

**Search for Higgs boson pair-production in the  $bb\tau\tau$  final state using proton-proton collisions at  $\sqrt{s} = 13$  TeV data with the ATLAS detector**

**Zhiyuan Li**

February 1, 2022



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theory and Motivation</b>	<b>2</b>
2.1	The Standard Model and the Higgs boson . . . . .	2
2.2	Beyond the Standard Model . . . . .	2
<b>3</b>	<b>The ATLAS experiment at the Large Hadron Collider</b>	<b>3</b>
3.1	The Large Hadron Collider . . . . .	3
3.1.1	The LHC Accelerator Complex . . . . .	3
3.1.2	Luminosity and pileup . . . . .	5
3.1.3	Operation schedule . . . . .	7
3.2	The ATLAS Detector . . . . .	9
3.2.1	Coordinate system . . . . .	11
3.2.2	Magnets . . . . .	13
3.2.3	Inner detector . . . . .	15
3.2.3.1	Pixel detector and IBL . . . . .	15
3.2.3.2	Semiconductor tracker . . . . .	16
3.2.3.3	Transition radiation tracker . . . . .	16
3.2.4	Calorimeter system . . . . .	17
3.2.4.1	Electromagnetic calorimeter . . . . .	18
3.2.4.2	Hadronic calorimeter . . . . .	19
3.2.4.3	Forward calorimeter . . . . .	19
3.2.5	Muon Spectrometer . . . . .	19
3.2.6	Trigger system . . . . .	21
<b>4</b>	<b>Physics Object Reconstruction</b>	<b>22</b>
4.1	Track and vertex reconstruction . . . . .	22
4.2	Electrons . . . . .	24
4.2.1	Reconstruction . . . . .	24
4.2.2	Identification . . . . .	26

---

4.3	Muons . . . . .	27
4.3.1	Reconstruction . . . . .	28
4.3.2	Identification . . . . .	29
4.4	Jets . . . . .	29
4.4.1	Reconstruction . . . . .	30
4.4.2	Calibration . . . . .	32
4.4.3	Identification of heavy-quark flavoured jets . . . . .	32
4.5	Hadronically decaying $\tau$ lepton . . . . .	34
4.5.1	Reconstruction . . . . .	35
4.5.2	Identification . . . . .	36
4.6	Missing transverse energy . . . . .	37
4.7	Overlap removal . . . . .	38
<b>5</b>	<b>Data and Monte Carlo Simulation</b>	<b>40</b>
5.1	Simulation of physics processes TODO: move this to theory section . . . . .	40
5.2	Data samples . . . . .	41
5.3	Simulated event samples . . . . .	42
5.3.1	Simulated signal samples . . . . .	44
5.3.2	Non-resonant signal reweighting and combination . . . . .	45
5.3.3	Simulated background samples . . . . .	47
<b>6</b>	<b>Charm jet mis-tagging calibration</b>	<b>50</b>
6.1	Calibration methods for $b$ -jet and light jet . . . . .	50
6.2	Calibration method for $c$ -jet . . . . .	51
6.3	Alternative Monte Carlo samples . . . . .	53
6.4	Kinematic Likelihood Fitter . . . . .	54
6.5	Maximising likelihood . . . . .	54
6.6	Event selection . . . . .	55
6.6.1	Standard selection . . . . .	55
6.6.2	Low- $p_T$ selection . . . . .	56
6.6.3	High- $p_T$ selection . . . . .	57
6.6.4	Combined selection . . . . .	58
6.7	Systematic uncertainties . . . . .	59
6.7.1	Experimental uncertainties . . . . .	59
6.7.2	Modelling uncertainties . . . . .	61
6.8	Under-estimation of $t\bar{t}$ + Heavy flavour background . . . . .	62
6.9	Results . . . . .	62
6.9.1	Overview . . . . .	62
6.9.2	$b$ -tagging algorithms output distribution . . . . .	63

---

6.9.3	Efficiencies and Scale Factors . . . . .	66
<b>7</b>	<b>Search for Higgs boson pair production in the <math>b\bar{b}\tau^+\tau^-</math> channel</b>	<b>71</b>
7.1	Trigger and event Selection . . . . .	72
7.1.1	Trigger selection . . . . .	72
7.1.2	Event pre-selection . . . . .	73
7.1.3	Anti- $\tau_{\text{had}}$ selection . . . . .	76
7.1.4	$Z + \text{HF}$ control region event selection . . . . .	79
7.2	Background estimation . . . . .	80
7.2.1	Fake- $\tau_{\text{had}}$ backgrounds estimation . . . . .	81
7.2.1.1	Fake factor method . . . . .	81
7.2.1.2	$t\bar{t}$ background reweighting . . . . .	83
7.2.1.3	Fake factor calculation . . . . .	84
7.2.1.4	Fake factor method validation . . . . .	90
7.3	Multivariate signal extraction . . . . .	92
7.3.1	MVA input variables . . . . .	92
7.3.2	MVA hyper parameters optimisation and training . . . . .	96
7.3.3	MVA output distributions . . . . .	98
7.4	Systematic uncertainties . . . . .	101
7.4.1	Experimental uncertainties . . . . .	101
7.4.2	Background modelling uncertainties for MC-based processes . . . . .	102
7.4.2.1	Uncertainties on $t\bar{t}$ . . . . .	105
7.4.2.2	Uncertainties on $Z + \text{HF}$ background . . . . .	109
7.4.2.3	Uncertainties on single top . . . . .	111
7.4.2.4	Uncertainties on single-Higgs . . . . .	116
7.4.2.5	Uncertainties on other minor backgrounds . . . . .	116
7.4.3	Signal uncertainties . . . . .	117
7.4.3.1	Resonant signals . . . . .	117
7.4.3.2	Non-resonant signal . . . . .	118
7.4.3.3	Uncertainties on $\kappa_\lambda$ scan . . . . .	120
7.4.4	Uncertainties on fake background . . . . .	122
7.4.5	Summary of Systematics Uncertainties . . . . .	123
7.5	Results . . . . .	125
7.5.1	Fit setup . . . . .	125
7.5.2	$\tau_{\text{lep}}\tau_{\text{had}}$ channel results . . . . .	127
7.5.2.1	Post-fit background events yields . . . . .	127
7.5.2.2	Post-fit MVA input and output distributions . . . . .	127
7.5.2.3	Limit results: $\tau_{\text{lep}}\tau_{\text{had}}$ only fit . . . . .	134
7.5.3	Combination results of $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$ . . . . .	136

---

7.5.4	Combination results with other $HH$ decay channels . . . . .	138
7.5.4.1	Combination results of $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$ . . . . .	138
7.5.4.2	Combination results of $b\bar{b}\tau^+\tau^-$ , $b\bar{b}\gamma\gamma$ and $b\bar{b}b\bar{b}$ . . . . .	140
<b>8</b>	<b>Summary</b>	<b>141</b>
<b>A</b>	<b>Supplementary material for <math>c</math>-jet calibration</b>	<b>155</b>
A.1	Additional plots for kinematic variables . . . . .	155
A.1.1	Standard selection . . . . .	155
A.1.2	Low- $p_T$ selection . . . . .	161
A.2	High- $p_T$ selection . . . . .	164
A.3	Combined selection . . . . .	171
A.4	Experimental uncertainties . . . . .	178
<b>B</b>	<b>Supplementary material for <math>HH</math> searches</b>	<b>179</b>
B.1	Additional plots for fake-background estimation . . . . .	179
B.2	Additional material for MVA signal extraction . . . . .	179
B.3	Additional material for systematics uncertainties . . . . .	179
B.4	Additional material for results . . . . .	200

# **Chapter 1**

## **Introduction**

## **Chapter 2**

# **Theory and Motivation**

**2.1 The Standard Model and the Higgs boson**

**2.2 Beyond the Standard Model**

## Chapter 3

# The ATLAS experiment at the Large Hadron Collider

### 3.1 The Large Hadron Collider

The Large Hadron Collider [1] is the world's largest and most powerful particle accelerator. It started operation in 2008 and retains its crucial role in the many accelerators in the world. The collider is located in a ring tunnel of circumference 26.7 km, which lies beneath the French-Swiss border near Geneva, with superconducting magnets along the tunnel to keep the particle beam in direction and a large number of accelerating structures to boost the beam to the desired energy. Inside the tunnel, *bunches* of up to  $1.15 \times 10^{11}$  protons travelling at close to the speed of light in opposite direction are collided 40 million times per second at four cross points, around which are positioned four main detectors: ATLAS (A Toroidal LHC ApparatuS) [2], CMS [3] (Compact Muon Solenoid), ALICE (A Large Ion Collider Experiment) [4] and LHCb (b stands for beauty) [5].

#### 3.1.1 The LHC Accelerator Complex

The proton beams colliding in the LHC are designed to carry energy of the order of a few TeV. To reach such high energy, a series of acceleration is required for the beams before entering the LHC ring. The protons are supplied by the injector chain Linac 2 — Proton Synchrotron Booster (PSB) — Proton Synchrotron (PS) — Super Proton Synchrotron (SPS), as shown in Figure 3.1.

The protons are produced by stripping off the electrons from hydrogen gas in an electric field. Linac 2, the first accelerator in the chain, is a linear accelerator which accelerates the protons to an energy of 50 MeV. They then enter the PSB, which accelerates the protons to 1.4 GeV, followed by the PS, where they reach 25 GeV. The series of radio frequency cavities in the PS splits the beam into discrete bunches of protons of 25 ns spacing. These

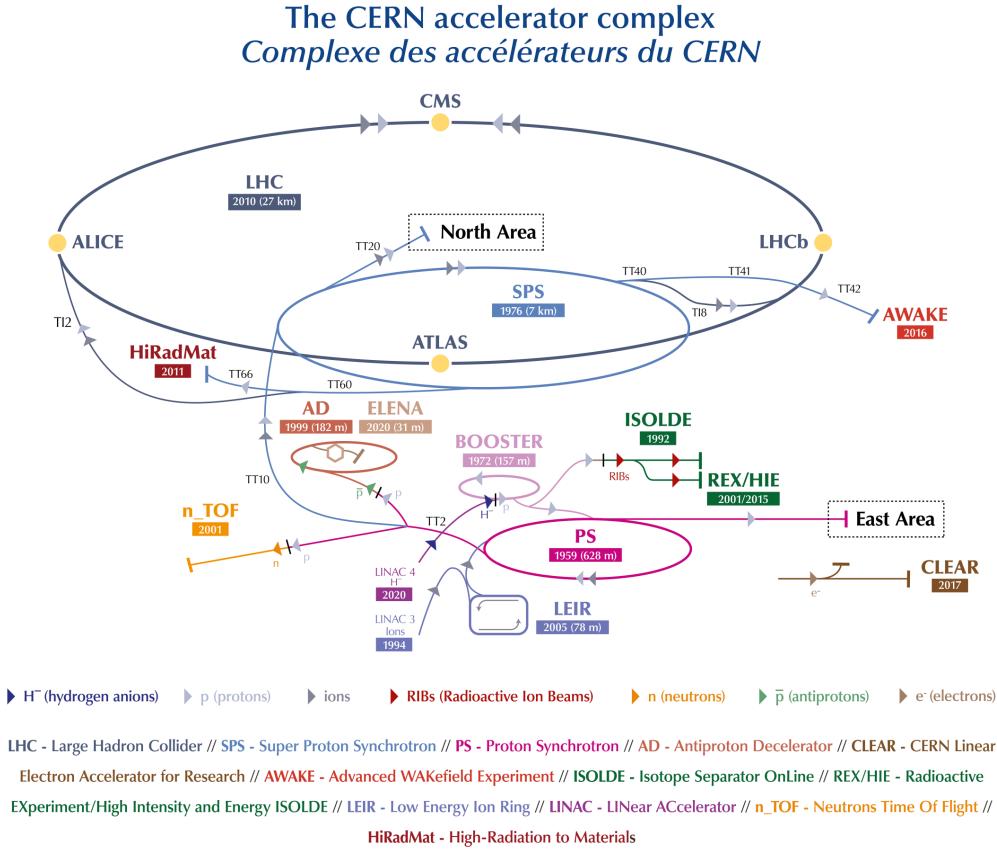


Figure 3.1: The LHC is the last ring (dark blue line) in a complex chain of particle accelerators. The smaller machines are used in a chain to help boost the particles to their final energies and provide beams to a whole set of smaller experiments.

bunches are then accelerated to 450 GeV in the SPS, from which they are finally injected into the beam pipes of the LHC. Each proton beam contains 2808 bunches, arranged in “trains” with 72 bunches each “carriage” of a gap of around 320 ns between each carriage. The beams are required to have well defined transverse and longitudinal emittance.

The beam pipes are kept at ultra-high vacuum, a vacuum thinner than interstellar void, maintained for 24 km of low-temperature section and 3 km of room-temperature section. For the low temperature section, the vacuum is achieved by pumping in 9000 m<sup>3</sup> of cryogenic gas, which later will be condensed and adhered to the surface of the beampipe. For the room temperature section, the vacuum is achieved by use of non-evaporable getter (NEG) that absorbs residue gas particles when heated. More residue is absorbed by an ion pumper.

The beam pipes are installed in the existing tunnel that was constructed between 1984 and 1989 for the Large Electron-Positron Collider (LEP) [6], which lies between 45 m and 170 m below the surface on a plane inclined at 1.4% sloping towards the Léman lake. There are advantages and disadvantages for a proton-proton collider, compared to the

electron-positron collider like LEP or proton-anti-proton collider. For the LHC, two rings are needed to accommodate the two counter-rotating beams, unlike particle-antiparticle colliders that can have both beams sharing the same phase space in a single ring, and therefore sharing the same magnets, same radio frequency cavities, etc. However, LHC is able to achieve very high collision energy which is not possible using electron-positron colliders, neither linear nor circular ones. This is due to the fact that in a circular electron collider, synchrotron radiation lost is proportional to the Lorentz factor  $\gamma = E/m$  to the power of four, where  $E$  and  $m$  are the energy and mass of the particle, respectively. Since electrons are about 2000 times lighter than protons, synchrotron radiation lost is at the order of  $10^{13}$  faster for electrons than for protons. For linear electron-positron collider, an extremely long acceleration section is required with current technologies, which makes it an impractical option. As for the proton-anti-proton collider, it would not be possible to achieve such high luminosity using anti-proton beams, since it is much more difficult to produce anti-protons than to produce protons. In addition, at high energies the proton anti-proton collider starts losing one of its advantages of having higher cross section, which is due to the quark sea and anti-quarks in protons become more “visible” at high energies (more details in the parton model section).

As explained above, two separate rings are required to accomodate the two beam pipes, while the internal diameter of the tunnel is only about 3.8 m. It's technically challenging to install them in such small space. LHC therefore adopted the twin-bore magnet design [7], as shown in Figure 3.2. It was first proposed by John Blewett at the Brookhaven laboratory in 1971 due to cost considerations [8], but in the case of the LHC the overriding reason for adopting this solution is the lack of space in the tunnel.

### 3.1.2 Luminosity and pileup

Luminosity is an important measure of a collider's performance. It comes closely to the number of events generated per second, given by:

$$N = \mathcal{L}\sigma,$$

where  $\sigma$  is the scattering cross section for the event under study and  $\mathcal{L}$  is the machine luminosity. For the cross-section, it is more common to use barn as the unit, where  $1\text{b} = 10^{-28} \text{ m}^2 = 10^{-24} \text{ cm}^2$ , since particle interactions usually have very small cross-sections. The machine luminosity depends on the beam parameters and can be written for a Gaussian beam distribution as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F,$$

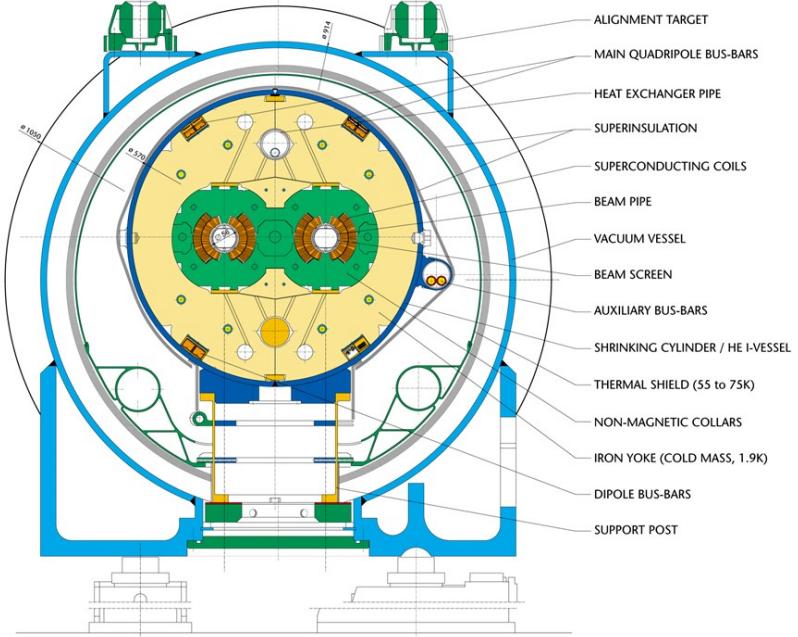


Figure 3.2: Double-bore magnet configuration of the LHC superconducting magnets [7].

where  $N_b$  refers to the number of protons per bunch,  $n_b$  is number of bunches per beam,  $f_{rev}$  is the revolution frequency,  $\gamma_r$  is the relativistic gamma factor,  $\epsilon_r$  is the normalised transverse beam emittance,  $\beta^*$  is the beta function at the collision point which describes the size of the beam, and  $F$  refers to the geometric luminosity reduction factor due to the crossing angle at the interaction point. While the instantaneous luminosity measures the rate of collisions (per unit cross-section), the total number of collisions (per unit cross-section) is measured by the integrated luminosity  $L$ , given by:

$$L = \int \mathcal{L} dt,$$

which is the integral of the instantaneous luminosity over time.

The two general purpose experiments, ATLAS and CMS are both aiming at a peak luminosity of  $\mathcal{L} = 10^{34} \text{ cm}^2\text{s}^{-1}$  for proton proton collision, which corresponds to about one billion  $pp$  collisions per second. The instantaneous luminosity was much improved at real-time operations reaching about twice the nominal (from 2015 to 2018) thanks to the effort of the LHC experts.

Another important parameter for the LHC experiments is the pileup, which is a measure of the number of inelastic  $pp$  interactions that occur per bunch crossing. Higher pileup gives more luminosity (for a fixed number of bunches) but makes physics analysis more difficult due to the signals in the detector from the additional interactions. The distribution of the recorded luminosity over the pileup is shown in Figure 3.3 for operations from

2015 to 2018 (Run 2), where the  $\langle\mu\rangle$  stands for the mean number of interactions per bunch crossing. There are two main sources of pileup: in-time pileup and out-of-time pileup. The former refers to the additional proton-proton collisions occurring in the same bunch-crossing as the collision of interest. The latter refers to the additional proton-proton collisions occurring in bunch-crossings just before and after the collision of interest.

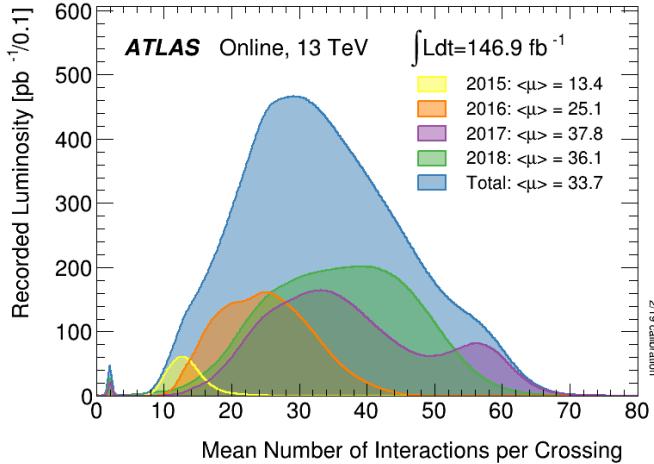


Figure 3.3: Shown is the luminosity-weighted distribution of the mean number of interactions per crossing for the data collected by the ATLAS from 2015 to 2018. All data recorded by ATLAS during stable beams is shown, and the integrated luminosity and the mean  $\mu$  value are given in the figure.

### 3.1.3 Operation schedule

The LHC operation and shutdown schedule is shown in Figure 3.4. Following the downtime after an incident in one of the main dipole circuits during the first commissioning in 2008 [9], the operation restarted at lower beam energy to minimise the risk. Therefore, the first proton run (2010–2013) [10] was carried out at 3.5–4 TeV per beam (centre of mass energy 7–8 TeV). Furthermore, a bunch spacing of 50 ns was used instead of the nominal 25 ns, with peak luminosity of  $0.8 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , and pileup larger than nominal.

In Run 1, the LHC delivered about  $30 \text{ fb}^{-1}$  of proton data and important physics results, most notably the discovery of the Higgs boson [11, 12]. Run 1 was followed by a long shutdown (LS1, 2013–2014) with a large number of consolidation and upgrade activities [13]. The bus-bar splices between the superconducting magnets were improved, in order to make sure that the LHC could operate at higher energy without risk of repeating the 2008 incident.

Run 2 (2016–2018) was carried out at 6.5 TeV per beam (center of mass energy 13 TeV) [14]. As shown in Figure 3.5, out of the  $156 \text{ fb}^{-1}$  of data LHC has delivered at 13 TeV centre of mass energy, the ATLAS detector has recorded  $147 \text{ fb}^{-1}$  and  $139 \text{ fb}^{-1}$  of data is



Figure 3.4: LHC operation schedule and luminosity targets.

certified to be good quality data. The  $156 \text{ fb}^{-1}$  data accounts for the luminosity delivered from the start of stable beams until the LHC requests ATLAS to put the detector in a safe standby mode to allow a beam dump or beam studies. The recorded luminosity is slightly smaller than the delivered luminosity, due to the inefficiency of the Data Acquisition and the so called “warm start”: when the stable beam flag is raised, the tracking detectors undergo a ramp of the high-voltage and, for the pixel system, turning on the pre-amplifiers. More details of the ATLAS detector can be found in the following sections. The recorded data is checked carefully to exclude possible hardware or software issues. This is achieved by monitoring detector-level quantities and reconstructed collision event characteristics at key stages of the data processing chain. This procedure led to high efficiency of good quality data: 95.6% [15].

In this thesis the  $139 \text{ fb}^{-1}$  data recorded by the ATLAS detector of Run 2 is used. The nominal bunch spacing of 25 ns was used, with slightly less bunches (2500) per beam. The LHC experts have continually improved the running scenario to increase the luminosity, and during Run 2 the luminosity surpassed the designed luminosity by a factor of 2. As well as improving the instantaneous luminosity, the availability of the machine was dramatically improved during Run 2 which is an important factor enabling the high efficiency of good quality data as mentioned above. During Run 2, the machine was providing physics collisions during 50% of the allocated physics time, which is very impressive for a superconducting collider.

The operation of CERN’s accelerators is subject to scheduled shutdowns to allow important repair and upgrade work to take place. The present shutdown, LS2, is devoted to

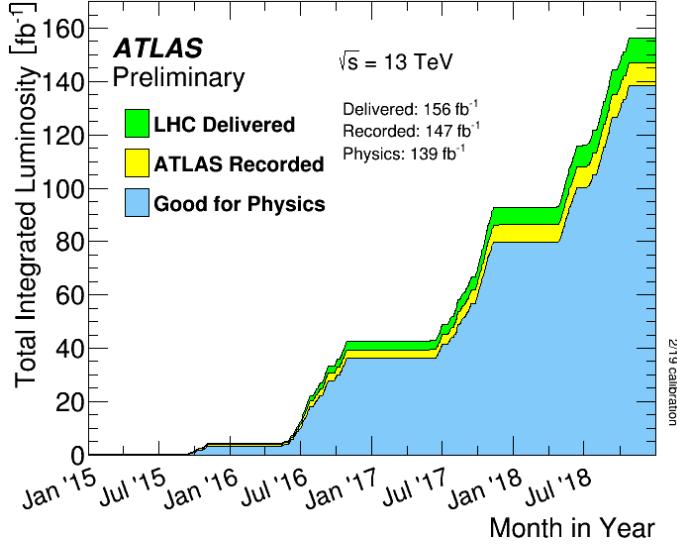


Figure 3.5: Cumulative luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for  $pp$  collisions at 13 TeV centre-of-mass energy in Run 2.

preparations for Run 3 of the LHC, which will have an integrated luminosity equal to the two previous runs combined, and for the High-Luminosity LHC (HL-LHC), the successor to the LHC, which will begin operation at the end of 2027, and eventually generate 10 times the integrated luminosity of all Run 1, 2 and 3 combined!

The LS2 schedule has had to be modified due to the COVID-19 pandemic, which the new schedule anticipates that the first test beams will circulate in the LHC at the end of September 2021, four months later than the date planned before the COVID-19 crisis, to give the LHC’s main experiments – ATLAS, CMS, ALICE and LHCb – time to prepare their own upgrade. Run 3 of the LHC will begin at the start of March 2022. The third long shutdown (LS3) will begin at the start of 2025 and end in mid-2027. This is when the equipment for the HL-LHC and its experiments will be installed.

## 3.2 The ATLAS Detector

ATLAS is one of the two general purpose detectors built for probing proton-proton collision. This detector represents the work of a large collaboration of several thousand physicists, engineers, technicians, and students over a period of fifteen years of dedicated design, development, fabrication, and installation. The overall layout of the detector is shown in Figure 3.6 [2]. It has the shape of a cylinder, 46 m long, 25 m in diameter, and sits in a cavern 100 m below ground. The ATLAS detector weighs 7000 tonnes, similar to the weight of the Eiffel Tower. The detector itself is a many-layered instrument designed to detect some of most energetic particles ever created on earth. It consists of six different de-

tecting subsystems wrapped concentrically in layers around the collision point of nearly  $4\pi$  solid angle coverage to record the trajectory, momentum, and energy of particles, allowing them to be individually identified and measured. These six subsystems are the pixel detector [16], the semiconductor tracker (SCT) [17], the transition radiation tracker (TRT) [17], the electromagnetic (EM) calorimeter [18], the hadronic calorimeter [19] and the muon spectrometer (MS) [20]. The first three sub-detectors are collectively known as the inner detector (ID), described in section 3.2.3, and it is used for tracking charged particles. The electromagnetic and the hadronic calorimeter, described in section 3.2.4, are responsible for measuring the energies of the electromagnetic and hadronic particles respectively. The MS, described in section 3.2.5, is a unique sub-detector used for measuring the momentum of muons leaving the calorimeters.

A huge magnet system bends the paths of the charged particles so that their momenta can be measured as precisely as possible.

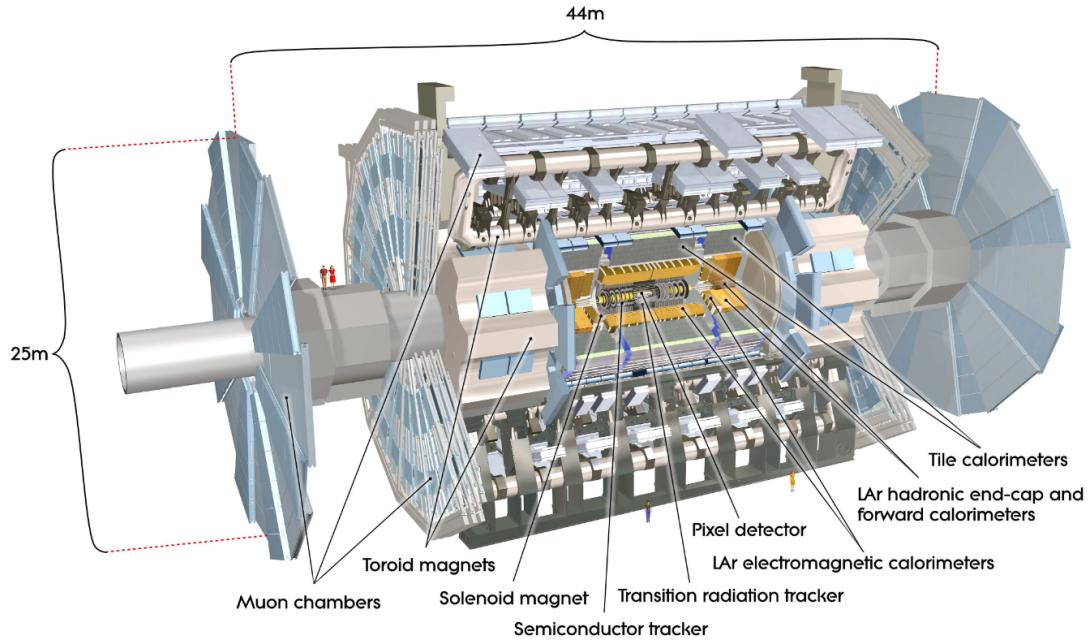


Figure 3.6: Cut-away view of the ATLAS detector. The dimensions of the detector are 25 m in height and 44 m in length. The overall weight of the detector is approximately 7000 tonnes. The figure is taken from reference [2].

The high interaction rates, radiation doses, particle multiplicities and energies, as well as the requirements for precision measurements have set stringent standards on the design of the ATLAS detector. Therefore the ATLAS detector is designed to fulfil the following requirements:

- Fast, radiation-resistant electronics and sensor elements and high detector granularity.

This is due to the high frequency of collisions, high particle fluxes and high radiation environment of the detector.

- Large acceptance in polar angle with almost full azimuthal angle coverage, due to the geometry of the detector (more details in section 3.2.1).
- Good energy resolution calorimetry, as required to enable accurate physical object reconstruction. The high resolution of energy can be obtained with very good electromagnetic calorimetry for electron and photon identification and measurements, complemented by full-coverage hadronic calorimetry for accurate jet and missing transverse energy measurements.
- Tracking of precision in the ID, as required to provide high momentum resolution and to allow the reconstruction of secondary vertices to identify  $b$ -hadrons and  $\tau$ -leptons.
- Good muon identification and momentum resolution over a wide range of momenta and the ability to determine unambiguously the charge of high- $p_T$  muons in the muon spectrometer.
- Trigger system with high efficiency on low  $p_T$  objects with sufficient background rejection, which is a prerequisite to achieve an acceptable trigger rate for most physics processes of interest.

The main performance goals of the detector are listed in Table 3.1.

Detector component	Required resolution	$\eta$ coverage	
		Measurement	Trigger
Tracking	$\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$	$\pm 2.5$	None
EM calorimetry	$\sigma_E/E = 10\% /E \oplus 0.7\%$	$\pm 3.2$	$\pm 2.5$
Hadronic calorimetry barrel and end-cap forward	$\sigma_E/E = 50\% /E \oplus 3\%$	$\pm 3.2$	$\pm 3.2$
	$\sigma_E/E = 100\% /E \oplus 10\%$	$3.1 <  \eta  < 4.9$	$3.1 <  \eta  < 4.9$
Muon spectrometer	$\sigma_{p_T}/p_T = 10\% \text{ at } p_T = 1 \text{ TeV}$	$\pm 2.7$	$\pm 2.4$

Table 3.1: General performance goals of the ATLAS detector. The units for energy of the particle,  $E$  and transverse momentum,  $p_T$  (detailed definition in section 3.2.1) are in GeV [2]. Note that, for high- $p_T$  muons, the muon-spectrometer performance is independent of the inner-detector system.

### 3.2.1 Coordinate system

The ATLAS coordinate system is a right-handed Cartesian system with the nominal interaction point defined as the origin of the coordinate system, while the beam direction defines the  $z$ -axis and the  $x$ - $y$  plane is transverse to the beam direction. The positive  $x$ -axis

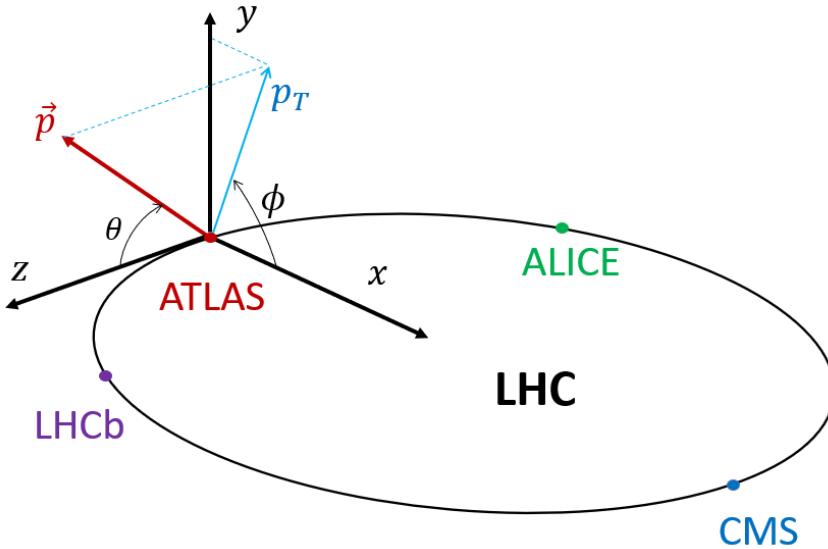


Figure 3.7: Illustration of the coordinate system used at the ATLAS experiment in the geographical context of the LHC.

is defined as pointing from the interaction point to the centre of the LHC ring and the positive  $y$ -axis is defined as pointing upwards, as shown in Figure 3.7.

The azimuthal angle  $\phi$  is measured as usual around the beam axis, and the polar angle  $\theta$  is the angle from the beam axis. In high energy physics, it's more common to use the pseudorapidity instead of the polar angle  $\theta$ , defined as:

$$\eta = -\ln \tan(\theta/2).$$

In the case of highly relativistic particles (which is the common case in high energy physics), the pseudorapidity approaches the rapidity,

$$y = 1/2 \ln[(E + p_z)/(E - p_z)],$$

where  $E$  is the energy of the particle,  $m$  is its mass and  $p_z$  is the momentum along the  $z$ -axis. There are two main reasons for using pseudorapidity rather than the polar angle  $\theta$  nor the rapidity. The reason for using pseudorapidity but not  $\theta$  is that the rapidity is invariant under Lorentz transformation, while capturing the characteristic of the particle direction of travel:  $y \rightarrow \pm\infty$  when the particle is travelling close to the beam pipe (positive for along the beam pipe, negative for the opposite direction) and  $y \rightarrow 0$  when  $p_z$  is small. The reason for using pseudorapidity but not rapidity is that due to the limited angle coverage of the detector, it's usually hard to determine the total energy and the momentum along the  $z$ -axis, especially when the direction of the particles are close to the beam pipe. While the pseudorapidity is determined only by the polar angle, which is much easier and faster

to compute. Another commonly used variable, transverse momentum  $p_T$ , is defined as the momentum of a particle transverse to the beam direction ( $z$ -direction):

$$\vec{p}_T = (p_x, p_y).$$

The reason for using transverse momentum is that, because the partons that make up a proton share the momentum, the initial longitudinal momentum is unknown; we do know, however, that the initial transverse momentum was zero. And hence we can look for the missing transverse momentum, defined as

$$\vec{E}_T^{\text{miss}} = - \sum_i \vec{p}_{T,i}$$

for visible particles  $i$ , where  $E_T^{\text{miss}}$  is the magnitude of  $\vec{E}_T^{\text{miss}}$  (Confusingly  $E_T^{\text{miss}}$  is commonly called missing transverse energy or MET. Missing transverse energy is equivalent to missing transverse momentum only if the missing particle(s) were massless.). Finally, the distance  $\Delta R$  in the pseudorapidity-azimuthal angle space is defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2},$$

where  $R$  is the radial distance from the particle position to the interaction vertex.

### 3.2.2 Magnets

ATLAS has a unique hybrid system of four large superconducting magnets [21]. This magnetic system is 22 m in diameter and 26 m in length, with a stored energy of 1.6 GJ. Figure 3.8 shows the real scale of the magnets system compared to a person. Figure 3.6 shows the general layout, the four main layers of detectors and the four superconducting magnets which provide the magnetic field over a volume of approximately 12000 m<sup>3</sup>.

The spatial arrangement of the coil windings is shown in Figure 3.9. The ATLAS magnet system consists of two parts:

- a solenoid, which is aligned on the beam axis and provides a 2 T axial magnetic field in the  $z$ -direction for the ID. Because the magnet is located in front of the EM calorimeter, it is imperative to minimise possible interactions between the magnet and the particles being studied. This is achieved by embedding over 9 km of niobium-titanium superconductor wires into strengthened, pure aluminum strips, which is capable to provide such powerful magnetic field in just 4.5 cm thickness.
- A barrel toroid and two end-cap toroids, which produce a toroidal magnetic field of approximately 0.5 T and 1 T for the muon detectors in the central and end-cap regions,



Figure 3.8: A picture showing the real size magnets compared to a person.

respectively. The barrel toroid generates the magnetic field in the central zone of the muon spectrometer, along the tangential direction of the circumferences centered on the  $z$ -axis ( $\phi$  direction). The end-cap toroids are two smaller toroids designed to provide the magnetic field in the forward areas of the muon spectrometer. This magnet configuration provides a field mostly orthogonal to the muon trajectories.

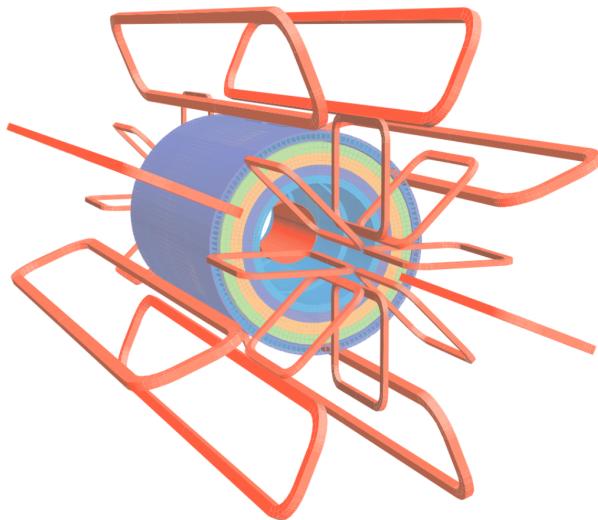


Figure 3.9: Geometry of magnet windings and tile calorimeter steel. Image taken from [2].

### 3.2.3 Inner detector

The inner detector (ID) is the closest sub-detector to the beam line, designed to track the early trajectories of charged particles for momentum calculations and locate their primary and secondary vertices with extremely high precision. The ID is required to deal with large numbers of tracks promptly, where 1000 particle collisions are taking place every 25 ns. The ID has full coverage in azimuthal angle  $\phi$  and  $|\eta| < 2.5$  acceptance in pseudorapidity. As mentioned briefly above, it consists of three parts: the pixel detector and the insertable B-Layer (IBL) [22] (as one part), the semiconductor tracker and the transition radiation tracker. The layout of the ID is shown in figure 3.10, with a charge track (in red) traversing the sensors and structural elements.

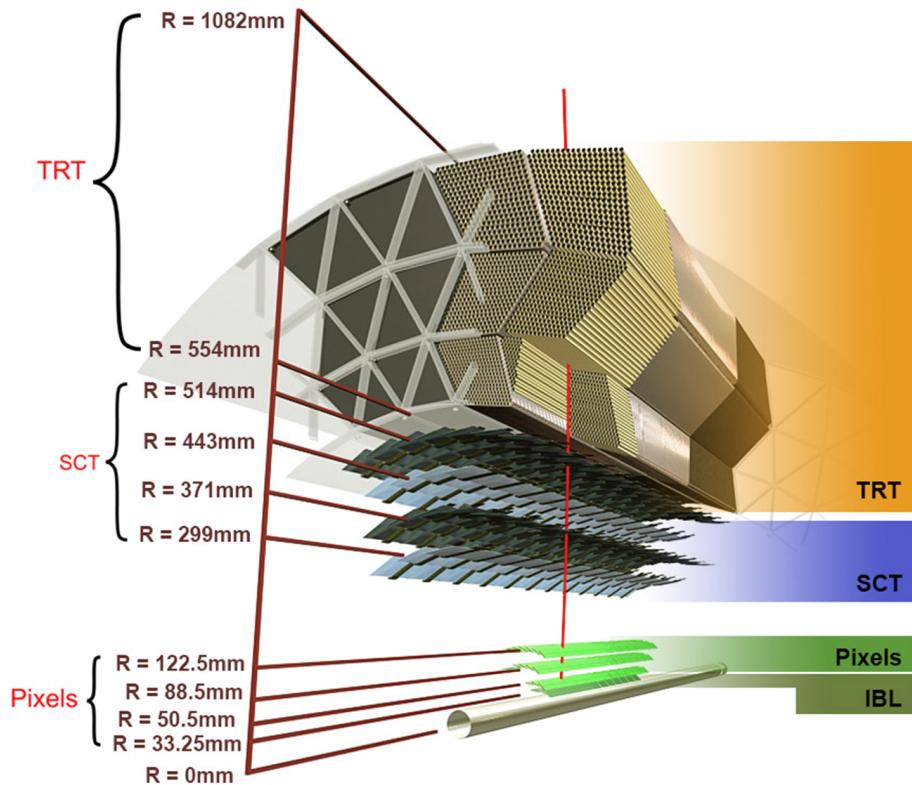


Figure 3.10: Cut-away view of the inner detector. Image taken from [23].

#### 3.2.3.1 Pixel detector and IBL

The silicon pixel detector is the closest ATLAS component to the collision. It is composed of layers of silicon pixels and designed to have a very high granularity for reconstructing primary and secondary interaction vertices. The detector layers are formed of silicon sensor modules and in total there are approximately 92 million pixels (consequently, 92 million

readout channels) in the system. It consists of three cylindrical layers in the barrel region, positioned at the radial distances of 50.5, 88.5 and 122.5 mm, and of disks perpendicular to the beams in the end-caps at the longitudinal distances of 49.5, 58.0 and 65.0 mm. In 2014, during the first LHC long shutdown, a fourth pixel layer was installed inside the existing detector, the insertable B-Layer (IBL) at a radius of 33 mm from the beam axis. The new pixel layer provides an additional space point very close to the interaction point, which significantly improves the identification of jets coming from  $b$ -quark hadronisation ( $b$ -jets). Particles with  $|\eta| < 2.5$  traverses the four layers usually produce four space-points. The pixel detector provides a resolution of  $\sigma_\phi = 10 \mu\text{m}$  in the bending direction ( $R - \phi$ ), and a resolution of  $\sigma = 115 \mu\text{m}$  in the  $z(R)$  direction in the barrel (end-cap) region.

### 3.2.3.2 Semiconductor tracker

The next constituent of the inner detector is the SCT. It is a silicon microstrip detector with over six million readout channels, which surrounds the pixel detector and covers the region of radius between 299 mm and 560 mm. It consists of four layers of strips located axially on the beam direction in the barrel region and placed along the  $z$ -direction in the end-cap region. This configuration allows the particles along the beam pipe to be constructed. Each layer of strips is glued back to back with an angle of 40 mrad to form a two-sided module and make possible the measurement of the second coordinate. The sensors are 285  $\mu\text{m}$  thick and are constructed of high-resistivity n-type bulk silicon with p-type implants. Readout strips are positioned every 80  $\mu\text{m}$ , providing a spatial resolution of  $\sigma_\phi = 17 \mu\text{m}$  in the bending direction ( $R - \phi$ ) and  $\sigma_\phi = 580 \mu\text{m}$  in the  $z$  (barrel) and  $R$  (end-cap) direction.

### 3.2.3.3 Transition radiation tracker

The outermost part of the inner detector is the TRT, which covers the radial region between 563 mm and 1066 mm. It is a straw drift tube tracker, which consists of modules of 4 mm diameter polyimide straws, filled with a mixture of gas of 70% Xe, 27% CO<sub>2</sub> and 3% O<sub>2</sub> and a gold-plated tungsten wire in the centre. The straws are interleaved with propylene fibres (foils) in the barrel (end-cap) region. With a spatial resolution of  $\sigma_\phi = 130 \mu\text{m}$ , the TRT measures the track position only in the bending direction ( $R - \phi$ ). This is because when a charged particle passes through a straw tube, electrons from the gas are liberated through ionisation; under high voltage, these electrons then drift toward the wire in the centre, where a current flow is created and registered as a hit. Since a hit can happen on any location along the wire, the information of the  $z$  position of the particle is lost. In addition, the TRT provides capability of distinguishing electrons from other charged particles. When a highly-relativistic charged particle traverses the polymer straws interface, the particle emits transition radiation which is then absorbed by the Xeon gas. The intensity of the

radiation depends on the gamma factor of the particle (strongest for lighter particles), hence this information can be exploited for electron identification.

### 3.2.4 Calorimeter system

Calorimeters are used to measure the energy of both charged and neutral particles. The ATLAS calorimeters [24], as shown in Figure 3.11, are consists of three major components, the Electromagnetic calorimeter, the Hadronic calorimeter and the Forward calorimeter (FCal). The fine granularity of the EM calorimeter is ideal for precision measurements of electrons and photons; the coarser granularity of hadronic calorimeter is sufficient for the hadronic jet reconstruction; the FCal provides coverage of large pseudorapidity region: these calorimeters cover the range  $|\eta| < 4.9$ . All the three calorimeters are sampling calorimeters. Sampling calorimeters use different material the absorber and the active part: the absorber (or passive material) is responsible for producing particle showers where the active part then measures their energy. Note well the a fraction of total particles energy deposited in the passive material and it's not measured; the overall energy must be deduced from the definite measurements taken in the active detector layers.

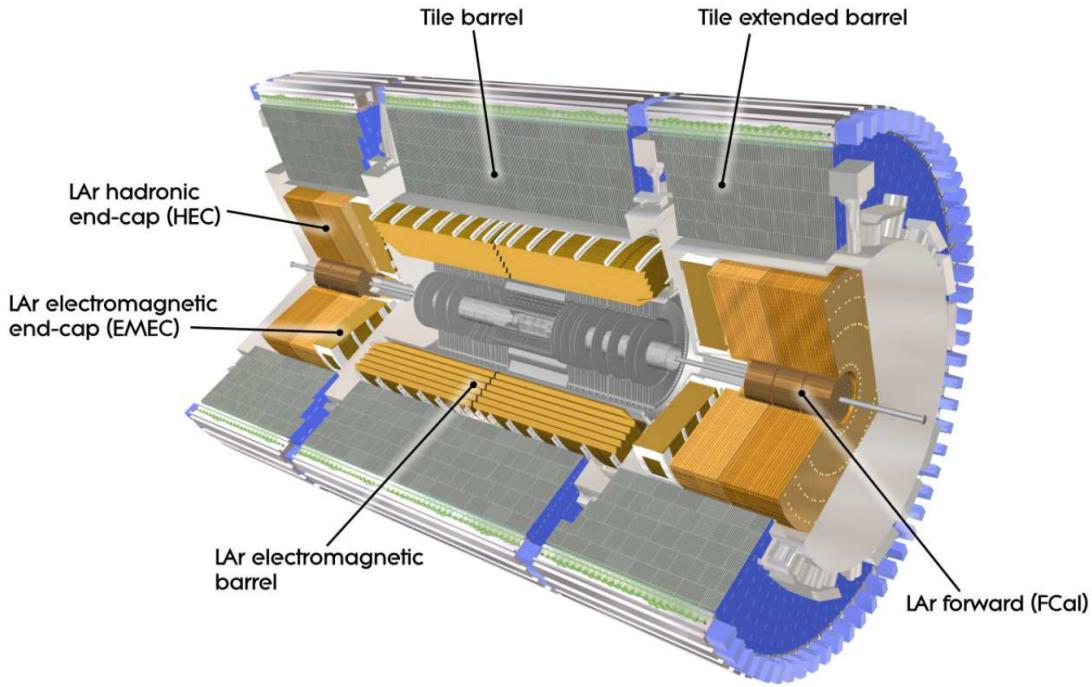


Figure 3.11: Cut-away view of the ATLAS calorimeter system. Image taken from [24].

The electromagnetic and hadronic showers must be contained in the calorimeter to ensure precise measurement of the total energy of the particle and to avoid punch-through into the muon system. The calorimeter depth is hence an important design consideration.

The thickness of the calorimeter is measured in radiation length  $X_0$ , which is the mean length of a material over which an electron will lose all but  $1/e$  of its initial energy through radiative processes; and nuclear interaction length  $\lambda$ , which is the mean distance travelled by a hadronic particle before undergoing an inelastic nuclear interaction. The total thickness of the EM calorimeter is greater than 22 radiation lengths ( $X_0$ ) in the barrel and greater than  $24 X_0$  in the end-caps. The total thickness of the calorimeters, including  $1.3 \lambda$  from the outer support, is  $11 \lambda$  at  $\eta = 0$  and has been shown both by measurements and simulations to be sufficient to reduce punch-through well below the irreducible level of prompt or decay muons.

### 3.2.4.1 Electromagnetic calorimeter

The EM calorimeter is a lead-Liquid argon (lead-LAr) detector [18] with accordion-shaped (as shown in Figure 3.12) kapton electrodes and lead absorber plates over its full coverage, while using liquid argon (LAr) as the active material. The lead thickness in the absorber plates has been optimised as a function of  $\eta$  in terms of EM calorimeter performance in energy resolution. The EM calorimeter is divided into a barrel part ( $|\eta| < 1.475$ ) and two end-cap components ( $1.375 < |\eta| < 3.2$ ), each housed in their own cryostat. Additional material needed to instrument and cool the detector creates a “crack” region at  $1.375 < |\eta| < 1.52$ , where the energy resolution is significantly degraded.



Figure 3.12: A figure of the accordion-shaped electrodes.

The barrel calorimeter consists of two identical half-barrels, separated by a small gap (4 mm) at  $z = 0$ . Each end-cap calorimeter is mechanically divided into two coaxial wheels: an outer wheel covering the region  $1.375 < |\eta| < 2.5$ , and an inner wheel covering the region  $2.5 < |\eta| < 3.2$ .

The calorimeter has three layers along the transverse direction: a pre-sampler with very high granularity in  $\eta$ , in order to reconstruct the neutral pions decaying to two photons and

particles which already starts showering in the inner detector. The pre-sampler is followed by longer towers of relatively high granularity, which is the major part of detecting EM showers, and responsible for measuring the  $\eta$  and  $\phi$  coordinates of the particles. The last layer detects showers generated from particles other than electrons or photons that start showering inside the EM calorimeter before leaving it.

### 3.2.4.2 Hadronic calorimeter

The Hadronic calorimeter is comprised of the Tile Hadronic calorimeter (HCAL) and the LAr hadronic end-cap calorimeter (HEC). The HCAL is placed directly outside the EM calorimeter envelope. Its barrel covers the region  $|\eta| < 1.0$ , and its two extended barrels the range  $0.8 < |\eta| < 1.7$ . It is using steel as the absorber and scintillating tiles as the active material. It is segmented in depth in three layers, approximately 1.5, 4.1 and 1.8  $\lambda$  thick for the barrel and 1.5, 2.6, and 3.3  $\lambda$  for the extended barrel. The total detector thickness at the outer edge of the tile-instrumented region is  $9.7 \lambda$  at  $\eta = 0$ .

The HEC is similar to the construction to the ECAL, using LAr as the active material, but instead of using lead it uses copper as the absorber. It consists of two independent wheels per end-cap, located directly behind the end-cap electromagnetic calorimeter and sharing the same LAr cryostats. The HEC covers the range of  $1.5 < |\eta| < 3.2$ , slightly overlapping with the forward calorimeter which will be described in the following paragraph (around  $|\eta|=3.1$ ) and the tile calorimeter ( $|\eta| < 1.7$ ). This overlap is to reduce the drop in material density at the transition between the different calorimeters.

### 3.2.4.3 Forward calorimeter

The Forward Calorimeter (FCal) covers  $3.1 < |\eta| < 4.9$  and is approximately 10 interaction lengths deep. It consists of three modules in each end-cap: the first, made of copper, is optimised for electromagnetic measurements, while the other two, made of tungsten, measure predominantly the energy of hadronic interactions. All three modules use LAr as active material. Due to high particle fluxes and energies in the forward region, the calorimeter must contain relatively long showers in the small volume allowed by design constraints, and thus must be very dense.

## 3.2.5 Muon Spectrometer

The muon spectrometer is the outermost and largest sub-detector of ATLAS. A cut-away view of the MS is illustrated in Figure 3.13. It fully covers the calorimeter system and occupies a large part of the ATLAS cavern. It is based on the magnetic deflection of muon tracks in the large superconducting air-core toroid magnets, instrumented with

separate trigger and high-precision tracking chambers. Over the range  $|\eta| < 1.4$ , magnetic bending is provided by the large barrel toroid. For  $1.6 < |\eta| < 2.7$ , muon tracks are bent by two smaller end-cap magnets inserted into both ends of the barrel toroid. Over  $1.4 < |\eta| < 1.6$ , usually referred to as the transition region, magnetic deflection is provided by a combination of barrel and end-cap fields. The configuration of magnets provides a field mostly orthogonal to the muon trajectories, hence minimising the degradation of resolution due to multiple scattering.

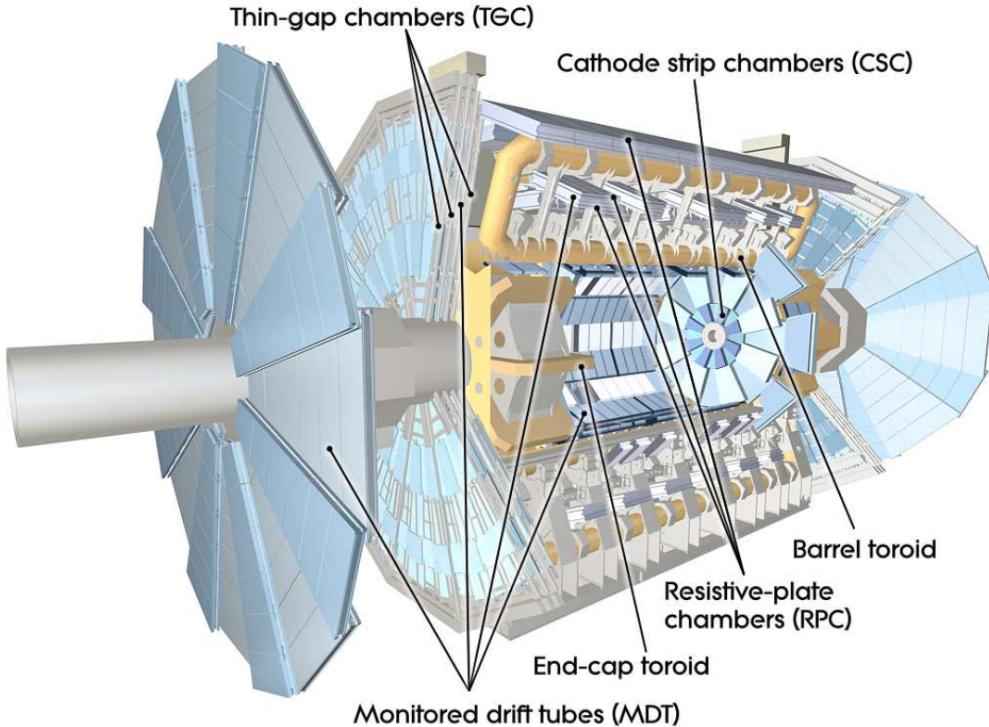


Figure 3.13: Cut-away view of the ATLAS muon spectrometer system. Image taken from [2].

The MS consists of four subsystems which rely on four different gas detector technologies. Two of them, the resistive plate chambers (RPC) in the barrel region and the thin gap chambers (TGC) in the end-cap region, provide trigger signals, while the other two, the monitored drift tubes (MDT) in the barrel and the cathode strip chambers (CSC) in the end-cap region provide the momentum measurement. The MDT chambers provide high precision measurements in the bending direction over most of the detector acceptance while the CSC are used in the forward region where the particle flux is too high for the MDT chambers. The muon chambers are arranged in the barrel ( $|\eta| < 1.05$ ) in three cylindrical layers around the beam axis, while in the end-cap regions ( $1.05 < |\eta| < 2.7$ ) they are placed in three wheels. The resolution of muons tracks momentum measurement varies from typically 2–3% over most of the kinematic range, to about 10% at  $p_T = 1$  TeV.

### 3.2.6 Trigger system

As mentioned in section 3.1.3, the spacing of each bunch is 25 ns, which translates to a 40 MHz of bunch-crossing frequency, with up to 80 collisions per bunch crossing. This is far beyond the data collection bandwidth and storage capacity of ATLAS. Therefore, it's necessary to adopts a trigger system that make fast decisions whether an event is high quality, rare or “interesting” and to save the event or not. The ATLAS trigger system consists of two consecutive parts: the Level 1 trigger [25] which is hardware-based, followed by the the software-based High Level Trigger (HLT) [26].

The L1 trigger searches for signatures from high- $p_T$  muons, electrons/photons, jets, and  $\tau$ -leptons decaying into hadrons. It also selects events with large MET and large total transverse energy. The L1 trigger uses reduced-granularity information from a subset of detectors: the RPC and TGC for high- $p_T$  muons, and all the calorimeter sub-systems for electromagnetic clusters, jets,  $\tau$ -leptons,  $E_{miss}^T$ , and large total transverse energy. As a result, the L1 trigger reduces the event rate from 40 MHz to a maximum of 100 kHz. The decision is made by Central Trigger Processor (CTP), which operates on signals from dedicated hardware in the calorimeter and muon detector systems. The decision time, at under 2.5  $\mu s$ , is faster than the ID can process events so ID information is omitted. For each data-taking period, the L1 trigger is loaded with a trigger menu, a list of up to 256 criteria used to determine whether an event is accepted. The trigger menus are designed to accomodate a broad physics programme, with high acceptance for both BSM searches and SM precision measurements. The L1 trigger also uses detector information with reduced granularity to identify Regions of Interest (RoI) [27] in  $\phi$  and  $\eta$ . The ROI information with full granularity and precision and all the available detector data (including the ID information) within the RoI's are provided to the HLT. This trigger level reduces the rate of events by two orders of magnitude, reaching an average of 1 kHz with a latency of 0.2  $\mu s$ . These events are passed on to a data storage system for offline analysis.

# Chapter 4

# Physics Object Reconstruction

Particles produced in the  $p p$  collisions inside ATLAS can interact with the detector sub-systems with each type of particle leaving a unique signature. Reconstructing and identifying these particles precisely and efficiently using the information recorded in each sub-detector is a building block of physics analysis. Therefore, the following chapter outlines the reconstruction procedure of the particles important for the analysis present in this thesis.

## 4.1 Track and vertex reconstruction

Track and vertex reconstruction is the starting point of physics objects reconstruction, which makes it crucial to understand how they are implemented in ATLAS. Track reconstruction [28, 29] is performed mainly with the so-called “inside-outside” procedure, complemented by the “outside-in” tracking and the reconstruction of TRT-standalone tracks. The inside-out stage starts by assembling the raw measurements into *clusters*: an algorithm called connected component analysis [30] groups pixels and strips in a given sensor, where the deposited energy yields a charge above threshold, with a common edge or corner into clusters.

From clusters, three-dimensional measurements, referred to as *space points*, are created (the yellow points in Figure 4.1). They represent the point where the charged particle traversed the active material of the ID. Each space point equates to one cluster in the pixel detector, while in the SCT, clusters from both sides of a strip layer must be combined to obtain a three-dimensional measurement. Three space points are combined to form track seeds (circled in blue in Figure 4.1). A combinatorial Kalman filter [31] is then used to build track candidates from the chosen seeds by incorporating additional space points from the remaining layers of the pixel and SCT detectors which are compatible with the preliminary trajectory (circled in a blue dashed line in Figure 4.1). A track

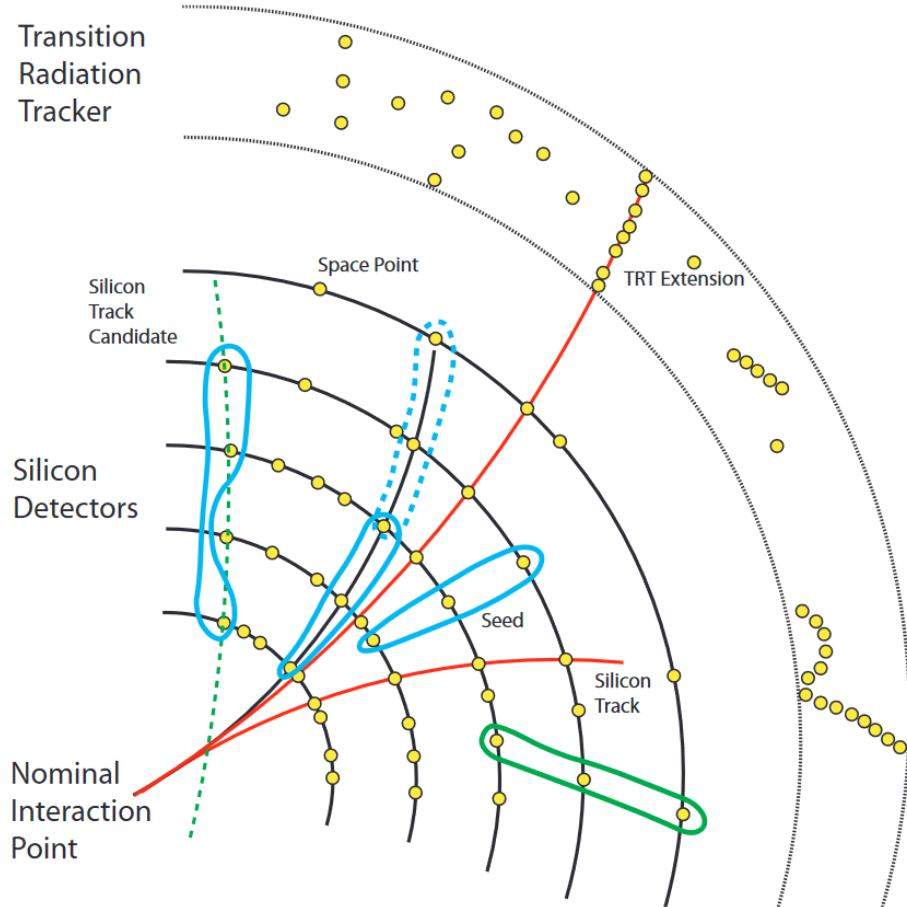


Figure 4.1: An example of track reconstruction. Image reproduced from Ref. [32].

score computed by the quantities of the fitted track is assigned to each track. The tracks candidates are processed in descending order of track score and those that fail a set of minimum requirements on  $p_T$ , the number of holes, and the number of clusters etc are rejected; candidates that have too many bad quality clusters are stripped down and re-scored, then returned to the list of remaining candidates. This process is referred to as “ambiguity solving” [29]. Finally, the tracks are extended into the TRT and, by using the full information of all three sub-detectors, the tracks are fitted again to determine the final track parameters. The track can be fully represented by five parameters measured at the perigee, which are the impact parameter (IP, transverse distance from the interaction point)  $d_0$ , the distance from the interaction point along the  $z$  axis  $z_0$ , the azimuthal angle  $\phi$ , the polar angle  $\theta$  and the charge-momentum ratio  $q/p_T$ .

The complementary “outside-in” algorithm starts with searching for tracks with segments reconstructed in the TRT, and extend the tracks inwards by adding silicon hits. The tracks are built with the combinatorial Kalman filter and passed to the ambiguity solving procedure. Finally, tracks with a TRT segment but no extension into the silicon detectors

are referred as TRT-standalone tracks.

After the tracks are reconstructed, primary vertices, which are the points where the hard-scattering processes occurred, are reconstructed in two steps [33]: a) the primary vertex finding algorithm, dedicated to associate reconstructed tracks to the vertex candidates, and b) the vertex fitting algorithm, dedicated to reconstruct the vertex position and its corresponding error matrix. It also refits the associated tracks constraining them to originate from the reconstructed interaction point. The vertex finding algorithm works as follows: first, vertex seeds are obtained from the  $z$ -position at the beamline of the reconstructed tracks; and then, an iterative  $\chi^2$  fit is then performed using the vertex seed and nearby tracks. Vertices are required to contain at least two tracks, and tracks displaced by more than  $7\sigma$  from the vertex are used to seed a new vertex. The procedure is repeated until no additional vertices can be found, and no unassociated tracks are left in the event. The primary vertex for each event is selected as the vertex with the highest  $\sum_{\text{tracks}}(p_T^{\text{track}})^2$ .

## 4.2 Electrons

When a high energy electron (or positron) enters the detector, it interacts with the detector material primarily via bremsstrahlung. This results in radiation of photons, which subsequently convert into electron–positron pairs that continue to interact with the detector material, leading to a cascade of particles of decreasing energy. These cascade particles are usually referred to as electromagnetic *shower*. They are very collimated and frequently create neighbouring signals in the calorimeter component. These interactions can occur inside the ID volume or even in the beam pipe, generating multiple tracks in the ID, or can instead occur downstream of the ID, only impacting the shower in the calorimeter.

### 4.2.1 Reconstruction

The reconstruction of electron is based on three fundamental components: a) localised clusters of energy deposits found within the EM calorimeter, b) charged tracks identified in the ID (as described in details in chapter 4.1), and c) close matching in  $\eta \times \phi$  space of the tracks to the clusters [34]. Figure 4.2 provides a schematic illustration of the elements that enter into the reconstruction and identification of an electron. The reconstruction starts from EM cluster seeding from localised energy deposits using a sliding-window algorithm [35]. The  $\eta \times \phi$  space of the EM is divided into *towers* of  $200 \times 256$  elements of size  $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ , consistent with the granularity of the second layer of the EM calorimeter. Then algorithm then “slides” a rectangular window of size  $3 \times 5$  towers whose summed transverse energy exceeds 2.5 GeV and is a local maximum to form a seed-cluster. The centre of the seed moves in steps of 0.025 in either the  $\eta$  or  $\phi$  direction

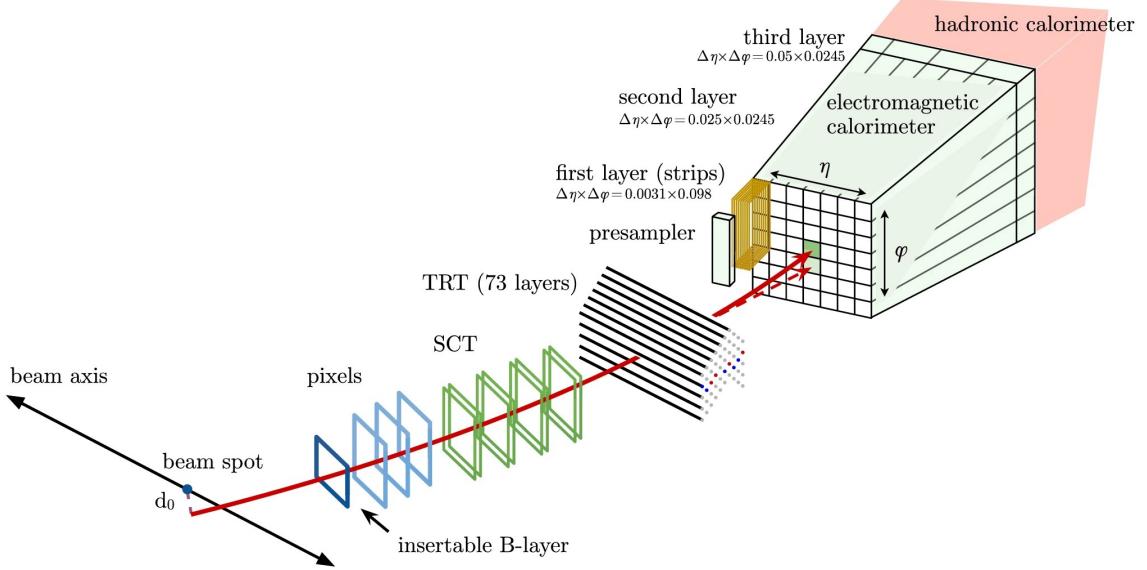


Figure 4.2: A schematic illustration of the path of an electron through the detector. The red trajectory shows the hypothetical path of an electron, which first traverses the tracking system (pixel detectors, then SCT and lastly the TRT) and then enters the electromagnetic calorimeter. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron with the material in the tracking system. Image reproduced from Ref. [34].

to search for localised energy deposits; this process is repeated until every element of the calorimeter has been covered. To better account for the energy loss of charged particles in material, a subsequent fitting procedure using an optimised Gaussian-sum filter [36] is performed on tracks which are “loosely” matched to the EM clusters. This matching requires the tracks and clusters to satisfy:

- $|\eta_{cluster} - \eta_{track}| < 0.05$ ,

and one of the two requirements:

- $-0.20 < \Delta\phi < 0.05$ , or
- $-0.10 < \Delta\phi_{res} < 0.05$ ,

where  $\Delta\phi \equiv -q \times (\phi_{cluster} - \phi_{track})$  with  $q$  being the charge of the particle, and  $\phi_{cluster}$ ,  $\phi_{track}$  and  $\eta_{cluster}$ ,  $\eta_{track}$  are the  $\phi$ ,  $\eta$  coordinates of the cluster barycentre and the position of the track extrapolated from the perigee to the second layer of the calorimeter, respectively;  $\Delta\phi_{res}$  is similar to  $\Delta\phi$  but with the momentum of the track rescaled to the energy of the cluster. The asymmetry in the condition is to account for the energy loss due to bremsstrahlung where tracks with negative (positive) electric charge bend due to the magnetic field in the positive (negative)  $\phi$  direction.

The matching of the fitted tracks to the candidate calorimeter seed-cluster is the final step of electron reconstruction. The matching requires  $-0.10 < \Delta\phi < 0.05$ , with the other

alternative requirement remaining the same. If several tracks fulfil the matching criteria, the track considered to be the primary electron track is selected using an algorithm that takes into account the distance in  $\eta$  and  $\phi$  between the extrapolated tracks and the cluster barycentres (again, measured in the second layer of the calorimeter), the number of hits in the silicon detectors and in the ID layer; a candidate with an associated track with at least four hits in the silicon layers and no association with a vertex from a photon conversion (photon converting into electron-positron pair) is considered as an electron candidate. However, if the primary candidate track can be matched to a secondary vertex and has no pixel hits, then this object is classified as a photon candidate (likely a conversion).

A further classification is performed using the candidate electron's  $E/p$  and  $p_T$ , the presence of a pixel hit, and the secondary-vertex information, to determine unambiguously whether the object is only to be considered as an electron candidate or if it should be ambiguously classified as potentially either a photon candidate or an electron candidate.

### 4.2.2 Identification

The reconstruction algorithm is very efficient in reconstructing electron, however, this is not necessarily what is needed for many ATLAS analysis, where they are interested in prompt electrons. Prompt electrons are electrons coming from the primary interaction of the event, while non-prompt electrons may come from the semileptonic decays of heavy quarks or from photon conversion. Other objects such as hadrons can be mis-reconstructed as electrons as well. This necessitates the identification of prompt electrons, making use of the differences between prompt electrons and non-prompt electrons/mis-reconstructed electrons. A multivariate likelihood technique, taking advantages of the correlations among the variables describing the differences, is employed to select prompt electrons [34]. The input variables to the likelihood describe the following characteristics: a) shower shape, b) properties of the track, c) matching of the track and clusters.

To quantify the performance of the identification, the identification efficiency is measured, which represents the probability of a true prompt electron passing the identification requirements. Different cuts are applied on the final discriminant to define *working points* (WP), which can specify the identification efficiency, as a function of  $E_T$  or  $\eta$  of the electron. Three WPs, *Loose*, *Medium* and *Tight* are defined. Each WP is utilising a separate multivariate discriminant formed from a different selection of discriminating variables, and applying a different requirement on the resulting discriminant output. An example of the electron identification efficiency is shown in Figure 4.3.

In addition, many analyses further require the electron to pass some *isolation* requirements. Isolation is built exploiting a characteristic signature of the prompt electrons that there is relatively little activity surrounding the prompt electrons, as compared to non-

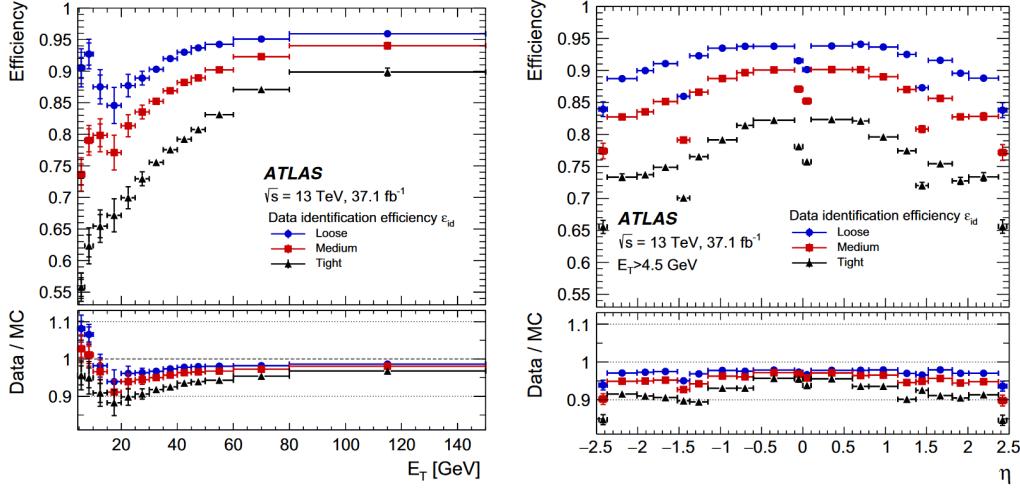


Figure 4.3: Measured electron identification efficiencies in  $Z \rightarrow ee$  events for the ‘Loose’ (blue circles), ‘Medium’ (red squares), and ‘Tight’ (black triangles) WPs as a function of  $E_T$  (left) and  $\eta$  (right). The vertical uncertainty bars (barely visible because they are small) represent the statistical (ID) and total (outer bars) uncertainties. For both plots, the bottom panel shows the data-to-simulation ratios. Reproduced from Ref. [34].

prompt electrons [37]. For the electron isolation, four WPs: *Gradient*, *HighPtCaloOnly*, *Loose* and *Tight* are defined, each targeting a fixed value of isolation efficiency or imposing fixed requirement on the isolation variables.

In the analysis presented in Chapter 7 the ‘Loose’ identification WP is used, which, in combination with additional track hit requirements, provides an electron efficiency of 95%. The electron candidates are also required to have  $p_T > 7$  GeV and  $|\eta| < 2.47$ . Further cuts are used to define signal electrons, as defined in section 7.1.2. In addition, the electron candidates are required to pass the ‘Loose’ isolation WP which has an efficiency of 99%. The isolation requirement is also inverted to provide control regions for estimating backgrounds, as described in section 7.2.1. In the FTAG calibration effort presented in Chapter 6 the ‘Medium’ identification WP is used, where the electron efficiency is roughly 85%, with additional requirements on the electron  $p_T > 27$  GeV and  $|\eta| < 2.47$  (while unlike requirements in Chapter 7, no further cuts are applied). Electrons are required to pass the ‘Gradient’ isolation WP, designed to give an efficiency of 90% at  $p_T = 25$  GeV and 99% at  $p_T = 60$  GeV.

### 4.3 Muons

The characteristic behaviour of a muon in the detector is a particle ionising minimally. The muon reconstruction is done taking advantage of this feature, with information from the ID and MS tracking detectors being combined, while information from the calorimeters is also

used to account for cases of large energy loss in the calorimeters, and for MS-independent tagging of ID tracks as muon candidates in regions of limited MS coverage [38]. On the other hand, the muon identification is performed by applying quality requirements that suppress background, mainly from pion and kaon decays, that muon candidates often leave a distinctive “kink” topology in the reconstructed track, and hence degrade the quality of the track.

### 4.3.1 Reconstruction

The muon reconstruction consists of two steps. The first step is to reconstruct the stand-alone track in the MS, followed by reconstruction with complete detector information. The first step of reconstruction starts with the identification of short straight-line local track segments reconstructed from hits in an individual MS *stations* (layers). At least two of these segments are then required to build preliminary track candidates, with information from precision measurements in the bending plane and measurements of the second coordinate from the MS trigger detectors to create three-dimensional track candidates. A global  $\chi^2$  fit of the muon trajectory through the magnetic field is performed, outlier hits are removed and hits along the trajectory that were not assigned to the original track candidate are added. Finally, the tracks are fitted again with the updated hits information, and ambiguities are resolved by removing tracks that share a large fraction of hits with higher-quality tracks.

The second step of reconstruction is to combine the track candidates with the complete information of all sub-detectors. The reconstruction proceeds according to five main reconstruction strategies, leading to the corresponding muon types:

- *combined muons* are identified by matching MS tracks to ID tracks and performing a combined track fit based on the ID and MS hits, taking into account the energy loss in the calorimeters, in the region  $|\eta| < 2.5$ .
- *Inside out muons* are reconstructed in the region  $|\eta| < 2.5$  using a complementary inside-out algorithm, which extrapolates ID tracks to the MS and searches for at least three loosely-aligned MS hits. The ID track, the energy loss in the calorimeters and the MS hits are then used in a combined track fit.
- *Muon-spectrometer extrapolated muons* are muons reconstructed when a MS track cannot be matched to an ID track. Its parameters are extrapolated to the beamline and used to define an Muon-spectrometer extrapolated muon. Such muons are used to extend the acceptance outside that of the ID ( $|\eta| < 2.5$ ), thus exploiting the full MS coverage up to  $|\eta| = 2.7$ .
- *segment-tagged muons* are identified by requiring that an ID track extrapolated to the

MS satisfies tight angular matching requirements to at least one reconstructed MS segment, in the  $|\eta| < 2.5$  region.

- *calorimeter-tagged muons* are identified by extrapolating ID tracks through the calorimeters to search for energy deposits consistent with a minimum-ionising particle, in the region  $|\eta| < 0.1$ .

### 4.3.2 Identification

As a consequence, it is expected that the fit quality of the resulting combined track will be poor and that the momentum measured in the ID and MS may not be compatible [39]. Therefore, a set of requirements are applied on the number of hits in the different ID subdetectors and different MS stations, on the track fit properties, and on variables that test the compatibility of the individual measurements in the two detector systems [38]. A given set of requirements for each of the muon types defined above is referred to as working point (WP). The main metrics considered for designing the WPs are the selection efficiency and purity in simulation, where the prompt muon efficiency of a selection WP represents the probability that a real prompt muon traversing the detector is reconstructed as a muon and satisfies the WP; the purity of a selection WP is one minus the hadron misidentification rate (the fraction of light hadrons reconstructed as muons and satisfying the WP). Three standard selection WPs: *Loose*, *Medium*, and *Tight* are designed to cover the majority of physics analysis, while two additional WPs, *High- $p_T$*  and *Low- $p_T$* , are designed to accommodate the analyses targeting extreme phase space. An example of the reconstruction and isolation efficiency of muon is shown in Figure 4.4.

In Chapter 7, muons are selected with  $p_T > 7 \text{ GeV}$  and  $|\eta| < 2.7$ , and passing the ‘loose’ identification criteria with an efficiency of about 95%, as well as the ‘pflowLoose\_VarRadIso’ isolation criteria [40] with an efficiency of roughly 97%. (further cuts are applied to define signal muons, as defined in section 7.1.2); while in Chapter 6 muons are selected with  $p_T > 27 \text{ GeV}$  and  $|\eta| < 2.5$ , and passing the ‘medium’ identification criteria as well as a track-based isolation criteria, ‘TightTrackOnly’ with an efficiency of around 94% [38] (with no further cuts applied).

## 4.4 Jets

A jet can be defined as a collimated spray of stable particles arising from the fragmentation and hadronisation (as described in section TODO: add link to theory section) of a parton (quark or gluon) after a collision. Jets provide a link between the observed colourless stable particles and the underlying physics at the partonic level. A basic illustration of a

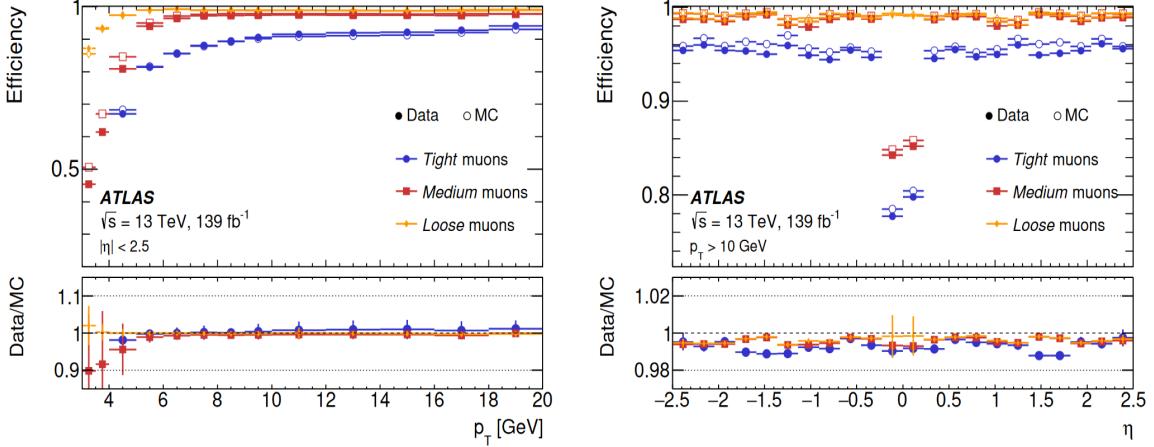


Figure 4.4: Muon reconstruction and identification efficiencies for the Loose, Medium, and Tight criteria. The left plot shows the efficiencies measured in  $J/\psi \rightarrow \mu\mu$  events as function of  $p_T$ . The right plot displays the efficiencies measured in  $Z \rightarrow \mu\mu$  events as a function of  $\eta$ , for muons with  $p_T > 10$  GeV. The predicted efficiencies are depicted as open markers, while filled markers illustrate the result of the measurement in collision data. The statistical uncertainty in the efficiency measurement is smaller than the size of the markers, and thus not displayed. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties. Reproduced from Ref. [38].

collision of two protons, the subsequent particle shower and a reconstructed jet is shown in Figure 4.5.

#### 4.4.1 Reconstruction

The jet reconstruction starts by forming clusters of energy deposit in the calorimeters by performing a three-dimensional topological clustering of individual calorimeter cell signals [42]. The clustering begins with a seed cell and builds a cluster by iteratively adding neighbouring cells, providing these cells have significant energy relative to the expected noise. This algorithm clusters the energy deposits into so called “topo-clusters” and combines their four-momenta. In Run 1 of the LHC, the ATLAS experiment used either solely the calorimeter or solely the tracker to reconstruct hadronic jets and soft particle activity, and the vast majority of analysis utilised jets that were built from topo-clusters, referred to *EMTopo jets*.

