$

Finally, we get the probability:

$$dP(p_\perp^2, y) = \alpha_s(p_\perp^2) D_{ij}(p_\perp^2, y) \exp \left( - \int_{p_\perp^2} \frac{p'_\perp^2}{p'^2_\perp} \int dy' \alpha_s(p'^2_\perp) D_{ij}(p'^2_\perp, y') \right) \frac{dp_\perp^2}{p_\perp^2} dy \quad (5.15)$$

(notice your old exp friend, the Sudakov). Also, hey, look, your intermediate partons are on shell, unlike in a  $1 \rightarrow 2$  cascade since your dipole absorbs recoil, and your inverse cascade is a well-behaved “jet clustering” algorithm. But  $g \rightarrow q\bar{q}$  has to be done by hand. Basically, you use this cascade/shower on your MEG partons to get scales that you reweight by  $\prod_i \alpha_s(p_{\perp i}) / \alpha_s(p_{\perp c})^n$  for some  $n$ .

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

Conventional Wisdom

# 6

## Experimental Systematic Uncertainties

ATL-COM-PHYS-2016-1724 section 6 points to 1674 (section II)

MUCH HAS BEEN SAID about the so-called Standard Model of particle physics

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

Conventional Wisdom

# 7

## Multivariate Analysis

MUCH HAS BEEN SAID

### 7.1 TRAINING SAMPLES AND VARIABLES

A subset of samples described in Section ?? was used for multivariate analysis (MVA) training, with  $qqZH \rightarrow \ell\ell b\bar{b}$  and  $ggZH \rightarrow \ell\ell b\bar{b}$  used as signal samples and  $Z + \text{jets}$ ,  $t\bar{t}$ , and  $VV$  used as background

samples. Truth-tagging is used on all samples in MVA training to improve training statistics and stability.

The standard set of variables taken as a baseline is the same as used in the fiducial analysis. The variables fall into several main categories: energy/momentum scales of composite objects ( $m_{bb}$ ,  $m_{bbj}$ ,  $p_T^V$ ,  $m_{\ell\ell}$ ), angles ( $\Delta R(b_1, b_2)$ ,  $\Delta\phi(V, H)$ ,  $\Delta\eta(V, H)$ ), transverse momenta of the jets in the event ( $p_T^{b_1}$ ,  $p_T^{b_2}$ ,  $p_T^{j_3}$ ), and  $E_T^{miss}$ .

In addition to the standard set of variables used for MVA training, two additional sets of variables were used: the Lorentz Invariants (LI) and RestFrames (RF) inspired variable sets. These will be discussed below. All figures quoted in this section scale distributions to a luminosity of  $36.1 \text{ fb}^{-1}$ .

## 7.2 LORENTZ INVARIANTS

The LI variables, first put forth by S. Hagebeck and others<sup>3</sup>, are based upon the fact that the four-vectors of an event are determined, all of the information in an event are encoded into 16 quantities: the ten inner products of the four vectors, the three Euler angles, and the three parameters specifying the boost of the  $ZH$  system. The masses of the four final state objects are not considered very useful and so can be removed to leave six meaningful inner products (the  ${}_4C_2$  combinations between distinct final state four vectors). Since these inner products can have an ill-defined physical interpretation and in order to help MVA training, each inner product is scaled by:

$$x \rightarrow \frac{x}{x + c} \quad (7.1)$$

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

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

Similar to the standard set there are masses (`j0_j1` corresponds to `mBB`, and `l0_l1` corresponds to `mLL`) and angles. Instead of individual final state object scales, there are the four jet-lepton inner products, though this correspondence (and indeed any physical interpretation) is far from clear. An important advantage of the LI variable set is that all of the variables are in it are orthogonal in the signal case by construction. A drawback of this framework in a completely closed final state is that there is no way to treat  $E_T^{miss}$  in a Lorentz invariant way. In practice, however, it is found that the  $E_T^{miss}$  has correlations comparable to other correlations between variables (cf. Figure 7.2 (b) and (e)). There is also no prescription for any additional jets in the event beyond the two  $b$ -tagged jets. They are simply ignored in these variable calculations.

### 7.3 RESTFRAMES VARIABLES

The RestFrames variables<sup>3</sup>, calculated using the software package of the same name, is based upon the idea that the most natural frame in which to analyze objects of the signal decay tree is in their individual production (rest) frames. The signal decay tree for  $ZH \rightarrow \ell\ell b\bar{b}$  is show in Figure 7.1. Gen-

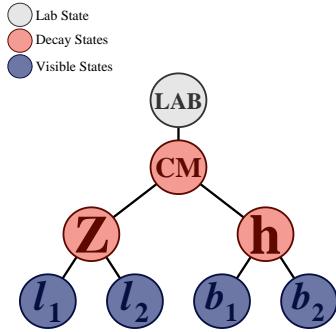


Figure 7.1: The  $ZH \rightarrow \ell\ell b\bar{b}$  decay tree.

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

Each frame has associated with it the boost from its immediate parent and a mass scale; that mass and the angles between the Euclidean three vector associated with boost and the axis of the decay products provide useful variables. In general, the polar angle (typically given as a cosine) is considered more useful than the azimuthal angle (typically just a  $\Delta\phi$ ), though this is dependent on the candidate decay tree. The  $Z$  frame, for example, has  $M_Z$ , which is just the usual  $m_{\ell\ell}$ ,  $\cos Z$ , the cosine

---

<sup>3</sup>There are ways to do this for generic decay trees, though, and this is the focus of much of<sup>2</sup> and the functionality of the RestFrames package.

of the polar angle between the lepton momentum axis in their production frame and the boost from the  $ZH$  center of mass (CM) frame, and the angle  $d\phi_{ICMZ}$ .

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

$$R_{p_T} = \frac{p_{T,CM}}{p_{T,CM} + M_{CM}}, \quad R_{p_z} = \frac{p_{z,CM}}{p_{z,CM} + M_{CM}}, \quad R_{met} = \frac{E_T^{miss}}{E_T^{miss} + p_{T,CM}} \quad (7.2)$$

denoted  $R_{pT}$ ,  $R_{pZ}$ , and  $R_{met}$ . These are used instead of the final state object scales and standard  $E_T^{miss}$  of the standard variable set.

For this analysis, the RF variables were chosen to be the masses  $M_{CM}$ ,  $M_H$ , and  $M_Z$ , the angles  $\cos CM$ ,  $\cosh$ ,  $\cos Z$ ,  $d\phi_{ICM H}$ , and the ratios  $R_{pT}$ ,  $R_{pZ}$ , and  $R_{met}$ .

#### 7.4 CORRELATIONS

The variable sets used in these studies are summarized in Table ??, while Figure 7.2 shows the correlations for the signal and background samples for the standard, LI, and RF variable sets in the most significant analysis signal region, the 2 jet,  $p_T^V > 150$  GeV bin (matrices, as well as input distributions, for all regions made be found in Appendix ??). As can be seen in the correlation matrices, variable correlations tend to be much lower, particularly for the signal hypothesis, for the LI and RF variables than for the standard set. Notable exceptions are jet/lepton inner products in the LI set and

$R_{met}$  and  $R_{pt}$  in the RF case (not surprising given variable definitions since  $p_{T,CM}$  shows up in both and one roughly expects this to be antiparallel to  $E_T^{miss}$  in the signal case). While these lower correlations are not so important for the stats only results in this section, they offer an explanation for some of the results in Section ??.

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

**Table 7.1:** Variables used in MVA training. Variables in parentheses are only used in the  $\geq 3$  jet regions.

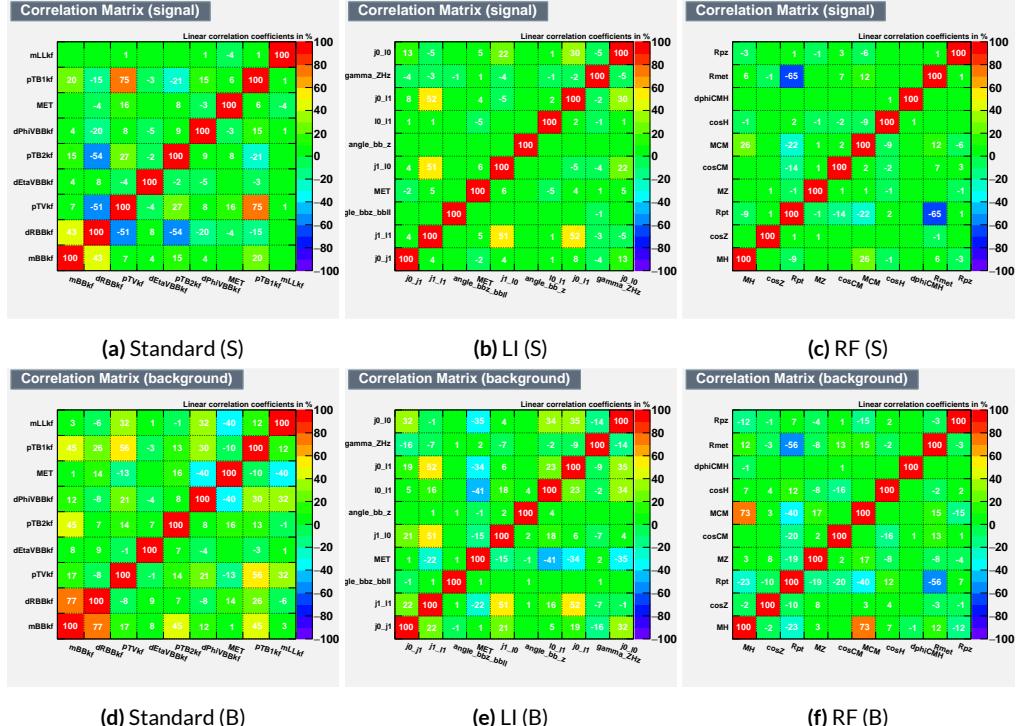
## 7.5 MVA TRAINING

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

The MVA discriminant used is a boosted descision tree (BDT). Training is done in TMVA using the training settings of the fiducial analysis<sup>31†</sup>. For the purposes of hyperparameterization and test-

---

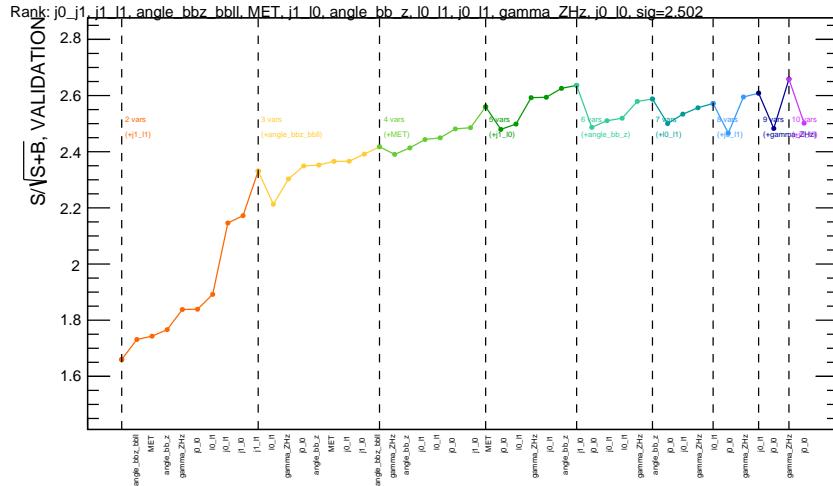
<sup>†</sup>Namely, `!H:!V:BoostType=AdaBoost:AdaBoostBeta=0.15:SeparationType=GiniIndex:-`



**Figure 7.2:** Signal and background variable correlations for the three variable sets in the 2 jet,  $p_T^V > 150$  GeV.

ing, transformation D with  $z_s = z_b = 10$  is applied to the BDT distributions, and the cumulative sum of the significance  $S/\sqrt{S+B}$  in each bin is calculated for each pair of distributions.

Variable ranking is done iteratively in each analysis region. In each set, the validation significance of a BDT using an initial subset of variables is calculated (**dRBB** and **mBB** for the standard set; **j0\_j1** for the **LI** set; and **MH** for the **RF** set). Each of the remaining unranked variables are then added separately, one at a time, to the BDT. The variable yielding the highest validation significance is then added to the set list of ranked variables and removed from the list of unranked variables. This process is repeated until no variables remain. A plot of the ranking for the **LI** set in the  $3+$  jet, low  $p_T^V$  region can be seen in Figure 7.3.



**Figure 7.3:** An example of a variable ranking to determine variable order in the BDT discriminant. The ranking for the 2 jet,  $p_T^V > 150$  GeV region for the **LI** set is shown.

Once variables have been ranked, the BDT may be used both to evaluate performance in a simplified analysis scenario in the absence of systematic uncertainties (described below in Section 7.6)

---

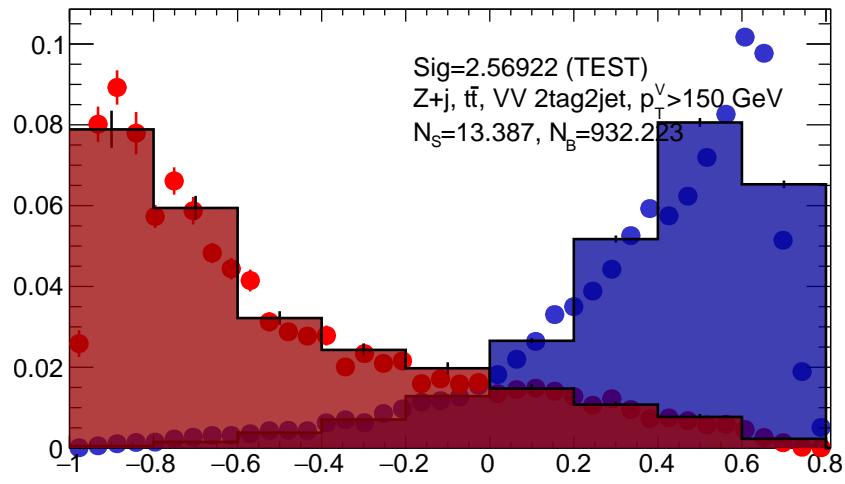
`PruneMethod=NoPruning:NTrees=200:MaxDepth=4:nCuts=100:nEventsMin=5%`

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

## 7.6 STATISTICS ONLY BDT PERFORMANCE

As described above, cumulative significances can be extracted from pairs of signal and background BDT output distributions in a given region. In order to evaluate performance of variable sets in the absence of systematic uncertainties, such pairs can be constructed by evaluating BDT score on the testing set of events using the optimal variable rankings in each region. An example of such an output is shown in Figure 7.4, with the training BDT distributions in points and the testing distributions after transformation D as block histograms. For better comparison of the distributions, all histograms have been scaled to have the same normalization.

As can be seen in the summary of cumulative significances for each of these analysis regions and variable sets in Figure 7.5, the performance of each of the variable sets is quite similar. The standard set performs best, with the LI (RF) set having a cumulative significance that is 7.9% (6.9%) lower. This suggests that the LI and RF variables, in the  $ZH \rightarrow \ell\ell b\bar{b}$  closed final state, have no more intrinsic descriptive power than the standard set. That these figures are all relatively high ( $\sim 4.5$ ) is due largely to the absence of systematics and possibly in part due to the fact that many of the most significant bins occur at high values of the BDT output, which, as can be seen in Figure 7.4, contain



**Figure 7.4:** Training (points) and testing (block histogram) MVA distributions used for stat only testing for the 2 jet,  $p_T^V > 150$  GeV region for the LI set is shown.

a small fraction of background events.

A full set of ranking and testing plots, as well as correlation matrices and input variable distributions, may be found in Appendix ??.

Correlation, ranking, and input variable plots for the standard, Lorentz Invariant, and Rest-Frames variable sets.

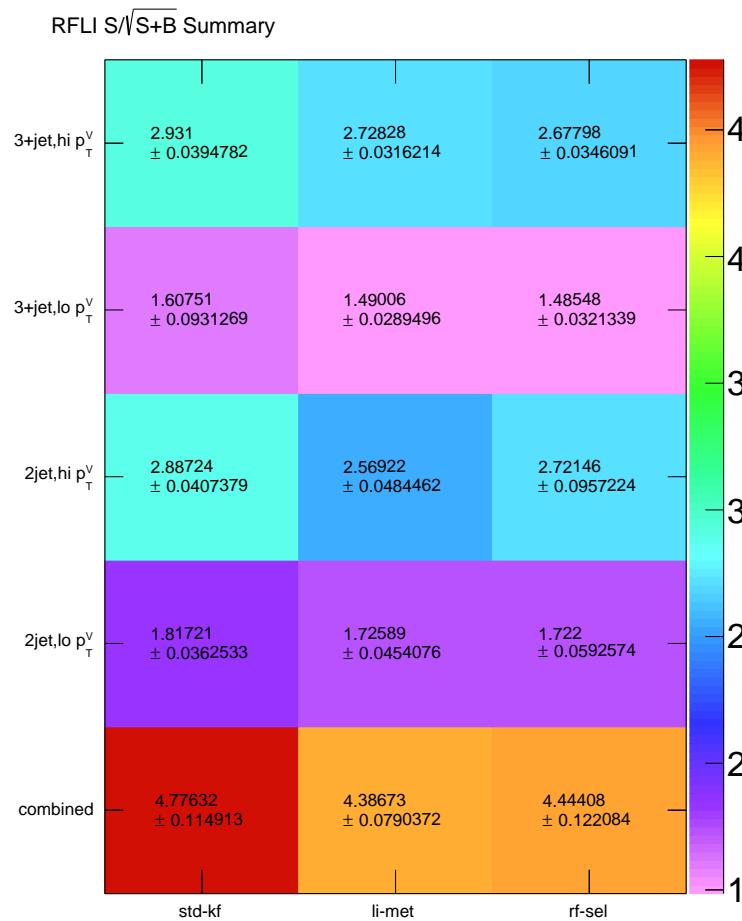
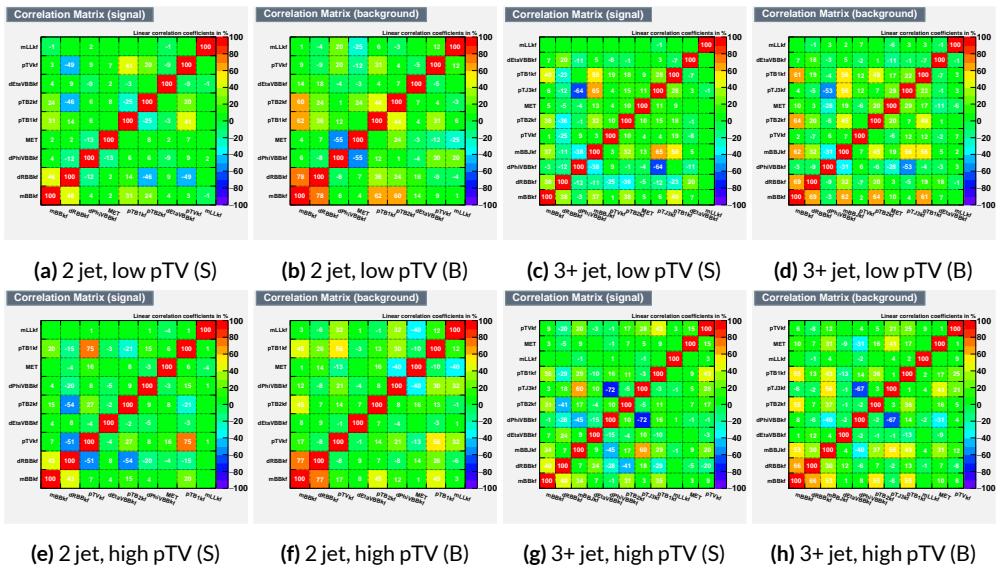
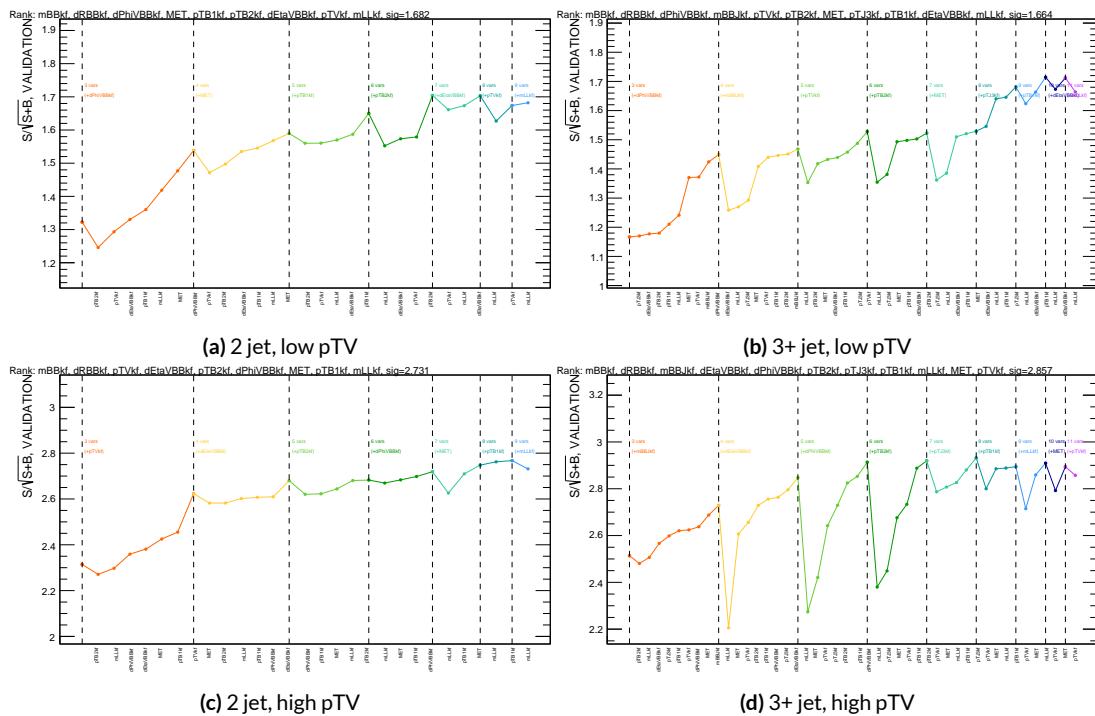


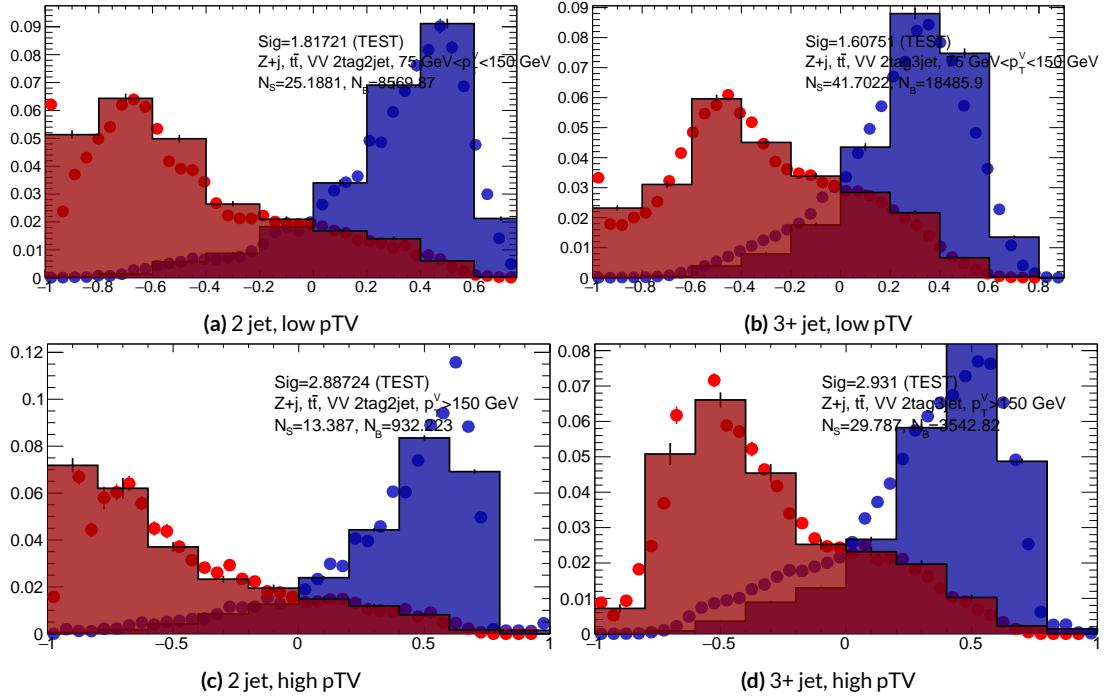
Figure 7.5: Results of testing significances sorted by analysis region and variable set.



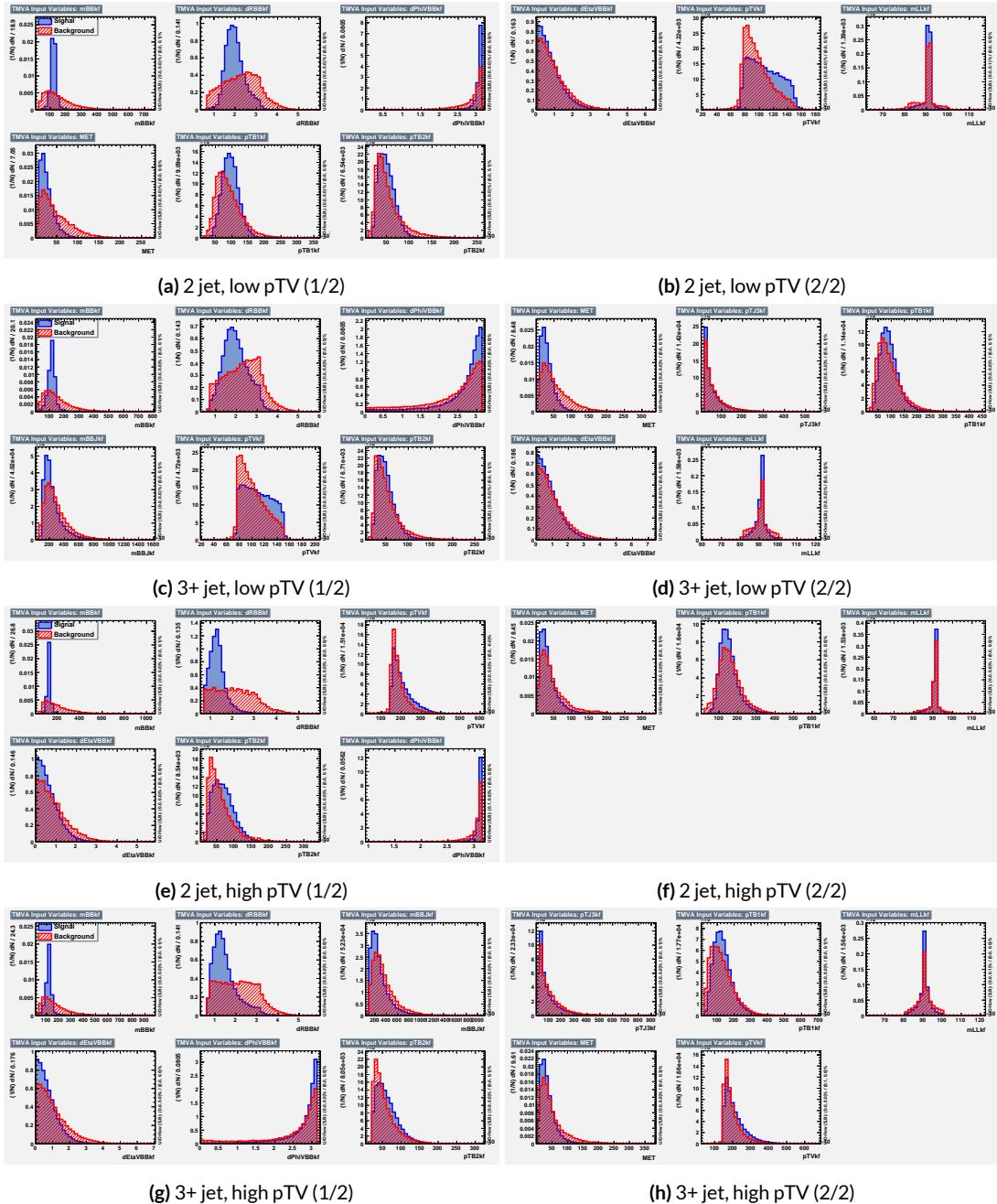
**Figure 7.6:** Signal and background variable correlations for the standard variable set.



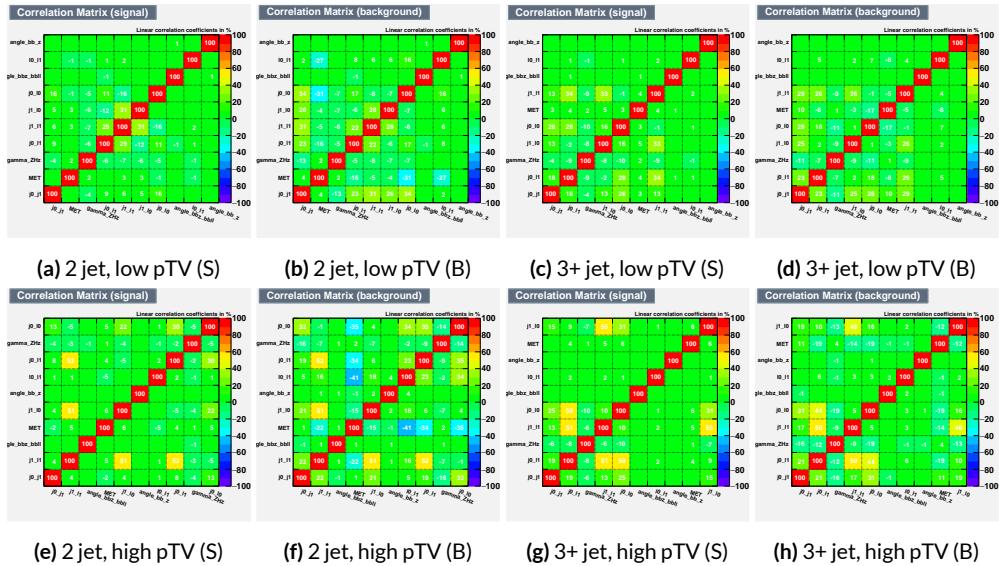
**Figure 7.7:** Signal and background variable correlations for the standard variable set.



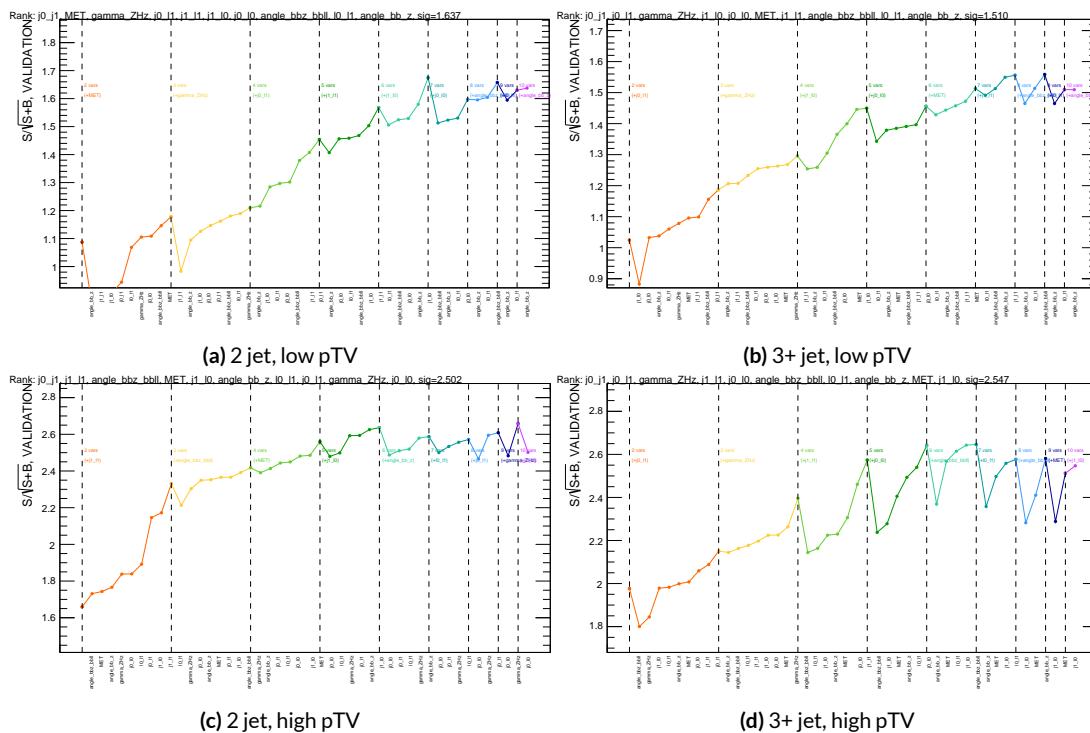
**Figure 7.8:** Training (points) and testing (block histogram) MVA distributions used for stat only testing for the standard variable set.