During Run-2, an alternative approach, called “*Particle flow*” (PFlow) became the default in ATLAS, as described in details in Ref. [43]. It is also the default approach used in this thesis. Measurements from both the tracker and the calorimeter are combined to form the signals. The particle flow algorithm provides a list of tracks and a list of topo-clusters. Then, well-measured tracks are selected following a set of stringent quality criteria. The algorithm then attempts to match each track to a single topo-cluster in the calorimeter. The

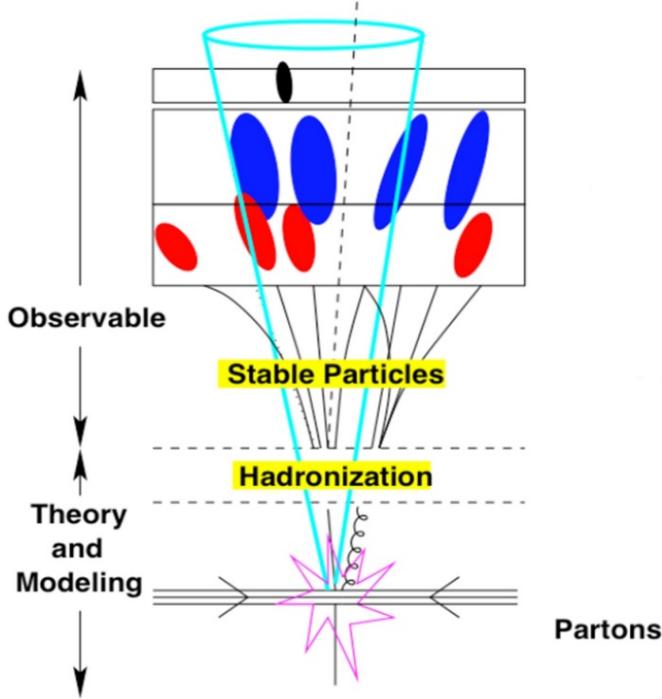


Figure 4.5: A simple example of an event showing the point of collision, the fragmentation and hadronization of the quarks and gluons and the resulting jet found through the detection of the stable particles. Image reproduced from Ref. [41].

expected energy in the calorimeter, deposited by the particle that also created the track, is computed based on the topo-cluster position and the track momentum. As a result, a new set of clusters called *PFlow clusters* are produced, matching the topo-clusters to the particles that created the high-quality tracks.

Using the PFlow clusters as inputs, PFlow jets are then reconstructed using the anti- $k_t$  algorithm [44]. The anti- $k_t$  algorithm sequentially combines PFlow clusters into larger clusters based on the momentum-weighted distance between clusters. For two clusters  $i$  and  $j$  the algorithm defines:

$$d_{i,j} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{i,j}^2}{R^2}$$

and

$$d_{i,beam} = p_{T,i}^{2p}$$

where  $p$  is a exponent parameter and the value of -1 is used for the anti- $k_t$  algorithm (as the name ‘anti’ suggests),  $\Delta R^2 i, j = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ , and  $p_{T,i}$ ,  $\eta_i$  and  $\phi_i$  are the transverse momentum, pseudorapidity and azimuthal coordinate of cluster  $i$ , respectively. The parameter  $R$  controls the size of the jet and for standard jets in ATLAS this is chosen

to be  $R = 0.4$ . The algorithm combines objects  $i, j$  with minimum value of  $d_{i,j}$  into one object iterartively, until no more  $d_{i,j}$  can be found greater than  $d_{i,beam}$ . The combined object is considered the final jet candidate. A validation study was performed by the author to compare the perfomance of EMTopo jets and PFlow jets in the content of the analysis presented in this thesis, as discussed in Sec TODO: add reference after selection is defined.

#### 4.4.2 Calibration

The calibration of jets starts prior to the reconstruction. Since the energy of the calorimeter cells is measured at the electromagnetic scale, the local cluster weighting (LCW) calibration is applied to topo-clusters (before they are used to form jets) to account for the differences in the detector response to hadronic and electromagnetic showers [45].

After the jets are reconstructed, the four-momenta of jets are calibrated with the jet energy scale (JES) calibration, which consists of several consecutive stages derived from a combination of MC-based methods and in-situ techniques [46]. MC-based calibrations correct the reconstructed jet four momentum to that found from the simulated stable particles within the jet, and in-situ techniques are used to measure the difference in jet response between data and MC simulation, with residual corrections applied to jets in data only. The calibrations account for features of the detector, the jet reconstruction algorithm, jet fragmentation, the busy data-taking environment resulting from multiple  $p p$  interactions, and the difference in jet response between data and MC simulation.

In order to calibrate the jet energy resolution (JER), the jet momentum must be measured precisely. As described in detail in Ref. [47], a *dijet balance* approach is used for this purpose, based on a well-defined dijet system, where the two jets are expected to have  $p_T$  that sum up to zero precisely.

Furthermore, jets arising from pileup are suppressed by using the *jet vertex tagger*, which is a multivariate combination of track-based variables developed to separate hard-scatter jets from pileup jets [48].

A jet cleaning selection is applied in order to veto any ‘fake’ jets, which arise from non-collision background events, such as cosmic rays, or from detector effects. Finally, all jets in the analysis are required to have  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.5$ .

#### 4.4.3 Identification of heavy-quark flavoured jets

Known as *flavour tagging*, the identification of jets containing  $b$ -hadrons ( $b$ -jets) against the large background of jets containing  $c$ -hadrons ( $c$ -jets) or jets coming from the hadronization of light ( $u,d,s$ ) quarks or gluons (light jets) is of major importance in many areas of the physics programme of the ATLAS experiment at the LHC. It is crucial in a large number of SM precision measurements, studies of the Higgs boson properties, and searches for new

phenomena [49–51]; it also plays an important role in the  $HH \rightarrow bb\tau\tau$  searches presented in Chapter 7.

The ATLAS Collaboration uses various algorithms to identify  $b$ -jets [52], referred to as  $b$ -tagging algorithms, when analysing data recorded during Run 2 of the LHC. These algorithms exploit the long lifetime, high mass and high decay multiplicity of  $b$ -hadrons, as well as the properties of the  $b$  quark fragmentation. Given a lifetime of the order of 1.5 ps,  $b$ -hadrons have a significant mean flight length ( $\langle c\tau \rangle \approx 450 \mu\text{m}$ ), in the detector before decaying, generally leading to at least one vertex displaced from the hard-scatter collision point, as illustrated in Figure 4.6.

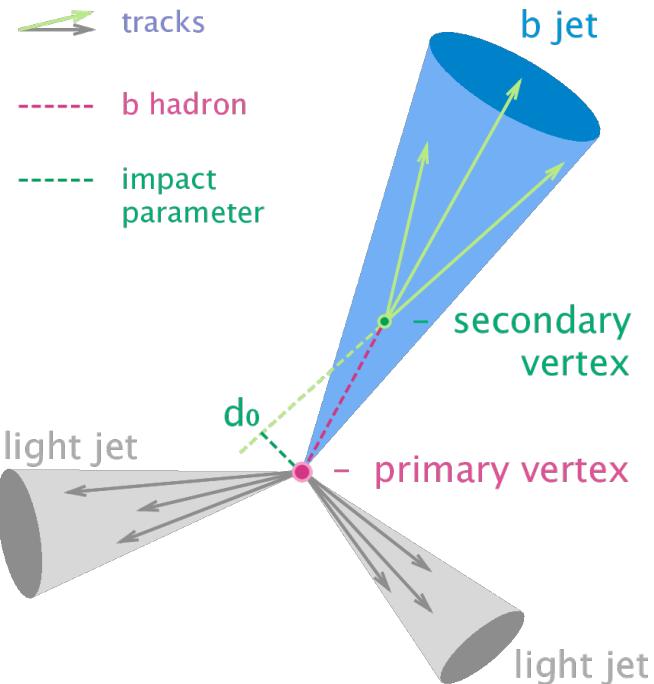


Figure 4.6: A diagram showing the  $b$  hadron decay initiated jets.

The strategy developed by the ATLAS Collaboration is based on a two-stage approach. Firstly, low-level algorithms reconstruct the characteristic features of the  $b$ -jets via two complementary approaches. A first approach, implemented in the IP2D and IP3D algorithms, is inclusive and based on exploiting the large impact parameters of the tracks originating from the  $b$ -hadron decay [53]. The second approach explicitly reconstructs displaced vertices. The SV1 algorithm [54], attempts to reconstruct an inclusive secondary vertex, while the JetFitter algorithm [55], aims to reconstruct the full  $b$ - to  $c$ -hadron decay chain. These algorithms, first introduced during Run 1 [52], have been improved and retuned for Run 2 [56]. Secondly, in order to maximise the  $b$ -tagging performance, the results of the low-level  $b$ -tagging algorithms are combined into high-level algorithms via multivariate classifiers.

The most performant algorithms presently in use in physics analyses at ATLAS are based on multivariate combinations of the available information (MV2) or additionally using a deep feed-forward neural network (DL1) [53, 57]. This is illustrated in Figure 4.7, where the performance is characterised by the probability of tagging a  $b$ -jet ( $b$ -jet tagging efficiency,  $\epsilon_b$ ) and the probability of mistakenly identifying a  $c$ -jet (light-flavour jet) as a  $b$ -jet labelled  $\epsilon_c(\epsilon_l)$ .

The distribution of the output discriminant of the MV2 and DL1 tagger for  $b$ -jets,  $c$ -jets, and light-flavour jets in simulated  $t\bar{t}$  events are shown in Figure 4.8. The evaluation of the performance of the algorithms is carried out using  $b$ -jet tagging single-cut operating points (OPs, or working points, WP). These are based on a fixed selection requirement on the  $b$ -tagging algorithm output discriminant distribution ensuring a specific  $b$ -jet tagging efficiency, for the  $b$ -jets present in simulated  $t\bar{t}$  sample. The discriminant distributions are also divided into five ‘pseudo-continuous’ bins, delimited by the selections used to define the  $b$ -jet tagging single-cut WPs for 85%, 77%, 70% and 60% efficiency, and bounded by the trivial 100% and 0% selections.

Depending on the low-level algorithm, the DL1 tagger can be further separated into two taggers: DL1 and DL1r, where the DL1 tagger uses traditional track-based impact parameter taggers IP2D and IP3D [58] and the DL1r tagger uses a Recurrent Neural Network Impact Parameter tagger (RNNIP) [53].

The calibration of DL1 and DL1r algorithms has been an original contribution of the author of this thesis and it is described in more detail in Chapter 6. The DL1r tagger is now the default  $b$ -tagging algorithm used for flavour tagging in ATLAS and is utilised in the analysis presented in this thesis.

## 4.5 Hadronically decaying $\tau$ lepton

With a mass of 1.777 GeV and a proper decay length of 87  $\mu\text{m}$  [59], tau leptons decay either leptonically ( $\tau_{lep} \rightarrow \ell\nu\ell\nu\tau$ ,  $\ell = e, \mu$ ) or hadronically ( $\tau_{had} \rightarrow \text{hadrons } \nu_\tau$ ) and do so typically before reaching active regions of the ATLAS detector. The leptonically decaying  $\tau_{lep}$  is simply reconstructed as either an electron or muon, with the neutrinos contributing to the real component of the  $E_T^{\text{miss}}$ . On the other hand, the hadronically decaying  $\tau_{tau}$  can be identified via their decay products. The hadronic tau lepton decays represent 65% of all possible decay modes [59]. In these decay modes, the hadronic decay products are one or three charged pions in 72% and 22% of all cases, respectively. Charged kaons are present in the majority of the remaining hadronic decays. In 78% of all hadronic decays, up to one associated neutral pion is also produced. This results in an experimental signature of a collimated calorimeter shower with either one or three associated tracks (*prongs*). The

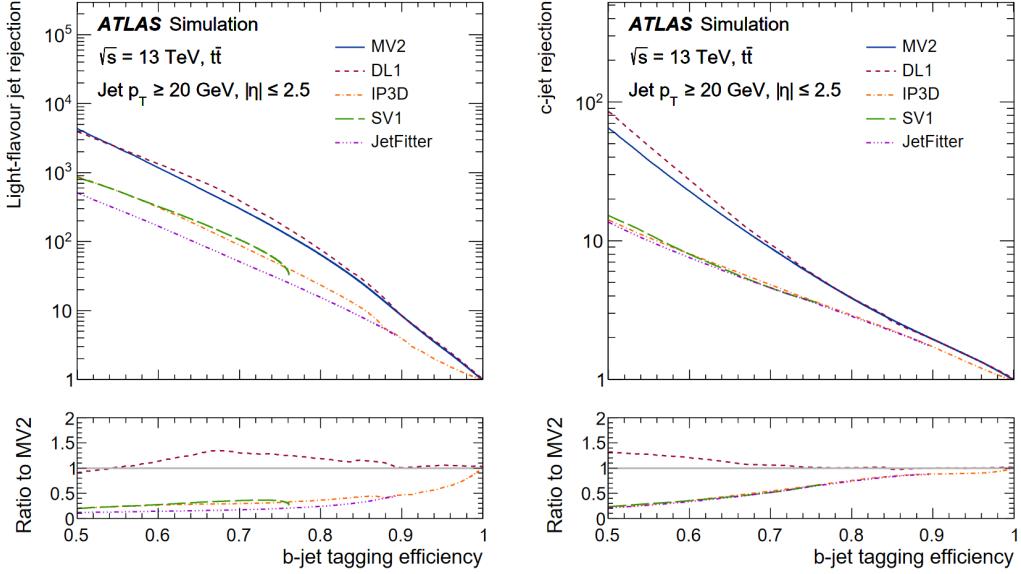


Figure 4.7: The light-flavour jet (left) and  $c$ -jet (right) rejections versus the  $b$ -jet tagging efficiency for the IP3D, SV1, JETFITTER, MV2 and DL1  $b$ -tagging algorithms evaluated on  $t\bar{t}$  events [56].

neutral and charged hadrons stemming from the tau lepton decay make up the visible decay products of the tau lepton, and are in the following referred to as  $\tau_{had-vis}$ .

### 4.5.1 Reconstruction

The  $\tau_{had-vis}$  candidates are seeded by jets formed using the anti- $k_t$  algorithm, with a jet-size parameter  $R = 0.4$ . For events with multiple interactions, the chosen primary may not be the one where the tau lepton originated. Here the *tau vertex association* algorithm is used with all tau candidates tracks within a region of  $\Delta R < 0.2$  around the jet seed direction as input. The  $p_T$  of these tracks is summed and the primary vertex candidate to which the largest fraction of the  $p_T$  sum is matched to is chosen as the tau vertex [48].

Tracks are associated with the  $\tau_{had-vis}$  if they are in the *core* region  $\Delta R < 0.2$  around the  $\tau_{had-vis}$  direction and satisfy the following criteria:  $p_T > 1$  GeV, at least two associated hits in the pixel layers of the inner detector, and at least seven hits in total in the pixel and the SCT layers. Furthermore, requirements are imposed on the distance of closest approach of the track to the vertex in the transverse plane,  $|d_0| < 1.0$  mm, and longitudinally,  $|z_0 \sin \theta| < 1.5$  mm. Tracks in the *isolation* region  $0.2 < \Delta R < 0.4$  are used for the calculation of identification variables and are required to satisfy the same selection criteria. The number of tracks (prongs) is susceptible to underestimation due to tracking inefficiency, or overestimation due to tracks from photon conversions passing the track selection criteria.

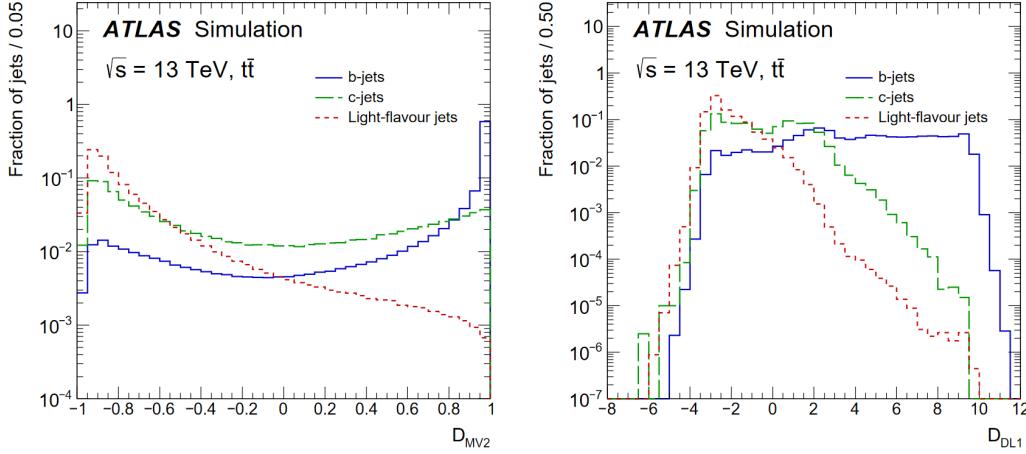


Figure 4.8: The fraction of light-flavour jets and  $c$ -jets versus the  $b$ -jets in the MV2 (left) and DL1 (right) b-tagging algorithms output distribution evaluated on  $t\bar{t}$  events [56].

Working point	Signal efficiency		Background rejection BDT		Background rejection RNN	
	1-prong	3-prong	1-prong	3-prong	1-prong	3-prong
Tight	60%	45%	40	400	70	700
Medium	75%	60%	20	150	35	240
Loose	85%	75%	12	61	21	90
Very loose	95%	95%	5.3	11.2	9.9	16

Table 4.1: List of defined working points with fixed true  $\tau_{had-vis}$  selection efficiencies and the corresponding background rejection factors for misidentified  $\tau_{had-vis}$  in dijet events for the BDT and RNN classifiers. Table reproduced from Ref. [60].

### 4.5.2 Identification

The  $\tau_{had-vis}$  reconstruction algorithm alone provides no discrimination against other particles that result in jet-like signatures in the detector. Therefore, dedicated algorithms are used to identify hadronic tau lepton decays. Here, a recurrent neural network (RNN) classifier is used as described in Ref. [60]. The RNN uses a combination of low-level input variables for individual tracks and clusters that are associated to the  $\tau_{had-vis}$  candidate as well as several high-level observables calculated from track and calorimeter quantities. Due to the distinct signatures of 1- and 3-prong  $\tau_{had}$  decays, the  $\tau_{had-vis}$  identification is split into dedicated algorithms for 1- and 3-prong  $\tau_{had-vis}$ . Four working points with increasing background rejection (*Very loose*, *Loose*, *Medium* and *Tight*) are defined to be used by physics analyses. The corresponding signal selection efficiencies and rejection powers are given in Table 4.1.

Compared to the ID that was used in the analysis of the  $36.1 \text{ fb}^{-1}$  data [61], which was based on a boosted decision tree, the RNN tau-ID shows better performance and allows to move to a looser WP gaining increased efficiency (about 24% and 11% in case of two

$\tau_{had}$  and one  $\tau_{had}$  in the final state, respectively) without losing jet rejection, as shown in Figure 4.9.

Selected  $\tau_{had-vis}$  candidates in the analysis are required to have  $p_T > 20$  GeV,  $|\eta| < 2.5$ , with candidates in the barrel-endcap transition region of the calorimeter ( $1.37 < |\eta| < 1.52$ ) vetoed due to poor detector instrumentation in this region, one or three tracks, unit charge, and to pass the ‘loose’  $\tau_{had}$ -ID working point. The loose WP corresponds to 85% efficiency for 1-prong and 75% efficiency for 3-prong (the efficiency is flat in  $p_T$  by definition).

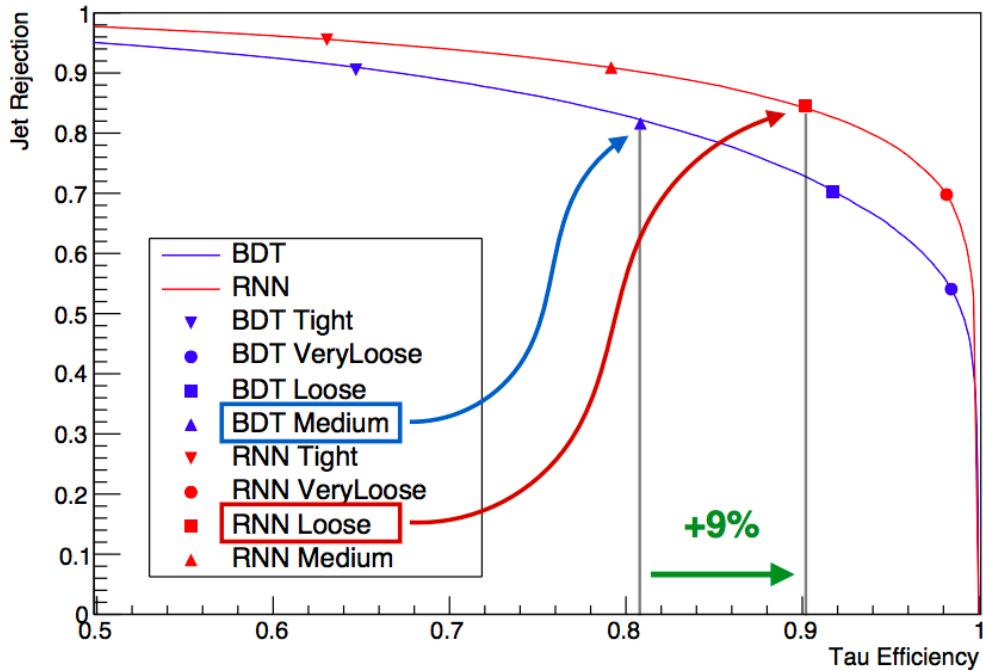


Figure 4.9: Jet rejection and tau efficiency of tau candidate, measured in  $\gamma^* \rightarrow \tau\tau$  sample for RNN-ID [60] and  $Z/\gamma^* \rightarrow \tau\tau$  sample for BDT-ID [62]. Jet rejection represents the probability of a jet not originating from a tau lepton being rejected by the identification algorithm. The green arrow indicates the increase in tau efficiency. Figure

Additional rejection of  $\tau_{had-vis}$  candidates originating from electrons is provided by a BDT employing track and shower shape information. The ‘loose’ working point is used for the analysis presented in Chapter 7, corresponding to a selection efficiency of about 95% efficiency for true  $\tau_{had-vis}$  [63].

## 4.6 Missing transverse energy

As defined in section 3.2.1, the  $\vec{E}_T^{miss}$  is given as  $\vec{E}_T^{miss} = -\sum_i \vec{p}_{T,i}$ , which is the negative vector sum of transverse momentum collected from the detector, from which one

or more “invisible” particle(s) can be inferred. The reconstruction of  $\vec{E}_T^{miss}$  is comprised of two contributions [64], within the region of  $|\eta| < 4.9$ . The first one is from *hard-event* signals, combining information from fully reconstructed and calibrated physics objects, i.e. electrons, muons, photons, jets, hadronically decaying  $\tau$ -leptons and jets. The second one is from the *soft-event* signals, consisting of reconstructed charged-particle tracks associated with the hard-scatter vertex but with no physics objects. Hence, the  $E_T^{miss}$  is calculated as:

$$E_T^{miss} = - \sum_{\text{selected electrons}} p_T^e - \sum_{\text{accepted photons}} p_T^\gamma - \sum_{\text{accepted } \tau\text{-leptons}} p_T^{\tau_{had}} - \sum_{\text{selected muons}} p_T^\mu - \sum_{\text{accepted jets}} p_T^{jet} - \sum_{\text{unused tracks}} p_T^{track}$$

A non-zero  $E_T^{miss}$  suggests the existence of a non-interacting particle. In the SM, the only particle that does not deposit energy in the detector are neutrinos, as they are weakly interacting leptons which do not undergo the strong or electromagnetic forces. Nevertheless, some BSM theories predict the existence of additional weakly-interacting particles.

## 4.7 Overlap removal

After the event is reconstructed, an overlap-removal procedure is applied to resolve ambiguities when a physical object is reconstructed as multiple particles in the ATLAS detector. The angular distance  $\Delta R$  is used to measure the overlap of two reconstructed objects. Overlaps between most of the detector objects used in the analysis are resolved by using the standard overlap removal tools AssociationUtils [65], with analysis specific procedure for the reconstructed  $\tau_{had-vis}$  anti- $\tau_{had-vis}$  objects and jets. The step-by-step procedure that is used to resolve ambiguities in the reconstructed objects is summarised in the following:

- $e_1-e_2$ : For two electrons  $e_1$  and  $e_2$  in an event, reject  $e_1$  if both electrons share the track and  $p_{T1} < p_{T2}$
- $\tau_{had-vis}-e$ : Reject  $\tau_{had-vis}$  if  $\Delta R < 0.2$
- $\tau_{had-vis}-\mu$ : Reject  $\tau_{had-vis}$  if  $\Delta R < 0.2$ :
  - Case 1 ( $\tau_{had-vis} p_T > 50\text{GeV}$ ):  $p_T, \mu > 2\text{GeV}$  and combined muon
  - Case 2 ( $\tau_{had-vis} p_T \leq 50\text{GeV}$ ):  $p_T, \mu > 2\text{GeV}$
- $\mu-e$ : Reject  $\mu$  if calo-muon and shared ID track
- $e-\mu$ : Reject  $e$  if shared ID track
- jet- $e$ : Reject jet if  $\Delta R < 0.2$

- $e$ -jet: Reject  $e$  if  $\Delta R < 0.4$
- jet- $\mu$ : Reject jet if  $N_{track} < 3$  ( $p_T^{track} > 500\text{MeV}$ ), and  $\Delta R < 0.2$
- $\mu$ -jet: Reject  $\mu$  if  $\Delta R < 0.4$

Additionally, an analysis-specific overlap-removal procedure for  $\tau_{had-vis}$  anti- $\tau_{had-vis}$  and jets is implemented:

- jet- $\tau_{had-vis}$  Reject jet if  $\Delta R < 0.2$
- anti- $\tau_{had-vis}$ -jet: Reject anti- $\tau_{had}$  if jet is  $b$ -tagged and  $\Delta R < 0.2$
- jet-anti- $\tau_{had-vis}$  Reject jet if  $\Delta R < 0.2$

This establishes the following priority:  $\tau_{had-vis} > b$ -tagged jet  $>$  anti- $\tau_{had-vis} >$  un-tagged jet.

# Chapter 5

## Data and Monte Carlo Simulation

### 5.1 Simulation of physics processes **TODO: move this to theory section**

A precise and reliable theoretical prediction is the key to interpret and analyse the data recorded by the ATLAS experiment. It enables the quantification of agreement between the data and the SM, and test of possible new physics beyond the SM. In this thesis, the signal and background processes are simulated by Monte Carlo event generators. The full simulation process is comprised of a few consecutive steps: starting with the simulation of the hard-scattering process, followed by the simulation of the parton-showering and finally the simulation of the interaction of the particles with the detector and the response of the detector.

The hard-scattering process happens when two partons which carry a fraction of the protons collide inelastically. The partons can be valence quarks, which are the quarks or anti-quarks that determine the quantum numbers of the proton, or gluons which mediate the strong force, or sea quarks which are virtual quark-anti-quark pairs that are created and annihilated promptly. The hard-scatter process is typically characterised by large momentum transfer, and the probability of a parton carrying a given fraction of the total proton momentum is described by the *parton distribution function* (PDF). In the MC simulation, the hard-scattering process is modelled by the *matrix element*, using leading order (LO) or next-to-leading order (NLO) Feynman diagrams. It can also be simulated at next-to-next-to-leading order (NNLO) for better theoretical approximation.

After the hard-scattering process, the parton will undergo the showering process where it hadronises or radiates further partons. This process is simulated by dedicated algorithms. The hadronisation process happens mostly in low-energy regime, which is non-perturbative and requires phenomenological modelling exploiting specific hadronisation models. On the other hand, the radiation process happens at higher-energy, and it stops when the

parton loses enough energy to reach the confinement energy-scale. The hard-scattering and showering processes of partons are described by the combination of the matrix element generator and the parton shower algorithms.

In addition to the hard-scattering and showering processes, interaction can occur prior or after them, as referred to initial state radiation (ISR) or final state radiation (FSR). The energy scale of these additional processes are typically a few GeV, much smaller compared to the hard-scattering energy scale. The radiated gluons or photons, together with other particles originating from soft scattering interactions, are described as *underlying event*.

In addition, pileup events, as described in section 3.1.2, need to be simulated. They are separate collisions and are simulated separately from the hard-scattering event. However, in the reconstruction of the event, they cannot be separated from the hard-scattering event, therefore they are later overlaid on the hard-scattering events.

A variety of MC generator programs were developed. Some of them are called multi-purpose generators and can simulate a full-event on their own, while some others are dedicated only to hard-scattering or parton shower and need to be used combining with other generators. After simulation, simulated events are passed to GEANT 4 [66, 67], a software package for simulating the interactions of particles with matter.

Finally, events generators require *tuning* to match data. The tuning parameters are based on phenomenological models and are applied on hadronisation simulation and underlying event.

## 5.2 Data samples

The data used in this search were collected at a centre-of-mass energy of 13 TeV between 2015 and 2018. For the FTAG calibration presented in Chapter 6, the data sample was collected using a set of single-electron [68] and single-muon triggers [69]. Requirements on  $p_T$  over a range of 24–300 GeV is applied to the single-electron triggers, with additional quality and isolation requirements depending on the  $p_T$  threshold and the data-taking period. While for the single-muon triggers, requirements on  $p_T$  over a range 20–26 GeV are applied on the isolated muons, and a tigher cut of 50 GeV is applied for muons without any isolation requirement.

For the  $HH$  searches presented in Chapter 7, single-lepton triggers and lepton-plus- $\tau_{\text{had}}$  triggers are used. Details about these triggers are discussed in Section 7.1.2.

Events are selected for analysis only if they are of good quality and if all the relevant detector components are known to be in operating conditions [70]. The total integrated luminosity of the data, after meeting the good quality criteria, is  $139.0 \pm 2.4 \text{ fb}^{-1}$  [71, 72]. The recorded events contain an average of 34 simultaneous inelastic  $pp$  collisions per

bunch-crossing.

### 5.3 Simulated event samples

As mentioned in section 5.1, the signal and background processes are modelled by Monte Carlo simulation. No signal sample is defined for the FTAG calibration. The signal targeted in the  $HH$  searches includes the SM-like non-resonant  $HH$  production via ggF and VBF, and the BSM resonant  $HH$  production. To simulate these processes, the full ATLAS detector simulation [67] (FS) is applied, passing to the GEANT 4 package. The only exception is the resonant signal samples, where a fast simulation (AF2) [67] is used, that instead of fully simulating the response of the calorimeters, it is replaced by pre-simulated shower to save computation time. For unstable hadrons i.e.  $b$ - and  $c$ -hadrons, the decay process is simulated by the EvtGEN v1.6.0 package[73], with the exception of the VBF non-resonant samples and samples generated by SHERPA [74]. The resulting events were then processed through the same reconstruction programs as the data. To simulate the pileup effects, minimum bias events are overlaid on the simulation, exploiting the PYTHIA 8.186 generator [75] for soft QCD processes using the A3 tune [76]. The NNPDF2.3LO [77] PDFs are used. In addition, the Higgs boson mass was fixed to 125 GeV for all simulated samples that contain this particle. The mass of the Higgs boson is assumed to be 125 GeV for all samples containing a SM Higgs boson. This value is also used for calculating the single- and pair- production cross-sections of the Higgs boson, as well as the decay branching ratio of the Higgs boson.

A summary of the event samples used for the simulation of the signal and background processes is shown in Table 5.1.

Process	ME generator	ME QCD	PDF	PS	Tuned parameters	Cross-section
<b>Signal</b>						
non-resonant $gg \rightarrow HH$ (ggF)	POWHEG-BOX v2	NLO	PDF4LHC15 NLO	PYTHIA 8.244	A14	NNLO FTApprox
non-resonant $qq \rightarrow qqHH$ (VBF)	MADGRAPH5_aMC@NLO v2.7.3	LO	NNPDF3.0NLO	PYTHIA 8.244	A14	N3LO(QCD)
resonant $gg \rightarrow X \rightarrow HH$	MADGRAPH5_aMC@NLO v2.6.1	LO	NNPDF2.3LO	HERWIG v7.1.3	H7.1-Default	–
<b>Top-quark</b>						
$t\bar{t}$	POWHEG-BOX v2	NLO	NNPDF3.0NLO	PYTHIA 8.230	A14	NNLO+NNLL
single top ( $t$ -, $s$ -, $Wt$ -channels)	POWHEG-BOX v2	NLO	NNPDF3.0NLO	PYTHIA 8.230	A14	NLO
$t\bar{t}Z$	SHERPA 2.2.1	NLO	NNPDF3.0NNLO	SHERPA 2.2.1	Default	NLO
$t\bar{t}W$	SHERPA 2.2.8	NLO	NNPDF3.0NNLO	SHERPA 2.2.8	Default	NLO
<b>Single Higgs boson</b>						
ggF	POWHEG-BOX v2	NNLO	NNPDF3.0NLO	PYTHIA 8.212	AZNLO	N3LO(QCD)+NLO(EW)
VBF	POWHEG-BOX v2	NLO	NNPDF3.0NLO	PYTHIA 8.212	AZNLO	NNLO(QCD)+NLO(EW)
$qq \rightarrow WH, ZH$	POWHEG-BOX v2	NLO	NNPDF3.0NLO	PYTHIA 8.212	AZNLO	NNLO(QCD)+NLO(EW)
$gg \rightarrow ZH$	POWHEG-BOX v2	NLO	NNPDF3.0NLO	PYTHIA 8.212	AZNLO	NLO+NLL
$t\bar{t}H$	POWHEG-BOX v2	NLO	NNPDF3.0NLO	PYTHIA 8.230	A14	NLO
<b>Vector boson + jets</b>						
$W/Z+jets$	SHERPA 2.2.1	NLO ( $\leq 2$ jets), LO (3,4 jets)	NNPDF3.0NNLO	SHERPA 2.2.1	Default	NNLO
<b>Diboson</b>						
$WW, WZ, ZZ$	SHERPA 2.2.1	NLO ( $\leq 1$ jet), LO (2,3 jets)	NNPDF3.0NNLO	SHERPA 2.2.1	Default	NLO

Table 5.1: The generators used for the simulation of the signal and background processes. The order of the cross-section calculation refers to the expansion in the strong coupling constant ( $\alpha_S$ ). The acronyms ME, PS and UE are used for matrix element, parton shower and underlying event, respectively. The terms ggF, VBF refer to gluon-gluon fusion and vector-boson fusion respectively. The cross-section of the resonant production is not shown as a dummy cross-section of 1 pb was chosen. Reproduced from Ref. [78].

### 5.3.1 Simulated signal samples

In the  $HH$  searches, contributions from both the ggF and VBF to the SM non-resonant  $HH$  signal production are included, each simulated with different generators and PDFs. The expansion order of Feynman diagrams are also different. While for the resonant  $HH$  signal, only the ggF contribution is considered. It was simulated for 20 values of the resonance mass,  $m_X$ , between 251 GeV and 1600 GeV (251, 260, 280, 300, 325, 350, 375, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1100, 1200, 1400 and 1600).

The ggF events were generated with the Powheg-Box v2 generator [79] at next-to-leading order (NLO) with finite top-quark mass, using the PDF4LHC15 NLO PDF set [80]. Parton showers and hadronisation were interfaced to Pythia 8.244 [75] with the A14 set of tuned parameters [81, 82] and the NNPDF2.3LO PDF set.

TODO: move this paragraph to theory The cross-section of the ggF non-resonant  $HH$  production is calculated at next-to-next-to-leading order (NNLO) FTApprox [83], taking into account the finite top-quark mass assumption. The cross-section is  $31.05^{+2.2\%}_{-5.0\%}(\text{scale}) \pm 2.1\%(\alpha_S) \pm 2.1\%(\text{PDF}) \pm 2.6\%(m_{\text{top}})$  fb at  $\sqrt{s} = 13$  TeV and  $m_H = 125$  GeV [84]. The scale uncertainty is due to the finite order of quantum chromodynamics (QCD) calculations, the  $\alpha_S$  and PDF terms account for the uncertainties on the strong coupling constant and parton distribution functions respectively, and the  $m_{\text{top}}$  uncertainty is related to the top-quark mass scheme. The normalisation of this process is set to the production cross-section times the  $b\bar{b}\tau^+\tau^-$  branching ratio (BR),  $\sigma_{ggF} \times BR_{b\bar{b}\tau^+\tau^-} = 31.05 \text{ fb} \times 0.0730 = 2.268 \text{ fb}$ .

On the other hand, the VBF events generated using the MadGraph5\_aMC@NLO v2.7.3 [85] generator at LO with the NNPDF3.0NLO [86] PDF set. Parton showering and hadronisation were simulated using Pythia 8.244 with the A14 tune and the NNPDF2.3LO PDF set. To simulate the decays of the  $b$ - and  $c$ -hadrons, the EvtGen v1.7.0 program was used.

TODO: move this paragraph to theory The cross-section of the VBF non-resonant  $HH$  production calculated at next-to-next-to-next-to-leading order (N3LO) in QCD in the limit in which there is no partonic exchange between the two protons [87] is  $1.726^{+0.03\%}_{-0.04\%}(\text{scale}) \pm 2.1\%(\text{PDF} + \alpha_S)$  fb [84]. This process is normalised to the cross-section times the  $b\bar{b}\tau^+\tau^-$  branching ratio,  $\sigma_{VBF} \times BR_{b\bar{b}\tau^+\tau^-} = 1.726 \text{ fb} \times 0.07306 = 0.1261 \text{ fb}$ . Other non-resonant  $HH$  production modes are not considered as their contributions to the analysis sensitivity are expected to be negligible.

Finally, the resonant signal of a heavy spin-0 narrow width resonance via ggF production was simulated with the MadGraph5\_aMC@NLO v2.6.1 generator using the NNPDF2.3LO PDF set at LO. The parton shower and hadronisation were simulated to HERWIG 7.1.3 [88, 89], using the H7.1-Default tune [90] and the NNPDF2.3LO PDF set.

TODO:move this paragraph to theory The normalisation of these resonant signal samples in the analysis is set to  $\sigma \times BR = 1 \text{ pb} \times BR(bb\tau\tau) = 0.073061 \text{ pb}$ . A dummy cross-section of 1 pb is chosen for these samples to make combination with other decay channels easier, and to ease scaling of the signal for the calculation of the cross-section limits (more details in Section 7.5).

### 5.3.2 Non-resonant signal reweighting and combination

In the BSM scenarios, the non-resonant di-Higgs production is sensitive to the self-coupling constant and other possible anomalous coupling, as described in section ?? (TODO: ref back to theory section about klambda dependence). In this thesis, a reweighting method is used to evaluate the non-resonant di-Higgs production with a range of possible values of self-coupling modifier,  $k_\lambda$ , and 7 benchmark models (BM) of a set of five HEFT couplings.

**$k_\lambda$  Reweighting** For the ggF non-resonant  $HH$  production, MC samples are generated at  $k_\lambda = 1$  and 10, while for VBF production, MC samples are generated at  $k_\lambda = 0, 1, 2$  and 10. A sample combination technique is used to model the signal hypothesis at different  $k_\lambda$  values. For the ggF di-Higgs production, a reweighting method described in Ref. [91] is used to obtain predictions at different  $k_\lambda$  values in the range  $k_\lambda \in [-30, 30]$  in increments of 0.2 based on a linear combination of generator samples at  $k_\lambda = 0, 1$  and 20. The remaining  $k_\lambda = 10$  sample is used to validate the method. For each  $k_\lambda$  value, a set of weights  $w(m_{HH}, k_\lambda)$  is evaluated by dividing the binned  $m_{HH}$  distribution of the  $k_\lambda$  target sample by the SM distribution. They can be used to reweight the SM non-resonant sample to any  $k_\lambda$  value, which is performed at analysis level—after reconstruction and the selection steps defined in section 7.1.2. Given the assumption that the kinematic of the ggF events and their acceptance depend only on  $m_{HH}$  variable, using the weights  $w(m_{HH}, k_\lambda)$  the reweighted sample describes correctly any kinematic distribution of a given target  $k_\lambda$  value. Good closure is found between the distributions obtained from  $k_\lambda$  generated and reweighted, as shown in section ?? TODO: add reference to systematics section.

For the VBF di-Higgs production, the above reweighting procedure is not valid, because the kinematic of the events can not be defined just using a single variable such as  $m_{HH}$ . Instead, three fully-reconstructed MC samples with  $k_\lambda = 1, 2, 10$  are used. The event distributions and the multivariate algorithm output (more details in section 7.3) for any  $k_\lambda$  value are obtained from the linear combination of the corresponding distributions of the three samples at analysis level. As defined in section TODO: ref back to theory chapter, the full cross-section for the VBF  $HH$  production involves three diagrams, and expanding

Benchmark	$c_t$	$c_{hhh}$	$c_{tt}$	$c_{ggh}$	$c_{gghh}$	Cross-sections (fb)
SM	1	1	0	0	0	31.05
BM1	0.94	3.94	-1/3	0.5	1/3	181.40
BM2	0.61	6.84	1/3	0.0	-1/3	135.24
BM3	1.05	2.21	-1/3	0.5	0.5	108.90
BM4	0.61	2.79	1/3	-0.5	1/6	50.45
BM5	1.17	3.95	-1/3	1/6	-0.5	117.04
BM6	0.83	5.68	1/3	-0.5	1/3	144.91
BM7	0.94	-0.10	1	1/6	-1/6	97.96

Table 5.2: The proposed coupling values and the corresponding cross-sections of each BMs. The physical meaning of each coupling is described in section ?? TODO: add reference to theory section.

the absolute squared of the amplitude yields six terms:

$$\sigma = \kappa_V^2 \kappa_\lambda^2 a_1 + \kappa_V^4 a_2 + \kappa_{2V}^2 a_3 + \kappa_V^3 \kappa_\lambda a_4 + \kappa_V \kappa_\lambda \kappa_{2V} a_5 + \kappa_V^2 \kappa_{2V} a_6.$$

In the case of  $\kappa_\lambda$  scan, this formula is reduced to

$$\sigma = \kappa_\lambda^2 a_1 + \kappa_\lambda a_2 + a_3.$$

Using the basis of  $\kappa_\lambda = 1, 2, 10$ , the linear coefficients for combining the three samples defined are then given by:

$$\begin{aligned} \sigma(\kappa_\lambda) &= \left( \frac{\kappa_\lambda^2}{9} - \frac{4\kappa_\lambda}{3} + \frac{20}{9} \right) \times \sigma(1) + \left( -\frac{\kappa_\lambda^2}{8} + \frac{11\kappa_\lambda}{8} - \frac{5}{4} \right) \times \sigma(2) \\ &\quad + \left( \frac{\kappa_\lambda^2}{72} - \frac{\kappa_\lambda}{24} + \frac{1}{36} \right) \times \sigma(10) \end{aligned} \tag{5.1}$$

**HEFT Benchmarks reweighting** The reweighting is only applied to the ggF non-resonant production. A similar approach is used: for each BM, a set of weights  $w(m_{HH}, BM)$  is evaluated by dividing the binned  $m_{HH}$  distribution of the BM sample by the SM distribution; these weights can then be used to reweight the SM non-resonant sample to any BM signals as defined in section TODO: ref back to theory chapter.

TODO: move this paragraph and table to theory The assumed couplings values and the corresponding cross-sections of these BMs are shown in Table 5.2. These BM points are defined to represent the different possible characteristic shapes in  $m_{HH}$  distribution, as described in Ref. [92].

### 5.3.3 Simulated background samples

The major background processes for the  $HH$  searches are the  $t\bar{t}$ , single-top, boson produced in association with jets and jets faking a  $\tau_{\text{had}}$  (more details in section 7.2.1). Minor background processes include the Drell-Yan processes, processes with dibon final states and single-Higgs processes. These background processes, except for the jets faking a  $\tau_{\text{had}}$  background, are modelled by the full ATLAS detector simulation. For the FTAG calibration, the events used for calibration mostly originate from the  $t\bar{t}$  processes. Other minor background includes the single-top, diboson, production of  $t\bar{t}$  in association with a boson and boson produced in association with jets. Likewise, all processes are passed through the full simulation. The generators, PDFs, expansion order and tune used in the  $HH$  searches and the FTAG calibration for the samples in common are in general identical, unless specified.

The  $t\bar{t}$  production and the single top-quark events in the  $Wt$ -,  $s$ - and  $t$ -channels were simulated by the PowHEG-Box v2 generator [79, 93–95] together with the NNPDF3.0NLO PDF set [86]. The showering, hadronisation and underlying event are modeled by PYTHIA 8.230, with parameters set according to the A14 tune [96] and using the NNPDF2.3LO PDF set. The top-quark mass was set to 172.5 GeV, with top-quark spin correlations preserved. To achieve the better accuracy, the  $t\bar{t}$  production cross-section is calculated at next-to-next-to-leading order and next-to-next-to-leading-logarithm (NNLO+NNLL) [97]. While for the single-top process, the cross-sections are calculated at NLO [98–100]. The  $t\bar{t}$ - $Wt$  interference is removed using the diagram removal scheme [101] (more details in Section ?? TODO: add reference to the systematics for singletop ).

The production of bosons in association with jets ( $W/Z+\text{jets}$ ) are simulated by SHERPA 2.2.1 generator [74] using the NNPDF3.0NNLO [86] PDF set. The tuning used for parton shower is developed by the SHERPA authors. The matrix elements are simulated for up to two partons at NLO and up to four partons at LO, calculated with the COMIX [102] and OPENLOOPS [103] libraries. The samples are normalised to NNLO prediction [104].

Similar settings apply to the diboson ( $WW$ ,  $WZ$  and  $ZZ$ ) events. SHERPA 2.2.1 generator is used to simulate these processes for the  $HH$  searches, while SHERPA 2.2.1 and SHERPA 2.2.2 are used for the FTAG calibration depending on the process. For both the  $HH$  searches and FTAG calirabtion, diboson samples are simulated using matrix elements at NLO accuracy in QCD for up to one additional parton and at LO accuracy for up to three additional parton emissions. The cross-section is calculated at NLO accuracy, while the rest of the settings remains the same as the ones for the boson+jets background.

In addition, the events where a vector boson is produced in association with  $t\bar{t}$  ( $t\bar{t}Z$  and  $t\bar{t}W$ ) are generated differently between the  $HH$  searches and the FTAG calibration. In the former, SHERPA 2.2.1 (2.2.8) is used to simulate the  $t\bar{t}Z$  ( $t\bar{t}W$ ) production at NLO. The cross-

sections are calculated at NLO accuracy and the rest remains the same as the settings for the boson+jets. While for the FTAG calibration, the  $t\bar{t}Z$  and  $t\bar{t}W$  events are modelled using the **MADGRAPH5\_aMC@NLO** v2.3.3 [85] generator at NLO with the **NNPDF3.0nlo** [86] parton distribution function (PDF). The events are interfaced to **PYTHIA8.210** [105] using the A14 tune [81] and the **NNPDF2.3l0** [86] PDF set. The decays of bottom and charm hadrons are simulated using the **EvtGEN** v1.2.0 program [73].

In the  $HH$  searches presented in this thesis, the SM single Higgs boson production is considered as part of the background. It has played a non-negligible role as a background to the analysis, especially in the non-resonant search. This due to the similar kinematics between the single- and double- Higgs production. The simulated events of the single Higgs boson production, in various processes, were generated using **POWHEG-Box v2** generator and **NNPDF3.0NNLO** PDF set.

The ggF single Higgs production is interfaced to the Powheg **NNLOPS** program [106, 107] at NNLO accuracy. The cross-section is calculated at Next-to-Next-to-Next-to-Leading order (N3LO) for the QCD processes and NLO for the electroweak expansion [108–112]. On the other hand, the VBF single Higgs production events are interfaced to **PYTHIA 8.212** using the **CTEQ6L1** PDF set [113], together with the **AZNLO** tune [114]. The cross-section of the VBF process is calculated at NNLO for QCD processes and NLO for electroweak expansion [108, 115–117].

The  $qq \rightarrow WH$ ,  $qq \rightarrow ZH$  and  $gg \rightarrow ZH$  simulated events were interfaced to **PYTHIA 8.212** for parton shower and hadronisation using the **AZNLO** tune together with the **CTEQ6L1** PDF set. The cross-sections are set to the **NNLO(QCD)+NLO(EW)** calculations for  $qq \rightarrow WH$  and  $qq \rightarrow ZH$  [118–124], and to the next-to-leading order and next-to-leading-logarithm (NLO+NLL) in QCD for  $gg \rightarrow ZH$  [125–129].

For processes where single Higgs produced in association with a vector boson, i.e. the  $qq \rightarrow ZH$ ,  $gg \rightarrow ZH$  and  $qq \rightarrow WH$ , processes, **PYTHIA 8.212** is used to simulate the parton shower and hadronisation using the **AZNLO** tune, together with the **CTEQ6L1** PDF set. For the  $qq \rightarrow WH$  and the  $qq \rightarrow ZH$  sample, the cross-section is calculated at NNLO for QCD processes and NLO for electroweak expansion [118–124]; while for  $gg \rightarrow ZH$  the cross-section is calculated at NLO for QCD and next-to-leading logarithmic (NLL) for electroweak expansion [125–129].

Finally, the simulated events with a single Higgs produced in association with a pair of top-quarks are interfaced to **PYTHIA 8.230**. The parton shower and hadronisation are set to the A14 tune, using the **NNPDF2.3LO** PDF set. The cross-section is calculated at NLO accuracy [108].

In the  $HH$  searches and FTAG calibration presented in this thesis, additional samples were produced with alternative generators or settings, in order to estimate systematic

uncertainties in the event modelling. These alternative will be defined once they are introduced in the content.

# Chapter 6

## Charm jet mis-tagging calibration

### 6.1 Calibration methods for $b$ -jet and light jet

MC simulations are not able to model exactly the performance of the  $b$ -tagging algorithms in data. For this reason calibration is required, i.e. correcting MC to recover the data in terms of  $b$ -tagging efficiency, charm jet mis-tagging and light jet mis-tagging rates [56]. The calibration is performed for all supported jet collections(TODO: refer back to the object definition chapter) and working points, which are cuts in the  $b$ -tagging algorithm output identifying the different tagging efficiencies and corresponding light jet and  $c$ -jetrejection rate. In general, the efficiency is calculated with data and simulations, and scale factors are then calculated to match the efficiency extracted from simulations to the data. The production of  $t\bar{t}$  pairs at the LHC provides an abundant source of  $b$ -jets by virtue of the high cross-section and the  $t \rightarrow Wb$  branching ratio being close to 100%. A very pure sample of  $t\bar{t}$  events can be selected by requiring that both  $W$  bosons decay leptonically, referred to as di-leptonic  $t\bar{t}$  decays in the following. For the  $b$ -jet calibration, the performance of the  $b$  tagging algorithms is evaluated in the simulation and the efficiency with which these algorithms identify jets containing  $b$ -hadrons is measured in collision data. The measurement uses a likelihood-based method in the di-leptonic  $t\bar{t}$  sample, where events with exactly 2 jets and 2 opposite-sign leptons are selected. The data  $b$ -jet efficiency is then extracted from a combined likelihood fit, and subsequently compared with that predicted by the simulation. Scale factors are then calculated to emulate the performance of the algorithms to the data [56].

For the light jet mis-tagging calibration, two methods are used to measure the mis-tagging rate from the data [130]. The first is the negative tag method, which uses a high statistics data sample enriched in light jets with the application of a modified algorithm which reverses some of the criteria used in the nominal identification algorithm. The second is the adjusted Monte Carlo (adjusted-MC) method, which adjusts the characteristic track

observables in the simulation to im the data, and then compares the adjusted simulation to the “standard” simulation. The scale factors are then calculated using the these two methods. The scale factors of the two different methods are in good agreement within the systematic uncertainties.

## 6.2 Calibration method for $c$ -jet

It is worth mentioning that the author’s qualification task to become an ATLAS author is to calibrate the rate of a charm jet being mis-identified as a  $b$ -jet which is a part of the calibration of the  $b$ -tagging algorithm. During the task the calibration range has been extended down to 20 GeV (previously 25 GeV) in jet  $p_T$  and a new selection category has been developed to increase the data statistics of the scale factors in the high- $p_T$  ( $p_T^{jet} > 70$  GeV) region. The calibration is performed on the PFlow jets (as defined in Section 4.4) and *VR-Track jets* reconstructed using the variable radius jet algorithm [131].

As determined by the CKM matrix [132, 133], the  $W$  boson decays dominantly to a pair of light quarks ( $u$  quark and  $d$  quark) or to a  $s$  quark and a  $c$  quark. The  $W$  boson decays very rarely to pairs containing a  $b$  quark. More specifically, the branching ratio of a  $W$  boson decays to a  $u$  quark and  $d$  quark pair or a  $s$  quark and  $c$  quark pair is 33.1%, and to pairs containing a  $b$  quark is only 0.057% [59]. Therefore,  $b$ -tagged jets from the  $W$  decay are most likely to be mis-tagged  $c$ -jetsor light jets.

Furthermore, given the ratio between the DL1 light jet rejection and the corresponding charm jet rejection ranges from 10 to 40 (Figure 4.7), the  $c$ -jet is much more likely to be mis-tagged than the light jet. This allows for a source of mis-tagged  $c$ -jetsto be obtained in the  $t\bar{t}$  events, requiring that one  $W$  boson decays leptonically and the other decay hadronically (referred to as semi-leptonic  $t\bar{t}$  decay in the following), where the  $b$ -tagged jets from the  $W$  decay are candidates of mis-tagged  $c$ -jets. Requiring a  $W$  boson decaying leptonically reduces the number of combinations of jets of different flavour, and allows triggering with the lepton.

The events kinematics are shown by the diagram in Figure 6.1, where the  $t\bar{t}$  pair decays to a  $b$  and a  $\bar{b}$  quark, circled in red. One of the  $W$  bosons, circled in blue, decays hadronically to quarks, and the other  $W$  boson decays leptonically to either an electron or a muon and the corresponding neutrinos, circled in green and purple, respectively. The lepton in the final state is used for triggering. The following notation will be used: the jets that are the decay products of the  $W$  boson are referred to as  $W$  jets and the remaining two jets are referred to as top-jets.

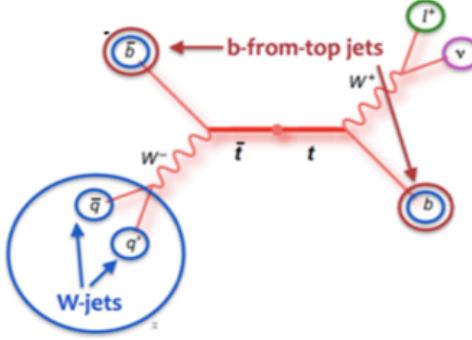


Figure 6.1: Feynman diagram of the semi-leptonic  $t\bar{t}$  events.

A kinematic likelihood technique, referred to as KLFitter [134], is used to assign jets to the proper  $t\bar{t}$  decay product (more details in Section 6.4). The following notation will be used: the jets that are assigned as the decay products of the  $W$  boson are referred to as  $W$  jets and the remaining two jets are referred to as top jets.

The charm jet efficiency is defined as the ratio of events with either of the  $W$ -jet is tagged. The efficiency is evaluated in four  $p_T$  intervals, with boundaries of 20, 40, 65, 140 and 250 GeV for the PFlow jets and 15, 20, 40, 140 GeV for the VR-Track jets; and for four tagging intervals with the boundaries of 85%, 77%, 70% and 60%.

The choice of the bin boundaries ensures enough statistics for each bin and hence relatively flat statistical uncertainty, given the underlying charm-jet  $p_T$  spectrum as shown in Figure 6.7. The boundaries for the VR-Track jets are lower than for PFlow jets, since the track jets miss the neutral particles the reconstructed energy is significantly below the true jet energy.

The main method described in the chapter is for the “fixed-cut” calibration where the efficiency is defined as the fraction of  $b$ -jets passing the tagger. Jets are said to be tagged (untagged) at particular working point if they have DL1r scores greater (less) than the DL1r score of that working point. The events with both  $W$ -jet are discarded to simplify the fit described in the following.

To extract the scale factors of the charm jet mis-tagging, a fit is performed by minimising the  $\chi^2$  defined as:

$$\begin{aligned} \chi^2 = & \sum_{t=1}^4 \sum_{i=1}^4 \sum_{j=1}^4 (N_{\text{data}}^t(i, j) - p(i, j)[c^t(i)N_C^t(i, j) + N_J^t(i, j) \\ & + \sum_k c^4(k)N_X^t(i, j, k)])^2 / N_{\text{data}}^t(i, j) \end{aligned}$$

$$+ \sum_{i=1}^4 \sum_{j=i}^4 [N_{\text{data}}^{\text{untag}}(i, j) - p(i, j) N_{\text{MC}}^{\text{untag}}(i, j)]^2 / N_{\text{data}}^{\text{untag}}(i, j). \quad (6.1)$$

The  $c^t(i)$  is the main floating parameter in the fit, which is the charm jet mis-tagging scale factor at working point  $t$  of  $p_T$  bin labelled  $i$ . The other main floating parameter is  $p(i, j)$  that is the normalisation factor scaling the MC to data. The  $N_{\text{data}}^t(i, j)$  is the number of data events with a tagged  $W$ -jet in the  $p_T$  bin labelled  $i$  and the other (untagged)  $W$ -jet in the  $p_T$  bin labelled  $j$ . Similarly the  $N_C^t(i, j)$  is the number of MC events with a tagged  $W$ -jet while the tagged  $W$ -jet is indeed a  $c$ -jet which can be seen as “signal”. In contrast the  $N_J^t(i, j)$  is the number of events with neither the tagged  $W$ -jet nor the top jets are  $c$ -jets and the  $N_X^t(i, j, k)$  is the number of events with one of the top jets is a  $c$ -jet. These two types of events can be seen as “background”. The later case is slightly more complicated, as the  $c$ -jet lies in a different  $p_T$  bin to the tagged jet, denoted as  $k$ , and it only depends on the  $c^4(k)$  (which is the scale factor of the 4th working point i.e. 60% ) as the top jets are tagged at 60% working point. The calibration is then given as the scale factors of the four working points in bins of  $p_T$  defined in the above text.

### 6.3 Alternative Monte Carlo samples

The events that are used in this study originate mostly due to  $t\bar{t}$  production. This process is produced with settings defined in section 5.3.3. While all samples were produced using the ATLAS simulation infrastructure and GEANT4 software, this process is also produced with fast simulation, using the same generators and settings. This alternative sample is used to estimate the uncertainty due to different choice of simulation.

For the nominal setting, the  $t\bar{t}$  process is modelled using the  $h_{\text{damp}}$  parameter<sup>1</sup> set to  $1.5 m_{\text{top}}$  [135]. An alternative sample is produced varying the  $h_{\text{damp}}$  and the renormalisation and factorisation simultaneously, and choosing the Var3c up/down variants of the A14 tune. This alternative sample is produced in order to estimate the uncertainty due to ISR. While the impact of the FSR is estimated by varying the renormalisation scale for emissions from the parton shower by a factor of two up or down.

Another alternative sample with replacing the PYTHIA 8.230 interface to the HERWIG 7.04, while keeping the same generator (POWHEGBOX v2) and same parameters for the generation. The HERWIG 7.04 is used together with the H7UE tune and the MMHT2014LO. This alternative sample is produced to estimate the uncertainty due different choice of parton shower and hadronisation model.

---

<sup>1</sup>The  $h_{\text{damp}}$  parameter is a resummation damping factor and one of the parameters that controls the matching of POWHEG matrix elements to the parton shower and thus effectively regulates the high- $p_T$  radiation against which the  $t\bar{t}$  system recoils.

All the alternative samples are generated using the fast simulation.

## 6.4 Kinematic Likelihood Fitter

The simulated  $t\bar{t}$  events are split according to the origin of  $W$  jets. The notation “ $t\bar{t}, \text{ll}$ ” denotes that both  $W$  jets are light flavour jets. Similarly, “ $t\bar{t}, \text{cl}$ ” (“ $t\bar{t}, \text{bl}$ ”) indicates that one of the  $W$  jets is a  $c$ -jet ( $b$ -jet) whereas the other is a light flavour jet.  $W$  jets with origin other than what is discussed above fall into the category denoted by “ $t\bar{t}, \text{other}$ ”. This category includes events in which at least one of the  $W$  jets comes from a hadronically decaying  $\tau$ -lepton.

The four-vectors of the four highest  $p_T$  jets, the lepton and the event  $E_T^{\text{miss}}$  are used as inputs to a likelihood-based  $t\bar{t}$  event reconstruction algorithm, which is described in more detail in Ref. [134]. This algorithm uses a likelihood function to assign the four jets to the  $t\bar{t}$  decay topology. In particular, the algorithm assigns one jet to be the  $b$ -jet from the leptonically decaying top quark ( $t \rightarrow Wb \rightarrow \ell vb$ ), another to the  $b$ -jet from the hadronically decaying top quark ( $t \rightarrow Wb \rightarrow qq'b$ , where  $qq'$  are the quarks in which the  $W$  boson decays) and the remaining two jets to the jets that come from the hadronic  $W$  boson decay. The jet assignment does not use any  $b$ -tagging information to avoid bias.

## 6.5 Maximising likelihood

Taking only four jets in the event limits the total number of possible jet orderings (permutations) in the event. In the semi-leptonic channel, four jets can be permuted a total number of times equal to  $4! = 24$ . However, the two  $W$  jets are kinematically indistinguishable. This reduces the possible number of permutations to 12. For every combination of jet ordering, the likelihood is maximised over its free parameters, the energy of the four jets, the lepton energy and the three components of the momentum of the neutrino, and provides a value based on how closely the kinematic information from the reconstructed objects for a specific jet ordering resembles the expected kinematic behaviour of the decay of a Standard Model semi-leptonic  $t\bar{t}$  event. The likelihood therefore distinguishes the possible permutations on an event-by-event basis. The best permutation, given by the largest log-likelihood value, is adopted as the jet ordering for the event. An additional requirement of  $\log\text{-likelihood} > -48$  is placed on the output of the likelihood value for the chosen event permutation. An example of the distribution of log-likelihood of the best permutations is shown in Figure 6.2. In this figure, the data events are compared against the simulation. The majority of the events come from  $t\bar{t}$  production. There is only a very small fraction of events, which is denoted as “non  $t\bar{t}$ ” on the figure, that come from other processes like  $W$  or  $Z$  production in association with jets or single-top production.

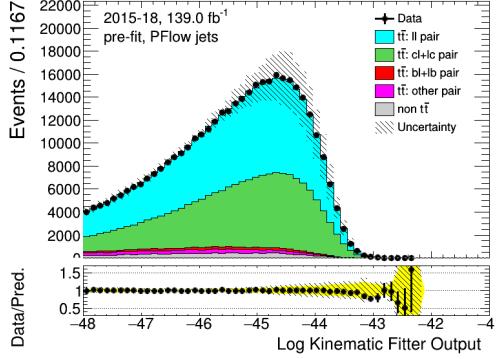


Figure 6.2: Distribution of the negative logarithm of the likelihood that is used to reconstruct the  $t\bar{t}$  decay.

## 6.6 Event selection

### 6.6.1 Standard selection

Events are required to contain exactly one trigger-matched lepton with  $p_T > 27$  GeV and exactly four jets with  $p_T > 25$  GeV. Leptons are required to have  $p_T$  above 27 GeV in order to avoid the turn-on curve for the single lepton triggers. Events which contain an additional lepton with  $p_T > 27$  GeV are rejected. The events are also required to have  $E_T^{\text{miss}} > 20$  GeV, which is assumed to be the result of the neutrino from the leptonically decaying  $W$  boson. The transverse mass  $m_T$  between the lepton and the  $E_T^{\text{miss}}$ , is constrained as follows:

$$m_T = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \Delta\phi)} > 40 \text{ GeV},$$

where  $\Delta\phi = \phi(E_T^{\text{miss}}) - \phi(\ell)$  is the azimuthal difference between the lepton and  $E_T^{\text{miss}}$ . The

	PFlow jets	Track jets
Data	227118	218351
$t\bar{t}$	$235670 \pm 200$	$223770 \pm 180$
Non $t\bar{t}$	$7610 \pm 120$	$7280 \pm 100$
Data/MC	$0.934 \pm 0.002$	$0.945 \pm 0.002$

Table 6.1: Standard selection: prefit comparison of the number of events in data and in simulation considering the PFlow jets and the VR-Track jets for events with exactly 4 jets.

yields of the data and the MC are given in Table 6.1. An example of the  $p_T$  distributions before any tagging or fitting and after the standard selection is shown in Figure 6.3. More plots can be found in Appendix A.1.1. The yellow band in the lower pad shows the overall systematic uncertainties, combining the experimental uncertainties and the  $t\bar{t}$  modelling uncertainties, as described in Section 6.7. The data/MC ratio shows good agreement within the systematic uncertainties.

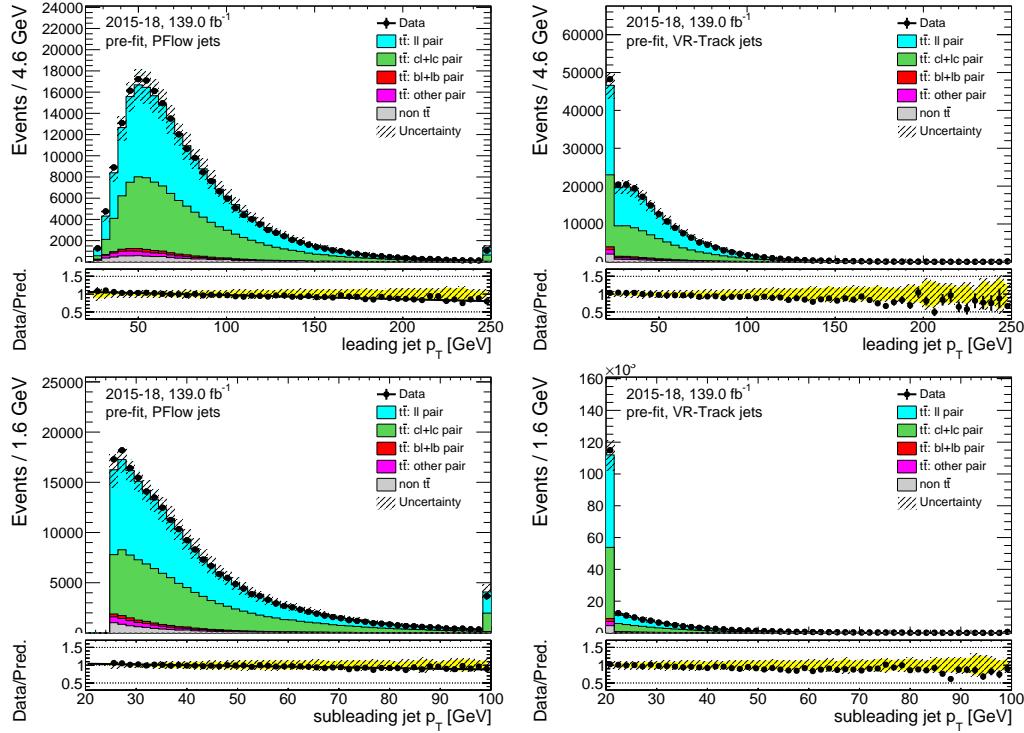


Figure 6.3: Standard selection: data versus simulation of the leading and sub-leading  $W$  jet  $p_T$  for the PFlow jets in the left column and for VR-Track jets in the right column. The leading jet and sub-leading jet refer to the highest  $p_T$   $W$  jet and the second highest  $p_T$  jet, respectively. The ‘non  $t\bar{t}$ ’ background indicates background comes from non- $t\bar{t}$  processes like  $W$  or  $Z$  production in association with jets or single-top production. The error in the table (and the following yields tables for different selection) is stats-only.

## 6.6.2 Low- $p_T$ selection

The author has developed an orthogonal selection to extend the calibration in the low- $p_T$  region so that the calibration can be applied to PFlow jets with  $20 < p_T < 25$  GeV. The  $p_T$  threshold of the VR-Track jets is 10 GeV therefore the low- $p_T$  selection is not needed. Instead of requiring events to have exactly 4 jets  $p_T > 25$  GeV, events are required to have exactly 3 jets with  $p_T > 25$  GeV and exactly 1 jet with  $25 \text{ GeV} > p_T > 20$  GeV. Other than that, all requirements for the selection are the same. This additional cut provides candidates for the PFlow  $W$  jet that is used for calibration in the  $20 - 25$  GeV region. The inclusive yields of the low- $p_T$  selection of the data and the MC are given in Table 6.2, and the  $p_T$  distributions of the  $W$  jets are shown in Figure 6.4. More plots of the kinematic distributions are shown in Appendix A.1.2. Good agreement between MC and data is shown in these distributions, and the  $p_T$  range of the sub-leading has gone down to 20 GeV.

	PFlow jets
Data	59987
$t\bar{t}$	$56530 \pm 90$
Non $t\bar{t}$	$3340 \pm 60$
Data/MC	$1.002 \pm 0.004$

Table 6.2: Low- $p_T$  selection: prefit comparison of the number of events in data and MC for the PFlow  $W$  jets. Events are required to have exactly 3 jets with  $p_T > 25$  GeV and one jet with  $20 < p_T < 25$  GeV.

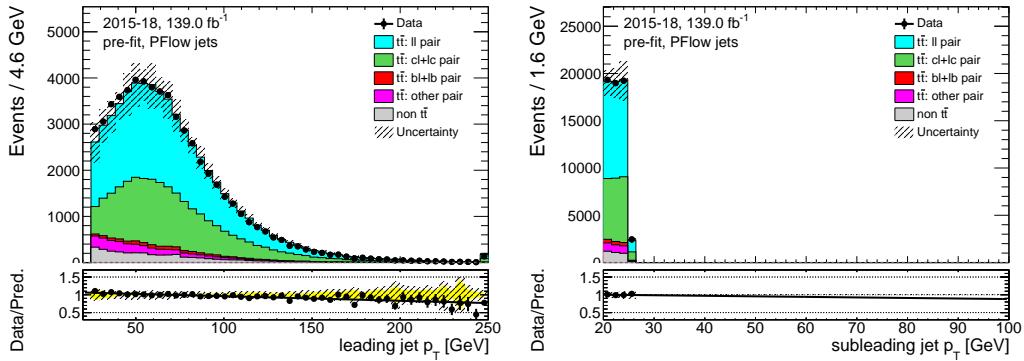


Figure 6.4: Low- $p_T$  selection: data versus simulation of the PFlow  $W$  jets  $p_T$ .

### 6.6.3 High- $p_T$ selection

It has been observed that in the previous calibrations that the statistics are relatively low for the high- $p_T$  region (e.g. jet  $p_T > 100$  GeV). Therefore, the author has worked on an orthogonal selection to improve this situation. Instead of requiring events to have exactly 4 jets, events are required to have at least 5 jets with  $p_T > 25$  GeV, in which at least 1 jet with  $p_T > 70$  GeV. Other than that, all requirements for the selection remain the same.

The choice of cut value at 70 GeV is based on the study shown in the following. The effect on the  $c$ -jet purity and the potential statistical gain is investigated, where the  $c$ -jet purity is defined as:

$$c \text{ jet purity} = \frac{N_{\text{true } c \text{ jet}}}{N_{\text{all}}}, \quad (6.2)$$

where  $N_{\text{true } c \text{ jet}}$  stands for the number of events with a true  $c$ -jet from the  $W$  decay, and  $N_{\text{all}}$  stands for the number of all events. The ideal situation is the high- $p_T$  selection will maximally increase the statistics while minimally decreasing the  $c$ -jet purity, therefore a figure of merit  $P^{\text{Cut}}$  is defined as:

$$P^{\text{Cut}} = \frac{\sum_i \text{Gain in stats}_i^2}{\sum_i c \text{ jet purity}_i^2},$$

where  $i$  stands for the number of bins. The “Gain in stats” stands for increase in statistics and it’s summed over all bins in Figure 6.5. The  $c$ -jet purity and the statistical gain are calculated for 4 different cut values as shown in Figure 6.5, comparing with the cut value of 0. The value of 70 GeV is chosen as it gives the highest value of  $P^{\text{Cut}}$ .

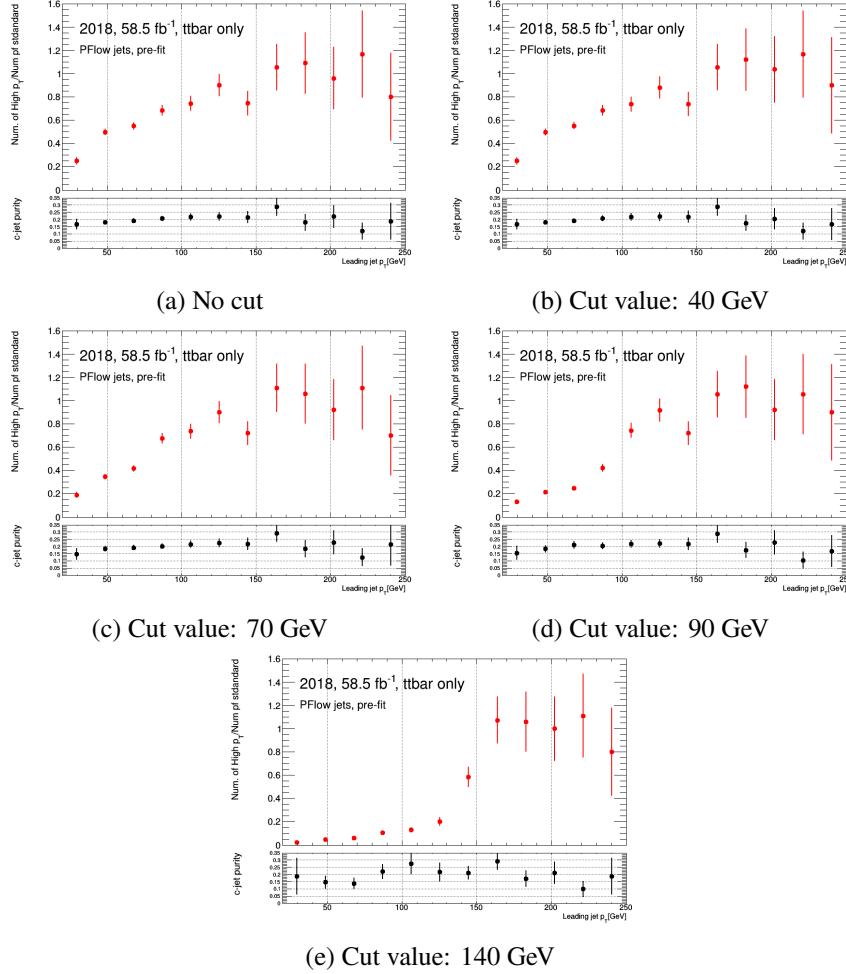


Figure 6.5: Comparison of different cut values in terms of gain in stats and  $c$ -jet purity.

The yields of the data and the MC are given in Table 6.3. An example of the  $p_T$  distributions before any tagging or fitting, applying the high- $p_T$  selection is shown in Figure 6.6. In general the event statistics improve about 80% in region with  $p_T > 70$  GeV as desired. More plots can be found in Appendix A.2.

#### 6.6.4 Combined selection

As the standard selections, low- $p_T$  selection and high- $p_T$  selection are orthogonal to each other, all the selections are combined to provide the maximum range and statistics for the calibration. The yields of the data and the MC are given in Table 6.4, an example of the

	PFlow jets	Track jets
Data	98273	83957
$t\bar{t}$	$99430 \pm 120$	$87476 \pm 110$
Non $t\bar{t}$	$1842 \pm 21$	$1570 \pm 20$
Data/MC	$0.97 \pm 0.003$	$0.94 \pm 0.003$

Table 6.3: High- $p_T$  selection: prefit comparison of the number of events in data and in simulation considering the PFlow  $W$  jets and the VR-Track jets.

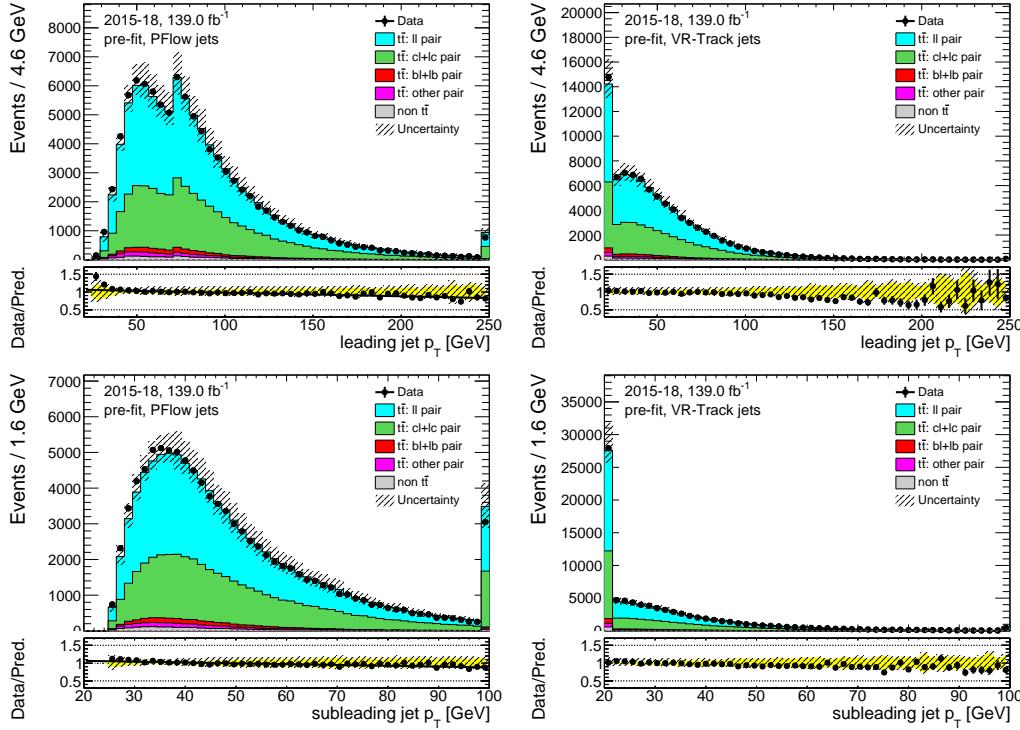


Figure 6.6: High- $p_T$  selection: data versus simulation of  $W$  jets  $p_T$  for PFlow jets in the left column and for VR-Track jets in the right column.

$p_T$  distributions before any tagging or fitting and after the combined selection is shown in Figure 6.7. More plots can be found in Appendix A.3.

## 6.7 Systematic uncertainties

The systematic uncertainties considered and propagated in this calibration can be broadly categorised into experimental and modelling systematic uncertainties.

### 6.7.1 Experimental uncertainties

TODO: refer to the analysis Part Experimental uncertainties are related to the detector and estimated using data-driven methods or MC simulations. The lepton energy scale and

	PFlow jets	Track jets
Data	385378	302308
$t\bar{t}$	$383520 \pm 230$	$302690 \pm 200$
Non $t\bar{t}$	$12420 \pm 120$	$8570 \pm 100$
Data/MC	$0.973 \pm 0.002$	$0.971 \pm 0.002$

Table 6.4: Combined selection: prefit comparison of the number of events in data and in simulation considering the PFlow jets and the VR-Track jets for an inclusive selection.

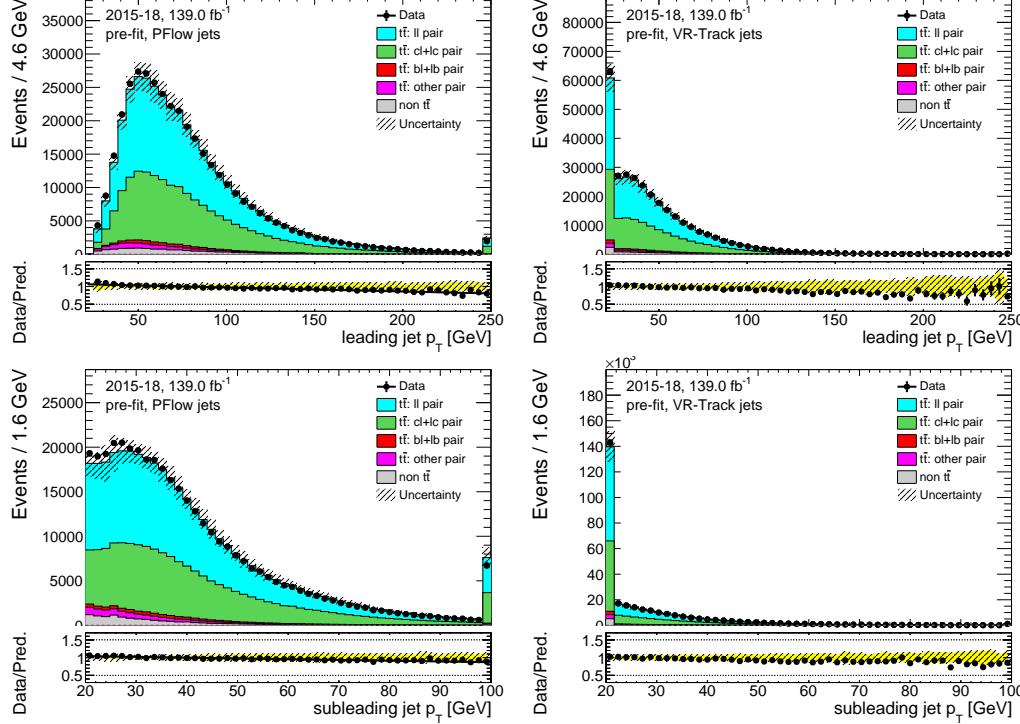


Figure 6.7: Combined selection: data versus simulation of  $W$  jets  $p_T$  for PFlow jets in the left column and for VR-Track jets in the right column.

resolution are corrected to provide better agreement between MC predictions and data, uncertainties due the corrections are considered. Uncertainties are taken into account on the electron and muon trigger, identification and reconstruction efficiencies, and for uncertainties associated with the isolation requirements.

The jet energy scale (JES) uncertainty depends on  $p_T$  and  $\eta$  and takes into account uncertainties due to pile-up effects. Uncertainties on the jet energy resolution (JER) are taken into account. Uncertainties on the energy scale and resolution of the electrons, muons, jets and taus are propagated to the calculation of the  $E_T^{\text{miss}}$ , which also has additional dedicated uncertainties on the scale, resolution, and reconstruction efficiency of tracks not associated to any of the reconstructed objects, along with the modelling of the underlying event. Uncertainties on the  $b$ -tagging (mis-tagging) probabilities for  $b$  (light) jets are

considered both for the tagging jets assigned to the  $b$  quark from the top decay and for the jets associated to the hadronically decaying  $W$  boson. Supporting material for this section can be found in the appendix, Tab.[A.1](#).

### 6.7.2 Modelling uncertainties

The uncertainty due to different choices of the parton shower models is estimated by comparing the MC samples generated with nominal parton shower model and with the alternative parton shower model. More specifically, it is derived by comparing the prediction from PowHEG interfaced either to PYTHIA8 or Herwig++. The uncertainty due to additional radiation in the initial state and the final state is estimated by comparing the nominal MC samples with the MC samples with alternative scale of renormalisation and factorisation. The uncertainty on modelling of initial state radiation (ISR) is assessed with two alternative PowHEG+PYTHIA8 samples. The samples include one with an increase in radiation which has the renormalisation and factorisation scales decreased by a factor of two and the  $hdamp$  parameter doubled (which controls the  $p_T$  of the first additional emission), while the sample with a decrease in radiation has the scales increased by a factor of two. In all cases, MC-to-MC SFs are taken into account. In addition, the uncertainty due to the variations samples being produced by fast simulation while the nominal samples being produced full simulation is also considered. The comparisons of the nominal  $t\bar{t}$  sample and the samples with each systematic uncertainty are shown in Table [6.5](#).

	PFlow jets		Track jets	
	Yields	Ratio of difference to nominal sample	Yields	Ratio of difference to nominal sample
$t\bar{t}$ Nominal Data/MC	$385378 \pm 230$ $0.973 \pm 0.002$		$302690 \pm 200$ $0.971 \pm 0.002$	
$t\bar{t}$ AF2 DATA/MC(AF2)	$386260 \pm 250$ $0.967 \pm 0.002$	0.716%	$304860 \pm 230$ $0.965 \pm 0.002$	0.716%
$t\bar{t}$ ISR DATA/MC(ISR)	$377130 \pm 220$ $0.989 \pm 0.002$	-1.665%	$297960 \pm 200$ $0.986 \pm 0.002$	-1.562%
$t\bar{t}$ Herwig DATA/MC(Herwig)	$331960 \pm 220$ $1.119 \pm 0.002$	-13.443%	$259940 \pm 190$ $1.126 \pm 0.002$	-14.123%

Table 6.5: Comparison of the number of events in data and in simulation considering the PFlow jets and the VR-Track jets for an inclusive selection. The uncertainty due to the variations samples being produced by fast simulation is included in the table as  $t\bar{t}$  AF2.

## 6.8 Under-estimation of $t\bar{t}$ + Heavy flavour background

Despite the fact that the true nature of most of the reconstructed  $W$  jets are either  $c$ -jets or light jets, there is still a very small amount of them are true  $b$ -jets.

There are two main sources of these true  $b$ -jets. The first is a  $W$  boson decays to a  $b$  and a  $c$  quark. The second is when the  $t\bar{t}$  plus a gluon process (referred to as  $t\bar{t}$  + heavy flavour process) is selected, and the gluon splits into a pair a  $b$  quarks and one of them is assigned as a  $W$  jet. The first source can be excluded by requiring no  $c$ -jets in the  $W$  jets, meaning the true  $b$ -jet in the  $W$  jets can only come from the  $t\bar{t}$  + heavy flavour process. This process is underestimated by the MC by about 30% for both the PFlow and VR-Track jets collections, as shown in Table 6.6 and Figure 6.8, where an extra cut requiring at least one  $W$  jet with  $DL1r > 8$  is added to the combined selection to reject most of the true  $c$ -jets and true light jet. A more thorough study is done in Ref. [136], where the mis-modelling factor is measured to be  $1.25 \pm 0.25$ , which is also consistent with the 30% mismodelling observed in the previous study. Therefore, events in the simulation in which the top jets and at least one of the  $W$  jets are  $b$ -jets (referred to as 3 true  $b$ -jets events), are scaled by  $1.25 \pm 0.25$ . All results shown in this chapter have this scale factor implemented, and the full difference between the simulation before applying this scale factor and after is taken as a systematic uncertainty. This uncertainty has been added in quadrature to the systematic uncertainties described in Section 6.7 in all the plots in this chapter.

	PFlow jets	VR-Track jets
Data	1589	1336
$t\bar{t}$	$1100 \pm 13$	$940 \pm 12$
Non $t\bar{t}$	$83 \pm 6$	$69 \pm 5$
Data/MC	$1.34 \pm 0.04$	$1.32 \pm 0.04$

Table 6.6: Yields of the 2018 data and MC of the combined selection, requiring at least 1 PFlow or track  $W$  jet with  $DL1r > 8$  to reject most of the light- and  $c$ -jets.

## 6.9 Results

### 6.9.1 Overview

Four rounds of calibrations have been carried out, containing different jet collections, Monte Carlo samples, analysis framework and  $b$ -jet identification algorithm. In the latest round, the calibration includes the PFlow jet and the VR-Track jet collection, and MV2c10, DL1 and DL1r taggers. The low- $p_T$  selection and the standard selection are carried out for all four calibrations, while the high- $p_T$  selection is only implemented in the latest calibration.

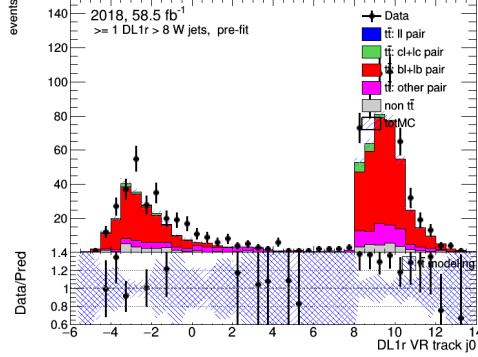


Figure 6.8: The DL1r score distribution of the leading VR-Track jet, requiring at least 1 VR-Track jets have  $DL1r > 8$  to reject most of the light and the  $c$  jets, with  $t\bar{t}$  modelling and statistical uncertainties.

### 6.9.2 $b$ -tagging algorithms output distribution

The distributions of the  $b$ -tagging algorithm' output of the MC and the data of the latest calibration (December 2020) are shown in Figure 6.9 for the PFlow jets and Figure 6.10 for the VR-Track jets, combining the standard selection, low  $p_T$  and the high- $p_T$  selection. In these figure, the data events are compared against the simulation. The majority of the events come from  $t\bar{t}$  production. There is only a very small fraction of non  $t\bar{t}$  events. The  $W$  jets pairs are mostly light jets pairs and  $c$ -jet light jet pairs, and a very small fraction of the pairs are  $b$ -jet light jet pairs or pairs containing one or more  $\tau$  hadron(s). The yellow band in the lower pad indicates the overall systematic uncertainties and the black band represents the  $t\bar{t}$  modelling systematic uncertainty, which dominates at low  $b$ -tagging discriminant (DL1 or  $DL1r < 4$ ). The experimental systematic uncertainty is in general very small. At high  $b$ -tagging discriminant (DL1 or  $DL1r > 4$ ), the uncertainty due to the  $1.25 \pm 0.25$  scale factor becomes more important.

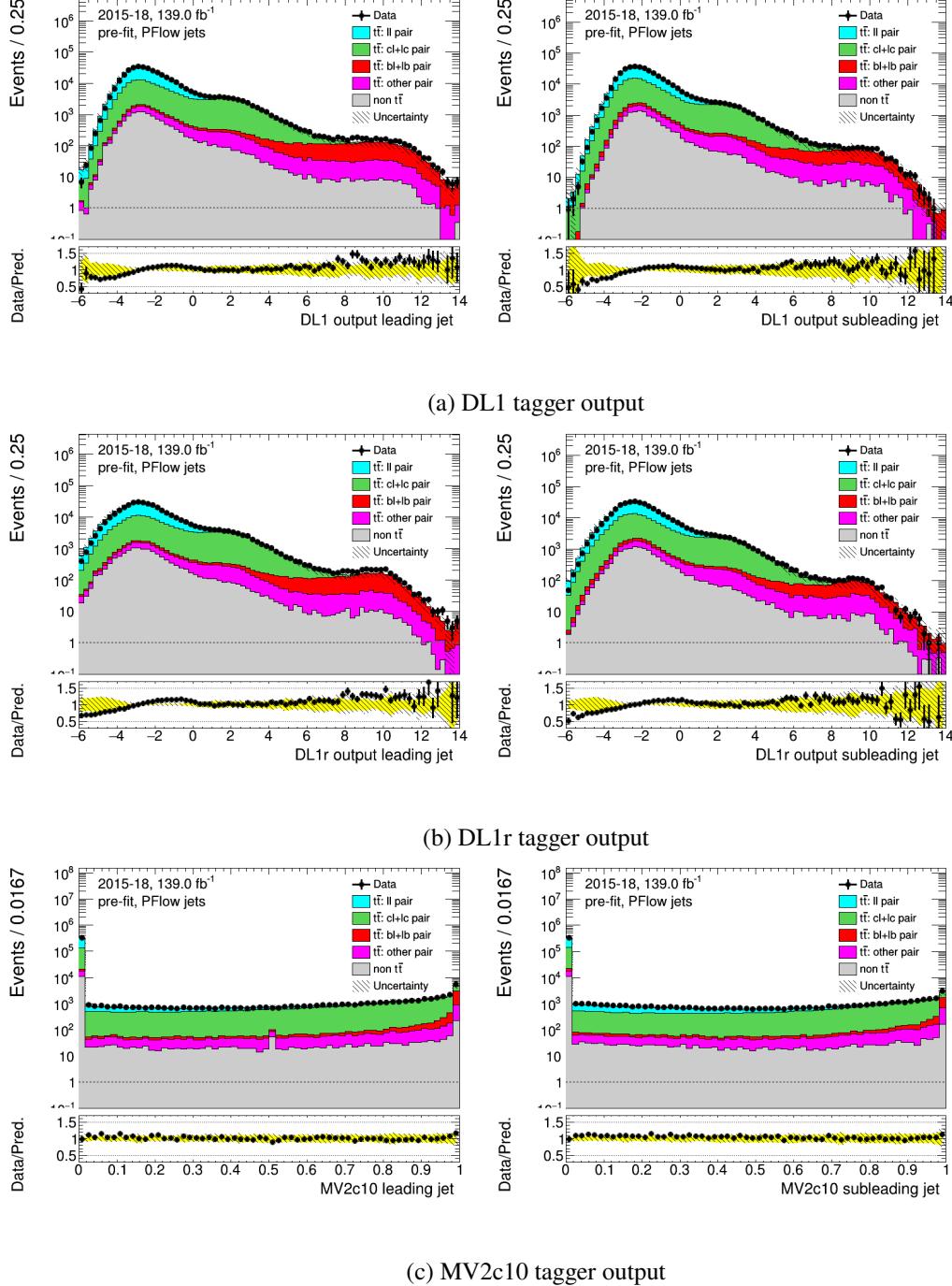


Figure 6.9: PFFlow jets: distributions of the DL1, DL1r and MV2c10 tagger outputs of the combined selection, leading jet in the left column and sub-leading jet in the right column, before fitting or tagging with full uncertainties.

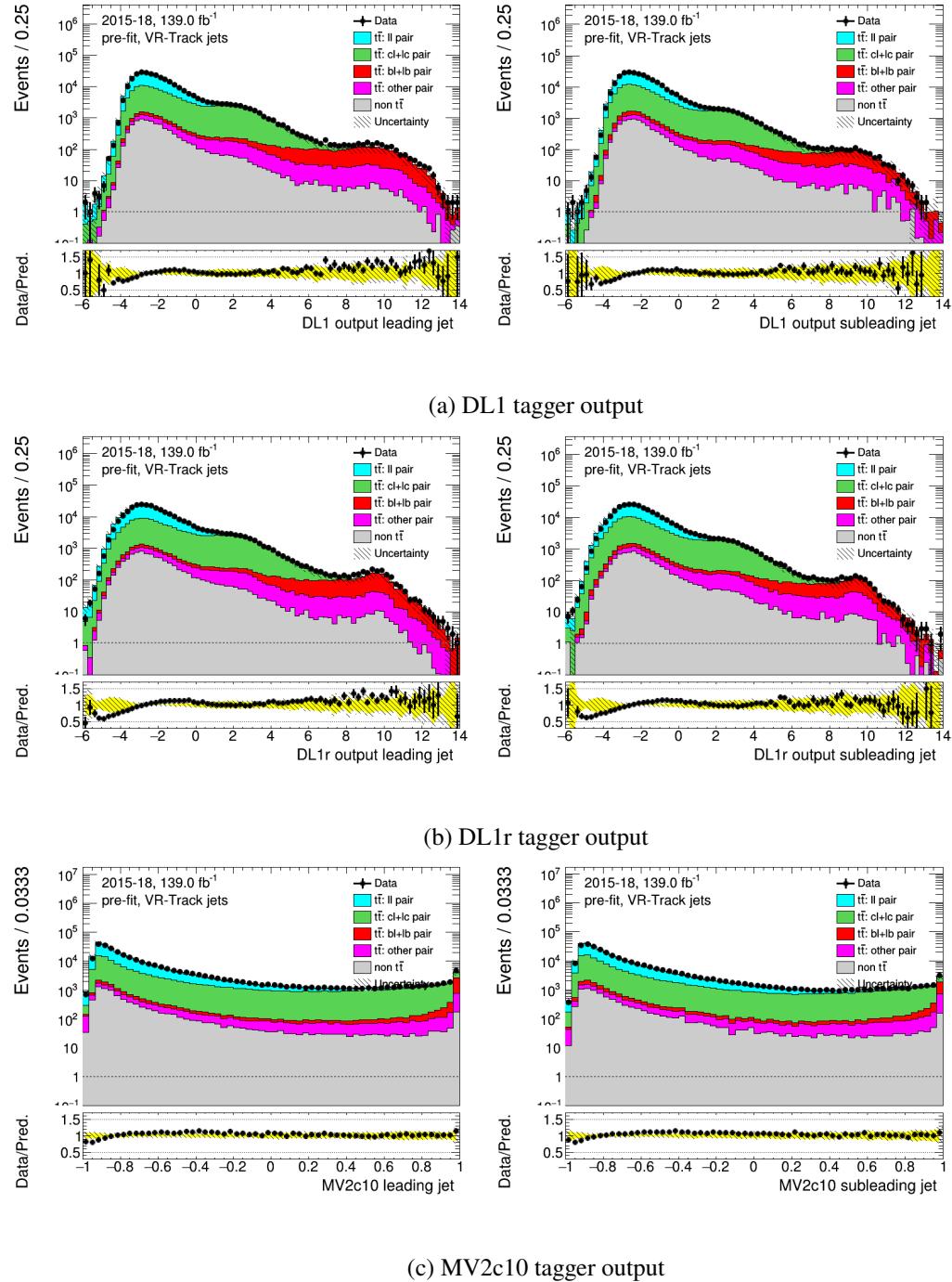


Figure 6.10: VR-Track jets: distributions of the DL1, DL1r and MV2c10 tagger outputs of the combined selection, leading jet in the left column and sub-leading jet in the right column, before fitting or tagging with full uncertainties.

### 6.9.3 Efficiencies and Scale Factors

The DL1 and DL1r  $c$ -jet efficiencies and scale factors with systematics uncertainties are calculated with four fixed cut working points for the PFlow and VR VR-Track jets collection in the latest derivation in December 2020.

The  $c$ -jet mis-tagging efficiencies are shown in Figure 6.11-6.14 for the PFlow jet collections and the VR-Track jets with the DL1 and the DL1r tagger. For PFlow jets, these results combine the standard selection, low- $p_T$  selection and the high- $p_T$  selection and for the VR-Track jets they combine the standard selection and the high- $p_T$  selection.

The  $1.25 \pm 0.25$  scale factor is applied on events with 3 true  $b$ -jets. The overall uncertainties are shown in the red band. The scale factors are shown in Figure 6.15-6.18 for the PFlow jets and the VR-Track jets with the DL1 and DL1r tagger. The tighter working points (60%, 70%) show larger uncertainties and bigger deviation from 1, while the looser working points (77%, 85%) have much smaller uncertainty and the simulation is able to recover the data well due to more abundant events statistics. For the PFlow jets, in most of the working points the systematic uncertainties dominate in the low- $p_T$  bins ( $p_T < 150$ ) and the statistical error, represented by the error bars on the markers, become more important in the last bin. For the VR-Track jets the statistical uncertainty is relatively constant for all bins while the systematic uncertainty increases as the  $p_T$  increases. To demonstrate the effect on statistics with the high- $p_T$  selection, the fractional statistical uncertainties of 60% working point scale factor are shown in Table 6.7 for the standard and the combined selection. In some bins the statistical uncertainty can decrease up to 30%, suggesting that the high- $p_T$  selection is successful at increasing events statistics.

	PFlow jets			VR-Track jets		
	Standard selection	High- $p_T$ selection	Fractional decrease	Standard selection	High- $p_T$ selection	Fractional decrease
Bin No.1	3.3%	3.3%	0.0%	5.6%	5.3%	5.7%
Bin No.2	3.1%	2.8%	10.7%	4.2%	3.7%	13.5%
Bin No.3	3.4%	2.6%	30.8%	5.8%	4.9%	18.4%
Bin No.4	12.1%	9.3%	30.1%	7.2%	5.6%	28.6%

Table 6.7: Comparison of the fractional statistical uncertainty in the DL1r 60% working point scale factor. The  $p_T$  range of each bin can be found in section 6.2.

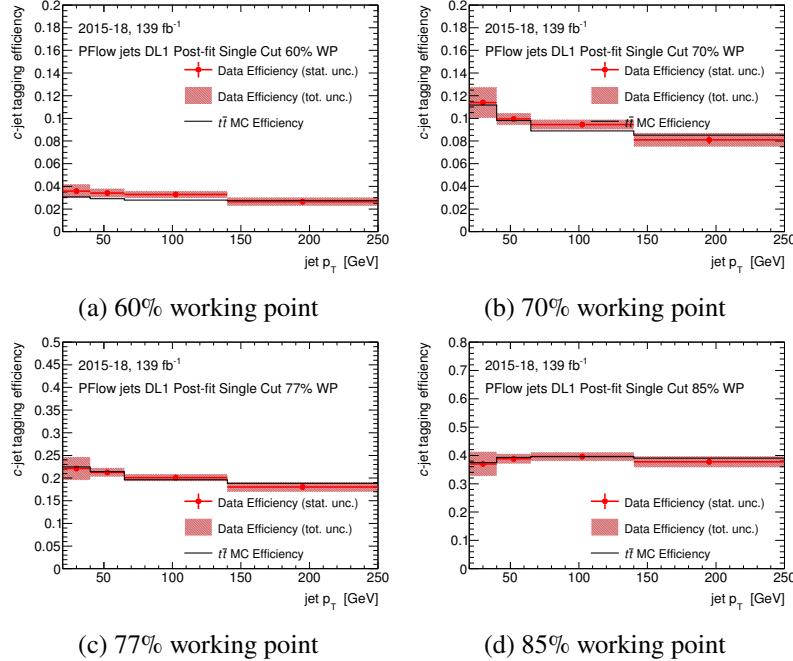


Figure 6.11: Charm-jet efficiencies for the PFlow jets collection with the DL1 tagger.

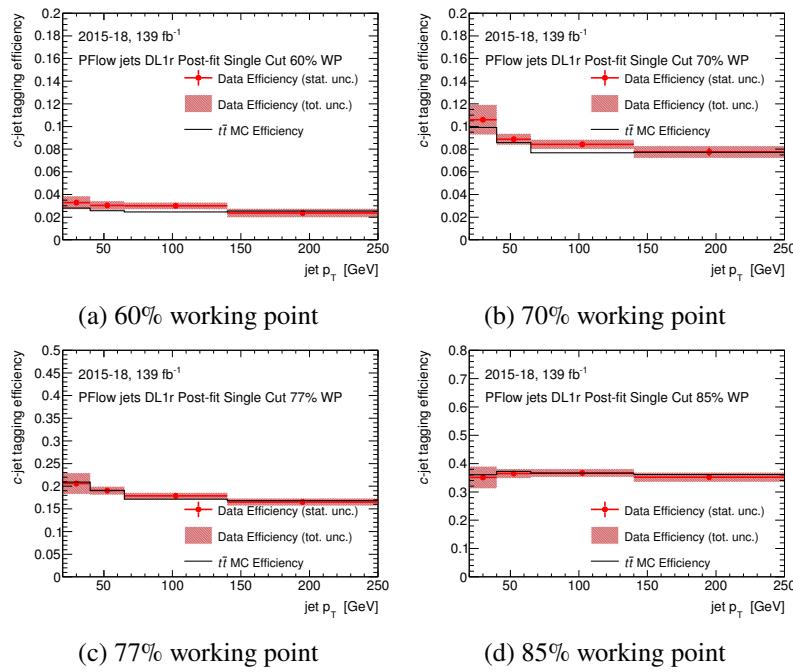


Figure 6.12: Charm-jet efficiencies for the PFlow jets collection with the DL1r tagger.

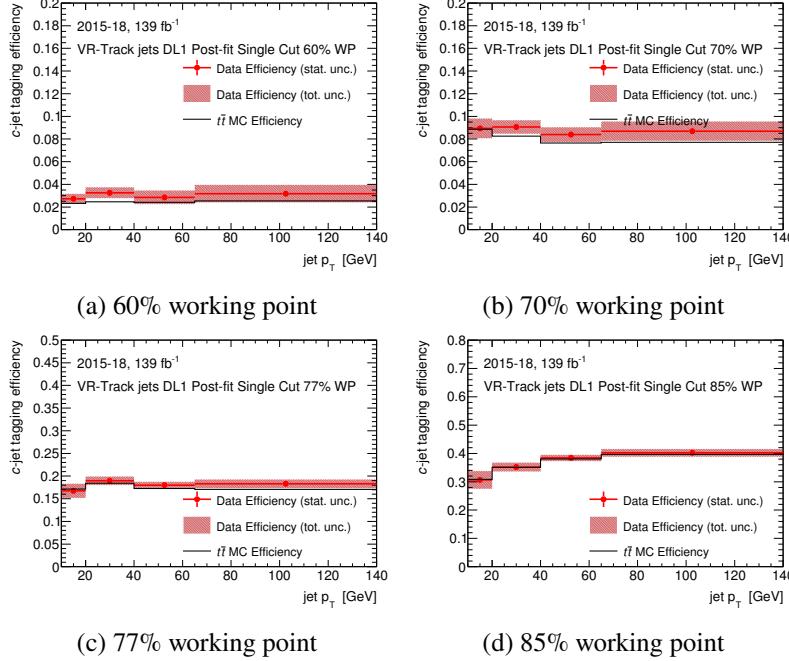


Figure 6.13: Charm-jet efficiencies for the VR-Track jets collection with the DL1 tagger.

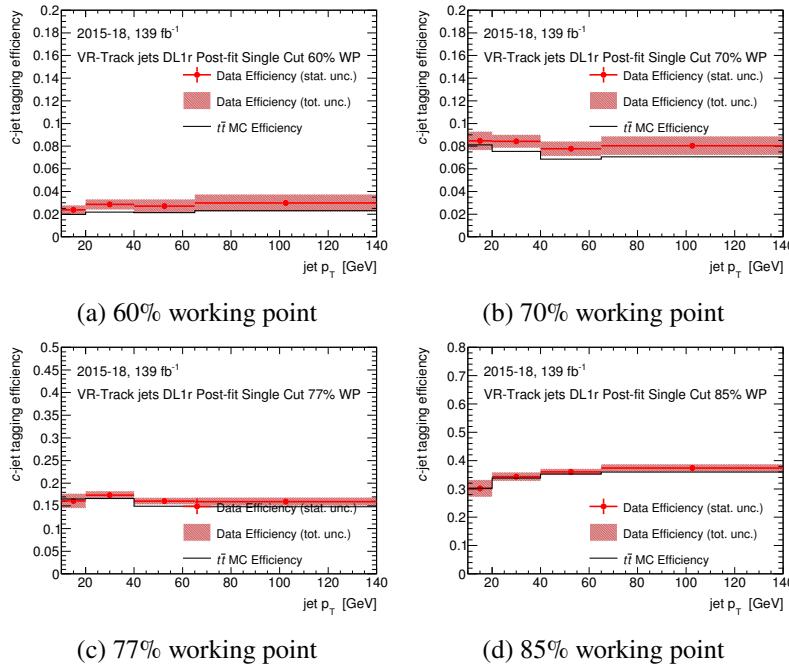


Figure 6.14: Charm-jet efficiencies for the VR-Track jets collection with the DL1r tagger.

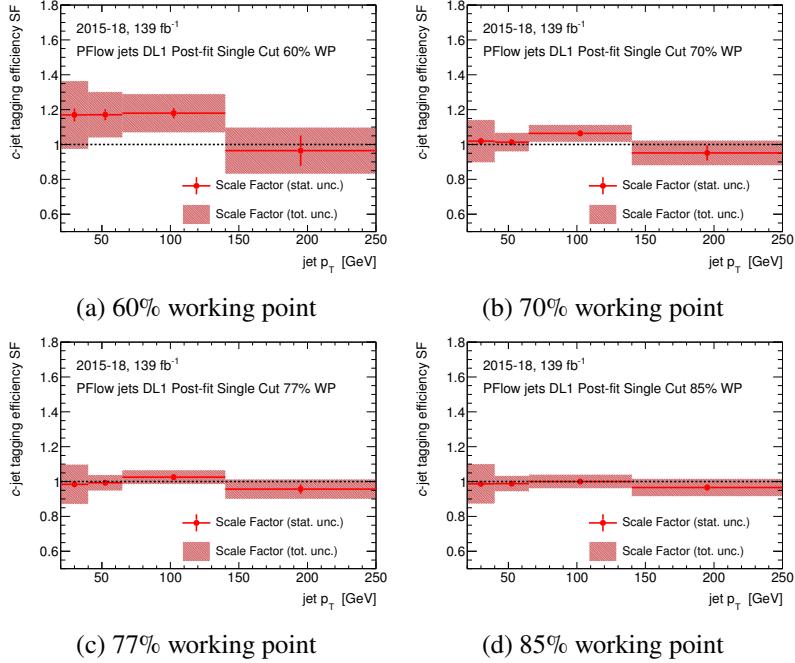


Figure 6.15: Charm-jet scale factors for the PFFlow jets collection with the DL1 tagger.

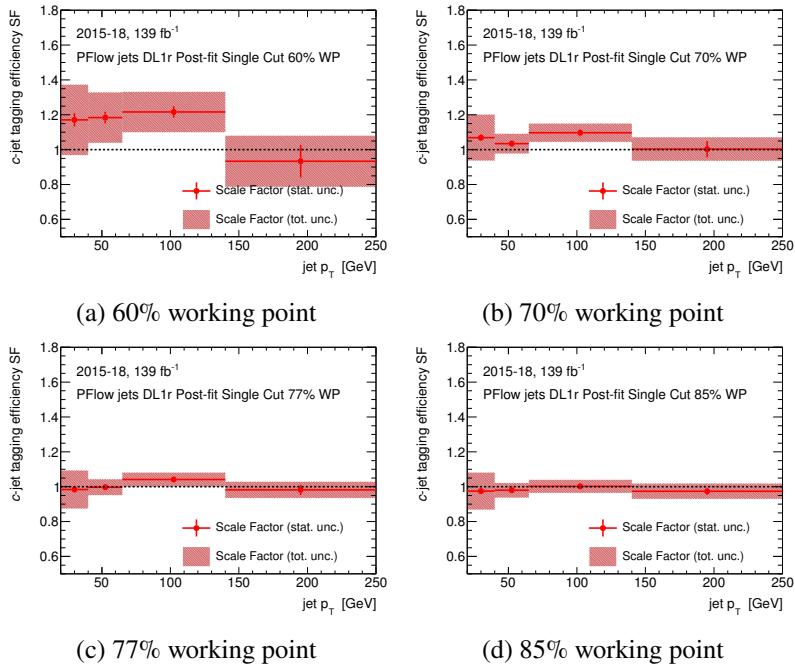


Figure 6.16: Charm-jet scale factors for the PFFlow jets collection with the DL1r tagger.

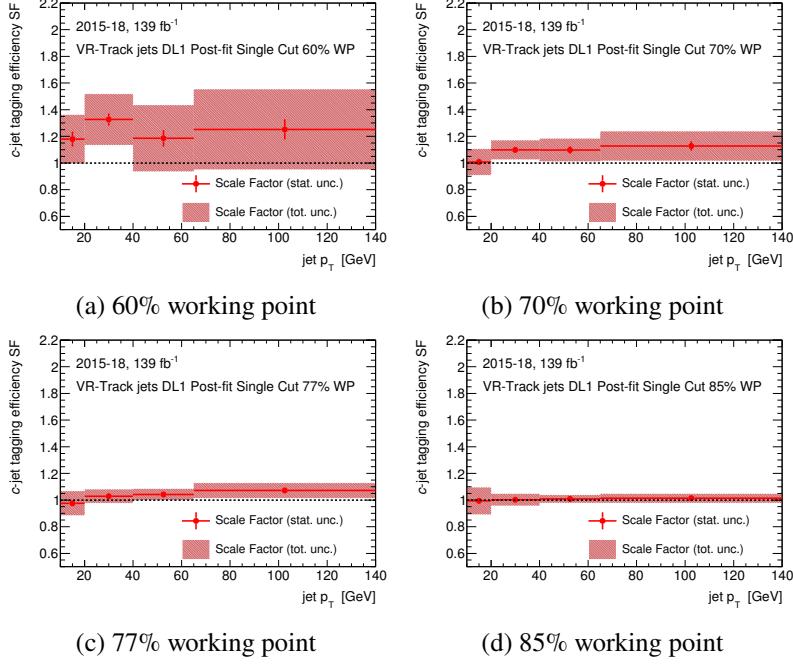


Figure 6.17: Charm-jet scale factors for the VR-Track jets collection with the DL1 tagger.

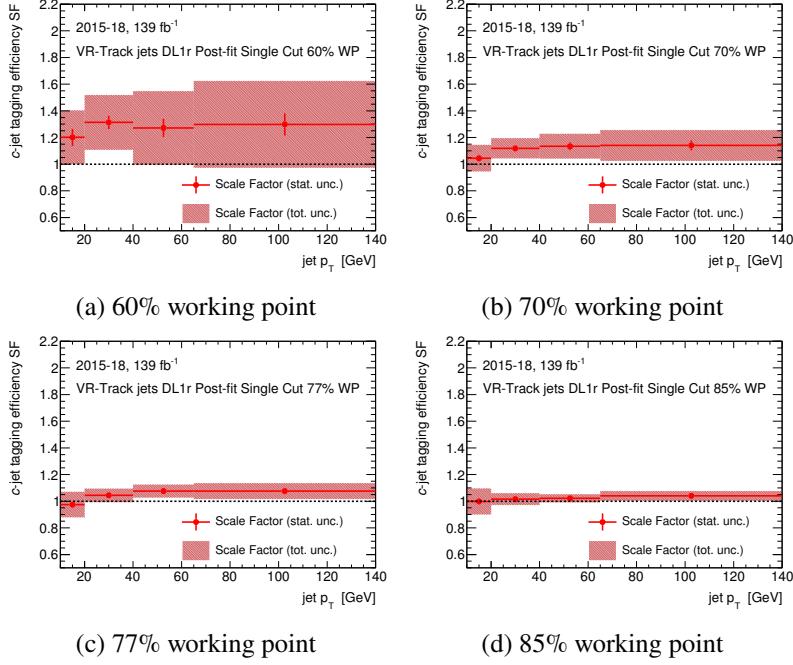


Figure 6.18: Charm-jet scale factors for the VR-Track jets collection with the DL1r tagger.

## Chapter 7

# Search for Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ channel

This chapter describes the search for Higgs boson pair production in the  $b\bar{b}\tau^+\tau^-$  channel, where one Higgs boson decays to a  $b$  quark pair and the other to a  $\tau$ -lepton pair. As the two  $\tau$ -leptons decay either leptonically or hadronically, the analysis is divided further into two sub-channels depending on their decay mode, the  $b\bar{b}\tau_{\text{lep}}^\pm\tau_{\text{had}}^\mp$  channel (for simplicity, referred to as  $\tau_{\text{lep}}\tau_{\text{had}}$  channel) where one of the **tau-leptons** decays leptonically and the other decays hadronically, and the  $b\bar{b}\tau_{\text{had}}^+\tau_{\text{had}}^-$  channel (or referred to as  $\tau_{\text{had}}\tau_{\text{had}}$  channel) where both of the **tau-leptons** decay hadronically. The decay mode of both **tau-leptons** decay leptonically is not considered in this analysis due to its insignificant contribution. In this thesis, the author will present his work in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel, and will also show the combination results with the  $\tau_{\text{had}}\tau_{\text{had}}$  channel.

The results of this search are interpreted in terms of resonant and non-resonant production of the di-Higgs. For the non-resonant production, upper limits are set on the SM di-Higgs production cross section, and exclusion limits are set on the Higgs self-coupling  $\lambda_{H\bar{H}H}$ . For the resonant search, upper limits are set on the resonance production cross section as a function of the resonance mass, constraining a model with an extended Higgs sector based on two Higgs doublet (2HDM) model (TODO: link to theory chapter). The  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  combination results are published in Reference [78] TODO: if paper is published cite the paper The  $\lambda_{H\bar{H}H}$  result is combined with other final states of the di-Higgs production including  $b\bar{b}b\bar{b}$  and  $b\bar{b}\gamma\gamma$ . The combination result is published in Reference [137] TODO: if paper is published cite the combination paper.

The analysis strategy is as follows: first, a set of selections are applied on the reconstructed physics objects and kinematics variables which defines the signal regions and the control region, as described in section 7.1.2; neural networks trained on the signal regions events are then used to extract the various signals, and the output of the algorithm is used as

the final discriminant, as described in section 7.3; the systematic uncertainties considered in this analysis is discussed in section 7.4.5; a profile likelihood fit is then performed simultaneously on all  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  signal regions and the control region, with all systematics uncertainties served as nuisance parameters. The statistical combination result of  $b\bar{b}\tau^+\tau^-$ ,  $b\bar{b}\gamma\gamma$  and  $b\bar{b}b\bar{b}$  is also studied. The setup and the results are shown in section 7.5.

The author's contributions to the analyses presented in this Chapter are as follows. As one of the main analysers in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel, the author produced all data and MC (including all signals and background) histograms with full experimental and theoretical systematics, which are the inputs to the likelihood fit. The author has derived the systematics uncertainties for the following major backgrounds:  $t\bar{t}$ ,  $Z + \text{HF}$  and single-top. The author has also contributed to the estimation of the fake- $\tau_{\text{had}}$  background and its uncertainties. Finally, the author implemented and validated a reweighting method in order to scan the non-resonant  $HH$  production with various  $\kappa_\lambda$  values, and derived the corresponding uncertainties.

## 7.1 Trigger and event Selection

### 7.1.1 Trigger selection

The ATLAS trigger system, as described in section 3.2.6, consists of the hardware-based level (L1) and the software-based level (HLT) to select characteristic events. Events in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel are recorded using a combination of single-lepton triggers (SLTs) and lepton-plus- $\tau_{\text{had}}$  triggers (LTTs). The offline objects are required to pass addition  $p_{\text{T}}$  cuts. For the leptons (*e or mu*), a threshold which is 1 GeV higher than the HLT  $p_{\text{T}}$  threshold is applied to reach its efficiency plateau region. For the  $\tau_{\text{had}}$ , the threshold is set at 5 GeV above the HLT  $p_{\text{T}}$  requirement for the same reason. The offline *e*, *mu* and  $\tau_{\text{had}}$  must be matched to the corresponding trigger objects.

The priority is given to SLT events if a lepton fulfils the offline lepton  $p_{\text{T}}$  requirements. A combination of various un-prescaled <sup>1</sup> single electron triggers (SET) and single muon triggers (MET) are used to maximise the efficiency. As the  $p_{\text{T}}$  threshold is raised for 1 GeV for the HLT  $p_{\text{T}}$  requirements, the single electron triggers (SET) require an event with an electron with  $p_{\text{T}} > 25$  or  $> 27$  GeV, whereas the single muon triggers (SMT) require an event with a muon with  $p_{\text{T}} > 21$  or  $> 27$  GeV, depending on the data-taking period.

The SLT triggers used are listed in Table 7.1. In this table, the following naming conventions are used: the trigger names begin with HLT, followed by the lepton type (*e* or *mu* for an electron or muon respectively) and the HLT  $p_{\text{T}}$  threshold. What comes

---

<sup>1</sup>The trigger rate is reduced by *prescaling* which randomly vetos events that pass the trigger.

Single Lepton Triggers (SLT)		
Period	Single Electron Triggers (SET)	Single Muon Triggers (SMT)
2015	HLT_e24_lhmedium_L1EM20VH	HLT_mu20_iloose_L1MU15
	HLT_e60_lhmedium	HLT_mu50
	HLT_e120_lhloose	
2016 & 2017 & 2018	HLT_e26_lhtight_nod0_ivarloose	HLT_mu26_ivarmedium
	HLT_e60_lhmedium_nod0	HLT_mu50
	HLT_e140_lhloose_nod0	

Table 7.1: SLT triggers used in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel, along with the trigger-dependent offline  $p_T$  thresholds, for each year/period are shown.

next is the identification WP (loose, medium or tight), prefixed with ‘lh’ which stands for likelihood-based trigger. The suffix of ‘nod0’ indicates that no impact parameter cuts are applied. For SMT, no identification is applied therefore it is not specified. The following is the isolation requirement, specified by ‘i’ and the working point (varloose, varmedium). If not specified, no isolation requirement is applied. Finally, the suffix which starts ‘L1EM’ (‘L1MU’) indicates the object is seeded with L1 electromagnetic (muonic) trigger items, with  $p_T$  threshold specified by the number after; the ‘VH’ stands for the  $p_T$  threshold varies with eta to account for energy loss and hadronic core isolation is applied. If not specified, the L1 seed is using the default setting.

The LTT triggers used are listed in Table 7.1. For the LTT triggers, an electron (muon) is required with  $p_T > 18$  (15) GeV, together with a  $\tau_{\text{had}}$  with  $p_T > 30$  GeV or 40 GeV depending on the period. Jets are cut at 80 GeV (45 GeV) for Level-1 trigger  $p_T$ -thresholds of 25 GeV (12 GeV) (the L1 jets  $p_T$  requirements are explained in the following). The same naming conventions are used for the LTT triggers. In addition, the trigger names are expanded by the ‘tau’ which stands for  $\tau_{\text{had}}$  followed by its  $p_T$ , identification requirements and pre-selection settings. Again, unless specified, the default L1 seeds are used. For dedicated L1 seeds, the lepton (‘EM’ for electron or MU for muon) followed by  $p_T$  requirements and specific settings. After the lepton, the number of  $\tau_{\text{had}}$  is specified by a number before ‘TAU’, followed by a number indicating the  $p_T$  threshold. For example, ‘2TAU12I’ stands for two  $\tau_{\text{had}}$  with  $p_T > 12$ , with electromagnetic isolation applied (‘I’). Finally, the number of jets and the  $p_T$  requirements are specified in the same fashion. If unspecified, at least one jets with  $p_T > 25$  GeV are required (the offline cut is raised to 80 GeV).

### 7.1.2 Event pre-selection

Before passing the events to the MVA algorithm, they are required to pass a loose pre-selection criteria to select events consistent with the signal topology and to remove obvious background. Events passing the pre-selection criteria is referred to as signal region (SR) events. Events are required to contain exactly one electron or muon, an oppositely charged

Lepton Tau Triggers (LTT)	
Electron Tau Triggers (ETT)	
HLT_e17_lhmedium_nod0_tau25_medium1_tracktwo	
HLT_e17_lhmedium_nod0_ivarloose_tau25_medium1_tracktwo	
HLT_e17_lhmedium_nod0_ivarloose_tau25_medium1_tracktwo_L1EM15VHI_2TAU12IM_4J12	
HLT_e17_lhmedium_nod0_ivarloose_tau25_medium1_tracktwoEF	
HLT_e17_lhmedium_nod0_ivarloose_tau25_medium1_tracktwoEF_L1EM15VHI_2TAU12IM_4J12	
HLT_e17_lhmedium_nod0_ivarloose_tau25_mediumRNN_tracktwoMVA	
HLT_e17_lhmedium_nod0_ivarloose_tau25_mediumRNN_tracktwoMVA_L1EM15VHI_2TAU12IM_4J12	
Muon Tau Triggers (MTT)	
HLT_mu14_tau25_medium1_tracktwo	
HLT_mu14_ivarloose_tau25_medium1_tracktwo	
HLT_mu14_ivarloose_tau25_medium1_tracktwo_L1MU10_TAU12IM_3J12	
HLT_mu14_ivarloose_tau25_medium1_tracktwoEF_L1MU10_TAU12IM_3J12	
HLT_mu14_ivarloose_tau25_mediumRNN_tracktwoMVA_L1MU10_TAU12IM_3J12	
HLT_mu14_ivarloose_tau35_medium1_tracktwo	
HLT_mu14_ivarloose_tau35_medium1_tracktwoEF	
HLT_mu14_ivarloose_tau35_mediumRNN_tracktwoMVA	

Table 7.2: LTT triggers used in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel. For different periods, a mixture of triggers from this list are used. Table reproduced from analysis internal notes.

$\tau_{\text{had}}$  and exactly two  $b$ -tagged jets. The selected electron (muon) must pass a tight (medium) identification requirement with an efficiency of around 80% (97%). If more than one primary vertex is present in the events, the one with the highest ( $p_T^{\text{track}}$ )<sup>2</sup> is chosen. To reject background events from low-mass Drell-Yan events, the invariant mass of the  $\tau$ -lepton pair ( $m_{\tau\tau}^{\text{MMC}}$ ), is required to be above 60 GeV, which is estimated from the four-momenta of the electron or muon, the  $\tau_{\text{had}}$  and the  $E_{\text{T}}^{\text{miss}}$  using the Missing Mass Calculator (MMC) [138]. In the calculation, the  $E_{\text{T}}^{\text{miss}}$  is assumed to be exclusively from the neutrinos produced in the  $\tau$ -lepton decays. To reject the  $t\bar{t}$  events, the  $b$ -tagged jet pair invariant mass ( $m_{bb}$ ) is required to be less than 150 GeV. And therefore, by reversing this cut, a  $t\bar{t}$ -enriched region can be defined which is used in the estimation of  $t\bar{t}$  backgrounds as described in Section 7.2.1. A  $\tau_{\text{had}}$  with  $p_{\text{T}} > 20$  GeV and  $|\eta| < 2.3$  is required in the SLT category. A  $\tau_{\text{had}}$  with  $p_{\text{T}} > 30$  GeV, or higher if required by the LTT triggers (as defined in the previous section 7.1.1), and  $|\eta| < 2.3$  is required in the LTT events. In both categories, the (sub-)leading  $b$ -tagged jet must have  $p_{\text{T}} > 45$  (20) GeV, in addition to any trigger-dependent jet requirements of LTT defined in the previous section. The full event pre-selection is summarised in Table 7.3.

The rate of events accepted by the detector is quantified by the *acceptance*, and the rate of events selected by the analysis selection is quantified by the *selection efficiency*. These two rates are usually multiplied together ( $A \times \epsilon$ ) to quantify the rate of a simulated event accepted by the detector and the analysis selection. The cumulative  $A \times \epsilon$  in each step of the pre-selection is summarised in Table 7.4 and Table 7.5 for the SLT and LTT channels, respectively, for the non-resonant signal and three example resonances mass points.

$\tau_{\text{lep}}\tau_{\text{had}}$ categories	
SLT	LTT
<b><math>e/\mu</math> selection</b>	
Exactly one tight $e$ or medium $\mu$	
$p_T^e > 25, 27 \text{ GeV}$	$18 \text{ GeV} < p_T^e < \text{SLT cut}$
$p_T^\mu > 21, 27 \text{ GeV}$	$15 \text{ GeV} < p_T^\mu < \text{SLT cut}$
$ \eta^e  < 2.47, \text{ not } 1.37 <  \eta^e  < 1.52$	
$ \eta^\mu  < 2.7$	
<b><math>\tau_{\text{had}}</math> selection</b>	
One loose $\tau_{\text{had}}$	
$ \eta  < 2.3$	
$p_T > 20 \text{ GeV}$	$p_T > 30 \text{ GeV}$
<b>Jet selection</b>	
$\geq 2$ jets with $ \eta  < 2.5$	
$p_T > 45 (20) \text{ GeV}$	Trigger dependent
<b>Event-level selection</b>	
Trigger requirements passed	
Collision vertex reconstructed	
$m_{\tau\tau}^{\text{MMC}} > 60 \text{ GeV}$	
Opposite-sign electric charges of $e/\mu/\tau_{\text{had-vis}}$ and $\tau_{\text{had}}$	
Exactly two $b$ -tagged jets	
$m_{bb} < 150 \text{ GeV}$	

Table 7.3: Summary of the event preselections, shown separately for the SLT and LTT. Thresholds on the (sub-)leading  $p_T$  object are given outside (within) parentheses. The possible values of the requirements in the SLT are separated by commas which depends on the year of the data-taking. For the jet selection in the LTT channel multiple selection criteria are used. The trigger  $p_T$  thresholds shown correspond to the offline requirements. Table reproduced from Ref. [78].

Selection	Non-resonant signal		Resonant signal		
	ggF $HH$	VBF $HH$	(300 GeV)	(500 GeV)	(1000 GeV)
Basic selection	19%	16%	13%	22%	30%
Trigger	12%	9.2%	6.0%	14%	22%
Object selections	9.7%	7.2%	5.0%	11%	20%
Trigger specific offline $p_T$ cuts	9.5%	6.8%	4.4%	11%	20%
Opposite-charged $\tau$ and lepton	9.4%	6.6%	4.4%	11%	20%
Two $b$ -tagged jets	4.4%	2.7%	1.8%	4.9%	10%
$m_{\tau\tau}^{\text{MMC}} > 60 \text{ GeV}$	4.3%	2.7%	1.8%	4.8%	9.7%
$m_{bb} < 150 \text{ GeV}$	4.1%	2.6%	1.7%	4.6%	9.2%

Table 7.4: Cumulative  $A \times \epsilon$  for simulated signal events to pass each stage of the event pre-selection in the SLT channel. The efficiencies are calculated with respect to  $HH \rightarrow b\bar{b}\tau^+\tau^-$  decays in which one  $\tau$ -lepton decays hadronically and one decays leptonically. The ‘Basic selection’ includes selections with at least one  $\tau_{\text{had}}$  candidate and one lepton passing loose kinematic requirements. The ‘Object selections’ requires exactly one  $\tau_{\text{had}}$  candidate, at least two jets with  $p_T > 25 \text{ GeV}$  and  $|\eta| < 2.5$ . The ‘Trigger specific offline  $p_T$  cuts’ are cuts placed on the  $p_T$  of the reconstructed jet or  $\tau_{\text{had}}$  that are geometrically matched to the HLT objects, to ensure the efficiencies of the HLT objects reach the plateau region.

Selection	Non-resonant signal		Resonant signal		
	ggF $HH$	VBF $HH$	(300 GeV)	(500 GeV)	(1000 GeV)
Basic selection	19%	16%	13%	22%	30%
Trigger	3.4%	3.2%	2.8%	4.0%	3.4%
Object selections	2.9%	2.3%	2.2%	2.8%	2.4%
Trigger specific offline $p_T$ cuts	2.3%	1.8%	1.4%	2.6%	2.3%
Opposite-charged $\tau$ and lepton	2.3%	1.8%	1.4%	2.6%	2.3%
Two $b$ -tagged jets	1.1%	0.76%	0.61%	1.2%	1.1%
$m_{\tau\tau}^{\text{MMC}} > 60 \text{ GeV}$	1.0%	0.75%	0.61%	1.2%	1.1%
$m_{bb} < 150 \text{ GeV}$	0.98%	0.71%	0.57%	1.1%	1.0%



Table 7.5: Cumulative  $A \times \epsilon$  for simulated signal events to pass each stage of the event pre-selection in the LTT channel. The efficiencies are calculated with respect to  $HH \rightarrow b\bar{b}\tau^+\tau^-$  decays in which one  $\tau$ -lepton decays hadronically and one decays leptonically. The ‘Basic selection’ includes selections with at least one  $\tau_{\text{had}}$  candidate and one lepton passing loose kinematic requirements. The ‘Object selections’ requires exactly one  $\tau_{\text{had}}$  candidate, at least two jets with  $p_T > 25 \text{ GeV}$  and  $|\eta| < 2.5$ . The ‘Trigger specific offline  $p_T$  cuts’ are cuts placed on the  $p_T$  of the reconstructed jet or  $\tau_{\text{had}}$  that are geometrically matched to the HLT objects, to ensure the efficiencies of the HLT objects reach the plateau region.

The  $A \times \epsilon$  for all resonances mass points are shown in Figure 7.1. The decrease in  $A \times \epsilon$  for  $m_X$  greater than about 1000 GeV is due to the Lorentz boost of the Higgs bosons causing their decay products to become highly collimated more often, and therefore harder to identify.

After the pre-selection, the event yields of the data, signal and background and the corresponding statistical error is shown in Table. 7.6 for the SLT channel, and Table. 7.7 for the LTT channel. The ‘Fake’ background in the table represents the background due to a jet faking a  $\tau_{\text{had}}$ , which is estimated by a data-driven background described in details in section 7.2.1.

### 7.1.3 Anti- $\tau_{\text{had}}$ selection

In order to provide fake- $\tau_{\text{had}}$ -enriched regions used for background estimation, an “anti- $\tau_{\text{had}}$ ” selection is defined: those  $\tau_{\text{had-vis}}$  objects that fail the RNN loose  $\tau_{\text{had-ID}}$  and have an RNN score greater than 0.01 are labelled as anti- $\tau_{\text{had}}$  candidates. The RNN cut is used and recommended by the Fake-Tau-Task-Force [139]. The minimum RNN score cut can make sure that the jet would have some features of the true  $\tau_{\text{had}}$ , and that the composition of the jet (either quark- or gluon-initiated) would be more similar to that in the SR.



Exactly one anti- $\tau_{\text{had}}$  object is selected when there are no  $\tau_{\text{had}}$  passing the offline  $\tau_{\text{had-ID}}$  requirement. This is to make sure only one  $\tau_{\text{had}}$  objects (either true  $\tau_{\text{had}}$  or anti- $\tau_{\text{had}}$ ) is selected. For the LTT channel where  $\tau_{\text{had-ID}}$  is applied at trigger level (more details in Section 7.1.1), only the anti- $\tau_{\text{had}}$  object that is matched to the trigger  $\tau_{\text{had}}$  is considered,



SampleName	Entries	Integral	Error	Error/Integ.
<b>Signal Samples</b>				
ggF Non-resonant	151522	5.872	0.018	0.3%
VBF Non-resonant	36457	0.2002	0.0013	0.6%
Resonant at $m_X = 251$	22541	60.6	0.4	0.7%
Resonant at $m_X = 260$	22252	60.9	0.4	0.7%
Resonant at $m_X = 280$	24595	67.9	0.4	0.7%
Resonant at $m_X = 300$	18398	77.2	0.6	0.8%
Resonant at $m_X = 325$	20904	90.9	0.6	0.7%
Resonant at $m_X = 350$	23593	106.4	0.7	0.7%
Resonant at $m_X = 400$	29280	140	0.8	0.6%
Resonant at $m_X = 450$	34665	175.3	1	0.6%
Resonant at $m_X = 500$	21332	210.7	1.5	0.7%
Resonant at $m_X = 550$	23571	240.9	1.6	0.7%
Resonant at $m_X = 600$	25917	273.9	1.8	0.6%
Resonant at $m_X = 700$	29596	327	2	0.6%
Resonant at $m_X = 800$	32606	74.5	2.1	0.6%
Resonant at $m_X = 900$	34392	405.6	2.3	0.6%
Resonant at $m_X = 1000$	35119	425.2	2.3	0.6%
Resonant at $m_X = 1100$	44157	416.2	2	0.5%
Resonant at $m_X = 1200$	32337	386.9	2.4	0.6%
Resonant at $m_X = 1400$	25539	251.1	1.6	0.7%
Resonant at $m_X = 1600$	14421	160.6	1.4	0.9%
<b>Background Samples</b>				
Fake	2089776	33920	110	0.3%
$t\bar{t}$	490058	61620	90	0.1%
single top	33158	3689	32	0.9%
ZttHF	23827	1316	59	4.5%
ZHF	19390	739	55	7.4%
ZttLF	1522	92	22	23.9%
ZLF	1171	65	18	27.6%
SM Higgs	234087	154	2	1.1%
diboson	8856	155	7	4.3%
W	747	75	8	10.7%
Wtt	76	5.4	0.8	14.8%
DYtt	9	2.7	1.8	66.7%
DY	68	15	4	26.7%
total bkg	3054267	102440		
data	98456	98456		

Table 7.6: Pre-fit event yields in the di-Higgs  $bb\tau_{\text{lep}}\tau_{\text{had}}$  SLT signal region for the data, background and signal. Here, ZttHF (ZHF) represents the processes as  $Z \rightarrow \tau\tau$  ( $Z \rightarrow ee/\mu\mu$ ) in association with a pair of heavy flavour jets ( $b$ -jets or  $c$ -jets), while ZttLF and ZLF represents the same decay of  $Z$  in association with light flavour jets. Similarly for Wtt and W, DYtt and DY.

SampleName	Entries	Integral	Error	Error/Integ.
<b>Signal Samples</b>				
ggF Non-resonant	35045	1.416	0.009	0.6%
VBF Non-resonant	10188	0.0548	0.0006	1.2%
Resonant at $m_X = 251$	6815	17.97	0.22	1.3%
Resonant at $m_X = 260$	6930	18.5	0.23	1.2%
Resonant at $m_X = 280$	8145	22.23	0.25	1.1%
Resonant at $m_X = 300$	6319	26.48	0.34	1.3%
Resonant at $m_X = 325$	7085	30.9	0.4	1.2%
Resonant at $m_X = 350$	7896	35.9	0.4	1.2%
Resonant at $m_X = 400$	8780	42.5	0.5	1.1%
Resonant at $m_X = 450$	9303	47.4	0.5	1.1%
Resonant at $m_X = 500$	5177	51.7	0.7	1.4%
Resonant at $m_X = 550$	5123	53.4	0.8	1.4%
Resonant at $m_X = 600$	5026	53.8	0.8	1.5%
Resonant at $m_X = 700$	4699	52.8	0.8	1.5%
Resonant at $m_X = 800$	4534	52.8	0.8	1.5%
Resonant at $m_X = 900$	4285	51.6	0.8	1.6%
Resonant at $m_X = 1000$	3820	47.4	0.8	1.7%
Resonant at $m_X = 1100$	4436	42.8	0.7	1.5%
Resonant at $m_X = 1200$	2878	34	0.7	2.1%
Resonant at $m_X = 1400$	2264	22.7	0.5	2.2%
Resonant at $m_X = 1600$	1412	16.1	0.4	2.8%
<b>Background Samples</b>				
Fake	41151	1752	33	1.9%
$t\bar{t}$	32873	4207	24	0.6%
single top	1715	197	8	3.9%
ZttHF	8079	439	29	6.6%
ZHF	1336	43	10	23.2%
ZttLF	549	47	15	32.5%
ZLF	65	2	2	74.9%
SM Higgs	30508	24	1	3.3%
Wtt	22	2.7	1	37.0%
W	16	1.31	0.34	26.0%
DYtt	3	0.26	0.17	65.4%
DY	6	0.63	0.31	49.2%
total bkg	117806	6876.79		
data	6351	6351		

Table 7.7: Pre-fit event yields in the di-Higgs  $bb\tau_{\text{lep}}\tau_{\text{had}}$  LTT signal region for signal, data and background. Here, ZttHF (ZHF) represents the processes as  $Z \rightarrow \tau\tau$  ( $Z \rightarrow ee/\mu\mu$ ) in association with a pair of heavy flavour jets ( $b$ -jets or  $c$ -jets), while ZttLF and ZLF represents the same decay of  $Z$  in association with light flavour jets. Similarly for Wtt and W, DYtt and DY.

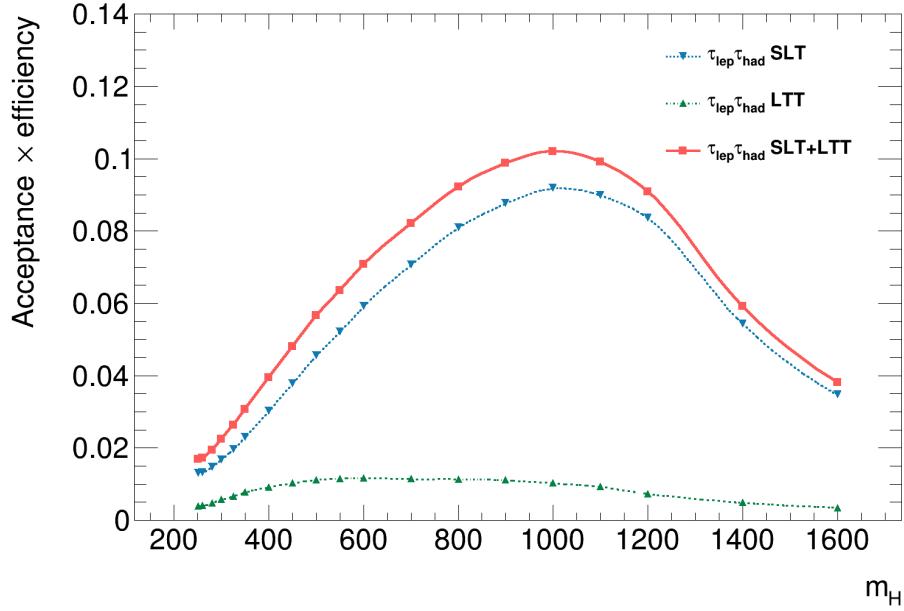


Figure 7.1:  $A \times \epsilon$  for the resonant Di-Higgs production as a function of the resonance mass  $m_X$  in SLT, LTT and SLT LTT combined.

and thus there are no multiple selection possibilities. However, for the SLT channel where a  $\tau_{\text{had}}$  trigger is not used, an anti- $\tau_{\text{had}}$  candidate is chosen randomly when there are more reconstructed  $\tau_{\text{had}}$  satisfying the anti- $\tau_{\text{had}}$  definition. Any anti- $\tau_{\text{had}}$  objects that are not selected in this process are also not considered when performing the overlap removal of detector objects, which is discussed in Section 4.7. Derived variables used in the analysis, such as the  $E_{\text{T}}^{\text{miss}}$ ,  $m_{\tau\tau}^{\text{MMC}}$  and  $E_{\text{T}}^{\text{miss}}\phi$  centrality (more details in section 7.3) are calculated in the same way as for signal events, but with the anti- $\tau_{\text{had}}$  taking the place of the loose  $\tau_{\text{had}}$  candidate.

#### 7.1.4 $Z + \text{HF}$ control region event selection

The cross-section of the  $Z$  boson production in association with heavy flavour jets, i.e.  $b$ -,  $c$ -jets background ( $Z + \text{HF}$  background) is known to be not well modelled by the SHERPA generator. Therefore, a dedicated control region event selection is defined to select  $Z \rightarrow \mu\mu/e e + \text{heavy flavour jets}$  events. Since the production of jets is independent of the decay mode of the  $Z$  boson, this selection provides an orthogonal region with high purity to the SR which requires two  $\tau$  leptons in the final state. The normalisation is extracted from this region, and it also provides constraints on the normalisation of the  $t\bar{t}$  background.

The event selection is described as follows:

- Events are selected using the single-lepton and di-lepton triggers, as specified in Ref. [140].

- Exactly two leptons (*e/mu*) with opposite-sign charges and  $p_T > 9$  GeV are required, these leptons are also required to be compatible with the primary vertex. In addition, they are required to pass the medium and loose isolation requirements.
- Exactly two *b*-tagged jets (tagged with DL1r 77% working point).
- The reconstructed mass of the two leptons are required to be between 75 GeV and 110 GeV (which is around  $Z$  mass peak).
- $m_{bb} < 40$  GeV or  $m_{bb} > 210$  GeV (to veto Higgs mass peak and to ensure orthogonality to  $bb\ell\ell$  signal region).

The data and background  $m_{\ell\ell}$  distributions are shown in Figure 7.2 before performing the likelihood fit and after in (a) and (b), respectively. In (b), the normalisation and shape of the backgrounds and the uncertainty on the total background are shown as determined from the likelihood fit to data in the non-resonant  $HH$  search, as described in detail in section 7.5.3.

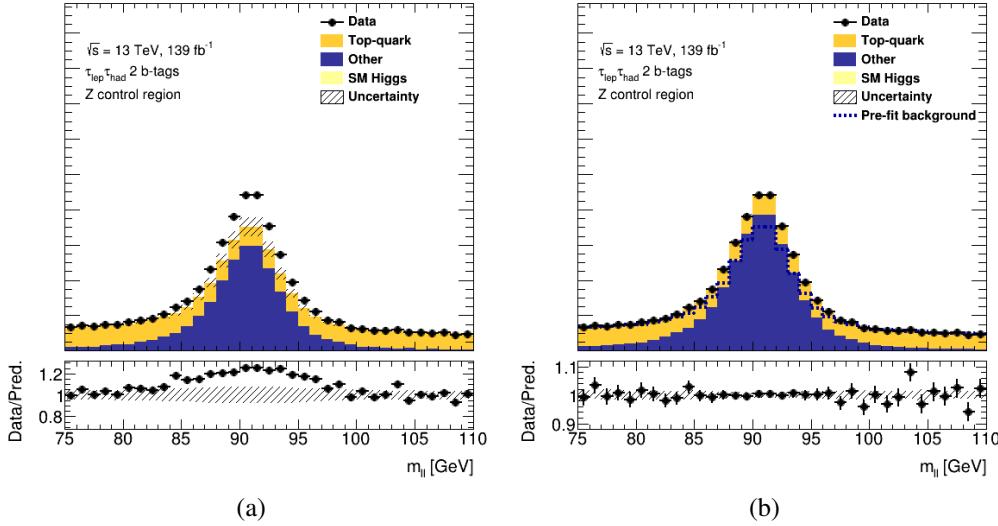


Figure 7.2: (a) Pre-fit background and data and (b) post-fit background and data  $m_{\ell\ell}$  distributions in the  $Z + \text{HF}$  control region. The uncertainty band includes statistical and systematics uncertainties on the total background. The ‘Top-quark’ background includes the  $t\bar{t}$  and the single-top background, the ‘other’ background includes  $W/Z$  bosons + jets, and the ‘SM Higgs’ background includes background in association with a single SM Higgs boson.

## 7.2 Background estimation

This section describes the background estimation methods used in the di-Higgs analysis. The simulated event samples summarised in Section 5.3 are used to model all background

processes, except for processes with fake- $\tau_{\text{had}}$  which are estimated using data-driven techniques, referred to fake- $\tau_{\text{had}}$  background, are discussed below (in particular, the lepton faking  $\tau_{\text{had}}$  background is modelled by simulation, and the contribution is very small). The  $t\bar{t}$  with true- $\tau_{\text{had}}$  and Z+HF templates are taken from the MC prediction but their normalisations are derived from data as included as freely floating parameters in the final fit, as described in Section 7.5. Events with electrons or muons that are misidentified as  $\tau_{\text{had}}$  objects, dominantly coming from the  $t\bar{t}$  production, represent a minor background in the analysis and they are estimated from simulation.

### 7.2.1 Fake- $\tau_{\text{had}}$ backgrounds estimation

The fake- $\tau_{\text{had}}$  can have different origins. In Figure 7.3, two Feynman diagrams are shown for the two dominant processes contributing to the fake- $\tau_{\text{had}}$  background, which are the  $t\bar{t}$  and multi-jet (QCD) processes. In the  $t\bar{t}$  events, the fake- $\tau_{\text{had}}$  typically originates from quark-initiated jets from top-quark decay; in multi-jet events jets initiated from both quarks and gluons can be misidentified as  $\tau_{\text{had}}$ . In the following text, the fake background initiated by the  $t\bar{t}$  (multi-jet) events is referred to as  $t\bar{t}$  (QCD) fakes.

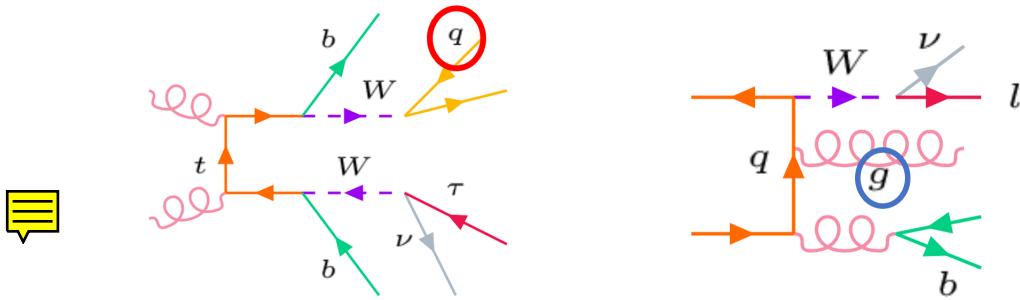


Figure 7.3: Feynman diagrams on the left (right) for the  $t\bar{t}$  (multi-jet) with quark (gluon) circled in red (blue) misidentified as a  $\tau_{\text{had}}$ .

#### 7.2.1.1 Fake factor method

The fake- $\tau_{\text{had}}$  background events are estimated using a data-driven method, the ‘fake factor’ (FF) method, due to imperfect MC modelling of these processes. In short, the FF is the ratio of the number of events with fake- $\tau_{\text{had}}$  in one region to another region. In this analysis, the numerator of the FF is the number of fake- $\tau_{\text{had}}$  events passing the nominal signal region Tau-ID selection (referred to as ID selection in the following), and the denominator is the number of fake- $\tau_{\text{had}}$  events passing the anti- $\tau_{\text{had}}$  selection (referred to as anti-ID selection, as defined in Section 7.1.3). The fake factors are calculated in different dedicated background enriched regions, and then they are combined and used to normalise the fake- $\tau_{\text{had}}$  events from the anti- $\tau_{\text{had}}$  selection to the signal region selection.

In formula form, the FF is given as:

$$\text{FF} = \frac{N(\text{ID selection})}{N(\text{anti-ID selection})}, \quad (7.1)$$

and to obtain the correct estimation of the  $N(\text{ID selection})$  ( $N(\text{anti-ID selection})$ ), events with a true  $\tau_{\text{had}}$  are subtracted from the data events passing the ID (anti-ID) selection.

Due to the different origins of the fake- $\tau_{\text{had}}$ , the FFs are calculated separately for the  $t\bar{t}$  and multi-jet, and for 1 and 3-prong  $\tau_{\text{had}}$  candidates. The FF is parameterised in bins of  $p_T$  of the  $\tau_{\text{had}}$ , while the dependence on  $\eta$  is also checked but no obvious trend is observed. For each process the  $FF$  are calculated in a dedicated background enriched region (referred to as FF-CR, not to be confused with the  $Z + \text{HF CR}$  used for likelihood fit). The FF-CRs for each process are defined as follows:

- $t\bar{t}$  FF-CR: same selection as the signal region selection but with  $m_{bb}$  cut reversed:  $m_{bb} > 150$  GeV.
- Multi-jet (QCD) FF-CR: same selection as the signal region selection but with lepton isolation requirements reversed: ‘tight’ electrons and ‘medium’ muons are required to fail their respective ‘loose’ isolation working points.

Individual fake factors for each process are then used to provide a combined fake factor, which is then applied to the fake- $\tau_{\text{had}}$  events from the anti-ID selection to estimate the fake- $\tau_{\text{had}}$  events from the ID selection. The FFs are derived separately for the SLT and LTT channels and so is the fake background estimation in the SR. The combined fake factor is calculated as:

$$FF(\text{comb}) = FF(\text{QCD}) \times r_{\text{QCD}} + FF(t\bar{t}) \times (1 - r_{\text{QCD}}), \quad (7.2)$$

where the  $FF(\text{QCD})$  ( $FF(t\bar{t})$ ) is the FF calculated in the QCD ( $t\bar{t}$ ) FF-CR. The  $r_{\text{QCD}}$  is defined as the fraction of QCD fakes in the total number of fake- $\tau_{\text{had}}$  background. It is measured as a function of the  $\tau_{\text{had}}$   $p_T$ , splitted into 1-prong and 3-prong, and into the type of light lepton ( $e$  or  $\mu$ ) since the QCD contents are different for them. The  $r_{\text{QCD}}$  is measured for events passing the anti-ID selection as:

$$r_{\text{QCD}} = \frac{N(\text{multi-jet, data})}{N(\text{data}) - N(\text{true } \tau_{\text{had}}, \text{MC})} \quad (7.3)$$

where the  $N(\text{multi-jet, data})$  is calculated by subtracting all background contributions apart from multi-jet, regardless of whether they contain fake or true- $\tau_{\text{had}}$  candidates, from the

data in the anti- $\tau_{\text{had}}$  selection:

$$N(\text{multi-jet, data}) = N(\text{data}) - N(\text{true } \tau_{\text{had}}, \text{MC}) - N(\text{fake } \tau_{\text{had}}, \text{MC}) \quad (7.4)$$

The subtracted backgrounds are taken from the MC predictions, and the MC simulated fake background has solely the contribution from the  $t\bar{t}$  fakes.

In graphical form, the various FF-CRs where the fake factors are measured and applied can be seen in Figure 7.4.

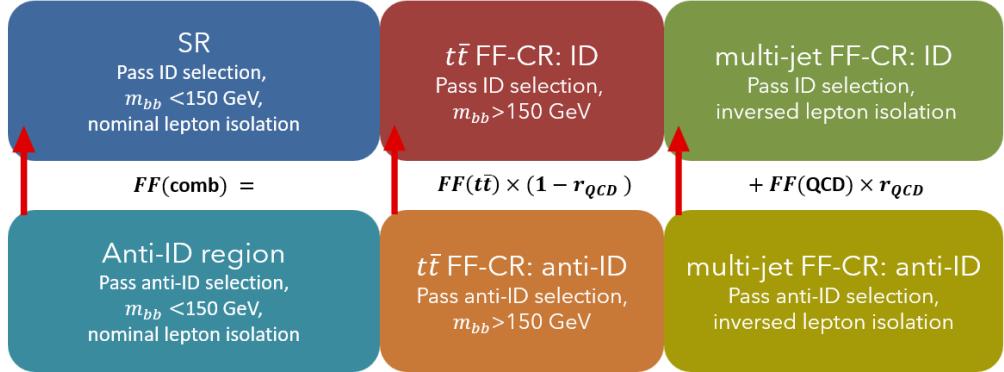


Figure 7.4: Graphical representation of the Fake Factor Method. The FFs are calculated independently for the  $t\bar{t}$  FF-CR and the multi-jet FF-CR, and the combined FF is calculated using these FFs and the  $r_{QCD}$ . The direction of the arrow indicates the direction of extrapolation, that the FF is applied on the bottom regions to extrapolate to the top regions.

### 7.2.1.2 $t\bar{t}$ background reweighting

The determination of the combined fake factor is sensitive to the modelling of simulated  $t\bar{t}$  events with true- $\tau_{\text{had}}$  given that this is the dominant background that is subtracted from data in the derivation of the FFs and  $r_{QCD}$ . The  $t\bar{t}$  modelling also affects the fake background estimation in the SR, since the FF is applied on the anti-ID events where true- $\tau_{\text{had}}$   $t\bar{t}$  background is subtracted from. It was observed that mismodeling in the true  $t\bar{t}$  background especially in the high jet multiplicity and high top-quark  $p_T$  can cause issues in the calculation of the fake factors, giving non-physical negative values at high  $\tau_{\text{had}}$   $p_T$  region. To mitigate this issue, simulated events from  $t\bar{t}$  production are differentially reweighted depending on the jet multiplicity and the scalar sum of the transverse momentum of all visible final state objects ( $H_T$ ) in the event.

These reweighting factors are determined bin-by-bin in distribution of jet multiplicity and  $H_T$ , from another  $t\bar{t}$  FF-CR,  $t\bar{t}$  FF-CR2, which is defined using a selection identical to the SR selection, but with the  $t\bar{t}$  FF-CR  $m_{bb}$  requirements ( $m_{bb} > 150$  GeV) and an additional  $m_T^W > 40$  GeV requirement. Furthermore, events in this region are required to

have a reconstructed  $\tau_{\text{had}}$  candidate, but this candidate is not required to pass any RNN  $\tau_{\text{had}}$  requirement. The  $m_T^W$  requirement is introduced to remove any potential contamination from multi-jet events. The reweighting factors are shown in Appendix B.1, Figure B.1

The reweighting method is validated in two additional validation regions. The reweighting only applies to the fake background estimation process while it's not applied to the true  $t\bar{t}$  background in the SR, as there is no significant mismodeling seen there, and the uncertainties due to the  $t\bar{t}$  modelling has been taken care using a different approach. 7.4.2.1.

### 7.2.1.3 Fake factor calculation

The data and MC events in  $t\bar{t}$  FF-CR passing the anti-ID selection before and after the  $t\bar{t}$  reweighting are shown in Figure 7.5 and 7.6, respectively. The fake background is simulated by MC, and it only includes the contributions from the  $t\bar{t}$ -initiated fakes. A clear mismodeling can be seen in the plots before the reweighting, and it is much mitigated after the reweighting process. At this stage, data and MC are not expected to agree well due to the  $t\bar{t}$  fake is known to be poorly modelled by the MC, which is also the main reason for using the data-driven FF method. Nevertheless, a large number of  $t\bar{t}$  fakes is observed, suggesting the high purity of  $t\bar{t}$  fakes in this region. With true  $\tau_{\text{had}}$  contributions subtracted from the data (and with true  $\tau_{\text{had}}$   $t\bar{t}$  reweighted), events in this region are used in  $t\bar{t}$  FF-CR fake factor calculation (as the denominator).

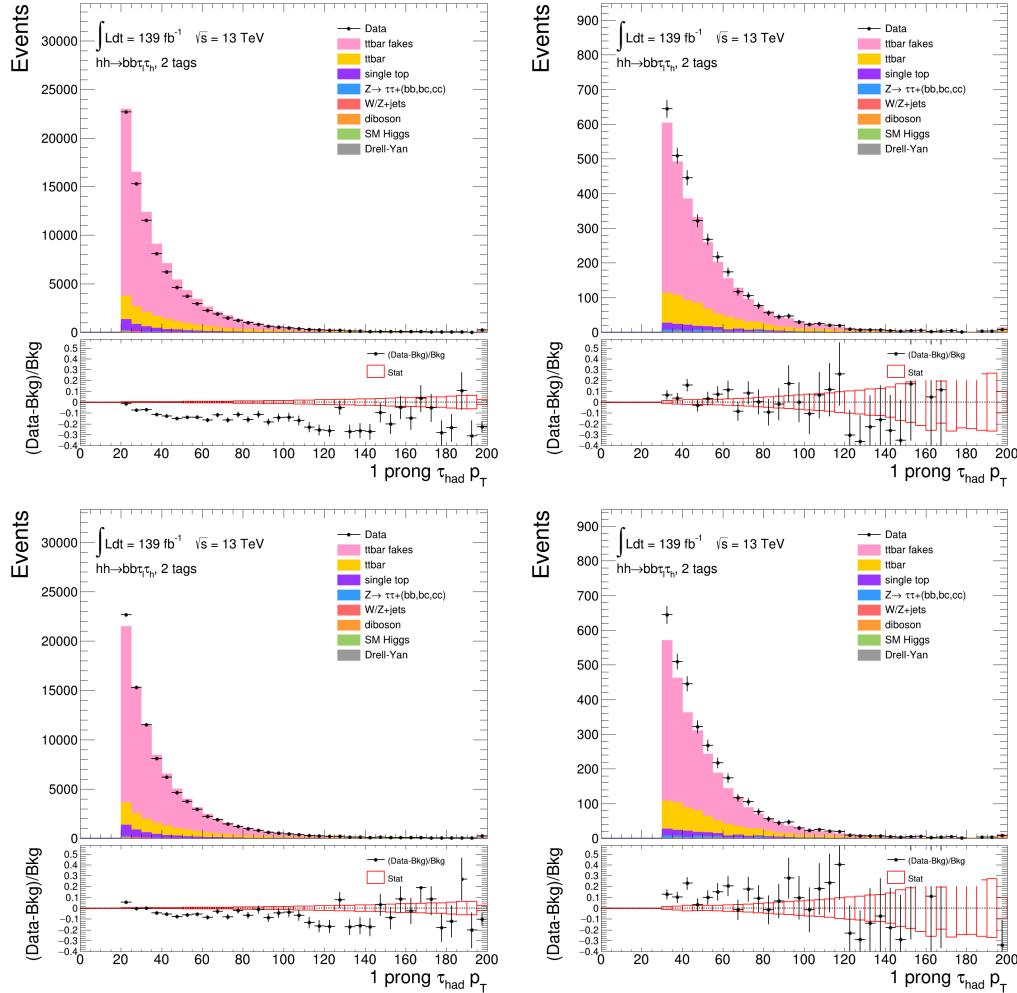


Figure 7.5: In the  $t\bar{t}$  FF-CR, the  $\tau_{\text{had}} p_T$  distributions are plotted for the SLT (left) and LTT channel (right) with  $t\bar{t}$  un-weighted (top) and with  $t\bar{t}$  re-weighted (bottom). The  $\tau_{\text{had}}$  in these events has 1 prong. The MC  $t\bar{t}$  fakes is labelled as ‘ttbar fakes’ in pink. Only statistical uncertainty is included in the error bands.

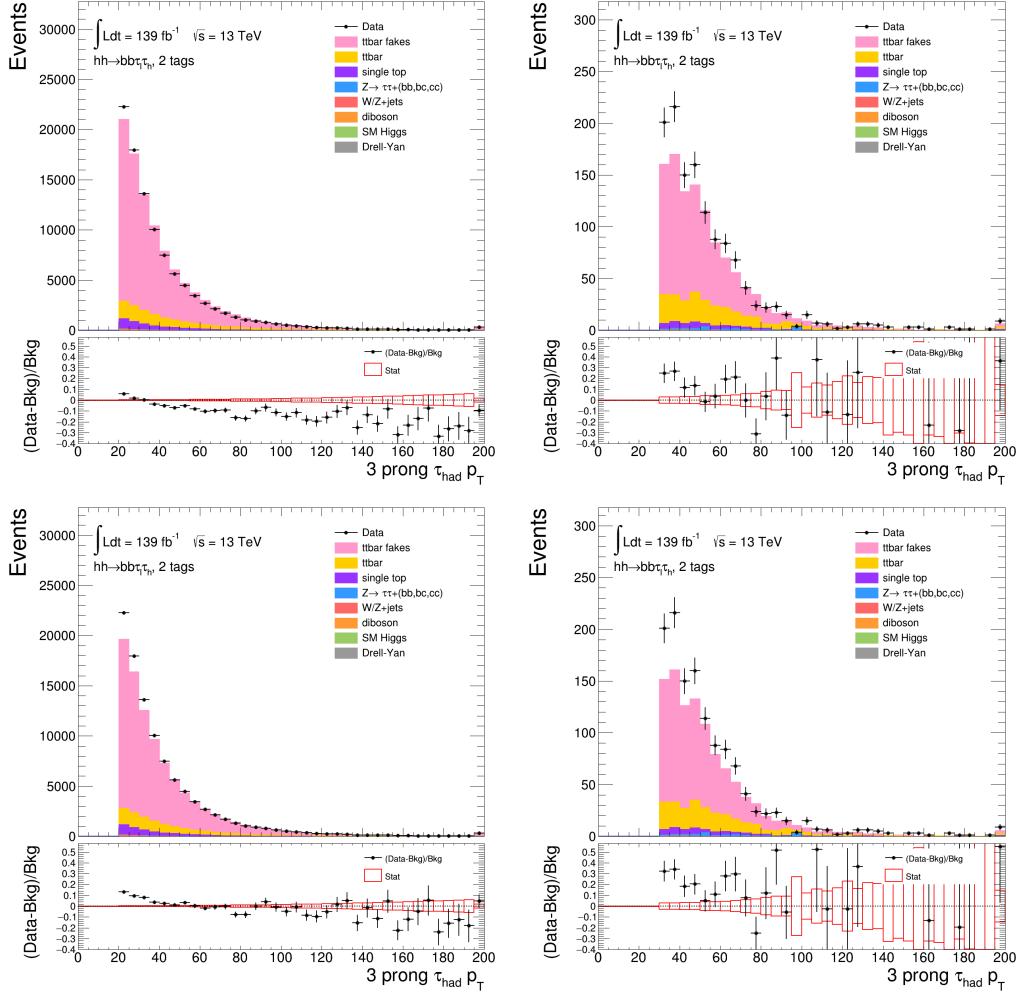


Figure 7.6: In the  $t\bar{t}$  FF-CR, the  $\tau_{\text{had}} p_T$  distributions are plotted for the SLT (left) and LTT channel (right) with  $t\bar{t}$  un-weighted (top) and with  $t\bar{t}$  re-weighted (bottom). The  $\tau_{\text{had}}$  in these events has 3 prongs. The MC  $t\bar{t}$  fakes is labelled as ‘ttbar fakes’ in pink. Only statistical uncertainty is included in the error bands.

Similarly, the data and MC events in QCD FF-CR passing the anti-ID selection before and after the  $t\bar{t}$  reweighting 7.7 and 7.8, respectively. As the QCD fakes is not modelled by the MC, a large discrepancy is shown between data and the MC distributions, which is comprised of the QCD fakes. For consistency, the  $t\bar{t}$  reweighting is applied, however, this region is dominated by the QCD fakes and the  $t\bar{t}$  contributions is negligible. The contribution from  $t\bar{t}$  fakes is also very small, suggesting the **purity of QCD fakes** is high in this region. With true  $\tau_{\text{had}}$  contributions subtracted from data, events in this region is used for the QCD FF-CR fake factor calculation (as the denominator).

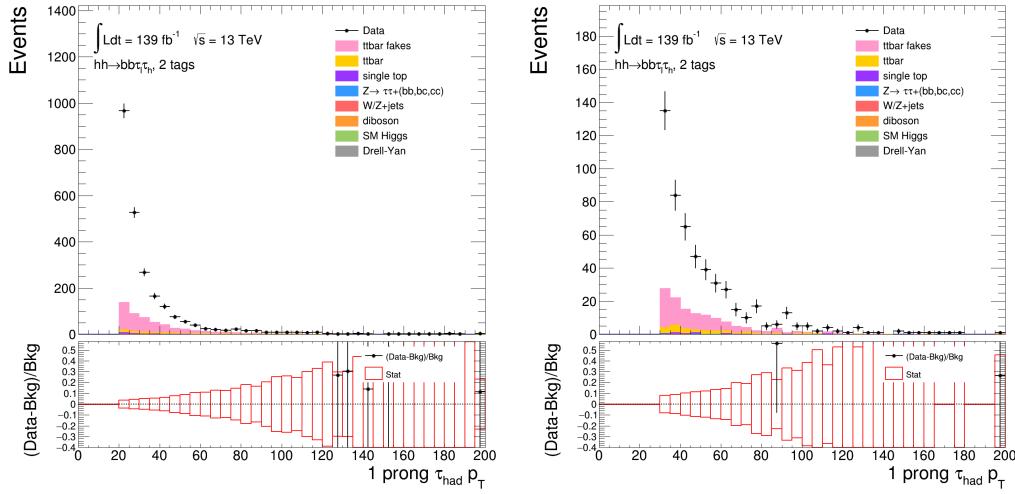


Figure 7.7: In the QCD FF-CR, the  $\tau_{\text{had}} p_T$  distributions are plotted for the SLT (left) and LTT channel (right) with  $t\bar{t}$  re-weighted. The  $\tau_{\text{had}}$  in these events has 1 prong. The MC  $t\bar{t}$  fakes is labelled as ‘ttbar fakes’ in pink. Only statistical uncertainty is included in the error bands.