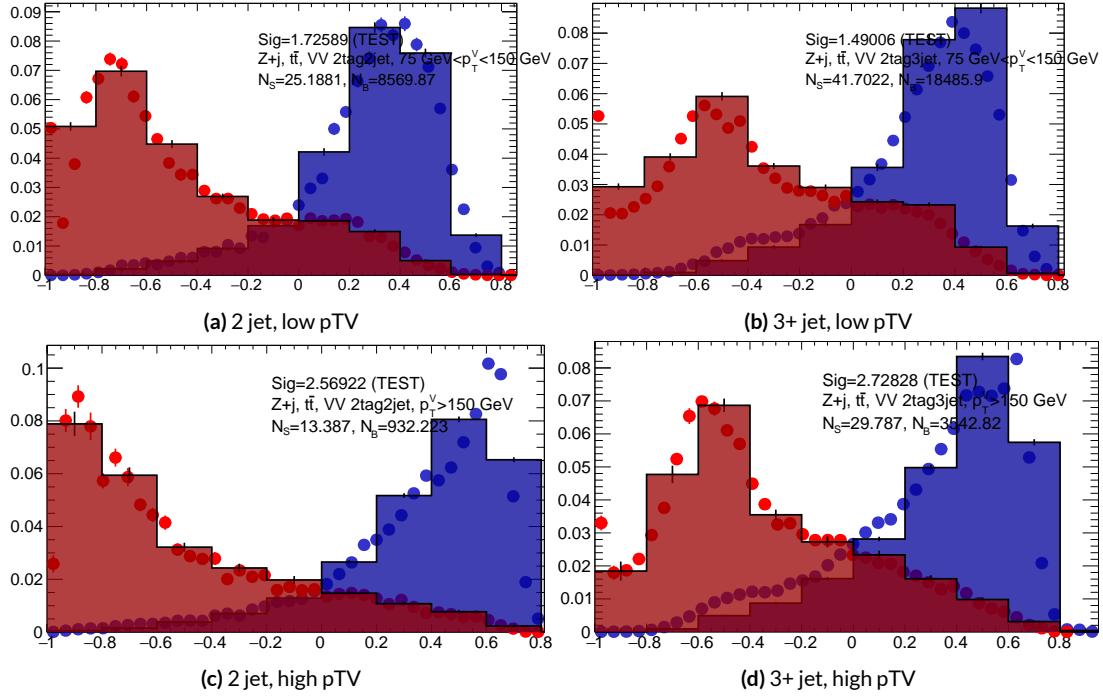
**Figure 7.9:** Input variables for the standard variable set.



**Figure 7.10:** Signal and background variable correlations for the LI variable set.



**Figure 7.11:** Signal and background variable correlations for the LI variable set.



**Figure 7.12:** Training (points) and testing (block histogram) MVA distributions used for stat only testing for the L1 variable set.

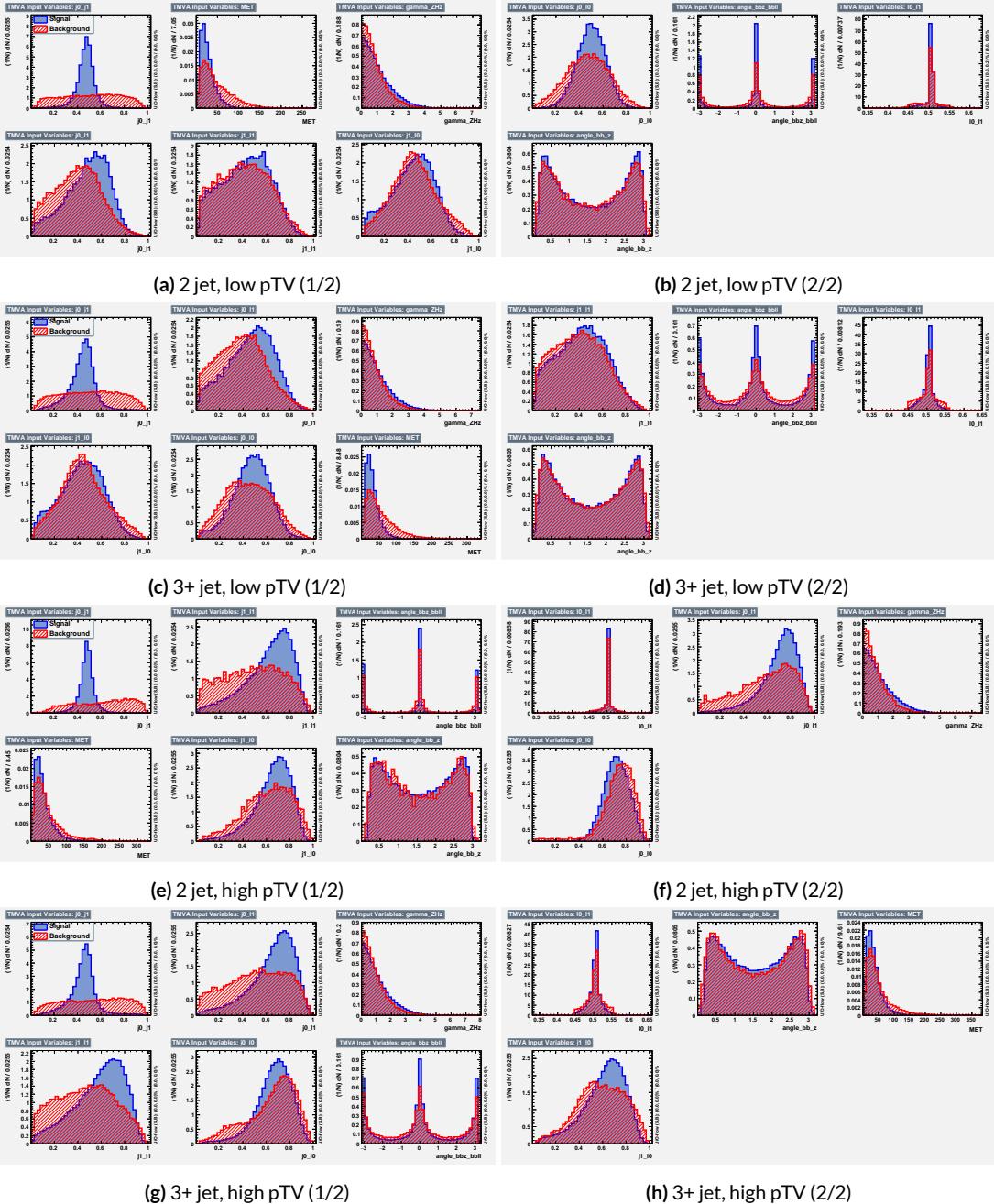


Figure 7.13: Input variables for the LI variable set.

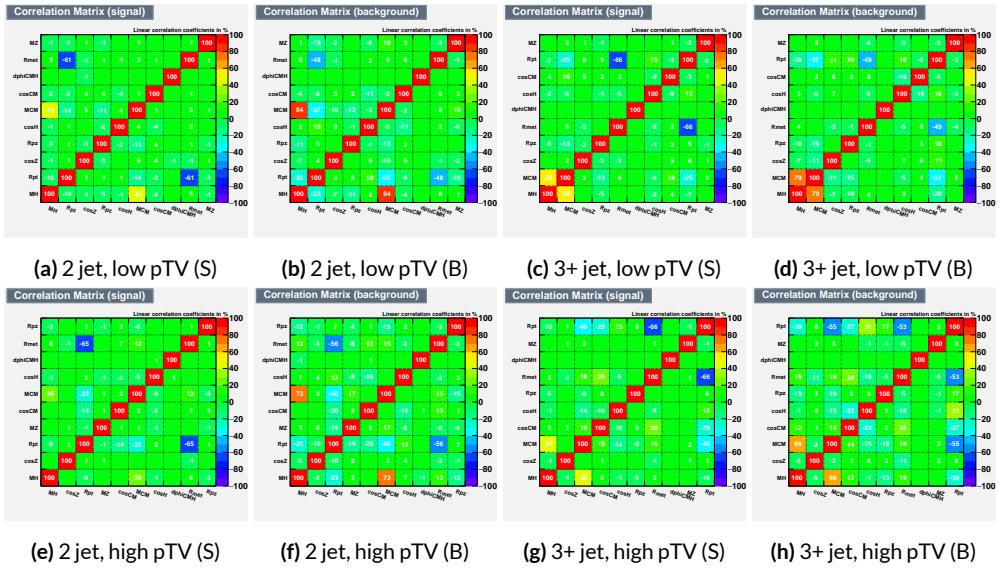


Figure 7.14: Signal and background variable correlations for the RF variable set.

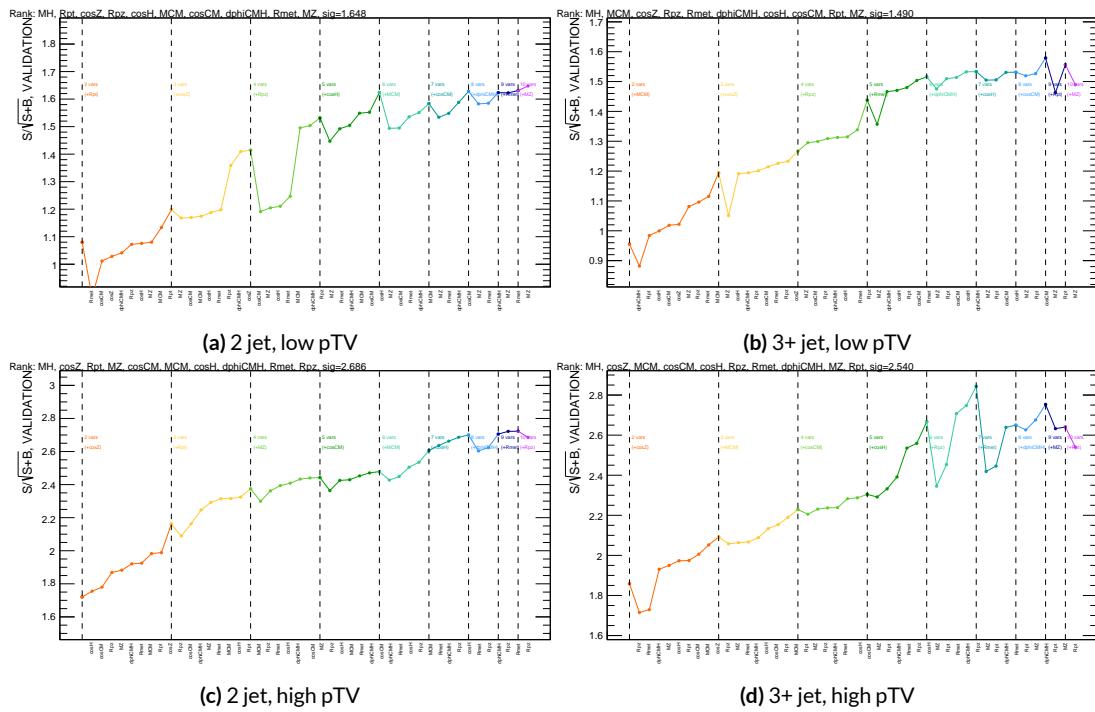
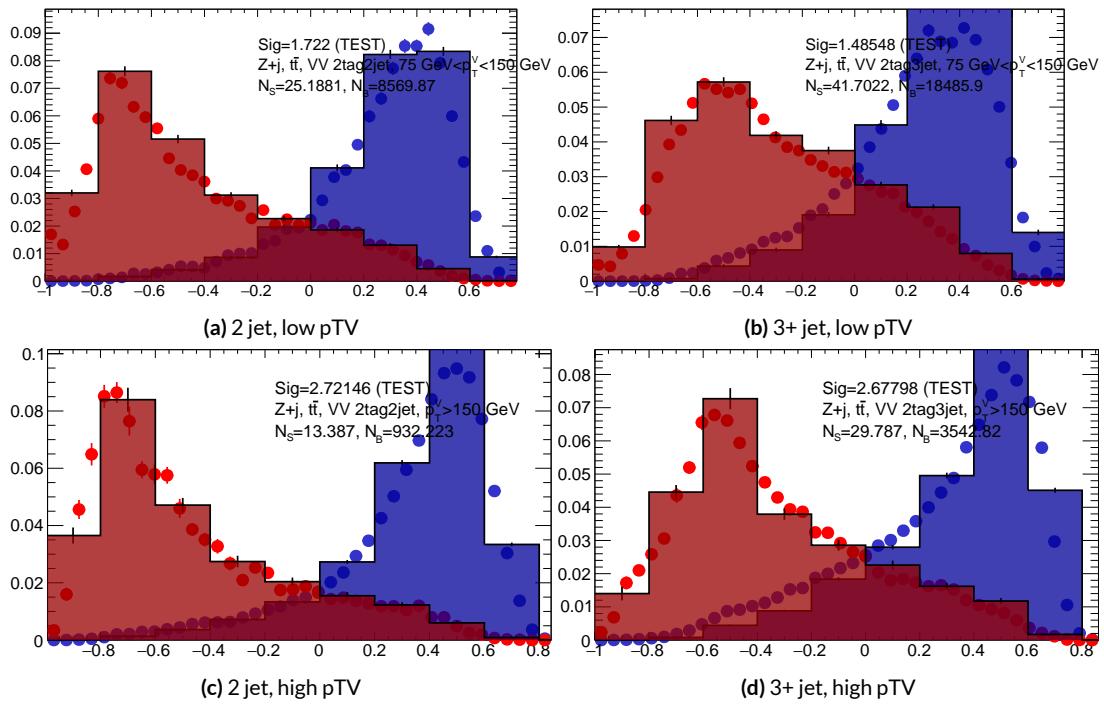


Figure 7.15: Signal and background variable correlations for the RF variable set.



**Figure 7.16:** Training (points) and testing (block histogram) MVA distributions used for stat only testing for the RF variable set.

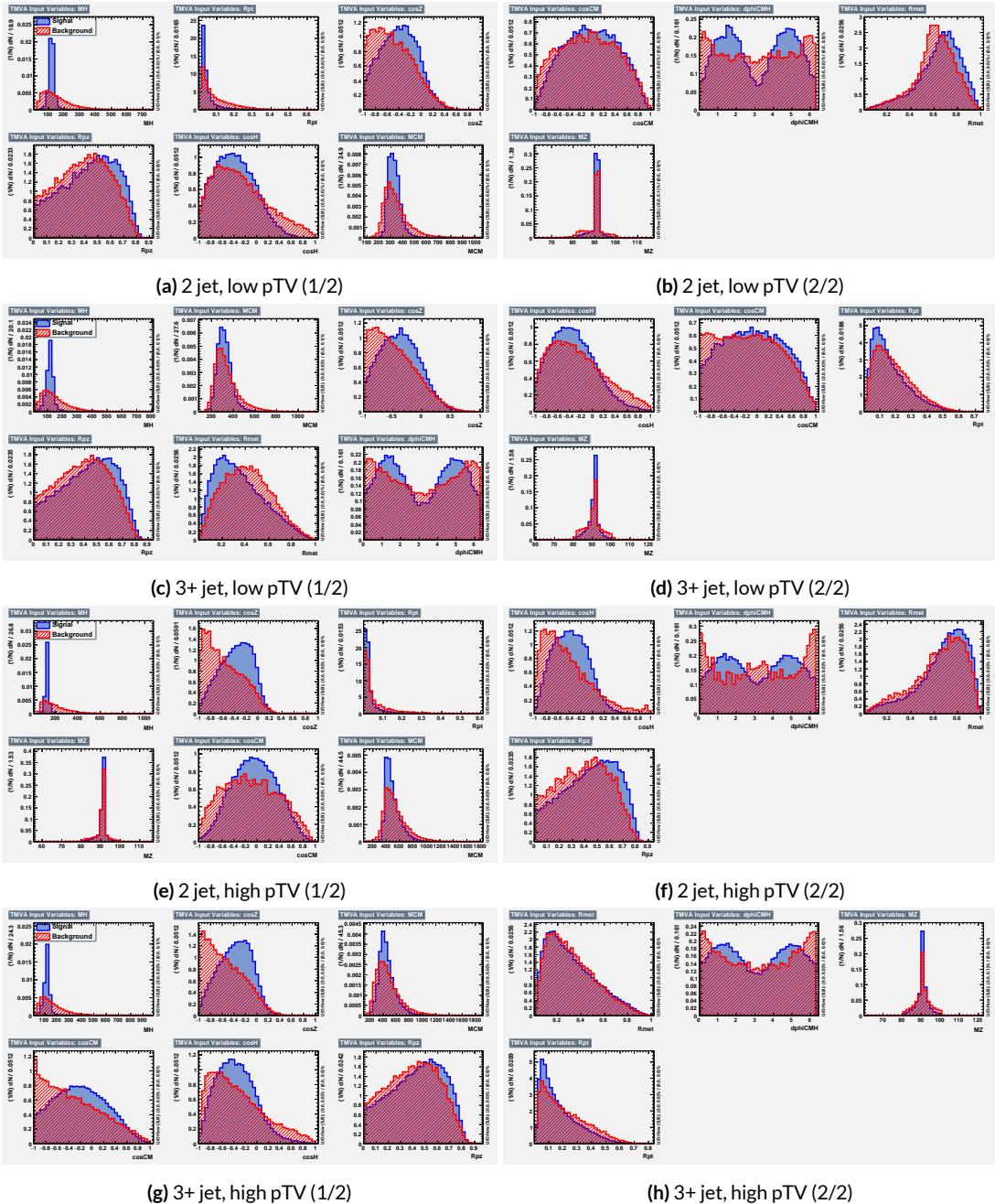


Figure 7.17: Input variables for the RF variable set.

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

Conventional Wisdom

# 8

## Statistical Fit Model and Validation

MUCH HAS BEEN SAID

### 8.1 THE FIT MODEL

In order to derive the strength of the signal process  $ZH \rightarrow \ell\ell b\bar{b}$ , denoted  $\mu$ , and other quantities of interest while taking into account systematic uncertainties or nuisance parameters (NP's, collec-

tively denoted  $\vartheta$ ), a binned likelihood function is constructed as the product over bins of Poisson distributions:

$$\mathcal{L}(\mu, \vartheta) = \text{Pois}(n | \mu S + B) \left[ \prod_{i \in \text{bins}} \frac{\mu s_i + b_i}{\mu S + B} \right] \quad (8.1)$$

where  $n$  is the total number of events observed,  $s_i$  and  $b_i$  are the number of expected signal and background events in each bin, and  $S$  and  $B$  are the total expected signal and background events. The signal and background expectations generally are functions of the NP's  $\vartheta$ . NP's related to the normalization of signal and background processes fall into two categories. The first set is left to float freely like  $\mu$  while the second set are parametrized as log-normally distributed to prevent negative predicted values. All other NP's are parametrized with Gaussian priors. This results in a “penalty” on the NLL discussed below of  $(\hat{\alpha} - \mu)^2 / \sigma^2$ , for NP  $\alpha$ , normally parametrized with mean  $\mu$  and variance  $\sigma^2$  for an MLE of  $\hat{\alpha}$ .

One can maximize<sup>\*</sup> the likelihood in Equation 8.1 for a fixed value of  $\mu$  to derive estimators for the NP's  $\vartheta$ ; values of  $\vartheta$  so derived are denoted  $\hat{\vartheta}$  to emphasize that these are likelihood maximizing for a given  $\vartheta$ . The profile likelihood technique finds the likelihood function's maximum by comparing the values of the likelihood over all possible values of  $\mu$  using these “profiles” and picking the one with the greatest  $\mathcal{L}(\mu, \hat{\vartheta})$  value; these values of  $\mu$  and  $\vartheta$  are denoted  $\hat{\mu}$  and  $\hat{\vartheta}$ . The profile likelihood can further be used to construct a test statistic<sup>†</sup>

$$q = -2 \left( \log \mathcal{L}(\mu, \hat{\vartheta}) - \log \mathcal{L}(\hat{\mu}, \hat{\vartheta}) \right) \quad (8.2)$$

---

<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.

This statistic can be used to derive the usual significance ( $p$  value), by setting  $\mu = 0$  to find the compatibility with the background-only hypothesis<sup>3</sup>. If there is insufficient evidence for the signal hypothesis, the  $CL_s$  method can be used to set limits<sup>3</sup>.

In order to both validate the fit model and study the behavior of fits independent of a given dataset, a so-called “Asimov” dataset can be constructed for a given fit model; this dataset has each bin equal to its expectation value for assumed values of the NP’s and a given  $\mu$  value (in this case,  $\mu = 1$ , the SM prediction).

## 8.2 FIT INPUTS

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

Input distributions in MC are further divided according to their physics process. The signal processes are divided based on both the identity of associated  $V$  and the number of leptons in the final state;  $ZH \rightarrow \ell\ell b\bar{b}$  events are further separated into distributions for  $qq$  and  $gg$  initiated processes.  $V+jets$  events are split according to  $V$  identity and into the jet flavor bins described in Section ??.

Due to the effectiveness of the 2  $b$ -tag requirement suppressing the presence of both  $c$  and  $l$  jets, truth-tagging is used to boost MC statistics in the  $cc$ ,  $cl$ , and  $l$  distributions.<sup>‡</sup> For top backgrounds, single top production is split according to production mode ( $s$ ,  $t$ , and  $Wt$ ), with  $t\bar{t}$  as single category. Diboson background distributions are also split according to the identity of the  $V$ 's ( $ZZ$ ,  $WZ$ , and  $WW$ ). Fit input segmentation is summarized in Table 8.1.

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 8.1:** Fit input segmentation.

### 8.3 SYSTEMATIC UNCERTAINTIES

A full discussion of systematic uncertainties can be found in<sup>31</sup>. A brief summary of the NP's considered in these studies is provided below.

#### 8.3.1 MODELING AND THEORETICAL UNCERTAINTIES

The signal and background physics processes considered in the final statistical fit and their nominal samples are described in Section ???. In addition to the nominal samples, alternate samples, described in<sup>32</sup>, are also used to derive systematic uncertainties, also described there—these are summarized in

---

<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.

Table 8.2 below for the 2-lepton case.  $p_T^V$  systematics are generally shape and normalization, whereas  $m_{bb}$  systematics are shape only; these shape systematics are typically parametrized as linear functions.

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
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 8.2:** Summary of modeling systematic uncertainties.

### 8.3.2 EXPERIMENTAL SYSTEMATICS

A full discussion may be found in <sup>21</sup>, and a full summary table may be found at Table 33 of <sup>33</sup>.

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/resolution
Muons	trigger eff, reco/ID eff, isolation eff, track to vertex association, momentum resolution/scale
Event	total luminosity, pileup reweighting

**Table 8.3:** Summary of experimental systematic uncertainties.

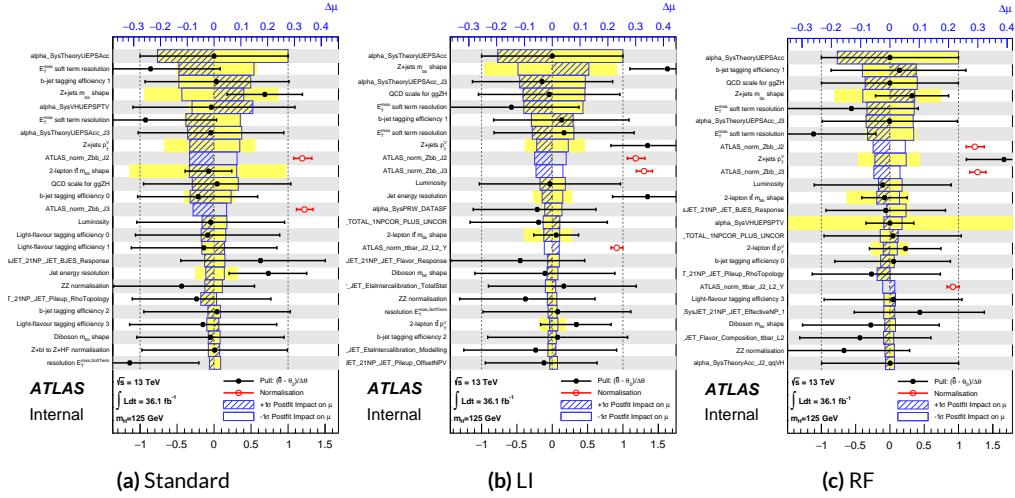
For VZ validations fit, cf. Appendix ??.

For fit model validation, cf. Appendix ??.

## 8.4 FULL BREAKDOWN OF ERRORS

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

sets may be found in Figure 8.1.



**Figure 8.1:** 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.

	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

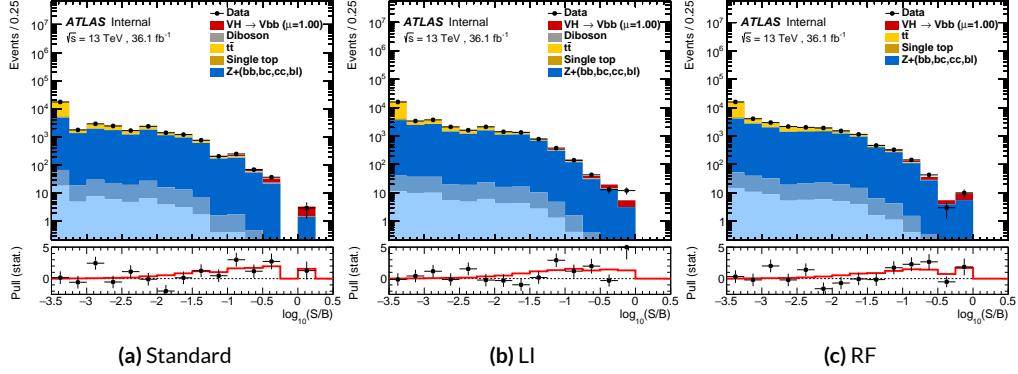
**Table 8.4:** Expected error breakdowns for the standard, LI, and RF variable sets

	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 8.5:** Observed signal strengths, and error breakdowns for the standard, LI, and RF variable sets

## 8.5 S/B PLOT

Plots for the binned S/B in signal region distributions may be found in Figure 8.2.



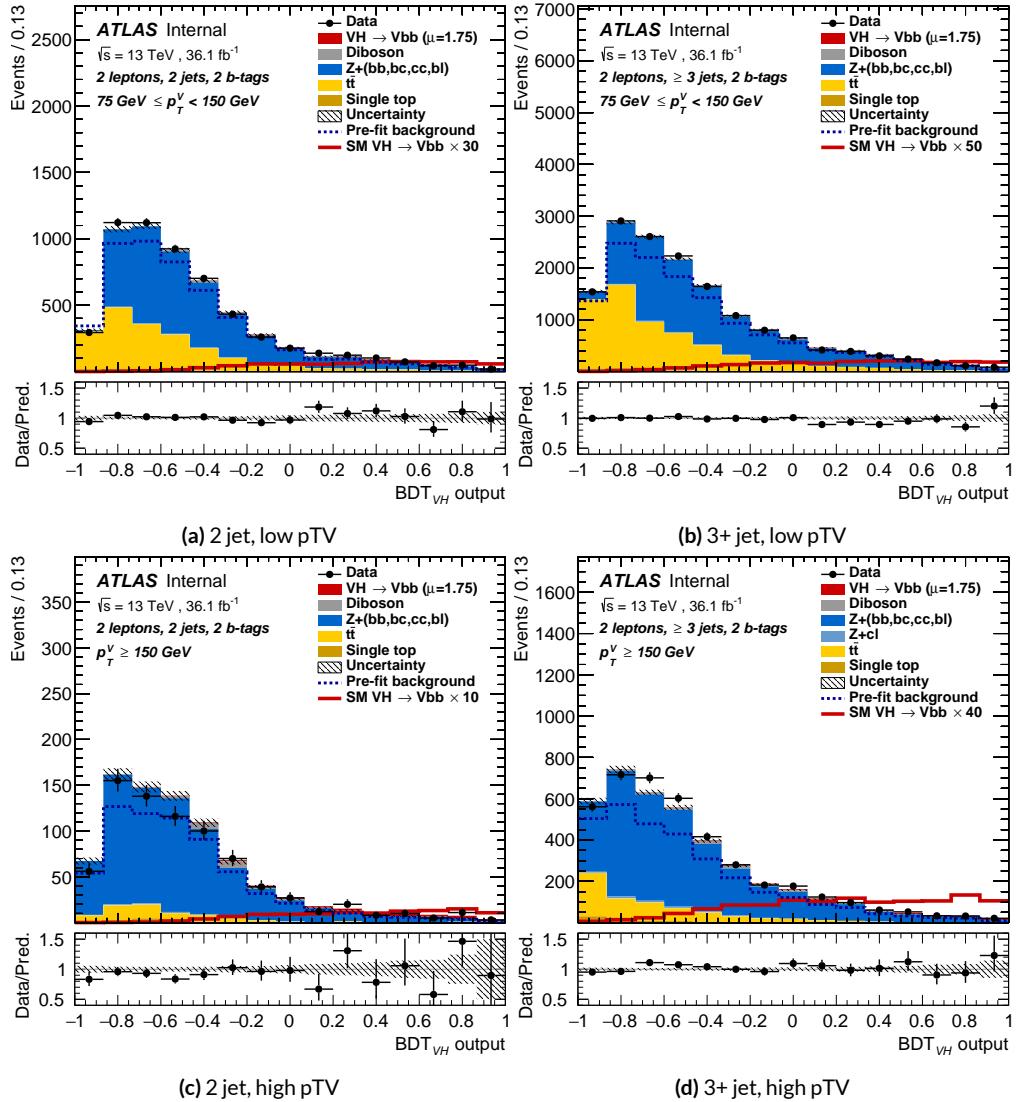
**Figure 8.2:** Binned S/B plots for the standard (a), LI (b), and RF (c) variable sets.

## 8.6 POSTFIT DISTRIBUTIONS

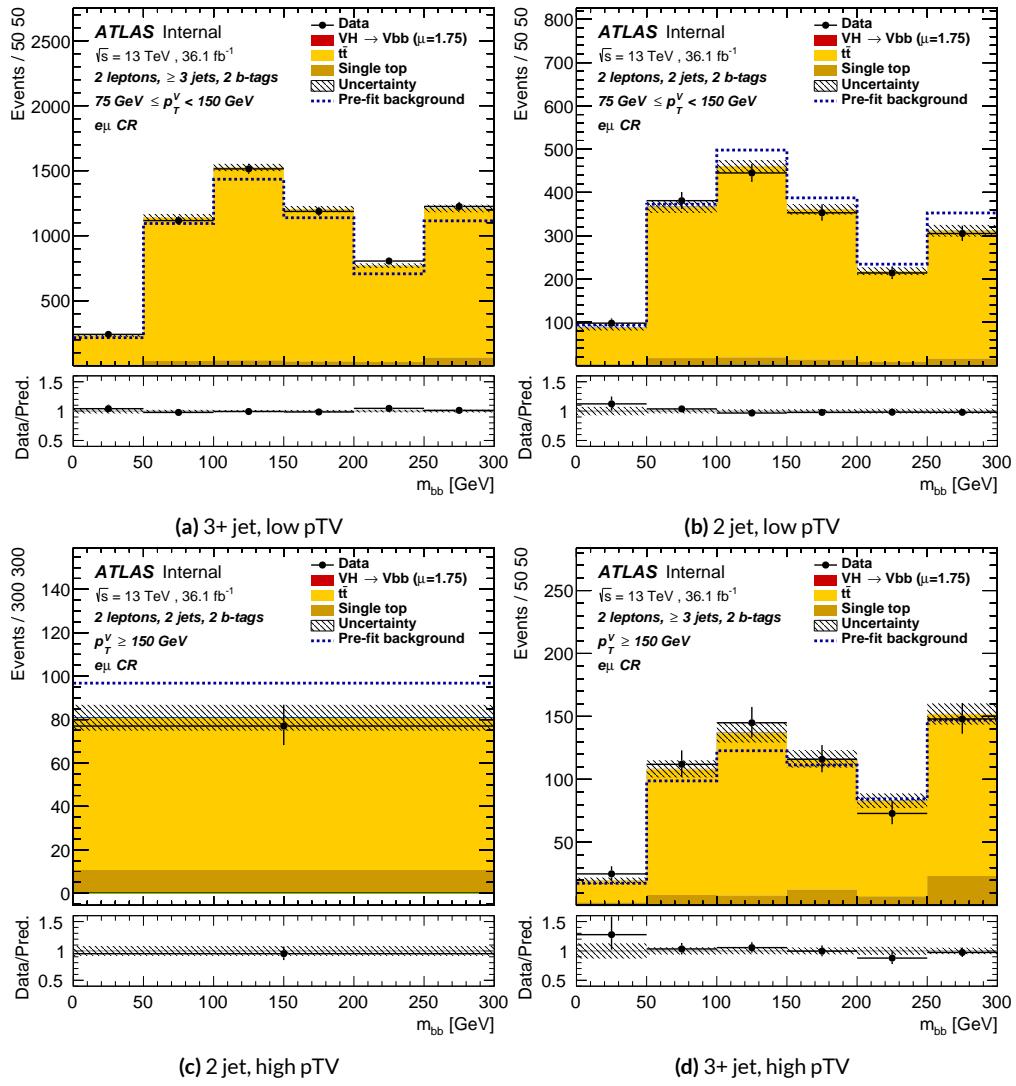
Postfit distributions for the MVA discriminant ( $m_{bb}$ ) distribution in the signal (top  $e - \mu$  control) region for the standard, Lorentz Invariant, and RestFrames variable sets.