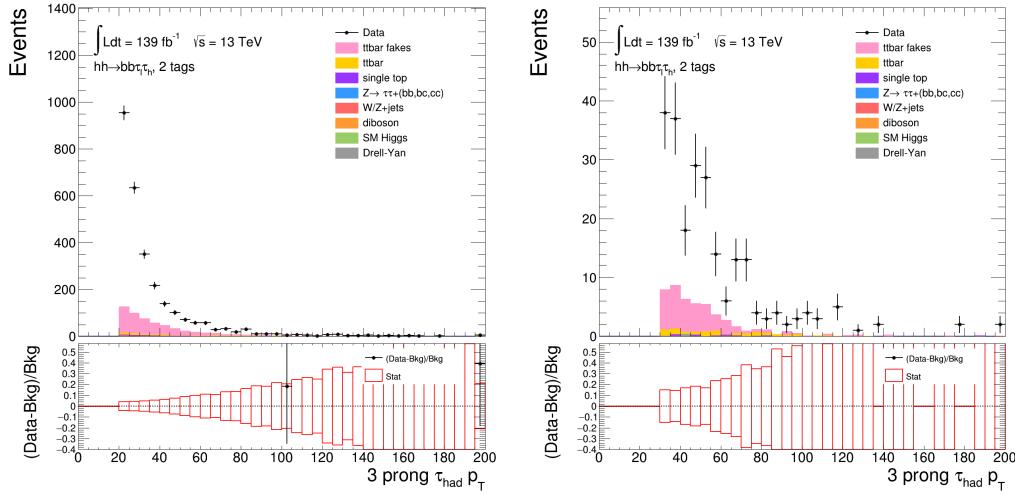


Figure 7.8: In the QCD FF-CR, the  $\tau_{\text{had}} p_T$  distributions are plotted for the SLT (left) and LTT channel (right) with  $t\bar{t}$  re-weighted. The  $\tau_{\text{had}}$  in these events has 3 prongs. Only statistical uncertainty is included in the error bands.

The calculated fake factors are shown on Figure 7.9 and Figure 7.10, for the SLT and LTT channels respectively. The **Fake factors** obtained with reweighted and un-weighted  $t\bar{t}$  contributions are shown on the same graph. The binning is optimised for the  $FF_{t\bar{t}}$  and the same binning is used for the  $FF_{QCD}$ . A **smooth trend is observed** in the  $FF_{t\bar{t}}$ , but some artifacts shapes at mid and high  $p_T$  range are shown in the  $FF_{QCD}$ . This issue has no visible impact on the fakes estimation, because the  $t\bar{t}$  FF dominates over the multi-jet FF in the combined FF which can be seen in the  $r_{QCD}$  distribution in Figure 7.11 and Figure 7.12. The  $r_{QCD}$  for 1-prong and 3-prong  $\tau_{had}$  candidates in  $e\tau_{had}$  and  $\mu\tau_{had}$  channels are shown in Figure 7.11 for the SLT category and in Figure 7.12 for the LTT category.

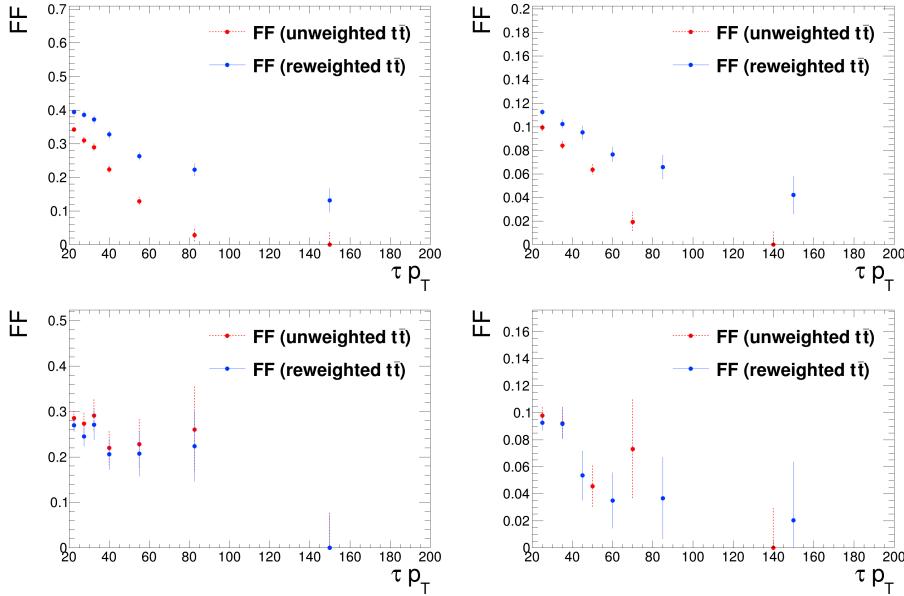


Figure 7.9: Fake factors calculated in  $t\bar{t}$  FF-CR (top) and QCD FF-CR (bottom) for 1-prong (left) and 3-prong (right)  $\tau_{had}$  candidates for the SLT category.

Statistical uncertainties in  $FF_{t\bar{t}}$ ,  $FF_{QCD}$  and  $r_{QCD}$  are evaluated and propagated to the final result, and a conservative 30% modelling uncertainty is assigned to simulated non- $t\bar{t}$  backgrounds which are subtracted from data. The uncertainties due to  $t\bar{t}$  modelling issue and its subtraction are discussed in more details in section 7.4.4.



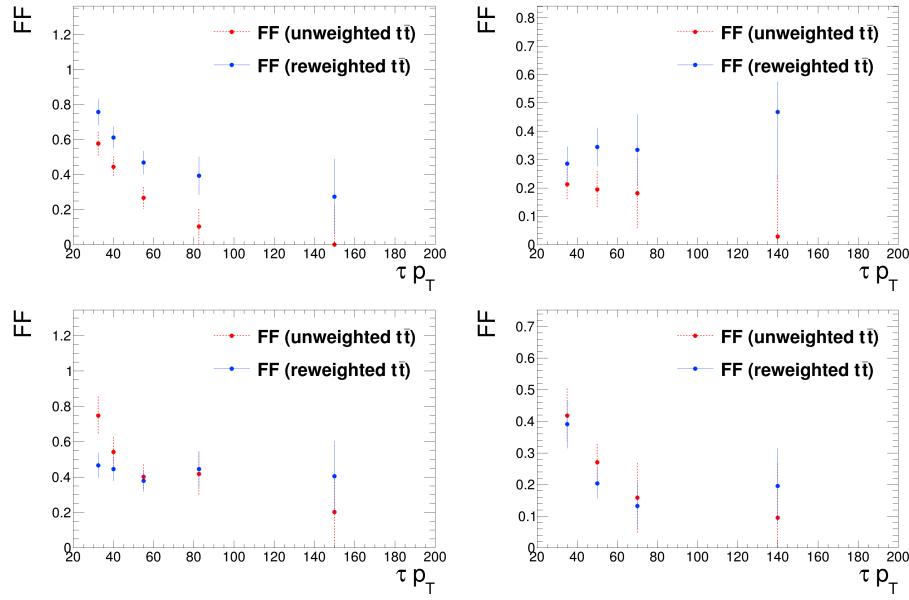


Figure 7.10: Fake factors calculated in  $t\bar{t}$  FF-CR (top) and QCD FF-CR (bottom) for 1-prong (left) and 3-prong (right)  $\tau_{\text{had}}$  candidates for the LTT category.

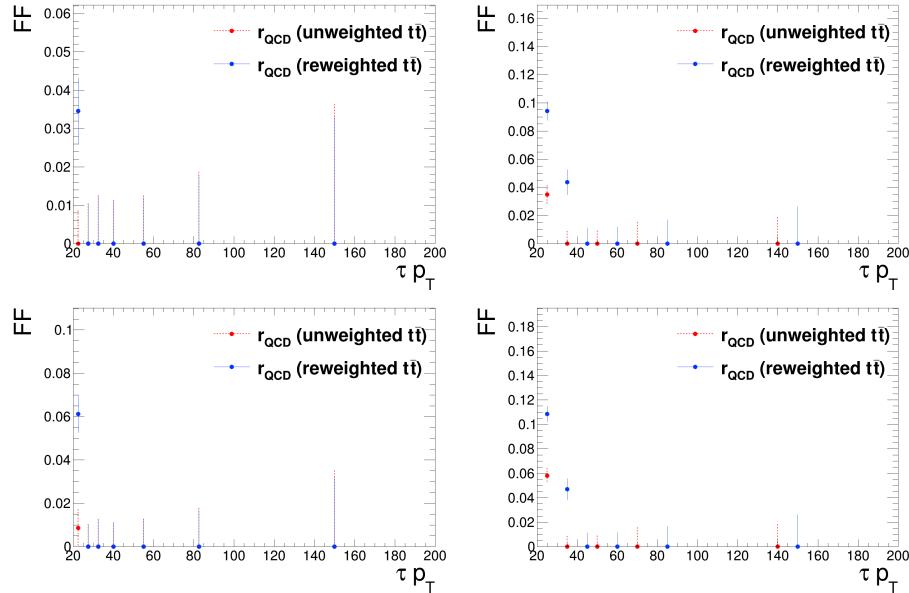


Figure 7.11:  $r_{\text{QCD}}$  for 1-prong (left) and 3-prong (right)  $\tau_{\text{had}}$  candidates with  $e\tau_{\text{had}}$  (top) and  $\mu\tau_{\text{had}}$  (bottom) final states for the SLT channel.

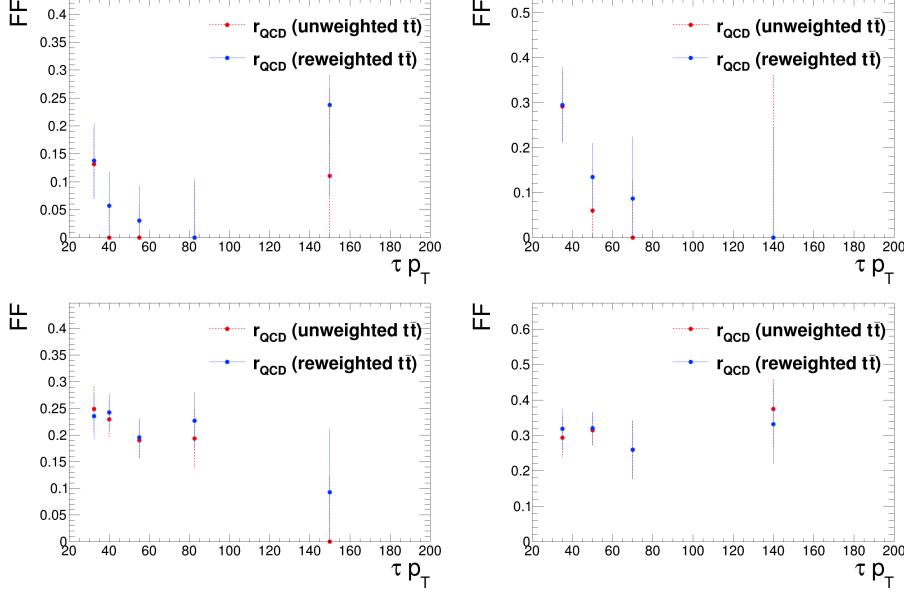


Figure 7.12:  $r_{\text{QCD}}$  for 1-prong (left) and 3-prong (right)  $\tau_{\text{had}}$  candidates with  $e\tau_{\text{had}}$  (top) and  $\mu\tau_{\text{had}}$  (bottom) final states for the LTT channel.

#### 7.2.1.4 Fake factor method validation

The combined FF method is validated in the 0- $b$ -tagged and 1- $b$ -tagged regions, where the same event selection as SR is applied but with different numbers of  $b$ -tagged jets required. The fakes in each validation region are estimated with FF calculated in each validation region with the same method used in the 2- $b$ -tagged region. These two validation regions are chosen as the signal contamination in the 0- $b$ -tagged and 1- $b$ -tagged regions is negligible, and the 0- $b$ -tagged region can benefit from its rich statistics and the 1- $b$ -tagged region can benefit from being closer to the SR. The estimated background distributions agree well with the observed distributions in all validation regions. The data and MC comparison with fakes estimated with the FF method are shown in Figure 7.13 for the SLT channel and Figure 7.14 for the LTT channel.

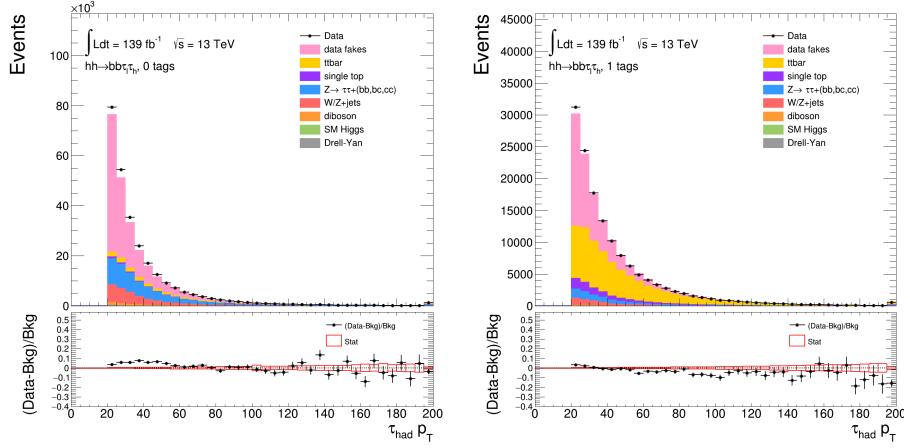


Figure 7.13: A comparison of data and background with fakes estimated with combined FF evaluated at the 0-*b*-tagged (left) and 1-*b*-tagged region (right) for the SLT channel. Only statistical uncertainty is included in the error bands.

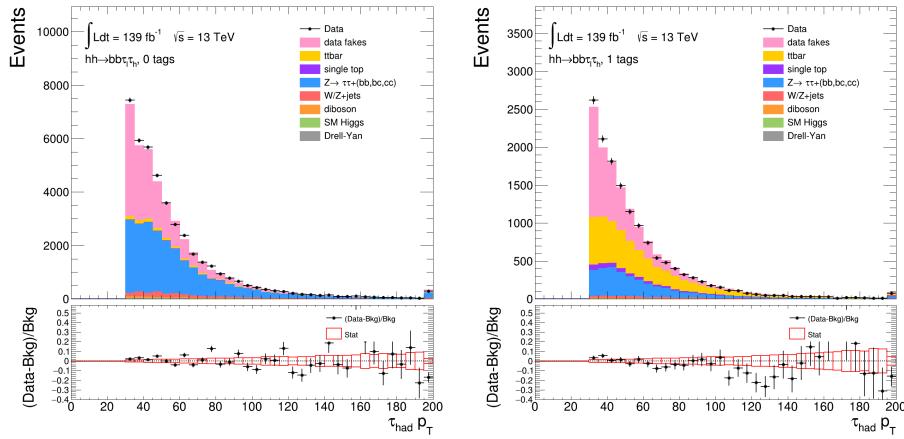


Figure 7.14: A comparison of data and background with fakes estimated with combined FF evaluated at the 0-*b*-tagged (left) and 1-*b*-tagged region (right) for the LTT channel. Only statistical uncertainty is included in the error bands.

## 7.3 Multivariate signal extraction

The signal extraction in the  $HH$  analysis is based on multivariate (MVA) techniques. As compared to the traditional cut-based approach, the MVA techniques can keep events which may fail the cut-based identification because it does not satisfy the selection criterion for a single quantity, while in the MVA selection this event can still satisfy the selection criteria, because the MVA combines the information of all of the discriminating variables. In this sense, the cut-based method is similar to slicing a single hypercube for one (or more) quantity in the parameter space, while the MVA approach is similar to splitting the parameter space into a large number of hypercubes and as a result, is capable of taking account of the correlations between variables.

To extract the non-resonant  $HH$  signal, a neural network is used, which is trained with this signal and background defined in section 5. In the training, the background due to a jet faking a  $\tau_{\text{had}}$  is modelled by MC simulation, in oppose to the data-driven method described in section 7.2. The resonant signal, however, is not a single signal hypothesis, but rather a set of continuous signal hypotheses parameterised by the mass of the heavy resonance which decays to Higgs boson pairs. As a result, a single classifier would not be enough; while using a set of single classifiers trained on each of the simulated resonant signal points would provide good discrimination for each of them, but can not interpolate well between these points. For this reason, and to reduce the number of algorithms that require training, Parametric Neural Networks (PNNs) [141] are used for extracting the resonant signal. PNNs are neural networks connected to one or more physics parameters which enable the optimal signal-background classification for a continuous spectrum of signal.

### 7.3.1 MVA input variables

The same set of MVA input variables is used for the resonant and non-resonant production modes, though different input variables are used for the SLT and LTT channels. The choice of input variables is based on Ref. [61]. Three new high-level variables,  $\Delta\phi(\ell\tau, E_{\text{T}}^{\text{miss}})$ ,  $\Delta\phi(\ell, E_{\text{T}}^{\text{miss}})$  and  $S_{\text{T}}$  are introduced, which have shown discriminating power in the LTT channel.

These variables are listed in Table 7.8 and are defined as follows:

- $m_{HH}$  is the invariant mass of the  $HH$  system as reconstructed from the  $\tau$ -lepton pair (calculated using the MMC) and the  $b$ -tagged jet pair;
- $\Delta R(\tau, \tau)$  is evaluated between the electron or muon and the  $\tau_{\text{had}}$ ;
- $\Delta R(b, b)$  is evaluated between the  $b$ -tagged jets;

- $\Delta p_T(\ell, \tau)$  is the difference between the transverse momenta of the lepton and the  $\tau_{\text{had}}$
- $m_T^W = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos \Delta\phi_{\ell, E_T^{\text{miss}}})}$  is the transverse mass of the lepton and the  $E_T^{\text{miss}}$ ;
- the  $E_T^{\text{miss}} \phi$  centrality specifies the angular position of the  $E_T^{\text{miss}}$  relative to the  $\tau_{\text{had}}$  in the transverse plane [142] and is defined as  $(A + B)/\sqrt{A^2 + B^2}$ , where  $A = \sin(\phi_{E_T^{\text{miss}}} - \phi_{\tau_2})/\sin(\phi_{\tau_1} - \phi_{\tau_2})$ ,  $B = \sin(\phi_{\tau_1} - \phi_{E_T^{\text{miss}}})/\sin(\phi_{\tau_1} - \phi_{\tau_2})$ , and  $\tau_1$  and  $\tau_2$  represent the electron or muon and  $\tau_{\text{had}}$ ;
- $\Delta\phi(\ell\tau, bb)$  is the azimuthal angle between the  $\ell + \tau_{\text{had-vis}}$  system and the  $b$ -tagged jet pair;
- $\Delta\phi(\ell, E_T^{\text{miss}})$  is the azimuthal angle between the lepton and the  $E_T^{\text{miss}}$ ;
- $\Delta\phi(\ell\tau, E_T^{\text{miss}})$  is the azimuthal angle between the electron or muon and  $\tau_{\text{had}}$  system and the  $E_T^{\text{miss}}$ ;
- $S_T$  is the total transverse energy in the event, summed over all jets,  $\tau_{\text{had}}$  and leptons in the event and  $E_T^{\text{miss}}$ .

Variable	SLT	LTT
$m_{HH}$	✓	✓
$m_{\tau\tau}^{\text{MMC}}$	✓	✓
$m_{bb}$	✓	✓
$\Delta R(\tau, \tau)$	✓	✓
$\Delta R(b, b)$	✓	
$\Delta p_T(\ell, \tau)$	✓	✓
Sub-leading $b$ -tagged jet $p_T$ ( $p_T^b 2$ )	✓	
$m_T^W$	✓	
$E_T^{\text{miss}}$	✓	
$E_T^{\text{miss}} \phi$ centrality	✓	
$\Delta\phi(\ell\tau, bb)$	✓	
$\Delta\phi(\ell, E_T^{\text{miss}})$		✓
$\Delta\phi(\ell\tau, E_T^{\text{miss}})$		✓
$S_T$		✓

Table 7.8: Variables used as inputs to the MVAs in the SLT and LTT channels. The same choice of input variables is used for the resonant and non-resonant production modes.

The data versus MC background comparison of these variables are shown in Figure 7.15 for the SLT channel and Figure 7.16 for the LTT channel, respectively. In both NN used for non-resonant signal and PNNs used for resonant signal, the three most significant and discriminating variables are  $m_{HH}$ ,  $m_{\tau\tau}^{\text{MMC}}$  and  $m_{bb}$ .

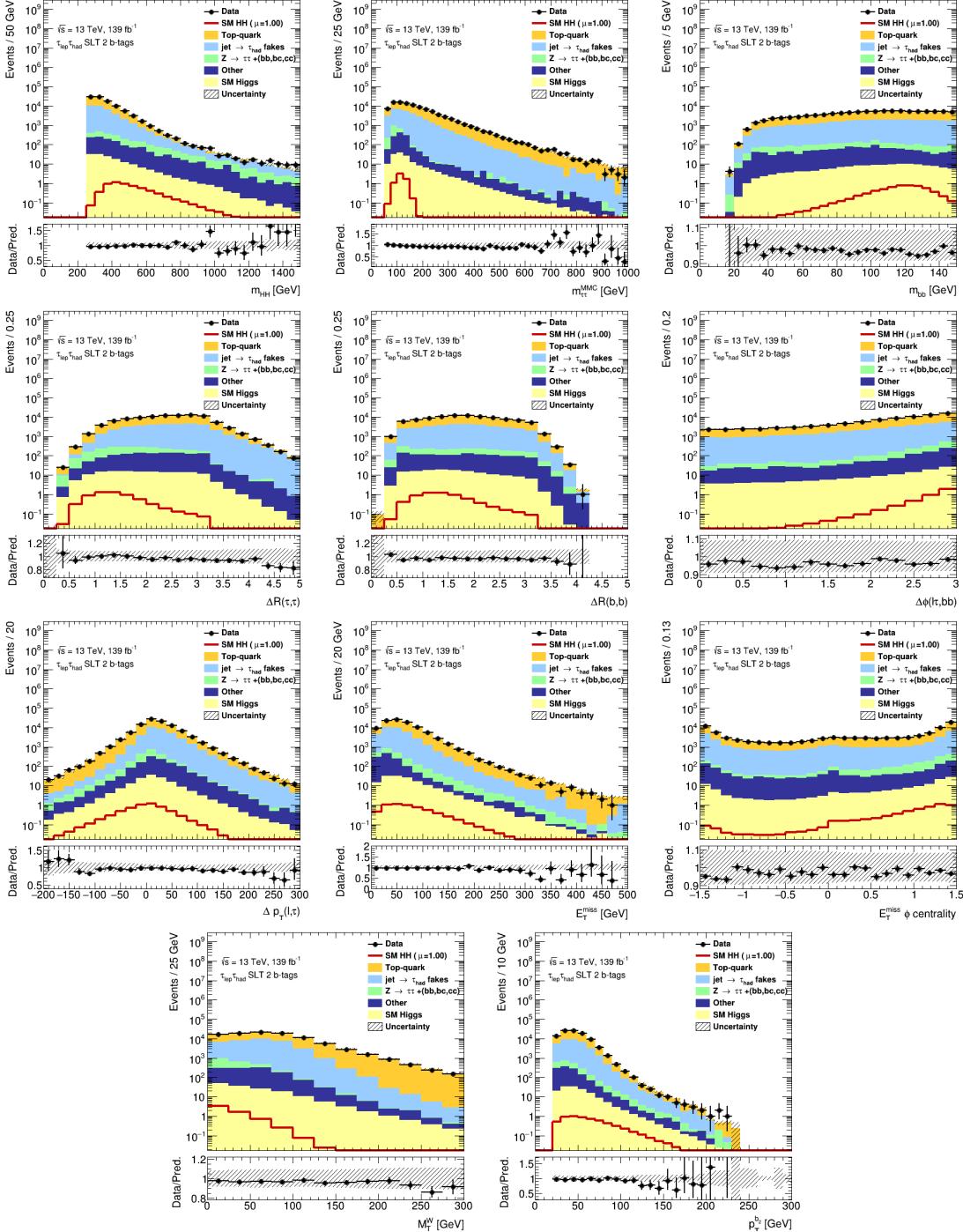


Figure 7.15: Pre-fit PNN/NN input variable distributions for the SLT signal region.

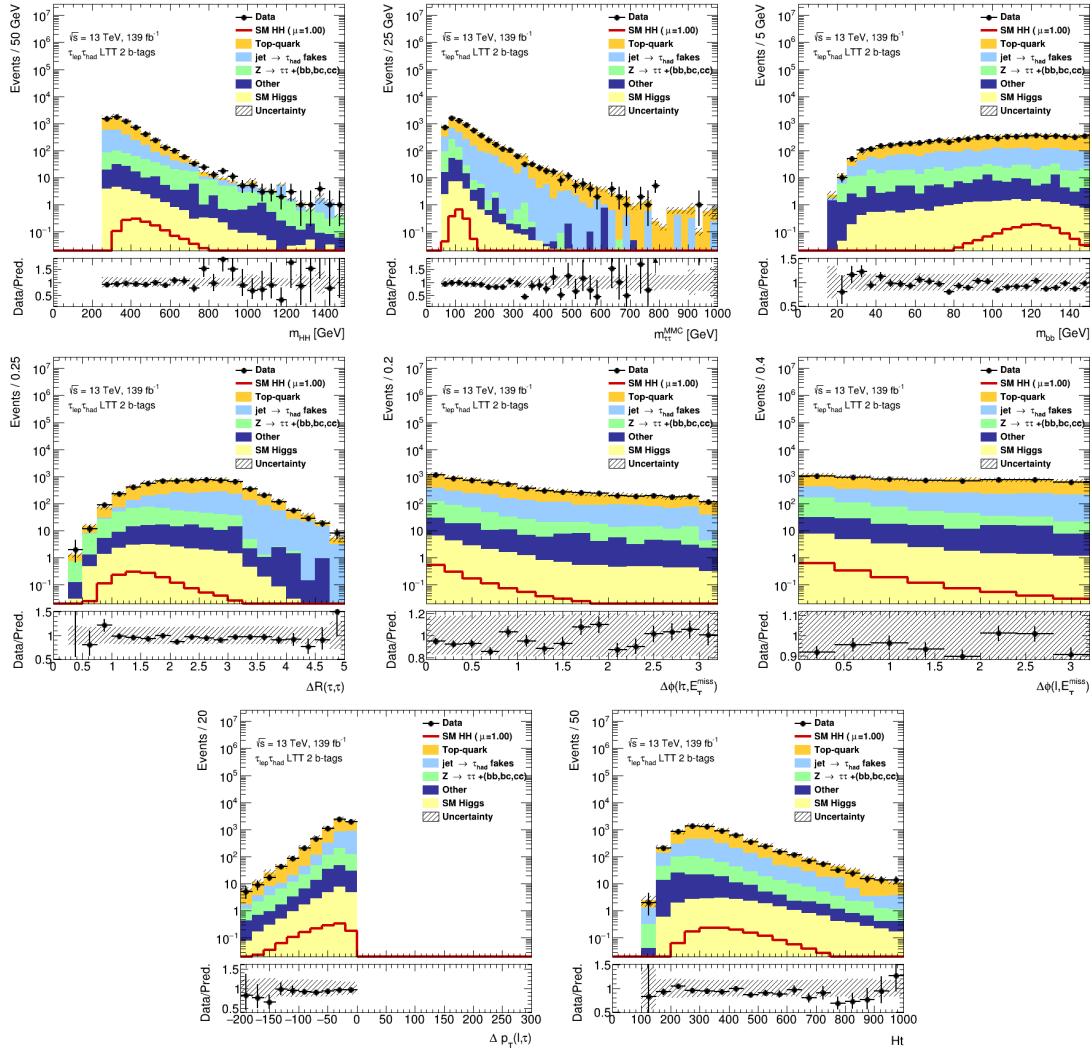


Figure 7.16: Pre-fit PNN/NN input variable distributions in the di-Higgs  $bb\tau_{lep}\tau_{had}$  LTT signal region.

Parameter	Value for SLT	Value for LTT
Layer sizes	512 512	512 512 512
Epochs	241	238
Batch-size	34	35
Learning rate	0.500	0.500
Learning rate decay	$4.15 \times 10^{-3}$	$4.20 \times 10^{-3}$
Nesterov momentum	0.944	0.791
L2 regularisation weight	$1.00 \times 10^{-5}$	$1.65 \times 10^{-5}$

Table 7.9: Training parameters used for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  non-resonant NNs. Table reproduced from Ref. [78].

### 7.3.2 MVA hyper parameters optimisation and training

Standardised variables are used for all NN trainings described in the following. The input variables are standardised by subtracting the median and dividing by the interquartile range. During training, the sum of all backgrounds normalised to their respective cross-sections is used. The neural networks are trained using Keras [143] and Tensorflow [144]. The hyperparameters used for training the NN (PNN) are listed in Table 7.9 (7.10). The rectified linear unit (ReLU) and sigmoidal activation functions are used for the hidden layers and output layers respectively. The loss function is the binary cross-entropy, using L2 regularisation ('Ridge regression', which adds squared amplitude for the penalty term), and the optimiser used is stochastic gradient descent [145]. The numbers of epoch, batch size, learning rate and learning rate decay are again listed in Table 7.9, 7.9. The Nesterov momentum is used in the gradient descent.

These hyperparameters were optimised separately for each MVA by a Bayesian optimisation procedure with a Gaussian process. A metric for the NN hyperparameters optimisation is defined as the quadrature sum of approximated significances of bins with at least 5 expected background events, while for the PNN, a metric is defined as the approximated significance after an optimised cut (a cut that maximises the significance), and the bins after cut are required to have at least 5 expected background events. The metric is evaluated on the validation and testing sets, and the hyper parameters giving the best value are chosen. The NN and PNN are retrained when the optimisation is done.

During training for the NN (PNNs), the two-fold validation is used. The data set is partitioned into two sets depending on the event numbers being even or odd, where one set is used for training with the other set being used for validation, vice versa. To check if the neural networks are overtrained, the loss functions as a function of the training epoch for the NN (PNNs) are shown in Figure 7.17 (Figure 7.18) for the SLT and LTT channels. A stable generalisation gap is seen between the training and validation, indicating that the overtraining is small. Further checks are done using a much simpler architecture consisting of three layers of 32 nodes for the LTT channel. A large gap is still seen,

Parameter	Value for SLT	Value for LTT
Layer sizes	512 512	256 256 256
Epochs	245	296
Batch-size	33	24
Learning rate	0.500	0.377
Learning rate decay	$4.55 \times 10^{-3}$	$3.38 \times 10^{-3}$
Nesterov momentum	0.975	0.975
L2 regularisation weight	$5.45 \times 10^{-7}$	$8.64 \times 10^{-6}$

Table 7.10: Training parameters used for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  resonant PNNs. Table reproduced from Ref. [78].

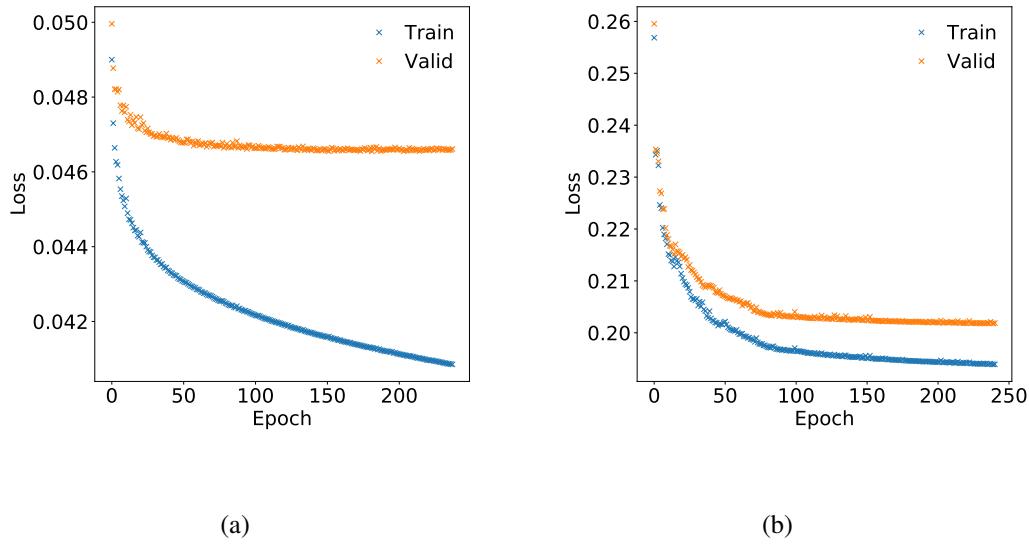


Figure 7.17: Non-resonant NN loss functions as a function of training epochs evaluated using the two-fold cross validation, for the (a) LTT and (b) SLT categories. Figures reproduced from analysis internal notes.

indicating the generalisation gap is due to limited data rather than effective capacity of the neural networks. Additional plots of the NN (PNNs) output distribution of the signal and summed background can be found in Appendix B.2, Figure B.2 and Figure B.3 for the SLT and LTT channel, respectively.

In addition, the PNNs are trained on all the signals simultaneously, and are parameterised by the mass of the resonance  $HH$  sample. The PNNs therefore provide near-optimal sensitivity over the range of signal masses considered. Since the background does not have a well-defined value of the resonance mass, the background events are assigned a random value generated from the signal mass parameter during training. In addition, the PNNs interpolate well between the mass points used for the training. This is checked by plotting the PNN response in terms of Asimov significance as a function of the PNN mass parameter, which is shown in Figure 7.19. The PNNs show very high significance over whole scanning mass spectrum. The minimum significance appears in the vicinity of low signal

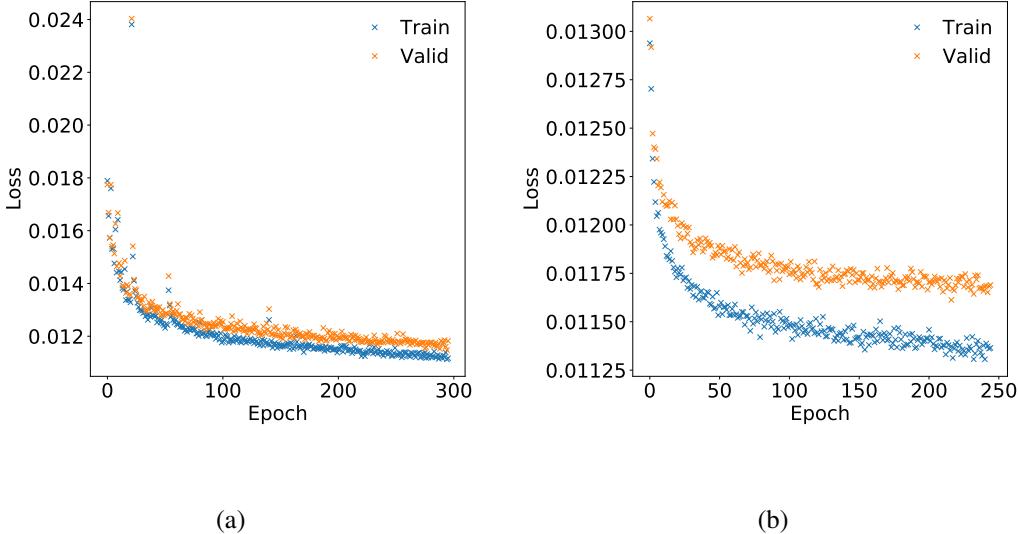


Figure 7.18: PNN loss functions as a function of training epochs evaluated using the two-fold cross validation, for the (a) LTT and (b) SLT categories. Figures reproduced from analysis internal notes.

masses, reaching a value of 0.7.

### 7.3.3 MVA output distributions

The output distributions of the NN and the PNNs are shown in Figure 7.20. The data distributions are not **blinded** here as the definition of the event pre-selection is very loose. To validate the MC modelling, the pre-fit PNN score distributions are shown in Figure B.4 in Appendix B.2 in the 1-tag control region. The binning of the MVA output discriminant is defined in section 7.5.1. The events yields in the last three bins of the MVA discriminant are shown in Table B.1 (Table B.2) in Appendix B.2 for SLT (LTT) channel, which are the most significant bins for signal extraction.

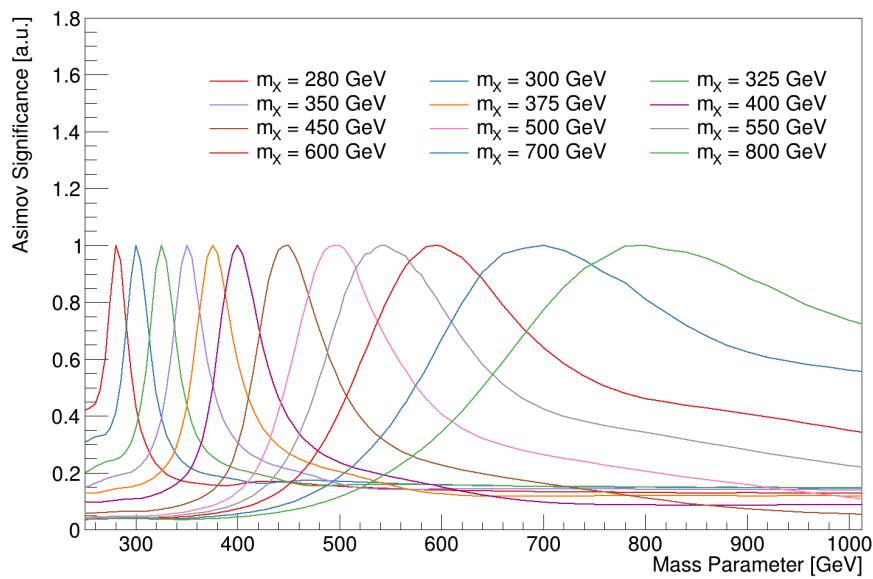


Figure 7.19: PNN response in terms of Asimov significance as a function of the PNN mass parameter obtained for signal samples with different masses. Due to a large difference in acceptance times efficiency of the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel signal region selection depending on the mass of the resonance, all significances are scaled so that the maximum significance is 1 for better visibility.

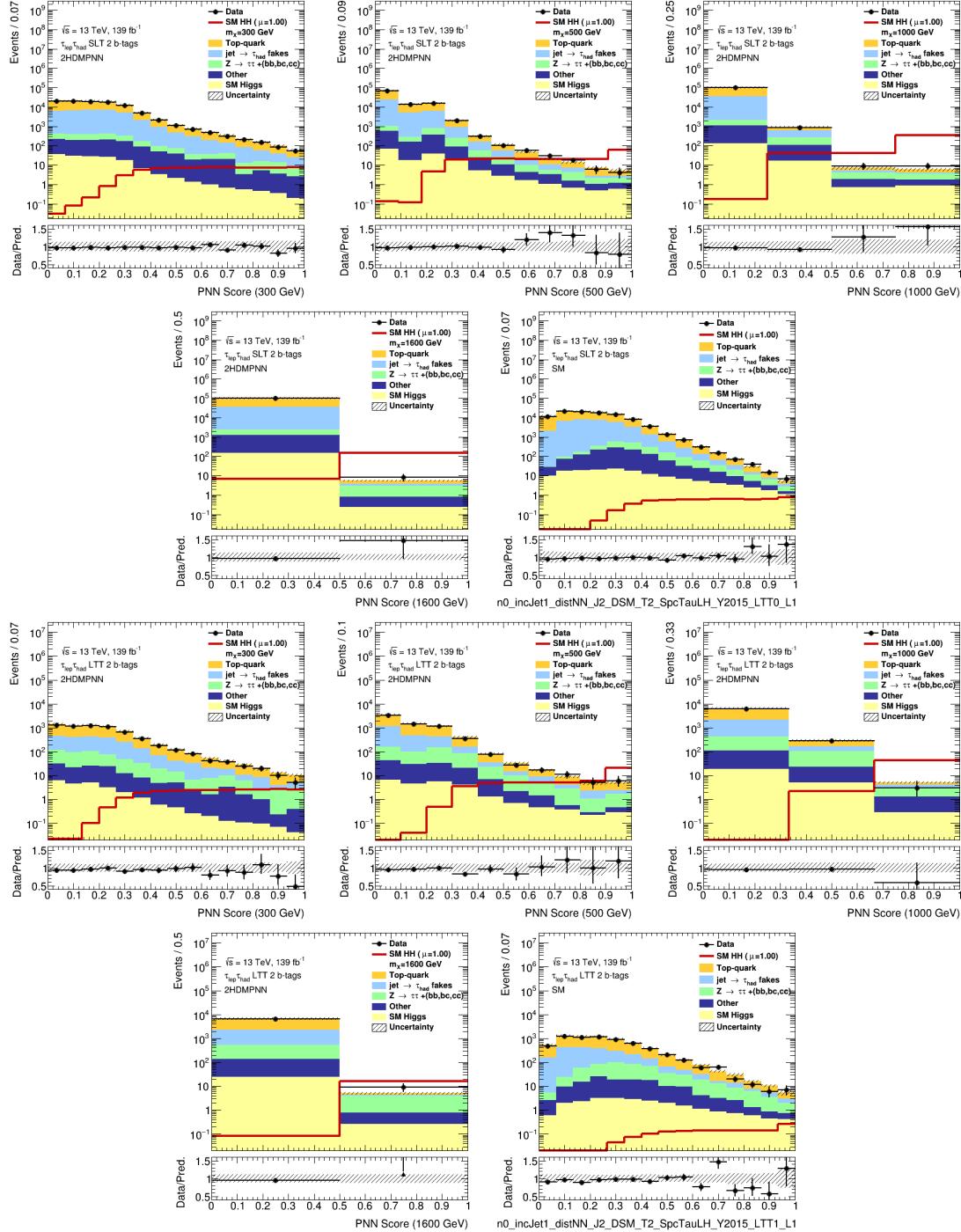


Figure 7.20: Pre-fit PNN score distributions for the 300, 500, 1000, 1600 GeV mass points and non-resonant NN in the di-Higgs  $bb\tau_{\text{lep}}\tau_{\text{had}}$  SLT (top two rows) and LTT (bottom two rows) signal regions.

## 7.4 Systematic uncertainties

The systematics considered in the  $HH$  analysis is outlined in this section. For the same reason mentioned in section 6.7, the systematic uncertainties can be broadly categorised into experimental and modelling systematic uncertainties. The former arise from experimental sources related to the detector response and the physics object reconstruction and identification, while the modelling systematic uncertainties account for theoretical uncertainties related to the limited knowledge of the MC simulated processes and uncertainties on data-driven backgrounds.

The experimental systematic uncertainties have been briefly discussed in section 6.7, while many of the systematics considered there will be included in this analysis as well. In table 7.11 and table 7.12, the full list of the experimental systematics are listed, which will be discussed in more details in the following section.

The experimental and modelling uncertainties are derived for all backgrounds and signals, manifesting the overall yields or the shape of the final discriminant, or both together. All uncertainties considered are propagated through the analysis, and are included in the statistical analysis fit discussed in section 7.5.

### 7.4.1 Experimental uncertainties

Uncertainty of measuring the luminosity arises from calibrating it with the van der Meer method [146]. The uncertainty on the full Run 2 data is  $\pm 1.7\%$ , obtained using the LUCID (LUminosity measurement using a Cherenkov Integrating Detector) detector [72].

A reweighting is applied on the MC simulations to match the pile-up condition in each data-taking period during Run 2 [147]. The uncertainty associated with the reweighting is estimated by varying the reweighting factors used in the MC simulation.

The uncertainties in single lepton triggers ( $e$  or  $\mu$ ), uncertainties due to leptons reconstruction, identification and isolation are estimated by varying the selection used in the tag-and-probe method. In the tag-and-probe approach, stringent selection criterias are applied on one lepton, and the other lepton is used for the efficiency measurement. For electrons,  $Z \rightarrow e^+e^-$  events are used [37] while for muons, both  $Z \rightarrow \mu^+\mu^-$  and  $J/\psi \rightarrow \mu^+\mu^-$  events are used [38].



The uncertainty on electron energy resolution is estimated by smearing the electron energies in MC simulations, while the momentum scale and energy resolution variations are derived in  $Z \rightarrow \mu^+\mu^-$  and  $J/\psi \rightarrow \mu^+\mu^-$  events.

Similarly for the  $\tau_{\text{had}}$  lepton, the tag-and-probe method is used to measure the reconstruction and identification efficiency scale factors, using  $Z \rightarrow \tau^+\tau^-$  events [63]. Uncertainties are estimated in the efficiencies of the electron veto for  $\tau_{\text{had}}$ , using  $Z \rightarrow e^+e^-$

events.

The systematic uncertainties associated to the jet energy scale calibration [47], as described in section 4.4, are grouped into a reduced-size set of nuisance parameters. The reduced set is referred to as ‘CategoryReduction’, and is recommended set of uncertainties for analyses aiming to perform combinations with other searches. Similarly, the jet energy resolution uncertainties [47] are grouped in to the ‘FullJER’ scheme for this reason. The jet energy resolution uncertainties therefore are comprised of 30 **nuisance parameters** and 13 nuisance parameters for the jer energy resolution uncertainties.

The uncertainty associated with the JVT algorithm (see section 4.4) is estimated by varying the requirements on the algorithm output discriminant.

As described in section 6, scale factors are measured to correct the  $b$ -tagging efficiency in MC simulations to match the one in data. Uncertainties in the  $b$ -tagging efficiency are considered as a function of  $p_T$  for  $b$ -jets and  $c$ -jets using  $t\bar{t}$  events [56, 148], and as a function of  $pT$  and  $\eta$  for light-flavour jets using di-jet events [130]. The ‘medium reduction’ scheme is used, which is the preferred reduction scheme used for analyses aiming for combination. It contains 13 nuisance parameters, 3 for  $b$ -jets, 4 for  $c$ -jets, 4 for light-jets and 2 for extrapolations.

The energy scale and resolution uncertainties mentioned above, in addition to an uncertainty in the tracks matched to the primary vertex but not associated to other reconstructed objects in the event, are propagated to the  $E_T^{\text{miss}}$  calculation [149].

#### 7.4.2 Background modelling uncertainties for MC-based processes

The modelling systematic uncertainties **include** the uncertainties due to theoretical cross-section calculations and acceptance. The former only affects the normalisation of the samples, and is applied on all MC backgrounds; the later affects both the expected event yields and the shape of kinematic distributions, and therefore the **MVA discriminants** shapes. In order to disentangle these effects, the impact from shape variation is considered separately from the normalisation effect, as it does not change the amount of expected events. The shape uncertainties are derived comparing the distributions of several kinematic variables or **MVA output discriminant** of the alternative samples to the ones of the nominal sample. The comparison is done by taking the normalised distributions (to exclude the normalisation effects that are considered separately) and calculating the ratio of the alternative and nominal number of events bin per bin.

Acceptance uncertainties for  $t\bar{t}$ ,  $Z$ +heavy flavour jets, single-top ( $Wt$ -channel),  $ttH$  and  $ZH$  are estimated in the SR and CR, using the nominal and alternative MC simulated samples. The acceptance uncertainties on minor backgrounds ( $Z$ +light flavour jets,  $W$ +jets, and di-boson) are taken from the **SM Vhbb** analysis [150], as described in Section 7.4.2.5.

NP Name	Description
Event	
ATLAS_LUMI_Run2	Uncertainty on the total integrated luminosity (1.7%)
ATLAS_PU_PRW_DATASF	Pile-up reweighting
electrons	
EL_EFF_TRIGGER_TOTAL_INPCOR_PLUS_UNCOR	electron-trigger
EL_EFF_RECO_TOTAL_INPCOR_PLUS_UNCOR	electron-reconstruction
EL_EFF_ID_TOTAL_INPCOR_PLUS_UNCOR	electron-identification
EL_EFF_ISO_TOTAL_INPCOR_PLUS_UNCOR	electron-isolation
muons	
MUON_EFF_TrigStatUncertainty	muon-trigger
MUON_EFF_TrigSystUncertainty	muon-reconstruction
MUON_EFF_RECO_STAT	
MUON_EFF_RECO_SYS	
MUON_EFF_RECO_STAT_LOWPT	
MUON_EFF_RECO_SYS_LOWPT	
MUON_EFF_ISO_STAT	muon-isolation
MUON_EFF_ISO_SYS	
MUON_EFF_TTVA_STAT	muon-TTVA
MUON_EFF_TTVA_SYS	
MUON_ID	
MUON_MS	
MUON_SCALE	
MUON_SAGITTA_RHO	
MUON_SAGITTA_RESBIAS	
$\tau$ -leptons	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATDATA161718	$\tau$ -trigger
TAUS_TRUEHADTAU_EFF_TRIGGER_STATDATA1718	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATDATA2016	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATDATA2018	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATDATA2018AFTTS1	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATMC161718	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATMC1718	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATMC2016	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATMC2018	
TAUS_TRUEHADTAU_EFF_TRIGGER_STATMC2018AFTTS1	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYST161718	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYST1718	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYST2016	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYST2018	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYST2018AFTTS1	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYSTMU161718	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYSTMU1718	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYSTMU2016	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYSTMU2018	
TAUS_TRUEHADTAU_EFF_TRIGGER_SYSTMU2018AFTTS1	
TAUS_TRUEHADTAU_EFF_RECO_HIGHPT	$\tau$ -reconstruction
TAUS_TRUEHADTAU_EFF_RECO_TOTAL	$\tau$ -identification
TAUS_TRUEHADTAU_EFF_RNNID_1PRONGSTATSYSTPT2025	
TAUS_TRUEHADTAU_EFF_RNNID_1PRONGSTATSYSTPT2530	
TAUS_TRUEHADTAU_EFF_RNNID_1PRONGSTATSYSTPT3040	
TAUS_TRUEHADTAU_EFF_RNNID_1PRONGSTATSYSTPTGE40	
TAUS_TRUEHADTAU_EFF_RNNID_3PRONGSTATSYSTPT2025	
TAUS_TRUEHADTAU_EFF_RNNID_3PRONGSTATSYSTPT2530	
TAUS_TRUEHADTAU_EFF_RNNID_3PRONGSTATSYSTPT3040	
TAUS_TRUEHADTAU_EFF_RNNID_3PRONGSTATSYSTPTGE40	
TAUS_TRUEHADTAU_EFF_RNNID_HIGHPT	
TAUS_TRUEHADTAU_EFF_RNNID_SYST	
TAUS_TRUEHADTAU_SME_TES_INSITUEXP	
TAUS_TRUEHADTAU_SME_TES_INSITUFIT	
TAUS_TRUEHADTAU_SME_TES_MODEL_CLOSURE	$\tau$ -energy scale
TAUS_TRUEHADTAU_SME_TES_PHYSICSLIST	
TAUS_TRUEELECTRON_EFF_ELEBDT_STAT	
TAUS_TRUEELECTRON_EFF_ELEBDT_SYST	
TAUS_TRUEHADTAU_EFF_ELEOLR_TOTAL	$\tau$ -electron veto
egamma	
EG_RESOLUTION_ALL	egamma resolution
EG_SCALE_ALL	egamma scale
EG_SCALE_AF2	
$E_T^{\text{miss}}$	
MET_SoftTrk_ResoPar a	
MET_SoftTrk_ResoPerp	
MET_SoftTrk_ScaleDown	
MET_SoftTrk_ScaleUp	

Table 7.11: List of experimental systematic uncertainties. Table reproduced from analysis internal notes.

NP Name	Description	
Jets		
JET_EtaIntercalibration_Modelling JET_EtaIntercalibration_TotalStat JET_EtaIntercalibration_NonClosure_highE JET_EtaIntercalibration_NonClosure_lowE JET_EtaIntercalibration_NonClosure_posEta JET_Pileup_OffsetMu JET_Pileup_OffsetNPV JET_Pileup_PTerm JET_Pileup_RhoTopology JET_Flavor_Composition JET_Flavor_Response JET_PunchThrough_MC16 JET_PunchThrough_AFII JET_EffectiveNP_Detector1 JET_EffectiveNP_Detector2 JET_EffectiveNP_Mixed1 JET_EffectiveNP_Mixed2 JET_EffectiveNP_Mixed3 JET_EffectiveNP_Modelling1 JET_EffectiveNP_Modelling2 JET_EffectiveNP_Modelling3 JET_EffectiveNP_Modelling4 JET_EffectiveNP_Statistical1 JET_EffectiveNP_Statistical2 JET_EffectiveNP_Statistical3 JET_EffectiveNP_Statistical4 JET_EffectiveNP_Statistical5 JET_EffectiveNP_Statistical6 JET_SingleParticle_HighPt JET_RelativeNonClosure_AFII JET_BJES_Response JET_EtaIntercalibration_NonClosure_2018data JET_JER_DataVsMC_MC16 JET_JER_DataVsMC_AFII JET_JER_EffectiveNP_1 JET_JER_EffectiveNP_2 JET_JER_EffectiveNP_3 JET_JER_EffectiveNP_4 JET_JER_EffectiveNP_5 JET_JER_EffectiveNP_6 JET_JER_EffectiveNP_7 JET_JER_EffectiveNP_8 JET_JER_EffectiveNP_9 JET_JER_EffectiveNP_10 JET_JER_EffectiveNP_11 JET_JER_EffectiveNP_12restTerm JET_JVT_EFF JET_FJVT_EFF	Jet energy scale (CategoryReduction)  Jet energy resolution (FullJER)  Jet vertex tagging Forward jet vertex tagging	
<i>b</i> -tagging	FT_EFF_Eigen_B_0 FT_EFF_Eigen_B_1 FT_EFF_Eigen_B_2 FT_EFF_Eigen_C_0 FT_EFF_Eigen_C_1 FT_EFF_Eigen_C_2 FT_EFF_Eigen_C_3 FT_EFF_Eigen_Light_0 FT_EFF_Eigen_Light_1 FT_EFF_Eigen_Light_2 FT_EFF_Eigen_Light_3 FT_EFF_extrapolation FT_EFF_extrapolation_from_charm	

Table 7.12: List of experimental systematic uncertainties. Table reproduced from analysis internal notes.

The normalisation acceptance uncertainties are derived comparing the acceptance obtained for each alternative sample to the nominal acceptance:

$$\sigma_A = \frac{A_{\text{var}} - A_{\text{nom}}}{A_{\text{nom}}} = \frac{N_{\text{var}} - N_{\text{nom}}}{N_{\text{nom}}}, \quad (7.5)$$

where  $A_{\text{var}}$  ( $A_{\text{nom}}$ ) is the acceptance of the variation (nominal) sample and  $N_{\text{var}}$  ( $N_{\text{nom}}$ ) is the expected events yields of the variation (nominal) sample. Acceptance uncertainties on the normalisation of the backgrounds are applied in all regions and treated as correlated across regions except for the  $t\bar{t}$  and  $Z + \text{HF}$  backgrounds whose normalisations are left unconstrained (floated) in the global likelihood fit. For these two processes, a relative acceptance uncertainty is needed to take into account the correlation. The relative acceptance uncertainty is estimated from the comparison of the relative amount of events predicted by the alternative model in one region ‘ $i$ ’ with respect to another region ‘ $j$ ’ and the same fraction in the nominal model:

$$\sigma_{A^i/A^j} = \frac{\frac{A_{\text{var}}^i}{A_{\text{var}}^j} - \frac{A_{\text{nom}}^i}{A_{\text{nom}}^j}}{\frac{A_{\text{nom}}^i}{A_{\text{nom}}^j}} = \frac{\frac{N_{\text{var}}^i}{N_{\text{var}}^j} - \frac{N_{\text{nom}}^i}{N_{\text{nom}}^j}}{\frac{N_{\text{nom}}^i}{N_{\text{nom}}^j}}. \quad (7.6)$$

The list of all uncertainties applied to all MC backgrounds is reported in Table B.3 and Table B.4 in Appendix B.3.

#### 7.4.2.1 Uncertainties on $t\bar{t}$

The normalisation of the  $t\bar{t}$  background is estimated from data as included as freely floating parameter in the final fit. This normalisation is correlated across all regions included in the fit and it is determined mainly from the tails of the  $m_{\ell\ell}$  distribution in the  $Z + \text{HF}$  CR which have high purity of  $t\bar{t}$  events.

Relative acceptance uncertainties on the normalisation are applied on  $t\bar{t}$  in the SR to take into account potential differences in the normalisation between the SRs and the CR. The shape variations are checked and applied, correlated with the relative acceptance uncertainties on the normalisations, where they are found to be relevant as described in the following. All these uncertainties are derived by MC-to-MC comparison. The alternative samples are all generated using the fast simulation (AF2), for consistency the fast simulation nominal  $t\bar{t}$  sample is also generated and is used for comparing. The derived uncertainties are then applied on the  $t\bar{t}$  sample generated with full simulation.

Uncertainties due to PDF,  $\alpha_s$ , FSR and scales are evaluated using internal alternative weights present in all the  $t\bar{t}$  nominal samples.

Uncertainties from the parton shower (PS), matrix element (ME) and ISR are estimated

Source	SLT	LTT
ME	+0.0026, -0.0026	-0.009, +0.009
PS	-0.072, +0.072	-0.088, +0.088
ISR	+0.0005, -0.0081	-0.0052, +0.013
FSR	+0.014, -0.0097	+0.0096, -0.032
PDF+ $\alpha_s$	-0.006, +0.006	-0.0073, +0.0073
Total	-0.074, +0.074	-0.094, +0.090

Table 7.13: Relative size of relative acceptance normalisation uncertainties on  $t\bar{t}$ .

using the differences between the nominal samples and the corresponding variations samples: PS hadronisation algorithmic uncertainties are evaluated comparing the nominal samples showered with PYTHIA 8 to alternative samples showered with HERWIG 7; uncertainty due to MEs evaluated comparing the nominal POWHEG +PYTHIA 8 samples to alternative aMC@NLO +PYTHIA 8 samples. ISR up variation is evaluated by comparing the fast simulation nominal POWHEG +PYTHIA 8 ( $h_{\text{damp}} = 1.5 m_{top}$ ) with the fast simulation samples with varied  $h_{\text{damp}}$  ( $h_{\text{damp}} = 3 m_{top}$ ), at the same time dividing by 2 the renormalization and factorization scales and the varying the showering; while the ISR down variation is evaluated by comparing the nominal POWHEG +PYTHIA 8 with the samples obtained doubling the renormalization and factorization scales and the varying the showering.

Relative acceptance uncertainties on the normalisation between the CR and the SR are calculated using Equation 7.6 as reported in Table 7.13.

The shapes of  $t\bar{t}$  PS and ME modelling uncertainties are checked in the MVA input variables and output distributions for the SLT and LTT channels. The shape in ME/PS is uncertainty is found to be insignificant in the NN/PNN scores in the LTT channel and therefore not considered in this analysis. In the SLT channel, in order to parametrise the PS/ME uncertainty in the MVA output distributions, two methods were studied: one using parametrisation in a combination of MVA input variables and one using parametrisation directly in the MVA output distributions. The later were adopted for the final result. Both methods will be discussed in the following.

The first method is based on parametrisation calculated as the ratio of the expected events yields of the variation sample to the nominal sample in bins of **MVA input variables**. For the **ME uncertainty**, obvious trends are seen in  $p_T$  of the 2  $b$ -jets ( $p_T^{bb}$ ) and  $E_T^{\text{miss}}$ . In order to mimic the MVA output distribution of the variation sample, the nominal sample is weighted by the parametrisation in these two distributions, and the weighted distributions are passed to the MVA classification to check the shape of the output. However, the parametrisation using neither the  $p_T^{bb}$  nor  $E_T^{\text{miss}}$  alone recover the shape of the MVA output of the variation sample very well. In addition, all other MVA input variables are checked for parametrisation, but none of them give good closure in MVA output either. Therefore, a sequential parametrisation in  $p_T^{bb}$  and  $E_T^{\text{miss}}$  is checked. The weights are applied on the

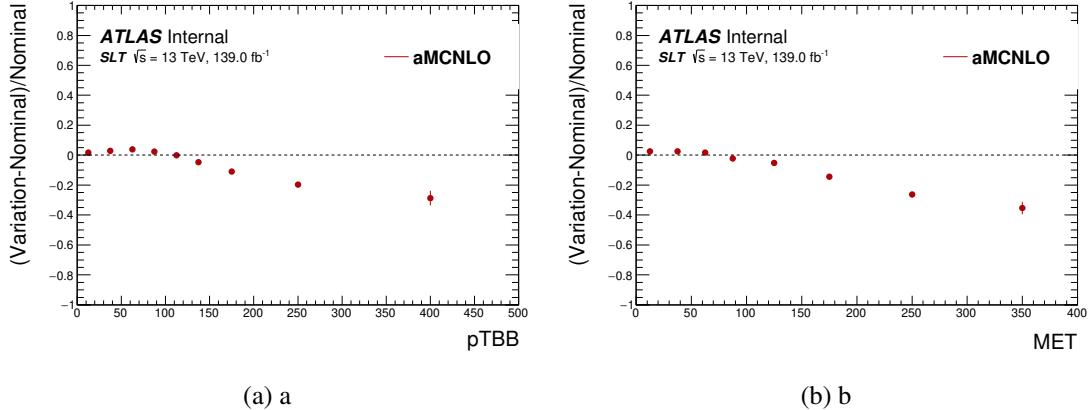


Figure 7.21: The shape-only parametrisation in bins of  $p_T^{bb}$  (a) and in  $E_T^{\text{miss}}$  for the ME uncertainty, derived in the SLT channel.

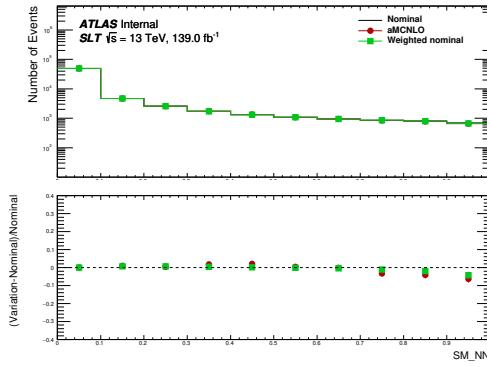


Figure 7.22: SLT channel: shape-only NN score of the ME variation sample and the weighted nominal, using sequential parametrisation in  $p_T^{bb}$  and  $E_T^{\text{miss}}$ . The binning shown here is in equal width bins of NN score.

nominal sample first with  $p_T^{bb}$  parametrisation. The  $E_T^{\text{miss}}$  parametrisation is then extracted by taking the ratio of the  $E_T^{\text{miss}}$  of the variation sample to the  $E_T^{\text{miss}}$  of the weighted nominal sample. The additional step is to take into account the possible change in  $E_T^{\text{miss}}$  after the first weighting step. This sequential parametrisation is shown in Figure 7.21 for the SLT channel.

The  $p_T^{bb}$  and  $E_T^{\text{miss}}$  distributions are checked again for the weighted samples as compared to the variation samples. Good closure is found in both distributions and in other MVA input variables. To check how well the weighted nominal sample can recover the MVA output distributions of the variation sample, the NN score distribution of the variation and the weighted nominal is shown in Figure 7.22, where the events yields of the variation sample are normalised to the ones of the nominal.

The PS uncertainty is first checked with parametrisation in single MVA variable but none of the MVA input alone give good closure in the MVA output distributions. Therefore,

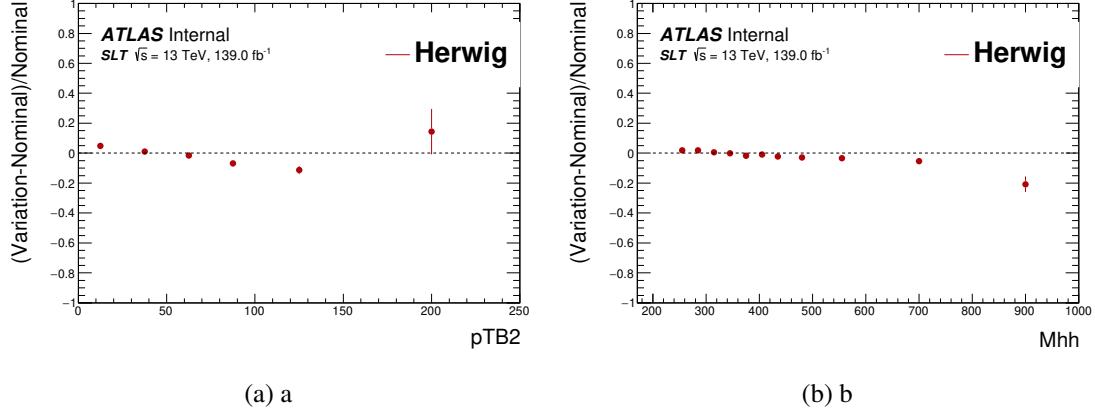


Figure 7.23: The shape-only parametrisation in bins of  $p_T^{b2}$  (a) in  $m_{HH}$  (b), for the PS uncertainty, derived in the SLT channel.

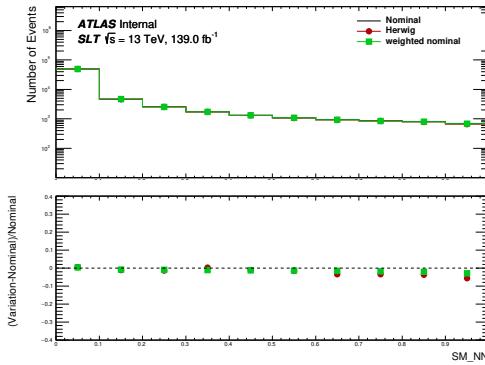


Figure 7.24: The shape-only NN score of the PS variation sample and the weighted nominal. The binning shown here is in equal width bins of NN score.

a similar sequential parametrisation approach is used. The parametrisation is done first in bins of the  $p_T$  of the subleading jet ( $p_T^{b2}$ ) followed by parametrising the residue in bins of mass of the di-Higgs system ( $m_{HH}$ ), as shown in Figure 7.23. The closure of the parametrisation in the NN score is shown in Figure 7.24.

The sequential parametrisation method is compared with the second method, which is based on parametrising directly in the ratio of MVA discriminants of the variation samples to the ones of the nominal. The binning used for parametrisation is the same as the binning used in the final fit. Due to the very fine binning in the high MVA output region, large statistical fluctuations are observed, therefore, smoothing is applied on these systematics in limit setting to reduce the fluctuations effect. The SM NN parametrisation for the ME and PS variation in the SLT channel is shown in Figure 7.25. Plots for more resonant mass points can be found in Appendix B.3.

The main reason for using the second method is to ease the combination to the other channels, and to reduce complexity, even though the second method has some drawbacks.

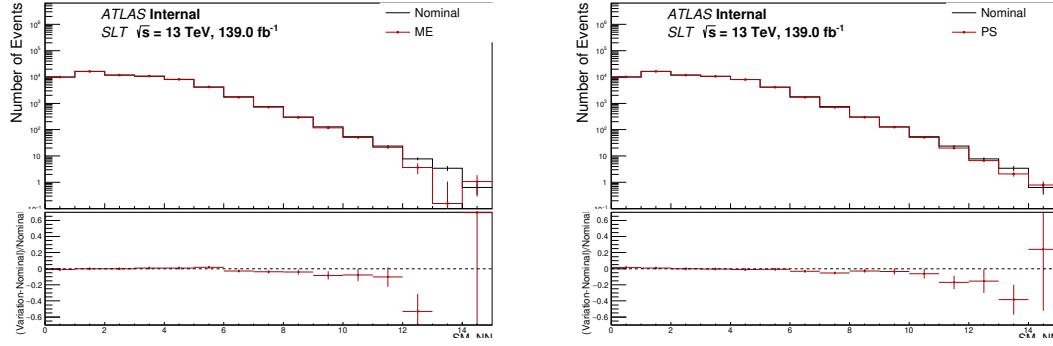


Figure 7.25: Parametrisation in SM NN score distribution for the ME (left) and PS (right) systematics. The binning is the same as the final fit binning.

Also in the high MVA score region, where the binning becomes much finer, the first method starts to fail to cover the large difference between the variation and the nominal samples.

The ISR up and down variations have not shown obvious shape in the NN/PNN distribution therefore only normalisation uncertainty is considered.

#### Uncertainties on $t\bar{t}$ in the $Z + \text{HF}$ control region

Shape uncertainties on the  $t\bar{t}$  background are neglected in the  $Z + \text{HF}$  control region as they are found to be negligible in the  $m_{\ell\ell}$  distribution included in the fit for this region, as shown in Figure B.7 in Apppedix B.3.

##### 7.4.2.2 Uncertainties on $Z + \text{HF}$ background

The normalisation of the  $Z + \text{HF}$  background is estimated from data as included as freely floating parameter in the final fit. This normalisation is mainly determined from the dedicated  $Z + \text{HF}$  dilepton control region. Only  $Z + \text{HF}$  background are considered, including  $Z + bb$ ,  $Z + bc$ ,  $Z + cc$ . Contributions from  $Z + \text{light flavour jets}$  such as  $Z + \text{light jets}$  ( $Z + l$ ),  $Z + bl$  and  $Z + cl$  are excluded.

Uncertainties due to choice of generators are evaluated by comparing the SHERPA 2.2.1 samples to the MADGRAPH +PYTHIA 8 samples. Uncertainties on the  $Z + \text{jets}$  background modelling related to the choice of renormalisation and factorisation scales are evaluated using event weights included in the SHERPA 2.2.1 samples, varying the scales either together or independently up and down by a factor of two, leading to 7-point scale variations. The uncertainty due to the choice of PDF set is evaluated in a similar way, using the event weights included in the samples. The PDF variations include 100 replicas of the nominal NNPDF3.0 PDF set as well as central values for two different PDF set, MMHT2014nnlo and CT14nnlo, and two NNPDF3.0nnlo  $\alpha_s$  variations. The NNPDF intra-PDF uncertainty is

Source	SLT	LTT
ME	+0.021, -0.021	+0.10, -0.10
Scales	-0.029, +0.053	-0.054, +0.085
CKKW	+0.07, -0.07	+0.071, -0.071
QSF	-0.016, +0.016	-0.016, +0.016
PDF+ $\alpha_s$	-0.0026, +0.0026	-0.0033, +0.0033
PDF choice	-0.0097, +0.0097	-0.011, +0.011
Total	-0.081, +0.092	-0.14, +0.15

Table 7.14: Relative size of relative acceptance normalisation uncertainties on  $Z + \text{HF}$  for the di-Higgs analysis.

estimated as the standard deviation of the set of 101 NNPDF3.0 sets. The envelope of the differences between the nominal NNPDF set and the other two PDF sets is taken as an additional uncertainty.

From the nominal Sherpa configuration there are two other parameters that can be varied to investigate uncertainties on the modelling of the  $W/Z+\text{jets}$  process: MEmatching ( $ckkw$ ), which varies the scale taken for the calculation for the overlap between jets from the MEand the PS. The nominal value for this parameter is 20 GeV. The up variation increases this to 30 GeV, while the down variation decreases the nominal value to 15 GeV. The second parameter is the resummation scale ( $qsf$ ), which varies the scale used for the resummation of soft gluon emission  $\mu_{qsf}$  is varied from 2 and with respect to the nominal.

Relative acceptance normalisation uncertainties between the CR and the SR are calculated from the single channel acceptances using Equation 7.6 as reported in Table 7.14.

For the scale variation, the ratios of the normalised MVA output distribution of the variation to the ones of the nominal are checked. Obvious shape is observed, and a similar parametrisation approach is used. All MVA input variables were considered for the parametrisation. Instead of parametrising all seven points of variation, the parametrisation is considered for the envelope encapsulating the maximum shift of the variation. The variable to parametrise is chosen such that the weighted nominal samples can best recover the envelop covering the maximum shift in the PNN scores. As a result, the  $Z + \text{HF}$  scale variation is parametrised by the  $p_T^{bb}$ , as shown in Figure 7.26. In the figure, the weighted nominal samples distributions are labelled as ‘para:pTBB Up(Down)’ and the maximum shift of the 7-point variations in the MVA output distributions are covered by the envelope labelled as ‘Envelop\_Up(Down)’.

the PNN score distribution is shown in Fig.7.27 for the SLT channel and Fig.7.28 for the LTT channel. Good closure within statistical uncertainty is observed using this parametrisation.

For the uncertainty due to generator variation, the ratios of the variance to the nominal are checked in the MVA output. For most of the bins and mass points, no significant

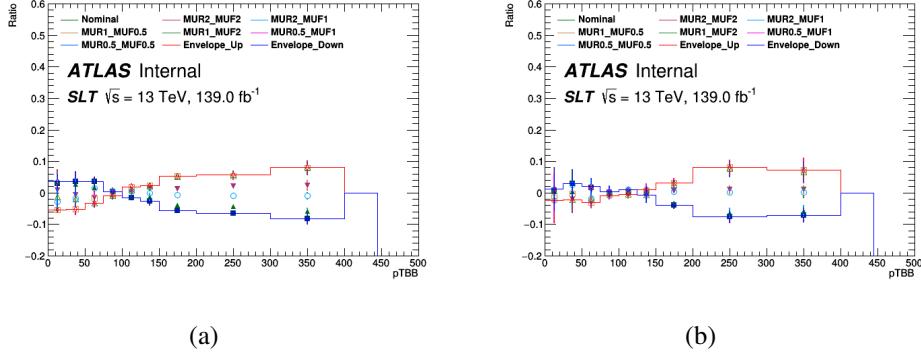


Figure 7.26: SLT (left) and LTT channel (right): Taking an envelope of the  $Z + \text{HF}$  scale variation with  $p_T$  of the 2  $b$ -jets.

deviation from unity is observed within the statistical uncertainty. Similarly, the shape dependence on the uncertainty due to choice of intra/inter-PDF set,  $\alpha_s$  variations, ckkw and qsf are checked. No obvious trend is observed. The normalised ratio of the PNN distribution of the variation mentioned above to the nominal are shown in Appendix B.3, in Figure B.8 (B.9) for the generator uncertainty, Figure B.10 (B.11) for the intra-PDF set uncertainties, Figure B.12 (B.13) for the inter-PDF set and  $\alpha_s$  uncertainty, Figure B.14 (B.15) for the ckkw and qsf uncertainties in the SLT (LTT) channel. The yields of the up and down variation for the ckkw and qsf uncertainties are both smaller than the nominal sample yields, and the solution for this issue is to normalize the central value of the up and down variation to the nominal sample, and take the difference from there.

## Uncertainties on $Z + \text{HF}$ background in the $Z + \text{HF}$ control region

Shape uncertainties on the  $Z + \text{HF}$  background are neglected in the  $Z + \text{HF}$  control region as they are found to be negligible in the  $m_{ll}$  distribution included in the fit for this region, as shown in Appendix B.3.

### 7.4.2.3 Uncertainties on single top

All single-top uncertainties are derived by MC-to-MC comparison, and split into the normalisation acceptance uncertainties and the shape uncertainties.

As the  $Wt$ -channel contribution dominates over the  $t$ -channel and  $s$ -channel contributions, only the contribution from the  $Wt$ -channel is considered. Uncertainties due to PDF, ISR and FSR are evaluated using internal alternative weights present in all singletop nominal samples. Using the PDF4LHC15\_30 PDF set, the uncertainties are estimated by combining the difference between the error sets and the nominal set following the recommendations in Ref.[80].

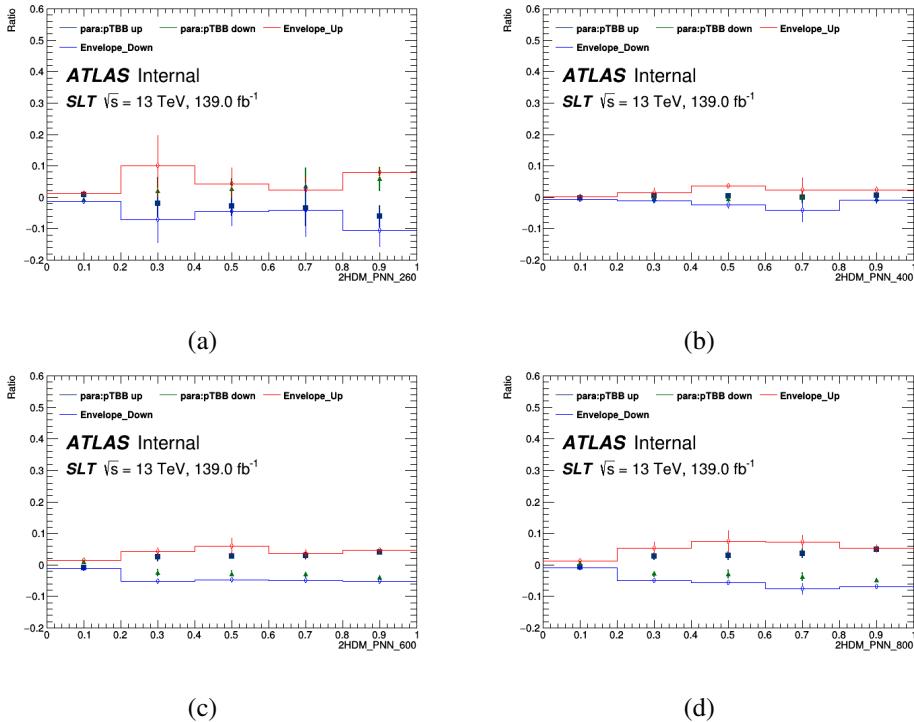


Figure 7.27: SLT channel: PNN output distributions of the nominal sample weighted by parametrisation in  $p_T^{bb}$ , labelled as ‘para:pTBB Up(Down)’ versus envelope encapsulating the maximum shift of the variation for various resonant mass.

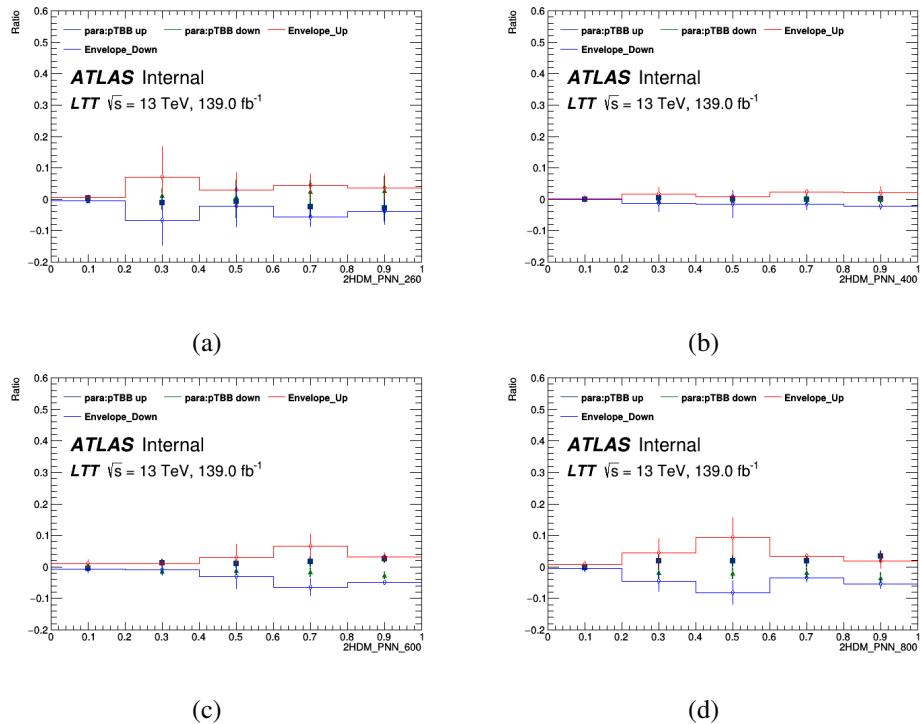


Figure 7.28: LTT channel: PNN output distributions of the nominal sample weighted by parametrisation in  $p_T^{bb}$ , labelled as ‘para:pTBB Up(Down)’ versus envelope encapsulating the maximum shift of the variation for various resonant mass.

Source	SLT	LTT
ME	-0.022, +0.022	-0.15, +0.15
PS	+0.077, -0.077	-0.093, +0.093
Single top interference	+0.078, -0.078	+0.11, -0.11
ISR	-0.047, +0.064	-0.045, +0.062
FSR	-0.054, +0.043	-0.069, +0.035,
PDF	-0.032, +0.032	-0.032, +0.032
Total	$\pm 13.7\%$	$\pm 21.1\%$

Table 7.15: Size of normalisation acceptance uncertainties for single-top background for the  $HH$  analysis.

Uncertainties from the PS, ME and single top interference are estimated using the differences between the nominal samples and the corresponding alternative samples. The PS uncertainty is evaluated comparing the AF2 nominal samples showered with **Pythia** 8 with alternative AF2 samples showered with **Herwig** 7. The ME uncertainty is evaluated comparing the AF2 nominal PowHEG +**Pythia** 8 samples to the alternative AF2 aMC@NLO +**Pythia** 8 samples. Uncertainty due single-top interference is evaluated comparing the full simulated nominal PowHEG +**Pythia** 8 samples with diagram removal (DR) scheme to the alternative full simulation PowHEG +**Pythia** 8 sample with diagram subtraction (DS) scheme.

The normalisation acceptance uncertainties are shown in Table 7.15.

The shape uncertainty is derived separately for the SLT channel and the LTT channels. Only normalisation acceptance uncertainty is considered for the PS and ME uncertainties as no obvious shape is observed in the NN or the PNN scores, as shown in Appendix B.3 in Figure B.17, B.19 in the SLT channel and in Figure B.18, B.20 in the LTT channel for the PS, ME uncertainties, respectively.

For the single-top interference uncertainty, the parametrisation is extracted from the ratio of the variation to the nominal fullsim samples in bins of  $p_T^{bb}$  where it shows the most obvious trend and largest shift from unity. Other MVA input are also tested, but most of them either do not have an obvious shape or the variation size is much smaller compared to the  $p_T^{bb}$ , and therefore cannot provide sufficient closure in the MVA output distributions. The parametrisation is shown in Figure 7.29 (a), (b) for the SLT and the LTT channels respectively. The parametrisation is applied on the nominal sample as weights and the weighted sample is passed through the MVA classification. The PNN score distributions of the weighted sample, and of the variation sample are shown in Fig 7.30. The closure in the PNN scores for various resonance mass are shown in Appendix B.3 in Figure B.21 and in Figure B.22 for the SLT and LTT channels, respectively. Good closure is seen across various mass points of PNN classification and in most of the bins, therefore the parametrisation in  $p_T^{bb}$  is adopted.

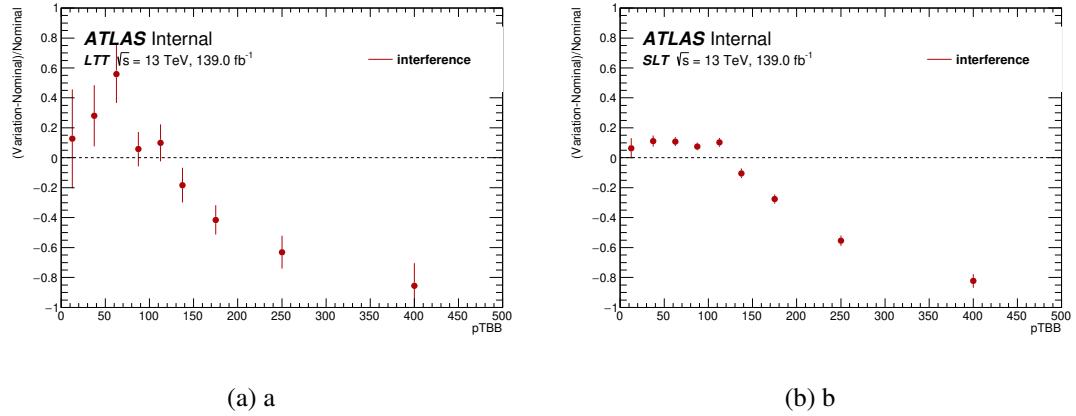


Figure 7.29: LTT (a) and SLT channels (b): shape only Parametrisation in bins of  $p_T^{bb}$  distribution for  $Wt$  DS/DR interference uncertainty.

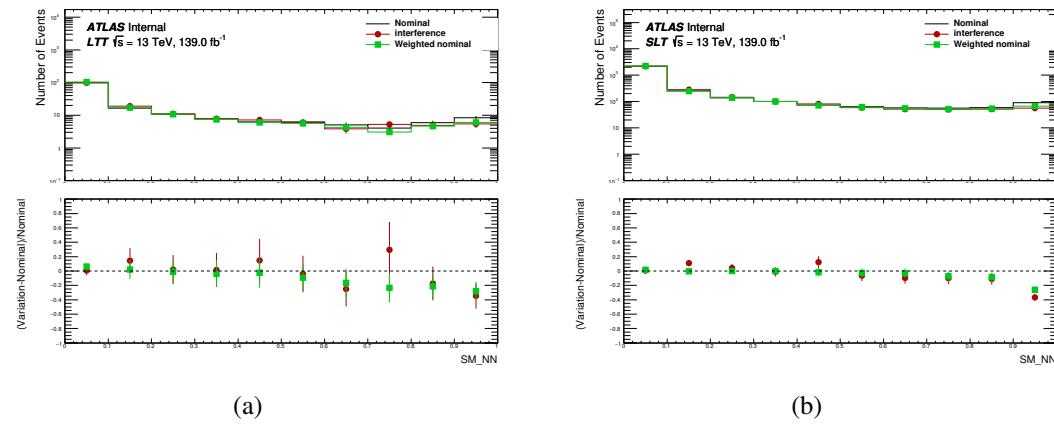


Figure 7.30: LTT (left) and SLT channels (right): shape only SM NN score of the  $Wt$  DS/DR interference uncertainty .

#### 7.4.2.4 Uncertainties on single-Higgs

Uncertainties on cross-section and BR are applied to all single-Higgs processes included in the analysis as backgrounds (following recommendations from LHC Higgs Working Group [84]).

For all single-Higgs processes, uncertainties arised from PDF,  $\alpha_s$  and scales are derived using internal alternative weights present in all the single-Higgs nominal samples.

For PDF and  $\alpha_s$  variations, the PDF4LHC\_NLO\_ 30 PDF set is used for evaluation. The PDF set is comprised of 30 eigenvectors parameterising the uncertainties for all the PDFs and 1 parameter for the  $\alpha_s$  variations. The variations of the PDF and  $\alpha_s$  are combined by taking the square root of the quadrature sum of the 30 eigenvectors variations and the  $\alpha_s$  variation in bins of PNN score distributions. The combined variation is then compared with the nominal PDF4LHC\_NLO\_ 30 PDF to evaluate the uncertainties due to PDF and  $\alpha_s$ . The method used here follows the PDF4LHC recommendations [80].

Variations of the renormalisation and factorisation scales are used to estimate the uncertainty due to missing higher order corrections. For these scale variations, the 7-point scale variations of the renormalisation ( $\mu_R$ ) and the factorisation scale ( $\mu_F$ ) are used and they are combined by taking an envelope of all of the variations, similar to the method used in section 7.4.2.2. PS hadronisation algorithmic uncertainties are evaluated comparing samples showered with Pythia 8 and Herwig 7 (with the ME provided by the same generator).

For the ttH process, the ME NLO matching uncertainty is evaluated comparing the nominal Powheg +Pythia 8 samples to alternative aMC@NLO +Pythia 8 samples. Internal alternative weights are present in the nominal samples to evaluate the eigentune variations for ISR and FSR variations. The Var3c eigenvariation of the A14 tune, representing the variation of the strong coupling in the initial state shower, is used to evaluate the ISR uncertainty. The FSR uncertainty is evaluated using variations of  $\mu_R$  by a factor 2 up and down around the nominal value in the PS.

For the ZHbb process, the ISR and FSR uncertainties are found to be negligible in the VHbb analysis as described in Ref [151], thus they are neglected in this analysis.

The size of the acceptance uncertainties on the normalisations are reported in Table 7.16. Normalisation uncertainties smaller than 1% are neglected in the analysis. After taking out the normalisation effects, shape effects were checked on the PNN score distributions and found to be negligible and therefore are not considered in the analysis.

#### 7.4.2.5 Uncertainties on other minor backgrounds

For minor backgrounds, i.e. single-top (s- and t-channels), Z+lf, W+jets and Diboson, acceptance uncertainties are only applied on the normalisation and their size is taken

Process	SLT	LTT	Description
ttH	-0.006,+0.006	-0.008,+0.008	PDF+ $\alpha_s$
ttH	-0.003,+0.003	-0.006,+0.006	Scales
ttH	-0.013,+0.013	-0.067,+0.067	PS
ttH	-0.001,+0.001	-0.01,+0.01	ISR
ttH	-0.051,+0.032	-0.15,+0.055	FSR
ttH	-0.0025,+0.0025	-0.019,+0.019	ME NLO matching
ZHbb	-0.006,+0.006	-0.005,+0.005	PDF+ $\alpha_s$
ZHbb	-0.030,+0.030	-0.025,+0.025	Scales
ZHbb	-0.11,+0.11	-0.037,+0.037	PS
ZHtautau	-0.0077,+0.0077	-0.012,+0.012	PDF+ $\alpha_s$
ZHtautau	-0.022,+0.022	-0.028,+0.028	Scales
ZHtautau	-0.055,+0.055	-0.15,+0.15	PS

Table 7.16: Relative size of single-Higgs acceptance uncertainties in the di-Higgs analysis. Table reproduced from analysis internal notes.

from the estimation performed in the SM VHbb analysis [150, 151]. On single-top production an acceptance uncertainty of 20% is applied in the s- and t-channels. An acceptance uncertainty of 23% is applied on Z+lf. On W+jets an acceptance uncertainty of 37% is applied in order to cover in addition for the fake- $\tau$  contribution (the uncertainty on the  $\tau$ -fakes in  $W$ +jets was estimated comparing the MC and the data-driven prediction for  $W$ +jets fakes in the 0b-tag region of the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel to be 31% in the previous iteration of this analysis). Acceptance uncertainties of 25%, 26% and 20% are applied on  $WW$ ,  $WZ$  and  $ZZ$  respectively.

### 7.4.3 Signal uncertainties

Acceptance uncertainties on the resonant and non-resonant signals are evaluated by comparing the nominal signal samples to alternative MC samples for PS, PDF and scale variations.

#### 7.4.3.1 Resonant signals

The PS uncertainties for the resonant signals are estimated by comparing the nominal samples showered with `Herwig` 7 to alternative samples showered with `Pythia` 8 which are available in full-simulation for the  $m_X = 500$  GeV and the  $m_X = 1000$  GeV mass points. The PS normalisation acceptance uncertainty is found to be 6% (3%) for  $m_X = 500$  ( $m_X = 1000$ ) GeV in the SLT and 4% (7%) for  $m_X = 500$  ( $m_X = 1000$ ) GeV in the LTT. Thus, a conservative 6% (7%) uncertainty is adopted on the normalisation for all mass points in the SLT (LTT) channel.

After taking out the normalisation effect, the acceptance uncertainty on the shape of the PNN distribution is evaluated. Figure B.23 and B.24 show the comparisons of the

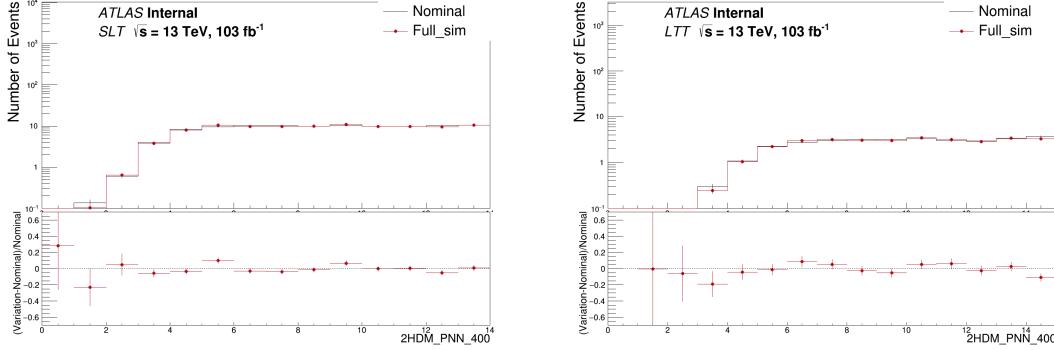


Figure 7.31: Comparison of the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  signal PNN distributions obtained from the nominal (AF2) and alternative (full simulation) signal samples for the  $m_X = 400$  GeV for the SLT (left) and the LTT (right) channel. The binning is the same as the final fit binning.

PNN distributions obtained from the nominal (black) and alternative (blue) signal samples for the  $m_X = 500$  GeV and the  $m_X = 1000$  GeV mass points for the SLT and LTT SRs respectively. A linear fit to the ratio of the two distributions is performed to parameterise the shape variation as a function of the PNN score. The linear function in the PNN score obtained from the fit performed at  $m_X = 500$  GeV which has the largest slope, and this function is used to obtain the templates for the variations for all mass points. Relevant plots can be found in Appendix B.3, Figure B.23 and Figure B.24.

Scale, PDF and  $\alpha_s$  uncertainties on signal are found to be negligible (less than 1%) and therefore not included in the analysis.

The resonant samples are generated with AF2 for masses up to 1 TeV (from 1.1 TeV full simulation is used). Therefore the difference between the AF2 samples and the full simulation samples is checked, using the 400 GeV mass point sample corresponding to 2017 and 2018 data-taking periods. The difference is less than 1% in the SLT channel and 2.9% in the LTT channel, and there is no obvious shape difference between the two types of simulation, as shown in Figure 7.31. The larger difference in acceptance times efficiency in LTT is originating from the  $\tau_{\text{had}}$  used in LTT triggers which no dedicated recommendations for AF2 exist. Therefore a 3% acceptance uncertainty is assigned to LTT of all signal samples simulated with AF2 (up to 1TeV mass point).

All acceptance uncertainties on the signal normalisation obtained in the different signal regions are shown in Table 7.17.

#### 7.4.3.2 Non-resonant signal

**ggF non-resonant** The PS uncertainties for the ggF non-resonant signal are estimated by comparing the nominal samples showered with **Pythia 8** to alternative samples showered with **Herwig 7**. The overall PS acceptance uncertainty on the normalisation is found to be

Process X, GeV	SLT	LTT	Description
$m_X = 400$	<0.01	0.029	AF2
$m_X = 500$	0.06	0.05	PS
$m_X = 1000$	0.03	0.07	PS

Table 7.17: List and relative size of the ggF HH resonant signal acceptance uncertainties in SLT and LTT. Table reproduced from analysis internal notes.

Process	SLT	LTT	Description
SM	0.076	0.075	PS
SM	0.0061	0.0072	PDF+ $\alpha_s$
SM	0.012	0.010	Scales

Table 7.18: List and relative size of ggF HH non-resonant signal acceptance uncertainties in SLT and LTT.

7.6% in SLT and 7.5% in LTT. Uncertainties from PDF,  $\alpha_s$  and scales are evaluated using alternative weights available in the nominal samples.

PDF and  $\alpha_s$  uncertainties are evaluated using the alternative weights for the PDF4LHC\_NLO\_30 uncertainty sets which consist of 30 eigenvectors for all the PDFs and 1 parameter for the  $\alpha_s$  variations. The variations are combined using the same method as mentioned in section 7.4.2.4 for combining the PDF and  $\alpha_s$  variations for single-Higgs processes. The combined variations have less than 1% shift from the nominal and therefore the uncertainties due to PDF and  $\alpha_s$  are neglected.

Scale uncertainties are evaluated using the 7-point scale variations of the renormalisation ( $\mu_R$ ) and the factorisation scale ( $\mu_F$ ), an envelope is used to encapsulate all of the variations. The normalisation uncertainty of the envelope is 1.2% (1%) in SLT (LTT).

The acceptance normalisation uncertainties are summarised in Table 7.18. After taking out the normalisation effects, shape effects were checked on the MVA score distributions and found to be negligible and therefore are not considered in the analysis.

**VBF non-resonant** The PS uncertainties for the VBF signal sample is evaluated by comparing the Pythia 8 nominal sample to an alternative sample showered using Herwig 7. The normalisation uncertainty is found to be 6.3% (2.1%) for the SLT (LTT) channel. No significant shape impact was observed on the final discriminants.

Scale uncertainties are evaluated using the internal weights of 7-point scale variations, the normalisation uncertainty are found be negligible.

PDF and  $\alpha_s$  uncertainties are evaluated using the alternative weights for the PDF4LHC\_NLO\_30 uncertainty sets which consist of 30 eigenvectors for all the PDFs and 1 parameter for the  $\alpha_s$  variations. The variations are combined using the same method as mentioned in the previous section. The normalisation uncertainty is less than 1% and therefore is neglected.

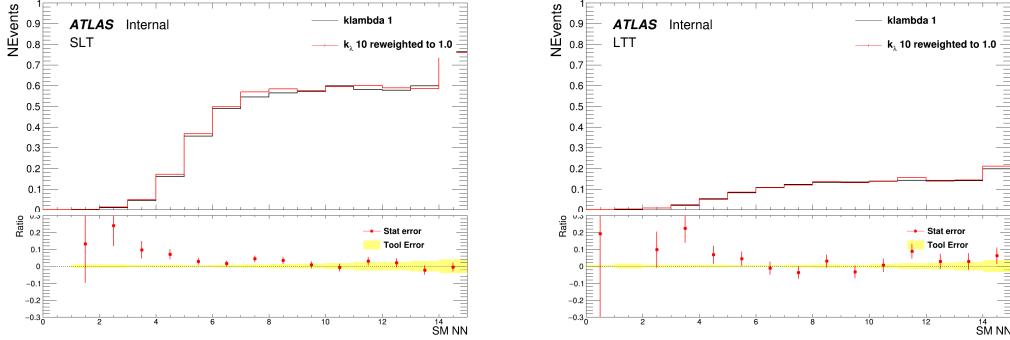


Figure 7.32: Comparison of the  $\kappa_\lambda = 1$  and 10 ggF samples in the NN output for the SLT (left) and the LTT (right) channel. The “Tool Error” band shows the statistical uncertainty on the applied weights, originating from the limited size of the samples used to derive them.

#### 7.4.3.3 Uncertainties on $\kappa_\lambda$ scan

As described in section 5.3.2, the non-resonant  $HH$  signal assuming various  $\kappa_\lambda$  values are produced by a reweighting method. The uncertainties of each sample need to be accounted for. This section outlines how the uncertainties on the various  $\kappa_\lambda$  samples via ggF and VBF production are calculated.

**Uncertainties on ggF** The reweighting method comes with an additional systematic uncertainty, covering the statistical uncertainty on the  $\kappa_\lambda$  weights. This uncertainty originates from the limited size of the samples used to derive them. As the weights are derived using truth-level samples, the reweighting method needs to be checked as additional shape might be introduced after the event selection. This is validated by comparing the SM non-resonant  $HH$  sample with the  $\kappa_\lambda = 10$  sample reweighted to  $\kappa_\lambda = 1$ . The comparison is shown in Figure 7.32. Nice closure is observed between the two distributions.

Nevertheless, it was found that the choice of sample used as input for the reweighting procedure affects the obtained signal acceptance at different values of  $\kappa_\lambda$ . This dependence is shown in Fig. 7.33, using either the generated  $\kappa_\lambda = 1$  or  $\kappa_\lambda = 10$  sample as input. The maximal discrepancy of 3% (SLT) or 8% (LTT) between the measured values is applied as an uncertainty on the normalisation of the ggF sample after the reweighting. At last, the acceptance uncertainties originating from the choice in PS, PDF and  $\alpha_s$  value or QCD scales have been derived for both the  $\kappa_\lambda = 1$  and the  $\kappa_\lambda = 10$  sample. They are summarised in Tab. 7.19. Since these could not be evaluated for all  $\kappa_\lambda$  values investigated in the scan, the greater uncertainty is applied for all signals with  $\kappa_\lambda$  not equal to 1 for each of the three considered sources of uncertainty.

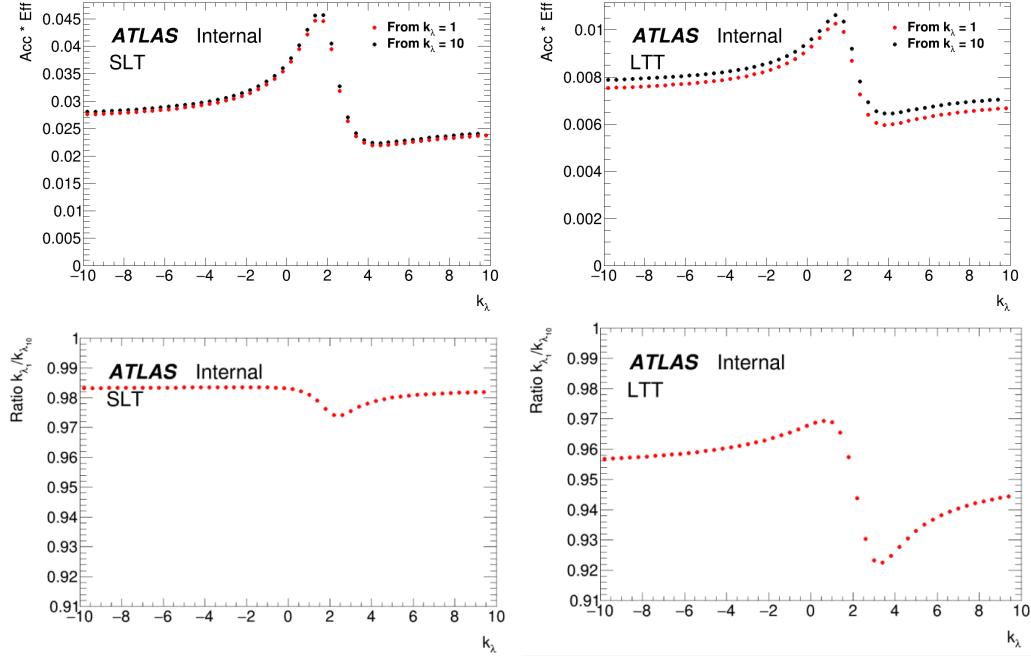


Figure 7.33: Comparison of acceptance  $\times$  efficiency at different  $\kappa_\lambda$  values depending on whether the reweighting procedure is applied on the ggF sample with  $\kappa_\lambda = 1$  (nominal) or  $\kappa_\lambda = 10$  (alternative).

Process	SLT	LTT	Description
SM	0.076	0.075	PS
SM	0.0061	0.0072	PDF+ $\alpha_s$
SM	0.012	0.010	Scales
$\kappa_\lambda = 10$	0.10	0.085	PS
$\kappa_\lambda = 10$	0.0074	0.0072	PDF+ $\alpha_s$
$\kappa_\lambda = 10$	0.0087	0.0074	Scales

Table 7.19: List and relative size of ggF HH non-resonant signal acceptance uncertainties in SLT and LTT.

**Uncertainties on VBF** As in the case of ggF, the acceptance uncertainties for VBF  $HH$  production are updated. The QCD scale and PDF+ $\alpha_s$  uncertainties for the SM and  $\kappa_\lambda \in \{0, 2, 10\}$  samples are listed in Tab. 7.20. Additionally, the PS uncertainty is evaluated for the SM sample and at  $\kappa_\lambda = 10$ . For each of these sources of uncertainty, the largest one is used to define the respective variation for the prediction in all non-SM cases.

An uncertainty on the closure of the linear combination of VBF samples is evaluated in both SLT and LTT and applied on the normalisation of the VBF sample after the linear combination. Four different bases of  $\kappa_\lambda$  values are tested. Closure plots are shown in Figure B.25 in Appendix B.3. The maximal difference is found to be of 2.2% (SLT) and 5.8% (LTT) between the multivariate discriminator distributions for the MC and reweighted  $\kappa_\lambda$  samples, and it comes from the linear combination of VBF samples ( $\kappa_\lambda = 1, 10, 0$ ) to

$\kappa_\lambda = 2$  in all cases.

Process	SLT	LTT	Description
SM	6.3 %	2.1 %	PS
SM	0.15 %	0.23 %	PDF+ $\alpha_s$ from NNPDF30NLO
SM	0.86 %	0.63 %	Scales
$\kappa_\lambda = 0$	0.28 %	0.21 %	PDF+ $\alpha_s$ from NNPDF30NLO
$\kappa_\lambda = 0$	1.16 %	0.91 %	Scales
$\kappa_\lambda = 2$	0.27 %	0.23 %	PDF+ $\alpha_s$ from NNPDF30NLO
$\kappa_\lambda = 2$	1.08 %	0.89 %	Scales
$\kappa_\lambda = 10$	7.79 %	5.73 %	PS
$\kappa_\lambda = 10$	0.26 %	0.37 %	PDF+ $\alpha_s$ from NNPDF30NLO
$\kappa_\lambda = 10$	1.25 %	0.99 %	Scales

Table 7.20: List and relative size of VBF HH non-resonant signal acceptance uncertainties in SLT and LTT. Table reproduced from analysis internal notes.

#### 7.4.4 Uncertainties on fake background

The following uncertainties are considered for the fake background:

- The statistical uncertainty on the  $FF_{t\bar{t}}$ ,  $FF_{\text{QCD}}$  and  $r_{\text{QCD}}$  are considered. They are propagated to the final estimations of the fake background.
- A conservative 30% uncertainty is assigned to all non- $t\bar{t}$  background being subtracted from the data. This uncertainty is derived by varying the FF up and down by 30% when applying on the non- $t\bar{t}$  background passing the anti-ID selection. A value of 50% was also considered, but was found to be over-conservative. As the anti-ID selection events are dominated by fake and  $t\bar{t}$  background, the 30% value is well enough to cover possible variations due to non- $t\bar{t}$  backgrounds.
- The uncertainty due to  $t\bar{t}$  modelling was estimated by the difference between the fake background derived with and without the  $t\bar{t}$  reweighting (more details in section 7.2.1). An alternative approach was studied by the author, using the  $t\bar{t}$  variation samples used for evaluating the PS, ME, ISR and FSR uncertainties to replace the nominal  $t\bar{t}$  samples, for deriving the FF and to be subtracted from data. The fake background derived using the variation samples are correlated to the corresponding ones in  $t\bar{t}$  modelling uncertainties estimations. The PNN score distributions of the fake background derived using the variation samples are shown in Appendix B.3, Figures B.26- B.33.

These two methods were compared by the author, where the first one was adopted at last due to the variation is observed to be higher and hence covering all fake background derived using variation samples. In addition, these two methods are

accounting for the same source of uncertainty originating from the MC modelling of the  $t\bar{t}$  background, therefore only the first method was used to avoid double counting.

- The  $r_{\text{QCD}}$  is highly sensitive to the  $t\bar{t}$  background. Given the fact that the  $FF_{\text{QCD}}$  and  $FF_{t\bar{t}}$  is very similar, the  $r_{\text{QCD}}$  in practice does not have big impact on the combined fake factor. Therefore the uncertainty on  $r_{\text{QCD}}$  is estimated by varying the value from 0 to 0.5. The impact of assigning such a conservative uncertainty was checked and was found to be very small.

#### 7.4.5 Summary of Systematics Uncertainties

In Figure B.49, the rankings of the nuisance parameters considered in the combined fit of SLT and LTT (more details in the next section 7.5.1) are shown. The nuisance parameters are arranged in descending order of their fractional impact on the non-resonant  $HH$  production signal strength  $\mu$ ,  $\Delta\mu/\Delta\mu_{tot}$ . The  $\Delta\mu/\Delta\mu_{tot}$  is represented in the horizontal bars in the plot with values reading on the top blue x-axis, with the dashed (un-dashed) blue area representing the fractional impact on  $\mu$  when the nuisance parameter is varied up (down) by one standard deviation (with this parameter fixed to the variation and fit again allowing all other nuisance parameters to vary). The black round dots reading on the bottom x-axis, represent the shift of the fitted value  $\theta$  from its nominal value  $\theta_0$ , with respect to its standard deviation  $\Delta\theta$ . The error bars corresponding to the dots are the fitted uncertainties, relative to the nominal uncertainties. The red open circle reading on the bottom x-axis represents the fitted value of the  $t\bar{t}$  and  $Z + \text{HF}$  normalisation which are freely floating in the fit, with the error bars representing the uncertainty. The statistical uncertainty is also included as nuisance parameters with Poissonian priors. The name ‘LepHad\_SLT(LTT)\_SR\_MVA\_bin\_14’ represents the statistical error on the 14<sup>th</sup> bin in the SM NN score distribution.

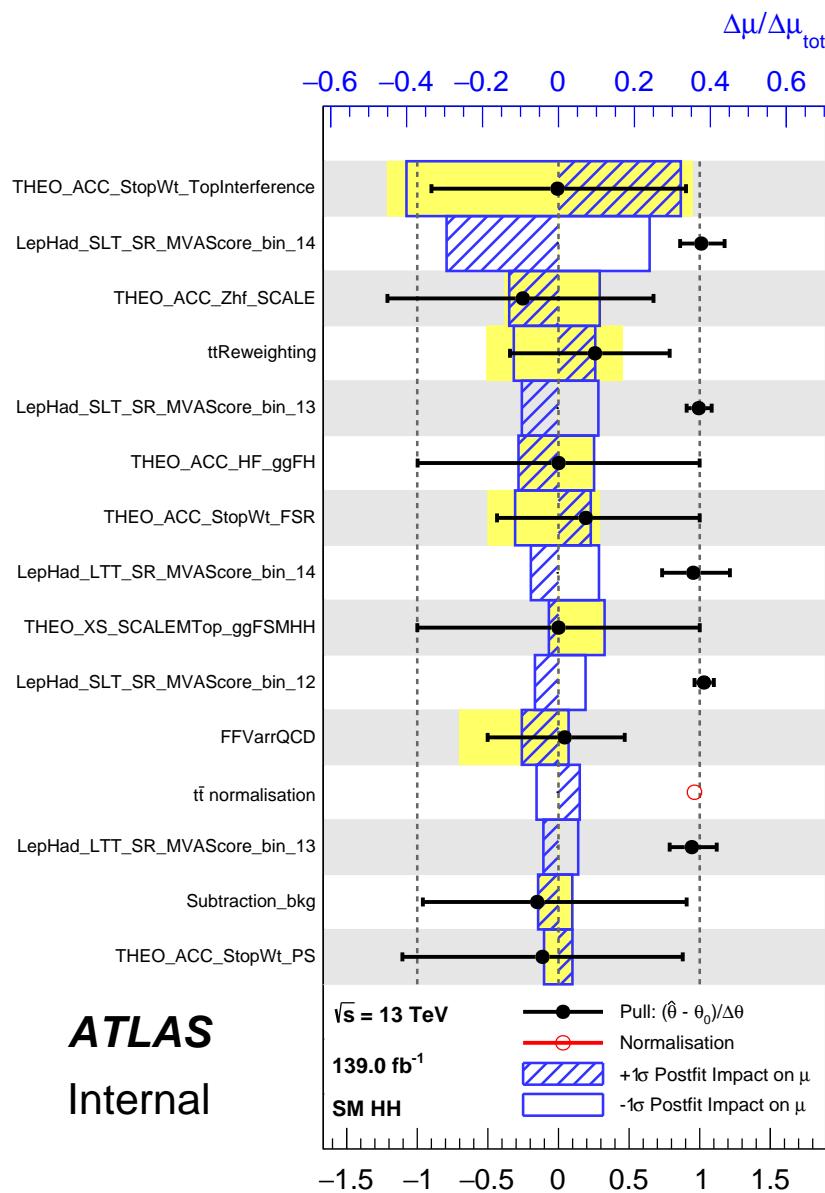


Figure 7.34: Nuisance parameters rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit to data for the SM fit.

## 7.5 Results

The set up used for the profile likelihood fit is described in section 7.5.1. In section 7.5.2, results obtained from  $\tau_{\text{lep}}\tau_{\text{had}}$  only fit (using only the  $\tau_{\text{lep}}\tau_{\text{had}}$  signal regions and the  $Z + \text{HF}$  control region) are presented. The result from the likelihood fit shows that the data is compatible with the null hypothesis for all signals. Therefore, the data is used to set upper limits at 95% C.L. on the signal cross-sections. The likelihood fit is performed again on the  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  signal regions and the  $Z + \text{HF}$  control region, and the results are presented in section 7.5.3. The fit results in  $b\bar{b}\tau^+\tau^-$  channel are then combined with the  $b\bar{b}\gamma\gamma$  channel for upper limits on the non-resonant SM  $HH$  production cross-section and limits on the self-coupling modifier  $\kappa_\lambda$ , and with the  $b\bar{b}\gamma\gamma$ ,  $b\bar{b}b\bar{b}$  channels for upper limits on the narrow-width scalar  $X$  production cross-section as a function of resonance mass. These combinations results are presented in section 7.5.4.

### 7.5.1 Fit setup

The profile likelihood fit is first performed on the MVA output in the  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT, LTT regions, together with the  $m_{ll}$  distribution in the  $Z + \text{HF}$  control region, referred to as  $\tau_{\text{lep}}\tau_{\text{had}}$  only fit. After checking the  $\tau_{\text{lep}}\tau_{\text{had}}$  only fit results, a likelihood fit is then performed simultaneously on the MVA output in all three signal regions, the  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT, LTT and the  $\tau_{\text{had}}\tau_{\text{had}}$  regions, together with the  $m_{ll}$  distribution in the  $Z + \text{HF}$  control region. As described in Section 7.3, for the resonant  $HH$  production, PNNs are used for signal extraction, and for the non-resonant production, a neural network is used in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel (a Boosted Decision Tree, BDT is used for the non-resonant production in the  $\tau_{\text{had}}\tau_{\text{had}}$  signal region).

The normalisation of the  $t\bar{t}$  background and the normalisation of the  $Z + \text{HF}$  background are determined in the fit as freely floating parameter. Relative acceptance uncertainties between CR and SRs are applied on these normalisations in the signal regions included in the fit as described in Section 7.4.5.

The Parameter Of Interest (POI) is the signal strength  $\mu$ . As described in section TODO: ref the theory section For the non-resonant SM case, the  $\mu$  is relative to the ggF+VBF input signal cross-section of  $(31.05 + 1.726)$   $\text{fb} \times \text{BR}_{HH \rightarrow b\bar{b}\tau\tau} = 32.776$   $\text{fb} \times \text{BR}_{HH \rightarrow b\bar{b}\tau\tau}$ , and for the resonant case, the  $\mu$  is relative to the input cross-section of  $1 \text{ pb} \times \text{BR}_{HH \rightarrow b\bar{b}\tau\tau}$ .

In the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel, the MVA outputs are binned in 1090 bins, where 990 bins with equal width are in the range from 0 to 0.99 and 100 bins with equal width are in the range from 0.99 to 1. A rebinning procedure is performed on the MVA output. First, a rebinning

metric is defined as:

$$Z(I[k, l]) = 10 \frac{n_s(I[k, l])}{N_s} + 5 \frac{n_b(I[k, l])}{N_b}, \quad (7.7)$$

where

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

The re-binning is then conducted using the following procedure:

1. Starting from the first bin  $k_0$  on the right end of the MVA output histogram (MVA output = 1), increase the range of the interval  $I(k_0, l)$  by adding one after the other to the left;
2. Calculate the value of  $Z$  at each step;
3. Once  $Z(I[k_0, l]) > 1$ , rebin all the bins in the interval  $I(k_0, l)$  into a single bin;
4. Repeat steps 1-3, starting this time from the last bin on the right (bin  $l + 1$ , the next bin to the left of bin  $l$ ) until reaches the left end (0) of the histogram.

In addition, binning algorithm requires at least 5 background events in each bin.

All sources of systematic uncertainties described in Section 7.4.5 are considered as nuisance parameters in the profile likelihood. Correlations of NPs across the four regions are included in the fit. All experimental uncertainties, modelling uncertainties from the same source (except for the fake background which was estimated using different method in the  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  channel),  $t\bar{t}$  and  $Z + \text{HF}$  background normalisations and relative acceptance uncertainties on  $t\bar{t}$  and  $Z + \text{HF}$  backgrounds are correlated across the three signal regions and one control region.

Each NP is split into shape and normalisation effects. The shape uncertainties are included as alternative histograms while the normalisation effects are implemented as either flat (for the floating backgrounds) or Gaussian priors. The NPs are then processed to be symmetrised, smoothed and pruned: for one-sided experimental systematics and systematics with up and down variation going to the same side, symmetrisation is applied to avoid under-constraint problems and to improve the stability of fitting; and then, smoothing is applied on systematics with large statistical fluctuations to avoid instabilities; finally, systematics which have less impact are pruned away to speed up the fitting process.

### 7.5.2 $\tau_{\text{lep}}\tau_{\text{had}}$ channel results

The profile likelihood fit is first performed on the two signal regions in the  $\tau_{\text{lep}}\tau_{\text{had}}$  and the  $Z + \text{HF}$  control region. This section describes the fit results obtained with this setting.

#### 7.5.2.1 Post-fit background events yields

The expected signal and background events yields passing the signal region event pre-selection after the  $\tau_{\text{lep}}\tau_{\text{had}}$  only fit are shown in Table 7.21 (7.22) for the SLT (LTT) channel, extracted from the fit result using NN output SM  $HH$  as input.

SampleName	Integral	Error
<b>Signal Samples</b>		
ggF+VBF Non-resonant	6	1
<b>Background Samples</b>		
Fake	33100	1600
$t\bar{t}$	58500	1500
single top	3730	480
Ztthf	1640	140
Zhf	921	83
Zttl	80	22
Zlf	57	19
SM Higgs	153	19
diboson	150	36
W	73	30
Wtt	5	2.2
DYtt	1.7	1.9
DY	12.8	3.7
total Bkg	98450	680
data	98456	

Table 7.21: Post-fit event yields in the SLT signal region for the data, background and signal. Background names conventions are the same as used in Table 7.6.

The determined floating normalisations of the  $t\bar{t}$  and  $Z + \text{HF}$  background are shown in Table 7.23. The normalisation factors depend on which MVA output is used as input for the fit, as a result of different binnings which were optimised for different MVA output.

#### 7.5.2.2 Post-fit MVA input and output distributions

The post fit MVA input variables distributions are shown in Figure 7.35 (7.36) for the SLT (LTT) channel. The post-fit PNN score distributions for the 300, 500, 1000, 1600 GeV resonant mass points and the post-fit SM NN distribution in the SLT and LTT channel are shown in Figure 7.37. As compared to the pre-fit MVA input and output distributions

SampleName	Integral	Error
<b>Signal Samples</b>		
ggF+VBF Non-resonant	1.44	0.24
<b>Background Samples</b>		
Fake	1540	180
$t\bar{t}$	3940	190
single top	203	40
Ztthf	537	65
Zhf	52.6	7.1
Ztlf	42	12
Zlf	2.1	1.1
SM Higgs	24	4.8
diboson	19.9	4.6
W	1.27	0.6
Wtt	2.6	1.1
DYtt	0.27	0.13
DY	0.62	0.28
total Bkg	6359	80
data	6351	

Table 7.22: Post-fit event yields in the LTT signal region for the data, background and signal. Background names conventions are the same as used in Table 7.7.

Mass point	$Z + \text{HF}$ norm. factor	$t\bar{t}$ norm. factor
SM	$1.38 \pm 0.11$	$0.96 \pm 0.03$
300 GeV	$1.38 \pm 0.11$	$0.97 \pm 0.04$
500 GeV	$1.37 \pm 0.11$	$0.97 \pm 0.04$
1000 GeV	$1.39 \pm 0.11$	$0.97 \pm 0.04$
1600 GeV	$1.37 \pm 0.11$	$0.97 \pm 0.04$

Table 7.23: Post-fit normalisation factors of  $t\bar{t}$  and  $Z + \text{HF}$  background obtained from the  $\tau_{\text{lep}}\tau_{\text{had}}$  only fit for SM  $HH$ , 300, 500, 1000 and 1600 GeV resonance fit.

presented in section 7.3, the post-fit distributions show much better MC and data agreement. The uncertainties are also more constrained with the help of the  $Z + \text{HF}$  control region.

To validate the background modelling, the post-fit NN/PNNs distributions are checked in the 1- $b$ -tag where the same signal region selection is applied except for requiring exactly one  $b$ -jet. The background distributions are plotted against the data, as shown in Figure B.34 in Appendix B.4. In this validation region, the data is well modelled by the MC simulation within uncertainties.

The non-resonant  $HH$  signal with  $\kappa_\lambda$  values are passed to the same NN classification used for the SM  $HH$  signal. The NN score of non-resonant  $HH$  with  $\kappa_\lambda = -5, -3, -1.6, 1, 2, 7$  and 10 samples are shown in Figure 7.38 (7.39) for the SLT (LTT) channel. As the NN has only been trained with SM signal, it is expected that the separation power will degrade for samples with  $\kappa_\lambda$  deviating far from 1.

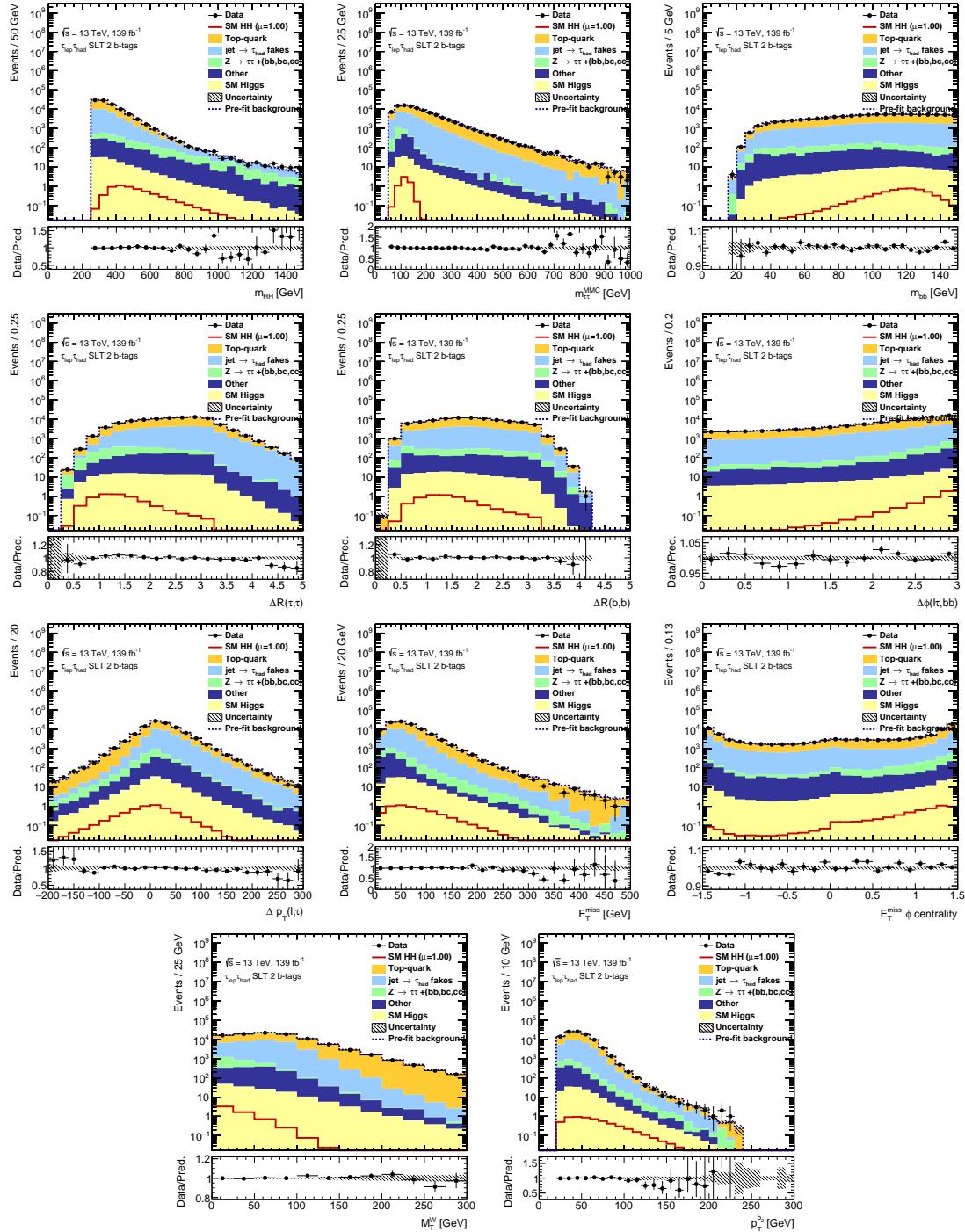


Figure 7.35: Post-fit PNN/NN input variable distributions in the SLT signal region.

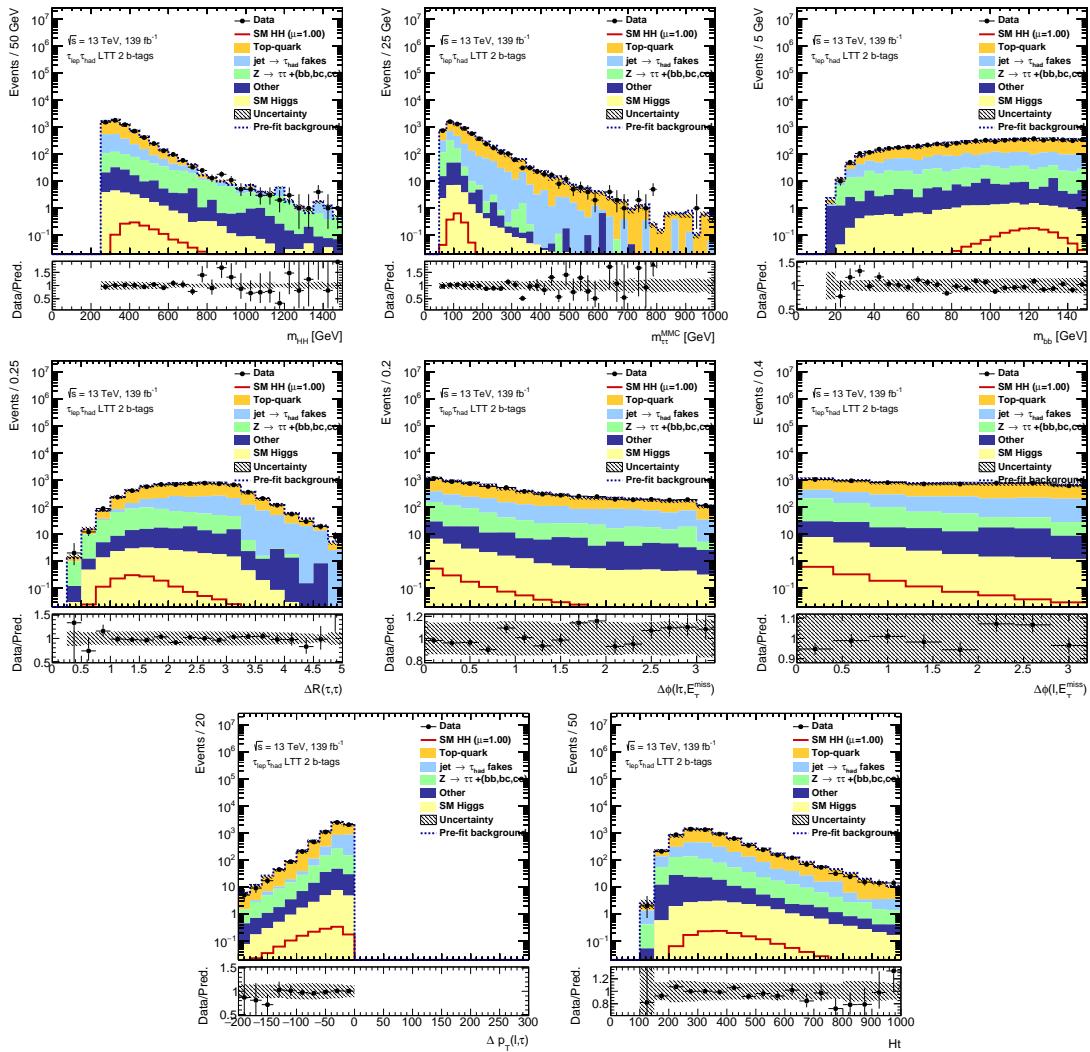


Figure 7.36: Post-fit PNN/NN input variable distributions in the LTT signal region.

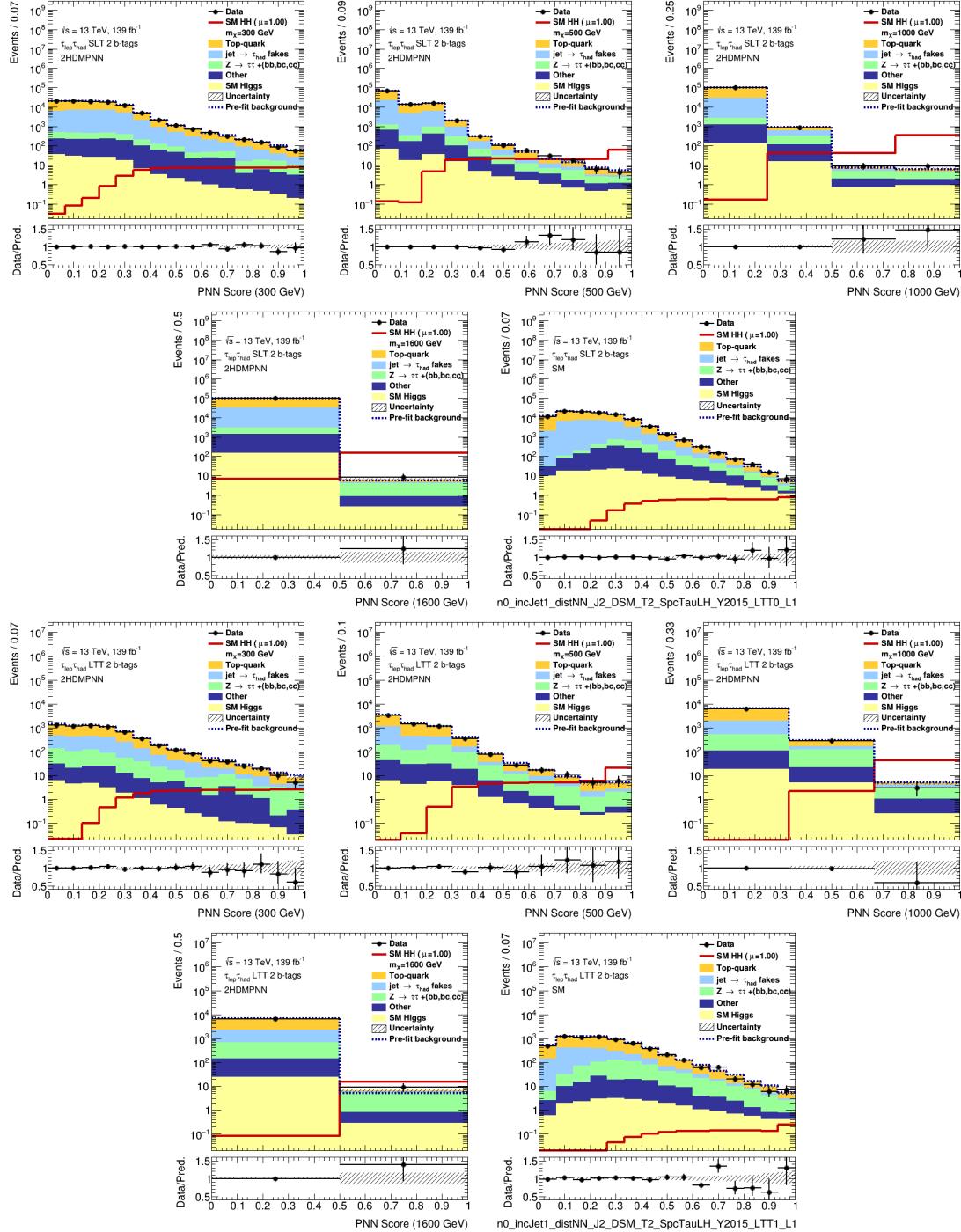


Figure 7.37: Post-fit PNN score distributions for the 300, 500, 1000, 1600 GeV mass points and non-resonant NN in the di-Higgs  $bb\tau_{lep}\tau_{had}$  SLT (top two rows) and LTT (bottom two rows) signal regions.

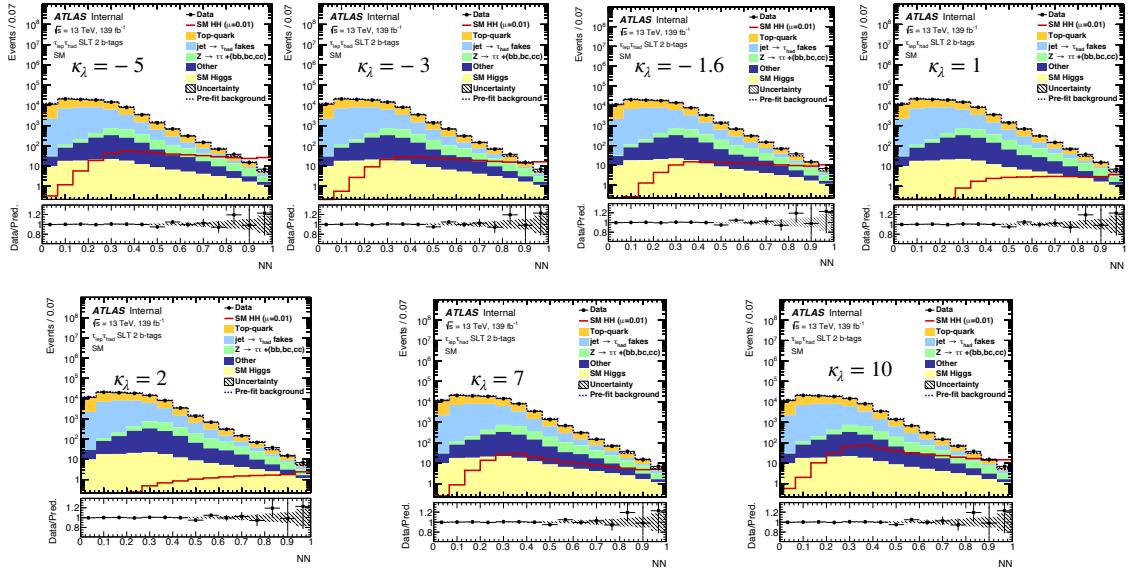


Figure 7.38: Post-fit plots of the neural network output in the  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT signal region, showing the agreement of the background-only hypothesis with the data as well as the signal prediction for different  $\kappa_\lambda$  values.

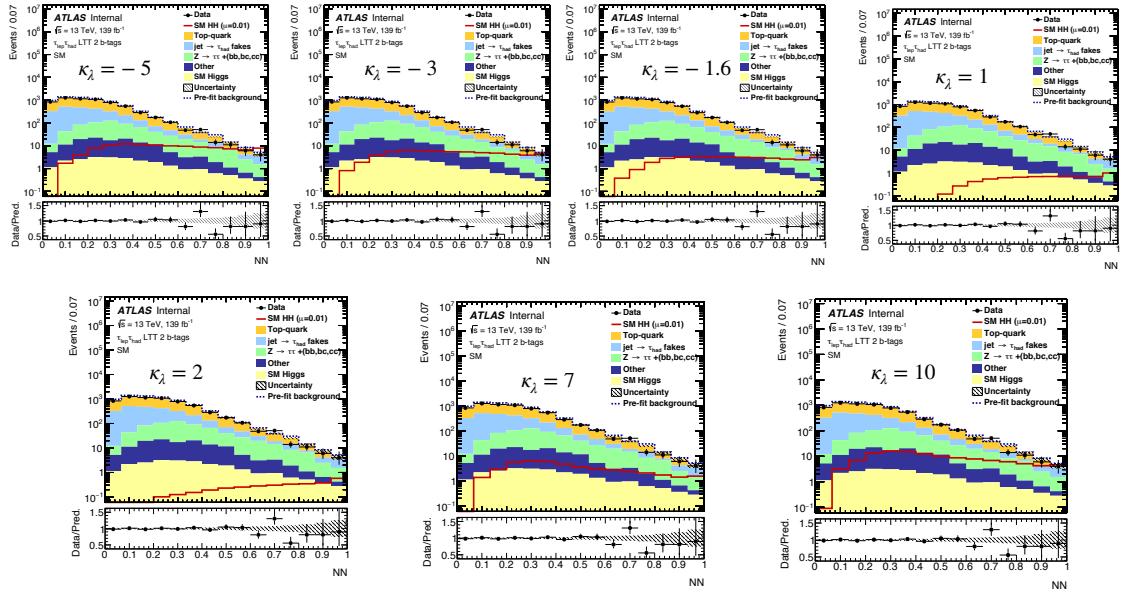


Figure 7.39: Post-fit plots of the neural network output in the  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT signal region, showing the agreement of the background-only hypothesis with the data as well as the signal prediction for different  $\kappa_\lambda$  values.

### 7.5.2.3 Limit results: $\tau_{\text{lep}}\tau_{\text{had}}$ only fit

Table 7.24 shows the limit results obtained using SLT only fit, LTT only fit, and SLT, LTT combined fit ( $\tau_{\text{lep}}\tau_{\text{had}}$  only fit).

	Observed	$-2\sigma$	$-1\sigma$	Median	$+1\sigma$	$+2\sigma$
<b>SLT only fit</b>						
$\sigma_{\text{ggF+VBF}} [\text{fb}]$	316.93	132.24	177.54	246.39	342.90	459.69
$\sigma_{\text{ggF+VBF}}$ StatOnly [fb]	222.82	102.04	136.99	190.11	264.58	354.69
$\mu_{\text{ggF+VBF}}$	10.91	4.48	6.01	8.34	11.60	15.56
$\mu_{\text{ggF+VBF}}$ StatOnly	6.80	3.11	4.18	5.80	8.07	10.82
<b>LTT only fit</b>						
$\sigma_{\text{ggF+VBF}} [\text{fb}]$	498.70	386.62	519.04	720.34	1002.50	1343.92
$\sigma_{\text{ggF+VBF}}$ StatOnly [fb]	322.27	339.61	455.93	632.75	880.60	1180.51
$\mu_{\text{ggF+VBF}}$	16.50	13.08	17.56	24.36	33.91	45.46
$\mu_{\text{ggF+VBF}}$ StatOnly	9.83	10.36	13.91	19.31	26.87	36.02
<b>SLT + LTT combined fit</b>						
$\sigma_{\text{ggF+VBF}} [\text{fb}]$	267.21	124.49	167.13	231.94	322.80	432.73
$\sigma_{\text{ggF+VBF}}$ StatOnly [fb]	169.05	96.33	129.33	179.48	249.79	334.86
$\mu_{\text{ggF+VBF}}$	9.16	4.22	5.66	7.86	10.93	14.66
$\mu_{\text{ggF+VBF}}$ StatOnly	5.16	2.94	3.95	5.48	7.62	10.22

Table 7.24: 95% C.L. expected and observed limit on the non-resonant HH production cross-section assuming SM kinematics using SLT only, LTT only and  $\tau_{\text{lep}}\tau_{\text{had}}$  channel fit. The signal strength is with respect to the SM cross-section, i.e.  $\mu_{\text{ggF+VBF}} = \sigma_{\text{ggF+VBF}}/\sigma_{\text{ggF+VBF}}^{\text{SM}}$ . The ‘StatOnly’ indicates that the limits are calculated with only statistical uncertainty, and if not specified, the limits are obtained with full systematics.

Figures 7.40 show the 95% C.L. expected limits on the resonant HH production cross-section as a function of the resonance mass from the  $\tau_{\text{had}}\tau_{\text{had}}$  and  $\tau_{\text{lep}}\tau_{\text{had}}$  channels, respectively. No significant excess is observed in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel (largest local significance at 1.1 TeV corresponding to  $1.6\sigma$ ).

Upper limits at 95% C.L. are set on the non-resonant  $HH$  production cross-section as a function of  $\kappa_\lambda$ , in the range of -10 to 10. The  $\kappa_\lambda$  is excluded at 95% C.L. outside the range of [-3.8, 12.2] ([-3.8, 11.9]) for observed (expected) limit. The results are shown in Figure 7.41.

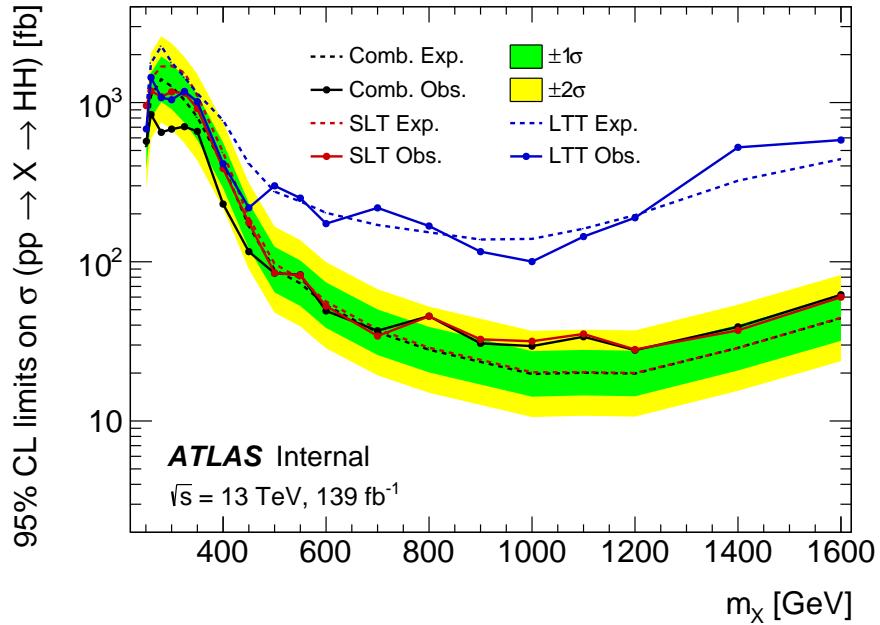


Figure 7.40: 95% C.L. expected and observed limits on the resonant  $HH$  production cross-section as a function of the resonance mass in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel. The black lines shows the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel combined limit whereas the red and blue lines show the limit for SLT and LTT channels separately. The solid and dashed lines represent the observed and expected limits respectively.

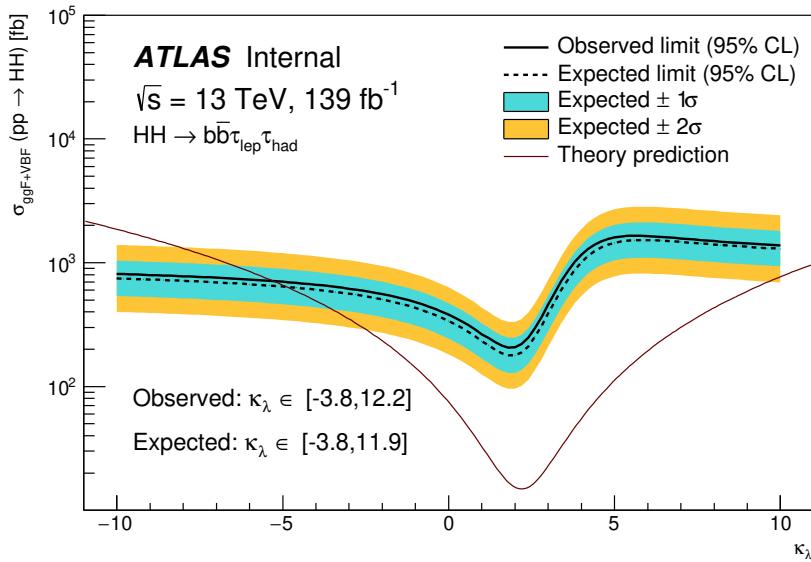


Figure 7.41: Observed and expected 95% CL exclusion limits on non-resonant  $HH$  production as a function of  $\kappa_\lambda$  in the using  $\tau_{\text{lep}}\tau_{\text{had}}$  only fit. The observed (expected) limit on  $\kappa_\lambda$  is [-3.8, 12.2] ([-3.8, 11.9]).

### 7.5.3 Combination results of $\tau_{\text{lep}}\tau_{\text{had}}$ and $\tau_{\text{had}}\tau_{\text{had}}$

A combined fit is then performed on all three signal regions and the  $Z + \text{HF}$  control region. The combined results are shown in this section.

Table 7.25 shows 95% C.L. expected limit on the non-resonant HH production cross-section assuming SM kinematics from the combined fit.

	Observed	$-2\sigma$	$-1\sigma$	Median	$+1\sigma$	$+2\sigma$
$\sigma_{\text{ggF+VBF}} [\text{fb}]$	135.27	61.35	82.36	114.31	159.08	213.26
$\sigma_{\text{ggF+VBF}} \text{ StatOnly} [\text{fb}]$	85.58	48.45	65.04	90.27	125.63	168.41
$\mu_{\text{ggF+VBF}}$	4.65	2.08	2.79	3.87	5.39	7.22
$\mu_{\text{ggF+VBF}} \text{ StatOnly}$	2.61	1.48	1.98	2.75	3.83	5.14

Table 7.25: 95% C.L. expected and observed limit on the non-resonant HH production cross-section assuming SM kinematics using  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  combined fit. The signal strength is with respect to the SM cross-section, i.e.  $\mu_{\text{ggF+VBF}} = \sigma_{\text{ggF+VBF}}/\sigma_{\text{ggF+VBF}}^{\text{SM}}$ . The ‘StatOnly’ indicates the limits are calculated with only statistical uncertainty, and if not specified, the limits are obtained with full systematics.

The 95% C.L. expected and the observed limits on the resonant  $HH$  production cross-section, as a function of the resonance mass from the  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  combined fit are shown in Figure 7.42. Observed (expected) upper limits range from 26 – 950 fb (12 – 850 fb), depending on the mass of the resonance. In the combined fit, excess (observed limit deviates greater than  $2\sigma$  from the expected limit) is present in the mass range between 800 and 1100 GeV. The maximum excess is found at the 1000 GeV which has a deviation of  $3.1\sigma$ . The local p-values as a function of resonance mass are shown in Figure 7.43. The look-elsewhere effect is accounted by calculating the global significance following Ref.[152], using the up-crossing method:

$$p_{\text{global}} = p_{\text{local}} + N_{\text{up}} e^{-1/2(Z_{\text{local}}^2 - Z_{\text{ref}}^2)},$$

where the  $p_{\text{local}}$  and  $p_{\text{global}}$  are the local and global p-values, and  $Z_{\text{ref}}$  is chosen for  $p=0.5$  (corresponding to  $0\sigma$  significance level), and  $N_{\text{up}}$  is the number of times the local p-value curve crossing the reference line (in this case,  $p=0.5$ ) in the upwawrd direction. The estimated global significance is  $2.1^{+0.4}_{-0.2}\sigma$ .

The 95% C.L. expected and the observed upper limits on the cross-section of non-resonant  $HH$  production as a function of  $\kappa_\lambda$  are shown in Figure 7.44. The 95% C.L. observed (expected) exclusion limit of  $\kappa_\lambda$  is  $-2.4 < \kappa_\lambda < 9.2$  ( $-2.0 < \kappa_\lambda < 9.0$ ).

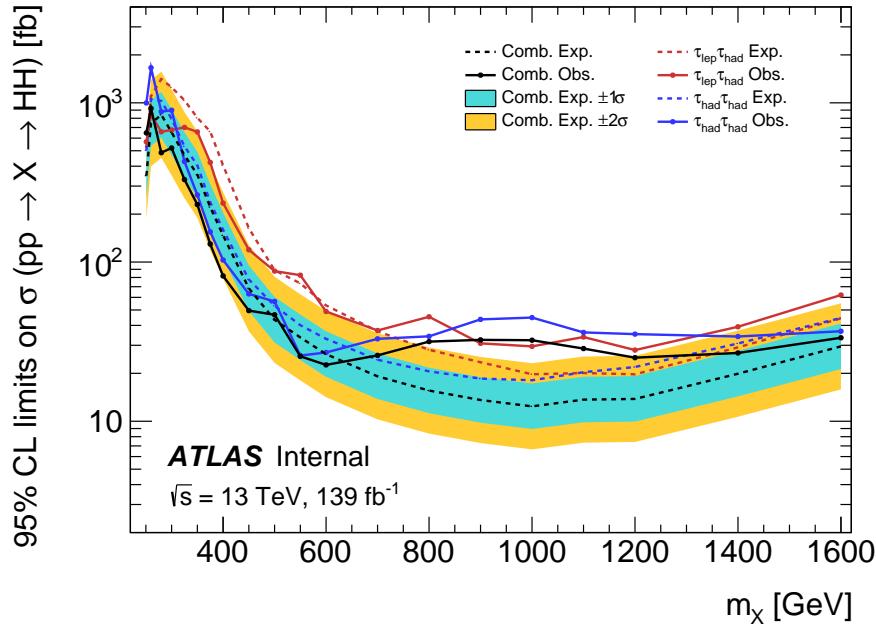


Figure 7.42: 95% C.L. expected and observed limits on the resonant HH production cross-section as a function of the resonance mass. The dashed lines show the expected limits, while the solid lines show the observed limits. Black, red and blue represents the upper limit from combination,  $\tau_{\text{lep}}\tau_{\text{had}}$  channel and  $\tau_{\text{had}}\tau_{\text{had}}$  channel, respectively.

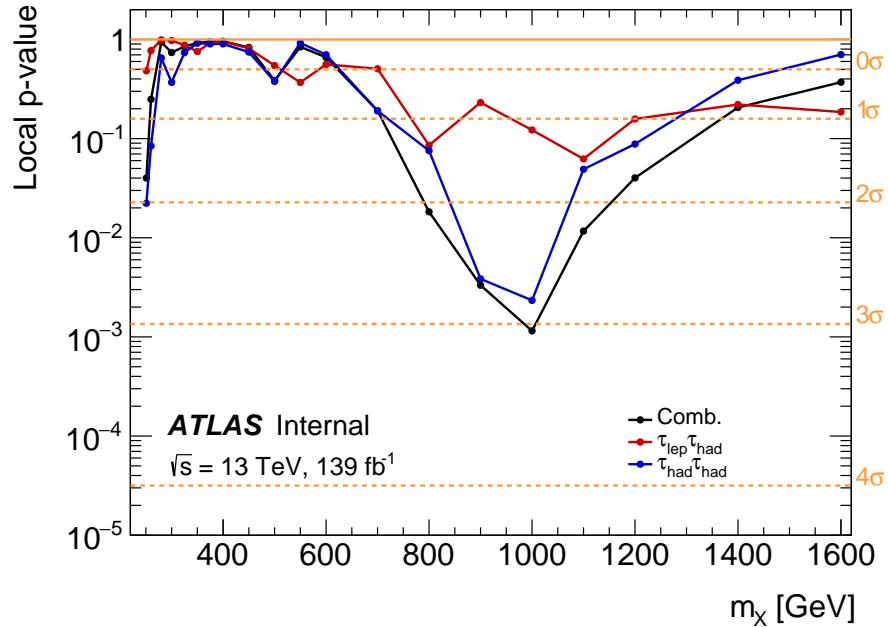


Figure 7.43: Local p-value of the background only hypothesis test as a function of the mass of the resonance. The black line represents the p-value from combination, while red and blue line are those for the  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  channel, respectively. The horizontal dashed orange lines show the corresponding (2-side) significance level.

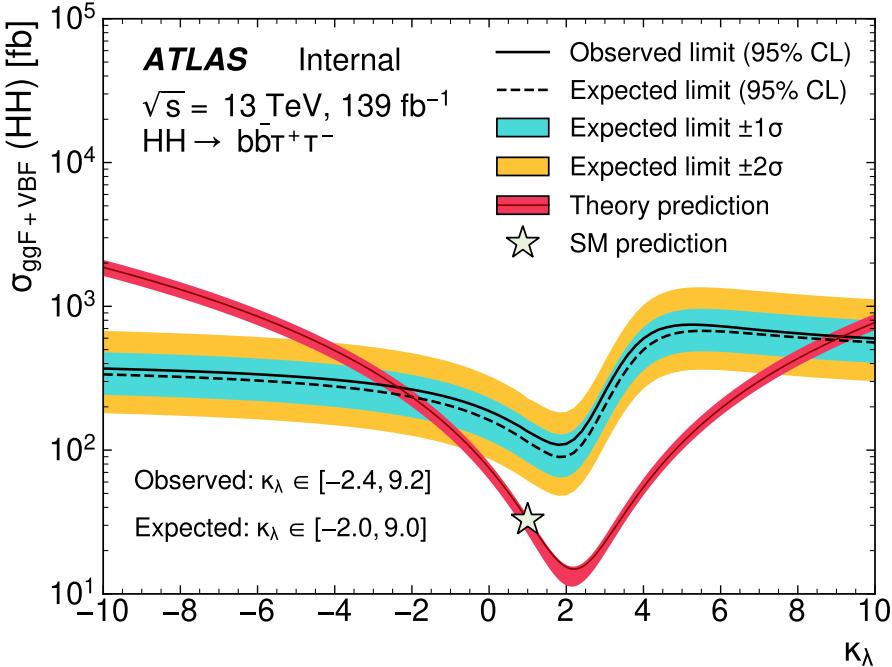


Figure 7.44: Observed and expected 95% CL upper limits on non-resonant  $\sigma(pp \rightarrow HH)$  cross-section as a function of  $\kappa_\lambda$  using  $\tau_{\text{lep}}\tau_{\text{had}}$  and  $\tau_{\text{had}}\tau_{\text{had}}$  combined fit. The observed (expected) exclusion limit at 95% C.L. on  $\kappa_\lambda$  in the  $b\bar{b}\tau^+\tau^-$  channel is: [-2.4, 9.2] ([-2.0, 9.0]).

#### 7.5.4 Combination results with other $HH$ decay channels

To maximise the sensitivity in the  $HH$  searches, the  $b\bar{b}\tau^+\tau^-$  results are combined with other  $HH$  decay channels results. The  $b\bar{b}\tau^+\tau^-$  non-resonant  $HH$  result is combined to the  $b\bar{b}\gamma\gamma$  channel, which also has high sensitivity due to the high energy resolution in di-photon systems. For the resonant searches, the  $b\bar{b}\tau^+\tau^-$  results are combined with  $b\bar{b}\gamma\gamma$  and  $b\bar{b}b\bar{b}$  channels. The  $b\bar{b}\gamma\gamma$  channel is expected to have high sensitivity in low energy regime and the  $b\bar{b}b\bar{b}$  channel dominates over the high-energy regime, which benefits from high selection efficiency in the more boosted events topology.

##### 7.5.4.1 Combination results of $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$

The  $b\bar{b}\tau^+\tau^-$  results are combined with the results in  $b\bar{b}\gamma\gamma$  channel in terms of upper limits on the non-resonant  $HH$  production signal strength  $\mu$ , and on cross-section as a function of  $\kappa_\lambda$  [137]. The upper limits on signal strength are shown in Table 7.26, and the limits on the cross-section are shown in Figure 7.45. This combination result sets the current world-best observed upper limit on the SM  $HH$  cross-section [153–156], 3.1 times SM cross-section, at 95% C.L.

	Obs.	$-2.0\sigma$	$-1.0\sigma$	Exp.	$1.0\sigma$	$2.0\sigma$
$b\bar{b}\gamma\gamma$	4.3	3.1	4.1	5.7	8.8	14.3
$b\bar{b}\tau^+\tau^-$	4.6	2.1	2.8	3.9	5.9	9.4
Combined	3.1	1.7	2.2	3.1	4.7	7.3

Table 7.26: Observed and expected 95 % C.L. upper limits on the signal strength for SM  $HH$  production derived from the  $b\bar{b}\tau^+\tau^-$  and  $b\bar{b}\gamma\gamma$  searches, and their statistical combination.

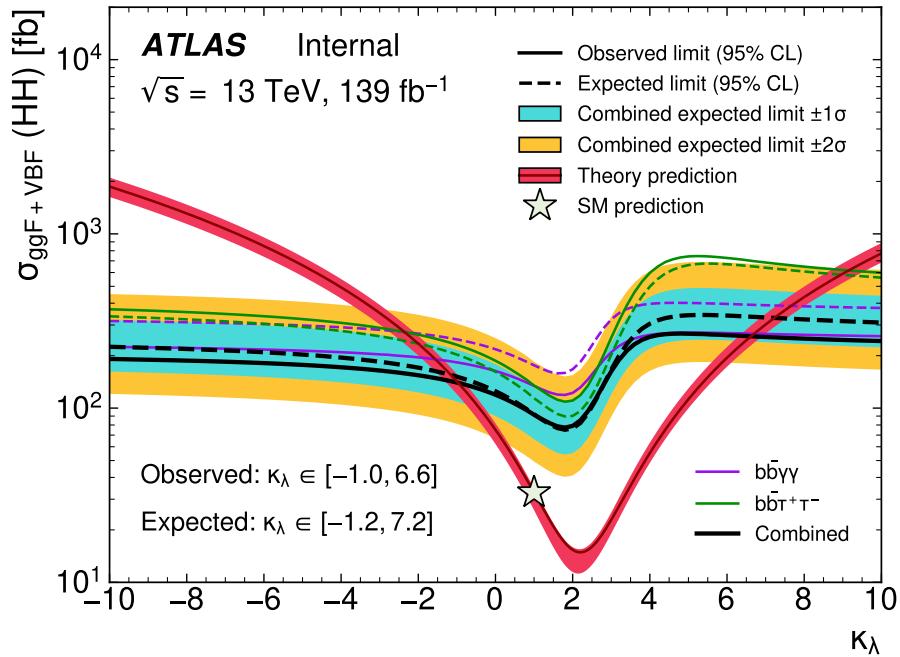


Figure 7.45: Observed and expected 95% CL exclusion limits on non-resonant  $\sigma(pp \rightarrow HH)$  as a function of  $\kappa_\lambda$  for the individual channels and the combination of  $b\bar{b}\tau^+\tau^-$  and  $b\bar{b}\gamma\gamma$ . The observed (expected) combined limit on  $\kappa_\lambda$  is:  $[-1.0, 6.6]$  ( $[-1.2, 7.2]$ ).

### 7.5.4.2 Combination results of $b\bar{b}\tau^+\tau^-$ , $b\bar{b}\gamma\gamma$ and $b\bar{b}b\bar{b}$

Observed and expected limits at 95% C.L. are set on the resonant  $HH$  production cross-section,  $\sigma(X \rightarrow HH)$ , for  $b\bar{b}\gamma\gamma$ ,  $b\bar{b}\tau^+\tau^-$ , and  $b\bar{b}b\bar{b}$  searches, and their statistical combination [137], as shown in Figure 7.46. The observed (expected) combined limits on cross-section range from 1.1 to 595 fb (1.2 to 392 fb) depending on the resonance mass. As mentioned before, the  $b\bar{b}\gamma\gamma$  results shows largest sensitive low  $m_X$ , and the  $b\bar{b}b\bar{b}$  limits dominates in high  $m_X$ . The  $b\bar{b}\tau^+\tau^-$  search is the most sensitive in the intermediate range, around 400–800 GeV. These three channels are therefore complementary to each other.

The largest excess is observed at 1.1 TeV, corresponding to a local significance of  $3.2\sigma$ . While the similar excess is observed in the  $b\bar{b}\tau^+\tau^-$  channel, the compatibility is checked with the  $b\bar{b}b\bar{b}$  excess. At 1.1 TeV, the level of compatibility of the fitted signal strength in the two channels corresponds to a p-value of 0.34 ( $>>0.05$ ), and hence the excess is compatible.

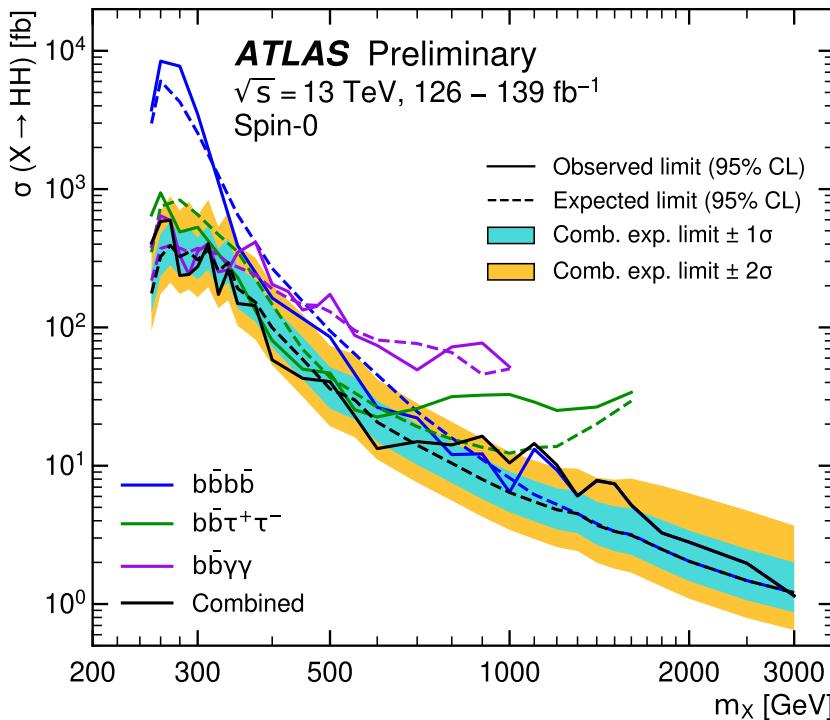


Figure 7.46: Expected and observed 95% C.L. upper limits on  $\sigma(X \rightarrow HH)$  for a spin-0 resonance as a function of its mass  $m_X$  in the  $b\bar{b}\gamma\gamma$ ,  $b\bar{b}\tau^+\tau^-$  and  $b\bar{b}b\bar{b}$  searches, and their statistical combination.

## **Chapter 8**

### **Summary**

# Bibliography

- [1] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001. doi: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [2] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003. doi: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [3] S. Chatrchyan et al. “The CMS Experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. doi: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [4] K. Aamodt et al. “The ALICE experiment at the CERN LHC”. In: *JINST* 3 (2008), S08002. doi: [10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [5] A. Augusto Alves Jr. et al. “The LHCb Detector at the LHC”. In: *JINST* 3 (2008), S08005. doi: [10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [6] Stephen Myers. “The LEP collider, from design to approval and commissioning”. In: (Oct. 1991). doi: [10.5170/CERN-1991-008](https://doi.org/10.5170/CERN-1991-008).
- [7] L Rossi. “The LHC superconducting magnets”. In: 1 (2003), pp. 141–145.
- [8] J. P. Blewett. “200-GeV Intersecting Storage Accelerators”. In: *eConf* C710920 (1971), p. 501.
- [9] Ph Lebrun et al. “Report of the Task Force on the Incident of 19 September 2008 at the LHC”. In: *CERN LHC Project Rep* 1168 (2009).
- [10] R Alemany-Fernandez et al. “Operation and Configuration of the LHC in Run 1”. In: (2013).
- [11] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), p. 1. doi: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- [12] ATLAS Collaboration. “A Particle Consistent with the Higgs Boson Observed with the ATLAS Detector at the Large Hadron Collider”. In: *Science* 338 (2012), p. 1576. doi: [10.1126/science.1232005](https://doi.org/10.1126/science.1232005).
- [13] F Bordry et al. “The first long shutdown (LS1) for the LHC”. In: (2013).
- [14] Jorg Wenninger. “Operation and Configuration of the LHC in Run 2”. In: (Mar. 2019). URL: <https://cds.cern.ch/record/2668326>.

- [15] Georges Aad et al. “ATLAS data quality operations and performance for 2015–2018 data-taking”. In: *JINST* 15.04 (2020), P04003.
- [16] ATLAS Collaboration. *ATLAS Pixel Detector: Technical Design Report*. ATLAS-TDR-11; CERN-LHCC-98-013. 1998. URL: <https://cds.cern.ch/record/381263>.
- [17] ATLAS Collaboration. *ATLAS Inner Detector: Technical Design Report, Volume 1*. ATLAS-TDR-4; CERN-LHCC-97-016. 1997. URL: <https://cds.cern.ch/record/331063>.
- [18] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter: Technical Design Report*. ATLAS-TDR-2; CERN-LHCC-96-041. 1996. URL: <https://cds.cern.ch/record/331061>.
- [19] ATLAS Collaboration. *ATLAS Tile Calorimeter: Technical Design Report*. ATLAS-TDR-3; CERN-LHCC-96-042. 1996. URL: <https://cds.cern.ch/record/331062>.
- [20] ATLAS Collaboration. *ATLAS Muon Spectrometer: Technical Design Report*. ATLAS-TDR-10; CERN-LHCC-97-022. CERN, 1997. URL: <https://cds.cern.ch/record/331068>.
- [21] ATLAS Collaboration. *ATLAS Magnet System: Magnet Project Technical Design Report, Volume 1*. ATLAS-TDR-6; CERN-LHCC-97-018. 1997. URL: <https://cds.cern.ch/record/338080>.
- [22] ATLAS Collaboration. *ATLAS Insertable B-Layer: Technical Design Report*. ATLAS-TDR-19; CERN-LHCC-2010-013. 2010. URL: <https://cds.cern.ch/record/1291633>.
- [23] J. Pequenao. *Computer generated image of the ATLAS inner detector*. 2008. URL: <https://cds.cern.ch/record/1095926>.
- [24] ATLAS Collaboration. *ATLAS Calorimeter Performance: Technical Design Report*. ATLAS-TDR-1; CERN-LHCC-96-040. 1996. URL: <https://cds.cern.ch/record/331059>.
- [25] ATLAS Collaboration. *ATLAS Level-1 Trigger: Technical Design Report*. ATLAS-TDR-12; CERN-LHCC-98-014. 1998. URL: <https://cds.cern.ch/record/381429>.
- [26] ATLAS Collaboration. *ATLAS High-Level Trigger, Data Acquisition and Controls: Technical Design Report*. ATLAS-TDR-16; CERN-LHCC-2003-022. 2003. URL: <https://cds.cern.ch/record/616089>.
- [27] Robert Blair et al. “The ATLAS High Level Trigger Region of Interest Builder”. In: *JINST* 3 (2008), P04001. doi: [10.1088/1748-0221/3/04/P04001](https://doi.org/10.1088/1748-0221/3/04/P04001). arXiv: [0711.3217](https://arxiv.org/abs/0711.3217) [physics.ins-det].
- [28] ATLAS Collaboration. “Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in High Pile-Up LHC Environment”. In: (2012). URL: <https://cds.cern.ch/record/1435196>.