## 8.7 NUISANCE PARAMETER PULLS

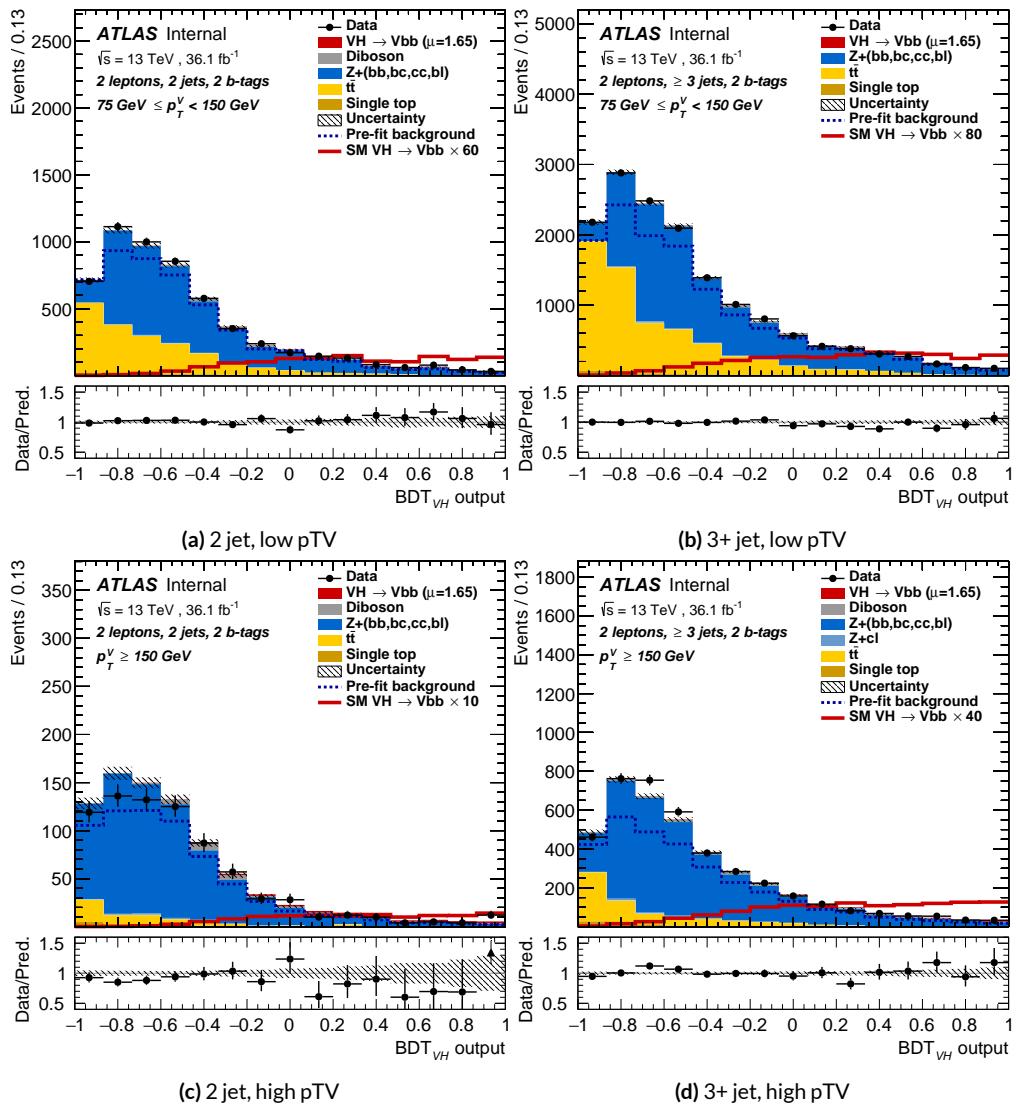
As can be seen in Figures 8.9–8.13, the fits for the three different variable sets are fairly similar from a NP pull perspective. Black is the standard variable set, red is the LI set, and blue is the RF set.



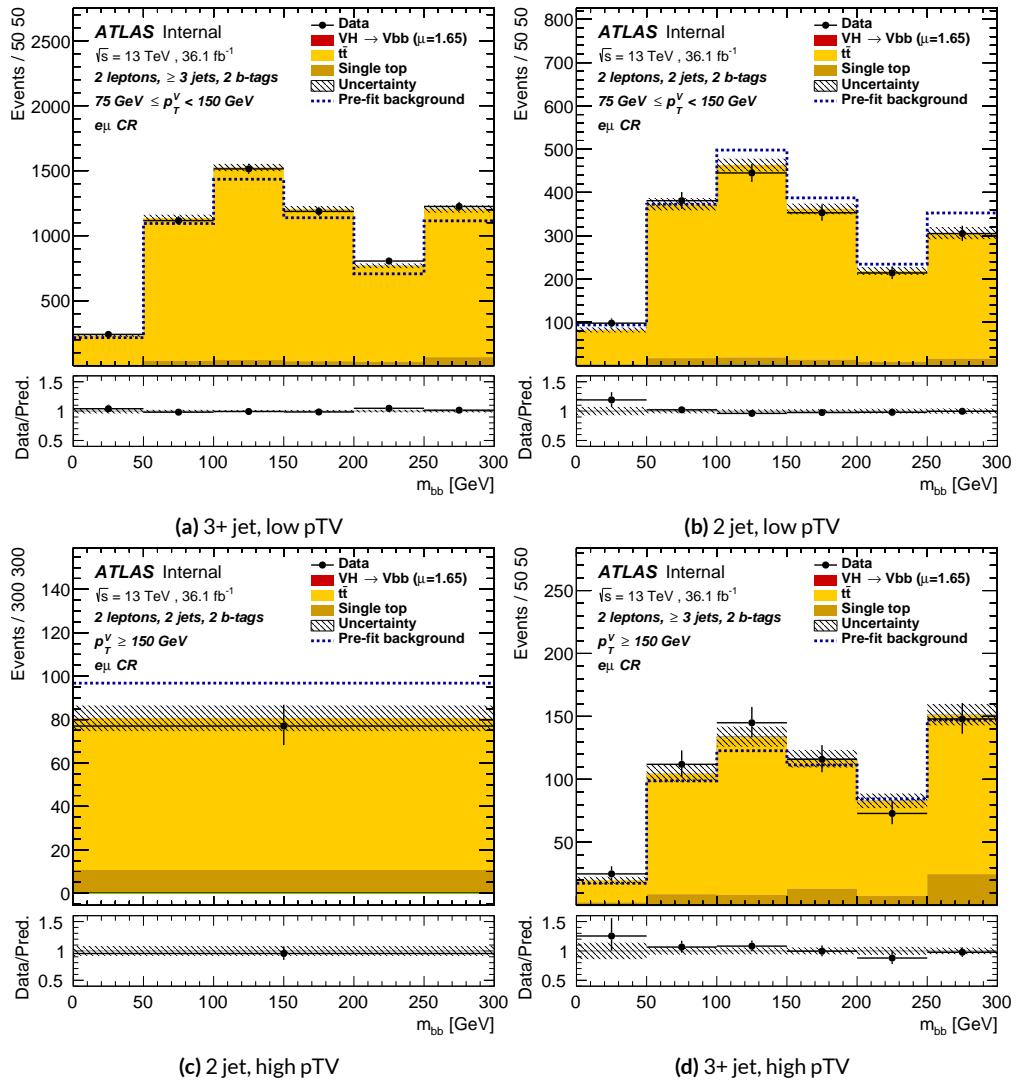
**Figure 8.3:** Postfit  $BDT_{VH}$  plots in the signal region for the standard variable set.



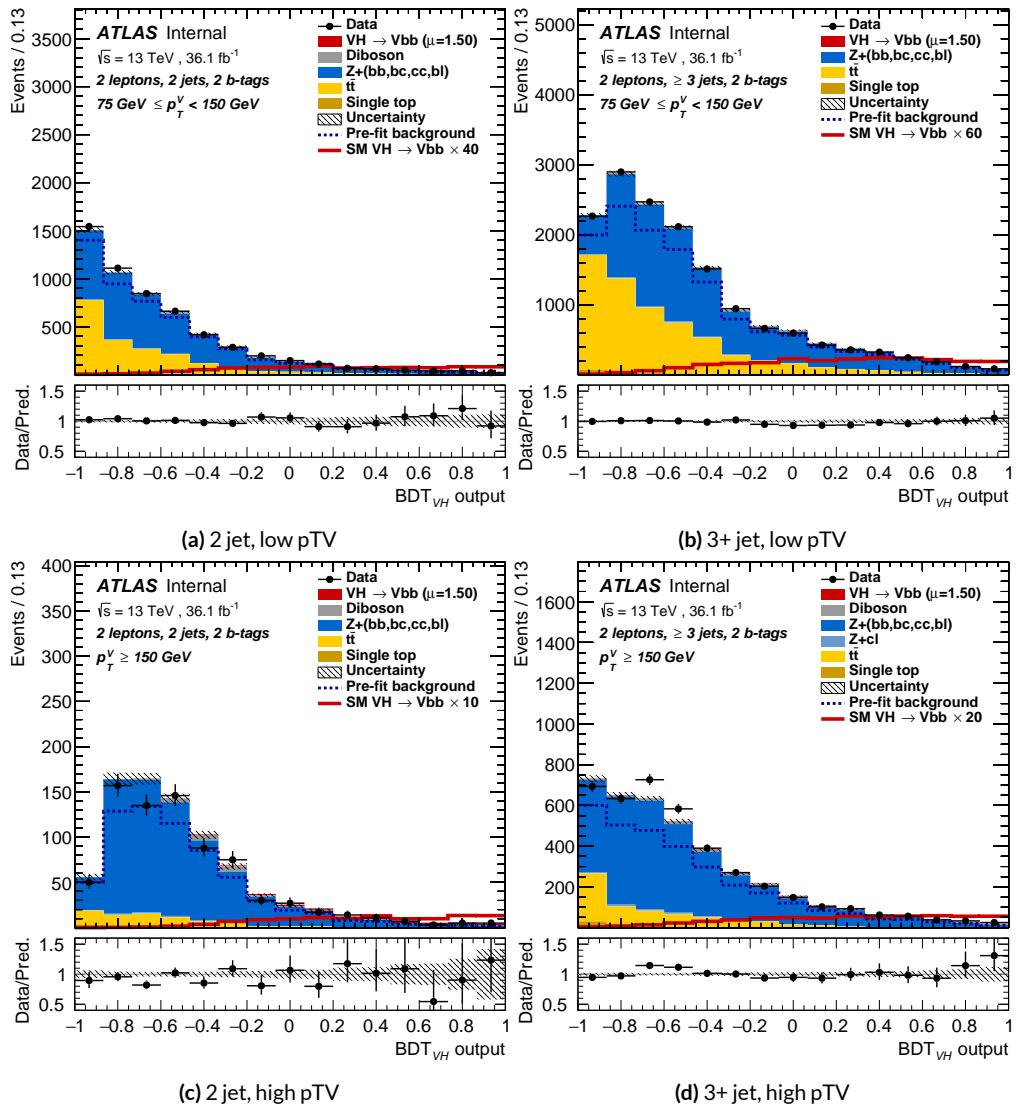
**Figure 8.4:** Postfit  $m_{bb}$  plots in the top  $e - \mu$  CR for the standard variable set.



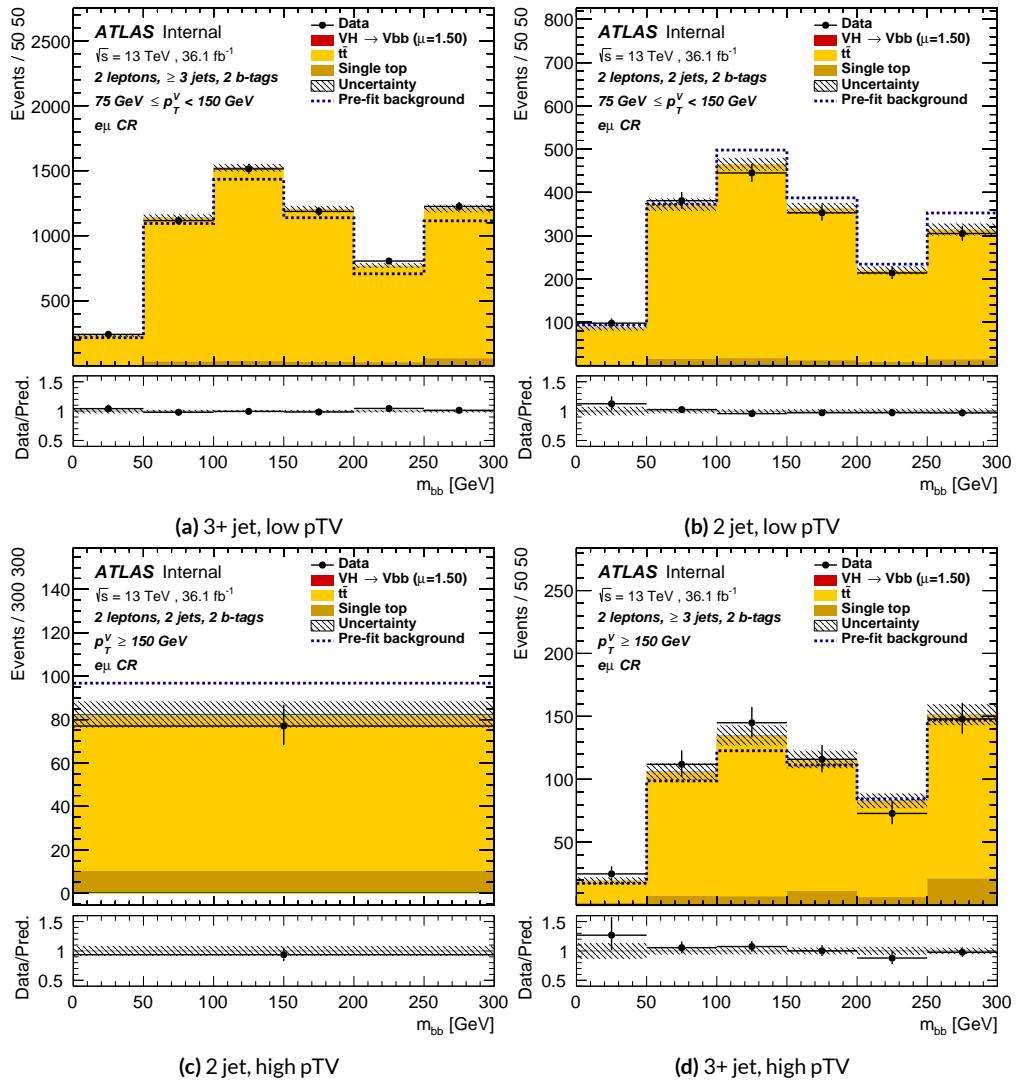
**Figure 8.5:** Postfit  $BDT_{VH}$  plots in the signal region for the LI variable set.



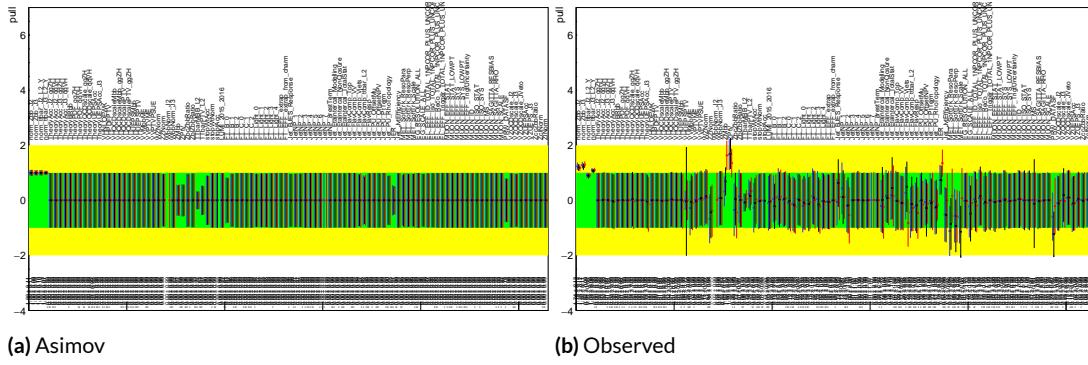
**Figure 8.6:** Postfit  $m_{bb}$  plots in the top  $e - \mu$  CR for the LI variable set.



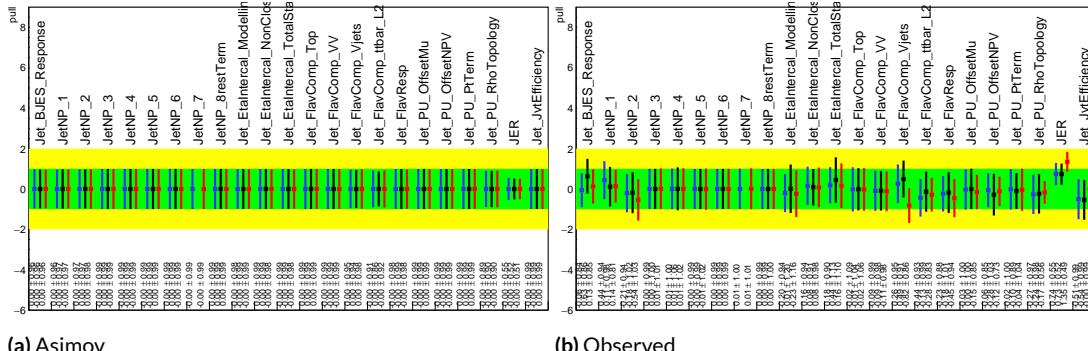
**Figure 8.7:** Postfit  $BDT_{VH}$  plots in the signal region for the RF variable set.



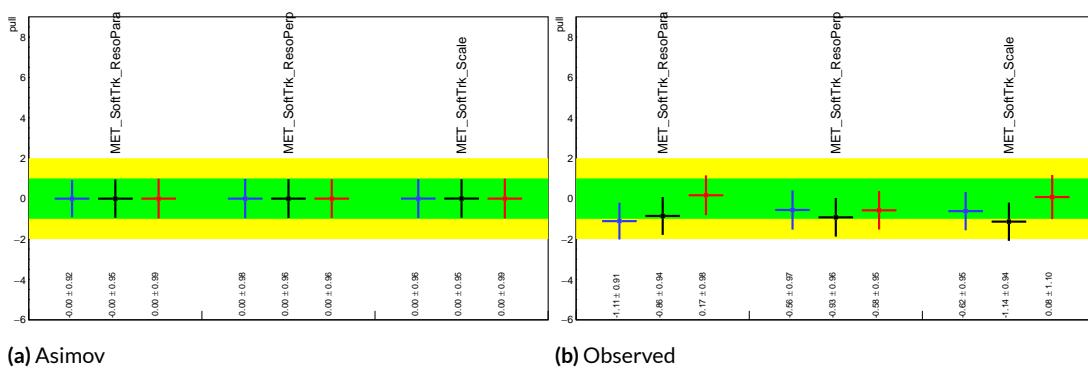
**Figure 8.8:** Postfit  $m_{bb}$  plots in the top  $e - \mu$  CR for the RF variable set.



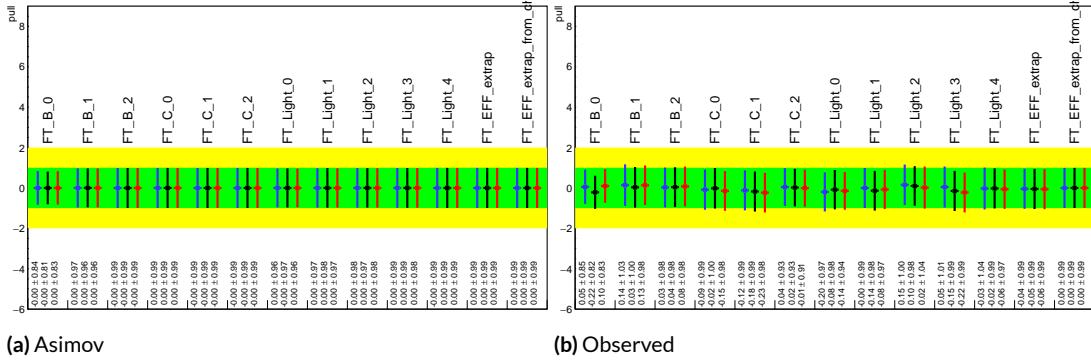
**Figure 8.9:** Pull comparison for all NP's but MC stats.



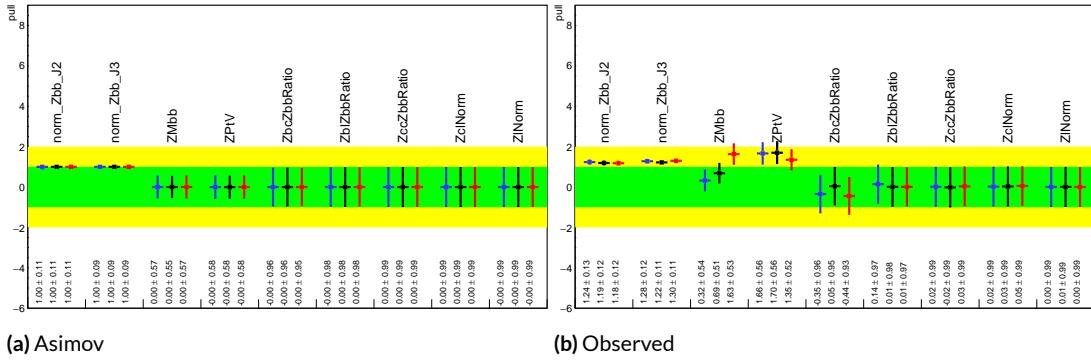
**Figure 8.10:** Pull comparison for jet NP's.



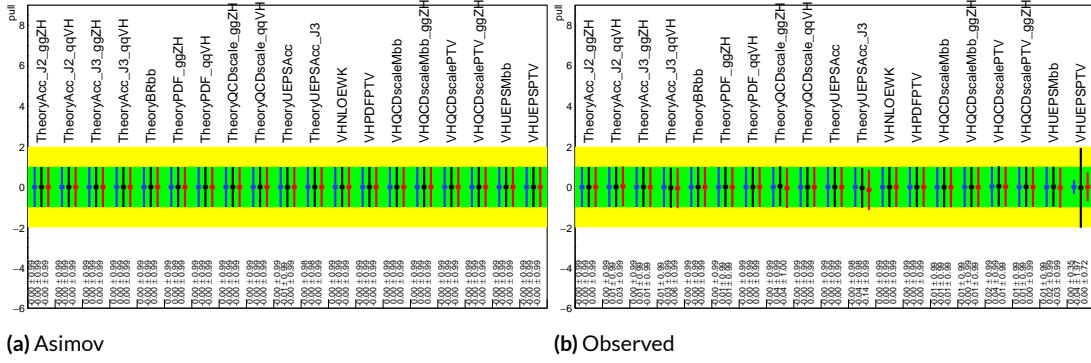
**Figure 8.11:** Pull comparison for MET NP's.



**Figure 8.12:** Pull comparison for Flavour Tagging NP's.



**Figure 8.13:** Pull comparison for  $Z + \text{jets}$  NP's.



**Figure 8.14:** Pull comparison for signal process modeling NP's.

## 8.8 NUISANCE PARAMETER CORRELATIONS

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

variable set fits can be found in Figure ?? for Asimov fits and Figure ?? for observed fits.

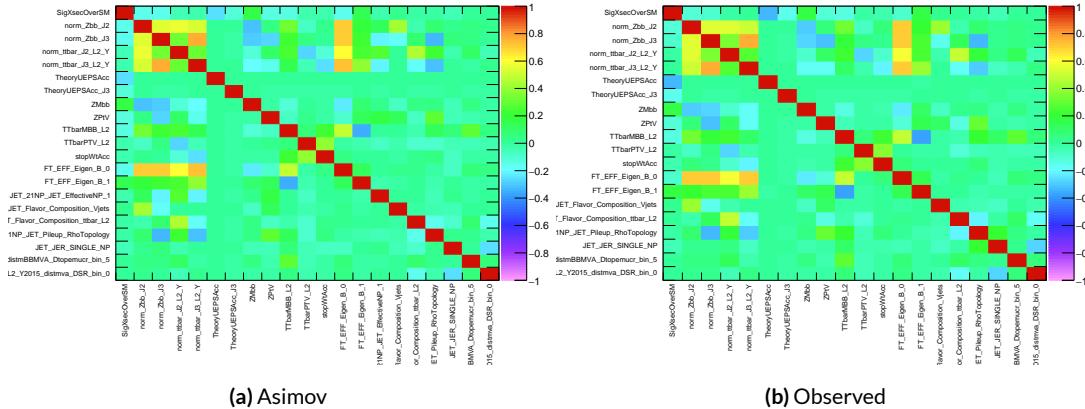


Figure 8.15: NP correlations for standard variable fits.

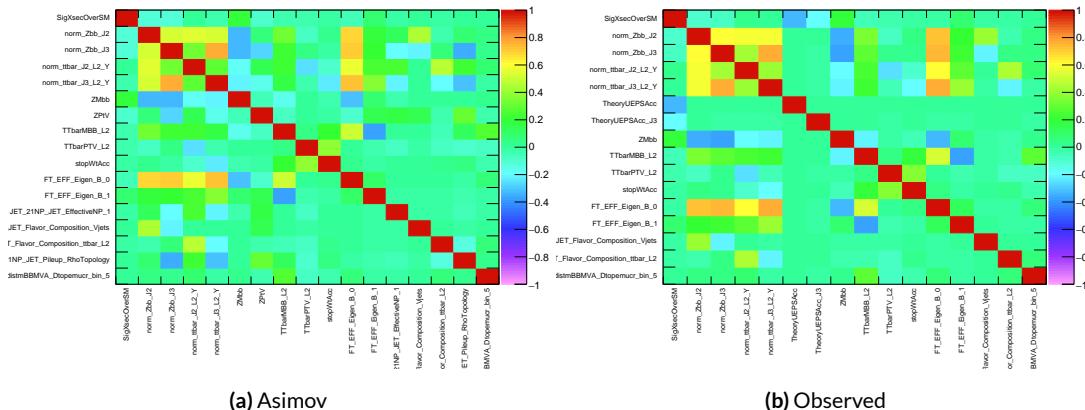
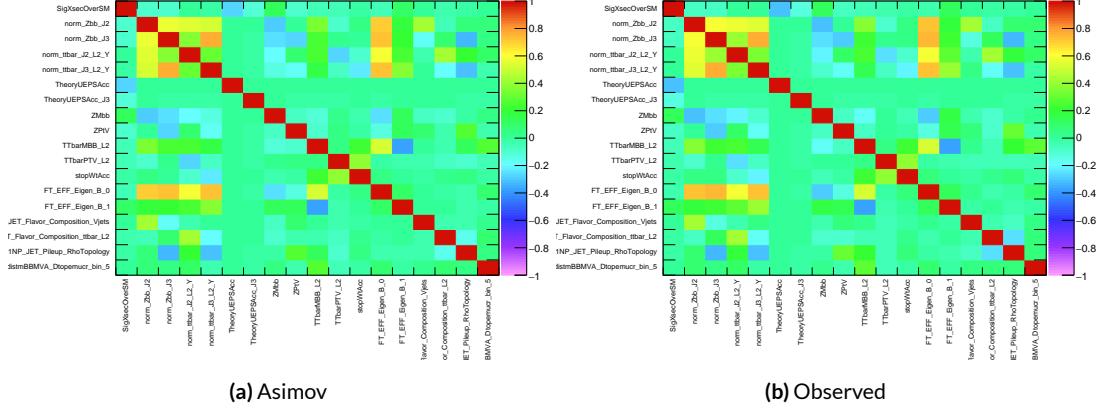


Figure 8.16: NP correlations for L1 variable fits.



**Figure 8.17:** NP correlations for RF variable fits.

## 8.9 SUMMARY OF RESULTS

One of the primary validation cross-checks for the fiducial analysis was a  $VZ$  fit—that is, conducting the entire analysis but looking for  $Z \rightarrow b\bar{b}$  decays instead of the Higgs. To do this, a new MVA discriminant is made by keeping all hyperparameter configurations the same (e.g. variable ranking) but using diboson samples as signal. For the 2-lepton case, this means using  $ZZ \rightarrow \ell\ell b\bar{b}$  as the signal sample. This new MVA is used to make the inputs described in Section 8.2, and the fit is then run as for the  $VH$  fit (again, with  $ZZ$  as signal).

The  $VZ$  fit sensitivities for the standard, LI, and RF fits are summarized in Table 8.6. The expected significances are all fairly comparable and about what was the case in the fiducial analysis.

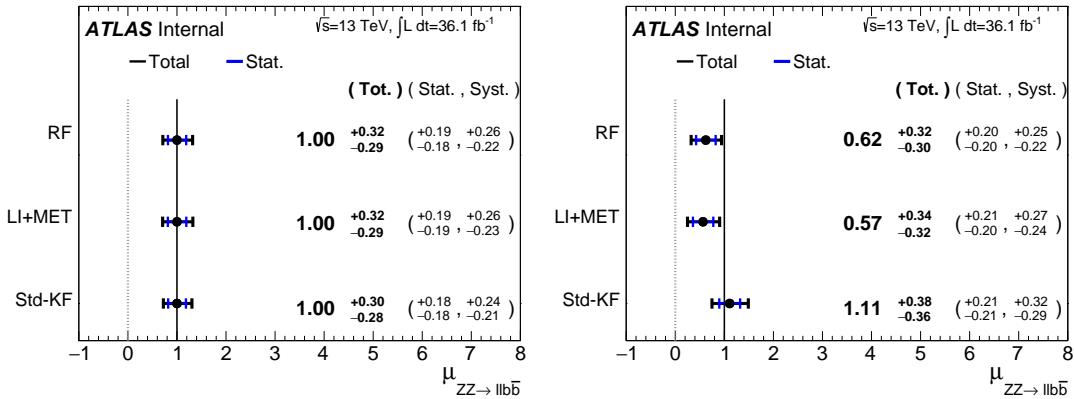
The observed significance for the standard set matches fairly well with the expected value on data, but the LI and RF observed significances are quite a bit lower.

These values, however, are consistent with the observed signal strength values, which can be seen in Figure 8.18 (b), with both the LI and RF fits showing a deficit of signal events with respect to the

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

SM expectation, though not by much more than one standard deviation (a possible explanation is explored in the following section). Just as in the  $VH$  fits, errors arising systematic uncertainties are lower in the fits to the observed dataset. That the effect is not noticeable in Asimov fits is not too surprising, since this analysis (and these variable configurations in particular), is not optimized for  $VZ$ .



**Figure 8.18:**  $\mu$  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).

## 8.10 2 AND $\geq 3$ JET FITS

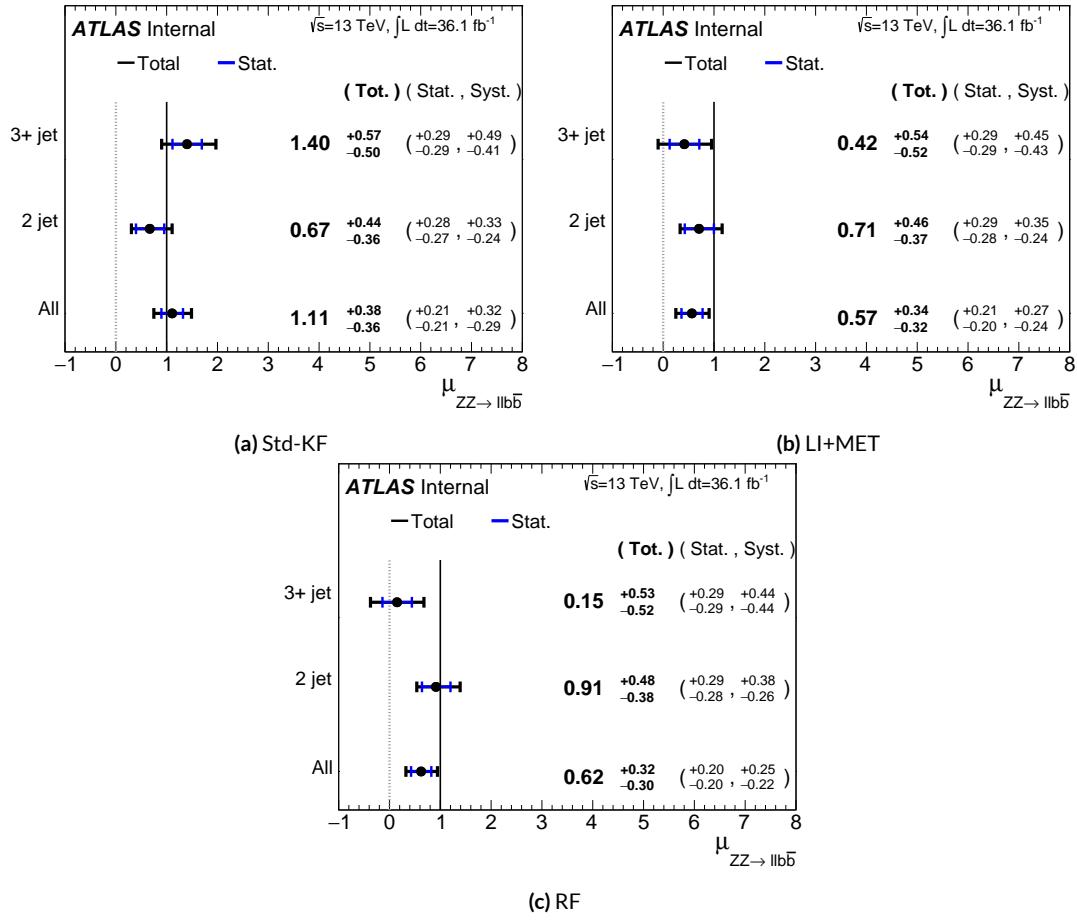
While the treatment of simply ignoring any additional jets in the event is fine for the treatment in the  $VH$  analysis, the potential shortcoming of this treatment appears in the  $VZ$  analysis when the  $2$  and  $\geq 3$  jet cases are fit separately<sup>§</sup>, as can be seen in Figure 8.19. Compared to the standard fit, the LI and RF fits have lower  $\hat{\mu}_{\geq 3 \text{jet}}$  values, consistent with the interpretation that the additional information in the  $\geq 3$  jet regions for the standard case is important for characterizing events in these regions for  $VZ$  fits.