- [29] ATLAS Collaboration. “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2”. In: *Eur. Phys. J. C* 77 (2017), p. 673. doi: [10.1140/epjc/s10052-017-5225-7](https://doi.org/10.1140/epjc/s10052-017-5225-7). arXiv: [1704.07983 \[hep-ex\]](https://arxiv.org/abs/1704.07983).
- [30] Azriel Rosenfeld and John L. Pfaltz. “Sequential Operations in Digital Picture Processing”. In: *J. ACM* 13.4 (Oct. 1966), pp. 471–494. ISSN: 0004-5411. doi: [10.1145/321356.321357](https://doi.org/10.1145/321356.321357). URL: <https://doi.org/10.1145/321356.321357>.
- [31] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. doi: [https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4). URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874>.
- [32] ATLAS Collaboration. “Performance of the ATLAS Silicon Pattern Recognition Algorithm in Data and Simulation at  $\sqrt{s} = 7$  TeV”. In: (2010). URL: <https://cds.cern.ch/record/1281363>.
- [33] ATLAS Collaboration. “Performance of primary vertex reconstruction in proton–proton collisions at  $\sqrt{s} = 7$  TeV in the ATLAS experiment”. In: (2010). URL: <https://cds.cern.ch/record/1281344>.
- [34] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at  $\sqrt{s} = 13$  TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 639. doi: [10.1140/epjc/s10052-019-7140-6](https://doi.org/10.1140/epjc/s10052-019-7140-6). arXiv: [1902.04655 \[hep-ex\]](https://arxiv.org/abs/1902.04655).
- [35] W Lampl et al. *Calorimeter Clustering Algorithms: Description and Performance*. Tech. rep. Geneva: CERN, Apr. 2008. URL: <https://cds.cern.ch/record/1099735>.
- [36] ATLAS Collaboration. “Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung”. In: (2012). URL: <https://cds.cern.ch/record/1449796>.
- [37] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data”. In: *JINST* 14 (2019), P12006. doi: [10.1088/1748-0221/14/12/P12006](https://doi.org/10.1088/1748-0221/14/12/P12006). arXiv: [1908.00005 \[hep-ex\]](https://arxiv.org/abs/1908.00005).
- [38] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2  $pp$  collision data set at  $\sqrt{s} = 13$  TeV”. In: *Eur. Phys. J. C* 81 (2020), p. 578. doi: [10.1140/epjc/s10052-021-09233-2](https://doi.org/10.1140/epjc/s10052-021-09233-2). arXiv: [2012.00578 \[hep-ex\]](https://arxiv.org/abs/2012.00578).
- [39] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton–proton collision data at  $\sqrt{s} = 13$  TeV”. In: *Eur. Phys. J. C* 76 (2016), p. 292. doi: [10.1140/epjc/s10052-016-4120-y](https://doi.org/10.1140/epjc/s10052-016-4120-y). arXiv: [1603.05598 \[hep-ex\]](https://arxiv.org/abs/1603.05598).

- [40] ATLAS Collaboration. “Recommended isolation working points”. In: (). URL: [https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/RecommendedIsolationWPs#Muon\\_working\\_points](https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/RecommendedIsolationWPs#Muon_working_points).
- [41] Ryan Atkin. “Review of jet reconstruction algorithms”. In: *Journal of Physics: Conference Series*. Vol. 645. 1. IOP Publishing. 2015, p. 012008.
- [42] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *The European Physical Journal C* 77.7 (July 2017). ISSN: 1434-6052. doi: [10.1140/epjc/s10052-017-5004-5](https://doi.org/10.1140/epjc/s10052-017-5004-5). URL: <http://dx.doi.org/10.1140/epjc/s10052-017-5004-5>.
- [43] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *Eur. Phys. J. C* 77 (2017), p. 466. doi: [10.1140/epjc/s10052-017-5031-2](https://doi.org/10.1140/epjc/s10052-017-5031-2). arXiv: [1703.10485 \[hep-ex\]](https://arxiv.org/abs/1703.10485).
- [44] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The anti- $k_t$  jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. doi: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). URL: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [45] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *Eur. Phys. J. C* 77 (2017), p. 490. doi: [10.1140/epjc/s10052-017-5004-5](https://doi.org/10.1140/epjc/s10052-017-5004-5). arXiv: [1603.02934 \[hep-ex\]](https://arxiv.org/abs/1603.02934).
- [46] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *Phys. Rev. D* 96 (2017), p. 072002. doi: [10.1103/PhysRevD.96.072002](https://doi.org/10.1103/PhysRevD.96.072002). arXiv: [1703.09665 \[hep-ex\]](https://arxiv.org/abs/1703.09665).
- [47] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: (2020). arXiv: [2007.02645 \[hep-ex\]](https://arxiv.org/abs/2007.02645).
- [48] ATLAS Collaboration. “Tagging and suppression of pileup jets with the ATLAS detector”. In: (2014). URL: <https://cds.cern.ch/record/1700870>.
- [49] ATLAS Collaboration. “Summary of the ATLAS experiment’s sensitivity to supersymmetry after LHC Run 1 — interpreted in the phenomenological MSSM”. In: *JHEP* 10 (2015), p. 134. doi: [10.1007/JHEP10\(2015\)134](https://doi.org/10.1007/JHEP10(2015)134). arXiv: [1508.06608 \[hep-ex\]](https://arxiv.org/abs/1508.06608).
- [50] ATLAS Collaboration. “Combination of searches for Higgs boson pairs in  $pp$  collisions at 13 TeV with the ATLAS experiment”. In: (2018). URL: <https://cds.cern.ch/record/2638212>.
- [51] Dean Carmi et al. “Interpreting LHC Higgs results from natural new physics perspective”. In: *Journal of High Energy Physics* 2012.7 (July 2012). ISSN: 1029-8479. doi: [10.1007/jhep07\(2012\)136](https://doi.org/10.1007/jhep07(2012)136). URL: [http://dx.doi.org/10.1007/JHEP07\(2012\)136](http://dx.doi.org/10.1007/JHEP07(2012)136).

- [52] “Performance of  $b$ -jet identification in the ATLAS experiment”. In: *Journal of Instrumentation* 11.04 (Apr. 2016), P04008–P04008. ISSN: 1748-0221. doi: [10.1088/1748-0221/11/04/p04008](https://doi.org/10.1088/1748-0221/11/04/p04008). URL: <http://dx.doi.org/10.1088/1748-0221/11/04/P04008>.
- [53] ATLAS Collaboration. “Optimisation and performance studies of the ATLAS  $b$ -tagging algorithms for the 2017-18 LHC run”. In: (2017). URL: <https://cds.cern.ch/record/2273281>.
- [54] ATLAS Collaboration. “Secondary vertex finding for jet flavour identification with the ATLAS detector”. In: (2017). URL: <https://cds.cern.ch/record/2270366>.
- [55] ATLAS Collaboration. “Topological  $b$ -hadron decay reconstruction and identification of  $b$ -jets with the JetFitter package in the ATLAS experiment at the LHC”. In: (2018). URL: <https://cds.cern.ch/record/2645405>.
- [56] ATLAS Collaboration. “ATLAS  $b$ -jet identification performance and efficiency measurement with  $t\bar{t}$  events in  $pp$  collisions at  $\sqrt{s} = 13$  TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 970. doi: [10.1140/epjc/s10052-019-7450-8](https://doi.org/10.1140/epjc/s10052-019-7450-8). arXiv: [1907.05120 \[hep-ex\]](https://arxiv.org/abs/1907.05120).
- [57] Luca Scodellaro. “ $b$  tagging in ATLAS and CMS”. In: (2017). arXiv: [1709.01290 \[hep-ex\]](https://arxiv.org/abs/1709.01290). URL: <https://arxiv.org/abs/1709.01290>.
- [58] “Optimisation of the ATLAS  $b$ -tagging performance for the 2016 LHC Run”. In: ATL-PHYS-PUB-2016-012 (June 2016). URL: <https://cds.cern.ch/record/2160731>.
- [59] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 Aug. 2018), p. 030001. doi: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001). URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [60] *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*. Tech. rep. Geneva: CERN, Aug. 2019. URL: <https://cds.cern.ch/record/2688062>.
- [61] ATLAS Collaboration. “A search for resonant and non-resonant Higgs boson pair production in the  $b\bar{b}\tau^+\tau^-$  decay channel in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *Phys. Rev. Lett.* 121 (2018), p. 191801. doi: [10.1103/PhysRevLett.121.191801](https://doi.org/10.1103/PhysRevLett.121.191801). arXiv: [1808.00336 \[hep-ex\]](https://arxiv.org/abs/1808.00336). Erratum: in: *Phys. Rev. Lett.* 122 (2019), p. 089901. doi: [10.1103/PhysRevLett.122.089901](https://doi.org/10.1103/PhysRevLett.122.089901).
- [62] ATLAS Collaboration. “Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC”. In: (2015). URL: <https://cds.cern.ch/record/2064383>.
- [63] ATLAS Collaboration. “Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using  $pp$  collisions at  $\sqrt{s} = 13$  TeV”. In: (2017). URL: <https://cds.cern.ch/record/2261772>.

- [64] M. *et. al.* Aaboud. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at  $\sqrt{s} = 13$  TeV”. In: *The European Physical Journal C* 78.11 (Nov. 2018). issn: 1434-6052. doi: [10.1140/epjc/s10052-018-6288-9](https://doi.org/10.1140/epjc/s10052-018-6288-9). URL: <http://dx.doi.org/10.1140/epjc/s10052-018-6288-9>.
- [65] ATLAS Collaboration. “ATLAS Overlap Removal Tool: AssociationUtils”. In: (). URL: <https://gitlab.cern.ch/atlas/athena/tree/21.2/PhysicsAnalysis/AnalysisCommon/AssociationUtils/>.
- [66] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. issn: 0168-9002. doi: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [67] ATLAS Collaboration. “The ATLAS Simulation Infrastructure”. In: *Eur. Phys. J. C* 70 (2010), p. 823. doi: [10.1140/epjc/s10052-010-1429-9](https://doi.org/10.1140/epjc/s10052-010-1429-9). arXiv: [1005.4568 \[physics.ins-det\]](https://arxiv.org/abs/1005.4568).
- [68] ATLAS Collaboration. “Performance of electron and photon triggers in ATLAS during LHC Run 2”. In: *Eur. Phys. J. C* 80 (2020), p. 47. doi: [10.1140/epjc/s10052-019-7500-2](https://doi.org/10.1140/epjc/s10052-019-7500-2). arXiv: [1909.00761 \[hep-ex\]](https://arxiv.org/abs/1909.00761).
- [69] ATLAS Collaboration. “Performance of the ATLAS muon triggers in Run 2”. In: *JINST* 15.09 (2020), P09015. doi: [10.1088/1748-0221/15/09/p09015](https://doi.org/10.1088/1748-0221/15/09/p09015). arXiv: [2004.13447 \[hep-ex\]](https://arxiv.org/abs/2004.13447).
- [70] ATLAS Collaboration. “ATLAS data quality operations and performance for 2015–2018 data-taking”. In: *JINST* 15 (2020), P04003. doi: [10.1088/1748-0221/15/04/P04003](https://doi.org/10.1088/1748-0221/15/04/P04003). arXiv: [1911.04632 \[physics.ins-det\]](https://arxiv.org/abs/1911.04632).
- [71] ATLAS Collaboration. “Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC”. In: (2019). URL: <https://cds.cern.ch/record/2677054>.
- [72] G. Avoni et al. “The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS”. In: *JINST* 13.07 (2018), P07017. doi: [10.1088/1748-0221/13/07/P07017](https://doi.org/10.1088/1748-0221/13/07/P07017).
- [73] D. J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152. doi: [10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4).
- [74] Enrico Bothmann et al. “Event generation with Sherpa 2.2”. In: *SciPost Phys.* 7.3 (2019), p. 034. doi: [10.21468/SciPostPhys.7.3.034](https://doi.org/10.21468/SciPostPhys.7.3.034). arXiv: [1905.09127 \[hep-ph\]](https://arxiv.org/abs/1905.09127).
- [75] T. Sjöstrand, S. Mrenna, and P. Skands. “A brief introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. doi: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036). arXiv: [0710.3820 \[hep-ph\]](https://arxiv.org/abs/0710.3820).

- [76] ATLAS Collaboration. “The Pythia 8 A3 tune description of ATLAS minimum bias and inelastic measurements incorporating the Donnachie–Landshoff diffractive model”. In: (2016). URL: <https://cds.cern.ch/record/2206965>.
- [77] Richard D. Ball et al. “Parton distributions with LHC data”. In: *Nucl. Phys. B* 867 (2013), p. 244. doi: [10.1016/j.nuclphysb.2012.10.003](https://doi.org/10.1016/j.nuclphysb.2012.10.003). arXiv: [1207.1303 \[hep-ph\]](https://arxiv.org/abs/1207.1303).
- [78] *Search for resonant and non-resonant Higgs boson pair production in the  $b\bar{b}\tau^+\tau^-$  decay channel using 13 TeV pp collision data from the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-030>. Geneva: CERN, July 2021. URL: <https://cds.cern.ch/record/2777236>.
- [79] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. doi: [10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043). arXiv: [1002.2581 \[hep-ph\]](https://arxiv.org/abs/1002.2581).
- [80] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *J. Phys. G* 43 (2016), p. 023001. doi: [10.1088/0954-3899/43/2/023001](https://doi.org/10.1088/0954-3899/43/2/023001). arXiv: [1510.03865 \[hep-ph\]](https://arxiv.org/abs/1510.03865).
- [81] ATLAS Collaboration. “ATLAS Pythia 8 tunes to 7 TeV data”. In: (2014). URL: <https://cds.cern.ch/record/1966419>.
- [82] ATLAS Collaboration. *Summary of ATLAS Pythia 8 tunes*. ATL-PHYS-PUB-2012-003. 2012. URL: <https://cds.cern.ch/record/1474107>.
- [83] Massimiliano Grazzini et al. “Higgs boson pair production at NNLO with top quark mass effects”. In: *JHEP* 05 (2018), p. 059. doi: [10.1007/JHEP05\(2018\)059](https://doi.org/10.1007/JHEP05(2018)059). arXiv: [1803.02463 \[hep-ph\]](https://arxiv.org/abs/1803.02463).
- [84] ATLAS. “LHC Higgs Cross Section Working Group DiHiggs subgroup webpage”. 2021. URL: [https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGH?redirectTo=LHCPhysics.LHCHXSWGHH#Latest\\_recommendations\\_for\\_gluon](https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGH?redirectTo=LHCPhysics.LHCHXSWGHH#Latest_recommendations_for_gluon).
- [85] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. doi: [10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079). arXiv: [1405.0301 \[hep-ph\]](https://arxiv.org/abs/1405.0301).
- [86] Richard D. Ball et al. “Parton distributions for the LHC run II”. In: *JHEP* 04 (2015), p. 040. doi: [10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040). arXiv: [1410.8849 \[hep-ph\]](https://arxiv.org/abs/1410.8849).
- [87] Frédéric A. Dreyer and Alexander Karlberg. “Vector-boson fusion Higgs pair production at N<sup>3</sup>LO”. In: *Phys. Rev. D* 98.11 (2018), p. 114016. doi: [10.1103/PhysRevD.98.114016](https://doi.org/10.1103/PhysRevD.98.114016). arXiv: [1811.07906 \[hep-ph\]](https://arxiv.org/abs/1811.07906).
- [88] M. Bähr et al. “Herwig++ physics and manual”. In: *Eur. Phys. J. C* 58 (2008), p. 639. doi: [10.1140/epjc/s10052-008-0798-9](https://doi.org/10.1140/epjc/s10052-008-0798-9). arXiv: [0803.0883 \[hep-ph\]](https://arxiv.org/abs/0803.0883).

- [89] Johannes Bellm et al. “Herwig 7.0/Herwig++ 3.0 release note”. In: *Eur. Phys. J. C* 76.4 (2016), p. 196. doi: [10.1140/epjc/s10052-016-4018-8](https://doi.org/10.1140/epjc/s10052-016-4018-8). arXiv: [1512.01178 \[hep-ph\]](https://arxiv.org/abs/1512.01178).
- [90] Stefan Gieseke, Christian Röhr, and Andrzej Siodmok. “Colour reconnections in Herwig++”. In: *Eur. Phys. J. C* 72 (2012), p. 2225. doi: [10.1140/epjc/s10052-012-2225-5](https://doi.org/10.1140/epjc/s10052-012-2225-5). arXiv: [1206.0041 \[hep-ph\]](https://arxiv.org/abs/1206.0041).
- [91] “Validation of signal Monte Carlo event generation in searches for Higgs boson pairs with the ATLAS detector”. In: (2019).
- [92] M. Capozi and G. Heinrich. “Exploring anomalous couplings in Higgs boson pair production through shape analysis”. In: *Journal of High Energy Physics* 2020.3 (Mar. 2020). issn: 1029-8479. doi: [10.1007/jhep03\(2020\)091](https://doi.org/10.1007/jhep03(2020)091). URL: [http://dx.doi.org/10.1007/JHEP03\(2020\)091](http://dx.doi.org/10.1007/JHEP03(2020)091).
- [93] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. “A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”. In: *JHEP* 09 (2007), p. 126. doi: [10.1088/1126-6708/2007/09/126](https://doi.org/10.1088/1126-6708/2007/09/126). arXiv: [0707.3088 \[hep-ph\]](https://arxiv.org/abs/0707.3088).
- [94] Paolo Nason. “A New method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. doi: [10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040). arXiv: [hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146).
- [95] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. doi: [10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070). arXiv: [0709.2092 \[hep-ph\]](https://arxiv.org/abs/0709.2092).
- [96] ATLAS Collaboration. “Studies on top-quark Monte Carlo modelling with Sherpa and MG5\_aMC@NLO”. In: (2017). url: <https://cds.cern.ch/record/2261938>.
- [97] Michal Czakon and Alexander Mitov. “Top++: A program for the calculation of the top-pair cross-section at hadron colliders”. In: *Comput. Phys. Commun.* 185 (2014), p. 2930. doi: [10.1016/j.cpc.2014.06.021](https://doi.org/10.1016/j.cpc.2014.06.021). arXiv: [1112.5675 \[hep-ph\]](https://arxiv.org/abs/1112.5675).
- [98] N. Kidonakis. “Next-to-next-to-leading logarithm resummation for  $s$ -channel single top quark production”. In: *Phys. Rev. D* 81 (2010), p. 054028. doi: [10.1103/PhysRevD.81.054028](https://doi.org/10.1103/PhysRevD.81.054028). arXiv: [1001.5034 \[hep-ph\]](https://arxiv.org/abs/1001.5034).
- [99] N. Kidonakis. “Next-to-next-to-leading-order collinear and soft gluon corrections for  $t$ -channel single top quark production”. In: *Phys. Rev. D* 83 (2011), p. 091503. doi: [10.1103/PhysRevD.83.091503](https://doi.org/10.1103/PhysRevD.83.091503). arXiv: [1103.2792 \[hep-ph\]](https://arxiv.org/abs/1103.2792).
- [100] Nikolaos Kidonakis. “Two-loop soft anomalous dimensions for single top quark associated production with a  $W^-$  or  $H^-$ ”. In: *Phys. Rev. D* 82 (2010), p. 054018. doi: [10.1103/PhysRevD.82.054018](https://doi.org/10.1103/PhysRevD.82.054018). arXiv: [1005.4451 \[hep-ph\]](https://arxiv.org/abs/1005.4451).

- [101] Stefano Frixione et al. “Single-top hadroproduction in association with a W boson”. In: *JHEP* 07 (2008), p. 029. doi: [10.1088/1126-6708/2008/07/029](https://doi.org/10.1088/1126-6708/2008/07/029). arXiv: [0805.3067 \[hep-ph\]](https://arxiv.org/abs/0805.3067).
- [102] Tanju Gleisberg and Stefan Höche. “Comix, a new matrix element generator”. In: *JHEP* 12 (2008), p. 039. doi: [10.1088/1126-6708/2008/12/039](https://doi.org/10.1088/1126-6708/2008/12/039). arXiv: [0808.3674 \[hep-ph\]](https://arxiv.org/abs/0808.3674).
- [103] Fabio Caccioli, Philipp Maierhöfer, and Stefano Pozzorini. “Scattering Amplitudes with Open Loops”. In: *Phys. Rev. Lett.* 108 (2012), p. 111601. doi: [10.1103/PhysRevLett.108.111601](https://doi.org/10.1103/PhysRevLett.108.111601). arXiv: [1111.5206 \[hep-ph\]](https://arxiv.org/abs/1111.5206).
- [104] Charalampos Anastasiou et al. “High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order”. In: *Phys. Rev. D* 69 (2004), p. 094008. doi: [10.1103/PhysRevD.69.094008](https://doi.org/10.1103/PhysRevD.69.094008). arXiv: [hep-ph/0312266](https://arxiv.org/abs/hep-ph/0312266).
- [105] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), p. 159. doi: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024). arXiv: [1410.3012 \[hep-ph\]](https://arxiv.org/abs/1410.3012).
- [106] Keith Hamilton et al. “NNLOPS simulation of Higgs boson production”. In: *JHEP* 10 (2013), p. 222. doi: [10.1007/JHEP10\(2013\)222](https://doi.org/10.1007/JHEP10(2013)222). arXiv: [1309.0017 \[hep-ph\]](https://arxiv.org/abs/1309.0017).
- [107] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. “Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator”. In: *JHEP* 05 (2015), p. 140. doi: [10.1007/JHEP05\(2015\)140](https://doi.org/10.1007/JHEP05(2015)140). arXiv: [1501.04637 \[hep-ph\]](https://arxiv.org/abs/1501.04637).
- [108] D. de Florian et al. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. In: (2016). doi: [10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002). arXiv: [1610.07922 \[hep-ph\]](https://arxiv.org/abs/1610.07922).
- [109] Charalampos Anastasiou et al. “High precision determination of the gluon fusion Higgs boson cross-section at the LHC”. In: *JHEP* 05 (2016), p. 058. doi: [10.1007/JHEP05\(2016\)058](https://doi.org/10.1007/JHEP05(2016)058). arXiv: [1602.00695 \[hep-ph\]](https://arxiv.org/abs/1602.00695).
- [110] Charalampos Anastasiou et al. “Higgs Boson Gluon-Fusion Production in QCD at Three Loops”. In: *Phys. Rev. Lett.* 114 (2015), p. 212001. doi: [10.1103/PhysRevLett.114.212001](https://doi.org/10.1103/PhysRevLett.114.212001). arXiv: [1503.06056 \[hep-ph\]](https://arxiv.org/abs/1503.06056).
- [111] Falko Dulat, Achilleas Lazopoulos, and Bernhard Mistlberger. “iHixs 2 – Inclusive Higgs cross sections”. In: *Comput. Phys. Commun.* 233 (2018), pp. 243–260. doi: [10.1016/j.cpc.2018.06.025](https://doi.org/10.1016/j.cpc.2018.06.025). arXiv: [1802.00827 \[hep-ph\]](https://arxiv.org/abs/1802.00827).
- [112] Stefano Actis et al. “NLO electroweak corrections to Higgs boson production at hadron colliders”. In: *Phys. Lett. B* 670 (2008), pp. 12–17. doi: [10.1016/j.physletb.2008.10.018](https://doi.org/10.1016/j.physletb.2008.10.018). arXiv: [0809.1301 \[hep-ph\]](https://arxiv.org/abs/0809.1301).
- [113] J. Pumplin et al. “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis”. In: *JHEP* 07 (2002), p. 012. doi: [10.1088/1126-6708/2002/07/012](https://doi.org/10.1088/1126-6708/2002/07/012). arXiv: [hep-ph/0201195 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0201195).

- [114] ATLAS Collaboration. “Measurement of the  $Z/\gamma^*$  boson transverse momentum distribution in pp collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector”. In: *JHEP* 09 (2014), p. 145. doi: [10.1007/JHEP09\(2014\)145](https://doi.org/10.1007/JHEP09(2014)145). arXiv: [1406.3660 \[hep-ex\]](https://arxiv.org/abs/1406.3660).
- [115] M. Ciccolini, Ansgar Denner, and S. Dittmaier. “Strong and Electroweak Corrections to the Production of Higgs + 2 Jets via Weak Interactions at the Large Hadron Collider”. In: *Phys. Rev. Lett.* 99 (2007), p. 161803. doi: [10.1103/PhysRevLett.99.161803](https://doi.org/10.1103/PhysRevLett.99.161803). arXiv: [0707.0381 \[hep-ph\]](https://arxiv.org/abs/0707.0381).
- [116] Mariano Ciccolini, Ansgar Denner, and Stefan Dittmaier. “Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC”. In: *Phys. Rev. D* 77 (2008), p. 013002. doi: [10.1103/PhysRevD.77.013002](https://doi.org/10.1103/PhysRevD.77.013002). arXiv: [0710.4749 \[hep-ph\]](https://arxiv.org/abs/0710.4749).
- [117] Paolo Bolzoni et al. “Higgs Boson Production via Vector-Boson Fusion at Next-to-Next-to-Leading Order in QCD”. In: *Phys. Rev. Lett.* 105 (2010), p. 011801. doi: [10.1103/PhysRevLett.105.011801](https://doi.org/10.1103/PhysRevLett.105.011801). arXiv: [1003.4451 \[hep-ph\]](https://arxiv.org/abs/1003.4451).
- [118] M.L. Ciccolini, S. Dittmaier, and M. Krämer. “Electroweak radiative corrections to associated  $WH$  and  $ZH$  production at hadron colliders”. In: *Phys. Rev. D* 68 (2003), p. 073003. doi: [10.1103/PhysRevD.68.073003](https://doi.org/10.1103/PhysRevD.68.073003). arXiv: [hep-ph/0306234 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0306234).
- [119] Oliver Brein, Abdelhak Djouadi, and Robert Harlander. “NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders”. In: *Phys. Lett. B* 579 (2004), pp. 149–156. doi: [10.1016/j.physletb.2003.10.112](https://doi.org/10.1016/j.physletb.2003.10.112). arXiv: [hep-ph/0307206 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0307206).
- [120] Giancarlo Ferrera, Massimiliano Grazzini, and Francesco Tramontano. “Associated Higgs-W-Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at NNLO”. In: *Phys. Rev. Lett.* 107 (2011), p. 152003. doi: [10.1103/PhysRevLett.107.152003](https://doi.org/10.1103/PhysRevLett.107.152003). arXiv: [1107.1164 \[hep-ph\]](https://arxiv.org/abs/1107.1164).
- [121] Oliver Brein et al. “Top-quark mediated effects in hadronic Higgs-Strahlung”. In: *Eur. Phys. J. C* 72 (2012), p. 1868. doi: [10.1140/epjc/s10052-012-1868-6](https://doi.org/10.1140/epjc/s10052-012-1868-6). arXiv: [1111.0761 \[hep-ph\]](https://arxiv.org/abs/1111.0761).
- [122] Giancarlo Ferrera, Massimiliano Grazzini, and Francesco Tramontano. “Higher-order QCD effects for associated  $WH$  production and decay at the LHC”. In: *JHEP* 04 (2014), p. 039. doi: [10.1007/JHEP04\(2014\)039](https://doi.org/10.1007/JHEP04(2014)039). arXiv: [1312.1669 \[hep-ph\]](https://arxiv.org/abs/1312.1669).
- [123] Giancarlo Ferrera, Massimiliano Grazzini, and Francesco Tramontano. “Associated  $ZH$  production at hadron colliders: The fully differential NNLO QCD calculation”. In: *Phys. Lett. B* 740 (2015), pp. 51–55. doi: [10.1016/j.physletb.2014.11.040](https://doi.org/10.1016/j.physletb.2014.11.040). arXiv: [1407.4747 \[hep-ph\]](https://arxiv.org/abs/1407.4747).
- [124] John M. Campbell, R. Keith Ellis, and Ciaran Williams. “Associated production of a Higgs boson at NNLO”. In: *JHEP* 06 (2016), p. 179. doi: [10.1007/JHEP06\(2016\)179](https://doi.org/10.1007/JHEP06(2016)179). arXiv: [1601.00658 \[hep-ph\]](https://arxiv.org/abs/1601.00658).

- [125] Lukas Altenkamp et al. “Gluon-induced Higgs-strahlung at next-to-leading order QCD”. In: *JHEP* 02 (2013), p. 078. doi: [10.1007/JHEP02\(2013\)078](https://doi.org/10.1007/JHEP02(2013)078). arXiv: [1211.5015](https://arxiv.org/abs/1211.5015) [hep-ph].
- [126] B. Hespel, F. Maltoni, and E. Vryonidou. “Higgs and Z boson associated production via gluon fusion in the SM and the 2HDM”. In: *JHEP* 06 (2015), p. 065. doi: [10.1007/JHEP06\(2015\)065](https://doi.org/10.1007/JHEP06(2015)065). arXiv: [1503.01656](https://arxiv.org/abs/1503.01656) [hep-ph].
- [127] Robert V. Harlander et al. “Soft gluon resummation for gluon-induced Higgs Strahlung”. In: *JHEP* 11 (2014), p. 082. doi: [10.1007/JHEP11\(2014\)082](https://doi.org/10.1007/JHEP11(2014)082). arXiv: [1410.0217](https://arxiv.org/abs/1410.0217) [hep-ph].
- [128] Robert V. Harlander, Stefan Liebler, and Tom Zirke. “Higgs Strahlung at the Large Hadron Collider in the 2-Higgs-doublet model”. In: *JHEP* 02 (2014), p. 023. doi: [10.1007/JHEP02\(2014\)023](https://doi.org/10.1007/JHEP02(2014)023). arXiv: [1307.8122](https://arxiv.org/abs/1307.8122) [hep-ph].
- [129] Oliver Brein, Robert V. Harlander, and Tom J. E. Zirke. “vh@nnlo - Higgs Strahlung at hadron colliders”. In: *Comput. Phys. Commun.* 184 (2013), pp. 998–1003. doi: [10.1016/j.cpc.2012.11.002](https://doi.org/10.1016/j.cpc.2012.11.002). arXiv: [1210.5347](https://arxiv.org/abs/1210.5347) [hep-ph].
- [130] ATLAS Collaboration. “Calibration of light-flavour  $b$ -jet mistagging rates using ATLAS proton–proton collision data at  $\sqrt{s} = 13$  TeV”. In: (2018). URL: <https://cds.cern.ch/record/2314418>.
- [131] David Krohn, Jesse Thaler, and Lian-Tao Wang. “Jets with variable R”. In: 2009.06 (June 2009), pp. 059–059. doi: [10.1088/1126-6708/2009/06/059](https://doi.org/10.1088/1126-6708/2009/06/059). URL: <https://doi.org/10.1088/1126-6708/2009/06/059>.
- [132] G. Abbiendi et al. “A measurement of the rate of charm production in W decays”. In: *Physics Letters B* 490.1-2 (Sept. 2000), pp. 71–86. ISSN: 0370-2693. doi: [10.1016/S0370-2693\(00\)00971-0](https://doi.org/10.1016/S0370-2693(00)00971-0). URL: [http://dx.doi.org/10.1016/S0370-2693\(00\)00971-0](http://dx.doi.org/10.1016/S0370-2693(00)00971-0).
- [133] K.A. Olive. “Review of Particle Physics”. In: *Chinese Physics C* 38.9 (Aug. 2014), p. 090001. doi: [10.1088/1674-1137/38/9/090001](https://doi.org/10.1088/1674-1137/38/9/090001). URL: <https://doi.org/10.1088/1674-1137/38/9/090001>.
- [134] Johannes Erdmann et al. “A likelihood-based reconstruction algorithm for top-quark pairs and the KLFitter framework”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 748 (2014), pp. 18–25. ISSN: 0168-9002. doi: <https://doi.org/10.1016/j.nima.2014.02.029>.
- [135] ATLAS Collaboration. “Studies on top-quark Monte Carlo modelling for Top2016”. In: (2016). URL: <https://cds.cern.ch/record/2216168>.

- [136] ATLAS Collaboration. “Measurements of inclusive and differential fiducial cross-sections of  $t\bar{t}$  production with additional heavy-flavour jets in proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *JHEP* 04 (2019), p. 046. doi: [10.1007/JHEP04\(2019\)046](https://doi.org/10.1007/JHEP04(2019)046). arXiv: [1811.12113 \[hep-ex\]](https://arxiv.org/abs/1811.12113).
- [137] *Combination of searches for non-resonant and resonant Higgs boson pair production in the  $b\bar{b}\gamma\gamma$ ,  $b\bar{b}\tau^+\tau^-$  and  $b\bar{b}b\bar{b}$  decay channels using  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-052>. Geneva: CERN, Oct. 2021. URL: <https://cds.cern.ch/record/2786865>.
- [138] A. Elagin et al. “A new mass reconstruction technique for resonances decaying to  $\tau\tau$ ”. In: *Nucl. Instrum. Meth. A* 654 (2011), pp. 481–489. doi: [10.1016/j.nima.2011.07.009](https://doi.org/10.1016/j.nima.2011.07.009). arXiv: [1012.4686 \[hep-ex\]](https://arxiv.org/abs/1012.4686).
- [139] ATLAS. “LHC Fake Taus Task Force webpage”. 2020. URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/FakeTausTaskForce>.
- [140] ATLAS Collaboration. “Search for non-resonant Higgs boson pair production in the  $b\bar{b}\ell\nu\ell\nu$  final state with the ATLAS detector in  $pp$  collisions at  $\sqrt{s} = 13$  TeV”. In: *Phys. Lett. B* 801 (2020), p. 135145. doi: [10.1016/j.physletb.2019.135145](https://doi.org/10.1016/j.physletb.2019.135145). arXiv: [1908.06765 \[hep-ex\]](https://arxiv.org/abs/1908.06765).
- [141] Pierre Baldi et al. “Parameterized neural networks for high-energy physics”. In: *Eur. Phys. J. C* 76.5 (2016), p. 235. doi: [10.1140/epjc/s10052-016-4099-4](https://doi.org/10.1140/epjc/s10052-016-4099-4). arXiv: [1601.07913 \[hep-ex\]](https://arxiv.org/abs/1601.07913).
- [142] ATLAS Collaboration. “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector”. In: *JHEP* 04 (2015), p. 117. doi: [10.1007/JHEP04\(2015\)117](https://doi.org/10.1007/JHEP04(2015)117). arXiv: [1501.04943 \[hep-ex\]](https://arxiv.org/abs/1501.04943).
- [143] François Chollet et al. *Keras*. <https://keras.io>. 2015.
- [144] *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: <https://www.tensorflow.org/>.
- [145] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016.
- [146] ATLAS Collaboration. “Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 8$  TeV using the ATLAS detector at the LHC”. In: *Eur. Phys. J. C* 76 (2016), p. 653. doi: [10.1140/epjc/s10052-016-4466-1](https://doi.org/10.1140/epjc/s10052-016-4466-1). arXiv: [1608.03953 \[hep-ex\]](https://arxiv.org/abs/1608.03953).
- [147] ATLAS Collaboration. “Measurement of the Inelastic Proton–Proton Cross Section at  $\sqrt{s} = 13$  TeV with the ATLAS Detector at the LHC”. In: *Phys. Rev. Lett.* 117 (2016), p. 182002. doi: [10.1103/PhysRevLett.117.182002](https://doi.org/10.1103/PhysRevLett.117.182002). arXiv: [1606.02625 \[hep-ex\]](https://arxiv.org/abs/1606.02625).

- [148] ATLAS Collaboration. “Measurement of the  $c$ -jet mistagging efficiency in  $t\bar{t}$  events using  $pp$  collision data at  $\sqrt{s} = 13$  TeV collected with the ATLAS detector”. In: (Sept. 2021). arXiv: [2109.10627 \[hep-ex\]](https://arxiv.org/abs/2109.10627).
- [149] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at  $\sqrt{s} = 13$  TeV”. In: *Eur. Phys. J. C* 78 (2018), p. 903. doi: [10.1140/epjc/s10052-018-6288-9](https://doi.org/10.1140/epjc/s10052-018-6288-9). arXiv: [1802.08168 \[hep-ex\]](https://arxiv.org/abs/1802.08168).
- [150] ATLAS Collaboration. “Measurements of  $WH$  and  $ZH$  production in the  $H \rightarrow b\bar{b}$  decay channel in  $pp$  collisions at 13 TeV with the ATLAS detector”. In: (2020). URL: <https://cds.cern.ch/record/2714885>.
- [151] Konie Al Khoury et al. *Data/Monte Carlo Modelling Studies for the Standard Model VH,  $H \rightarrow b\bar{b}$  Analysis*. Tech. rep. ATL-COM-PHYS-2019-1205. Geneva: CERN, Sept. 2019. URL: <https://cds.cern.ch/record/2690042>.
- [152] Eilam Gross and Ofer Vitells. “Trial factors for the look elsewhere effect in high energy physics”. In: *The European Physical Journal C* 70.1-2 (Oct. 2010), pp. 525–530. ISSN: 1434-6052. doi: [10.1140/epjc/s10052-010-1470-8](https://doi.org/10.1140/epjc/s10052-010-1470-8). URL: <http://dx.doi.org/10.1140/epjc/s10052-010-1470-8>.
- [153] CMS Collaboration. “Combination of Searches for Higgs Boson Pair Production in Proton-Proton Collisions at  $\sqrt{s} = 13$  TeV”. In: *Phys. Rev. Lett.* 122 (12 Mar. 2019), p. 121803. doi: [10.1103/PhysRevLett.122.121803](https://doi.org/10.1103/PhysRevLett.122.121803). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.122.121803>.
- [154] *Search for nonresonant Higgs boson pair production in the 4 leptons plus 2 b jets final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV*. Tech. rep. Geneva: CERN, 2020. URL: <https://cds.cern.ch/record/2725691>.
- [155] *Search for Higgs boson pair production in the four b quark final state*. Tech. rep. Geneva: CERN, 2021. URL: <https://cds.cern.ch/record/2771912>.
- [156] A. M. Sirunyan et al. “Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at  $\sqrt{s} = 13$  TeV”. In: *Journal of High Energy Physics* 2021.3 (Mar. 2021). ISSN: 1029-8479. doi: [10.1007/jhep03\(2021\)257](https://doi.org/10.1007/jhep03(2021)257). URL: [http://dx.doi.org/10.1007/JHEP03\(2021\)257](http://dx.doi.org/10.1007/JHEP03(2021)257).

$$\sqrt{s}$$

$= 13$  TeV”. In: *Journal of High Energy Physics* 2021.3 (Mar. 2021). ISSN: 1029-8479. doi: [10.1007/jhep03\(2021\)257](https://doi.org/10.1007/jhep03(2021)257). URL: [http://dx.doi.org/10.1007/JHEP03\(2021\)257](http://dx.doi.org/10.1007/JHEP03(2021)257).

## **Appendix A**

# **Supplementary material for $c$ -jet calibration**

### **A.1 Additional plots for kinematic variables**

#### **A.1.1 Standard selection**

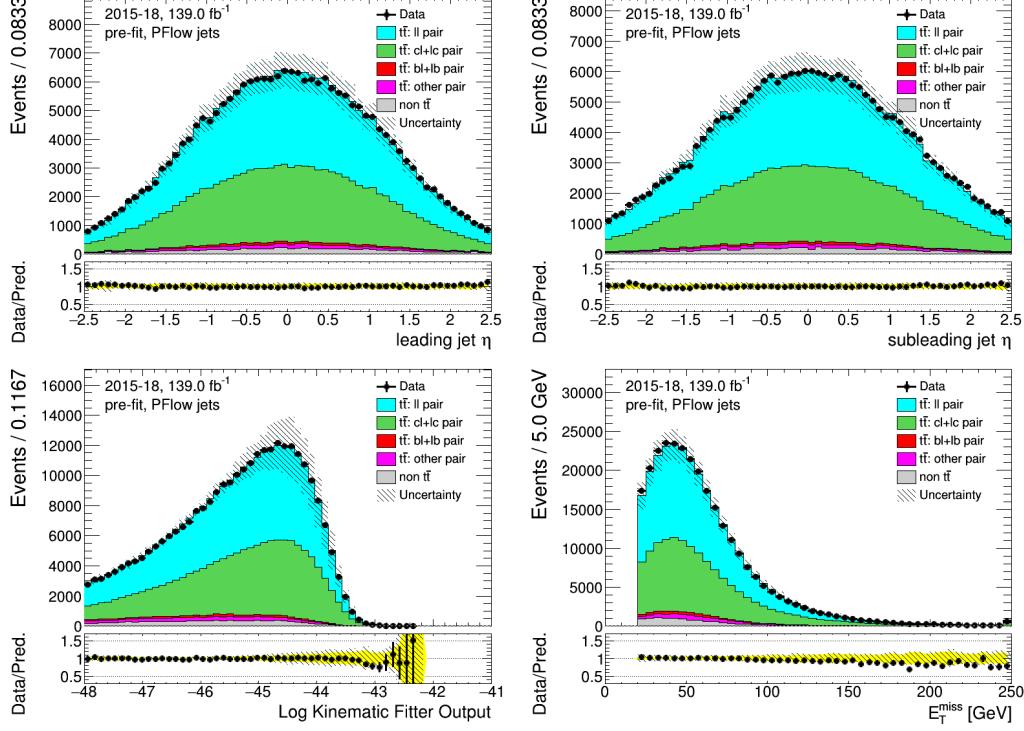


Figure A.1: PFFlow jets: distributions of the leading and sub-leading jets from W decay, KLFitter output and the transverse missing transverse energy of the standard selection, before fitting or tagging with full uncertainties.

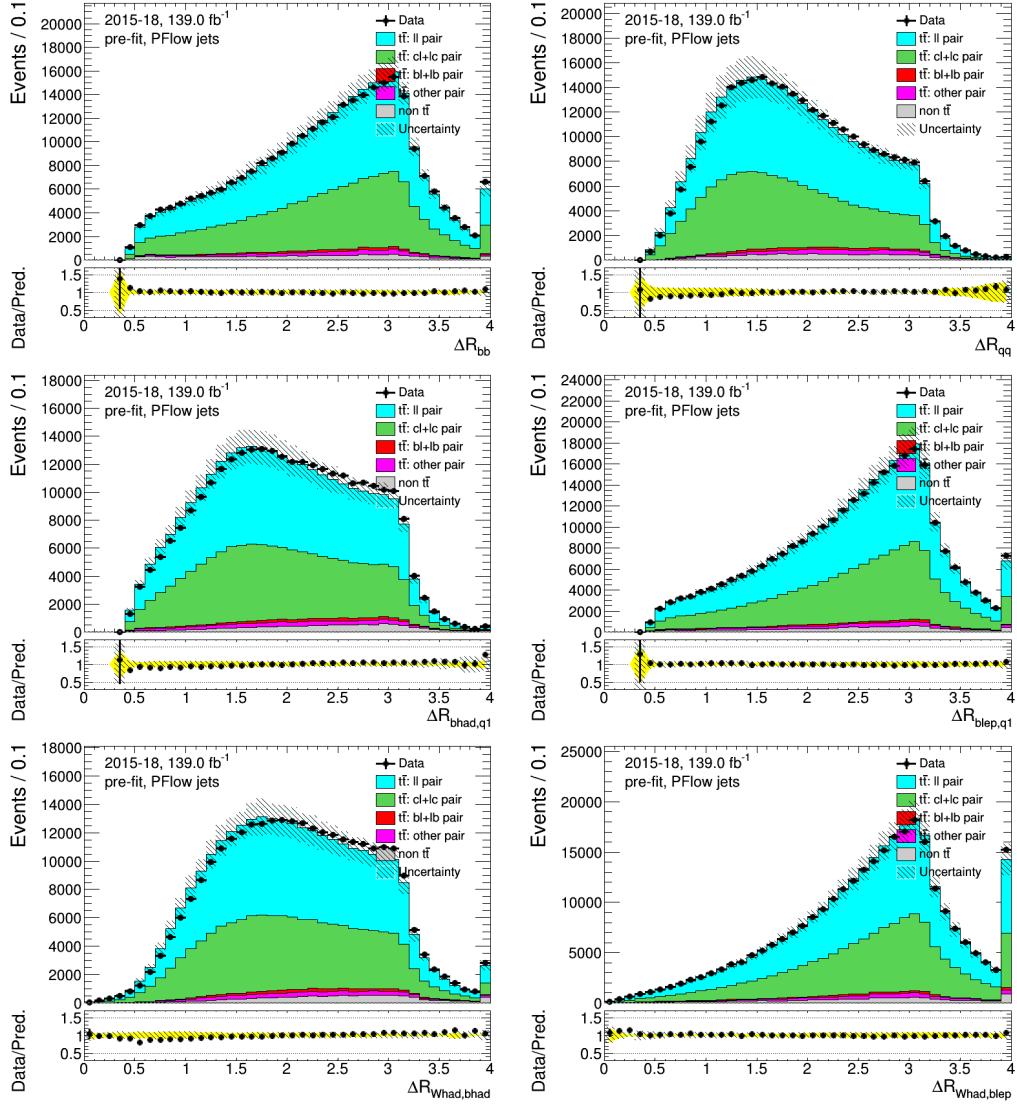


Figure A.2: PFlow jets: distributions of angle related variables of the combination of the standard selection, before fitting or tagging with full uncertainties.

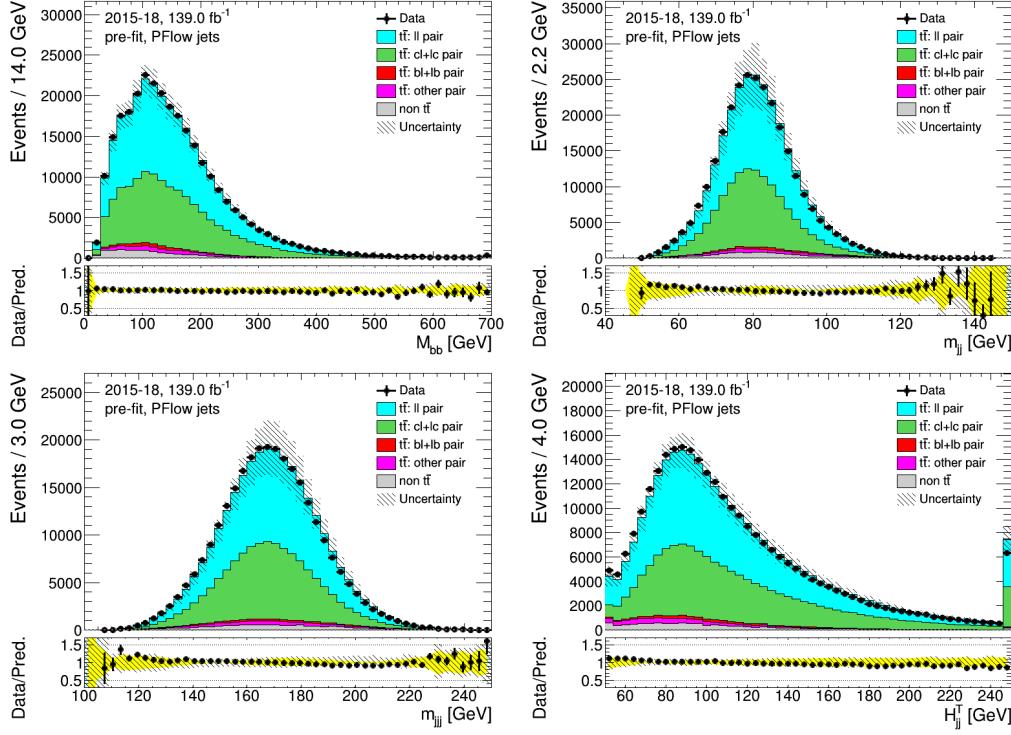


Figure A.3: PFlow jets: distributions of mass related variables of the standard selection, before fitting or tagging with stat-only uncertainties.

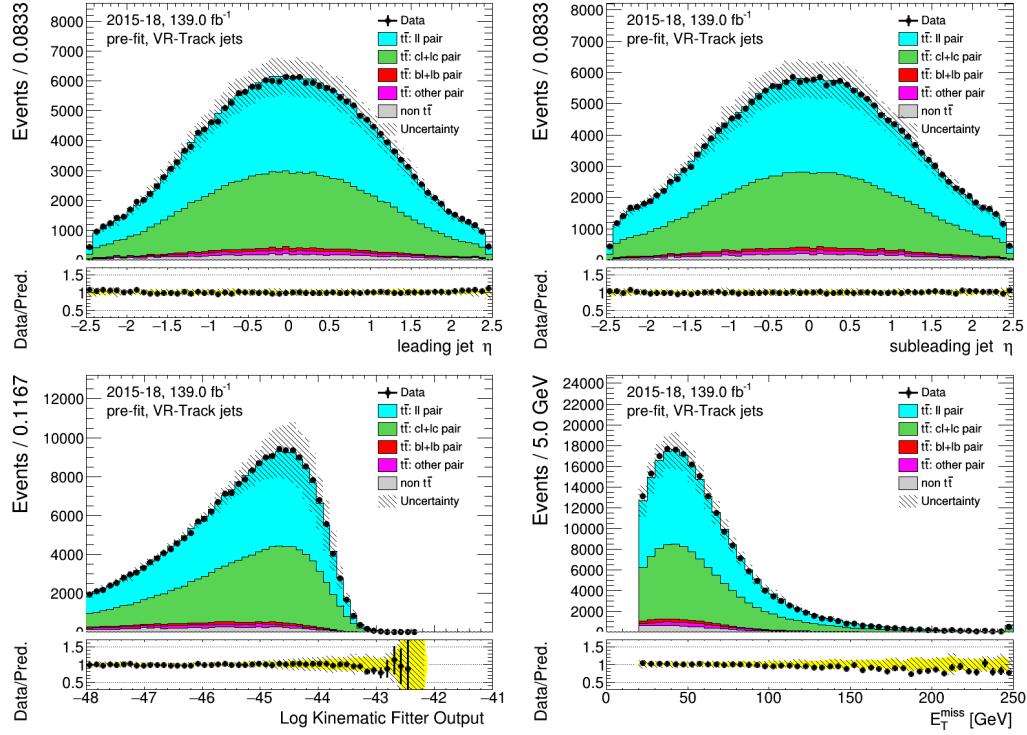


Figure A.4: VR-Track jets: distributions of the leading and sub-leading jets from W decay, KLFitter output and the transverse missing transverse energy of the standard selection, before fitting or tagging with full uncertainties.

## A.1. ADDITIONAL PLOTS FOR KINEMATIC VARIABLES

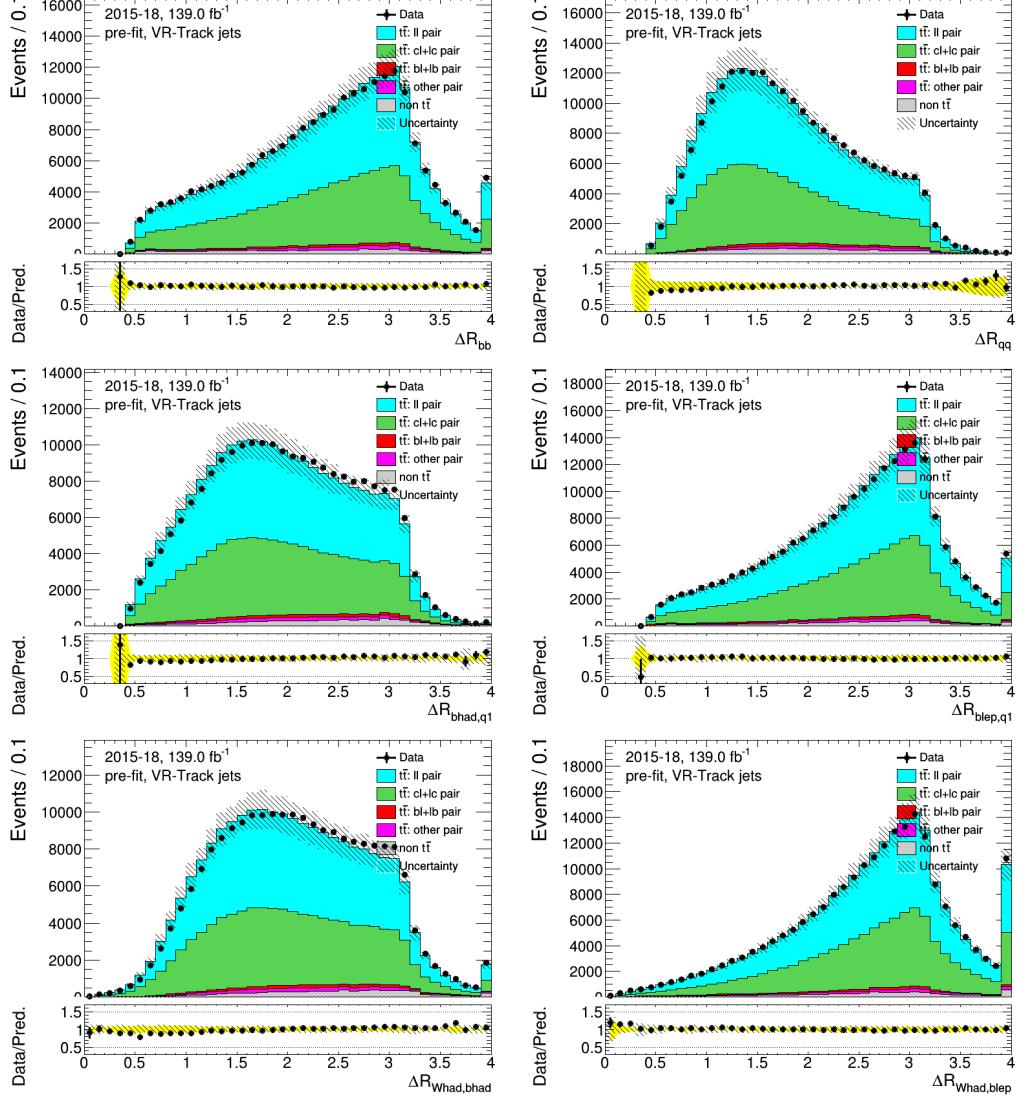


Figure A.5: VR-Track jets: distributions of angle related variables of the combination of the standard selection, before fitting or tagging with full uncertainties.

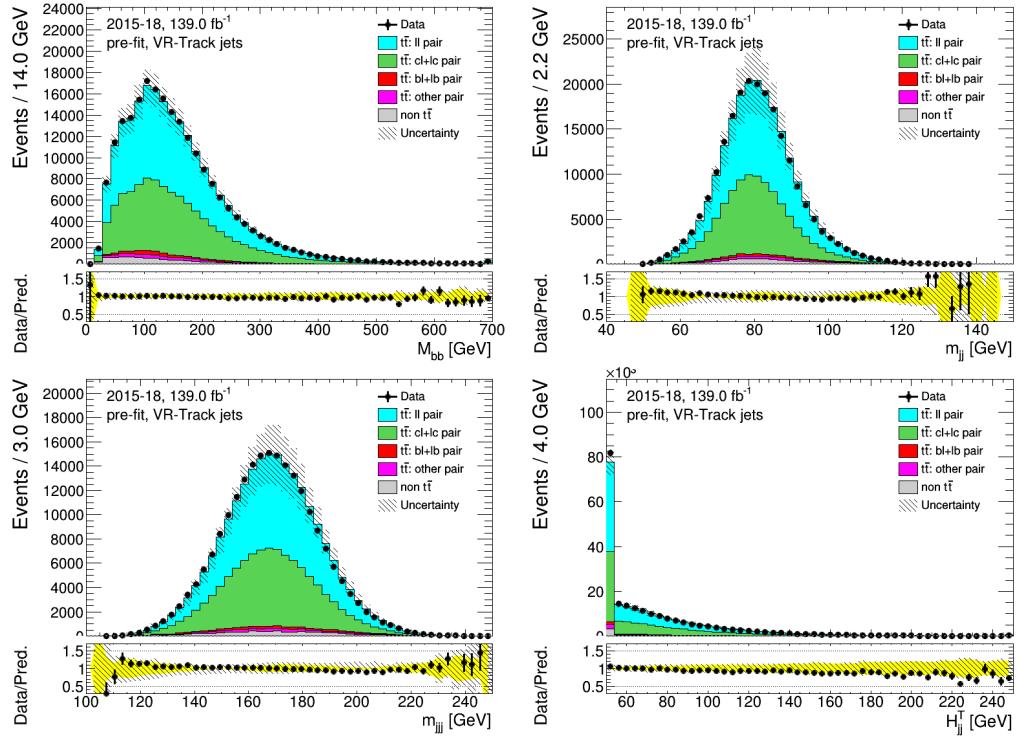


Figure A.6: VR-Track jets: distributions of mass related variables of the standard selection, before fitting or tagging with stat-only uncertainties.

### A.1.2 Low- $p_T$ selection

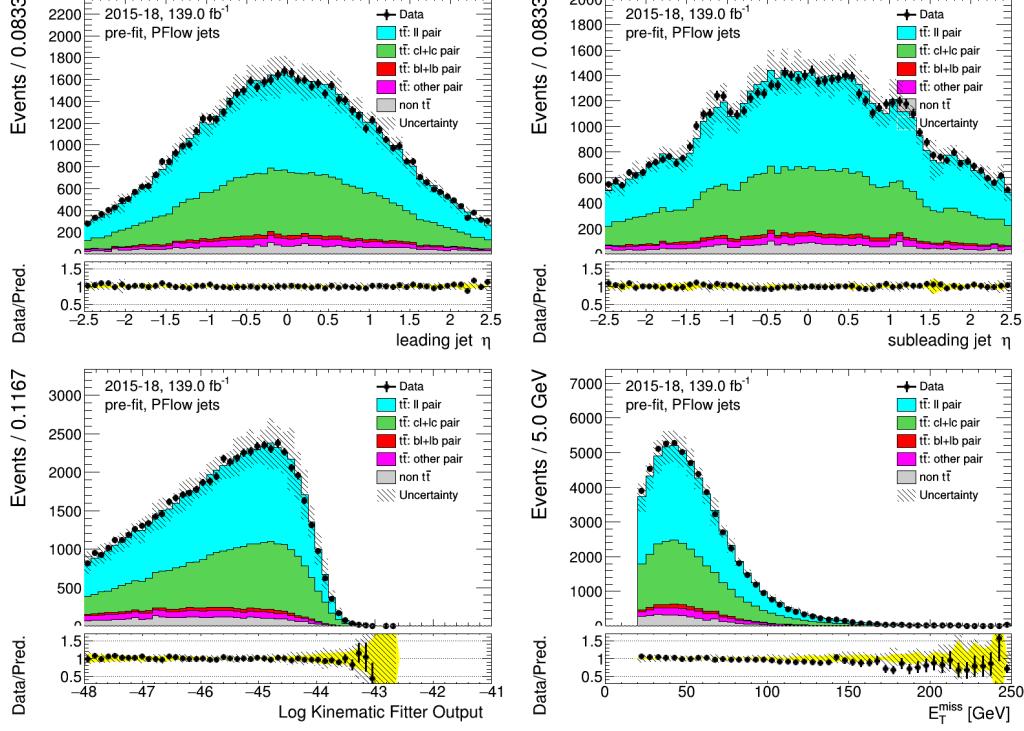


Figure A.7: PFFlow jets: distributions of the leading and sub-leading jets from W decay, KLFitter output and the transverse missing transverse energy of the low- $p_T$  selection, before fitting or tagging with full uncertainties.

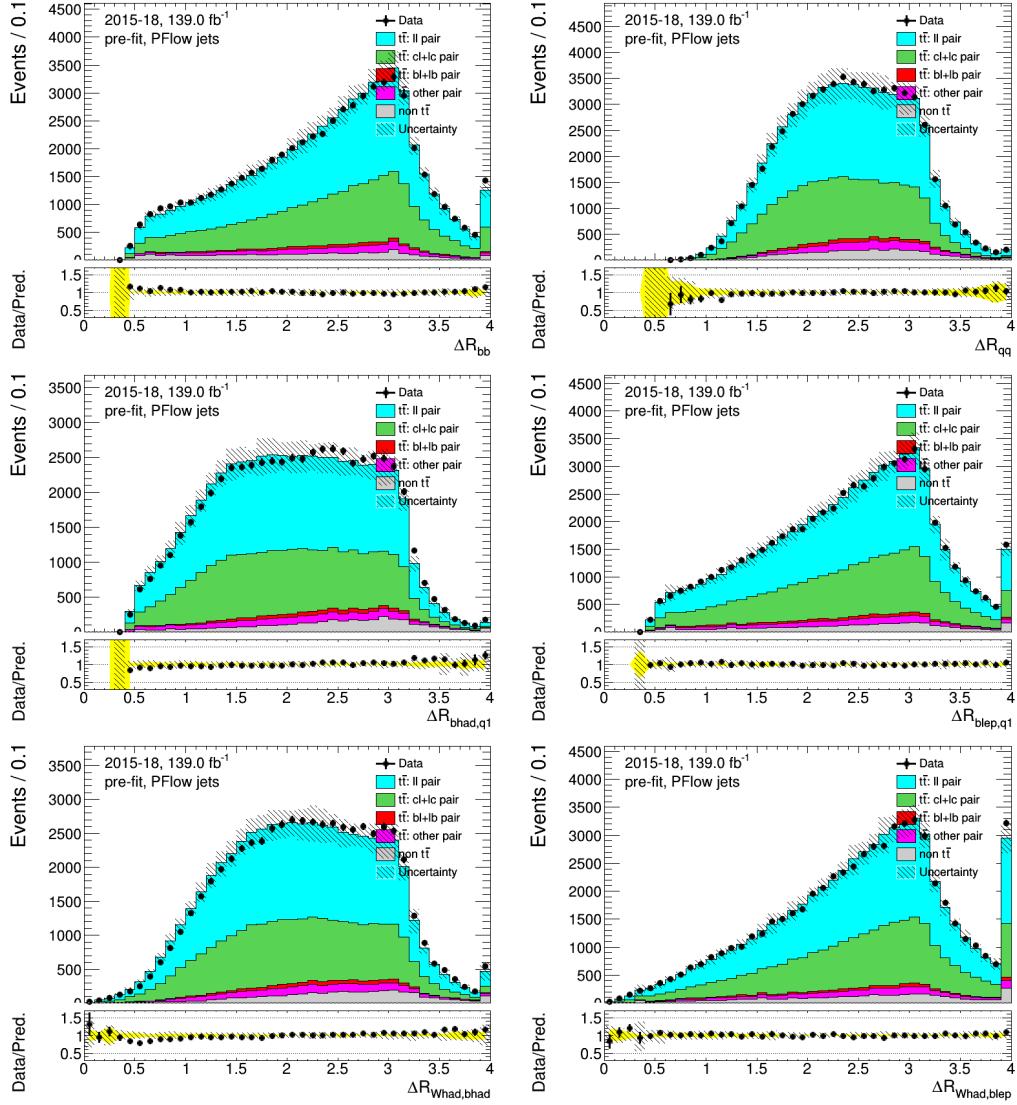


Figure A.8: PFlow jets: distributions of angle related variables of the combination of the low- $p_T$  selection, before fitting or tagging with full uncertainties.

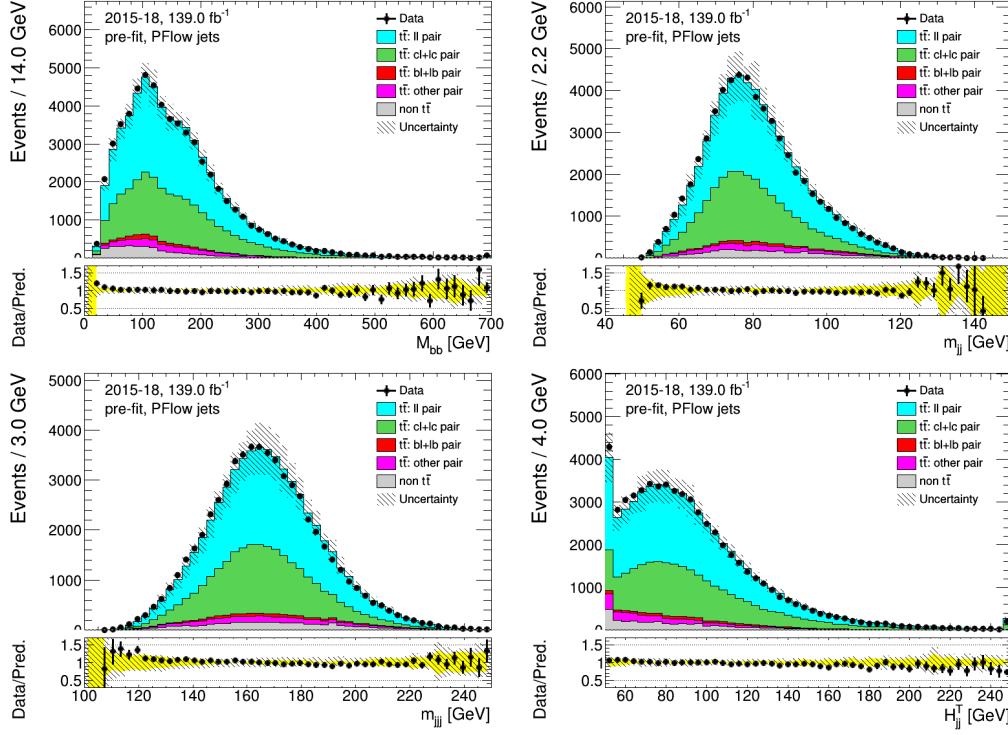


Figure A.9: PFlow jets: distributions of mass related variables of the low- $p_T$  selection, before fitting or tagging with stat-only uncertainties.

## A.2 High- $p_T$ selection

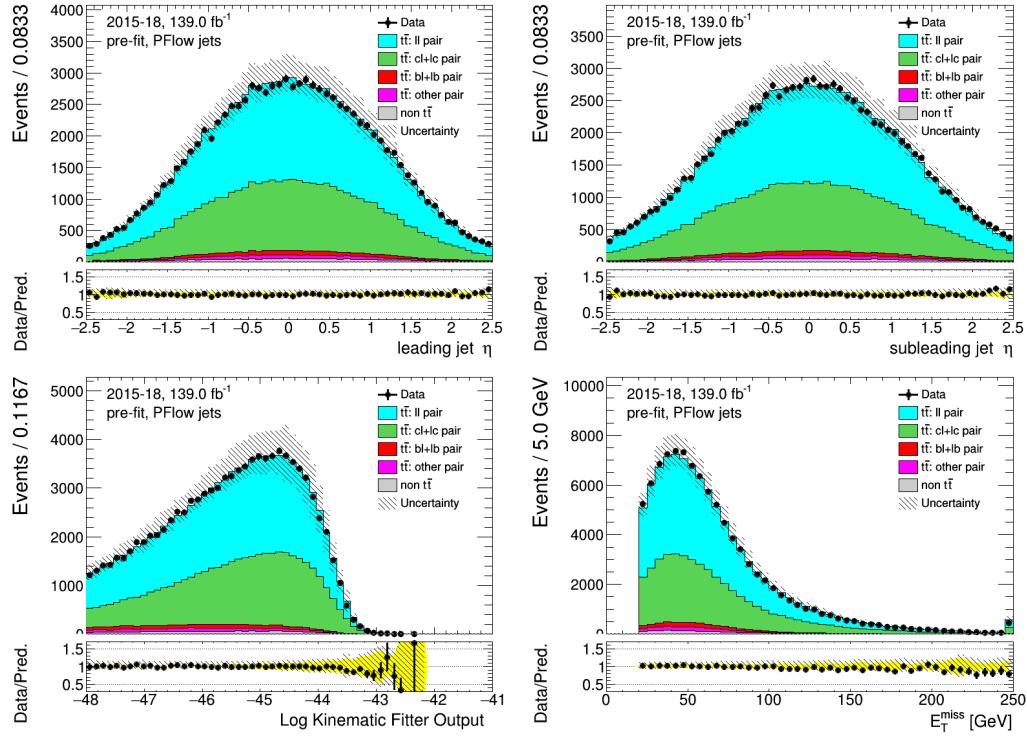


Figure A.10: PFlow jets: distributions of the leading and sub-leading jets from W decay, KLFitter output and the transverse missing transverse energy of the high- $p_T$  selection, before fitting or tagging with full uncertainties.

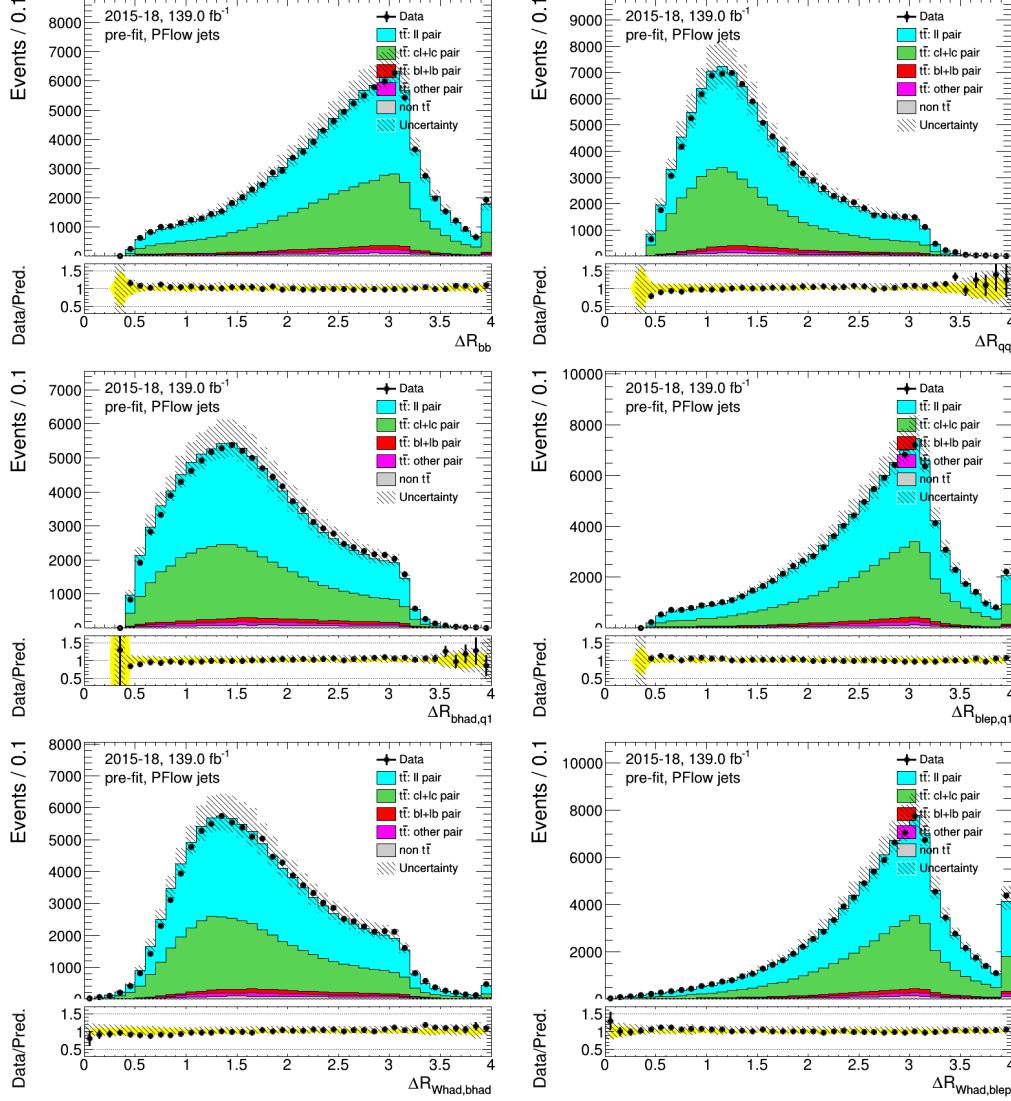


Figure A.11: PFlow jets: distributions of angle related variables of the combination of the high- $p_T$  selection, before fitting or tagging with full uncertainties.

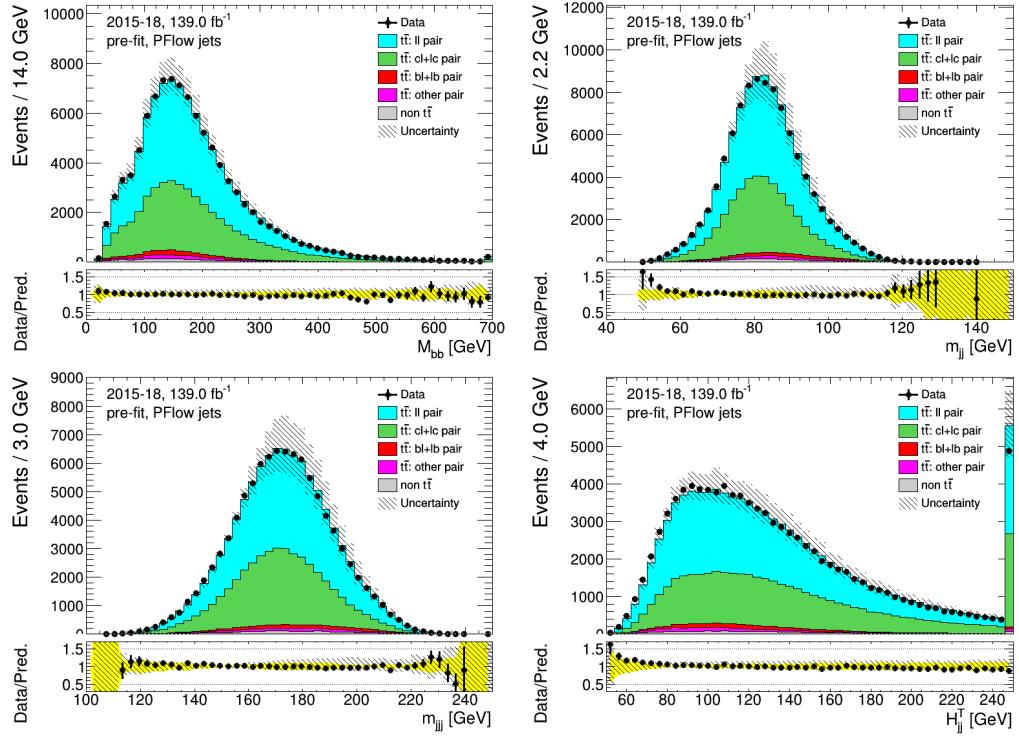


Figure A.12: PFlow jets: distributions of mass related variables of the high- $p_T$  selection, before fitting or tagging with stat-only uncertainties.

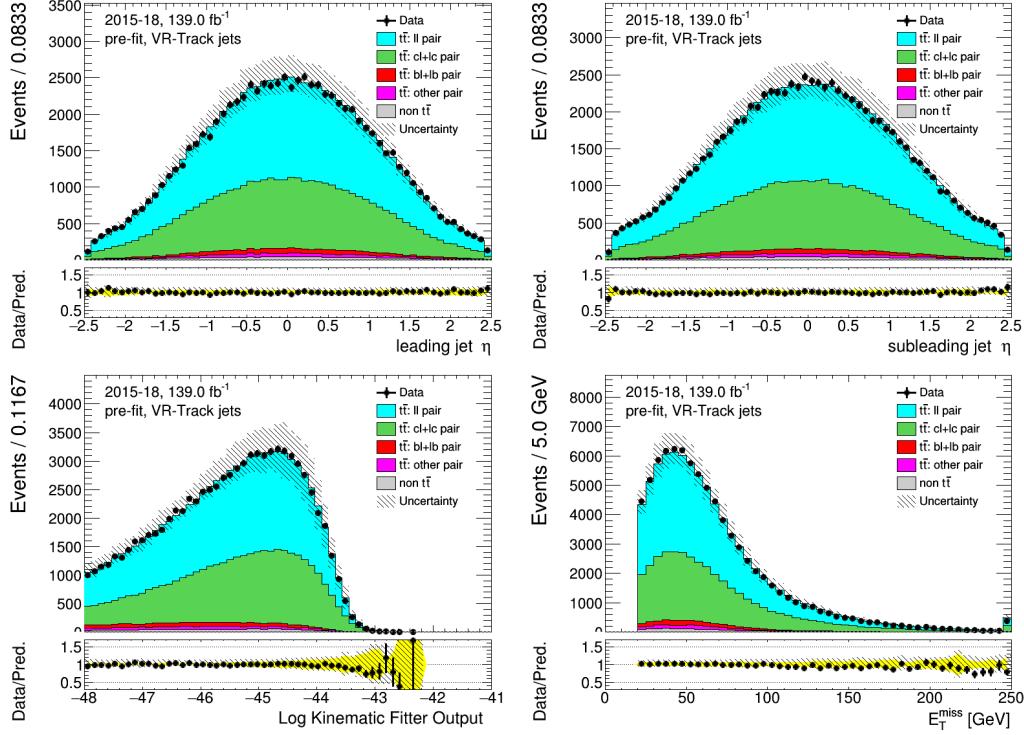


Figure A.13: VR-Track jets: distributions of the leading and sub-leading jets from W decay, KLFitter output and the transverse missing transverse energy of the high- $p_T$  selection, before fitting or tagging with full uncertainties.

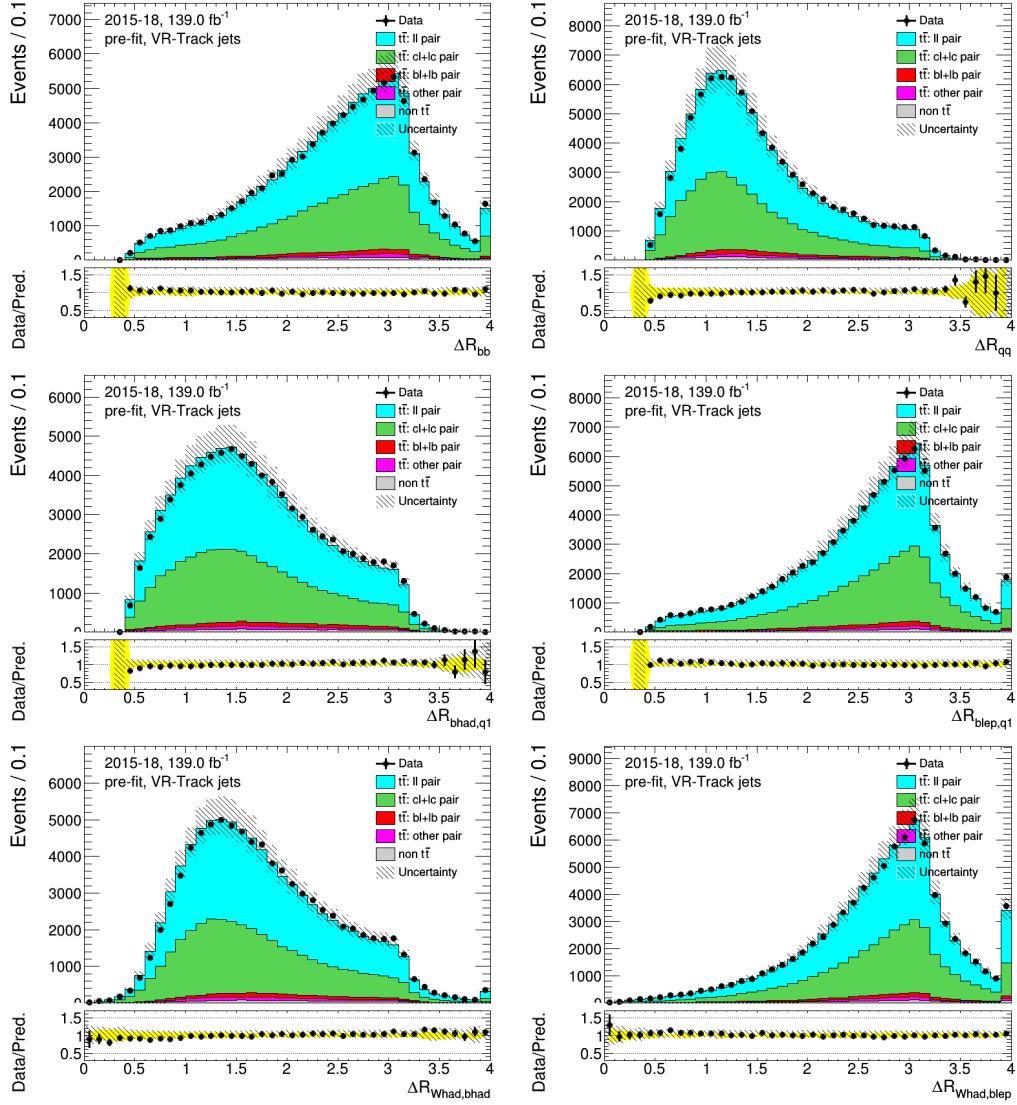


Figure A.14: VR-Track jets: distributions of angle related variables of the combination of the high- $p_T$  selection, before fitting or tagging with full uncertainties.

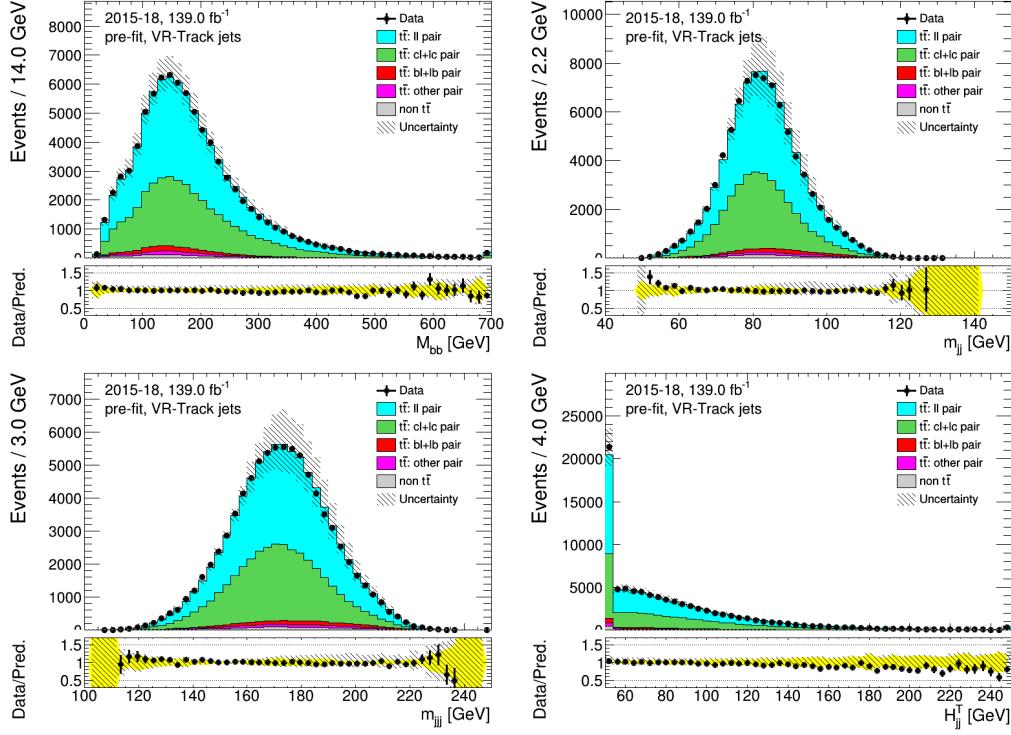


Figure A.15: VR-Track jets: distributions of mass related variables of the high- $p_T$  selection, before fitting or tagging with stat-only uncertainties.

### **A.3 Combined selection**

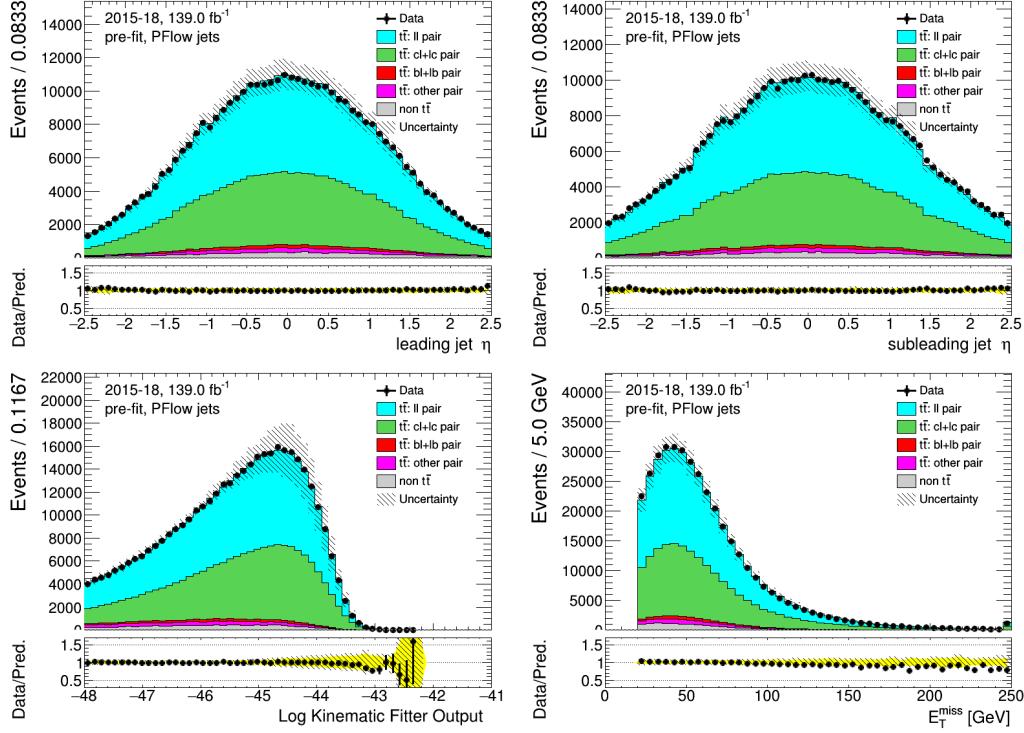


Figure A.16: PFlow jets: distributions of the leading and sub-leading jets from W decay, KLFitter output and the transverse missing transverse energy of the combined selection, before fitting or tagging with full uncertainties.

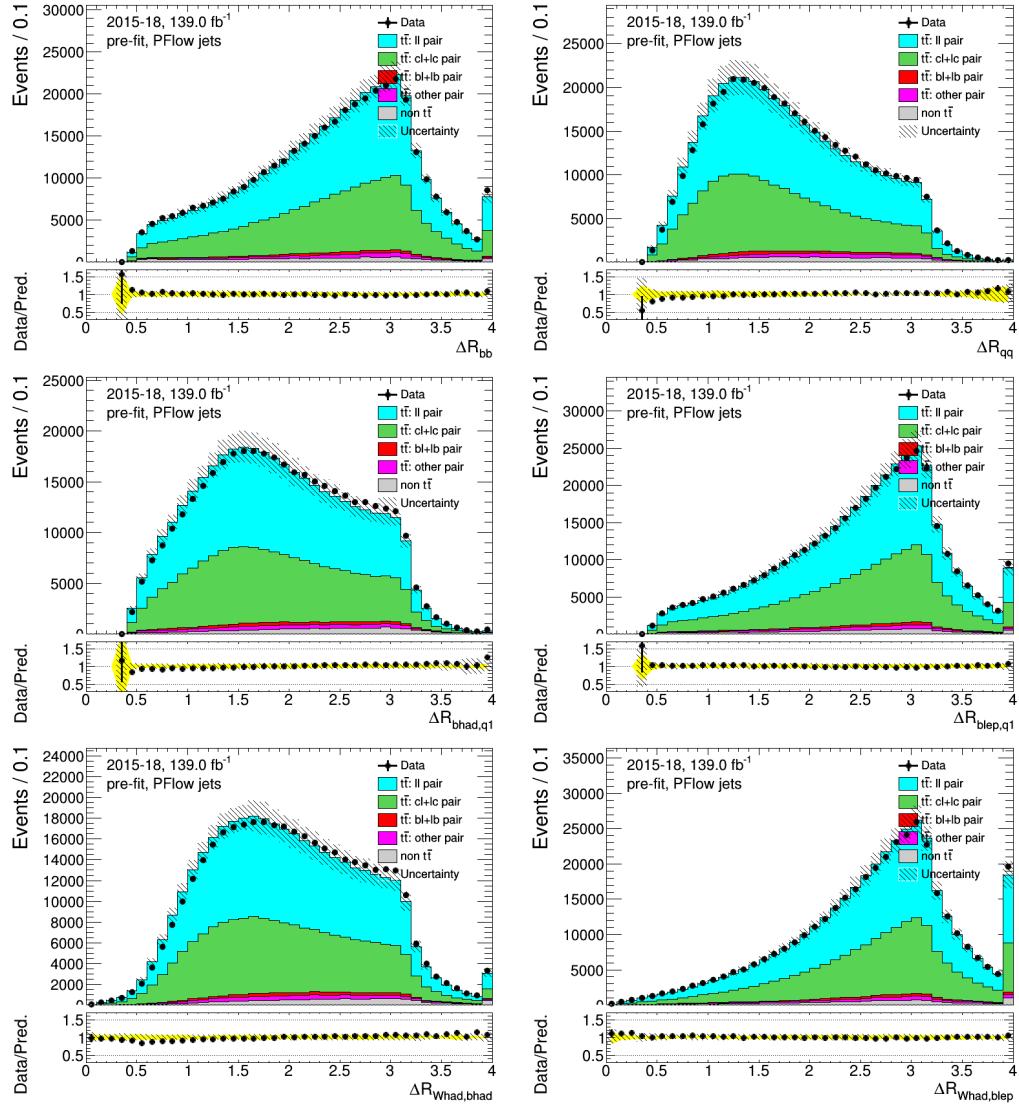


Figure A.17: PFlow jets: distributions of angle related variables of the combination of the combined selection, before fitting or tagging with full uncertainties.

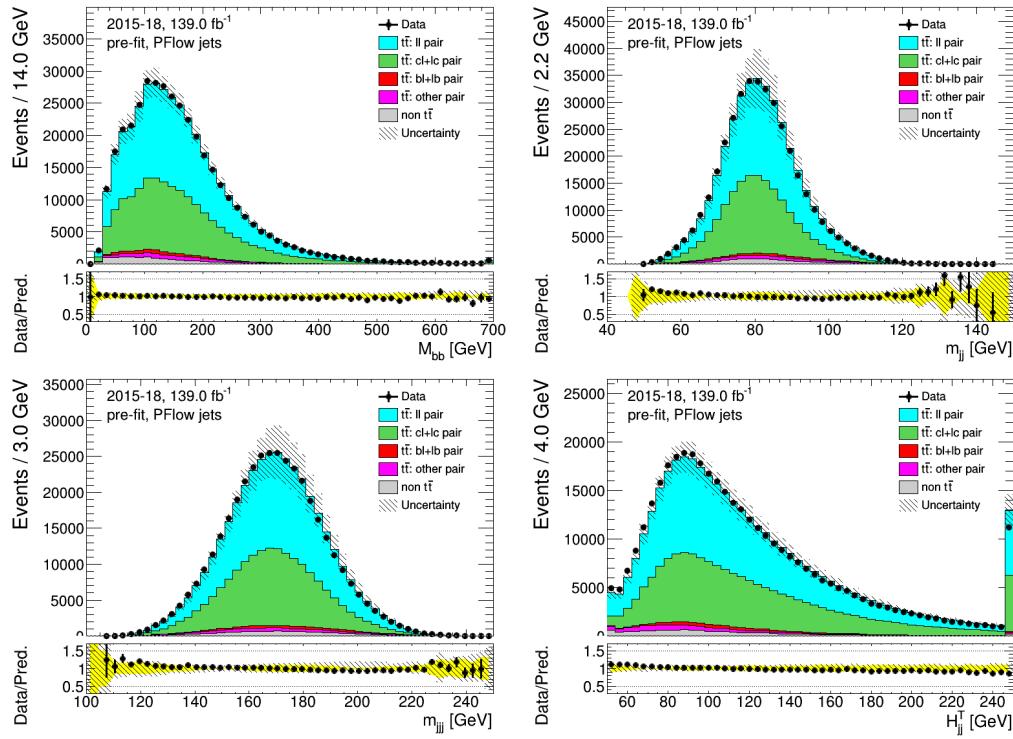


Figure A.18: PFflow jets: distributions of mass related variables of the combined selection, before fitting or tagging with stat-only uncertainties.

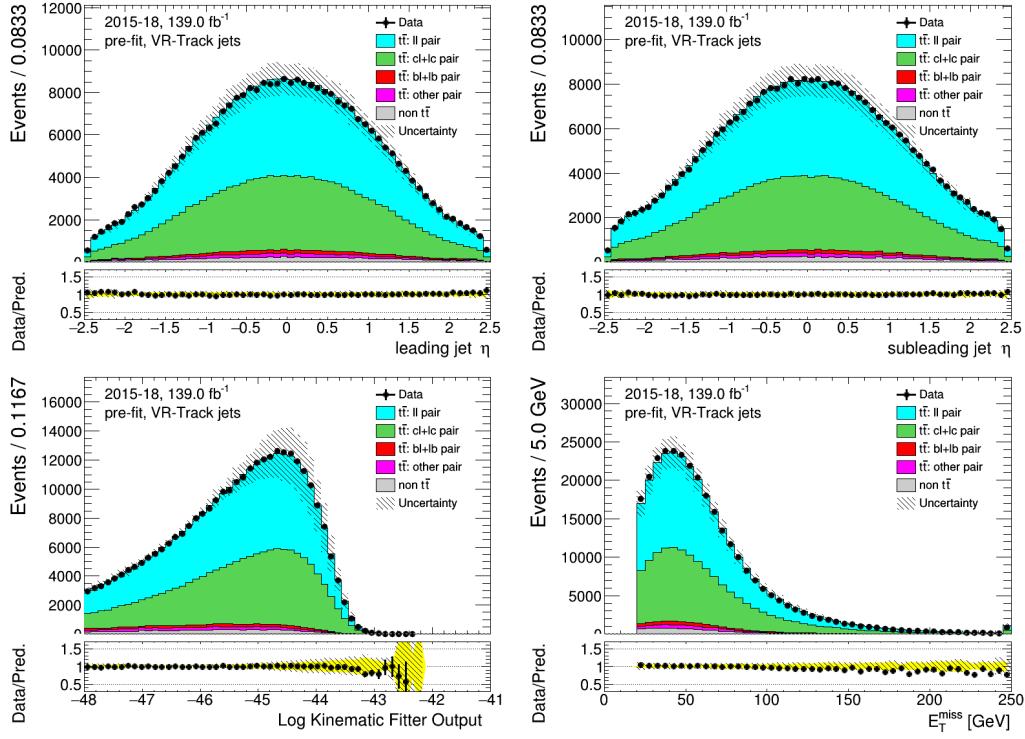


Figure A.19: VR-Track jets: distributions of the leading and sub-leading jets from W decay, KLFitter output and the transverse missing transverse energy of the combined selection, before fitting or tagging with full uncertainties.

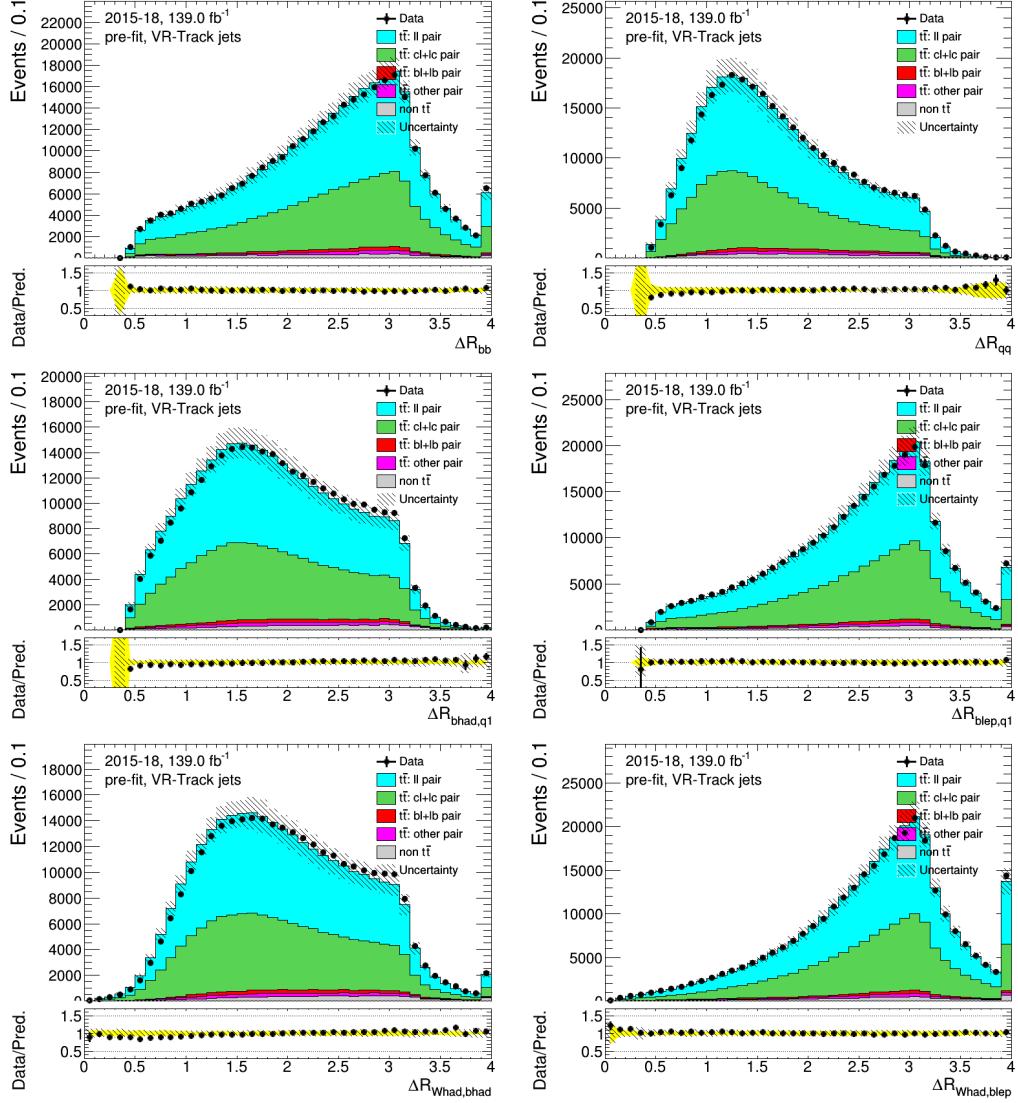


Figure A.20: VR-Track jets: distributions of angle related variables of the combination of the combined selection, before fitting or tagging with full uncertainties.

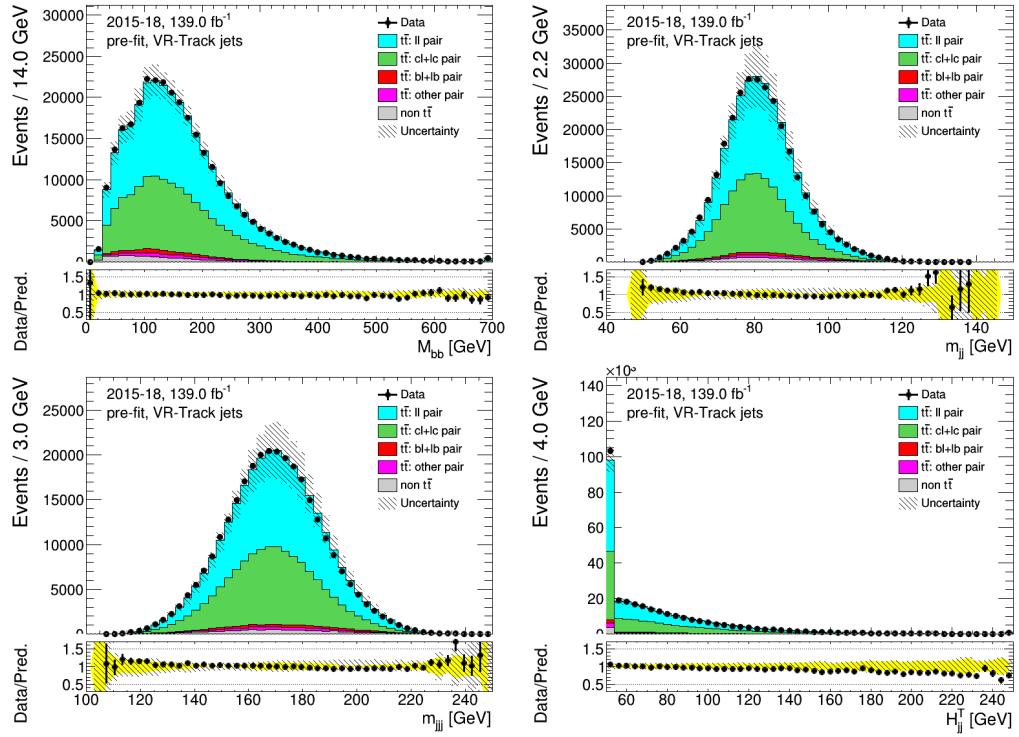


Figure A.21: VR-Track jets: distributions of mass related variables of the combined selection, before fitting or tagging with stat-only uncertainties.

## A.4 Experimental uncertainties

<b>Systematic uncertainty</b>
EG_RESOLUTION_ALL
MUON_ID
MUON_MS
MET_SoftTrk_ResoPara
MET_SoftTrk_ResoPerp
MET_SoftTrk_ScaleDown
MET_SoftTrk_ScaleUp
JET_Pileup_OffsetNPV
JET_Pileup_RhoTopology
JET_EffectiveNP_Modelling1
JET_EffectiveNP_Modelling2
JET_EffectiveNP_Modelling3
JET_EffectiveNP_Modelling4
JET_EffectiveNP_Statistical4
JET_EffectiveNP_Detector1
JET_JER_EffectiveNP_1
JET_JER_EffectiveNP_2
JET_JER_EffectiveNP_3
JET_JER_EffectiveNP_4
JET_BJES_Response
JET_Flavor_Composition
JET_Flavor_Response

Table A.1: List of experimental systematics.

## Appendix B

# Supplementary material for $HH$ searches

### B.1 Additional plots for fake-background estimation

### B.2 Additional material for MVA signal extraction

The following figures are additional material for the MVA section.

### B.3 Additional material for systematics uncertainties

In the following plots, the parametrisation of different MVA scores are shown. In Fig. B.5 (B.6), the shape only ratio of the PS (ME) variation sample to the nominal sample (both AF2) in the final fit binning is shown. This ratio is applied on the full sim nominal sample to mimic the effect from the systematic variations.

The resonant signal PS uncertainties are parametrised by linear functions. The function used in the SLT SR is:

- Down variation =  $1.82 - 0.95 * \text{PNN Score}$ .
- Up variation =  $0.18 + 0.95 * \text{PNN Score}$ ;

The function used in the LTT SR is:

- Down variation =  $1.84 - 1.01 * \text{PNN Score}$ ;
- Up variation =  $0.16 + 1.01 * \text{PNN Score}$ .

The data driven fakes estimated with  $t\bar{t}$  variation samples are shown in Fig.B.26 for the SLT channel and Fig.B.27 for the LTT channel. In Fig.B.28 and Fig.B.29, the y-axis

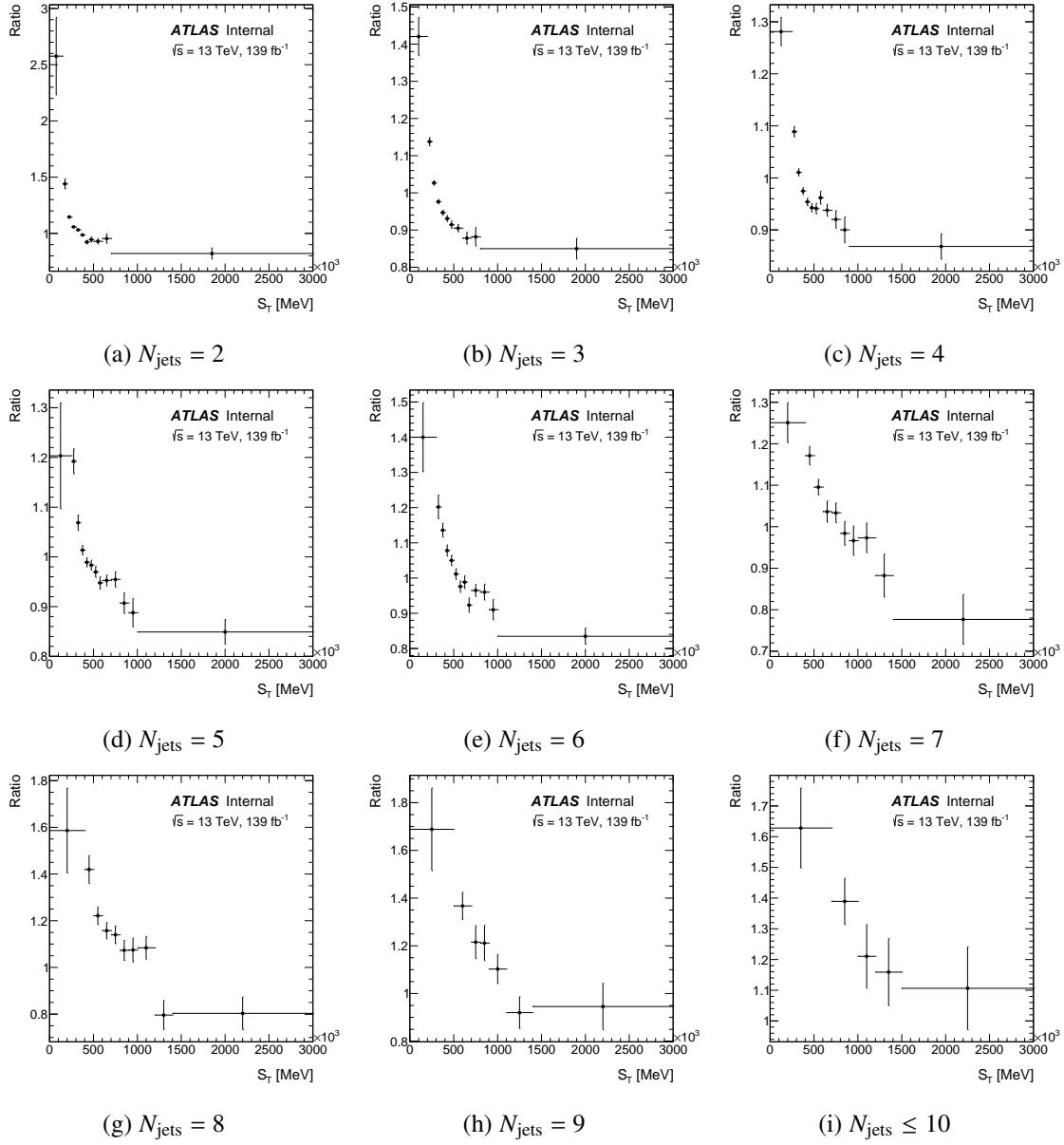


Figure B.1: The  $t\bar{t}$  shape correction scale factor as functions of  $H_T$  in different  $N_{\text{jets}}$ . The error bars are calculated from the statistical uncertainties of data and simulated samples. Figures reproduced from analysis internal notes.

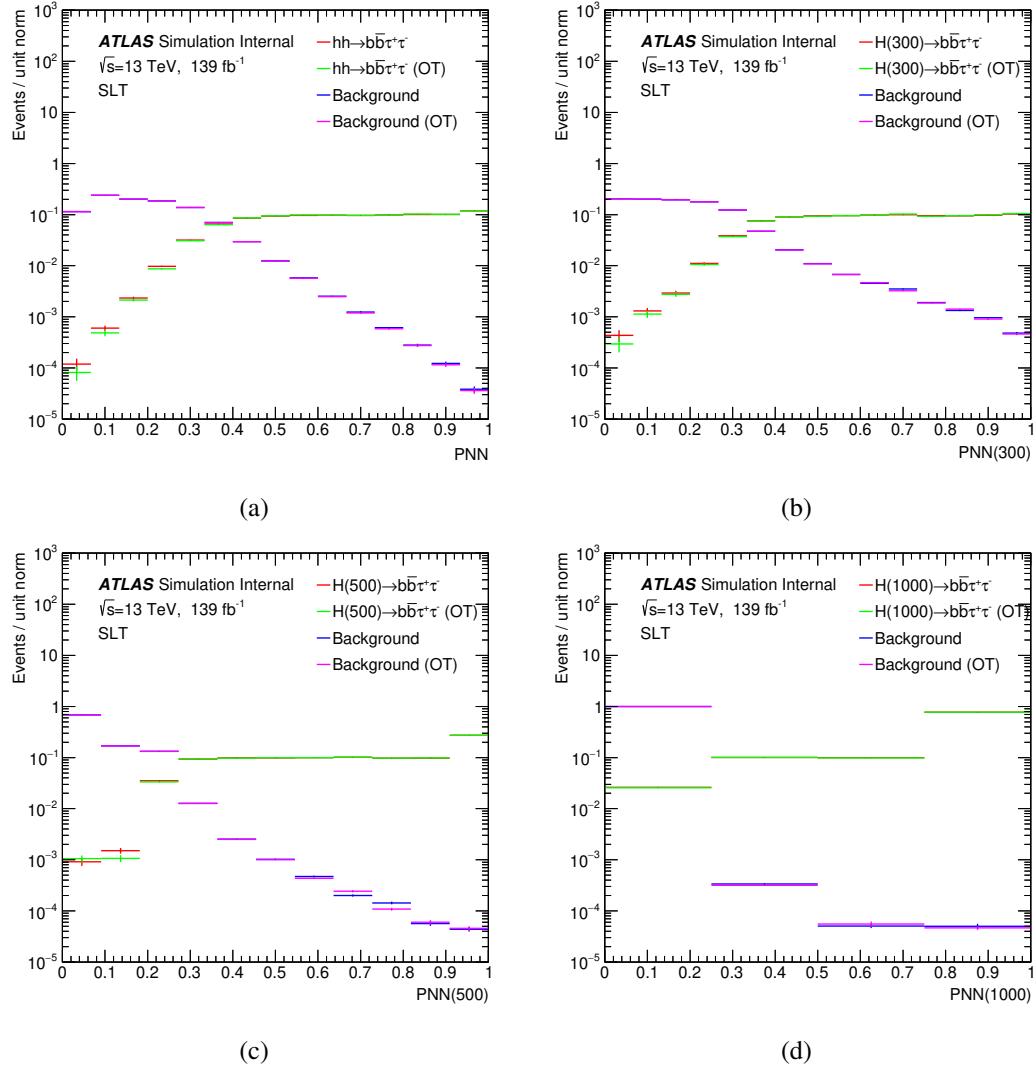


Figure B.2: (P)NN output distributions for (a) non-resonant signal, and target signal masses of (b) 300 GeV, (c) 500 GeV, and (d) 1000 GeV obtained evaluating the MVA with even-odd events crossing or evaluating the MVA on the training datasets for the SLT category. The “OT” histograms refer to the “OverTraining” check in which the (P)NN is applied to the same data on which it was trained, the other histograms are obtained evaluating the MVA with even-odd event crossing. Images reproced from analysis internal notes.

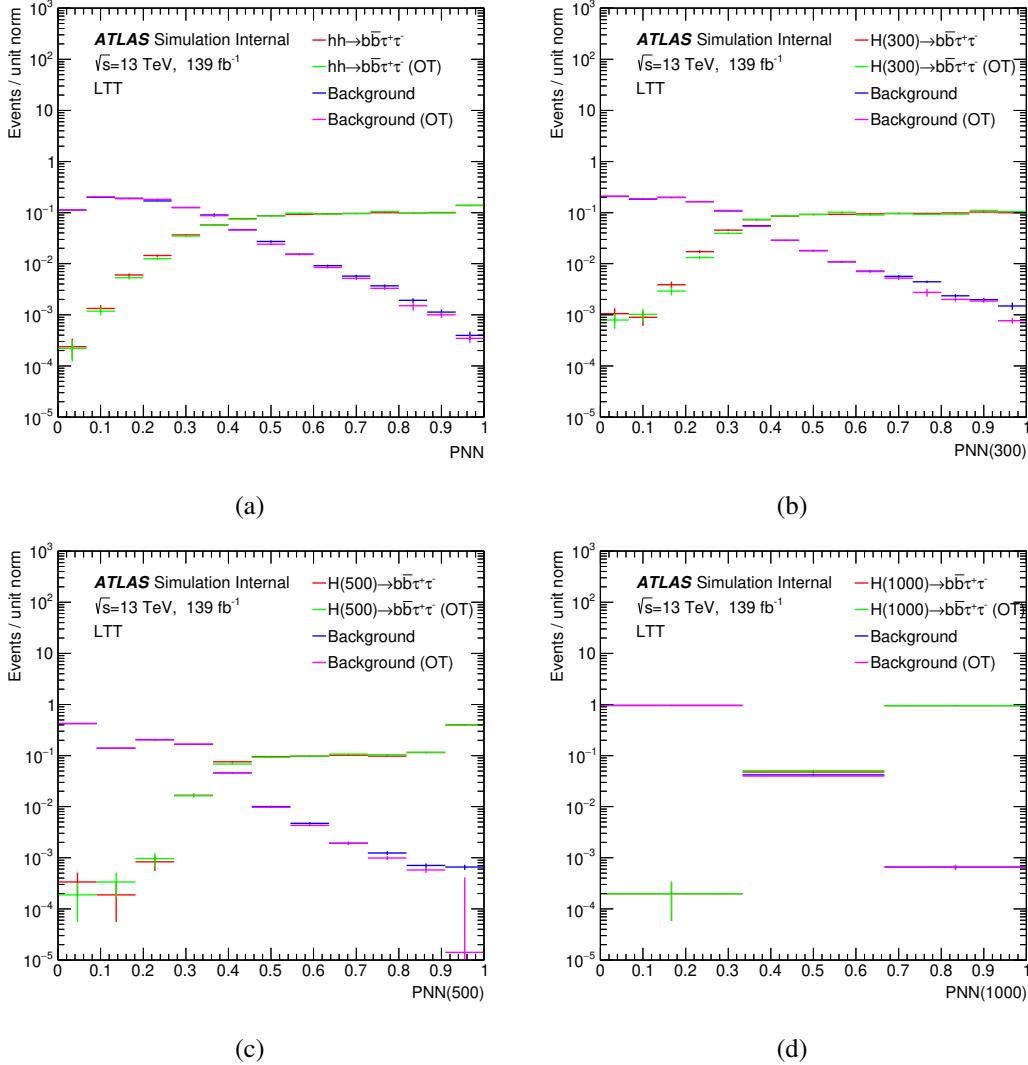


Figure B.3: (P)NN output distributions for (a) non-resonant signal, and target signal masses of (b)  $300\text{ GeV}$ , (c)  $500\text{ GeV}$ , and (d)  $1000\text{ GeV}$  obtained evaluating the MVA with even-odd events crossing or evaluating the MVA on the training datasets for the LTT category. The “OT” histograms refer to the “OverTraining” check in which the (P)NN is applied to the same data on which it was trained, the other histograms are obtained evaluating the MVA with even-odd event crossing. Images reprocod from analysis internal notes.

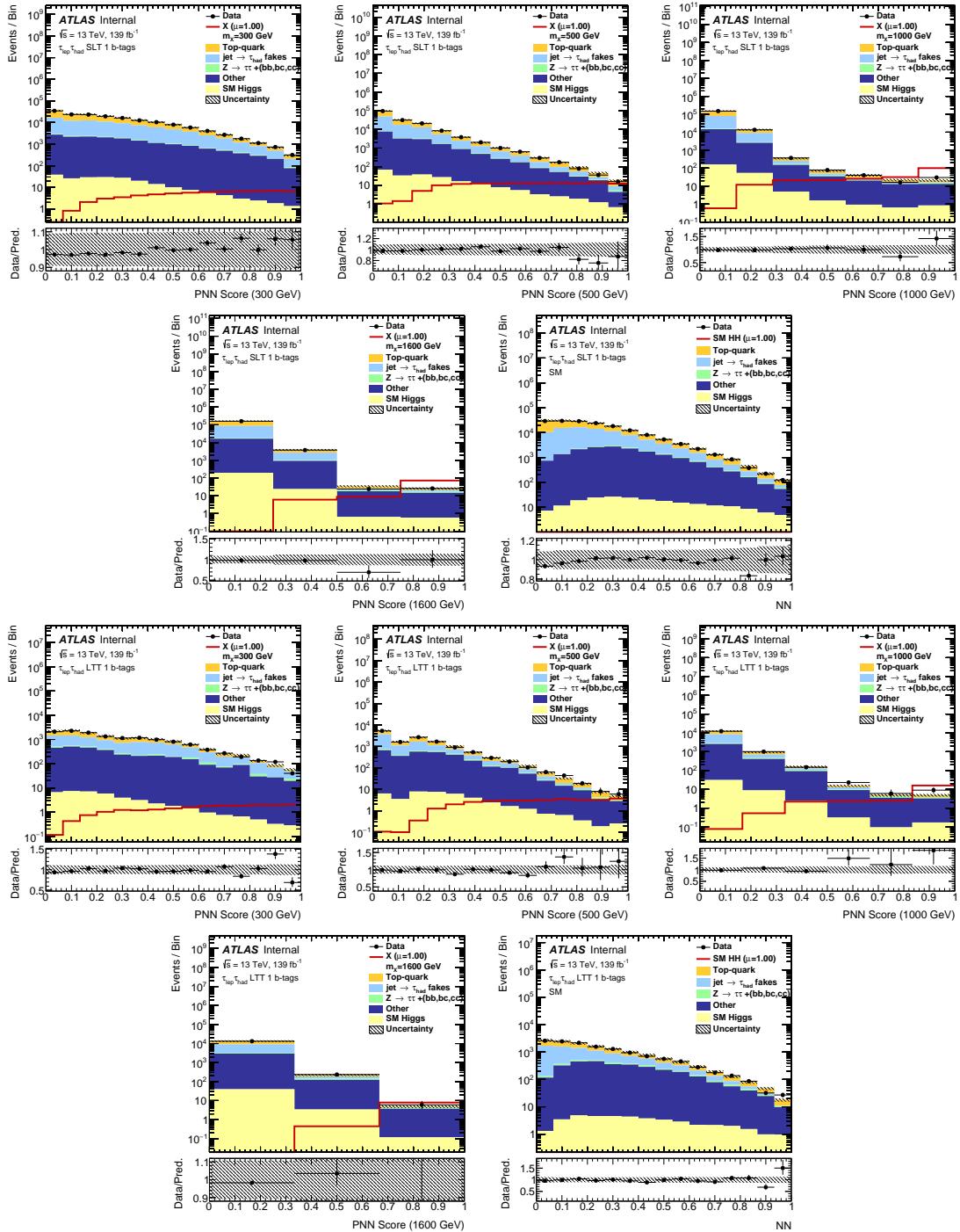


Figure B.4: Pre-fit PNN score distributions for the 300, 500, 1000, 1600 GeV mass points and non-resonant NN in the di-Higgs  $b\bar{b}\tau_{\text{lep}}\tau_{\text{had}}$  SLT (top two rows) and LTT (bottom two rows) 1tag control regions.

Process	Last bin			
	Last bin SM	Last bin 300 GeV	Last bin 500 GeV	Last bin 1000 GeV
ttbar	0.46 ± 0.22	30.9 ± 2	1.3 ± 0.4	0.68 ± 0.28
Fake	1.2 ± 0.7	16.7 ± 2.4	1.1 ± 0.6	0.8 ± 0.6
Stop	0.93 ± 0.35	3 ± 0.6	0.96 ± 0.35	0.9 ± 0.34
ZHF	1.32 ± 0.26	8.6 ± 3.2	1.36 ± 0.28	1.94 ± 0.25
ZLF	0.04 ± 0.025	-0.25 ± 0.25	0.029 ± 0.029	0.27 ± 0.18
WHbb	0.0005 ± 0.00035	0	0.0021 ± 0.0012	0.0033 ± 0.0013
WHTautau	0	0	0.006 ± 0.006	0
qqZHbb	0.278 ± 0.006	0.048 ± 0.008	0.078 ± 0.004	0.393 ± 0.007
ggZHbb	0.054 ± 0.004	0.0008 ± 0.0005	0.09 ± 0.05	0.0342 ± 0.0034
qqZHTautau	0.28 ± 0.04	0.059 ± 0.016	0.084 ± 0.023	0.158 ± 0.028
ggZHTautau	0.052 ± 0.014	0.00016 ± 0.00016	0.061 ± 0.015	0.032 ± 0.011
ggFHTautau	0.25 ± 0.05	0	0.111 ± 0.034	0.15 ± 0.04
VBFHTautau	0.0047 ± 0.0028	0.0025 ± 0.0019	0.0018 ± 0.0018	0.011 ± 0.004
ttH	0.221 ± 0.019	0.09 ± 0.012	0.166 ± 0.016	0.082 ± 0.012
Wjets	0	0	0	0
Diboson	0.2 ± 0.09	0.24 ± 0.09	0.15 ± 0.09	0.53 ± 0.13
DY	0	0	0	0
signal ggF	0.764 ± 0.007	8.03 ± 0.19	61.2 ± 0.8	342.5 ± 2.1
signal VBF	0.00743 ± 0.00024	NA	NA	NA
Second-to-last bin				
Process	Last-1 bin SM	Last-1 bin 300 GeV	Last-1 bin 500 GeV	Last-1 bin 1000 GeV
ttbar	4.5 ± 0.8	61.4 ± 2.9	3.1 ± 0.6	0.82 ± 0.34
Fake	1.9 ± 0.8	25.4 ± 3.2	0.8 ± 0.6	1.4 ± 0.8
Stop	2.4 ± 0.6	3 ± 0.6	1.3 ± 0.4	1.05 ± 0.34
ZHF	4 ± 0.6	13 ± 4	1.44 ± 0.33	2.64 ± 0.32
ZLF	0.3 ± 0.12	0.29 ± 0.27	0.11 ± 0.08	-0.03 ± 0.08
WHbb	0.0044 ± 0.0015	0	0.0026 ± 0.0015	0.0029 ± 0.001
WHTautau	0.013 ± 0.009	0	0	0
qqZHbb	0.448 ± 0.008	0.086 ± 0.01	0.075 ± 0.004	0.232 ± 0.005
ggZHbb	0.16 ± 0.05	0.0022 ± 0.0008	0.0331 ± 0.0034	0.06 ± 0.04
qqZHTautau	0.44 ± 0.05	0.07 ± 0.018	0.099 ± 0.021	0.14 ± 0.026
ggZHTautau	0.126 ± 0.022	0	0.059 ± 0.015	0.017 ± 0.008
ggFHTautau	0.25 ± 0.05	0.069 ± 0.029	0.093 ± 0.034	0.17 ± 0.04
VBFHTautau	0.012 ± 0.005	0.009 ± 0.004	0.0031 ± 0.0022	0.01 ± 0.004
ttH	0.27 ± 0.021	0.11 ± 0.012	0.128 ± 0.014	0.07 ± 0.01
Wjets	0	0	0	0
Diboson	0.56 ± 0.17	0.55 ± 0.2	0.12 ± 0.07	0.82 ± 0.18
DY	0	0	0	0
signal ggF	0.599 ± 0.006	7.53 ± 0.18	21 ± 0.5	42 ± 0.7
signal VBF	0.00869 ± 0.00027	NA	NA	NA
Third-to-last bin				
Process	Last-2 bin SM	Last-2 bin 300 GeV	Last-2 bin 500 GeV	Last-2 bin 1000 GeV
ttbar	9.1 ± 1.1	88 ± 3.4	5.2 ± 0.9	209 ± 5
Fake	4.7 ± 1.5	44 ± 4	3.6 ± 1.2	395 ± 13
Stop	5.3 ± 0.8	4.3 ± 0.7	1.2 ± 0.4	104 ± 4
ZHF	7.1 ± 0.7	11 ± 4	3.3 ± 0.7	212 ± 6
ZLF	0.54 ± 0.28	0.9 ± 0.7	0.04 ± 0.11	21.9 ± 2.9
WHbb	0.021 ± 0.004	0.013 ± 0.008	0.0053 ± 0.002	0.511 ± 0.018
WHTautau	0	0	0.005 ± 0.005	0.101 ± 0.029
qqZHbb	0.677 ± 0.011	0.123 ± 0.012	0.121 ± 0.006	3.355 ± 0.022
ggZHbb	0.24 ± 0.06	0.0024 ± 0.0009	0.043 ± 0.004	1.31 ± 0.11
qqZHTautau	0.45 ± 0.05	0.063 ± 0.018	0.159 ± 0.028	1.71 ± 0.09
ggZHTautau	0.161 ± 0.024	0	0.059 ± 0.015	0.57 ± 0.05
ggFHTautau	0.41 ± 0.07	0.089 ± 0.033	0.092 ± 0.031	4.23 ± 0.22
VBFHTautau	0.033 ± 0.008	0.009 ± 0.004	0.01 ± 0.004	0.402 ± 0.026
ttH	0.482 ± 0.028	0.159 ± 0.015	0.188 ± 0.017	4.45 ± 0.08
Wjets	0.08 ± 0.08	0.07 ± 0.07	0	6 ± 0.9
Diboson	1.03 ± 0.19	0.95 ± 0.22	0.26 ± 0.08	22.2 ± 1.5
DY	0	0	0	0.7 ± 0.4
signal ggF	0.596 ± 0.006	7.36 ± 0.18	184 20.6 ± 0.5	40.5 ± 0.7
signal VBF	0.00997 ± 0.00029	NA	NA	NA

Table B.1: Pre-fit event yields in the last, second-to-last and third-to-last MVA bin of the di-Higgs

## Chapter. B

---

Process	Last bin			
	Last bin SM	Last bin 300 GeV	Last bin 500 GeV	Last bin 1000 GeV
ttbar	0.76 ± 0.31	5 ± 0.8	2.3 ± 0.6	0.57 ± 0.29
Fake	2.6 ± 1.4	3.7 ± 1.7	1 ± 0.8	0.9 ± 0.9
Stop	0.32 ± 0.19	0.14 ± 0.14	0.17 ± 0.12	0.47 ± 0.25
ZHF	1.03 ± 0.23	3.3 ± 2.2	1.3 ± 0.4	2.1 ± 0.5
ZLF	0.07 ± 0.07	0	0.05 ± 0.09	0.52 ± 0.24
WHbb	0.0013 ± 0.0009	0	0	0.004 ± 0.004
WHTautau	0	0	0	0
qqZHbb	0.0656 ± 0.0032	0.0073 ± 0.0028	0.0416 ± 0.0027	0.0785 ± 0.0032
ggZHbb	0.0191 ± 0.0026	0.0004 ± 0.0004	0.0182 ± 0.0024	0.0063 ± 0.0014
qqZHTautau	0.065 ± 0.018	0.019 ± 0.01	0.051 ± 0.016	0.052 ± 0.018
ggZHTautau	0.02 ± 0.009	0	0.022 ± 0.009	0
ggFHTautau	0.076 ± 0.034	0	0.076 ± 0.03	0.11 ± 0.04
VBFHTautau	0.005 ± 0.0029	0	0.0018 ± 0.0018	0.0035 ± 0.0025
ttH	0.044 ± 0.008	0.014 ± 0.005	0.065 ± 0.01	0.031 ± 0.007
Wjets	0	0	0	0
Diboson	-0.01 ± 0.05	-0.05 ± 0.04	0.03 ± 0.04	0.44 ± 0.14
DY	0	0	0	0
signal ggF	0.1977 ± 0.0035	2.61 ± 0.11	20.8 ± 0.5	45.1 ± 0.8
signal VBF	0.00344 ± 0.00017	NA	NA	NA
Second-to-last bin				
Process	Last-1 bin SM	Last-1 bin 300 GeV	Last-1 bin 500 GeV	Last-1 bin 1000 GeV
ttbar	3.8 ± 0.7	8.5 ± 1.1	2.5 ± 0.6	108 ± 4
Fake	2 ± 1.4	1.4 ± 1.1	1.7 ± 1.2	84 ± 9
Stop	0.37 ± 0.21	0.45 ± 0.24	0.08 ± 0.08	18 ± 1.7
ZHF	1.6 ± 0.4	2.5 ± 1.1	0.82 ± 0.19	86.4 ± 3.4
ZLF	0.15 ± 0.1	0.07 ± 0.07	0	11 ± 4
WHbb	0	0.0027 ± 0.0027	0.0016 ± 0.0011	0.0074 ± 0.002
WHTautau	0	0	0	0.047 ± 0.021
qqZHbb	0.083 ± 0.004	0.022 ± 0.005	0.0384 ± 0.0033	0.826 ± 0.013
ggZHbb	0.0262 ± 0.0029	0.00022 ± 0.00016	0.0154 ± 0.0023	0.44 ± 0.06
qqZHTautau	0.088 ± 0.021	0.011 ± 0.008	0.06 ± 0.018	0.57 ± 0.05
ggZHTautau	0.027 ± 0.01	0	0.028 ± 0.01	0.212 ± 0.028
ggFHTautau	0.094 ± 0.032	0.015 ± 0.011	0.022 ± 0.015	1.48 ± 0.14
VBFHTautau	0.005 ± 0.0029	0.0025 ± 0.0024	0.0043 ± 0.0027	0.12 ± 0.014
ttH	0.062 ± 0.01	0.02 ± 0.005	0.046 ± 0.009	1.45 ± 0.05
Wjets	0	0	0	0.83 ± 0.31
Diboson	0.14 ± 0.08	0.1 ± 0.09	0 ± 0.04	3.7 ± 0.4
DY	0	0	0.15 ± 0.15	0.15 ± 0.15
signal ggF	0.1402 ± 0.0028	2.72 ± 0.11	5.87 ± 0.25	2.28 ± 0.17
signal VBF	0.00294 ± 0.00016	NA	NA	NA
Third-to-last bin				
Process	Last-2 bin SM	Last-2 bin 300 GeV	Last-2 bin 500 GeV	Last-2 bin 1000 GeV
ttbar	6.3 ± 0.9	9.7 ± 1.2	3.4 ± 0.7	4104 ± 24
Fake	1.4 ± 1.3	6.6 ± 2.4	2.4 ± 1.3	2060 ± 40
Stop	0.85 ± 0.34	0.3 ± 0.21	0.47 ± 0.25	179 ± 5
ZHF	4.8 ± 0.9	2.6 ± 1.1	2.1 ± 0.4	368 ± 14
ZLF	0.18 ± 0.09	0.25 ± 0.2	0.05 ± 0.04	37 ± 9
WHbb	0.0012 ± 0.0008	0	0.001 ± 0.0007	0.162 ± 0.026
WHTautau	0	0.006 ± 0.006	0	0.131 ± 0.033
qqZHbb	0.124 ± 0.006	0.041 ± 0.007	0.0457 ± 0.0032	2.95 ± 0.05
ggZHbb	0.0323 ± 0.0032	0.0013 ± 0.0006	0.0221 ± 0.0026	1.01 ± 0.13
qqZHTautau	0.129 ± 0.026	0.015 ± 0.009	0.052 ± 0.016	1.56 ± 0.09
ggZHTautau	0.039 ± 0.012	0.004 ± 0.004	0.028 ± 0.01	0.53 ± 0.04
ggFHTautau	0.098 ± 0.029	0.005 ± 0.005	0.085 ± 0.035	2.98 ± 0.2
VBFHTautau	0.014 ± 0.005	0.0019 ± 0.0019	0.0033 ± 0.0023	0.248 ± 0.02
ttH	0.111 ± 0.013	0.024 ± 0.006	0.043 ± 0.008	9.52 ± 0.11
Wjets	0	0	0	3.2 ± 1
Diboson	0.18 ± 0.08	0 ± 0.08	0.09 ± 0.07	16.3 ± 1
DY	0	0	0	0.68 ± 0.29
signal ggF	0.1382 ± 0.0028	18.51 ± 0.11	4.93 ± 0.23	0.009 ± 0.009
signal VBF	0.00331 ± 0.00017	NA	NA	NA

Table B.2: Pre-fit event yields in the last, second-to-last and third-to-last MVA bin of the di-Higgs

Process	Name	LepHad SLT	LepHad LTT	Comment
tbar	THEO_ACC_TTBAR_ME	N: +0.0026, -0.0026, S	N: -0.009, +0.009	Matrix element acceptance
tbar	THEO_ACC_TTBAR_PS	N: -0.072, +0.072, S	N: -0.088, +0.088	Parton shower acceptance
tbar	THEO_ACC_TTBAR_ISR	N: +0.0005, -0.0081	N: -0.0052, +0.013	ISR acceptance
tbar	THEO_ACC_TTBAR_FSR	N: +0.014, -0.0097	N: +0.0096, -0.032	FSR acceptance
tbar	THEO_ACC_TTBAR_PDFalphas	N: -0.006, +0.006	N: -0.0011, +0.0011	PDF+ $\alpha_s$ acceptance
Z+hf	THEO_ACC_Zhf_GENERATOR	N: +0.021, -0.021	N: +0.10, -0.10	Matrix element acceptance
Z+hf	THEO_ACC_Zhf_SCALE	N: -0.029, +0.053, S	N: -0.054, +0.085, S	Scale acceptance
Z+hf	THEO_ACC_Zhf_CKKW	N: +0.07, -0.07	N: +0.071, -0.071	CKKW acceptance
Z+hf	THEO_ACC_Zhf_QSF	N: -0.016, +0.016	N: -0.016, +0.016	QSF acceptance
Z+hf	THEO_ACC_Zhf_PDFalphas	N: -0.0026, +0.0026	N: -0.0033, +0.0033	PDF+ $\alpha_s$ acceptance
Z+hf	THEO_ACC_Zhf_PDFchoice	N: -0.0097, +0.0097	N: -0.011, +0.011	PDF choice acceptance
stopWt	THEO_XS_Stop	N: -0.054, +0.054	N: -0.054, +0.054	cross section
stopWt	THEO_ACC_StopWt_ME	N: -0.022, +0.022	N: -0.15, 0.15	Matrix element acceptance
stopWt	THEO_ACC_StopWt_PS	N: +0.077, -0.077	N: -0.093, +0.093	Parton shower acceptance
stopWt	THEO_ACC_StopWt_ISR	N: -0.047, +0.064	N: -0.045, +0.062	ISR acceptance
stopWt	THEO_ACC_StopWt_FSR	N: -0.054, +0.043	N: -0.069, +0.035	FSR acceptance
stopWt	THEO_ACC_StopWt_PDF	N: -0.032, +0.032	N: -0.032, +0.032	PDF acceptance
stopWt	THEO_ACC_StopWt_TopInterference	N: +0.078, -0.078, S	N: +0.11, +0.11, S	top interference acceptance
tH	THEO_XS SCALE_tH	N: -0.092, +0.058	N: -0.092, +0.058	Scale cross section
tH	THEO_XS_PDFalphas_tH	N: -0.036, +0.036	N: -0.036, +0.036	PDF+ $\alpha_s$ cross section
tH	THEO_ACC_GEN_tH	-	N: -0.019, +0.019	Matrix element acceptance
tH	THEO_ACC_PS_tH	N: -0.013, +0.013	N: -0.067, +0.067	Parton shower acceptance
tH	THEO_ACC SCALE_tH	-	N: -0.01, +0.01	Scale acceptance
tH	THEO_ACC_ISR_tH	N: -0.051, +0.032,	N: -0.15, +0.055	ISR acceptance
tH	THEO_ACC_FSR_tH	N: -0.039, +0.039	N: -0.039, +0.039	FSR acceptance
ggFHTautau	THEO_XS_SCALE_ggFH	N: -0.032, +0.032	N: -0.032, +0.032	Scale cross section
ggFHTautau	THEO_XS_PDFalphas_ggFH	N: -0.02, +0.02	N: -0.02, +0.02	PDF+ $\alpha_s$ section
ggF, ZH, WH, VBFH with Htautau	THEO_BR_Htautau	N: -1.0, +1.0	N: -1.0, +1.0	Htautau BR
ggFHTautau	THEO_ACC_HF_ggFH	N: -0.006, +0.005,	N: -0.006, +0.005	Higgs + HF mod unc
qqZHbb, qqZHTautau	THEO_XS_SCALE_qqZH	N: -0.019, +0.019	N: -0.019, +0.019	Scale cross section
qqZHbb, qqZHTautau	THEO_XS_PDFalphas_qqZH	N: -0.19, +0.25,	N: -0.19, +0.25	PDF+ $\alpha_s$ cross section
ggZHbb, ggZHTautau	THEO_XS_SCALE_ggZH	N: -0.024, +0.024	N: -0.024, +0.024	Scale cross section
ggZHbb, ggZHTautau	THEO_XS_PDFalphas_ggZH	N: -0.013, +0.013	N: -0.013, +0.013	PDF+ $\alpha_s$ cross section
ZHbb, WHbb	THEO_BR_Hbb	N: -0.11, +0.11	N: -0.037, +0.037	Hbb BR
ZHbb	THEO_ACC_PS_ZHbb	N: -0.030, +0.030	N: -0.025, +0.025	Parton shower acceptance
ZHtautau	THEO_ACC_SCALE_ZHtautau	N: -0.055, +0.055	N: -0.15, +0.15	Scale acceptance
ZHtautau	THEO_ACC_PDFalphas_ZHtautau	-	N: -0.012, +0.012	PDF+ $\alpha_s$ acceptance
WHbb, WHtautau	THEO_ACC_SCALE_ZHTautau	N: -0.022, +0.022	N: -0.028, +0.028	Scale acceptance
WHbb, WHtautau	THEO_XS_PDFalphan_WH	N: -0.007, +0.005	N: -0.007, +0.005	Scale cross section
WHtautau	THEO_XS_SCALE_WH	N: -0.019, +0.019	N: -0.019, +0.019	PDF+ $\alpha_s$ cross section
VBFHTautau	THEO_ACC_HF_WH	N: -1.0, +1.0	N: -1.0, +1.0	Higgs + HF mod unc
VBFHTautau	THEO_XS_SCALE_VBFH	N: -0.003, +0.004	N: -0.003, +0.004	Scale cross section
VBFHTautau	THEO_XS_PDFalphan_VBFH	N: -0.021, +0.021	N: -0.021, +0.021	PDF+ $\alpha_s$ cross section
VBFHTautau	THEO_ACC_HF_VBFH	N: -1.0, +1.0	N: -1.0, +1.0	Higgs + HF mod unc

Table B.3: List of MC background uncertainties for major backgrounds. The table shows the process, the name of the uncertainty, the relative size of the normalisation uncertainty (N) and whether the uncertainty includes also a shape variation (S) in the different SRs.

Process	Name	Size	Comment
V+jets	THEO_XS_V	-0.05, +0.05	cross section
Z+lf	THEO_ACC_Zlf	-0.23, +0.23	Acceptance from VHbb
W+jets	THEO_ACC_W	-0.37, +0.37	Acceptance from VHbb in LepHad SRs
W+jets	THEO_ACC_W	-0.50, +0.50	Acceptance inflated from VHbb analysis for tau fakes in HadHad SR
WW, WZ, ZZ	THEO_XS_Diboson	-0.06, +0.06	cross section
WW	THEO_ACC_Diboson	-0.25, +0.25	Acceptance from VHbb
WZ	THEO_ACC_Diboson	-0.26, +0.26	Acceptance from VHbb
ZZ	THEO_ACC_Diboson	-0.20, +0.20	Acceptance from VHbb

Table B.4: List of MC background uncertainties for minor backgrounds. The table shows the process, the name of the uncertainty, the relative size of the normalisation uncertainty (N). Table reproduced from analysis internal notes.

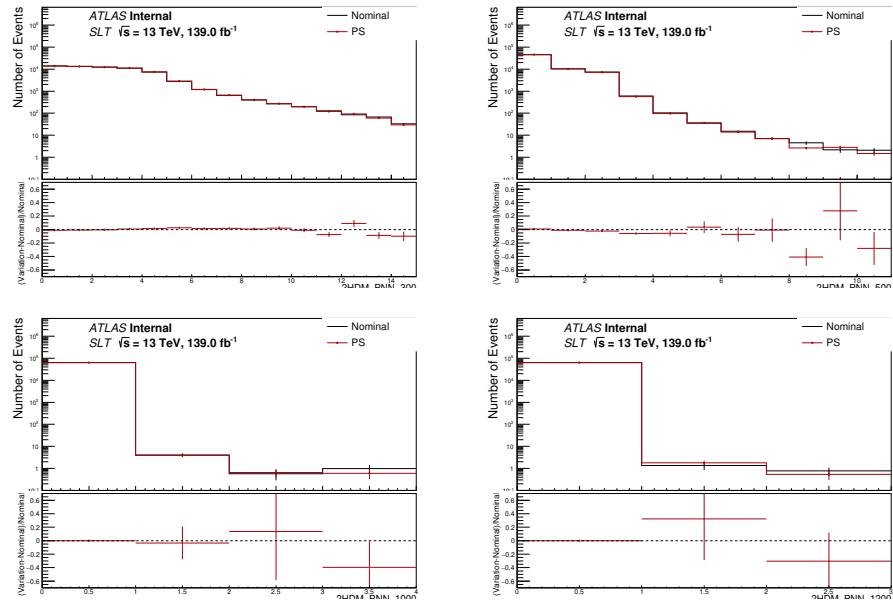


Figure B.5: Shape-only ratio of the PS variation to the nominal in PNN score of various resonant mass.

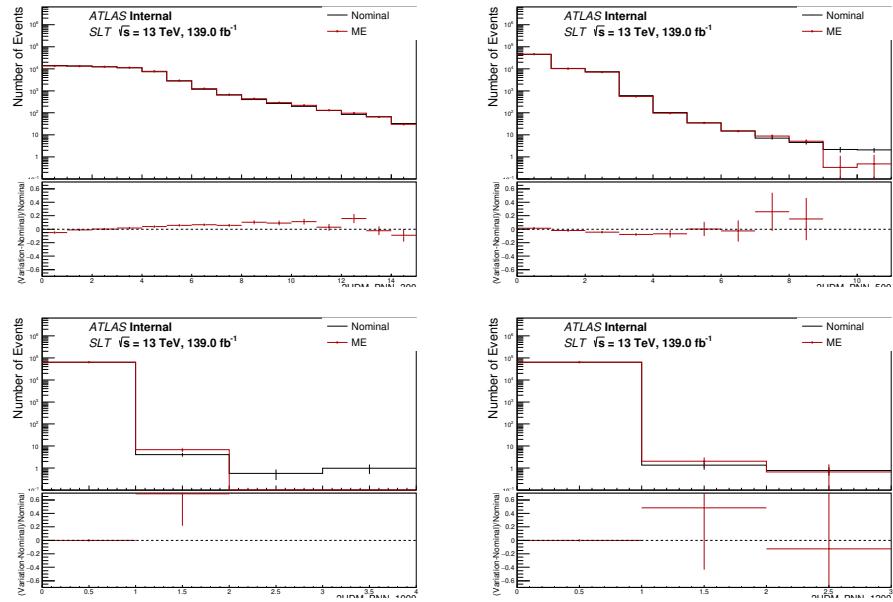


Figure B.6: Shape-only ratio of the ME variation to the nominal in PNN score of various resonant mass.

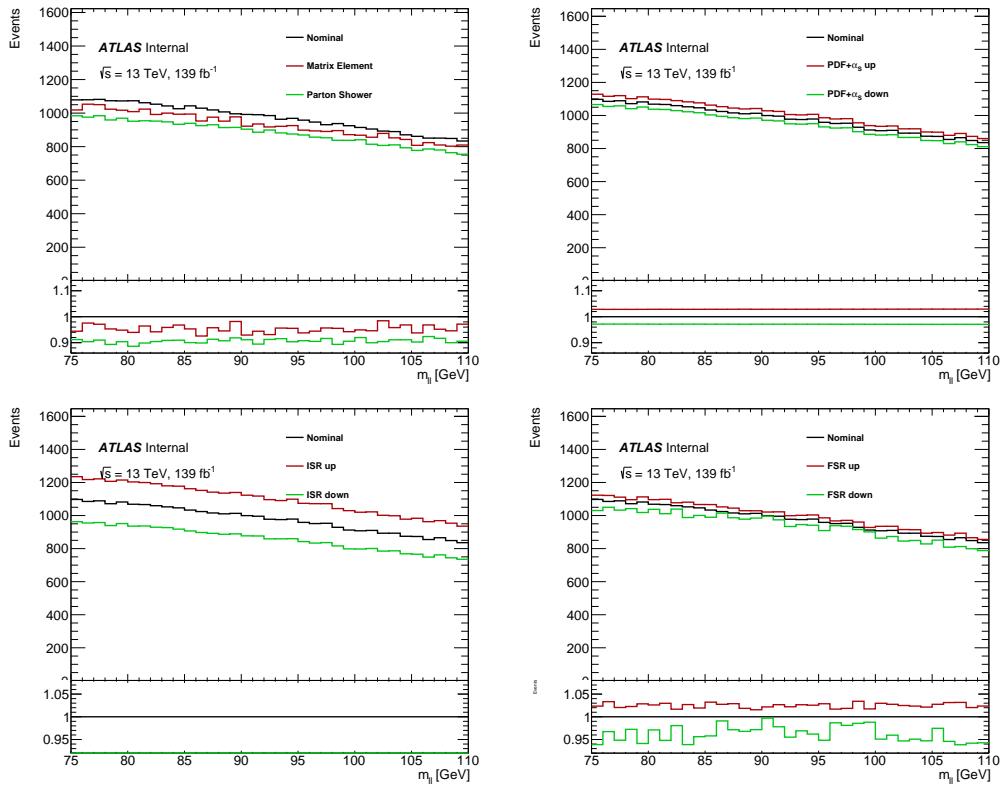


Figure B.7:  $m_{ll}$  spectra comparisons for nominal and systematics from matrix element and parton shower (top left), PDF+ $\alpha_S$  (top right), ISR (bottom left) and FSR (bottom right). Images reprodced from analysis internal notes.

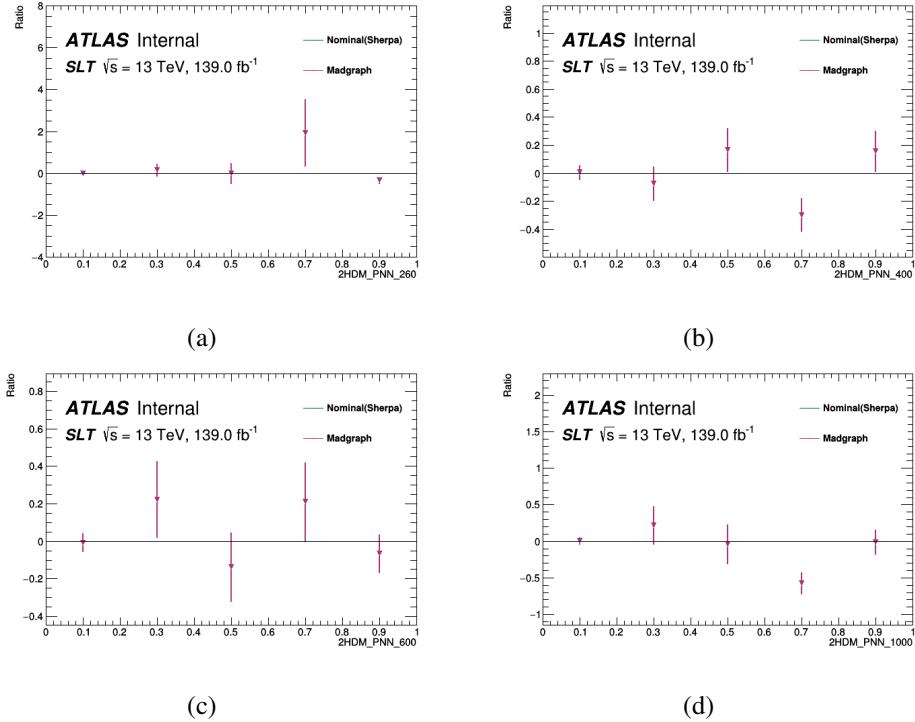


Figure B.8: SLT channel: comparision of the nominal Sherpa sample verse the variation Madgraph in the PNN score distribtuion.

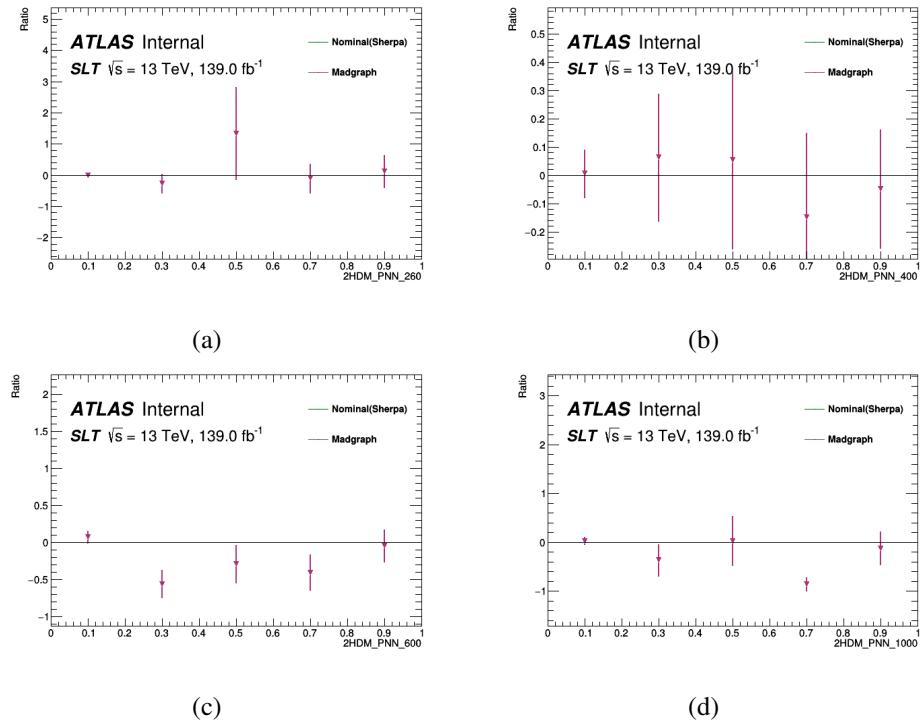


Figure B.9: LTT channel: comparision of the nominal Sherpa sample verse the variation Madgraph in the PNN score distribtuion.

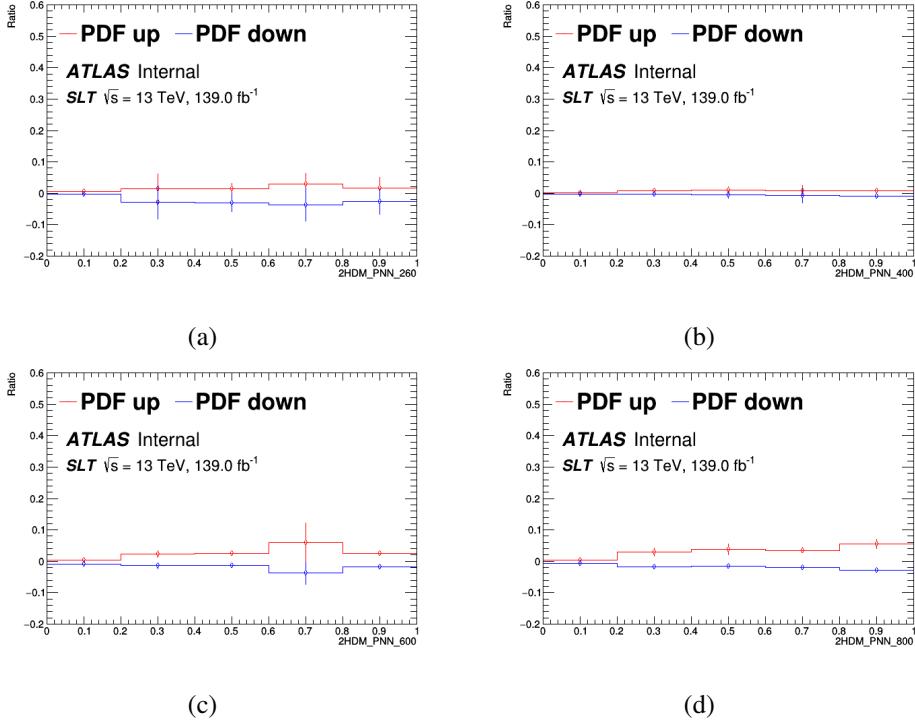


Figure B.10: SLT channel: Z+HF intra-PDF variation in PNN score of various resonant mass.

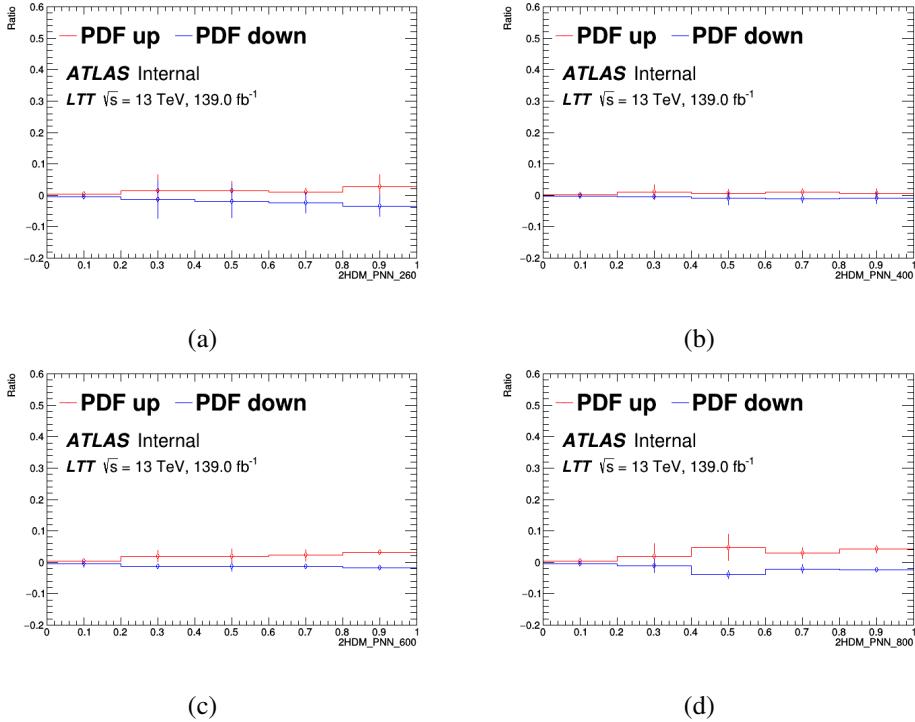


Figure B.11: LTT channel: Z+HF intra-PDF variation in the PNN score of various resonant masses.

## Chapter. B

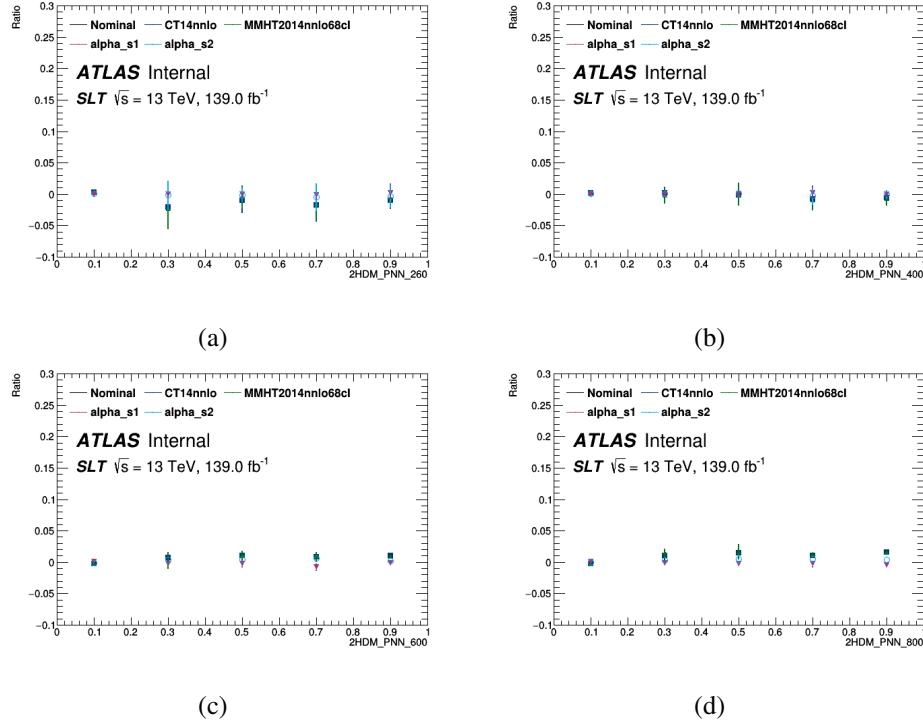


Figure B.12: SLT channel: Z+HF inter-PDF variations and  $\alpha_s$  variations in PNN score of various resonant mass.

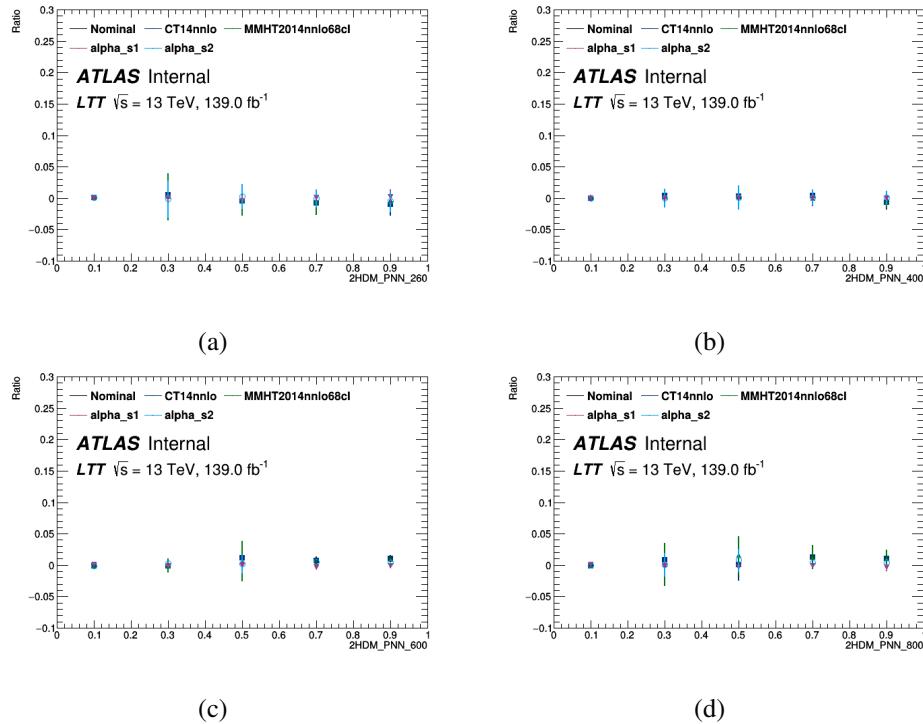


Figure B.13: LTT channel: Z+HF inter-PDF variations and  $\alpha_s$  variations in the PNN score of various resonant masses.

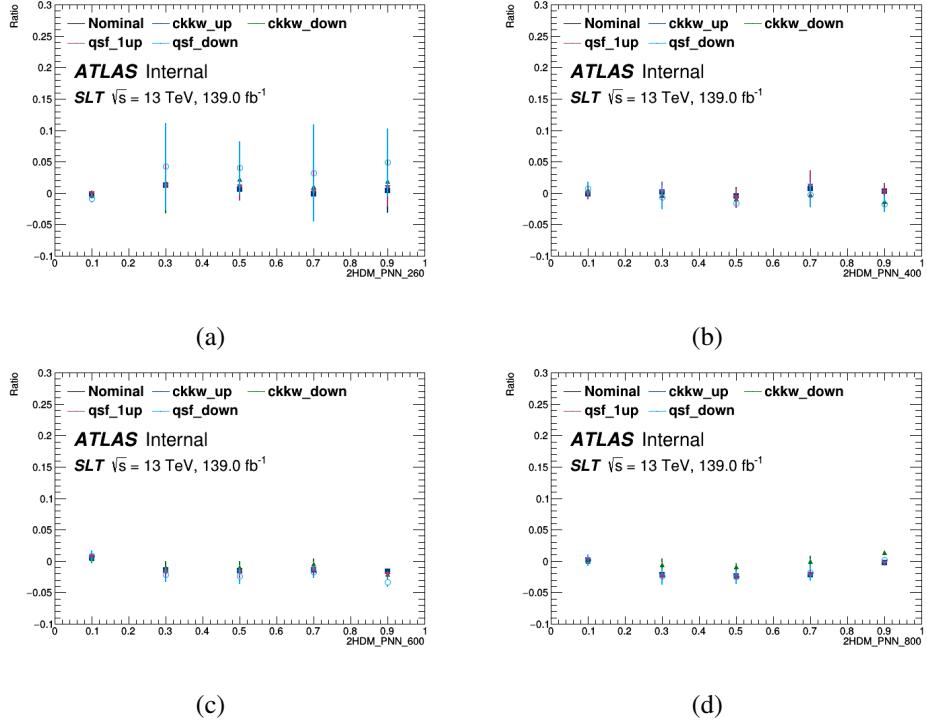


Figure B.14: SLT channel: Z+HF ckkw and qsf variations in PNN score of various resonant mass.

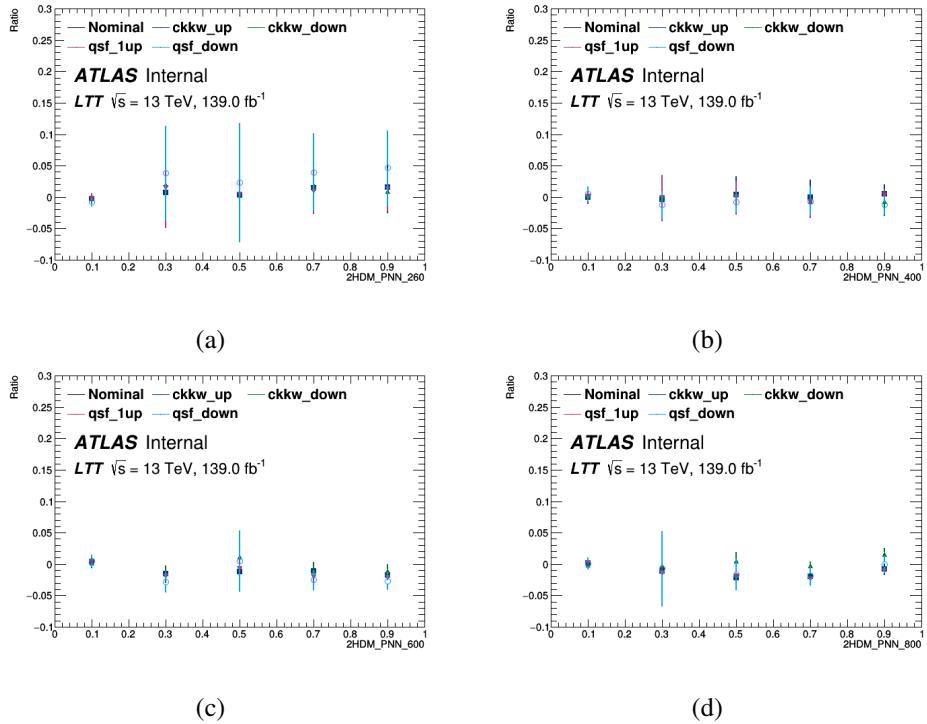


Figure B.15: LTT channel:Z+HF ckkw and qsf variations in the PNN score of various resonant masses.