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

If either the LI or RF schemes were to be used in a mainstream analysis, these validation fits suggest that perhaps events with 4 or more jets should be excluded (as in the 0 and 1-lepton cases) or that the third jet ought to be included in variable schemes (e.g. by adding the third jet to the Higgs

---

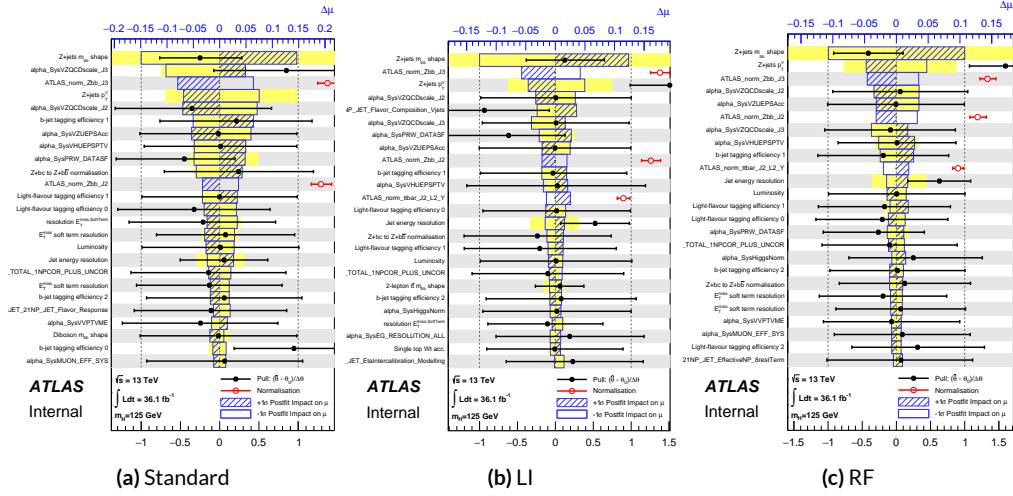
<sup>§</sup>standalone fits, with half the regions each, not 2 POI fits



**Figure 8.19:**  $\hat{\mu}$  summary plots with standalone fits for the different  $n_{jet}$  regions for the standard, LI, and RF variable sets.

in the high  $p_T^V$  case). Nevertheless, this optimization is beyond the scope of these studies, which aim to preserve as much of the fiducial analysis as possible for as straightforward a comparison as possible.

## 8.II NUISANCE PARAMETER RANKING PLOTS AND BREAKDOWNS



**Figure 8.20:** 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.

	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 8.7:** Summary of impact of various nuisance parameter categories on the error on  $\mu$  for Asimov fits for the standard, LI, and RF variable sets.

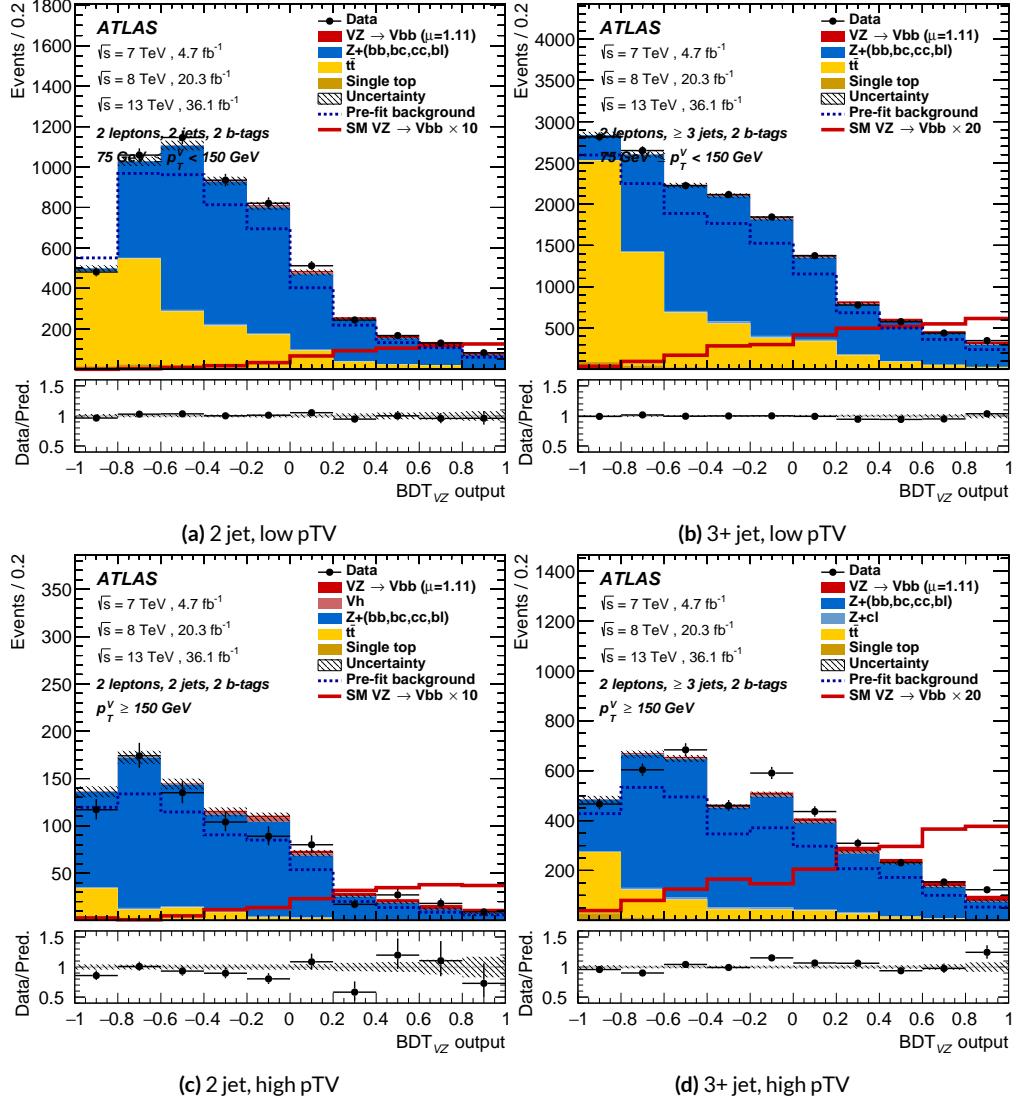
	Std-KF	LI+MET	RF
$\hat{\mu}$	1.1079	0.5651	0.6218
Total	+0.381 / -0.360	+0.339 / -0.316	+0.322 / -0.299
DataStat	+0.214 / -0.211	+0.210 / -0.205	+0.201 / -0.197
FullSyst	+0.315 / -0.292	+0.267 / -0.241	+0.252 / -0.225
Floating normalizations	+0.120 / -0.122	+0.095 / -0.089	+0.082 / -0.079
All normalizations	+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 8.8:** 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.

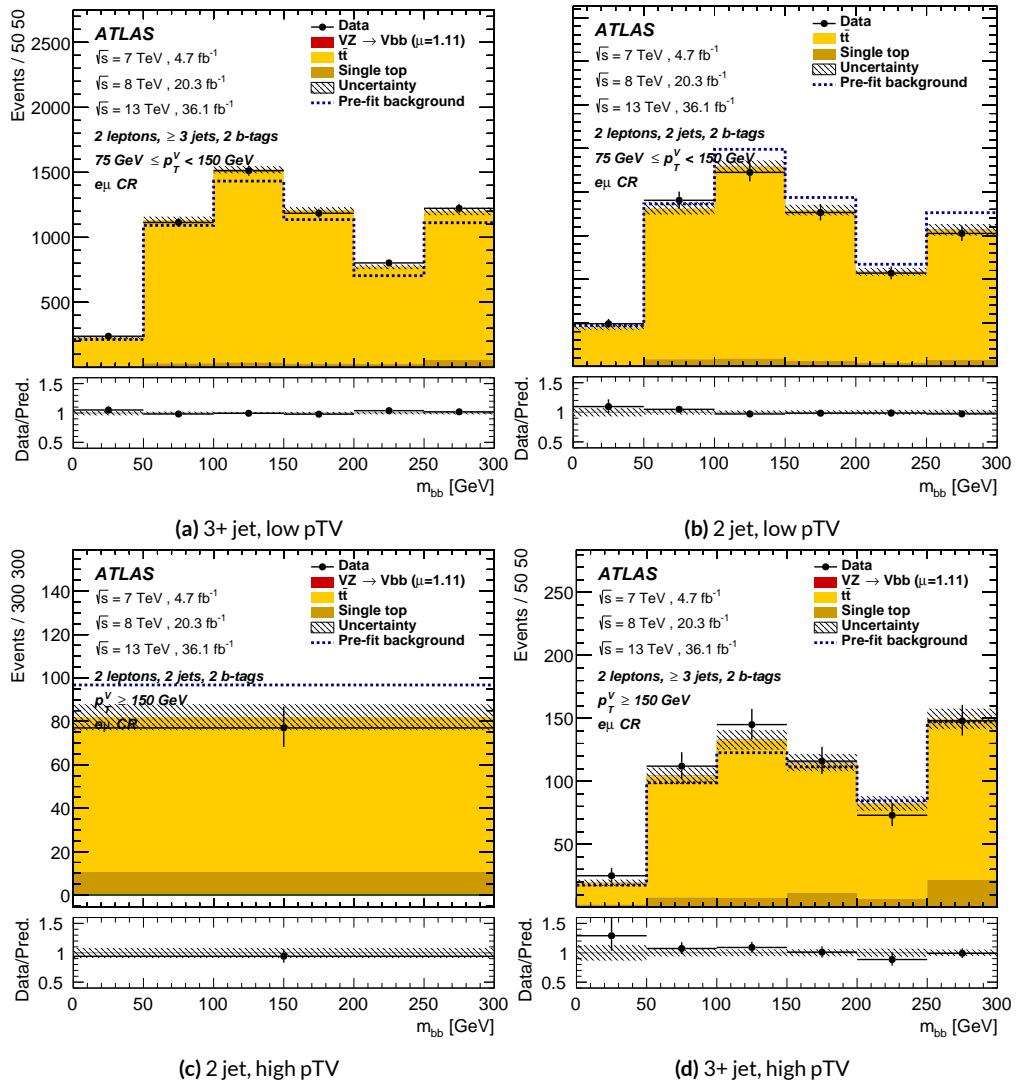
## 8.12 POSTFIT DISTRIBUTIONS

Postfit distributions for the MVA discriminant ( $m_{bb}$ ) distribution in the signal (top  $e - \mu$  control)

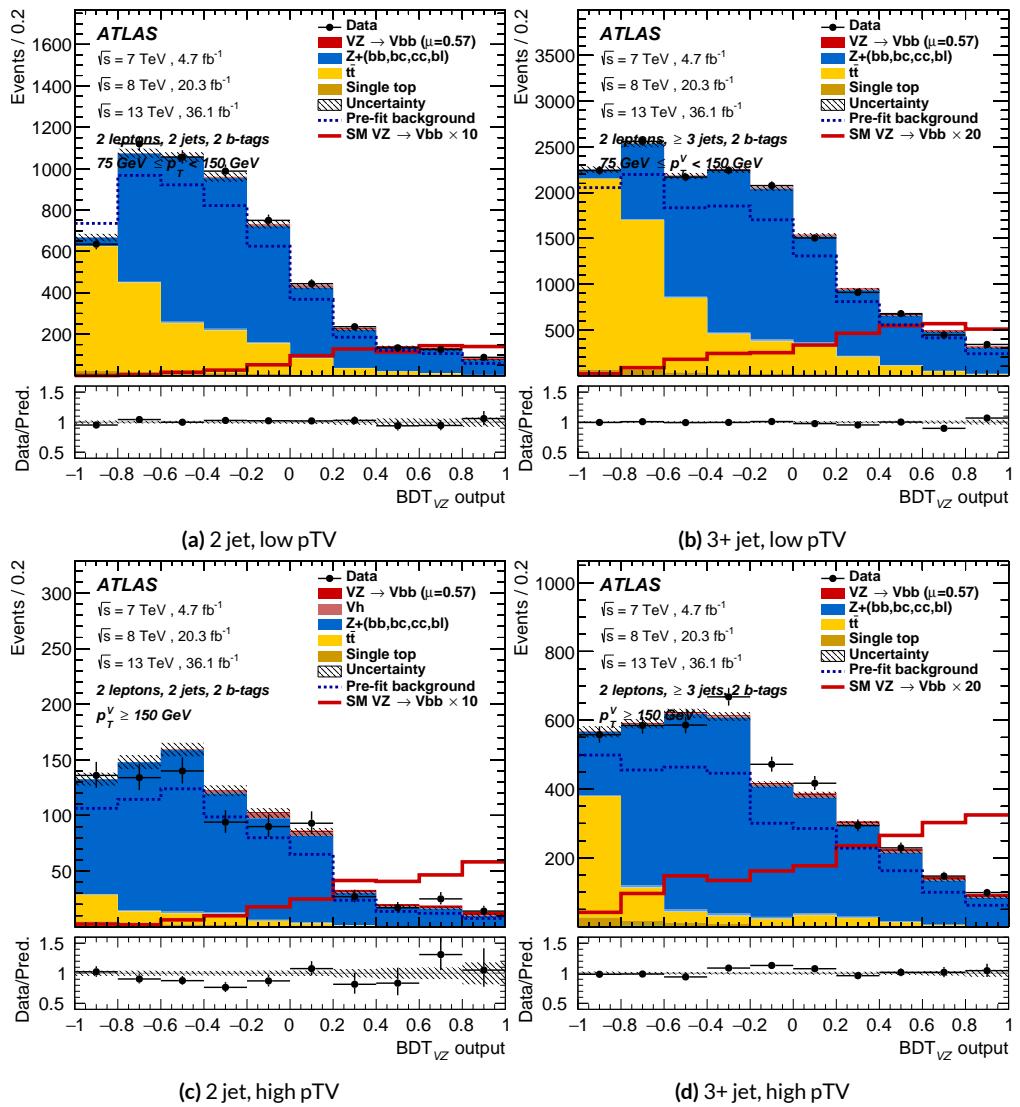
region for the standard, Lorentz Invariant, and RestFrames variable sets.



**Figure 8.21:** Postfit  $BDT_{VH}$  plots in the signal region for the standard variable set.



**Figure 8.22:** Postfit  $m_{bb}$  plots in the top  $e - \mu$  CR for the standard variable set.



**Figure 8.23:** Postfit  $BDT_{VH}$  plots in the signal region for the LI variable set.

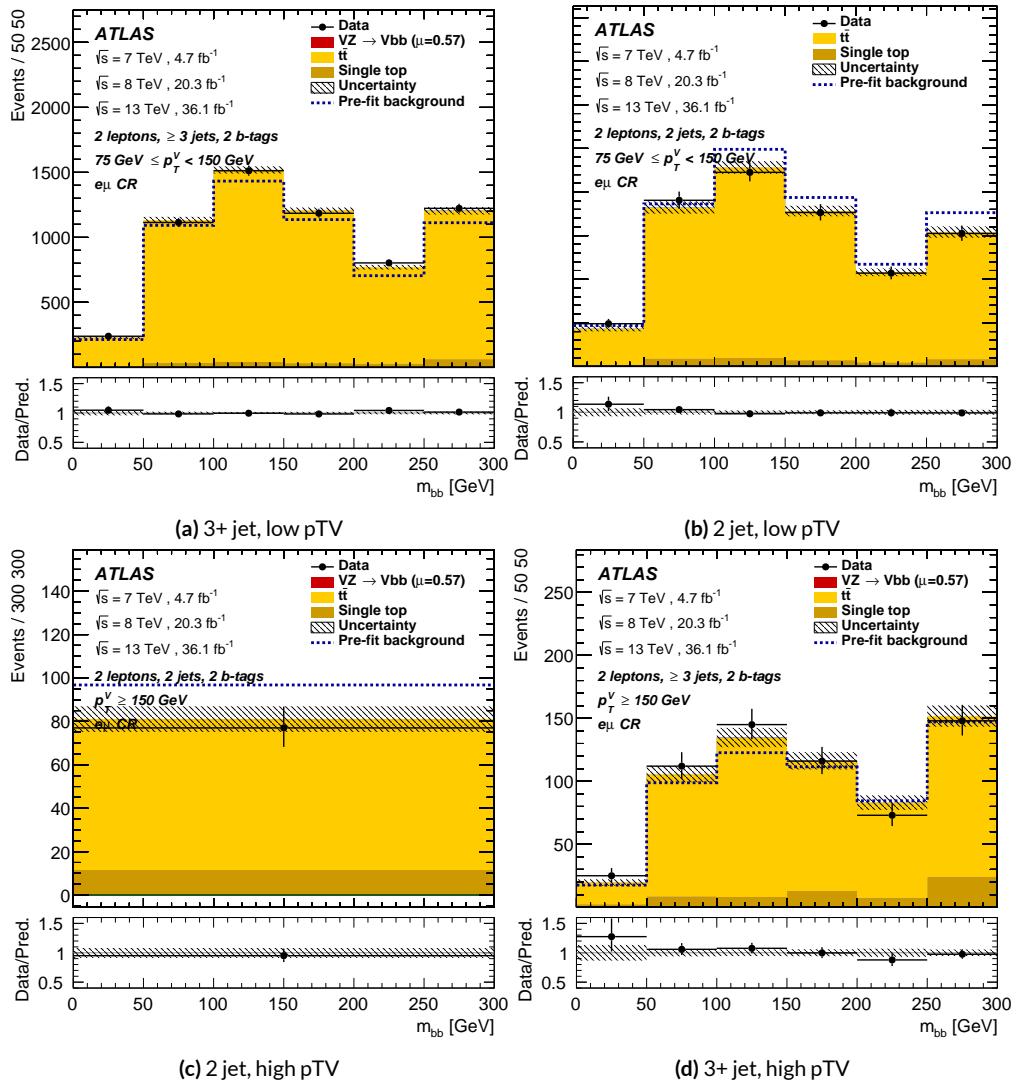
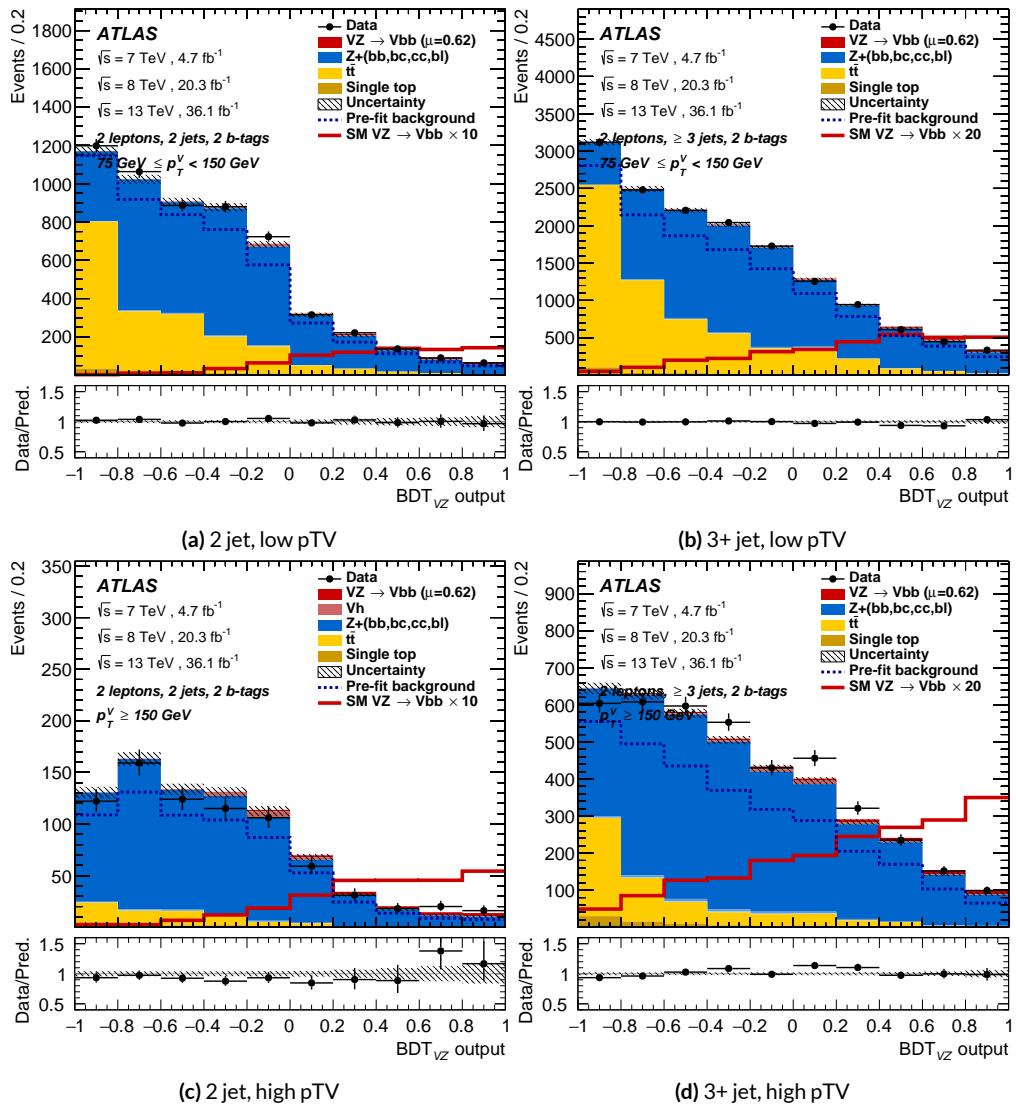


Figure 8.24: Postfit  $m_{bb}$  plots in the top  $e - \mu$  CR for the LI variable set.



**Figure 8.25:** Postfit  $BDT_{VH}$  plots in the signal region for the RF variable set.

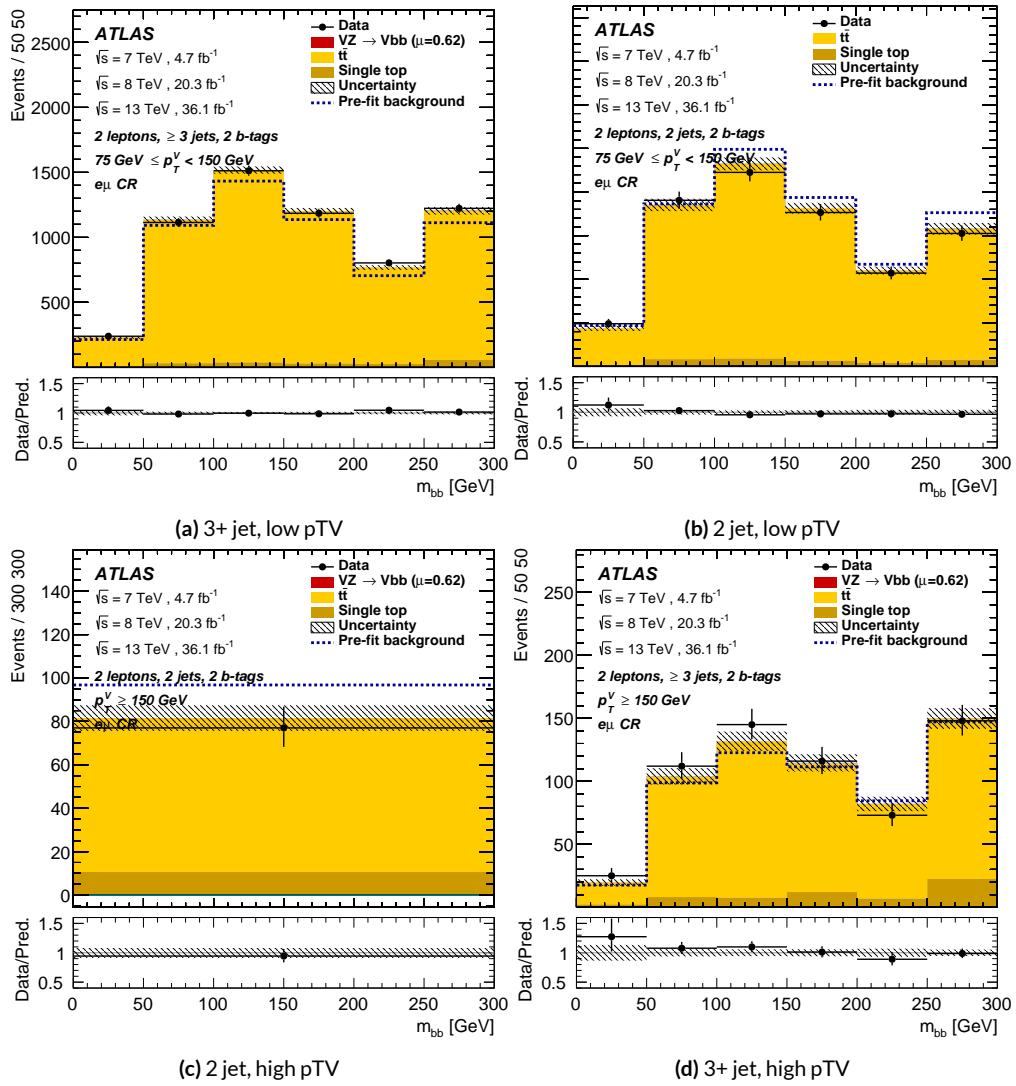
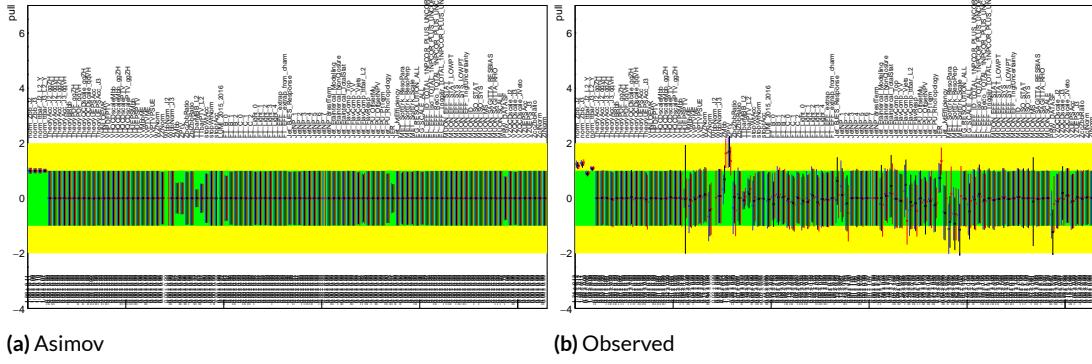
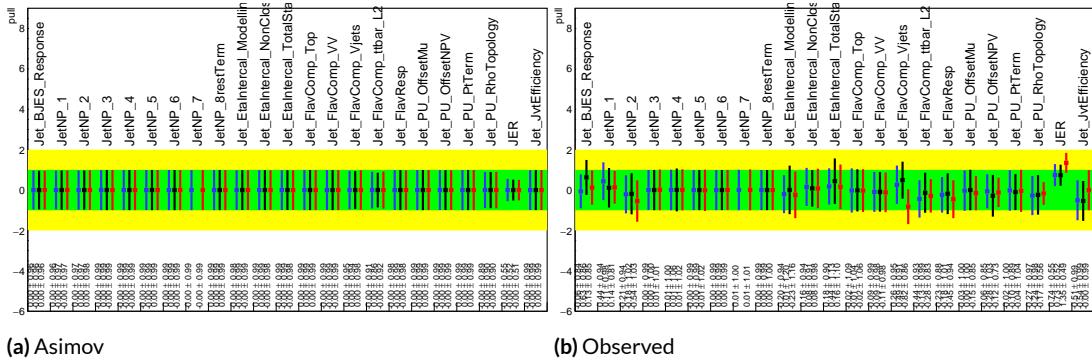


Figure 8.26: Postfit  $m_{bb}$  plots in the top  $e - \mu$  CR for the RF variable set.



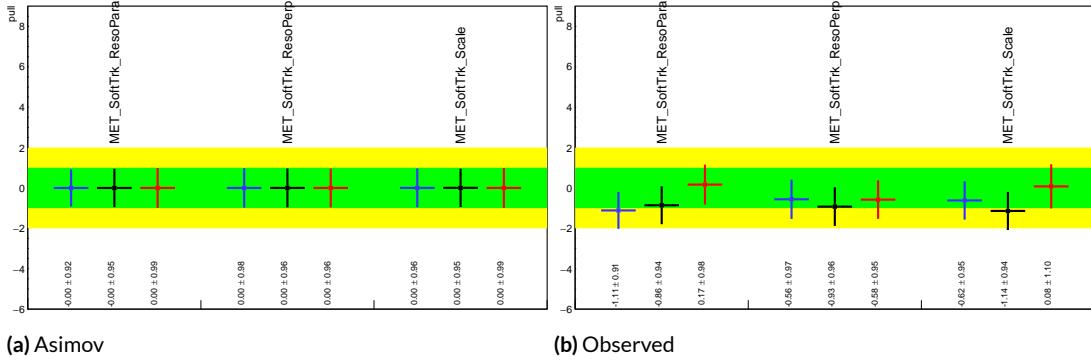
**Figure 8.27:** Pull comparison for all NP's but MC stats.



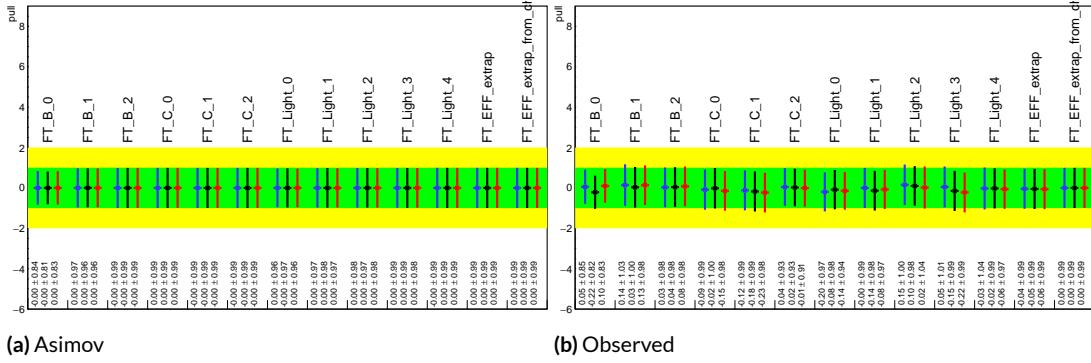
**Figure 8.28:** Pull comparison for jet NP's.

### 8.13 NUISANCE PARAMETER PULLS

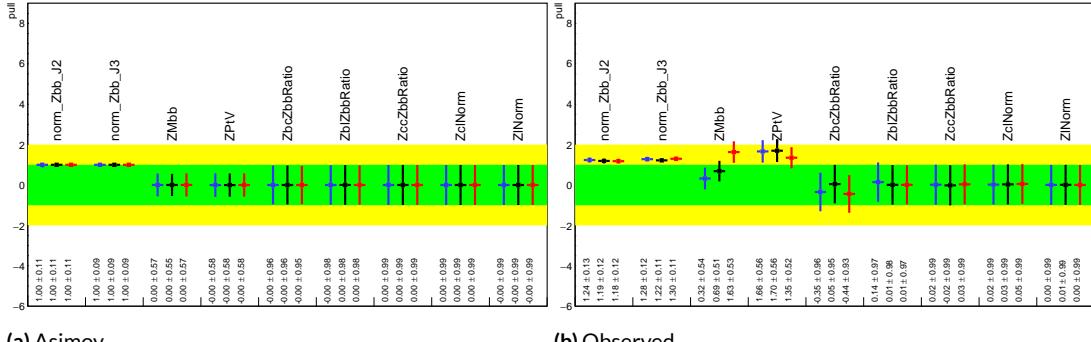
As can be seen in Figures 8.27–8.31, the fits for the three different variable sets are fairly similar from a NP pull perspective. Black is the standard variable set, red is the LI set, and blue is the RF set.



**Figure 8.29:** Pull comparison for MET NP's.



**Figure 8.30:** Pull comparison for Flavour Tagging NP's.



**Figure 8.31:** Pull comparison for  $Z$ +jets NP's.

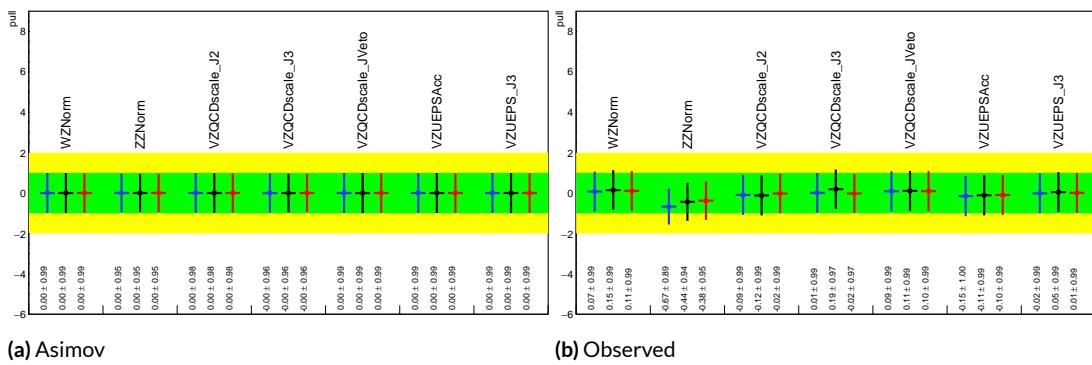


Figure 8.32: Pull comparison for signal process modeling NP's.

## 8.14 NUISANCE PARAMETER CORRELATIONS

Nuisance parameter correlation matrices (for correlations with magnitude at least 0.25) for all three variable set fits can be found in Figure ?? for Asimov fits and Figure ?? for observed fits.

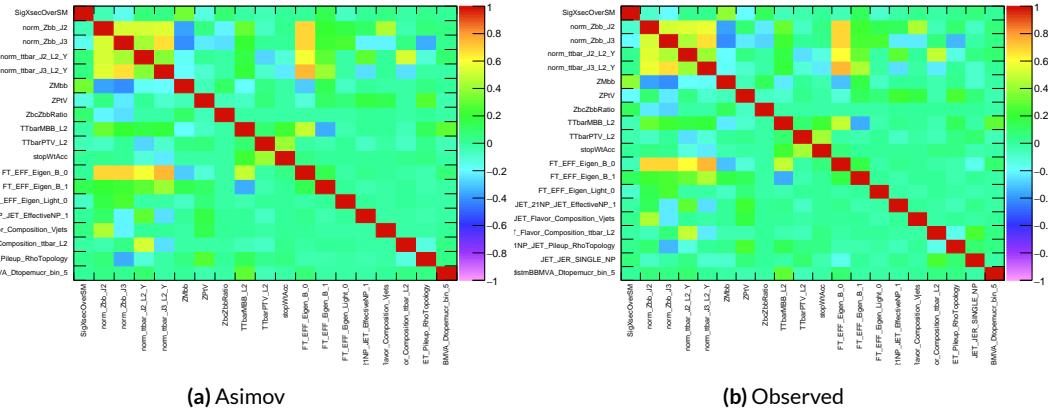


Figure 8.33: NP correlations for standard variable fits.

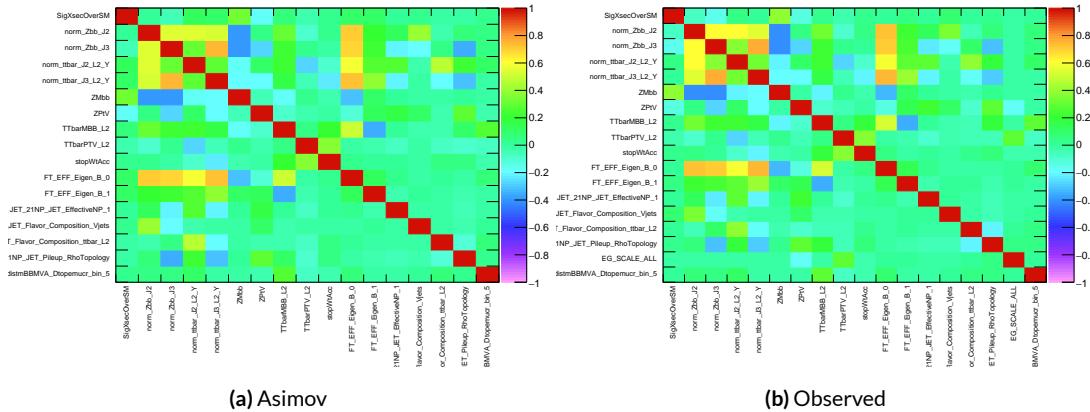
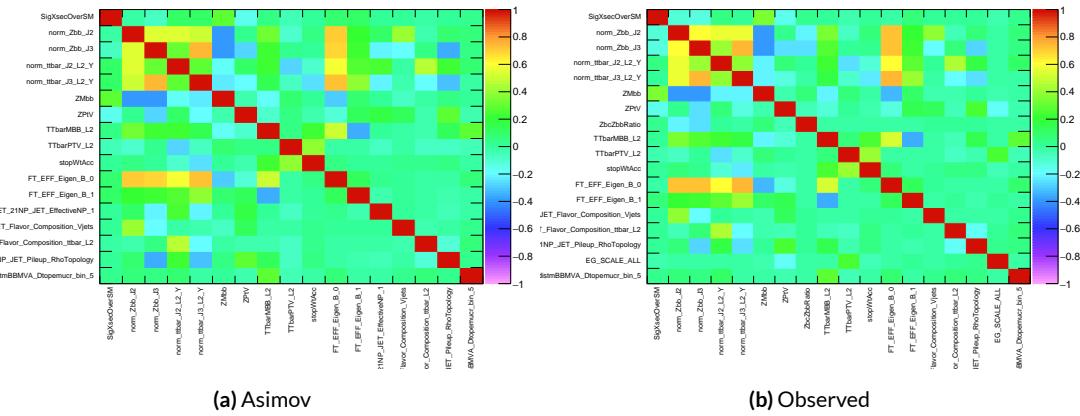


Figure 8.34: NP correlations for L1 variable fits.



**Figure 8.35:** NP correlations for RF variable fits.

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

Conventional Wisdom

# 9

## Fit Results

MUCH HAS BEEN SAID Expected and observed sensitivities for the different variable sets may be found in Table 9.1. The RF fits feature the highest expected sensitivities, outperforming the standard set by 3.5% and 3.4% for fits to Asimov and observed datasets, respectively. The LI variable has a lower significance than both for expected fits to both Asimov and data with a 6.7% (1.7%) signifi-

cance than the standard set for the Asimov (observed) dataset. While the fit using standard variables does have a higher observed significance than both the LI and RF fits, by 2.8% and 8.6%, respectively, these numbers should be viewed in the context of the best fit  $\hat{\mu}$  values, discussed below. That is, the standard set may yield the highest sensitivity for this particular dataset, but this is not necessarily (and likely is not) the case for 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 9.1:** Expected (for both data and Asimov) and observed significances for the standard, LI, and RF variable sets.

A summary of fitted signal strengths and errors for both the Asimov (a) and observed (b) datasets are shown in Figure 9.1.<sup>\*</sup> A summary of error breakdowns is given in Tables 9.2 (Asimov) and 9.3 (observed) for total error, data statistics contributions, total systematic error contributions, and categories for which the total impact is  $\geq 0.1$  for the standard fit. As is to be expected for both the Asimov and observed dataset fits, the contribution to the total error on  $\mu$  arising from data statistics is nearly identical, since each set of fits uses the same selections and data.<sup>†</sup>

The contribution from systematic uncertainties, however, does vary considerably across the variable sets. The Asimov fits are a best case scenario in the sense that, by construction, all NP's are equal to their predicted values (and so no “penalty” is paid for pulls on Gaussian NP's). The systematics

<sup>\*</sup>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>†</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.

	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

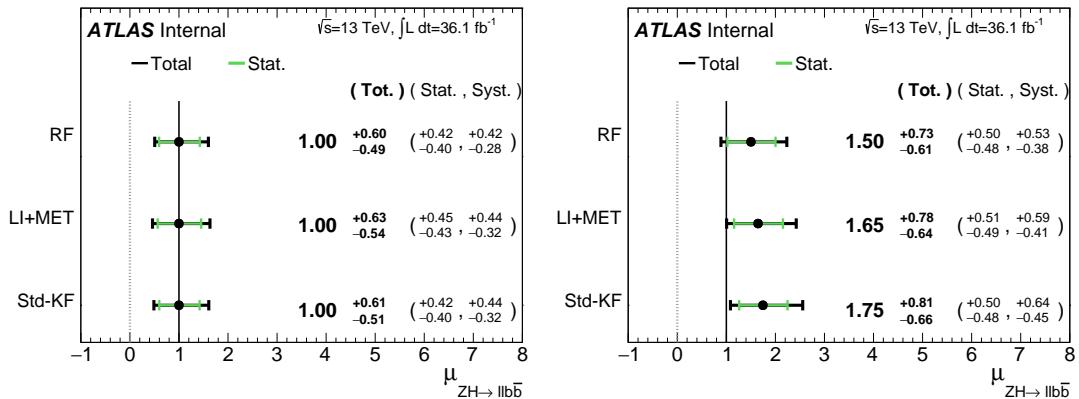
**Table 9.2:** Summary of error impacts on total  $\mu$  error for principal categories in the Asimov standard, LI, and RF fits.

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

error from the LI fit is slightly higher (subpercent) than that from the standard fit, and 4.6% higher error overall due to differences in data stats. The RF Asimov fit, however, has a 6.5% lower total error 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 9.3, 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 9.1:**  $\mu$  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).

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

Conventional Wisdom

# 10

## Conclusions

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

information in collision events.

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

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

Moreover, the LI and RF variable sets generally perform better in the context of the error on  $\mu$ . The LI fit is comparable to the standard set on Asimov data and has a 7.5% lower total systematics error on  $\hat{\mu}$  on observed data, while the RF fit is lower in both cases, with systematics error being 6.5% (16%) lower on Asimov (observed) data. A summary of performance metrics in this document may be found in Table 10.1.

These figures of merit suggest that both the LI and RF variables are more orthogonal than the standard variable set used in the fiducial analysis. Moreover, the RF variable set does seem to consistently perform better than the LI set. Furthermore, both variable sets have straightforward extensions to the one lepton channel in the  $VH(b\bar{b})$  analysis, and the RF set has a straightforward extension to the zero lepton channel as well. The magnitude of any gain from the more sophisticated

	Standard	LI	RF
$\hat{\mu}$	$1.75^{+0.24}_{-0.23}$ (stat.) $^{+0.34}_{-0.28}$ (syst.)	$1.65^{+0.24}_{-0.23}$ (stat.) $^{+0.34}_{-0.28}$ (syst.)	$1.50^{+0.24}_{-0.23}$ (stat.) $^{+0.34}_{-0.28}$ (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.

treatment of  $E_T^{miss}$  in these extensions is beyond the scope of these studies, but the performance in this closed final state do suggest that there is some value to be had in these non-standard descriptions independent of these considerations.

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

Conventional Wisdom

A

# Micromegas Trigger Processor Simulation

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

rithm for the MMTP, described in detail with initial studies in<sup>25</sup>. The goal of this note is to describe the MMTP algorithm performance under a variety of algorithm settings with both nominal and misaligned chamber positions, as well as addressing a number of performance issues.

This note is organized as follows: the algorithm and its outputs are briefly described in Section A.1; Monte Carlo samples used are in Section B.1; nominal algorithm performance and certain quantities of interest are described in Section A.3; algorithm performance under misalignment, misalignment corrections, and corrected performance are shown in Section A.8; and conclusions are presented in Section A.16.

## A.1 ALGORITHM OVERVIEW

The MMTP algorithm is shown schematically in Figure A.1, taken from<sup>25</sup>, where a more detailed description may be found. The algorithm begins by reading in hits, which are converted to slopes. These slopes are calculated under the assumption that the hit originates from the IP; slopes calculated under this assumption are denoted by a superscript  $g$  for global in order to distinguish them from local slopes calculated using only hits in the wedge. In the algorithm simulation, events are screened at truth level to make sure they pass certain requirements. The track's truth-level coordinates must place it with the wedge since some generated tracks do not reach the wedge. These hits are stored in a buffer two bunch crossings (BCs) in time deep that separates the wedge into so-called “slope-roads.” If any given slope-road has sufficient hits to pass what is known as a coincidence threshold, a fit proceeds. A coincidence threshold is a requirement for an event expressed as  $aX+bUV$ , which means that an slope-road must have at least  $a$  hits in horizontal (X) planes and at

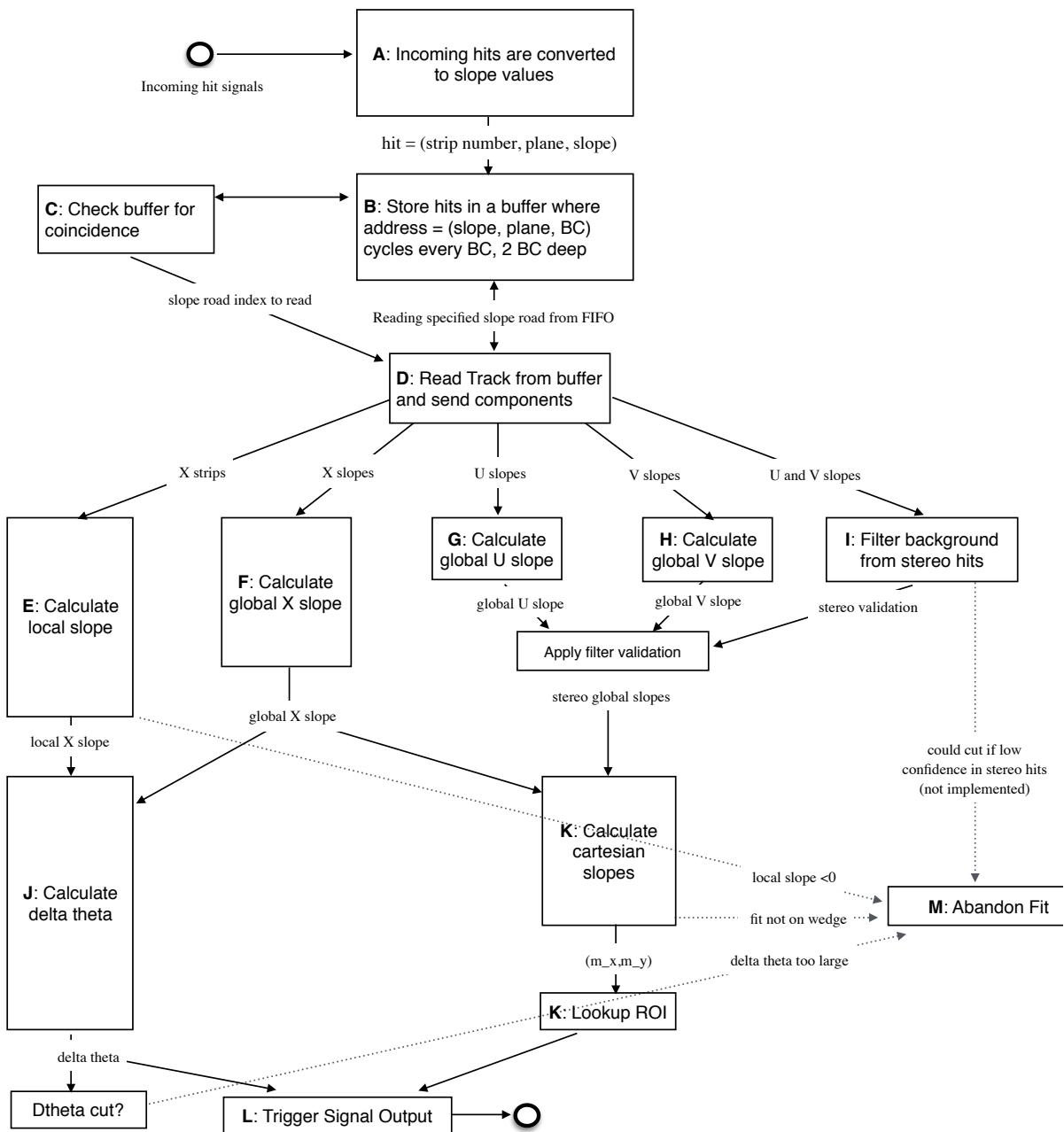


Figure A.1: A flow chart describing the algorithm steps, taken from<sup>25</sup>.

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

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

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

In the fit, individual hit slopes in a slope-road are used to calculate global slopes associated with each plane type, which are averages (e.g.  $M_X^{\ell}$  for the average slope of horizontal planes). These in turn are used to calculate the three composite slopes: slopes associated with the horizontal ( $m_x$ ) and vertical coordinates ( $m_y$ ) and the local slope of hits in the horizontal planes ( $M_X^l$ ), all of which are shown in Equation A.4. Note that the expression for  $M_X^l$  differs but is equivalent to the expression given in <sup>25</sup>. This is due to a procedural change in the algorithm. The local X slope is expressed in <sup>25</sup> as:

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

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

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

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

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

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

and  $z$  coordinates of hits in X planes.

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

From these composite slopes, the familiar expressions for the fit quantities  $\vartheta$  (the zenith),  $\phi$  (the azimuth<sup>†</sup>), and  $\Delta\vartheta$  (the difference in  $\vartheta$  between the direction of the segment extrapolated back to the interaction point and its direction when entering the detector region; the following is an approximation) may be calculated, as noted in<sup>25</sup>:

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

Looking at Equations A.4 and A.5, the dependence of fit quantities on input hit information becomes clear.  $\Delta\vartheta$  relies exclusively on information from the horizontal (X) planes. Both  $\vartheta$  and  $\phi$  rely on both horizontal and stereo slope information. However, the sum in quadrature of  $m_x$  and  $m_y$  in the arctangent for  $\vartheta$  means that  $\vartheta$  is less sensitive to errors in stereo hit information than  $\phi$ . Given that  $\vartheta_{st}$  is small,  $\cot \vartheta_{st}$  is large ( $\sim 38$ ), so  $m_x$  multiplies small differences in  $\mathcal{M}_U$  and  $\mathcal{M}_V$ , where  $m_y$  is simply an average over slopes. This means that while errors in horizontal hit information will affect all three fit quantities, comparable errors in stereo hits will have a proportionately larger effect on  $\vartheta$  and particularly on  $\phi$ . The  $\Delta\vartheta$  cut after step J in Figure A.1 has been implemented, requiring all fits to have  $|\Delta\vartheta| < 16$  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  $\phi = 0$

algorithm efficiency.

### A.2 MONTE CARLO SAMPLES

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

### A.3 NOMINAL PERFORMANCE

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

of the following studies.

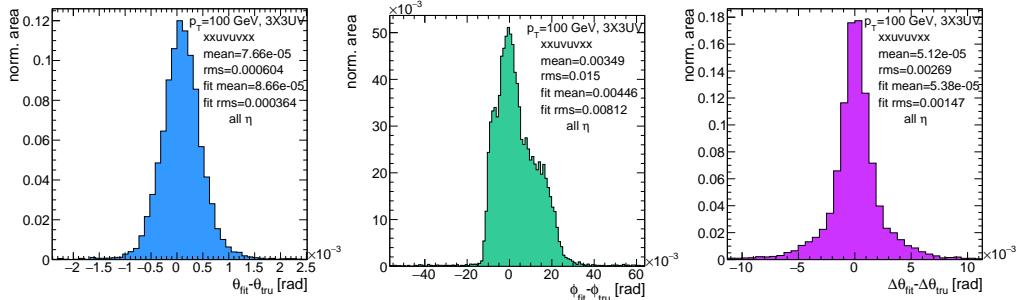
#### A.4 FIT QUANTITIES

In order to evaluate the performance of the algorithm’s fit quantities  $\vartheta$ ,  $\phi$ , and  $\Delta\vartheta$ , fit values are compared to truth-level MC values. The residual of the three fit quantities,  $\vartheta_{fit} - \vartheta_{tru}$ ,  $\phi_{fit} - \phi_{tru}$ , and  $\Delta\vartheta_{fit} - \Delta\vartheta_{tru}$ , are recorded for every fitted track. The distributions of these quantities, in particular their biases and standard deviations, are then used to evaluate performance. In most cases, following<sup>25</sup>, the mean and standard deviation of a  $3\sigma$  Gaussian fit are quoted, as they capture the main features of the algorithm and generally behave like the raw mean and rms. Nevertheless, discussion of the raw quantities will be included when their behavior deviates markedly from that of the  $3\sigma$  fit quantities.

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

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

$\eta$ -stations, as explained in Appendix B of<sup>24</sup>. The fitted (raw) rms for the  $\varphi$  distribution is 8.67 (16.6) mrad.

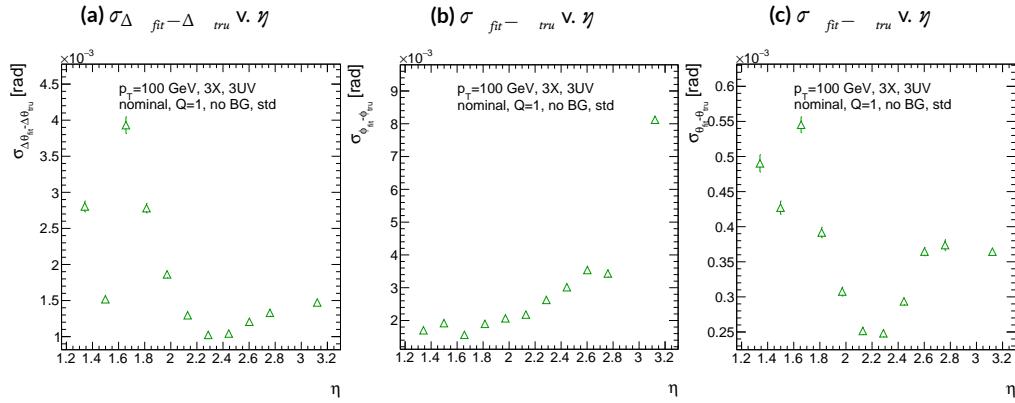


**Figure A.2:** Nominal residual plots;  $\vartheta$ ,  $\varphi$ ,  $\Delta\vartheta$  for  $p_T = 100 \text{ GeV}$  muons

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

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

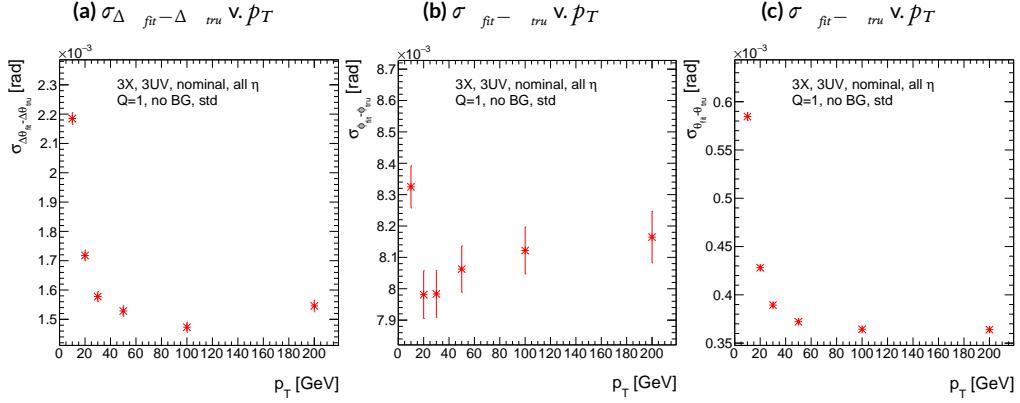
The interplay of these effects on the residual standard deviations can be seen in their dependences on  $\eta$  (Figure A.3; note that the final point in each of these plots is the rms of the distribution overall  $\eta$ ) and  $p_T$  (Figure A.4). For  $p_T = 100$  GeV muons,  $\Delta\vartheta$  performance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



**Figure A.3:** The rms distributions of  $\Delta\vartheta$ ,  $\varphi$ , and  $\vartheta$  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.

---

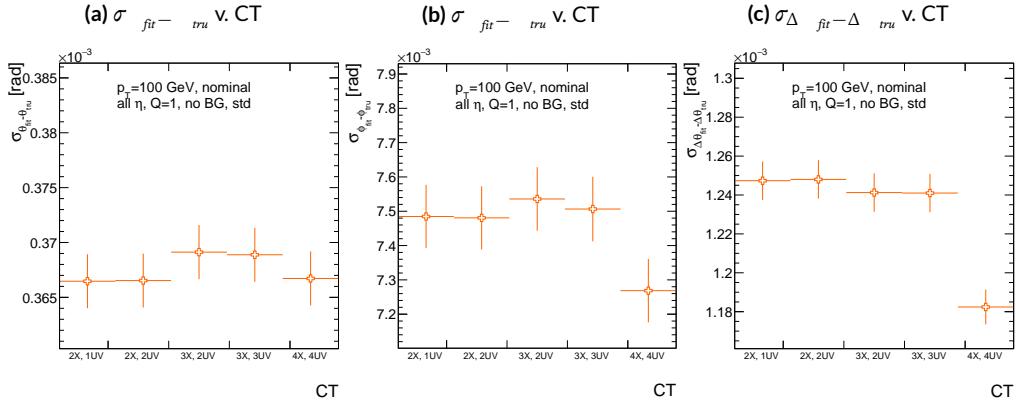
<sup>§</sup>The much worse overall performance for  $\varphi$  is due to the  $\eta$  dependent bias and other effects



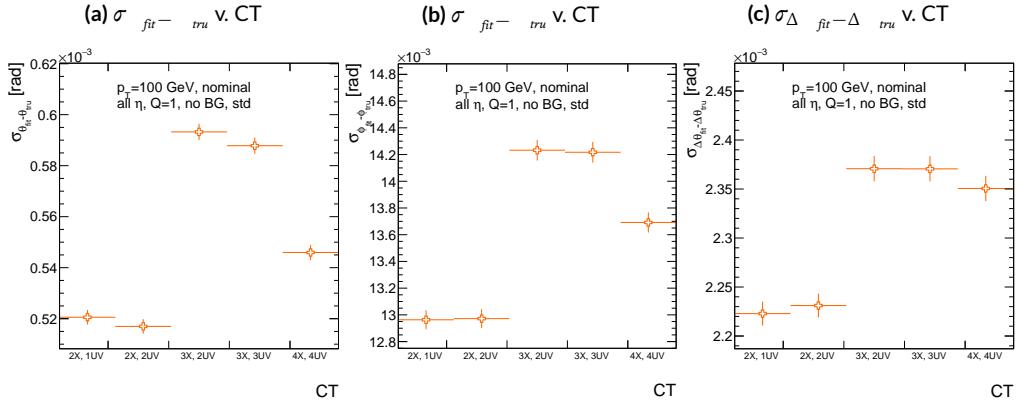
**Figure A.4:** The rms distributions of  $\Delta\emptyset$ ,  $\phi$ , and  $\emptyset$  as a function of  $p_T$ .

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

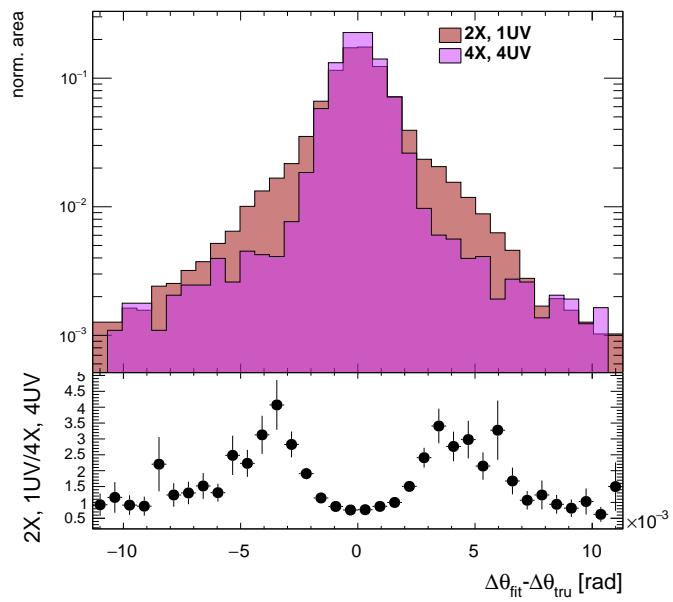
tribution in Figure A.7. As both the overlayed normalized curves and ratio distribution show, while the most central regions are fairly similar, the  $2X+1UV$  distribution is much more prominent in the tails but not the extreme tails, which means that, though the overall  $2X+1UV$  raw rms goes down, the overall quality of algorithm fits is worse.



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



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

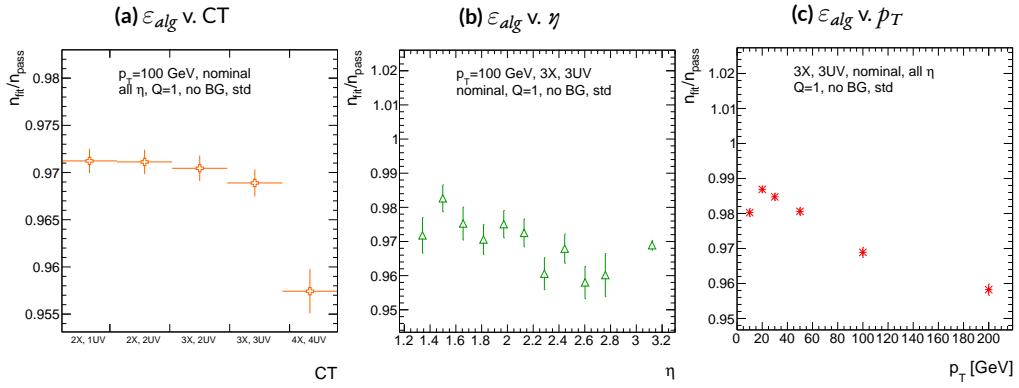


**Figure A.7:** Nominal  $\Delta\theta$  residual distribution for  $p_T = 100$  GeV muons with coincidence thresholds 2X+1UV and 4X+4UV normalized to the same area and plotted together (top) as well as the ratio of the 2X+1UV distribution and the 4X+4UV per bin.

## A.5 EFFICIENCIES

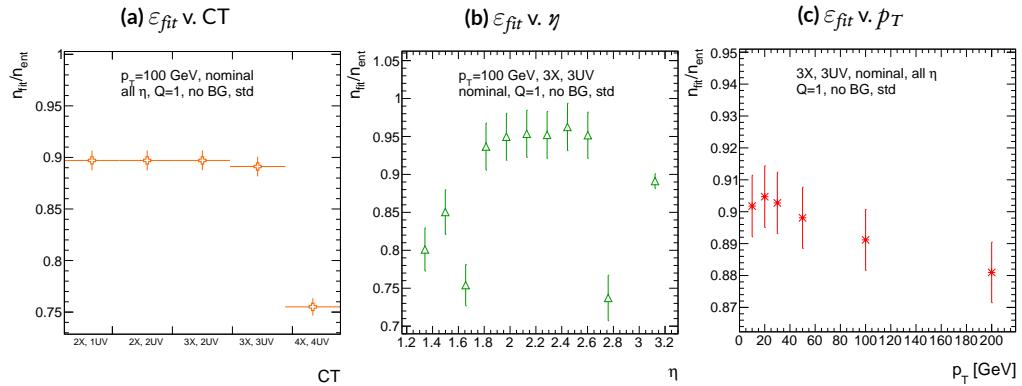
Two general efficiencies have been formulated to study the performance of the MMTP algorithm.

The first, denoted  $\varepsilon_{alg}$ , is the fraction of tracks that pass some (slope-road) coincidence threshold configuration that are successfully fit. An event that passes a slope-road coincidence but does not fit fails because some of the hits included are of sufficiently poor quality to throw off the fit. This efficiency answers the question of how often the algorithm performs fits when technically possible, giving a measure of overall algorithm performance for a given configuration. For example,  $\varepsilon = 95\%$  for  $3X+2UV$  means that 95% of tracks that produce at least  $3X$  hits and  $2UV$  hits in at least one slope-road will be successfully fitted 95% of the time. The performance of this efficiency as a function of coincidence threshold,  $\gamma$  (with the final point once again being the efficiency integrated over all  $\eta$ ), and  $p_T$  is shown in Figure A.8.  $\varepsilon_{alg}$  is fairly constant in  $\gamma$  and decreases with increased  $p_T$ , which can be attributed to the increased likelihood of secondaries introducing lower quality hits that cause the fit to fail.



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

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

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

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

## A.6 INCOHERENT BACKGROUND

The default slope-road size and tolerances associated with horizontal and stereo hits used in the above studies were configured to optimize algorithm performance, similar to studies in <sup>25</sup>. In order to evaluate algorithm performance under conditions with more limited resources, as might be expected at run-time, additional studies were conducted with the slope-road size and hit tolerances set equivalent to the sensitive area of a single VMM chip<sup>¶</sup> both with and without generation of incoherent background.

Incoherent background is generated based on the assumption that the intensity only varies as a function of the distance from a point to the beamline,  $r$ . The number of hits per unit area per unit time as a function of  $r$  is given in Equation A.6 and taken from <sup>25</sup>.

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

where  $r_0 = 1000$  mm and  $I_0 = 0.141$  kHz/mm<sup>2</sup>

Background generation happens per event as follows:

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

---

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

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

6. Calculate hit information according to these values.

The expectation value for the Poisson distribution is determined by integrating Equation A.6 over the surface area of the wedge to get the total hit rate for the wedge,  $\Gamma$ , and then multiplying this by the length of the time window over which hits may be generated. With  $H = 982$  mm,  $b_i = 3665$  mm, and  $\vartheta_w = 33\pi/180$ , we find<sup>||</sup>:

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

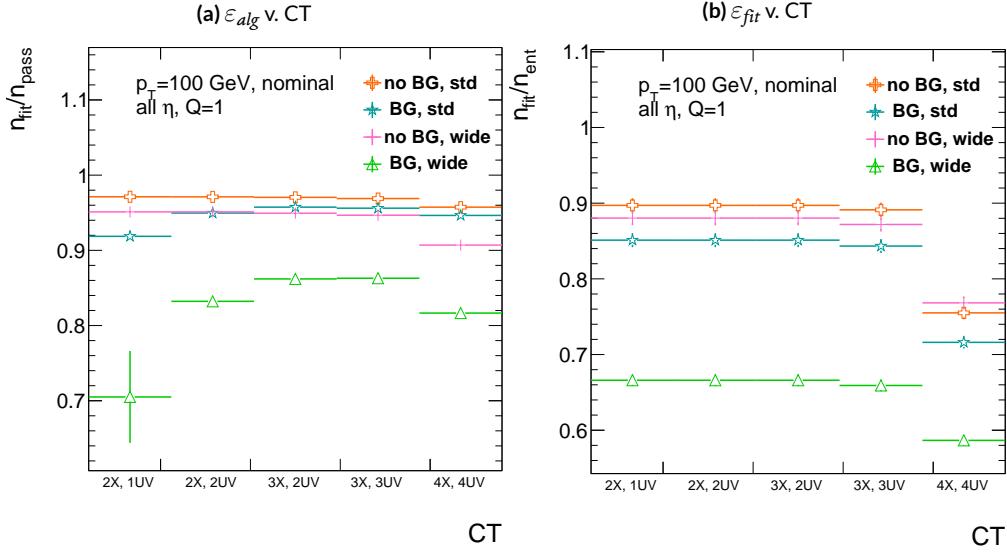
In this case, we have taken the nominal values of the MM sector geometry for  $H$  (wedge base),  $b_i$  (the wedge height), and  $\vartheta_w$  (the wedge opening angle).

The effects of incoherent background and larger slope road size are summarized in Figure A.11 for efficiencies and in Figure A.13 and Table A.1 for residual of fit quantities.

Figure A.11 show the effect of both wider slope-roads and the introduction of background on efficiencies. The introduction of wider slope-roads increases the chance that an early errant hit (either from secondaries/ionization or background) will be introduced into the fit, and the presence of incoherent background greatly increases the number of such errant hits. Both wider slope-roads and background drive down the number of fits (numerator) in both efficiencies, and background can artificially inflate the denominator of  $\varepsilon_{alg}$ , a reco-level, slope-road coincidence threshold. The shape of the  $\varepsilon_{fit}$  versus coincidence threshold distributions remains fairly constant with each complicating

---

<sup>||</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)).

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



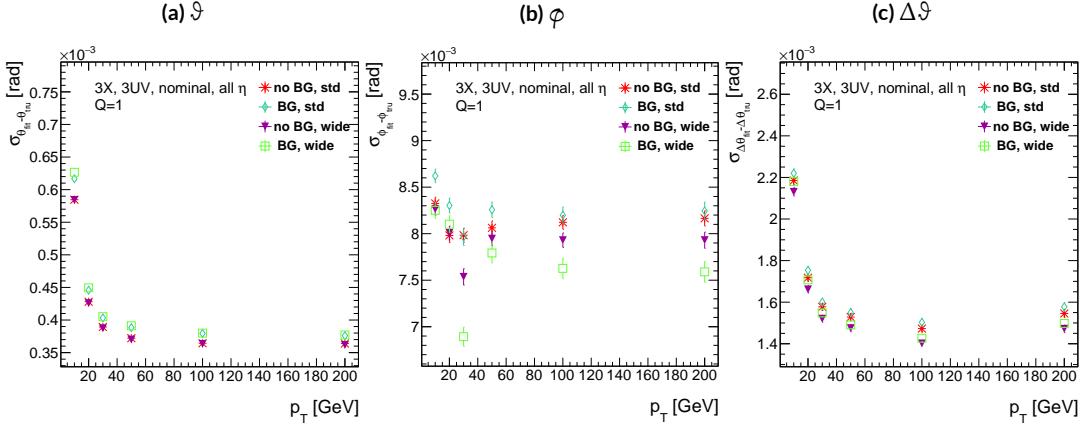
**Figure A.12:** The distribution of highest slope-road coincidence thresholds in events for the algorithm with wide slope-roads (width of 1 VMM) both without (a) and with (b) incoherent background; the 0,0 bin is the number of events passing selection requirements that fail to form the minimum  $2X+1UV$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

The effect of increasing slope-road size and incoherent background on fit quantity residual rms values as a function of  $p_T$  is shown in Figure A.13. As the figure shows, the fitted rms values are fairly robust against increased slope-road size and background. This does not hold for all of the raw rms values, however, as shown in Table A.1. Just as with the efficiencies, the introduction of background has a larger effect than that of increased slope-road size, which does not seem to have an overly large impact on any of the fit quantities on its own. While  $\Delta\vartheta$  remains robust to both increased slope-road size and background (likely due to the  $\Delta\vartheta$  cut of 16 mrad built into the algorithm),  $\vartheta$  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.

	No BG, std	No BG, wide	BG, std	BG, wide
$\vartheta$	0.364 (0.604)	0.363 (0.542)	0.379 (0.886)	0.380 (1.07)
$\varphi$	8.12 (15.0)	7.93 (13.2)	8.20 (24.6)	7.63 (24.8)
$\Delta\vartheta$	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.

As Table A.1 shows, rms values appear to be robust to an increase in slope-road size. Nevertheless, though the fitted  $\sigma$  residual values are also fairly robust to the introduction of background, the



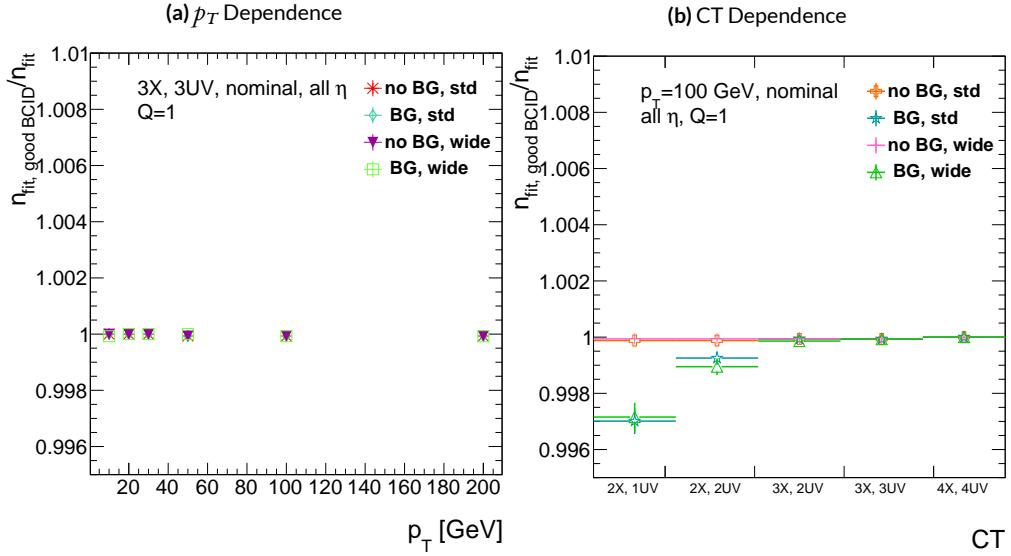
**Figure A.13:** The three fit quantity residual rms values as a function of  $p_T$  for different background settings and slope-road sizes (standard and wide (one slope road as 1 VMM chip)).

raw rms values are not. While the raw  $\Delta\vartheta$  rms stays stable, both  $\vartheta$  and  $\varphi$  suffer noticeable degradation, which suggests that the introduction of background has a detrimental effect on horizontal slope residual (i.e. on stereo strips in particular). This level of degradation is likely acceptable for  $\vartheta$ , though further steps may need to be taken to address  $\varphi$ .

### A.6.1 BCID

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

simulation.



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

## A.7 CHARGE THRESHOLD

The MMTP uses the first hits registered passing a charge threshold requirement given in units of electron charge. In principle, it would be beneficial to be able to use any hits that are registered regardless of deposited charge, but in the high rate environment envisioned for the NSW, this requirement might need to be raised. Nominal algorithm settings have this charge threshold requirement set to 1, and studies were conducted on algorithm performance for charge threshold values of 0, 1, and 2. Efficiencies as a function of coincidence threshold for different charge thresholds are shown in Figure A.15. Increasing the charge threshold lowers both efficiencies, particularly at high coincidence threshold, which suggests that energetic muons with secondaries create both very many hits

and hits with higher charge. While the shapes of the fit quantity distributions as a function of  $p_T$  in Figure A.16 are fairly constant across charge threshold, performance is not.  $\vartheta$  and  $\Delta\vartheta$  show some improvement with higher charge threshold, particularly at low  $p_T$ , suggesting that resolution improves in the vertical direction, but  $\phi$  shows degradation at higher charge threshold, which is a symptom of more highly charged particles experiencing greater bending in the ATLAS magnetic field in the  $\phi$  direction.

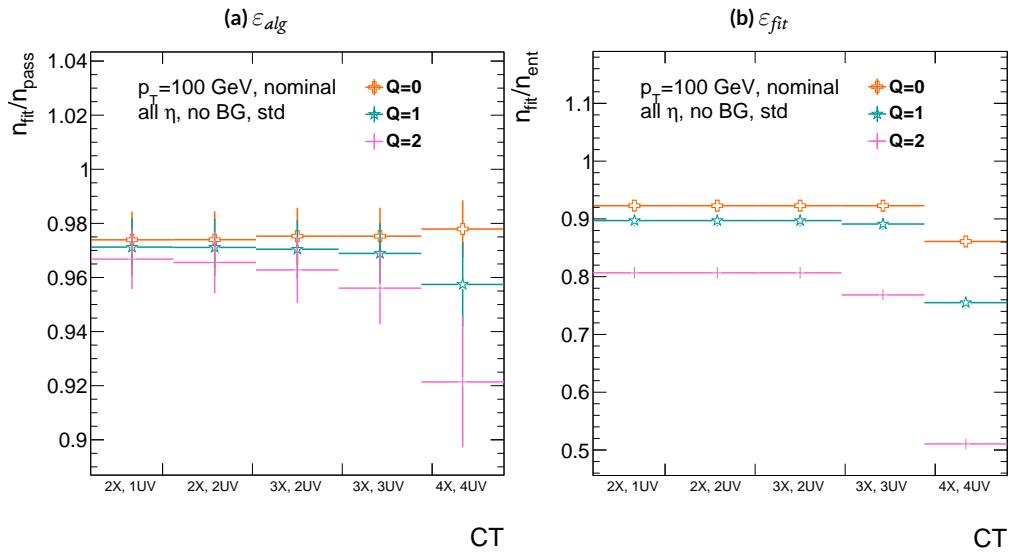


Figure A.15: The efficiencies as a function of coincidence threshold for charge thresholds 0, 1, and 2.

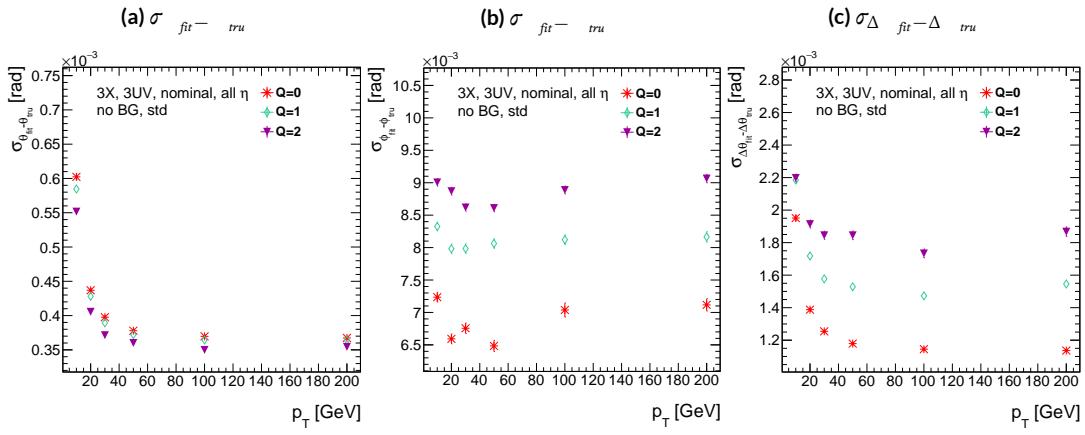


Figure A.16: The fit quantity residual rms values as a function of  $p_T$  for charge thresholds of 0, 1, and 2.

## A.8 MISALIGNMENTS AND CORRECTIONS

The performance of the trigger algorithm under misalignment has been studied for each of the six alignment quantities (three translations and three rotations all along the principal axes) described in<sup>3</sup> and<sup>2</sup>, whose convention we will follow here. For the simulated wedge studied here the local coordinates described in<sup>2</sup> are taken to be centered at the center of the base of the wedge<sup>\*\*</sup>, the local  $t$  axis corresponds to the axis of the beam line, the local  $z$  axis corresponds to the direction orthogonal to both the beam line and the horizontal strips, and the local  $s$  axis completes the right-handed coordinate system. The rotation angles  $\alpha$ ,  $\beta$ , and  $\gamma$  correspond to rotations around the local  $t$ ,  $z$ , and  $s$  axes, respectively. Note that the local  $s$ ,  $z$ , and  $-t$ , axes correspond to the usual global  $x$ ,  $y$ , and  $z$  axes. Misalignments were studied in twenty evenly spaced increments from nominal positions to misalignments of 1.5 mrad for the rotations (-1.5 mrad to +1.5 mrad for the  $\gamma$  case), and of 5 mm (a roughly corresponding linear shift) for the translations. In all cases, the front quadruplet is misaligned while the rear quadruplet remains in its nominal position. While only the front quadruplet of a single wedge is misaligned, the framework for misalignment presented below could be used to study generic local and global misalignments. The six misalignments are schematically represented in Figure A.17.

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

---

<sup>\*\*</sup>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.

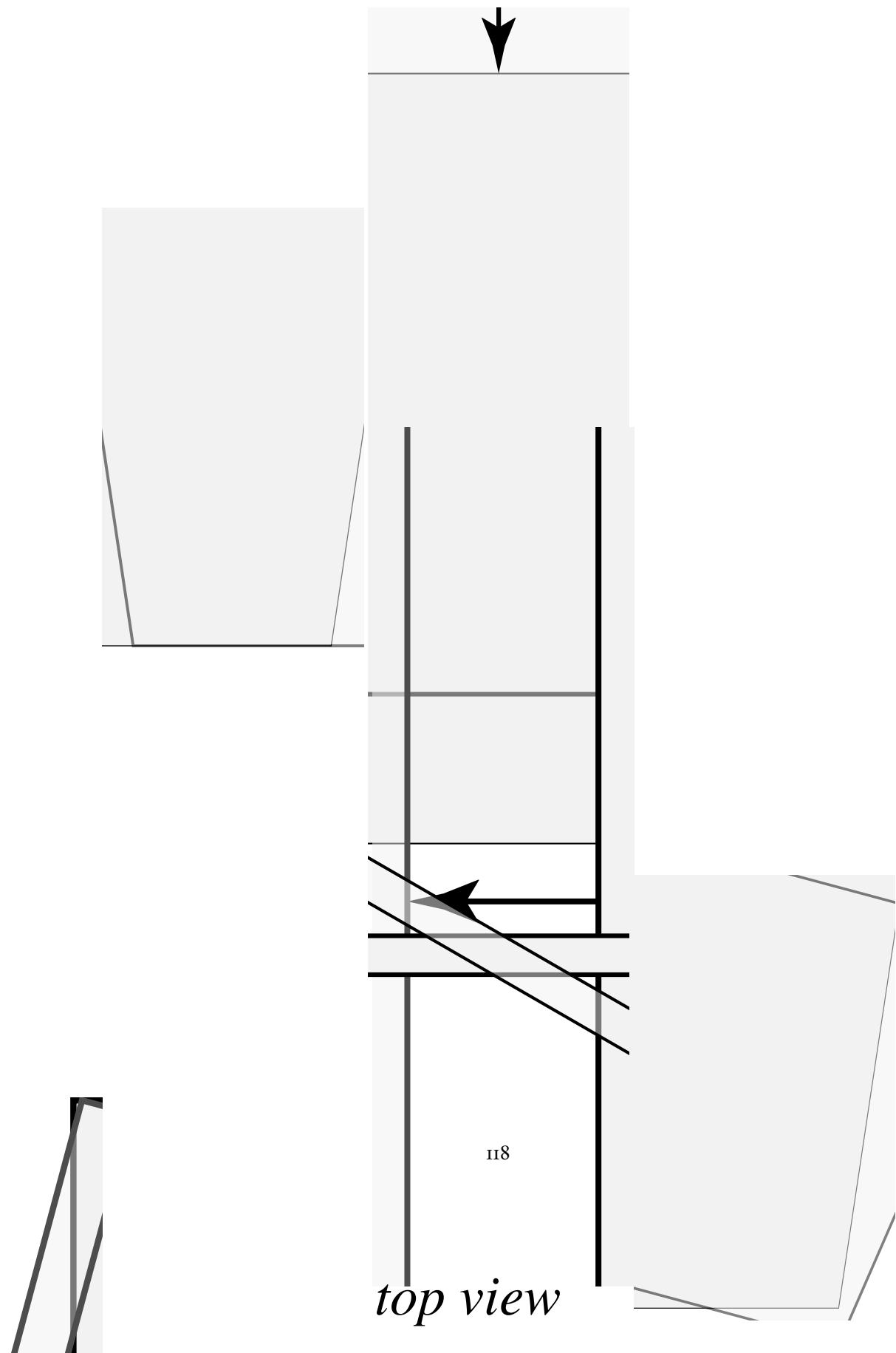


Figure A.17: The different misalignment cases as defined in the AMDR manual.

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

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

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

Symbol	Definition
$s_x, s_y, s_z, \vec{s}$	Position of the muon hit in ATLAS global coordinates; the infinite momentum muon track
$\hat{n}$	Vector normal to the plane; taken to be $\hat{z}$ (the beamline) in the nominal case
$\vec{\mathcal{O}}_{IP}^{g,l}$	Position of the interaction point in ATLAS global ( $g$ ) or wedge local ( $l$ ) coordinates
$\vec{\mathcal{O}}_{base}^{g,l}$	Position of the plane base in ATLAS global ( $g$ ) or wedge local ( $l$ ) coordinates; $(0, y_{base}, z_{pl})$ ( $(0, 0, 0)$ ) for the nominal case in global (local) coordinates
$\vec{\xi}$	$\vec{s} - \vec{\mathcal{O}}_{base}$
primed quant.	quantities after misalignment

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

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

difference is the choice of origin.

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

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

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

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

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

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

Rotations are done before translations, according to the order prescribed in the AMDB guide for chamber alignment, so the axes the principal axes of the plane are rotated according to the following matrix (where  $s$ ,  $c$ , and  $t$  are the obvious trigonometric substitutions)

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

---

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

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

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

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

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

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

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

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

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

The correction will be plane dependent since (denoting the stereo angle  $\omega$ ):

$$\hat{y}_x = \hat{y} \rightarrow \hat{y}'_x = -\sin\beta\hat{x} + (\cos\gamma - \sin\beta\sin\gamma)\hat{y} + (\cos\gamma + \sin\beta\cos\gamma)\hat{z}$$

(A.17)

and

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

### A.8.1 INDIVIDUAL CASES

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

zero.<sup>††</sup>

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

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

If the translations are zero,

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

### A.8.2 $ds \neq 0$

$k'/k = 1$  (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  $\vec{\theta}'_{base} = ds\hat{x}$  is:

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

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

### A.8.3 $dz \neq 0$

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

### A.8.4 $dt \neq 0$

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

### A.8.5 $\alpha \neq 0$

$k'/k = 1$  and

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

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

### A.8.6 $\beta \neq 0$

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

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

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

### A.8.7 $\gamma \neq 0$

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

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

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

In order to evaluate algorithm performance under misalignment and corrections for misalignment, the absolute means and relative resolutions of the fit quantities  $\vartheta$ ,  $\phi$ , and  $\Delta\vartheta$  are measured as a function of misalignment. In the following, results will only be shown for which the effects of misalignment are significant. “Significant,” for misalignments of 1 mm (0.3 mrad) for translations (rotations) means more than a 5% degradation in rms and/or bias shifts in  $\vartheta$ ,  $\phi$ , and  $\Delta\vartheta$  of 0.01 mrad, 1 mrad, and 0.1 mrad, respectively.

While corrections are typically done on a case-by-base basis, they fall under two general categories, analytic and simulation based. Analytic corrections rely upon specific knowledge of the misalignment, with each case being handled separately; as such, the additional resources required, both extra constants and operations, if any, vary accordingly. Simulation based corrections are all done in the same manner. The algorithm is run over a training MC sample (same setup but with  $p_T = 200$  GeV instead of the normal 100 GeV sample so as not to overtrain the corrections), and the mean biases for  $\vartheta$ ,  $\phi$ , and  $\Delta\vartheta$  are saved for different, equally spaced regions in the  $\eta - \varphi$  plane over the wedge based on the fitted  $\vartheta$  and  $\phi$  values. Currently, these values are saved for 10  $\eta$  and 10  $\varphi$  bins (100  $\eta, \varphi$

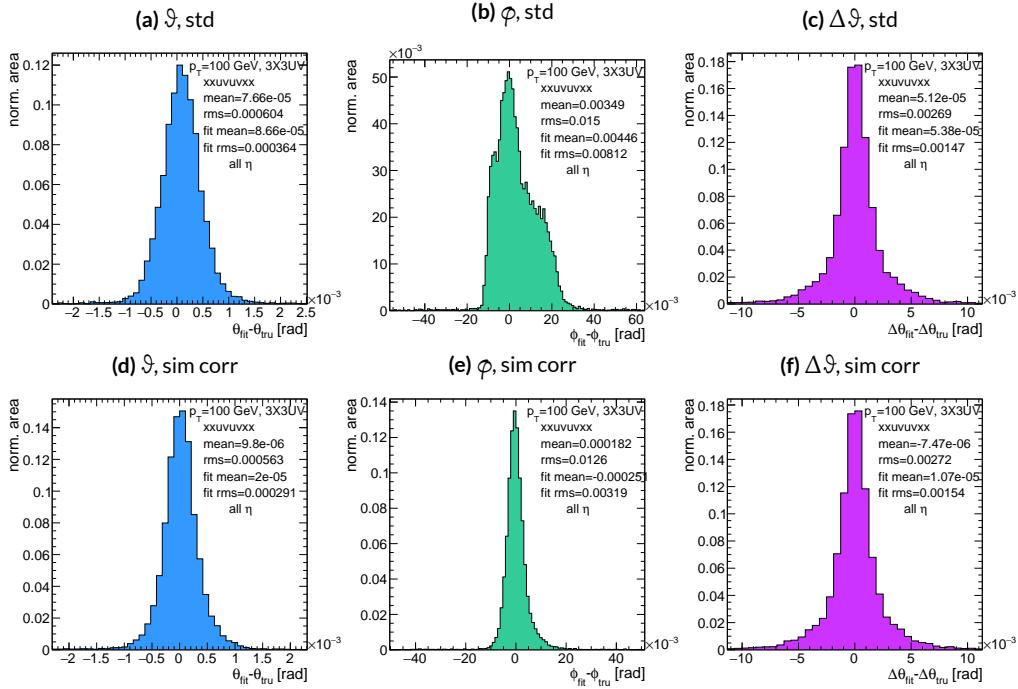
**Table A.3:** A summary of corrections with additional constants/operations (written as  $n_{const}c/n_{ops}\text{op}$ ;  $n_X$  is the number of X hits in a fit) necessary for analytic corrections. Yes means a correction exists but might not entirely remove misalignment effects, while yes+ means a quality of correction is only limited by knowledge of misalignment and memory

	$\Delta s$	$\Delta z$	$\Delta t$	$\gamma_s$	$\beta_z$	$\alpha_t$
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+

bins total), with the number of bins in each direction being a configurable parameter. When the algorithm runs with simulation based correction, this table of constant corrections is saved in a LUT before runtime, and corrections are added to final fit quantities based on the (uncorrected)  $\vartheta$  and  $\phi$  fit values. With the settings mentioned, this is 300 extra constants ( $10\vartheta\text{-bins} \times 10\phi\text{-bins} \times 3$  fit quantities) and two extra operations (a lookup and addition for each quantity done in parallel). The simulation correction can, in principle, also be applied to the algorithm in nominal conditions with non-trivial improvements, as detailed below in Section A.9. Depending on the misalignment case in question, different approaches work better. A summary of correction methods, including resources necessary for the individual analytic cases, is shown in Table A.3.

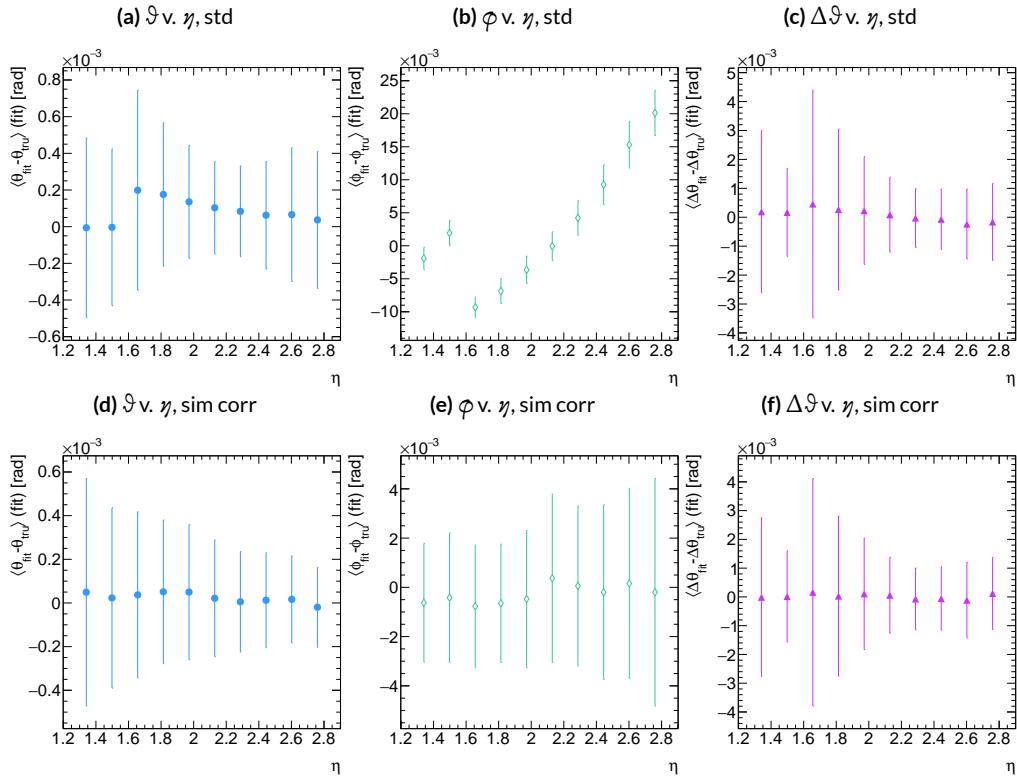
## A.9 SIMULATION CORRECTION OF THE ALGORITHM UNDER NOMINAL CONDITIONS

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



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

As can be seen in Figure A.19, the simulation based correction also removes the  $\gamma$  dependence to fit quantity resolution distributions, as expected. One consequence of this is that simulation-based corrections applied to the misalignment cases below will restore performance to the “sim” and not the “std” distributions of Figure A.18. Hence, when making comparisons between simulation corrected 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  $\gamma$  with points as means and error bars as rms values in each  $\gamma$  bin for the angles  $\vartheta$ ,  $\phi$ ,  $\Delta\vartheta$  for  $p_T = 100$  GeV muons in the uncorrected and simulation corrected cases.

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

rithm in nominal conditions most for the quantities that depend most on stereo information ( $\phi$  and  $\vartheta$ ) and remove the  $\gamma$  dependence of fit quantity resolutions suggests that there could, in principle, be analytic corrections that could be applied to the nominal algorithm. One possible solution is to introduce an additional set of constants, having the  $y_{base}$  depend on the strip number, similar to the  $\gamma_s$  correction for  $z_{plane}$  described in Section A.13, which would add a lookup per hit and  $8 \times n_{bins,y}$  extra constants that would be optimized as the  $\gamma_s$  correction was.

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

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

## A.10 TRANSLATION MISALIGNMENTS ALONG THE HORIZONTAL STRIP DIRECTION ( $\Delta s$ )

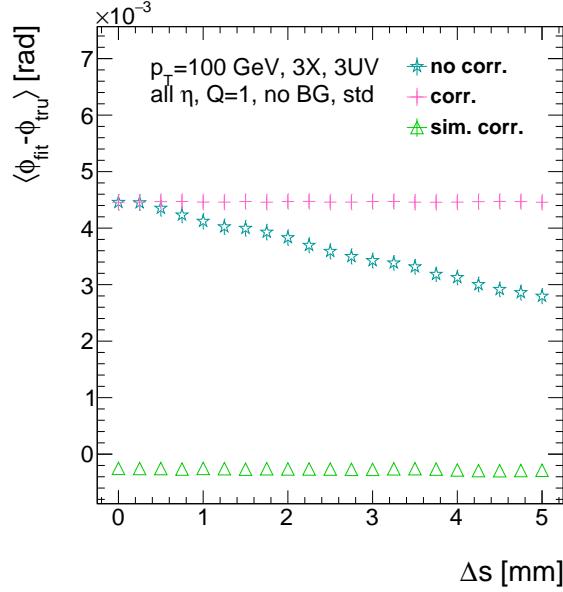
A translation in  $s$  (i.e. along the direction of a horizontal strip) only affects the stereo strips, and, since the stereo angle is small, a very large misalignment is necessary for effects to be noticeable (a misalignment of roughly 17 mm corresponds to one strip's misalignment in the stereo planes).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

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

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

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

discrepancy between the simulated and analytic correction is a natural consequence of the fact that the simulation correction, as previously mentioned, restores the  $\phi$  residual distribution to an overall 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.

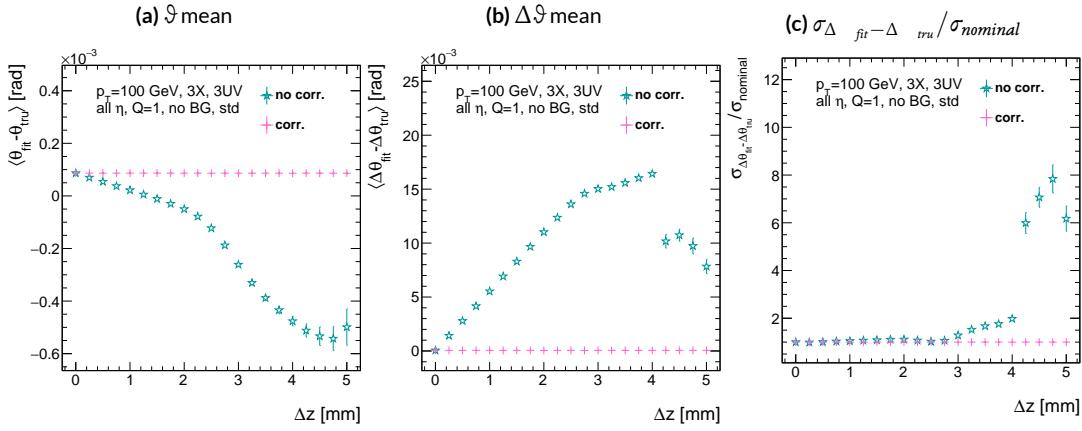
## A.II TRANSLATION MISALIGNMENTS ORTHOGONAL TO THE BEAMLINE AND HORIZONTAL STRIP DIRECTION ( $\Delta z$ )

A translation in AMDB  $z$ , the direction orthogonal to both the beamline and the horizontal strip direction, corresponds to a translation in the  $y$  of Equation A.1, affecting all slope calculations. This has a large impact on the  $\vartheta$  residual bias and both the bias and rms of  $\Delta\vartheta$  residual, as can be seen in Figures A.21 (a)–(c). The marked degradation and non-linear behavior in performance at very high levels of misalignments is a result of low statistics; there are fewer fits at high level of misalignments since for  $\Delta z \gtrsim 3$  mm, most fits will fail the  $\Delta\vartheta$  cut. The  $\vartheta$  bias shifts by about 0.075 mrad at  $\Delta z = 1$  mm, and  $\Delta\vartheta$  shifts by about 5 mrad for the same level of misalignment. While the fitted rms of the  $\Delta\vartheta$  residual remains fairly stable for  $\Delta z < 1$  mm or so, between  $\Delta z = 2$  mm and  $\Delta z = 3$  mm, the rms increases by 15% before the  $\Delta\vartheta$  cut issue mentioned above intervenes.

Fortunately, these misalignments are straightforward to correct with knowledge of the misalignment. The only modification necessary for this correction is to change the definitions of  $y_{base}$  in Equation A.1 for the individual hit slope addressing. This is done before runtime and adds no overhead to the algorithm, and the correction quality is only limited by knowledge of the misalignment. The results of this correction are also shown in Figures A.21 (a)–(c) and restore nominal performance.

---

Since  $\Delta\vartheta = \frac{M'_X - M''_X}{1 + M'_X M''_X}$  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 ( $\Delta\vartheta$  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  $\sigma_{\Delta_{\text{fit}} - \Delta_{\text{true}}} / \sigma_{\text{nominal}}$  for both the misaligned and corrected cases.

## A.12 TRANSLATION MISALIGNMENTS PARALLEL TO THE BEAMLINE ( $\Delta t$ )

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

Figure A.22.

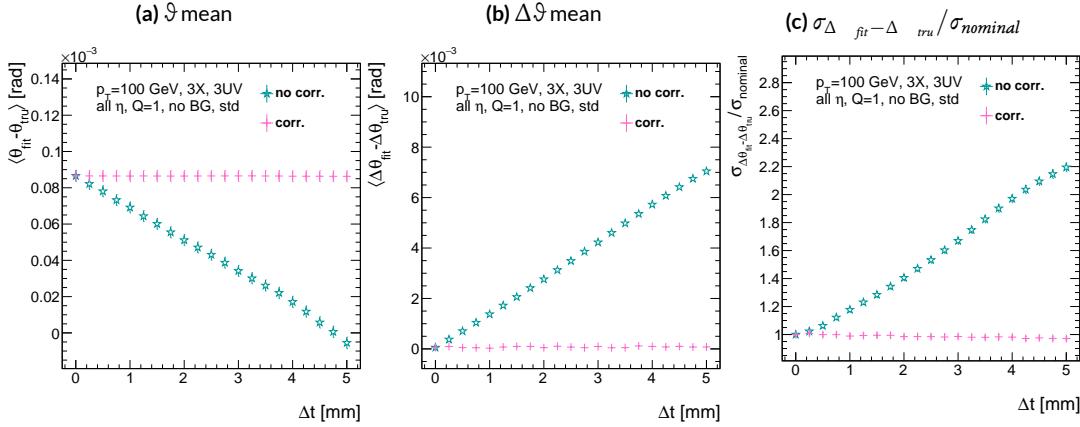


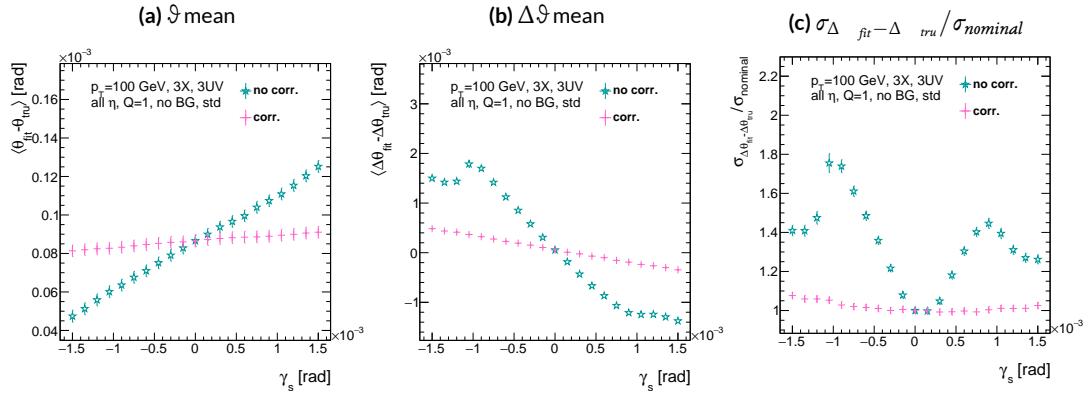
Figure A.22: The affected quantities of  $\Delta t$  misalignments:  $\vartheta$  bias,  $\Delta\vartheta$  bias, and  $\sigma_{\Delta_{fit} - \Delta_{true}} / \sigma_{nominal}$  for both the misaligned and corrected cases.

### A.13 CHAMBER TILTS TOWARDS AND AWAY FROM THE IP ( $\gamma_s$ ROTATION)

Chamber misalignment due to rotations around the  $s$  axis act effectively like a translation in  $t$  that depends on strip number. These rotations tilt misaligned chambers away from (towards) the IP for positive (negative) values of  $\gamma_s$ . Since, unlike for the other two rotation cases that will be studied, positive and negative rotation values are not symmetric, this misalignment is studied for both positive and negative  $\gamma_s$  values. The divergent effect at the tails is a result of a large population of fits not having fit quantities within the cores, and so not appearing in the fit rms. Once again, affected quantities of interest  $\vartheta$  bias,  $\Delta\vartheta$  bias, and  $\sigma_{\Delta\vartheta_{fit} - \Delta\vartheta_{true}}$ . The effects of misalignment can be seen in Figures A.23 (a)–(c). The relationship between biases and  $\gamma_s$  is roughly linear with  $\Delta\gamma_s = 0.3$  mrad (the angular scale corresponding to linear shifts of  $\sim 1$  mm) corresponding to  $0.005$  mrad ( $0.12$  mrad) for  $\vartheta(\Delta\vartheta)$ . For  $\sigma_{\Delta\vartheta_{fit} - \Delta\vartheta_{true}}$ , degradation is not symmetric. For negative (positive)  $\gamma_s$ , with the quadruplet tilted towards (away from) the IP, slope-roads are artificially expanded (shrunk), decreasing (increasing) the granularity of the trigger, explaining the asymmetry in Figure A.23 (c), with the degradation being a  $10\%$  ( $25\%$ ) effect for  $\gamma_s$  of  $+(-)0.3$  mrad.

Corrections are less simple in this case. In principle, corrections of the same accuracy of the translations could be calculated per strip, but the overhead of one correction per strip (many thousands of constants) is prohibitive. Instead, each plane was divided into eight equal segments with a  $t$  value ( $z$  in the slope calculation) assigned to strips in each region to correct for the misalignment. This amounts to 56 extra constants and a 2D instead of a 1D LUT for  $z$  positions while the algorithm runs. The corrected distributions can also be seen in Figures A.23 (a)–(c). The corrections, while not

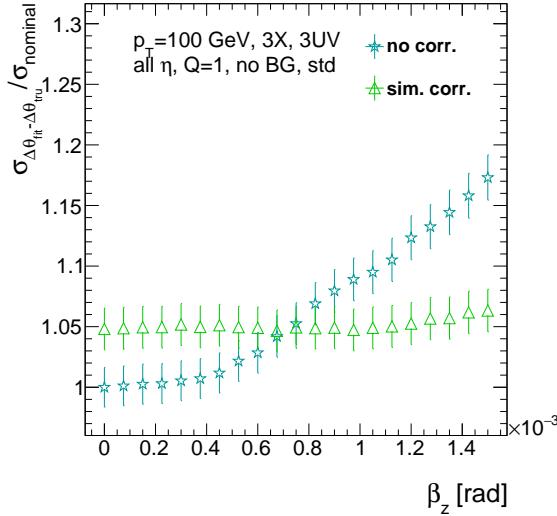
as effective as for the simple translation cases, are still very effective with the quoted misalignment values for bias shifts down to 0.001 mrad (0.25 mrad) for  $\langle \vartheta \rangle$  ( $\Delta\vartheta$ ) and no more than a 2% degradation in  $\sigma_{\Delta_{fit} - \Delta_{true}}$  for  $|\gamma_s| = 0.3$  mrad.



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

### A.14 ROTATION MISALIGNMENTS AROUND THE WEDGE VERTICAL AXIS ( $\beta_z$ )

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

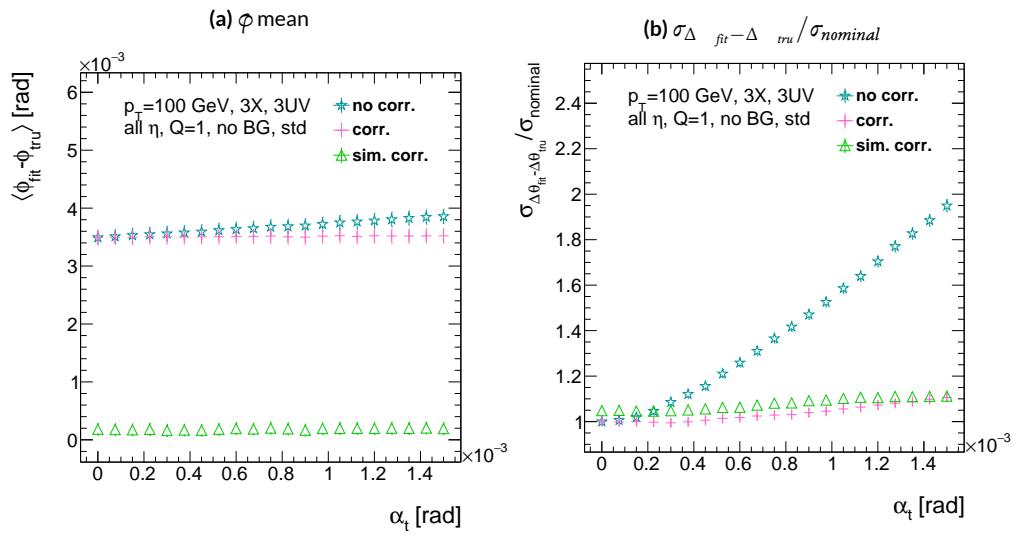


**Figure A.24:** The effects of rotations in the  $z$  axis on  $\sigma_{\Delta \theta_{fit} - \Delta \theta_{true}} / \sigma_{nominal}$  as a function of  $\beta_z$  both with and without misalignment corrections.

### A.15 ROTATION MISALIGNMENTS AROUND THE AXIS PARALLEL TO THE BEAMLINE ( $\alpha_t$ )

Misalignments arising from rotations around the  $t$  axis (parallel to the beamline at the center of the base of the wedge) are essentially rotations in the  $\varphi$  direction. The quantities of interest most affected are the  $\varphi$  bias and  $\sigma_{\Delta_{fit} - \Delta_{tru}}$ , as shown in Figures A.25 (a) and (b), respectively, and correspond to a shift in  $\varphi$  bias of 0.2 mrad and a 10% degradation in  $\sigma_{\Delta_{fit} - \Delta_{tru}}$  for  $\alpha_t = 0.3$  mrad (corresponding to a linear shift of  $\sim 1$  mm). The raw instead of fitted mean  $\varphi$  biases is used in Figure A.25 (a) to better illustrate the effect of misalignment.

Since the effect of misalignment is dependent on horizontal (along the strip direction,  $\hat{s}$ ) in addition to vertical information, corrections cannot be applied before a fit takes place. The  $\varphi$  bias shift is uniform over the entire wedge, so a constant additive correction to  $\varphi$  based on the level of misalignment can be applied to all fits depending on how many misaligned stereo planes enter in the fit.  $\Delta\vartheta$  is less straightforward, but corrections to the  $y$  and  $z$  information used in the local slope calculation in Equation A.4 can be applied once  $\vartheta_{fit}$  and  $\varphi_{fit}$  are known. These corrections are calculated ahead of time in bins of uniform  $\gamma$  and  $\varphi$  as with the simulation corrections using the same framework as the misalignment calculation. The results of both types of correction can be seen in Figure A.22. The apparent discrepancy between the simulation and analytic corrections in the  $\varphi$  bias happens for the same reason as in the  $\Delta s$  misalignment correction cases, as simulation correction restores a more Gaussian shape to the  $\varphi$  residual distribution opposed to the uncorrected nominal case, as discussed in Section A.9.



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

## A.16 CONCLUSION

The algorithm for Micromegas detectors in the NSW Trigger Processor performs well in a variety of conditions and has proven robust to a number of effects to deliver measurements on muon tracks of the three angles  $\vartheta$ ,  $\phi$ ,  $\Delta\vartheta$ . Under nominal conditions, the rms values for the residuals of these quantities are 0.364 mrad for  $\vartheta$ , 8.12 mrad for  $\phi$ , and 1.47 mrad for  $\Delta\vartheta$ . Algorithm performance was found to be largely independent of the charge threshold setting, and a hit majority BCID association was found to provide proper timing information over 99.7% even in the most relaxed settings ( $_{2X+1}UV$  coincidence threshold requirement+wide slope-road+background). The introduction of wide slope-roads to better mimic potentially limited algorithm resources at run time and the introduction of incoherent background was found to have a manageable effect on fit quantity residual rms values and on total algorithm efficiency for sufficiently stringent coincidence threshold. The effects of the three translation and three rotation misalignments specified by AMDB convention were studied, and correction methods for each of the six cases was developed. Simulation-based corrections were found to improve nominal algorithm performance to residual rms value of 0.291 mrad for  $\vartheta$ , 3.19 mrad for  $\phi$ , and 1.54 for  $\Delta\vartheta$ , which represent improvements of 20%, 62%, and -4.7%, respectively. Misalignment corrections were found to restore nominal performance for all but the rotation around the  $s$  axis, and a summary of tolerances may be found in Table A.4.

**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  $\vartheta$ , 1 mrad for  $\phi$ , and 0.25 mrad for  $\Delta\vartheta$  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.

	No Correction	Correction
$\Delta s$	4 mm ( $\phi$ bias)	$> 5$ mm
$\Delta z$	0.25 mm ( $\Delta\vartheta$ )	$> 5$ mm
$\Delta t$	0.25 mm ( $\Delta\vartheta$ )	$> 5$ mm
$\gamma_s$	0.15 mrad ( $\Delta\vartheta$ bias)	0.75 mrad
$\beta_z$	0.9 mrad ( $\Delta\vartheta$ rms)	$> 1.5$ mrad
$\alpha_t$	0.375 mrad ( $\Delta\vartheta$ rms)	$> 1.5$ mrad

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

Conventional Wisdom

# B

## Telescoping Jets

ANOTHER APPROACH TO IMPROVING  $ZH \rightarrow \ell\ell b\bar{b}$  is the use of telescoping jets<sup>23</sup>, which harnesses the power of multiple event interpretations. The use of multiple event interpretations was originally developed with non-deterministic jet algorithms like the Q-jets (“quantum” jets) algorithm<sup>2</sup>. When a traditional or “classical” algorithm, such as the Cambridge-Aachen<sup>2</sup> and anti- $k_t$ <sup>2</sup> 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>?</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>. Additionally,  $t\bar{t}$  production and di-boson ( $ZZ$ ) production processes were studied in validation. The  $t\bar{t}$  samples are generated by the Powheg gen-

erator<sup>??</sup>, using CT10 PDFs, interfaced with PYTHIA6, and the ZZ samples are generated with the HERWIG<sup>?</sup> generator using the CTEQ6L1 PDFs and the AUET2 tune<sup>??</sup>.

## B.2 JET RECONSTRUCTION AND CALIBRATION

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

After the telescoping jet axes have been established, telescoping jets are constructed using topological clusters in the calorimeters at a variety of jet cone sizes. Including the original anti- $k_t$  jets used for the  $R = 0.4$  case, twelve total sets of jets of cone sizes ranging from  $R = 0.4\text{--}1.5$  are constructed, with each successive size having a radius 0.1 larger than the preceding set. For each jet axis, telescoping jets consist of any topological cluster lying within  $R$  of the axis. In the event of overlap, clusters are assigned to the closer jet axis. If a given cluster is equidistant from the two jet axes, the cluster

is assigned to whichever jet axis is associated with the anti- $k_t$  jet with higher  $p_T$ . Calibration for the telescoping jets is conducted using corrections for anti- $k_t$  calorimeter topological cluster jets; the  $R = 0.4$  corrections are used for telescoping  $R = 0.5$ , and the  $R = 0.6$  corrections are used for telescoping  $R \geq 0.6$  (cf. Sec. B.4). The telescoping cone jets ( $R \geq 0.5$ ) at reconstructed level are trimmed using Cambridge-Aachen jets with  $R = 0.3$  and  $f_{cut} = 0.05$  with respect to the untrimmed jet  $p_T$ <sup>2</sup>. Since these jets are trimmed, the active area correction is not applied. In the event a  $Z$  candidate electron falls within  $R$  of the axis of a telescoping jet, its 4-momentum is subtracted from that of the jet vectorially.

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

### B.3 EVENT RECONSTRUCTION AND SELECTION

Events are selected on the basis of a combination of leptonic, jet, and missing  $E_T$  requirements, which are outlined in Table B.1. Leptons are categorized by three sets of increasingly stringent quality requirements, which include lower limits on  $E_T$ , upper limits on  $|\eta|$ , impact-parameter requirements, and track-based isolation criteria. The requirements differ for electrons<sup>?</sup> and muons<sup>?</sup>. Events are selected with a combination of single lepton, dielectron, and dimuon requirements. Each event must contain at least one lepton passing medium requirements and at least one other lepton passing loose requirements. These leptons are used to create a dilepton invariant mass cut to ensure the presence of a  $Z$  boson and suppress multijet backgrounds.

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

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

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

the validation.

**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]	$\Delta R(b, b)$
0–90	0.7–3.4
90–120	0.7–3.0
120–160	0.7–2.3
160–200	0.7–1.8
$> 200$	$< 1.4$

#### B.4 VALIDATION OF JET CALIBRATION

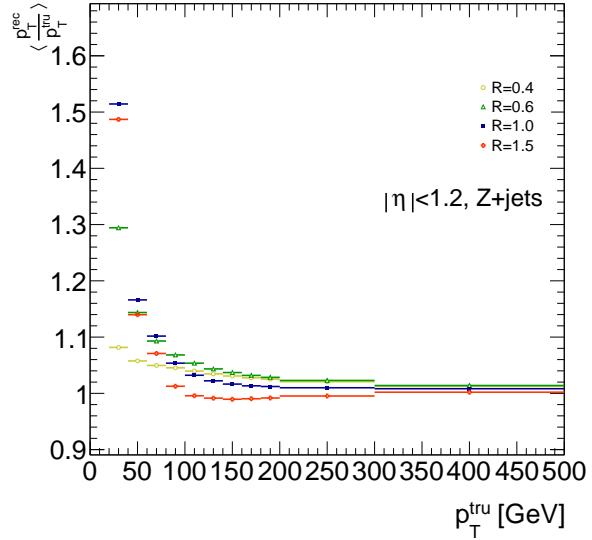
In order to validate the jet energy scale and resolution of jets constructed with this telescoping-jets algorithm, values of  $p_T^{\text{rec}}/p_T^{\text{true}}$  are studied for each value of  $R$  for the  $Z$ +jets MC sample. In a given event, all jets, not just the two hardest  $b$ -tagged jets, are telescoped. These jets are constructed in the same way as in the reconstructed and truth-level analyses; reconstructed-level jets are made from calorimeter topological clusters within  $R$  of the anti- $k_t$  with  $R = 0.4$  jet axes and then trimmed, and

truth-level jets are made from stable truth particles within  $R$  of the anti- $k_t$  with  $R = 0.4$  jet axes.

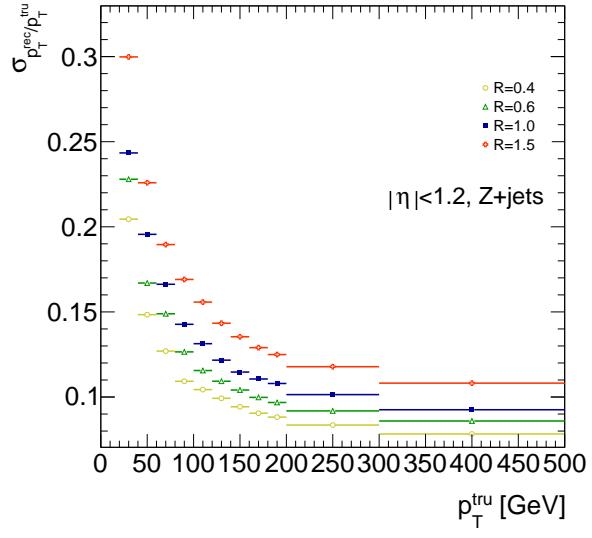
The reconstructed and truth-level telescoping jet ensembles are matched according to the separation in the  $(\eta, \phi)$  plane of their corresponding anti- $k_t$  with  $R = 0.4$  jets used as seeds. Only jets with  $|\eta| < 1.2$  are examined here, and the results of studies on the  $ZH_{125}$ ,  $ZZ$ , and  $t\bar{t}$  samples, as well as over other  $|\eta|$  ranges, are outlined in <sup>23</sup>. Any reconstructed jets not within  $\Delta R = 0.3$  of a truth jet are discarded. In the event that multiple reconstructed jets are the same distance away from a given truth jet, the reconstructed jet with the highest  $p_T$  gets matched. Matching is retained for all  $R$  values (i.e. telescoping seeds are matched, and telescoping jets are assumed to match if the anti- $k_t$  jets from which their seeds are derived match).

Once anti- $k_t$  with  $R = 0.4$  reconstructed and truth jets are matched, response functions are created 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 bin for 200–300 GeV, and one bin for 300–500 GeV for each  $R$ , with bins chosen for purposes of statistics. Ensembles with  $p_T^{tru} < 20$  GeV are ignored since no calibration exists for jets with transverse 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 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, and the error on the mean is taken from the error of this parameter in the fit. The resolutions are the values of the square root of the variance on this fit. As the total response distributions in Figure B.1 show, performance is best for low  $R$  values and high values of  $p_T^{tru}$ . Figure B.1 shows the  $R = 0.4$  (antik $_t$ ) case to show a baseline for performance,  $R = 0.6$  to show the deviations with “correct” calibrations, and  $R = 1.0, 1.5$  to show how big those deviations get with larger  $R$  jets. The resolutions,  $\sigma_{p_T^{rec}/p_T^{tru}}$ , as a function of  $p_T^{tru}$  are shown in Figure B.1(b). For  $p_T^{tru} > 60$  GeV, response is fairly

consistent over various  $R$  values. Resolution, as might naïvely be expected, is worse for increasingly larger values of  $R$ . For  $p_T^{tru} < 60$  GeV, resolution degrades, and response degrades in particular for increasing  $R$ ; this is likely a result from residual pileup effects.



(a)



(b)

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

## B.5 TRUTH-LEVEL ANALYSIS

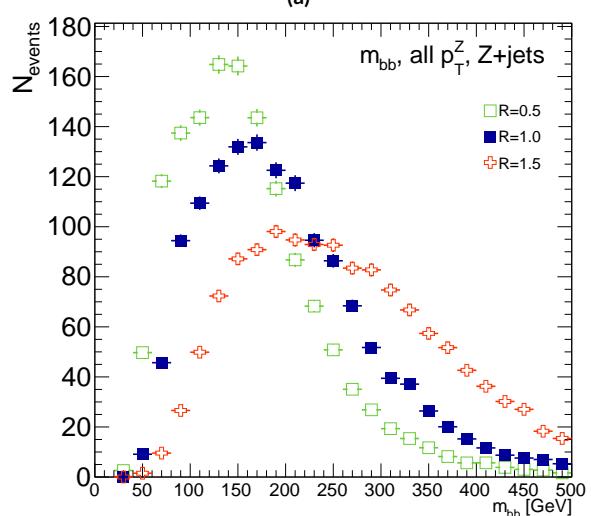
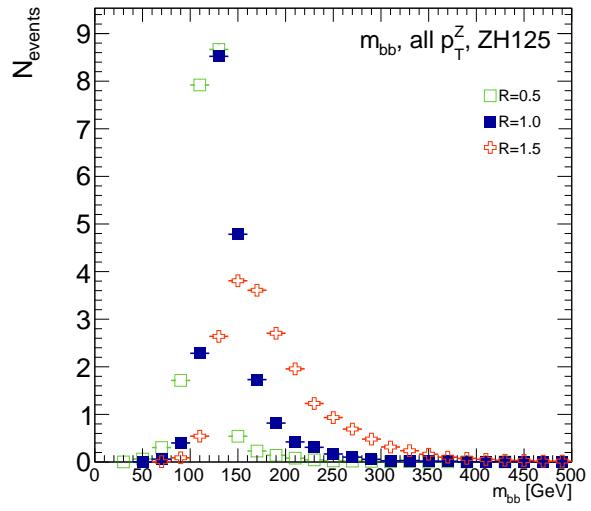
To understand the limits and sources of any potential improvements, a truth-level analysis was conducted on MC samples with a  $ZH_{125}$  signal sample and a  $Z+jets$  background sample. Distributions for the dijet invariant mass,  $m_{bb}$ , were made for each telescoping radius.\* Both signal and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

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

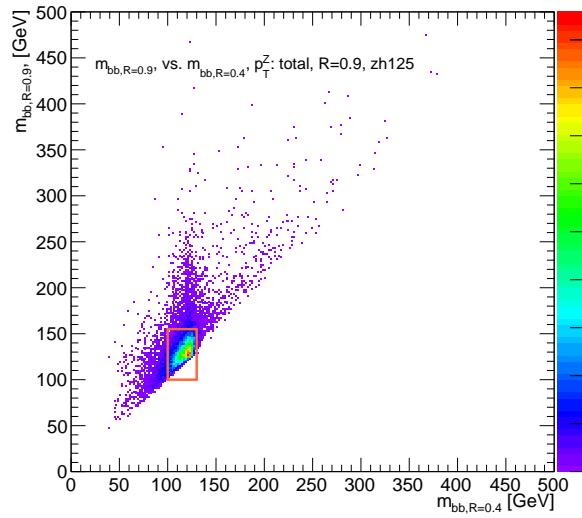
Another way to take advantage of multiple event interpretations is to make use of an event's cut-weight, denoted  $z$  and defined as the fraction of interpretations in a given event that pass a certain set of cuts (in this note, a cut on  $m_{bb}$ ). The distribution of cut-weights for a sample of events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

---

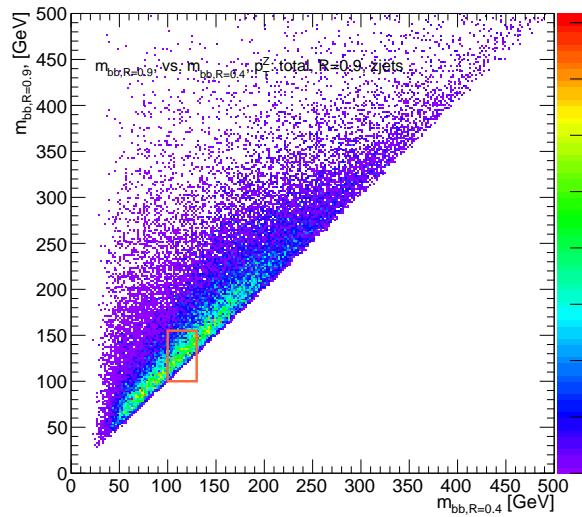
\*Distributions for  $m_{bb}$  at truth and reconstructed level for all telescoping radii studied may be found in Appendix ??



**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.

## B.6 ERRORS ON TELESCOPING SIGNIFICANCES

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

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

where  $0.5 + \sqrt{0.25 + B}$  is the characteristic upward fluctuation expected in a Poissonian data set using the Pearson chi-square test<sup>2</sup>.

## B.7 COUNTING

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

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

The error for the Poissonian case is:

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

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

## B.8 MULTIPLE EVENT INTERPRETATIONS

Using multiple event interpretations changes the formulae used in with simple counting. That is,  $S$  is not necessarily merely  $N_S$ , the number of events passing some signal cuts, and similarly for  $B$  and  $N_B$ . Using an event weighting by some function of the cut-weight,  $z$ , denoted  $t(z)$ ,  $S = N_S \langle t \rangle_s$  and  $B = N_B \langle t^2 \rangle_B$ . So

$$\mathcal{S}_{t,gaus} = \frac{N_S \langle t \rangle_s}{\sqrt{N_B \langle t^2 \rangle_B}} \rightarrow \mathcal{S}_{t,pois} = \frac{N_S \langle t \rangle_s}{0.5 + \sqrt{0.25 + N_B \langle t^2 \rangle_B}} = \frac{N_S \int_0^1 dz t(z) \xi_S(z)}{0.5 + \sqrt{0.25 + N_B \int_0^1 dz t^2(z) \xi_B(z)}} \quad (\text{B.4})$$

For histograms, everything is done bin-wise. The notation used below is as follows:  $\xi_i$  is the value of  $\xi(z)$  at bin  $i$  (where the bins run from 0 to  $n_{tel}$ , where  $n_{tel}$  is the total number of telescoping radii).  $t_i = t_i(\xi_{S,i}, \xi_{B,i}, i/n_{tel})$  is the value of  $t(z)$  at bin  $i$ , which can depend, in principle, on  $\xi_{S,i}$ ,  $\xi_{B,i}$  and  $i/n_{tel}$  (the last of which is  $z$  in bin  $i$ ). Explicitly,

$$N = \sum_{i=0}^{n_{tel}} \xi_i, \quad \int_0^1 dz t(z) \xi_S(z) = \sum_{i=0}^{n_{tel}} t_i \xi_{S,i}, \quad \int_0^1 dz t^2(z) \xi_B(z) = \sum_{i=0}^{n_{tel}} t_i^2 \xi_{B,i}$$

For the calculations that follow, let  $\xi = \sum_{i=0}^{n_{tel}} t_i \xi_{S,i}$ ,  $\psi = 0.5 + \sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \xi_{B,i}}$ ,  $\partial_S = \frac{\partial}{\partial \xi_{S,i}}$  (and similarly for  $B$ ), so  $\mathcal{S}_t = N_S \xi / \psi$

Some partial derivatives:

$$\partial_S N_S = 1, \quad \partial_{B,i} N_B = 1$$

$$\begin{aligned}\partial_S \xi &= t_i + (\partial_S t_i) \xi_{S,i}, & \partial_B \xi &= (\partial_B t_i) \xi_{B,i} \\ \partial_S \psi &= \frac{N_B t_i (\partial_S t_i) \xi_{B,i}}{\sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \xi_{B,i}}}, & \partial_B \psi &= \frac{\sum_{i=0}^{n_{tel}} t_i^2 \xi_{B,i} + N_B (t_i^2 + 2t_i (\partial_B t_i) \xi_{B,i})}{2\sqrt{0.25 + N_B \sum_{i=0}^{n_{tel}} t_i^2 \xi_{B,i}}} \\ \partial_S \mathcal{S}_t &= \frac{\xi}{\psi} + \frac{N_S}{\psi} \partial_S \xi - \frac{N_S \xi}{\psi^2} \partial_S \psi, & \partial_B \mathcal{S}_t &= N_S \left( \frac{1}{\psi} \partial_B \xi - \frac{\xi}{\psi^2} \partial_B \psi \right)\end{aligned}$$

Thus,

$$\Delta \mathcal{S}_{t,i} = \left[ \frac{\xi}{\psi} + \frac{N_S}{\psi} \partial_S \xi - \frac{N_S \xi}{\psi^2} \partial_S \psi \right] \Delta \xi_{S,i} \oplus N_S \left[ \frac{1}{\psi} \partial_B \xi - \frac{\xi}{\psi^2} \partial_B \psi \right] \Delta \xi_{B,i} \quad (\text{B.5})$$

and the total error is given by the sum in quadrature over all bins  $i$  of  $\Delta \mathcal{S}_{t,i}$ .

$$\text{B.9} \quad t(z) = z$$

With  $t(z) = z$ ,  $t_i = i/n_{tel}$ , so  $\partial_S t_i = \partial_B t_i = 0$ . So:

$$\begin{aligned}\partial_S \psi &= \partial_B \xi = 0 \\ \partial_S \xi &= \frac{i}{n_{tel}} \\ \partial_B \psi &= \frac{\sum_i i^2 \xi_{B,i} + N_B i^2}{n_{tel} \sqrt{n_{tel}^2 + N_B \sum_i i^2 \xi_{B,i}}}\end{aligned}$$

so  $\Delta\mathcal{S}_{z,i}$  reduces to

$$\Delta\mathcal{S}_{t,i} = \left[ \frac{\xi + N_S t_i}{\psi} \right] \Delta\varrho_{S,i} \oplus \left[ \frac{N_S \xi}{\psi^2} \partial_B \psi \right] \Delta\varrho_{B,i} \quad (\text{B.6})$$

$$\text{B.10} \quad t(z) = \varrho_S(z) / \varrho_B(z)$$

With the likelihood optimized<sup>†</sup>  $t^*(z) = \varrho_S(z) / \varrho_B(z)$ ,  $t_i = \varrho_{S,i} / \varrho_{B,i}$ , so  $\partial_S t_i = 1 / \varrho_{B,i}$  and  $\partial_B t_i = -\varrho_{S,i} / \varrho_{B,i}^2$ . So:

$$\begin{aligned} \partial_S \xi &= 2 \frac{\varrho_{S,i}}{\varrho_{B,i}} = 2t_i \\ \partial_B \xi &= -\frac{\varrho_{S,i}}{\varrho_{B,i}} = -t_i \\ \partial_S \psi &= \frac{N_B t_i}{\sqrt{0.25 + N_B \sum_i \varrho_{S,i}^2 / \varrho_{B,i}}} \\ \partial_B \psi &= \frac{\sum_i \varrho_{S,i}^2 / \varrho_{B,i} - N_B \left( \varrho_{S,i} / \varrho_{B,i} \right)^2}{\sqrt{1 + 4N_B \sum_i \varrho_{S,i}^2 / \varrho_{B,i}}} \end{aligned}$$

simplifying somewhat the terms in the per bin error in Equation B.6.

The new significance figure using multiple event interpretations becomes, with  $\varrho_S$  and  $\varrho_B$  denoting the cut-weight distributions in signal and background, respectively

$$\frac{S}{\delta B} = \frac{N_S \langle t \rangle_s}{\sqrt{0.5 + \sqrt{0.25 + N_B \langle t^2 \rangle_B}}} \quad (\text{B.7})$$

Of particular interest is the likelihood optimized  $t(z)$ ,<sup>‡</sup>  $t^*(z) = \varrho_S(z) / \varrho_B(z)$ .  $m_{bb}$  windows are

---

<sup>†</sup>for the Gaussian statistics case

<sup>‡</sup>Derived under the assumption of Gaussian statistics in Ref<sup>?</sup>

chosen separately for each scheme studied to maximize total significances and are summarized in

Table B.3.

$$\left(\frac{S}{\delta B}\right)_z = \frac{N_S \epsilon_S}{0.5 + \sqrt{0.25 + N_B (\epsilon_B^2 + \sigma_B^2)}} \quad (B.8)$$

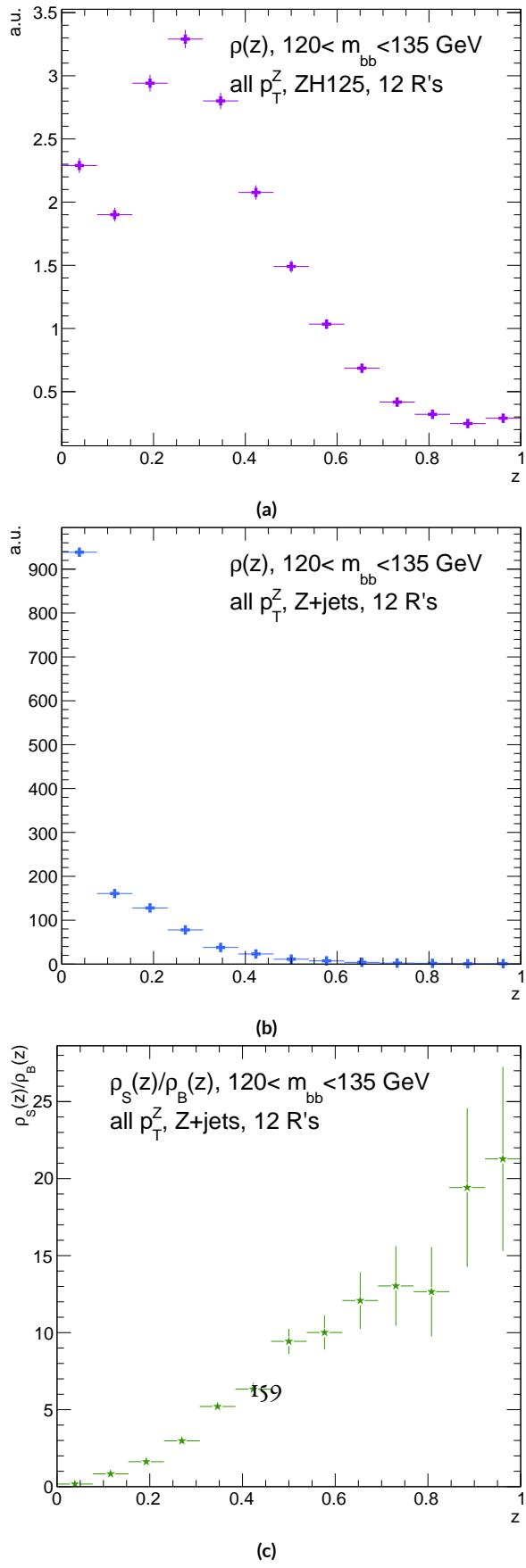
$$\left(\frac{S}{\delta B}\right)_{t^*(z)} = \frac{N_S \int_0^1 dz \frac{\dot{\xi}_S(z)}{\xi_B(z)}}{0.5 + \sqrt{0.25 + N_B \int_0^1 dz \frac{\dot{\xi}_S(z)}{\xi_B(z)}}} \quad (B.9)$$

where  $\epsilon_{S,B}$  are the means of  $\xi_{S,B}(z)$  and  $\sigma_B^2$  is the variance of  $\xi_B(z)$ . Further details can be found in Refs.<sup>23</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	$\text{anti-}k_t R = 0.4$ $t(z) = z$ $t(z) = \xi_S(z) / \xi_B(z)$ $\text{anti-}k_t R = 0.4, \text{telescoping } R = 0.6$	$90-140 \text{ GeV}$ $110-155 \text{ GeV}$ $110-155 \text{ GeV}$ $95-140 \text{ GeV} (R = 0.4), 105-160 \text{ GeV} (R = 0.6)$
Truth	$\text{anti-}k_t R = 0.4$ $t(z) = z$ $t(z) = \xi_S(z) / \xi_B(z)$ $\text{anti-}k_t R = 0.4, \text{telescoping } R = 0.9$	$100-130 \text{ GeV}$ $115-140 \text{ GeV}$ $120-135 \text{ GeV}$ $100-130 \text{ GeV} (R = 0.4), 100-155 \text{ GeV} (R = 0.9)$

The truth-level distributions  $\xi_S(z)$ ,  $\xi_B(z)$ , and  $\xi_S(z) / \xi_B(z)$  are shown for the  $m_{bb}$  window that optimizes  $(S/\delta B)_{t^*(z)}$  in Figure B.4, and significance improvements as a function of  $p_T^Z$  are summarized in Figure B.5. Uncertainties in Figures B.5 and B.9 are statistical uncertainties. JES systematics 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



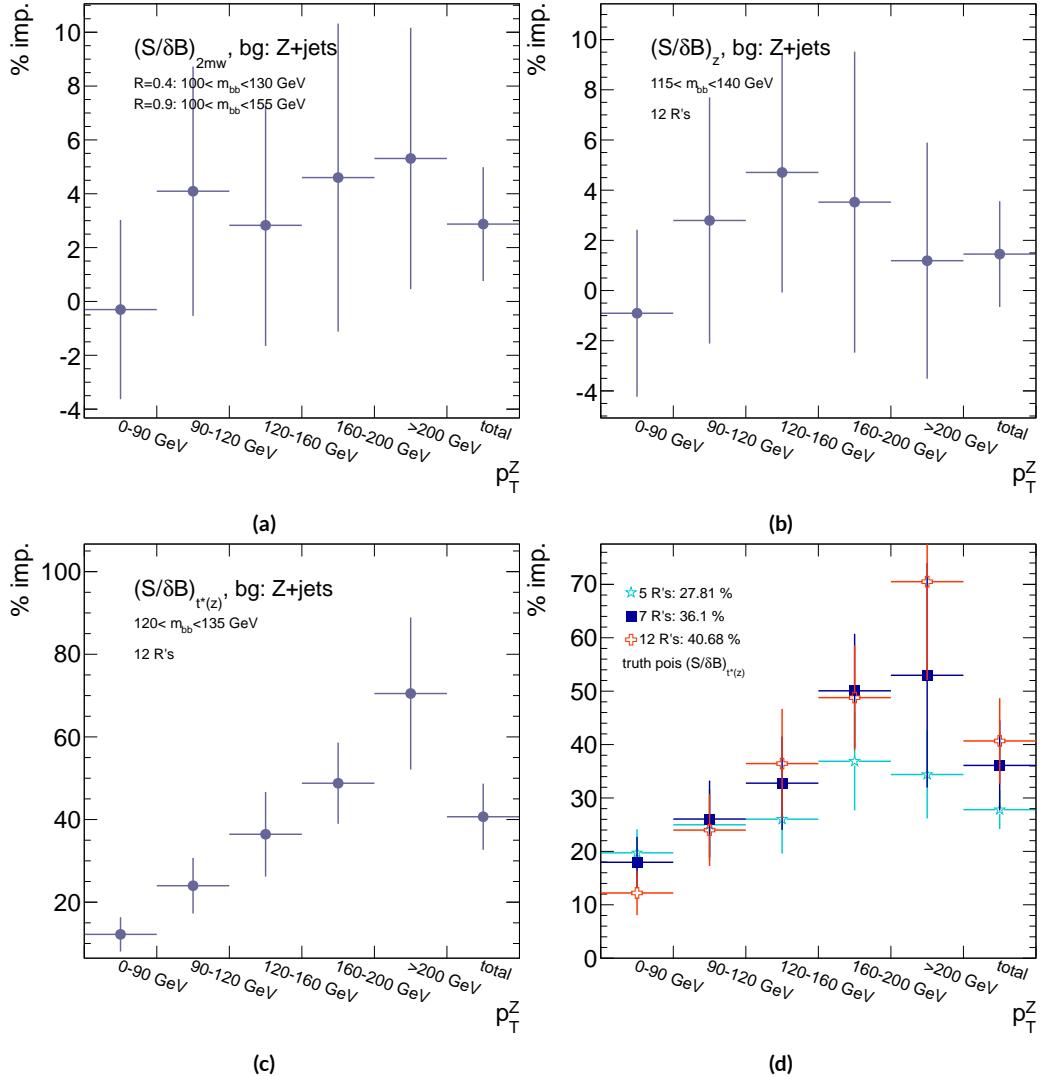
**Figure B.4:** Truth-level  $\rho(z)$  distributions for the  $m_{bb}$  window optimizing  $(S/\delta B)_{t^*(z)} \cdot \rho_S(z)$  for the signal  $ZH125$  sample is shown in (a), and  $\rho_B(z)$  for the background  $Z+jets$  sample is shown in (b). The distribution of  $\rho_S(z) / \rho_B(z)$

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 increase in improvement with increasing  $p_T^Z$ . Figure B.5 (d) summarizes the improvements with respect to  $p_T^Z$  for the  $t^*(z)$  event weight for five, seven, and twelve telescoping radii (interpretations) per event. Improvements increase with a greater number of interpretations and are more pronounced at higher  $p_T^Z$  for this scheme.<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,  $\xi_B(z)$  in Figure B.4 (b) behaves as one might with a marked peak at  $z = 0$ , the signal  $\xi_S(z)$  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  $\xi_S/\xi_B$  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>?</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).

## B.II RECONSTRUCTED-LEVEL ANALYSIS

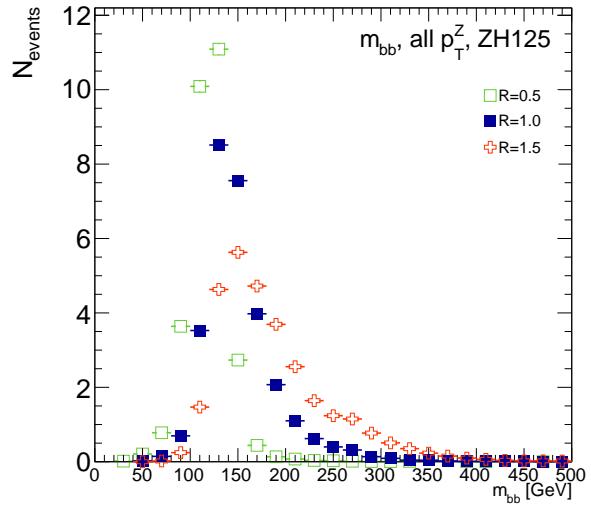
At reconstructed level, the same overall effect of introducing a high tail in  $m_{bb}$  distributions with increasing  $R$  is evident in comparing Figures B.2 and B.6. The optimal  $m_{bb}$  windows, however, grow larger, due to the lack of truth-level information.

Total significance gains at reconstructed level for the two dimensional  $m_{bb}$  cut and the  $t(z) = z$  case are similar, at 2.87% and 1.45%, respectively. The optimal two-dimensional  $m_{bb}$  cut at reconstructed level is  $95 < m_{bb,R=0.4} < 140$  GeV,  $105 < m_{bb,R=0.6} < 160$  GeV. Just as at truth level, the  $R = 0.4$   $m_{bb}$  cut is comparable to the optimal single  $R = 0.4$   $m_{bb}$  cut, and the second  $m_{bb}$  cut is at similar values (cf. Table B.3 and Figures B.3 and B.7). However, the optimal second telescoping radius is markedly smaller at  $R = 0.6$  versus the optimal truth-level second radius of  $R = 0.9$ , which suggests that effects like pileup at reconstructed level obscure correlations between the  $R = 0.4$  interpretations and limit the usefulness of larger  $R$  interpretations in this particular scheme. The  $t(z) = z$  case has a wider optimal window and yields about half the improvement it does at truth level.<sup>||</sup>

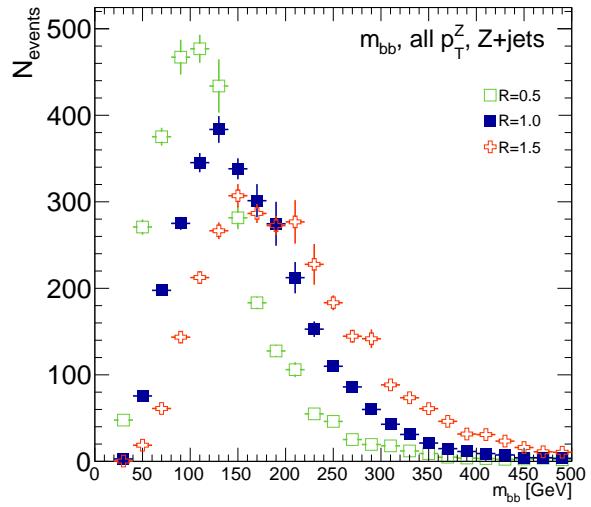
The optimal  $m_{bb}$  window for the  $t^*(z)$  case is also markedly wider at reconstructed level, at  $110 < m_{bb} < 155$  GeV in comparison to the truth-level optimal  $120 < m_{bb} < 135$  GeV. The  $\varrho(z)$  distributions for the signal  $ZH_{125}$  and background  $Z + \text{jets}$  as well as the  $\varrho_S(z) / \varrho_B(z)$  in this window are shown in Figure B.8. Compared with the truth-level distributions in Figure B.4, both the signal and background optimal  $\varrho(z)$  distributions have higher values at higher  $z$ . The peak in  $\varrho_S(z)$  at

---

<sup>||</sup>A fuller treatment of this scheme is given in Section ??.

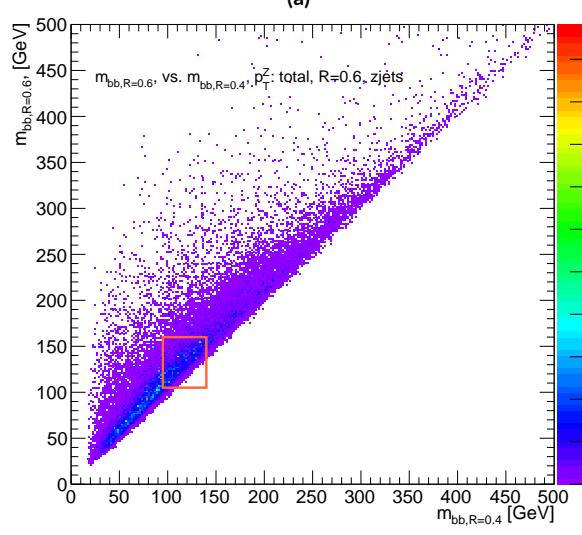
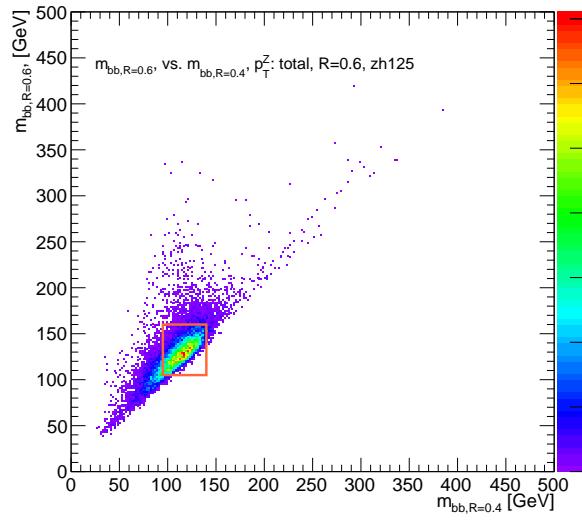


(a)

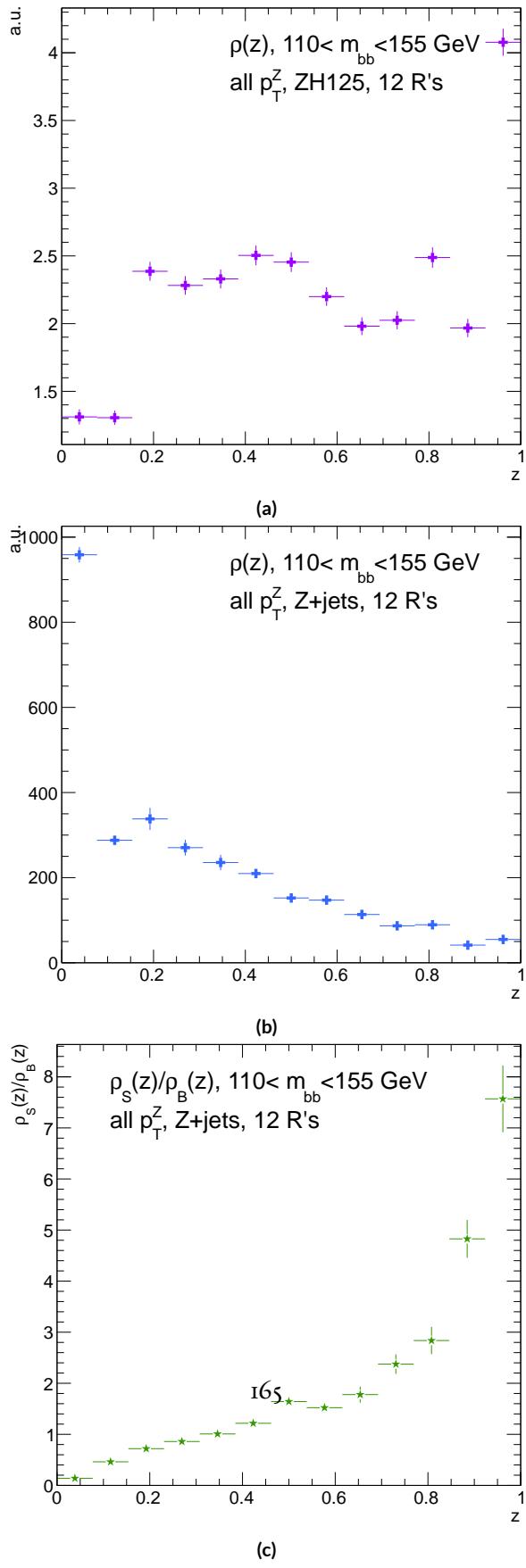


(b)

**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  $m_{bb,R=0.6}$  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.

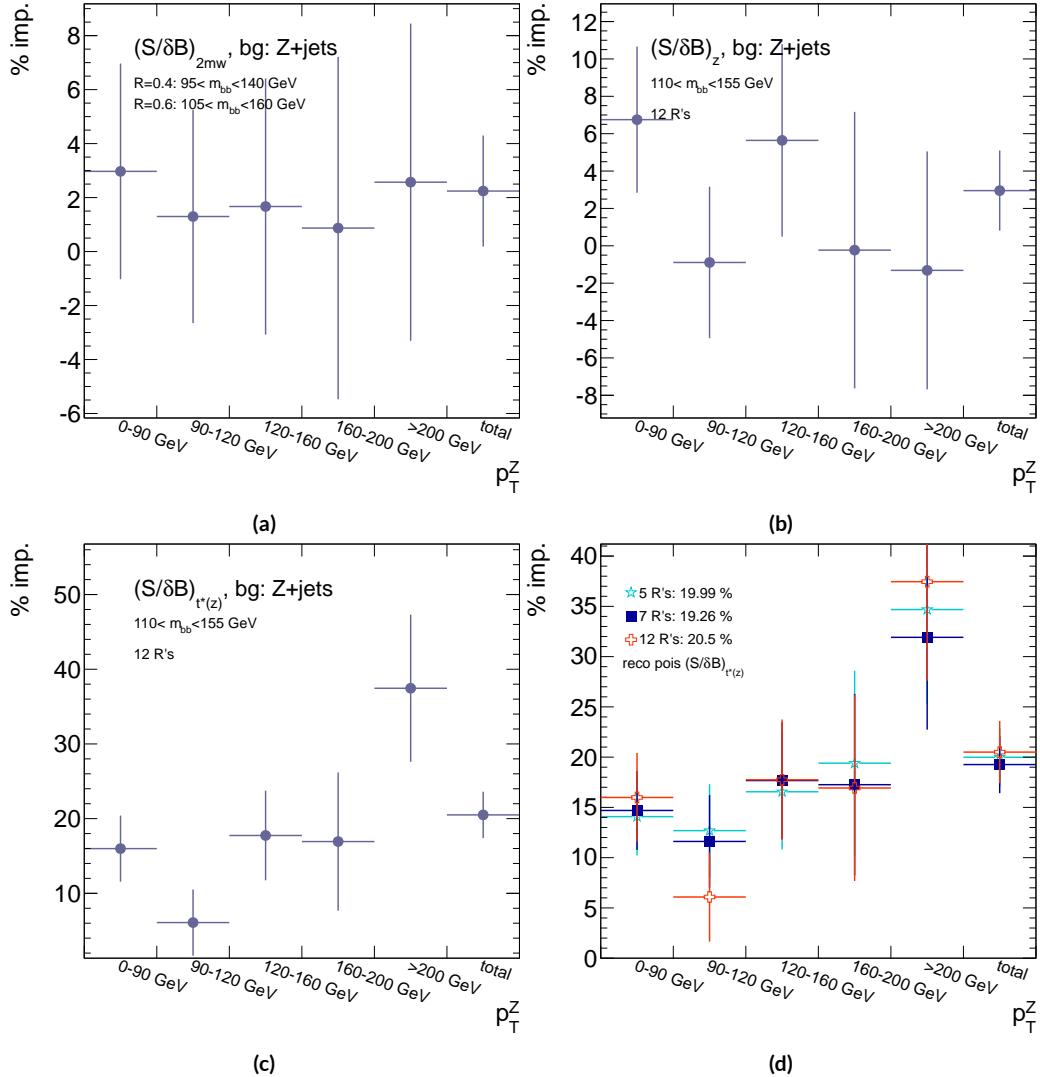


**Figure B.8:** Reconstructed-level  $\xi(z)$  distributions for the  $m_{bb}$  window optimizing  $(S/\delta B)_{t^*(z)} \cdot \xi_S(z)$  for the signal  $ZH125$  sample is shown in (a), and  $\xi_B(z)$  for the background  $Z+jets$  sample is shown in (b). The distribution of  $\xi_S(z)/\xi_B(z)$  for these samples is shown in (c).

$z = 1$  suggests that at reconstructed level, maximizing the number of more “signal-like” events is the key to optimizing significances, as opposed to the optimal, background suppressing  $\xi(z)$  distributions at truth level. The use of a greater number of interpretations per event (telescoping radii) does appear to result in overall greater improvement as at truth level, as twelve radii performed better than five, but this is less clear at reconstructed level, as shown in Figure B.9 (d). The improvement at reconstructed level using an event weight of  $t^*(z)$  is 20.5%, just over half the improvement at truth level but still quite significant. Summaries of improvements as a function of  $p_T^Z$  for all three cases 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).

## B.12 CONCLUSIONS AND PROSPECTS

The use of telescoping jets to provide multiple event interpretations shows promise as an avenue to increase significances in the  $H \rightarrow b\bar{b}$  search in ATLAS and make an observation in the systematics-limited environment of early Run 2. A preliminary study using the telescoping jets algorithm with 12 telescoping radii to build 12 event interpretations on 2012 Monte Carlo based on the full, cut-based Run 1 analysis yielded a 20.5% improvement in  $S/\delta B$  over using anti- $k_t$  with  $R = 0.4$  alone at reconstructed level using a likelihood maximized event weighting to study the  $ZH \rightarrow llb\bar{b}$  process. The jets used in this note at reconstructed level were trimmed in order to guarantee reasonable resolution in the large- $R$  interpretations. The algorithm, in particular, showed discriminating power at high  $p_T^Z$ , so better performance can be expected in Run 2 with a higher  $\sqrt{s}$  and higher numbers of events with large  $p_T^Z$ . Additionally, the many simplifying assumptions regarding jet calibration and the relatively basic use of information<sup>\*\*</sup> from multiple invariant masses in this note suggest that even further improvements than those quoted are possible. While this note did not explore the correlations between multiple event interpretations and the variables used in the BDT of the latest multivariate version of the  $H \rightarrow b\bar{b}$  analysis<sup>19</sup>, new phenomenological studies suggest that such correlations are not strong<sup>3</sup>. The corresponding reconstructed-level study, using a BDT, is left for future work. Also left for future work are better understanding the effects of jet trimming and which interpretations are the most useful.

---

<sup>\*\*</sup>For examples of more sophisticated treatments compared to the treatment in this note, see Ref<sup>3</sup>.

# References

- [18] (2016). *Monte Carlo Generators for the Production of a W or Z/ $\gamma^*$  Boson in Association with Jets at ATLAS in Run 2*. Technical Report ATL-PHYS-PUB-2016-003, CERN, Geneva.
- [19] Ahmadov, F., Alio, L., Allbrooke, B., Bristow, T., Buescher, D., Buzatu, A., Coadou, Y., Debenedetti, C., Enari, Y., Facini, G., Fisher, W., Francavilla, P., Gaycken, G., Gentil, J., Goncalo, R., Gonzalez Parra, G., Grivaz, J., Gwilliam, C., Hageboeck, S., Halladjian, G., Jackson, M., Jamin, D., Jansky, R., Kiuchi, K., Kostyukhin, V., Lohwasser, K., & Lopez Mateos, D, e. a. (2014). *Supporting Document for the Search for the bb decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector*. Technical Report ATL-COM-PHYS-2014-051, CERN, Geneva.
- [20] Buckley, A., Butterworth, J., Grellscheid, D., Hoeth, H., Lonnblad, L., Monk, J., Schulz, H., & Siegert, F. (2010). Rivet user manual.
- [21] Buzatu, A. & Wang, W. (2016). *Object selections for SM Higgs boson produced in association with a vector boson in which H-gt;bb and V decays leptonically with Run-2 data: Object support note for VH(bb) 2015+2016 dataset publication*. Technical Report ATL-COM-PHYS-2016-1674, CERN, Geneva. This is a support note for the VH(bb) SM publication using the 2015+2016 datasets.
- [22] Catani, S., Dokshitzer, Y. L., Seymour, M. H., & Webber, B. R. (1993). Longitudinally-invariant  $k_{\perp}$ -clustering algorithms for hadron-hadron collisions. *Nucl. Phys. B*, 406(CERN-TH-6775-93, LU-TP-93-2), 187–224. 38 p.
- [23] Chan, S., Huth, J., Lopez Mateos, D., & Mercurio, K. (2015a). *ZH → llbb Analysis with Telescoping Jets*. Technical Report ATL-PHYS-INT-2015-002, CERN, Geneva.
- [24] Chan, S. K.-w., Lopez Mateos, D., & Huth, J. (2015b). *Micromegas Trigger Processor Algorithm Performance in Nominal, Misaligned, and Corrected Misalignment Conditions*. Technical Report ATL-COM-UPGRADE-2015-033, CERN, Geneva.

- [25] Clark, B., Lopez Mateos, D., Felt, N., Huth, J., & Oliver, J. (2014). *An Algorithm for Micromegas Segment Reconstruction in the Level-1 Trigger of the New Small Wheel*. Technical Report ATL-UPGRADE-INT-2014-001, CERN, Geneva.
- [26] Collaboration, T. A., Aad, G., Abat, E., Abdallah, J., & A A Abdelalim, e. a. (2008). The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08), S08003.
- [27] Gleisberg, T., Höche, S., Krauss, F., Schönher, M., Schumann, S., Siegert, F., & Winter, J. (2009). Event generation with sherpa 1.1. *Journal of High Energy Physics*, 2009(02), 007.
- [28] Heinemeyer, S. & Mariotti, e. a. (2013). *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. Technical Report CERN-2013-004, CERN-2013-004, Geneva. Comments: 404 pages, 139 figures, to be submitted to CERN Report. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>.
- [29] Lavesson, N. & Lonnblad, L. (2005). W+jets matrix elements and the dipole cascade.
- [30] Lonnblad, L. (2001). Correcting the colour-dipole cascade model with fixed order matrix elements.
- [31] Masubuchi, T., Benitez, J., Bell, A. S., Argyropoulos, S., Arnold, H., Amaral Coutinho, Y., Sanchez Pineda, A. R., Buzatu, A., Calderini, G., & Chan, Stephen Kam-wah, e. a. (2016). *Search for a Standard Model Higgs boson produced in association with a vector boson and decaying to a pair of b-quarks*. Technical Report ATL-COM-PHYS-2016-1724, CERN, Geneva.
- [32] Mellado Garcia, B., Musella, P., Grazzini, M., & Harlander, R. (2016). CERN Report 4: Part I Standard Model Predictions.
- [33] Robson, A., Piacquadio, G., & Schopf, E. (2016). *Signal and Background Modelling Studies for the Standard Model VH,  $H \rightarrow b\bar{b}$  and Related Searches: Modelling support note for VH(bb) 2015+2016 dataset publication*. Technical Report ATL-COM-PHYS-2016-1747, CERN, Geneva. This is a support note for the VH(bb) SM publication using the 2015+2016 datasets.
- [34] Stewart, I. W. & Tackmann, F. J. (2011). Theory uncertainties for higgs and other searches using jet bins.

## References

- [18] (2016). *Monte Carlo Generators for the Production of a W or Z/ $\gamma^*$  Boson in Association with Jets at ATLAS in Run 2*. Technical Report ATL-PHYS-PUB-2016-003, CERN, Geneva.
- [19] Ahmadov, F., Alio, L., Allbrooke, B., Bristow, T., Buescher, D., Buzatu, A., Coadou, Y., Debenedetti, C., Enari, Y., Facini, G., Fisher, W., Francavilla, P., Gaycken, G., Gentil, J., Goncalo, R., Gonzalez Parra, G., Grivaz, J., Gwilliam, C., Hageboeck, S., Halladjian, G., Jackson, M., Jamin, D., Jansky, R., Kiuchi, K., Kostyukhin, V., Lohwasser, K., & Lopez Mateos, D, e. a. (2014). *Supporting Document for the Search for the bb decay of the Standard Model Higgs boson in associated (W/Z)H production with the ATLAS detector*. Technical Report ATL-COM-PHYS-2014-051, CERN, Geneva.
- [20] Buckley, A., Butterworth, J., Grellscheid, D., Hoeth, H., Lonnblad, L., Monk, J., Schulz, H., & Siegert, F. (2010). Rivet user manual.
- [21] Buzatu, A. & Wang, W. (2016). *Object selections for SM Higgs boson produced in association with a vector boson in which H-gt;bb and V decays leptonically with Run-2 data: Object support note for VH(bb) 2015+2016 dataset publication*. Technical Report ATL-COM-PHYS-

2016-1674, CERN, Geneva. This is a support note for the VH(bb) SM publication using the 2015+2016 datasets.

- [22] Catani, S., Dokshitzer, Y. L., Seymour, M. H., & Webber, B. R. (1993). Longitudinally-invariant  $k_\perp$ -clustering algorithms for hadron-hadron collisions. *Nucl. Phys. B*, 406(CERN-TH-6775-93, LU-TP-93-2), i87–224. 38 p.
- [23] Chan, S., Huth, J., Lopez Mateos, D., & Mercurio, K. (2015a). *ZH → llbb Analysis with Telescoping Jets*. Technical Report ATL-PHYS-INT-2015-002, CERN, Geneva.
- [24] Chan, S. K.-w., Lopez Mateos, D., & Huth, J. (2015b). *Micromegas Trigger Processor Algorithm Performance in Nominal, Misaligned, and Corrected Misalignment Conditions*. Technical Report ATL-COM-UPGRADE-2015-033, CERN, Geneva.
- [25] Clark, B., Lopez Mateos, D., Felt, N., Huth, J., & Oliver, J. (2014). *An Algorithm for Micromegas Segment Reconstruction in the Level-1 Trigger of the New Small Wheel*. Technical Report ATL-UPGRADE-INT-2014-001, CERN, Geneva.
- [26] Collaboration, T. A., Aad, G., Abat, E., Abdallah, J., & A A Abdelalim, e. a. (2008). The atlas experiment at the cern large hadron collider. *Journal of Instrumentation*, 3(08), S08003.
- [27] Gleisberg, T., Höche, S., Krauss, F., Schönherr, M., Schumann, S., Siegert, F., & Winter, J. (2009). Event generation with sherpa 1.1. *Journal of High Energy Physics*, 2009(02), 007.
- [28] Heinemeyer, S. & Mariotti, e. a. (2013). *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*.

Technical Report CERN-2013-004. CERN-2013-004, Geneva. Comments: 404

pages, 139 figures, to be submitted to CERN Report. Working Group web page:

<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>.

- [29] Lavesson, N. & Lonnblad, L. (2005). W+jets matrix elements and the dipole cascade.
- [30] Lonnblad, L. (2001). Correcting the colour-dipole cascade model with fixed order matrix elements.
- [31] Masubuchi, T., Benitez, J., Bell, A. S., Argyropoulos, S., Arnold, H., Amaral Coutinho, Y., Sanchez Pineda, A. R., Buzatu, A., Calderini, G., & Chan, Stephen Kam-wah, e. a. (2016). *Search for a Standard Model Higgs boson produced in association with a vector boson and decaying to a pair of b-quarks*. Technical Report ATL-COM-PHYS-2016-1724, CERN, Geneva.
- [32] Mellado Garcia, B., Musella, P., Grazzini, M., & Harlander, R. (2016). CERN Report 4: Part I Standard Model Predictions.
- [33] Robson, A., Piacquadio, G., & Schopf, E. (2016). *Signal and Background Modelling Studies for the Standard Model VH,  $H \rightarrow b\bar{b}$  and Related Searches: Modelling support note for VH(bb) 2015+2016 dataset publication*. Technical Report ATL-COM-PHYS-2016-1747, CERN, Geneva. This is a support note for the VH(bb) SM publication using the 2015+2016 datasets.
- [34] Stewart, I. W. & Tackmann, F. J. (2011). Theory uncertainties for higgs and other searches using jet bins.