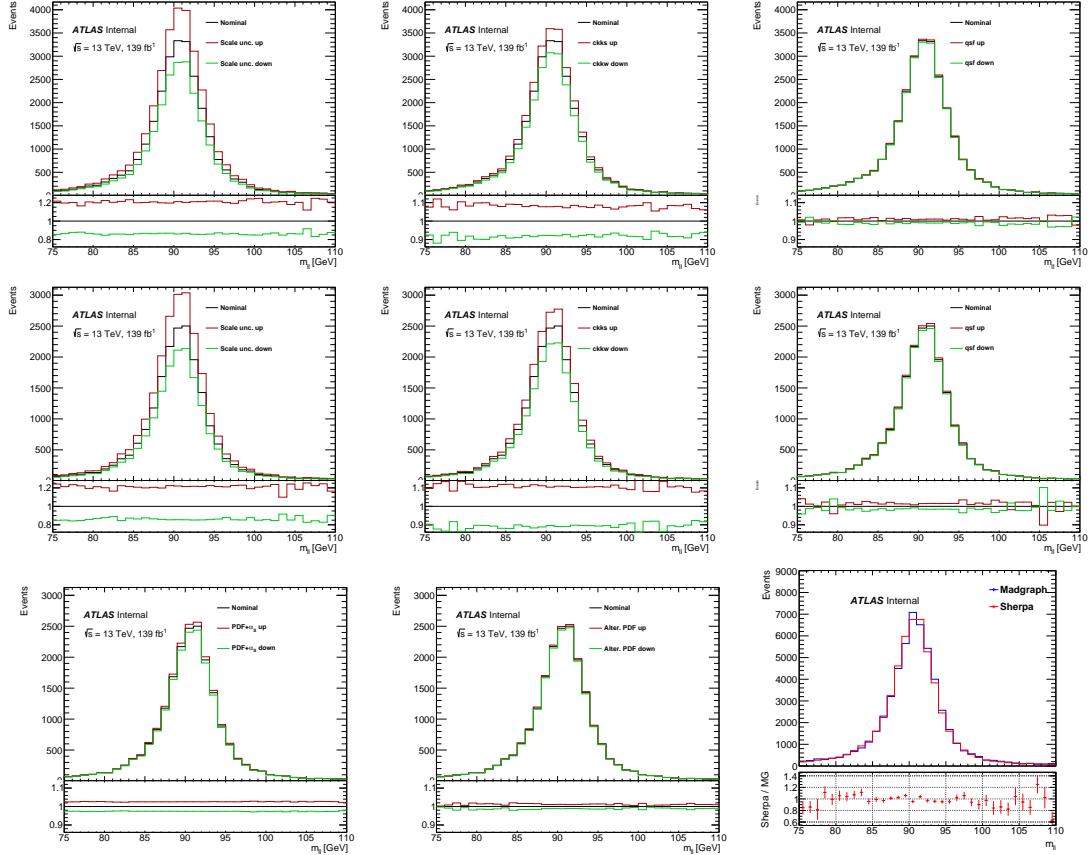


Figure B.16: Top row: the  $m_{ll}$  spectra comparisons for nominal and systematics from scale variations (left), ckkw (middle) and qsf (right) on  $Z(\rightarrow \mu\mu) + \text{HF}$  (first row) and  $Z(\rightarrow \ell\ell) + \text{HF}$  (second row) processes. Bottom row: the  $m_{ll}$  spectra comparisons for nominal and systematics from  $\text{PDF}+\alpha_S$  (left) and alternative PDF (middle) variations for  $Z(\rightarrow ee) + \text{HF}$  process and on the right for the nominal Sherpa generator and alternative MadGraph generator for  $Z(\rightarrow \ell\ell) + \text{HF}$  process. Images reproduced from analysis internal notes.

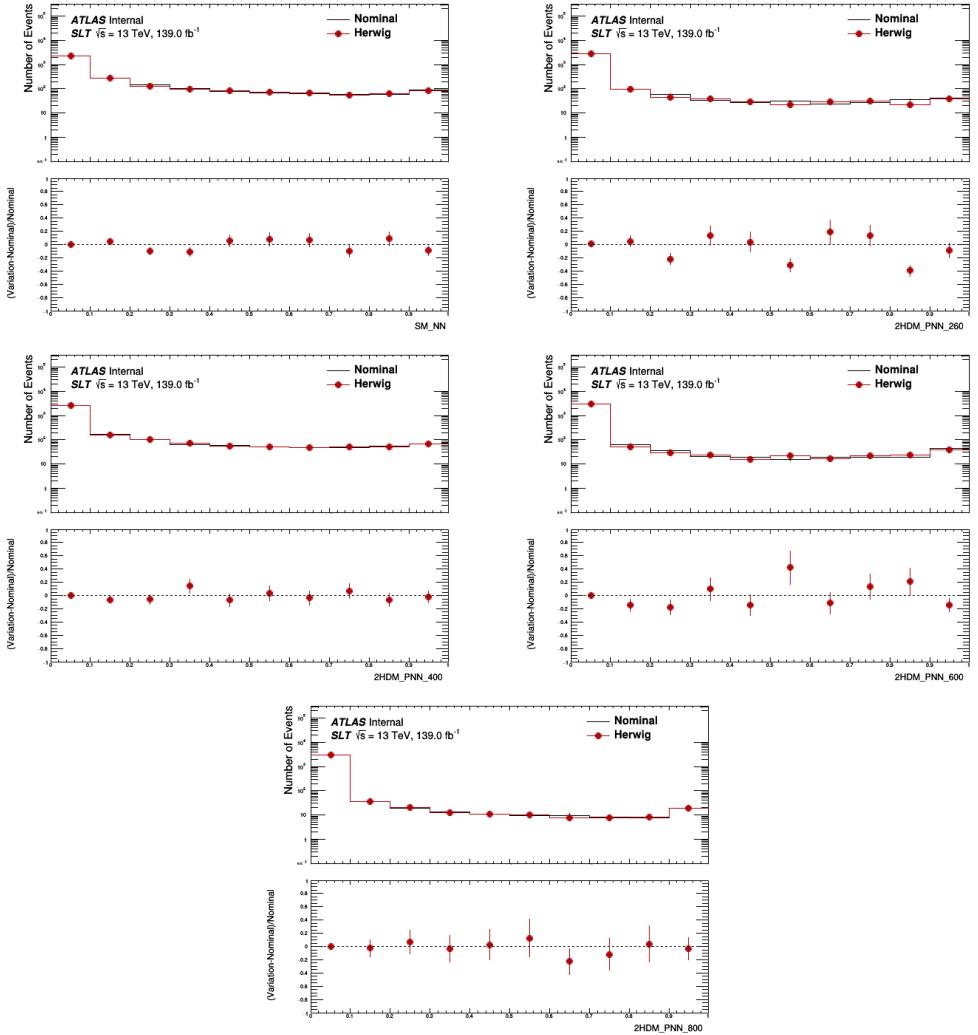


Figure B.17: SLT channel: shape-only PS uncertainty in the NN and PNN score of various resonant masses.

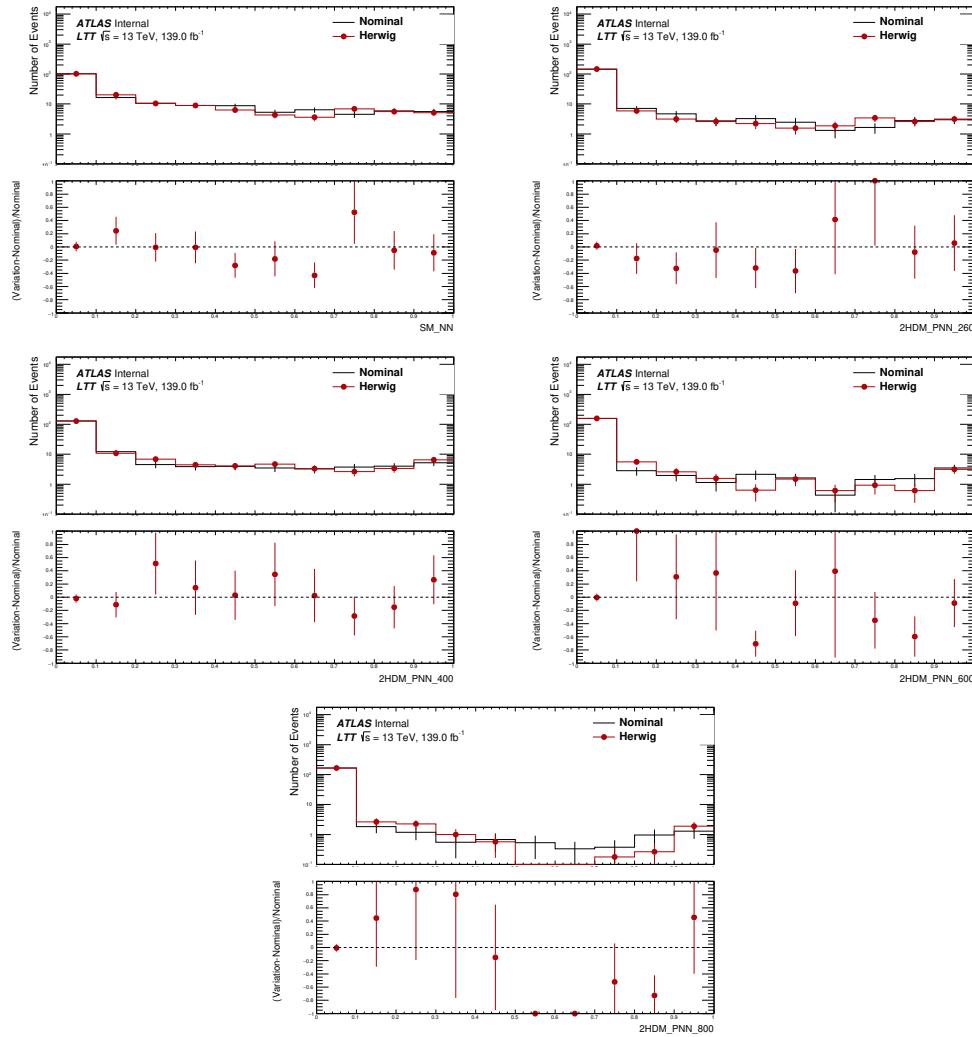


Figure B.18: LTT channel: shape-only PS uncertainty in the NN and PNN score of various resonant masses.

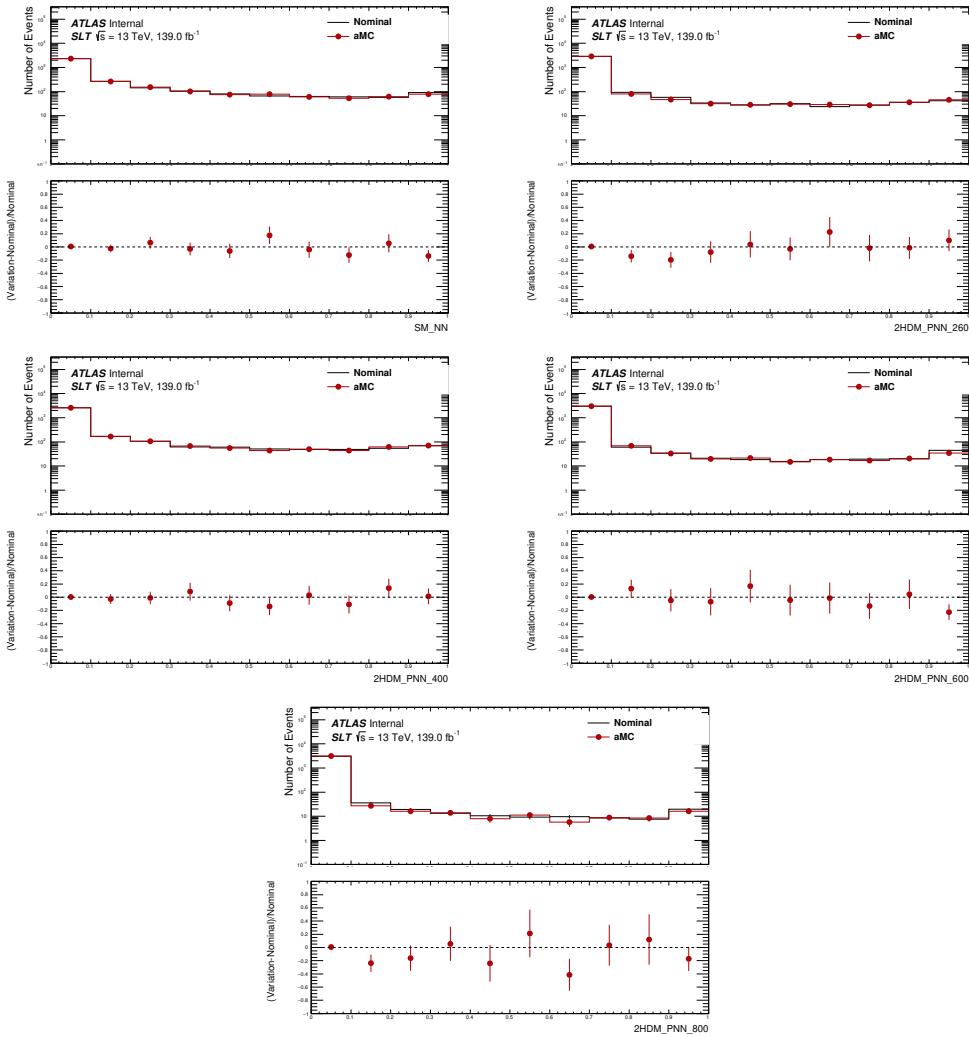


Figure B.19: SLT channel: shape-only ME uncertainty in the NN and PNN score of various resonant masses.

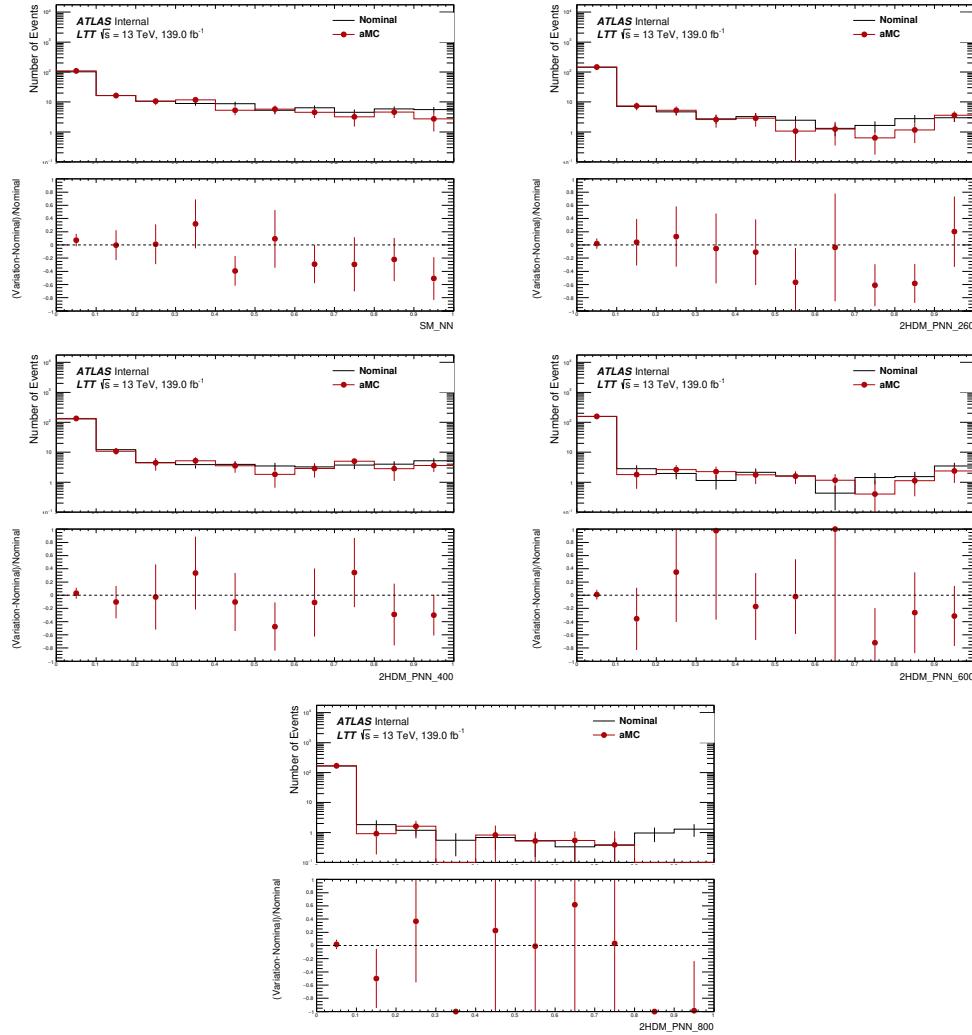


Figure B.20: LTT channel: shape-only ME uncertainty in the NN and PNN score of various resonant masses.

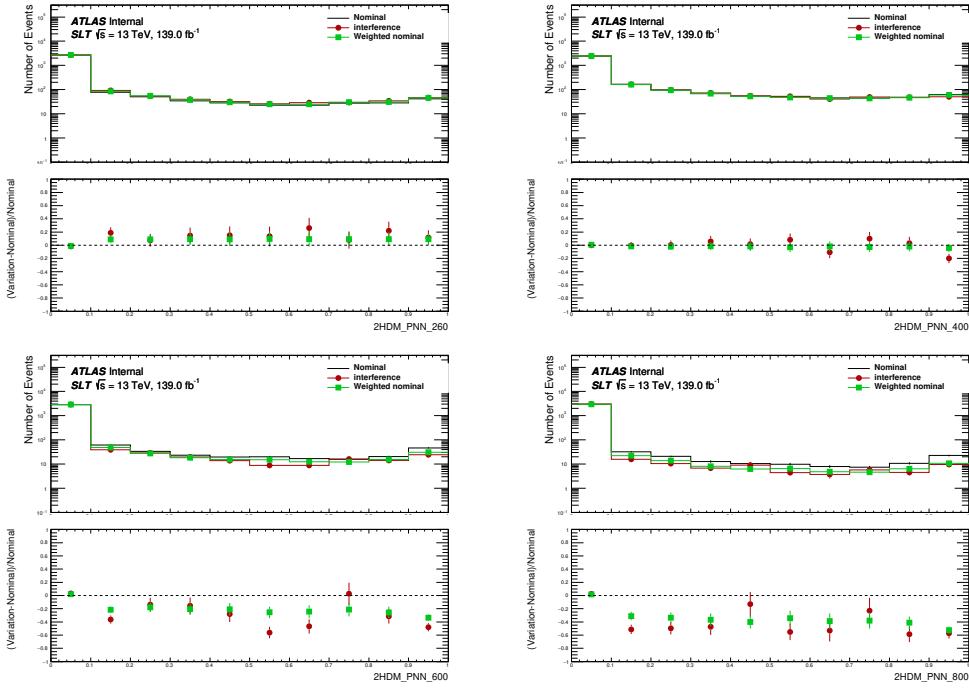


Figure B.21: SLT channel: shape-only interference uncertainty in the PNN score of various resonant masses.

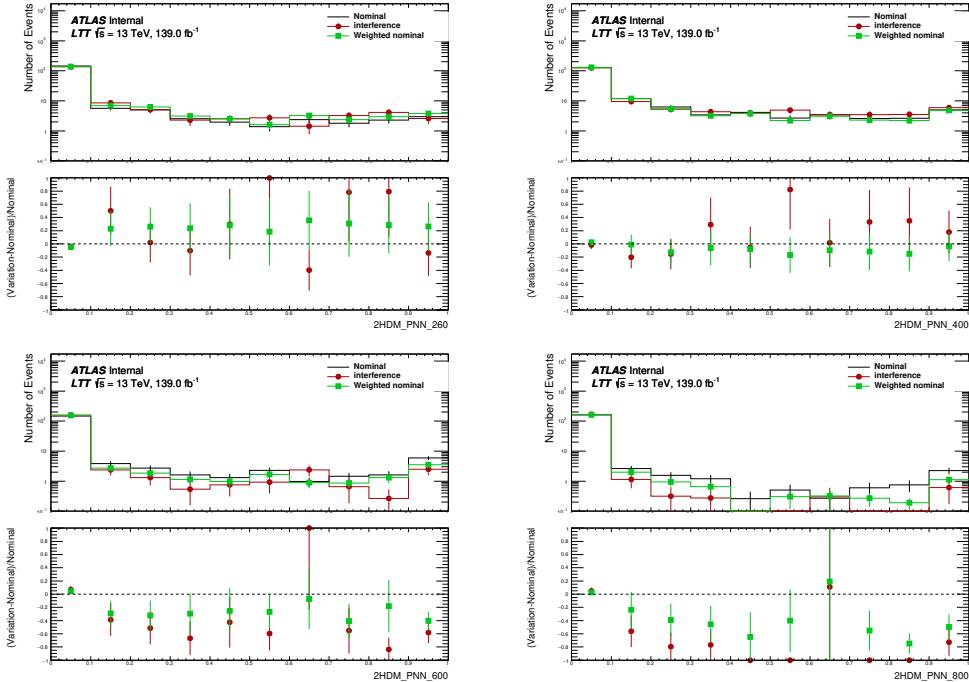


Figure B.22: LTT channel: shape-only interference uncertainty in the PNN score of various resonant masses.

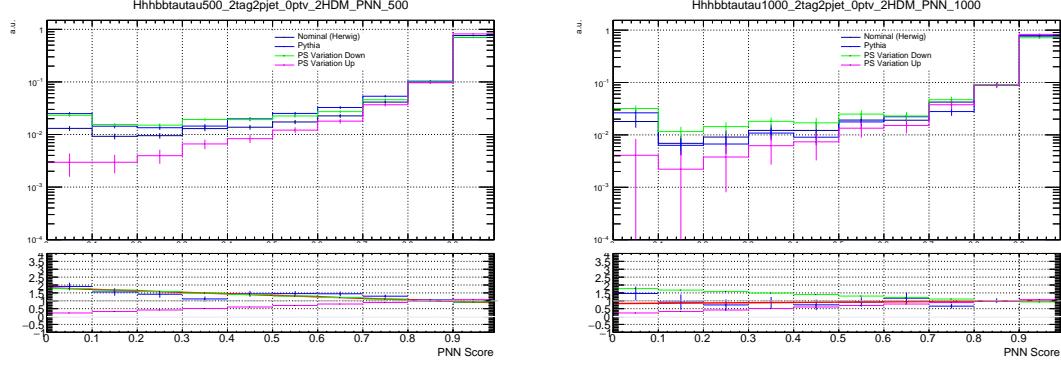


Figure B.23: Comparison of the SLT signal PNN distributions obtained from the nominal (black) and alternative (blue) signal samples for the PS variations for the  $m_X = 500$  GeV (left) and the  $m_X = 1000$  GeV (right) mass points. A linear fit to the ratio of the two distributions is performed and shown in the lower panel of the figures (red line). The variations obtained from the linear function obtained from the  $m_X = 500$  GeV fit are also shown in the figures (green and magenta). Image reproduced from analysis internal notes.

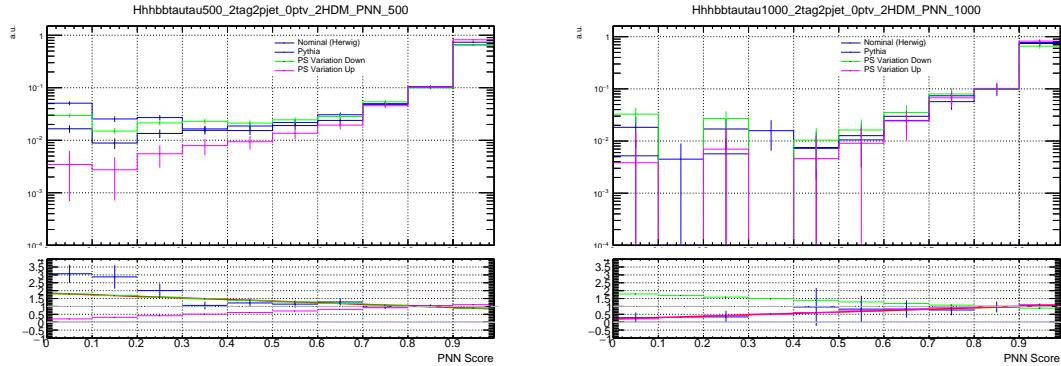


Figure B.24: Comparison of the LTT signal PNN distributions obtained from the nominal (black) and alternative (blue) signal samples for the PS variations for the  $m_X = 500$  GeV (left) and the  $m_X = 1000$  GeV (right) mass points. A linear fit to the ratio of the two distributions is performed and shown in the lower panel of the figures (red line). The variations obtained from the linear function obtained from the  $m_X = 500$  GeV fit are also shown in the figures (green and magenta). Image reproduced from analysis internal notes.

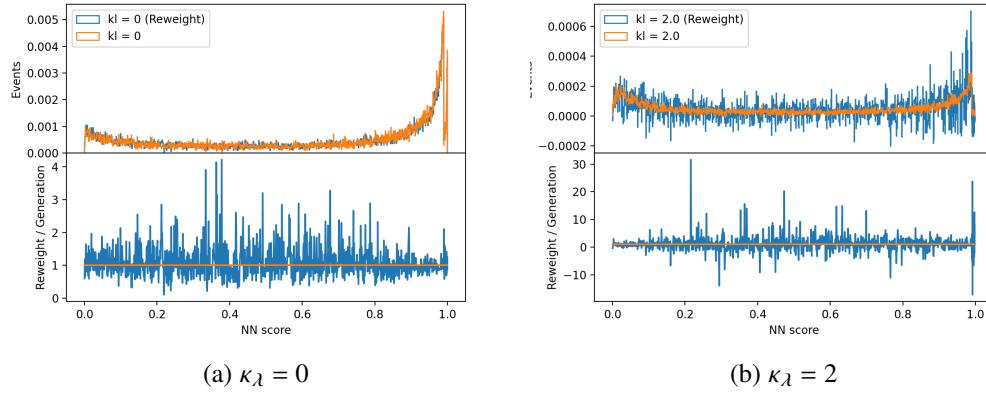


Figure B.25: Closure plots for the linear combination of VBF samples for two different bases in the  $\tau_{\text{lep}}\tau_{\text{had}}$  channel. NN score distributions for the MC  $\kappa_\lambda = 0, 2$  samples (orange) and the reweighted ones (blue) are shown together with their ratio. The largest difference between the two distributions is of 2.2% (SLT) and stems from the linear combination of VBF samples ( $\kappa_\lambda = 1, 10, 0$ ) to  $\kappa_\lambda = 2$ . Figure reproduced from analysis internal notes.

shows the rate of differences between the fakes estimated with  $t\bar{t}$  variation samples and the nominal  $t\bar{t}$  sample, to the nominal  $t\bar{t}$  sample. The nominal  $t\bar{t}$  sample is fast simulation sample and it's compared to the fast simulation variation  $t\bar{t}$  samples. These variation samples includes the uncertainties from the matrix elements, parton shower selection and the initial state radiation. The final state radiation radiation on the other hand is estimated with the full simulation  $t\bar{t}$  sample internal weights, as shown in Fig.B.30 for the SLT channel and Fig.B.31 for the LTT channel. Similarly the ratios are shwon in Fig.B.32 for the SLT channel and Fig.B.33 for the LTT channel.

## B.4 Additional material for results

Figures B.35- B.39 show the NP rankings for the  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit to data. Figures B.40- B.44 show the NP rankings for the  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit to data. Figures B.45- B.49 show the NP rankings for the  $\tau_{\text{lep}}\tau_{\text{had}}$  fit to data.

Figures B.50- B.54 show the NPs for the  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT channel. Figures B.55- B.59 show the NPs for the  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT channel.

Figures B.60- B.64 show the equivalent NP pulls for the combined  $\tau_{\text{lep}}\tau_{\text{had}}$  channel.

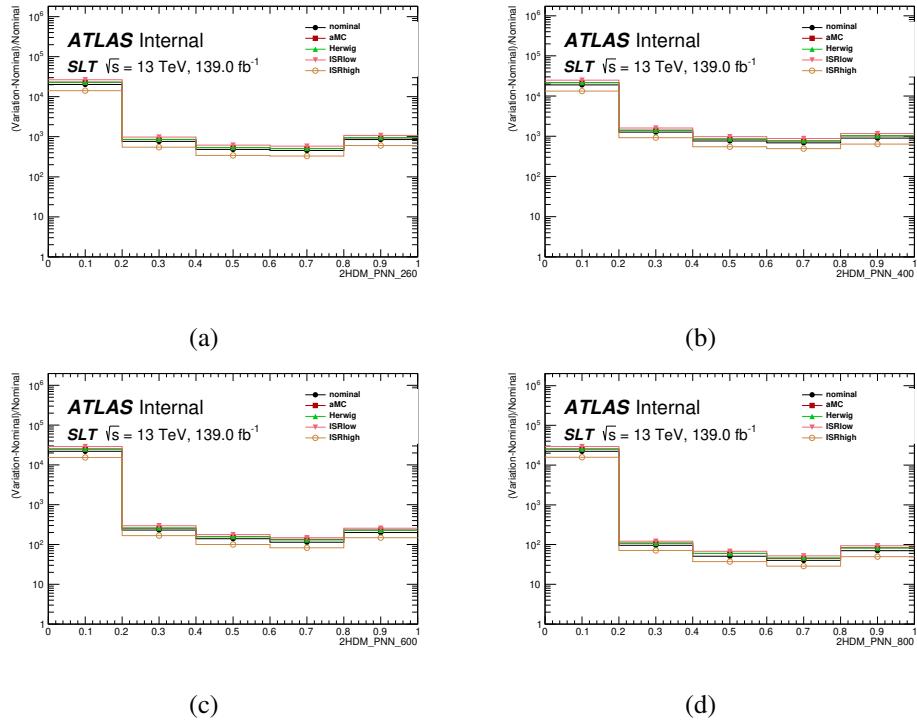


Figure B.26: SLT channel: data driven fakes estimated with the nominal  $t\bar{t}$  sample and the variation  $t\bar{t}$  sample.

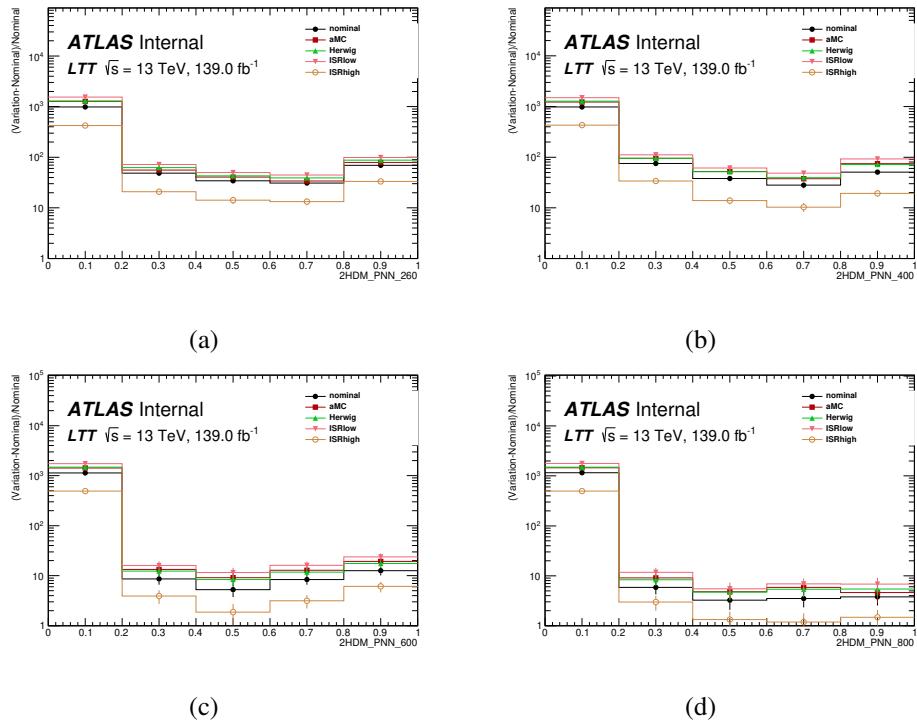


Figure B.27: LTT channel: data driven fakes estimated with the nominal  $t\bar{t}$  sample and the variation  $t\bar{t}$  sample.

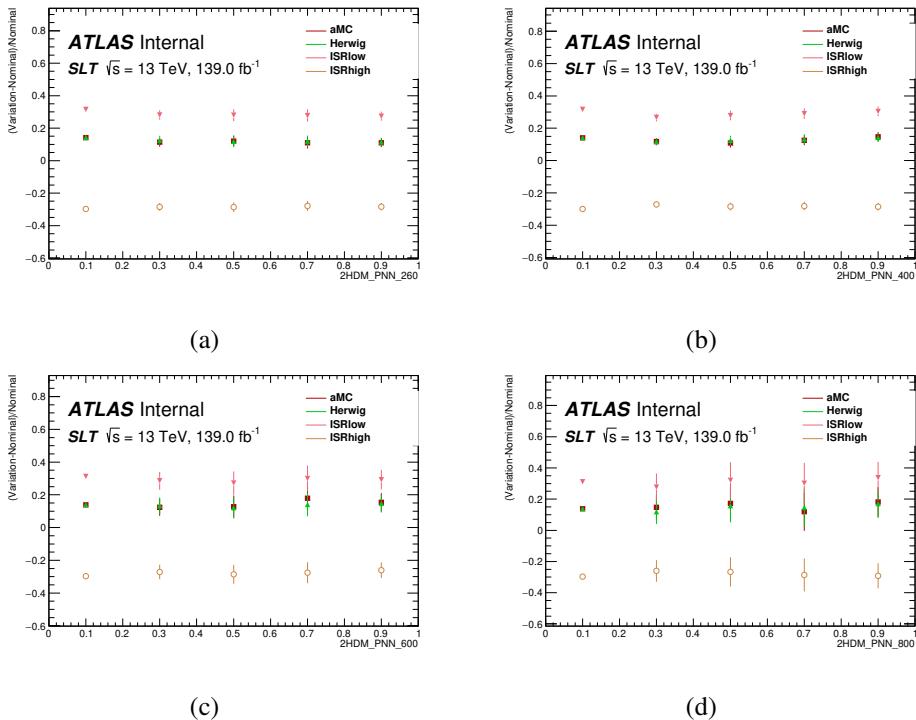


Figure B.28: SLT channel: data driven fakes estimated with the nominal  $t\bar{t}$  sample and the variation  $t\bar{t}$  sample.

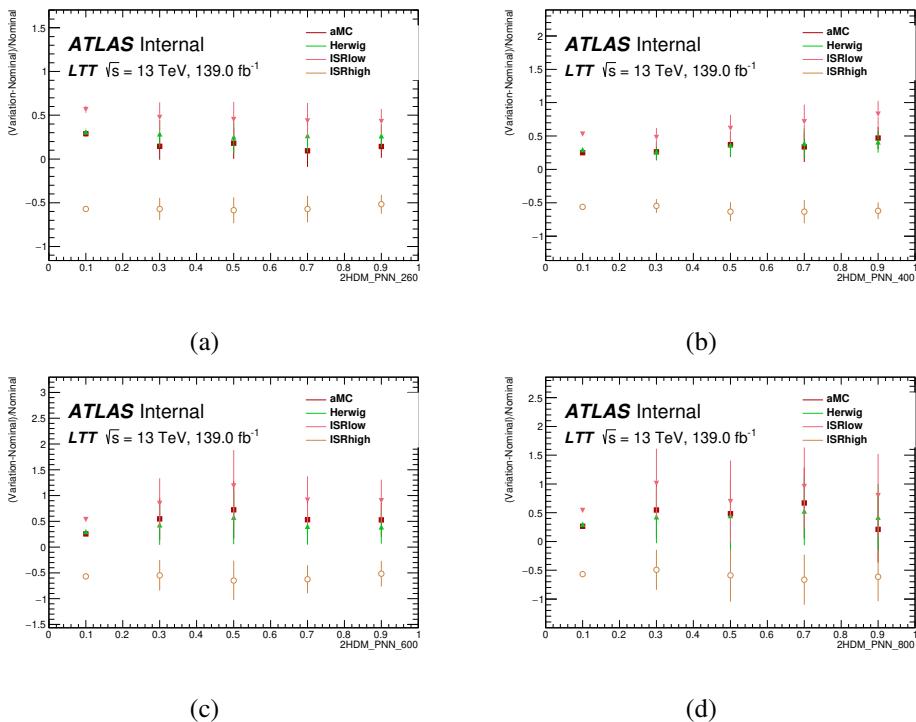


Figure B.29: LTT channel: data driven fakes estimated with the nominal  $t\bar{t}$  sample and the variation  $t\bar{t}$  sample.

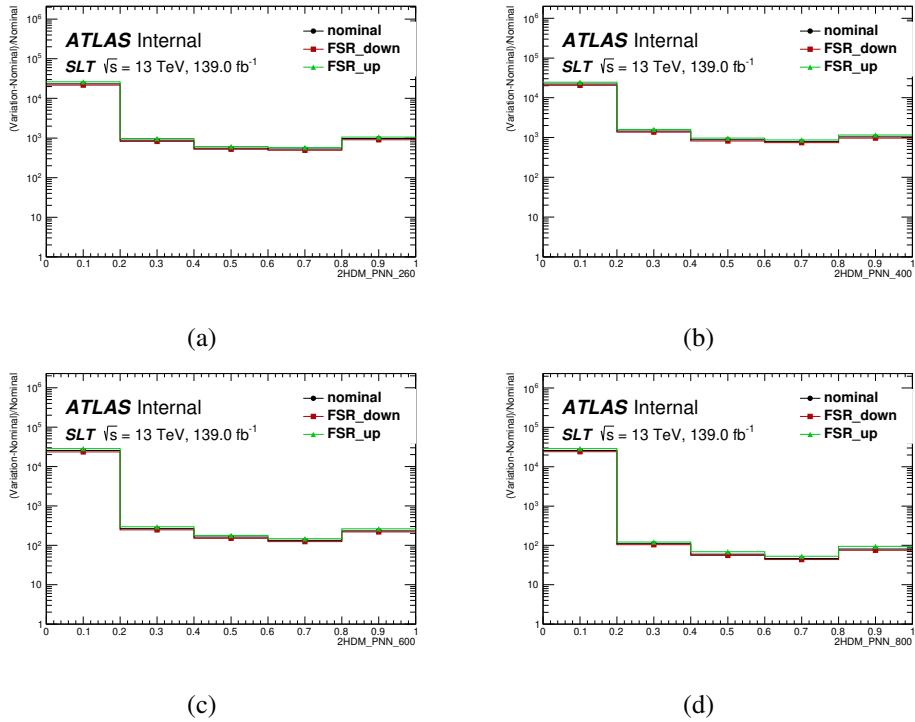


Figure B.30: SLT channel: data driven fakes estimated with the nominal  $t\bar{t}$  sample and the variation  $t\bar{t}$  sample.

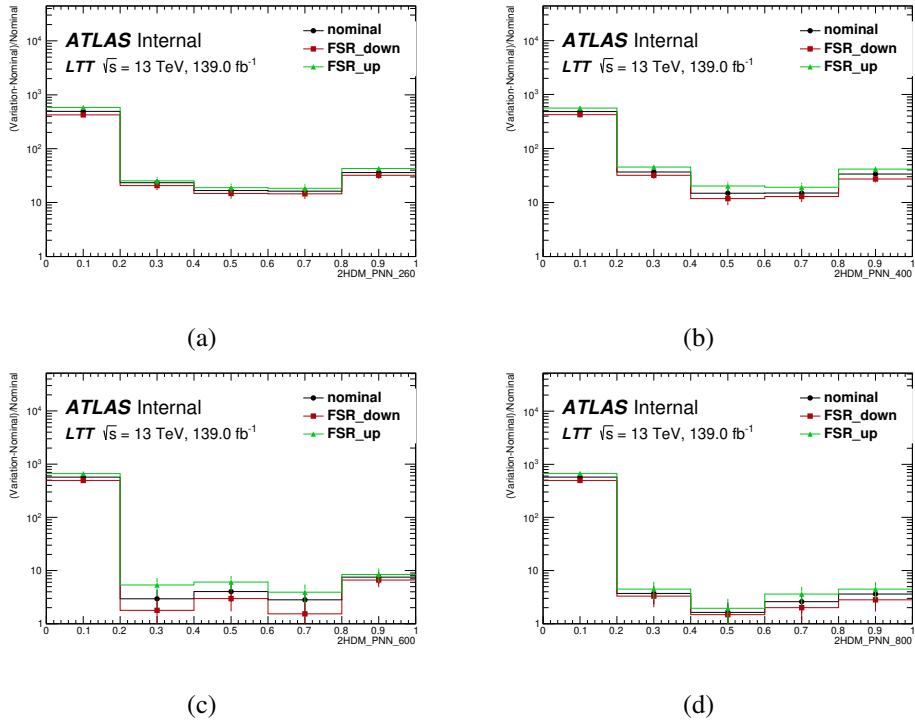


Figure B.31: LTT channel: data driven fakes estimated with the nominal  $t\bar{t}$  sample and the variation  $t\bar{t}$  sample.

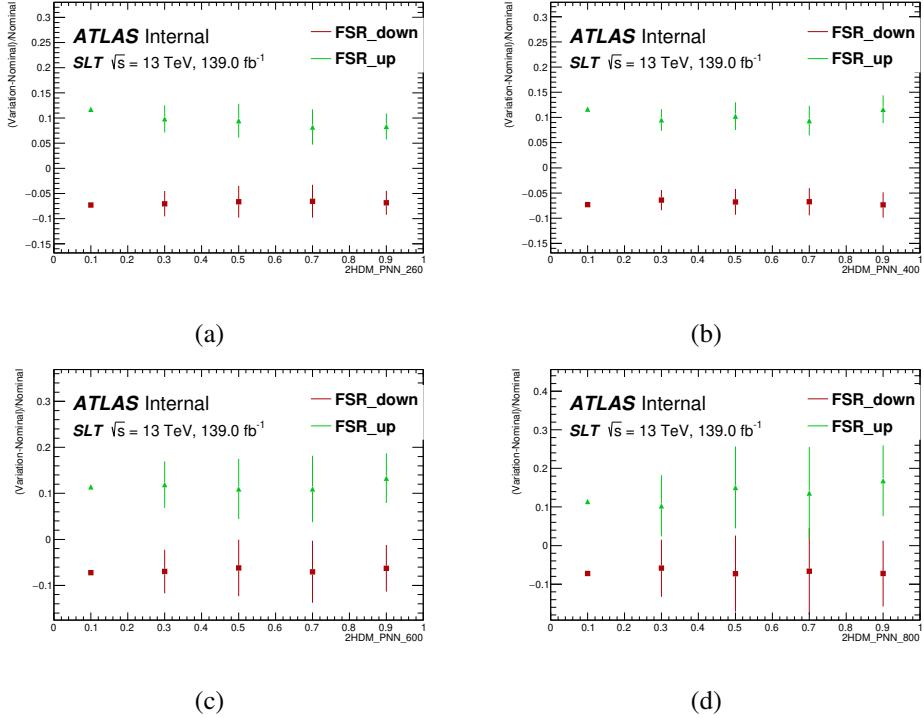


Figure B.32: SLT channel: data driven fakes estimated with the nominal  $t\bar{t}$  sample and the variation  $t\bar{t}$  sample.

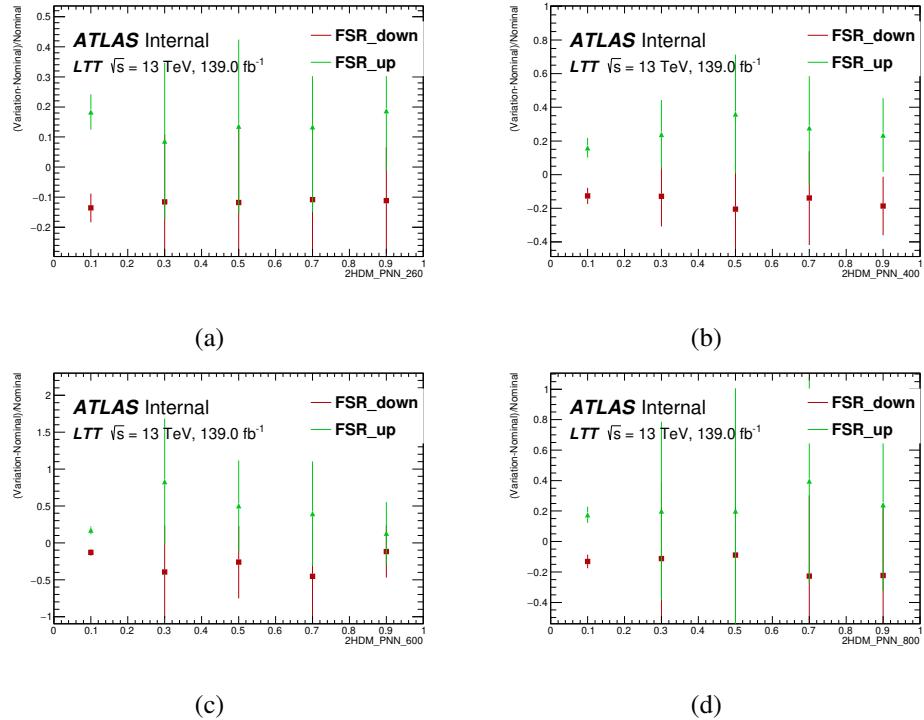


Figure B.33: LTT channel: data driven fakes estimated with the nominal  $t\bar{t}$  sample and the variation  $t\bar{t}$  sample.

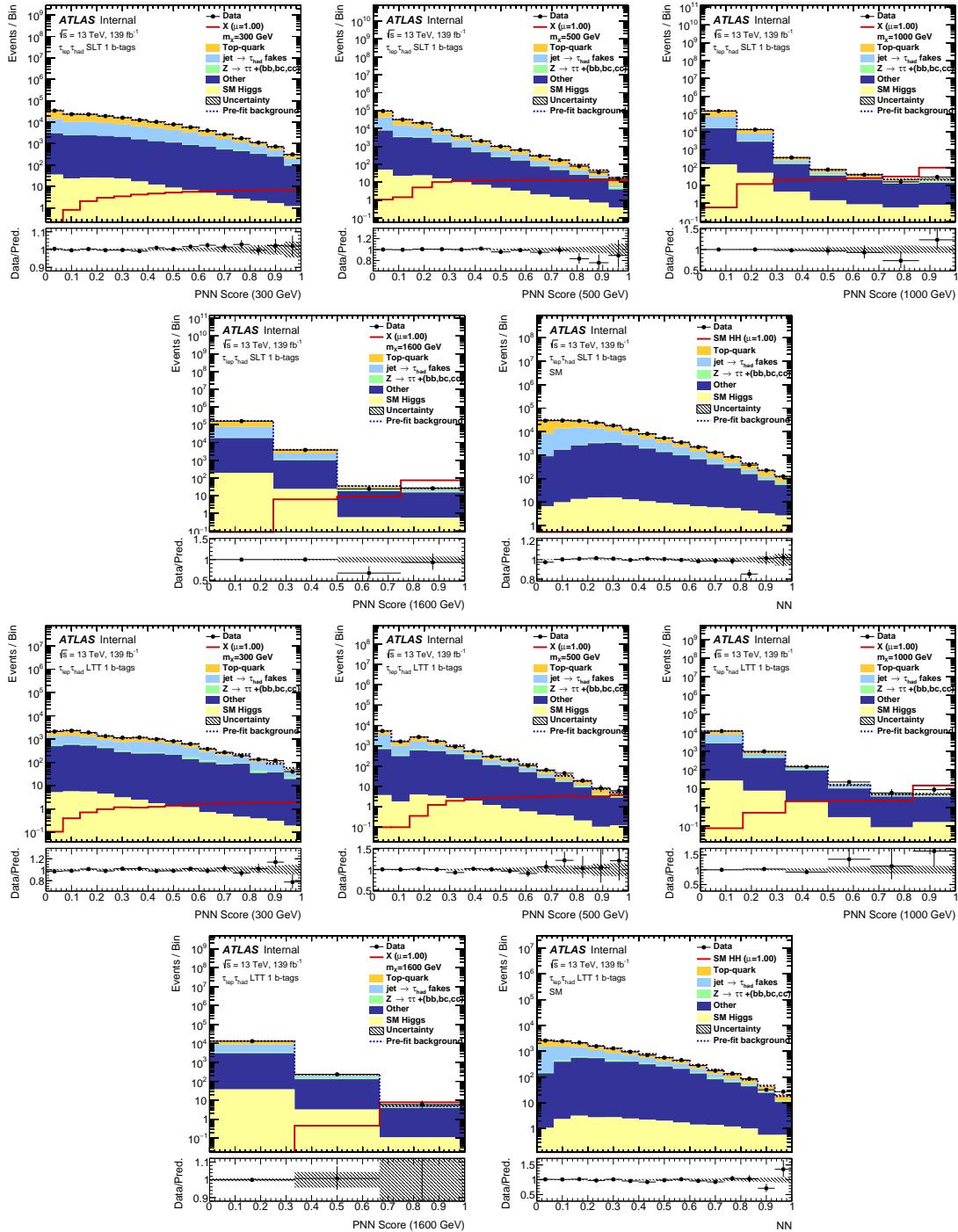


Figure B.34: Post-fit PNN score distributions for the 300, 500, 1000, 1600 GeV mass points and non-resonant NN in the di-Higgs  $bb\tau_{\text{lep}}\tau_{\text{had}}$  SLT (top two rows) and LTT (bottom two rows) 1 b-tag validation regions.

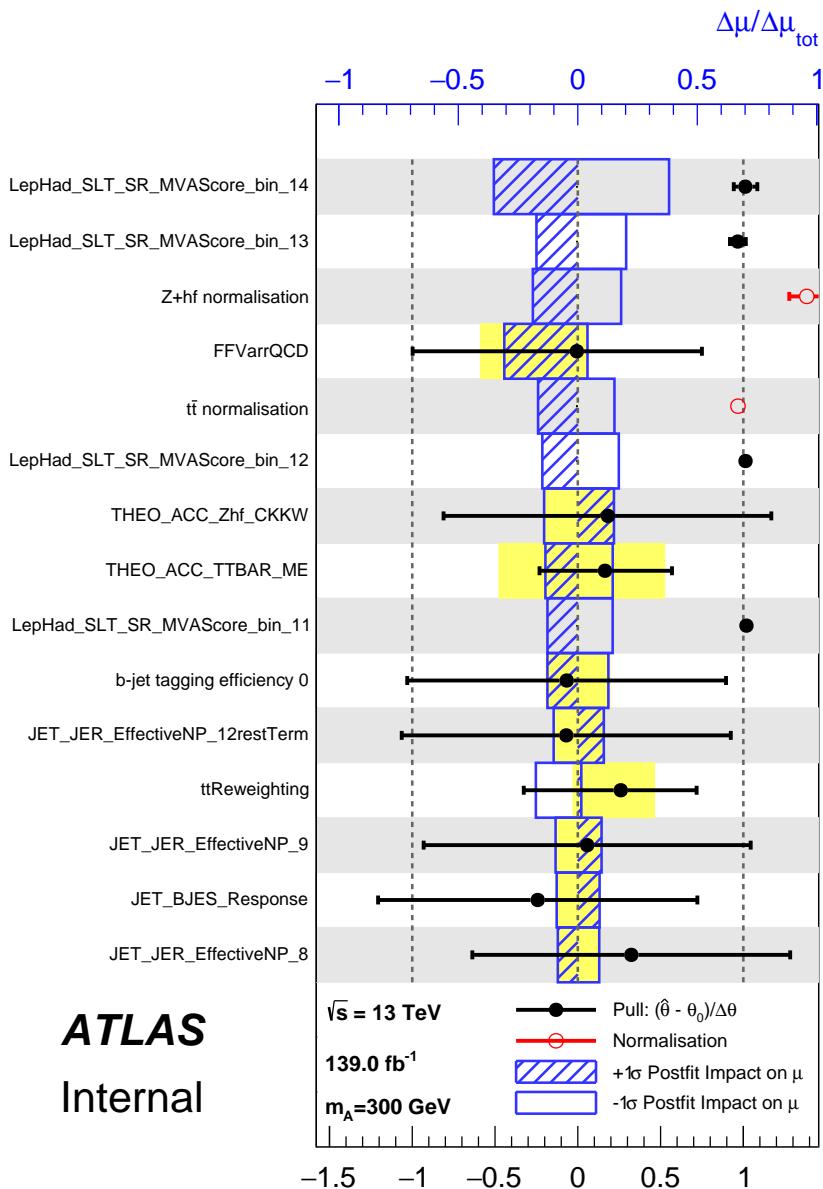


Figure B.35: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit to data for the 300 GeV resonant fit.

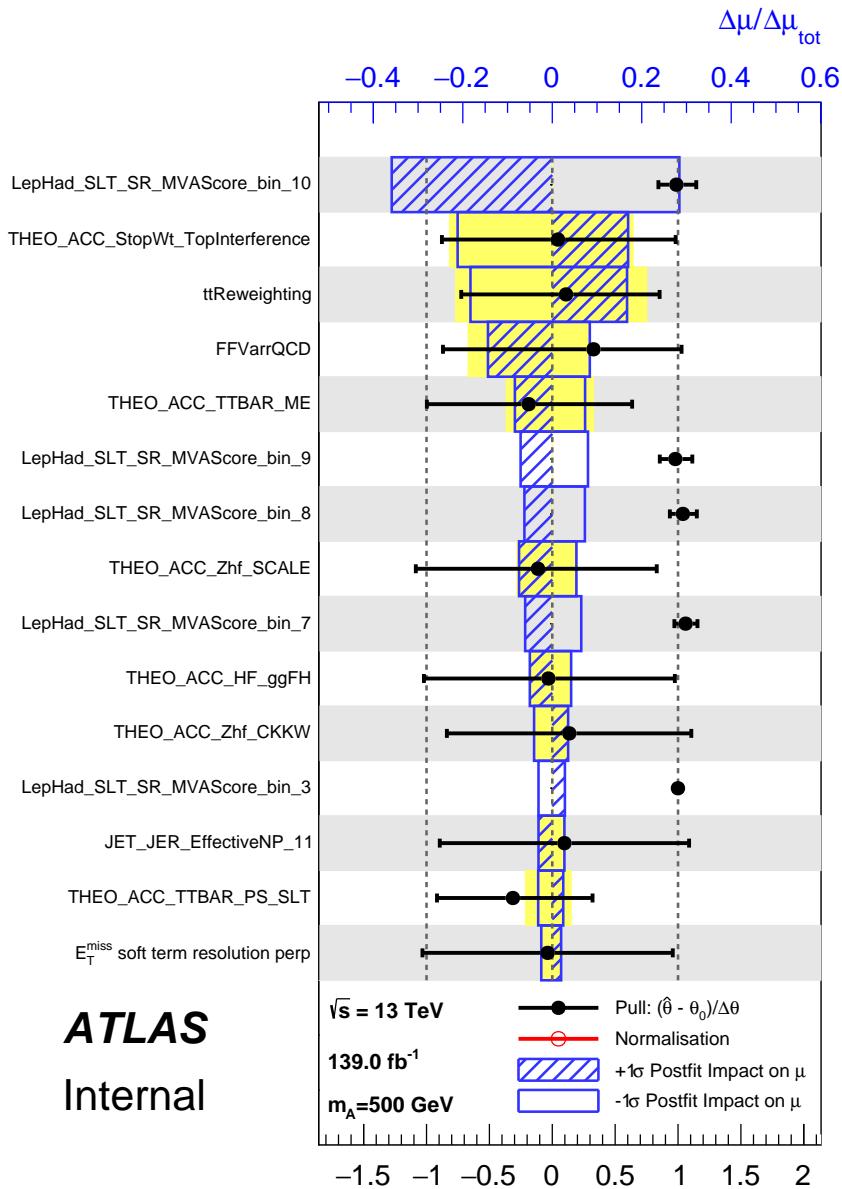


Figure B.36: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit to data for the 500 GeV resonant fit.

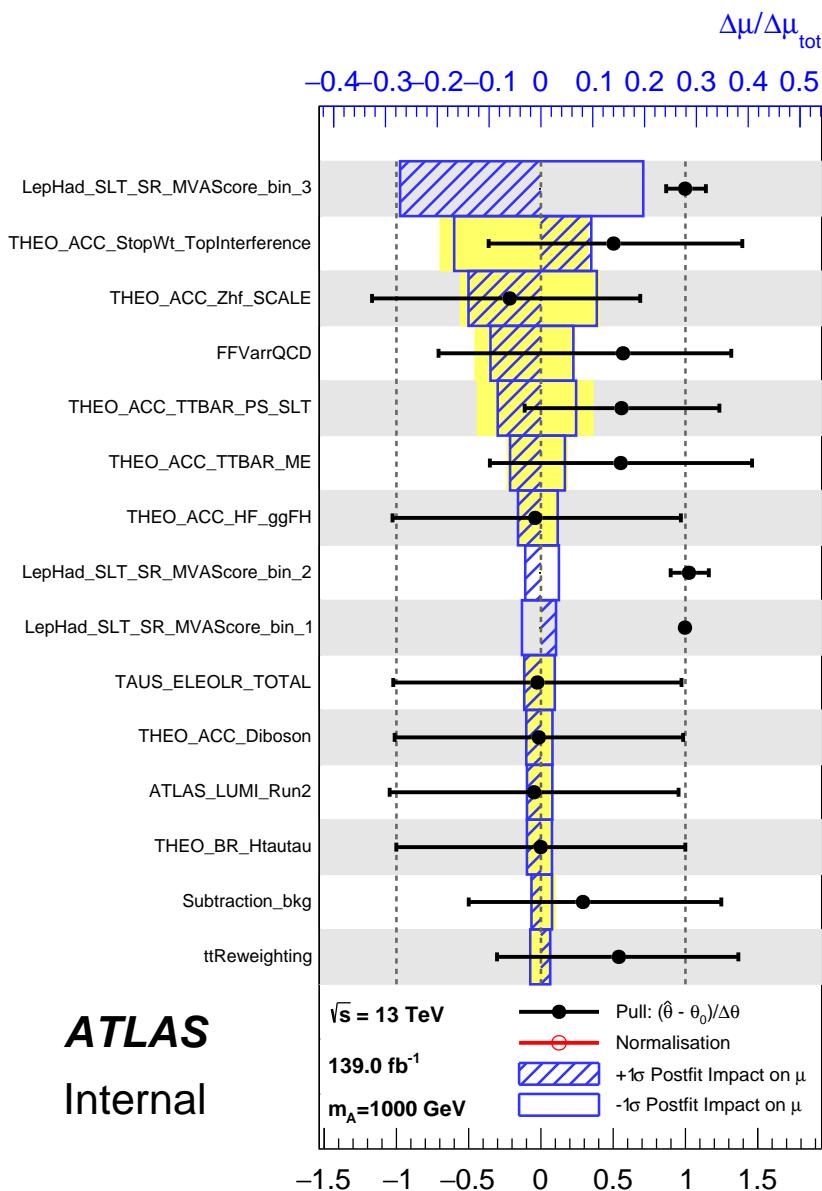


Figure B.37: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit to data for the 1000 GeV resonant fit.

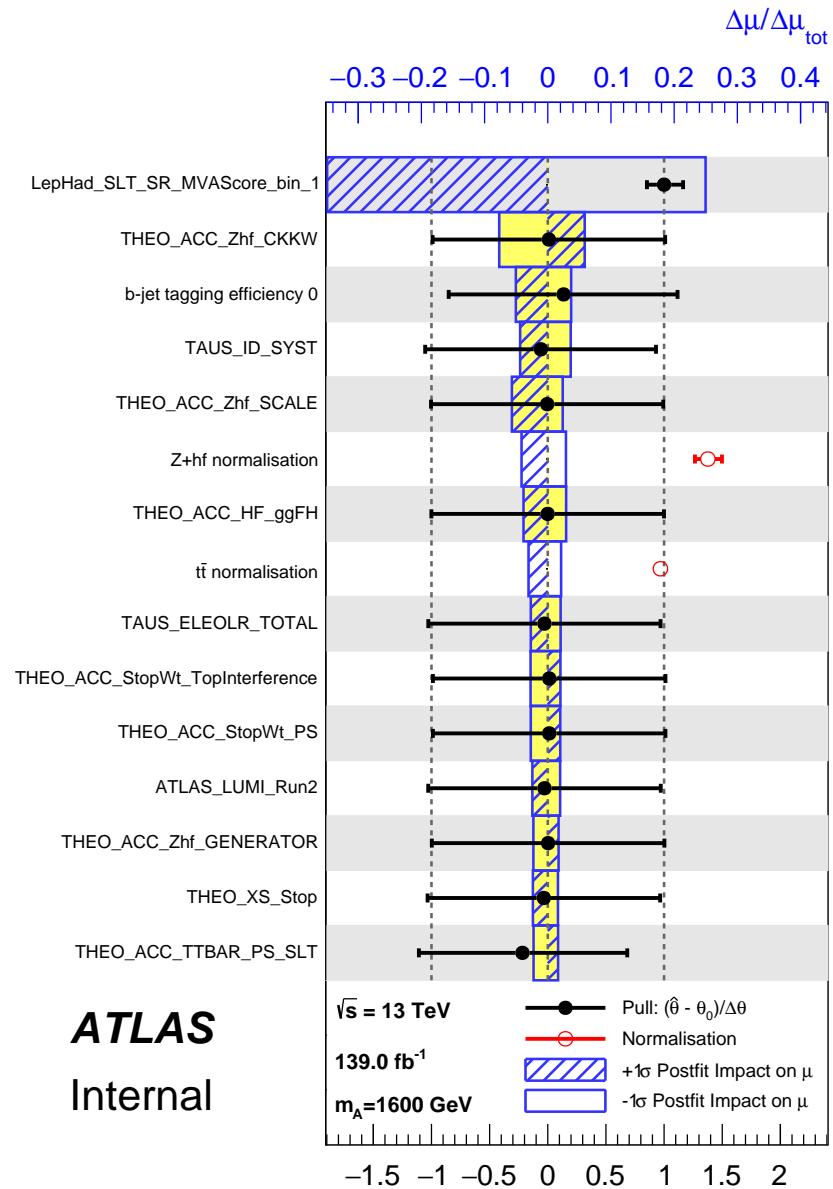
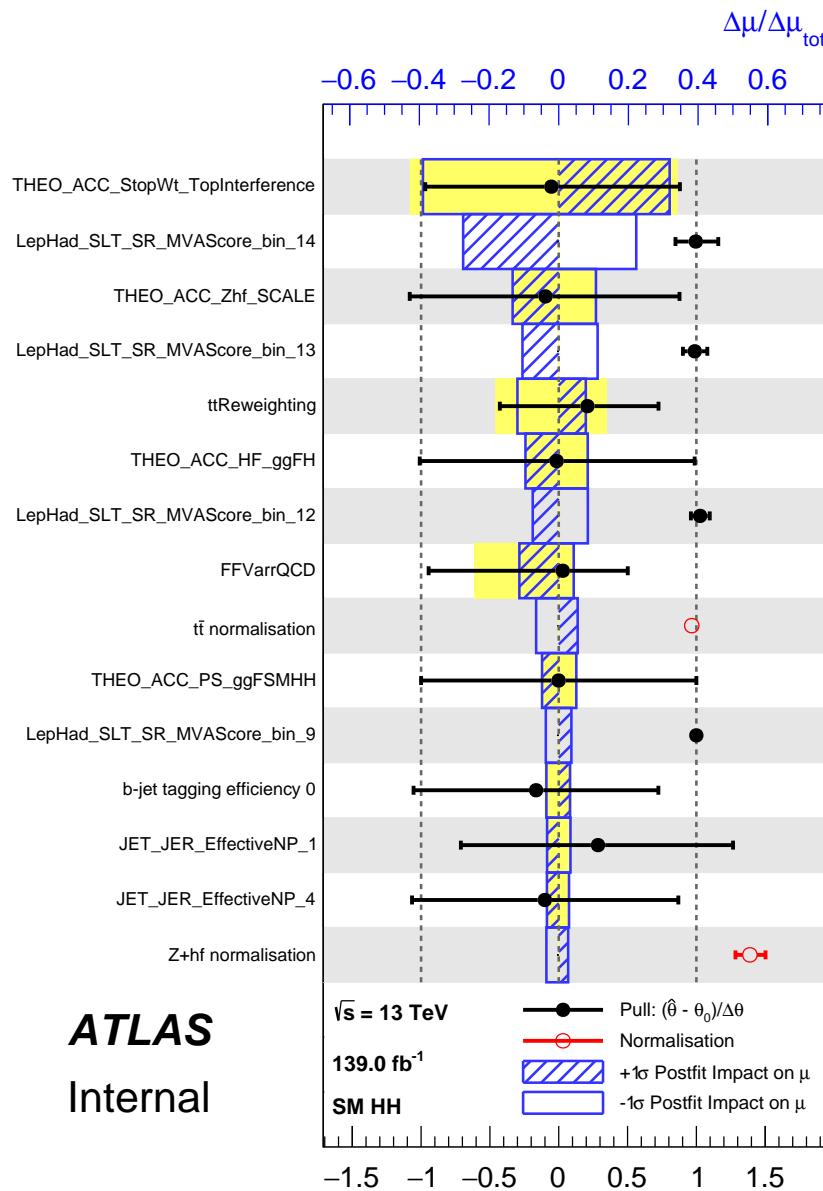


Figure B.38: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit to data for the 1600 GeV resonant fit.

Figure B.39: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit to data for the SM fit.

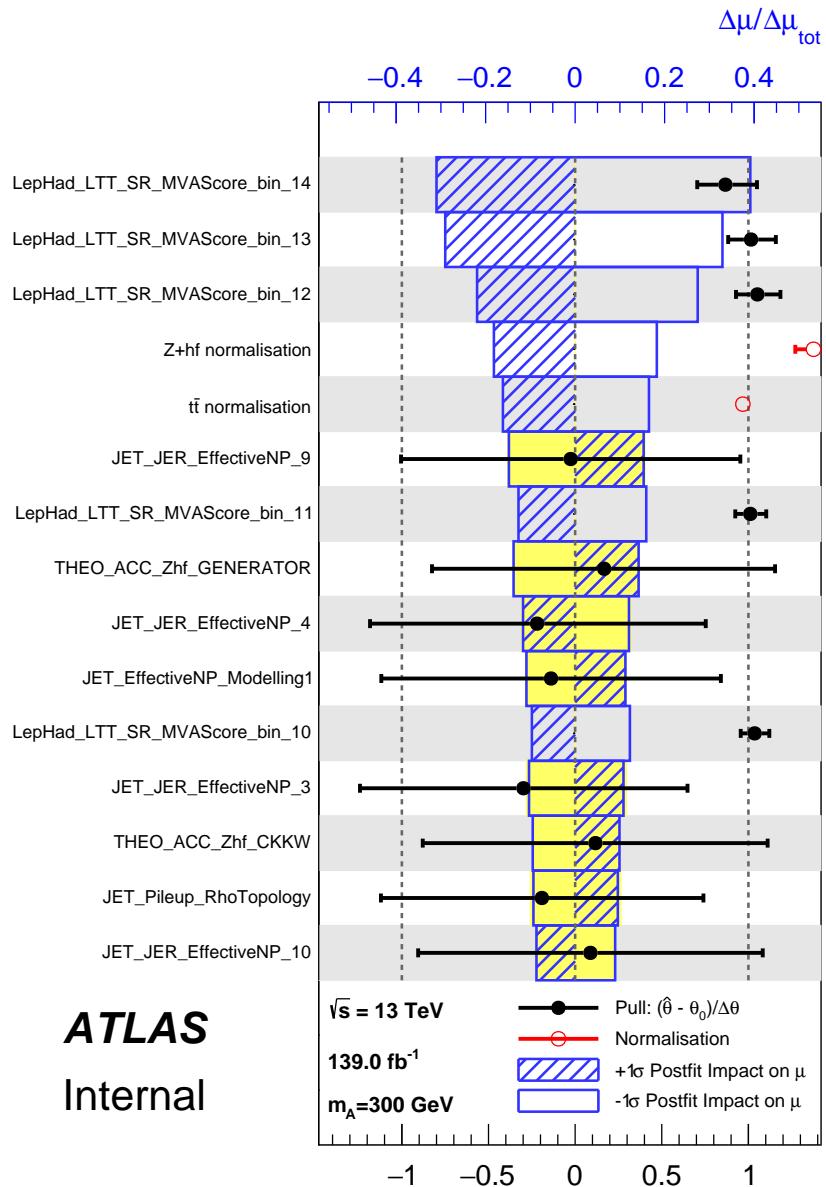


Figure B.40: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit to data for the 300 GeV resonant fit.

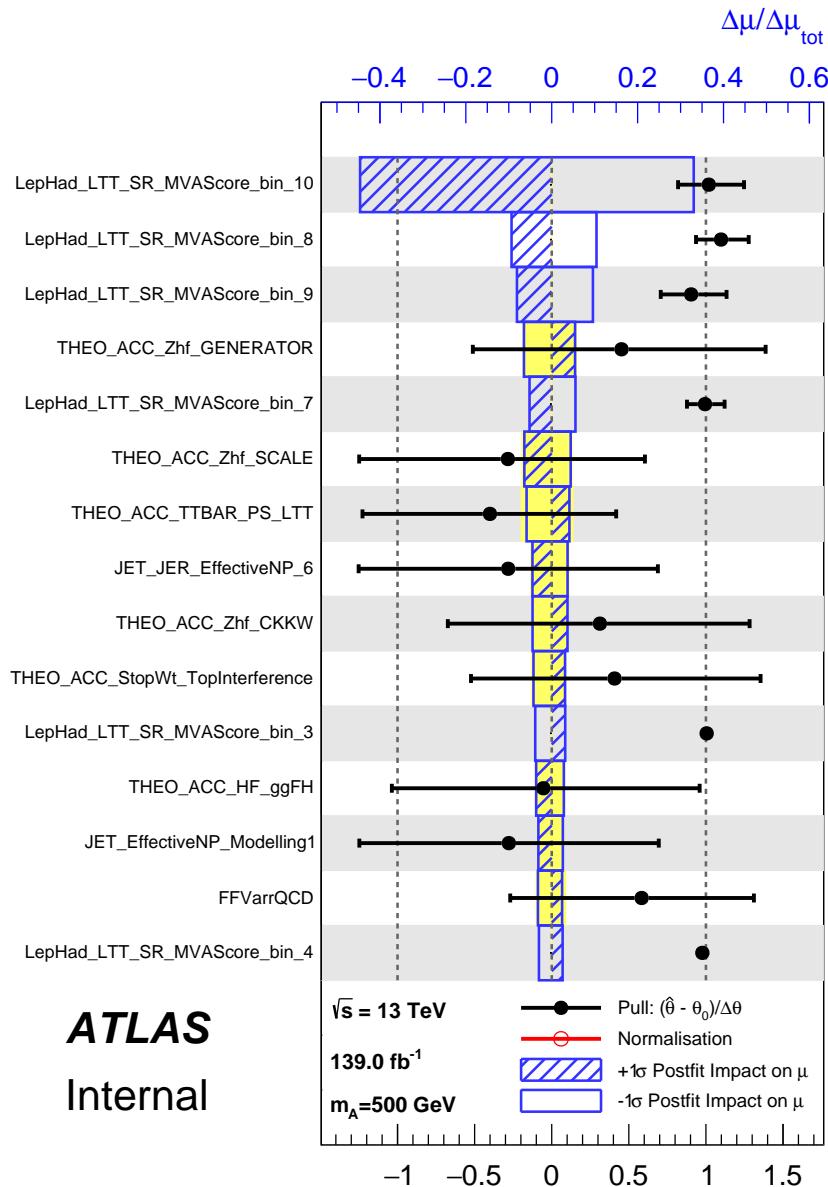


Figure B.41: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit to data for the 500 GeV resonant fit.

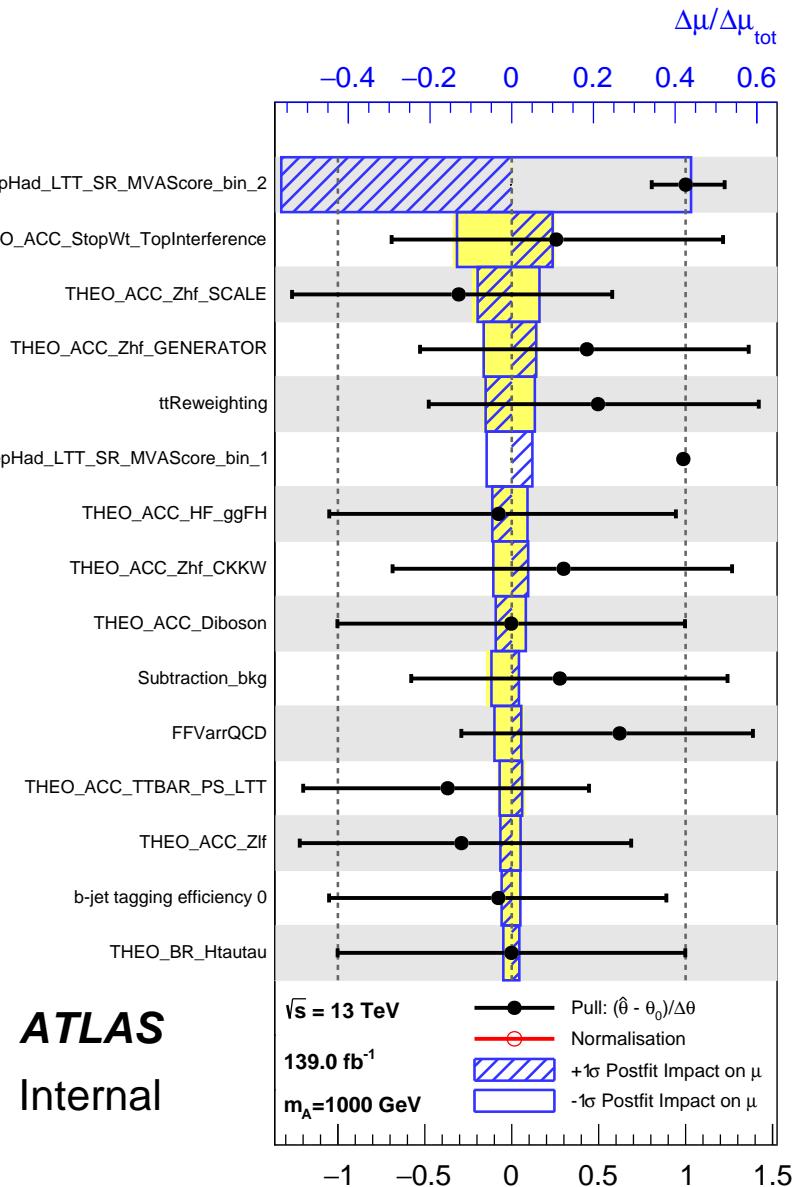


Figure B.42: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit to data for the 1000 GeV resonant fit.

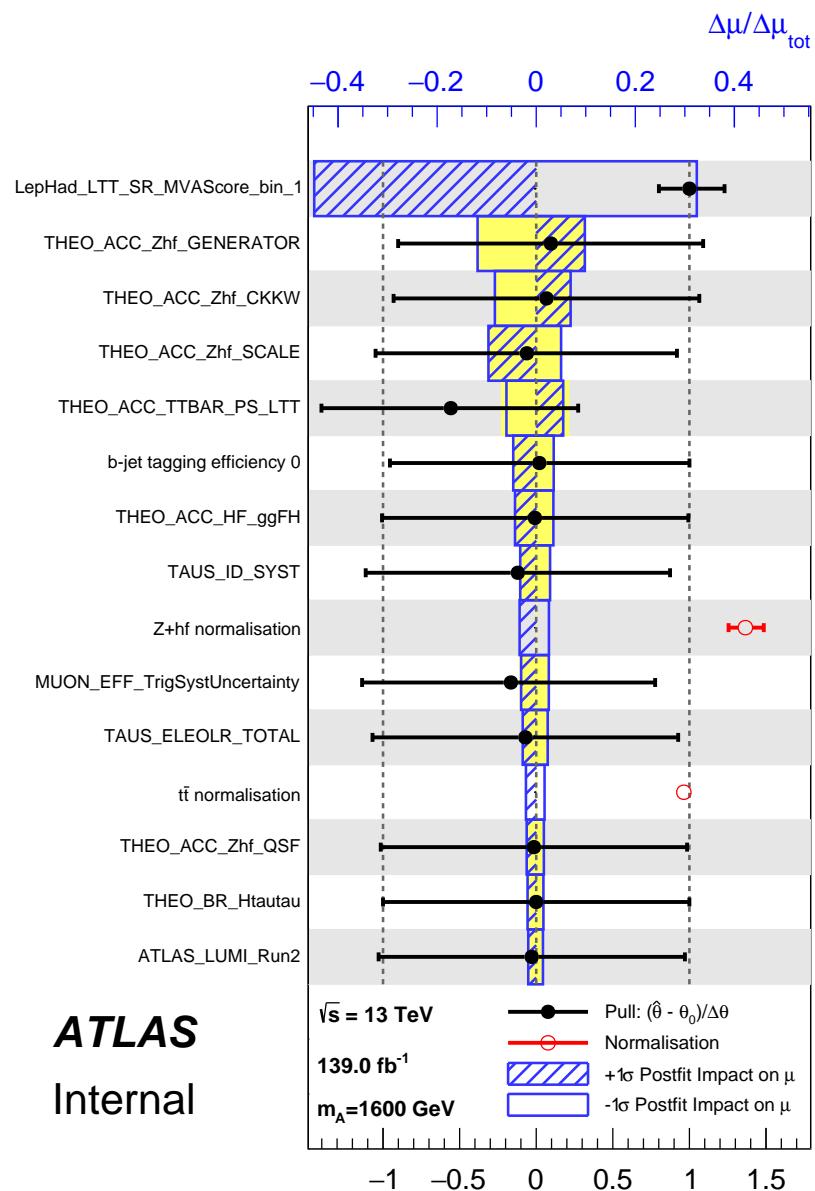


Figure B.43: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit to data for the 1600 GeV resonant fit.

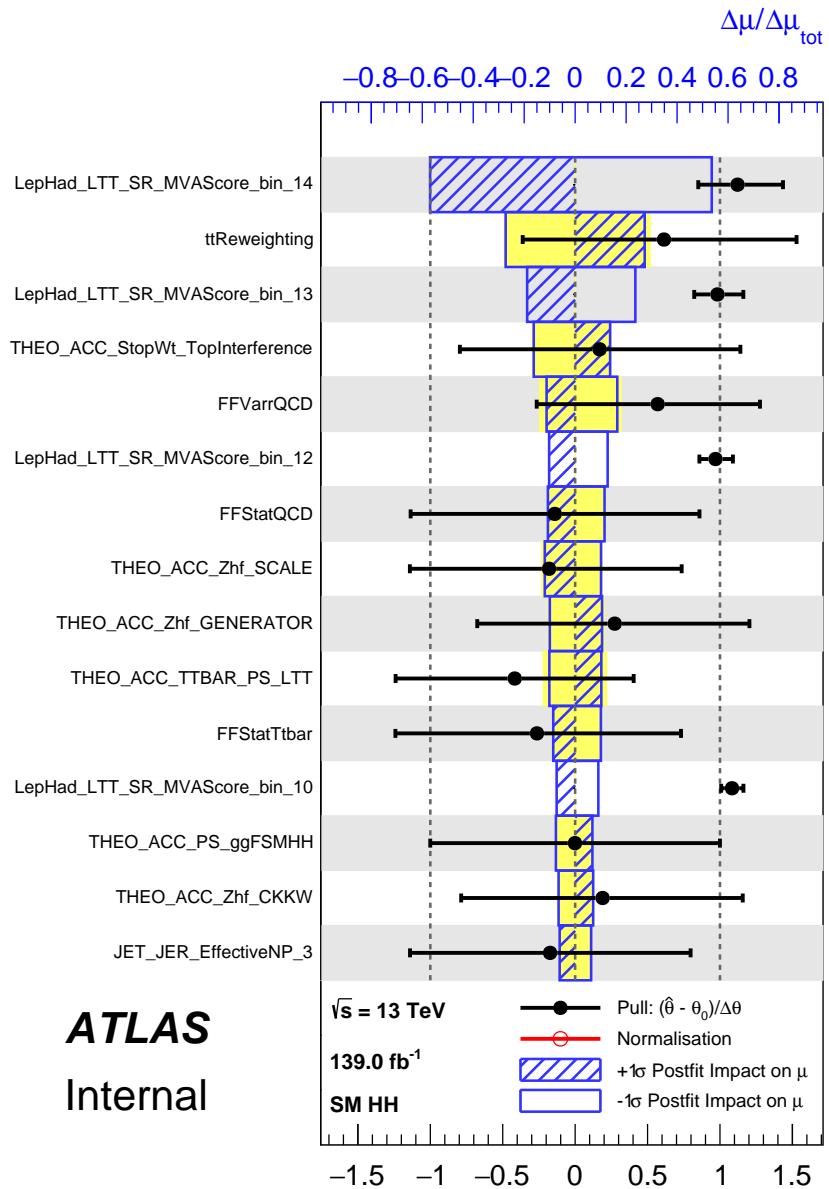
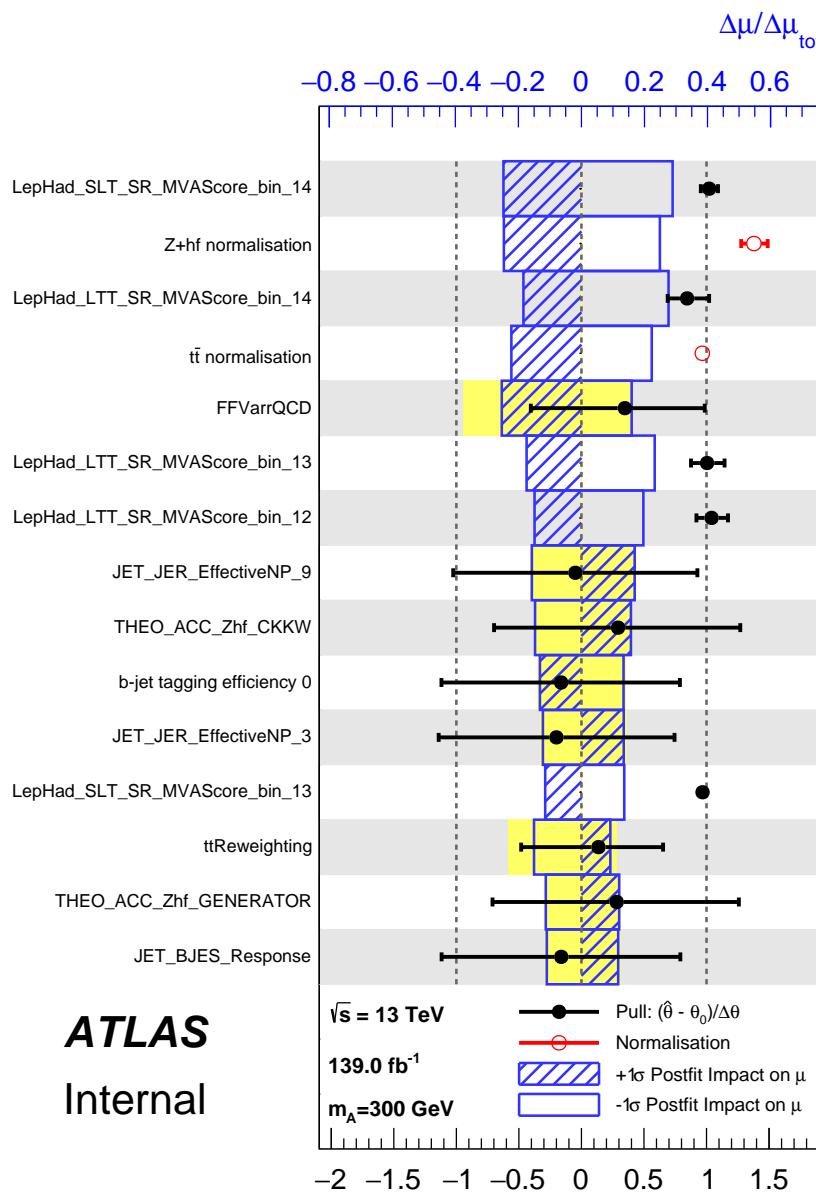


Figure B.44: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit to data for the SM fit.

Figure B.45: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit to data for the 300 GeV resonant fit.

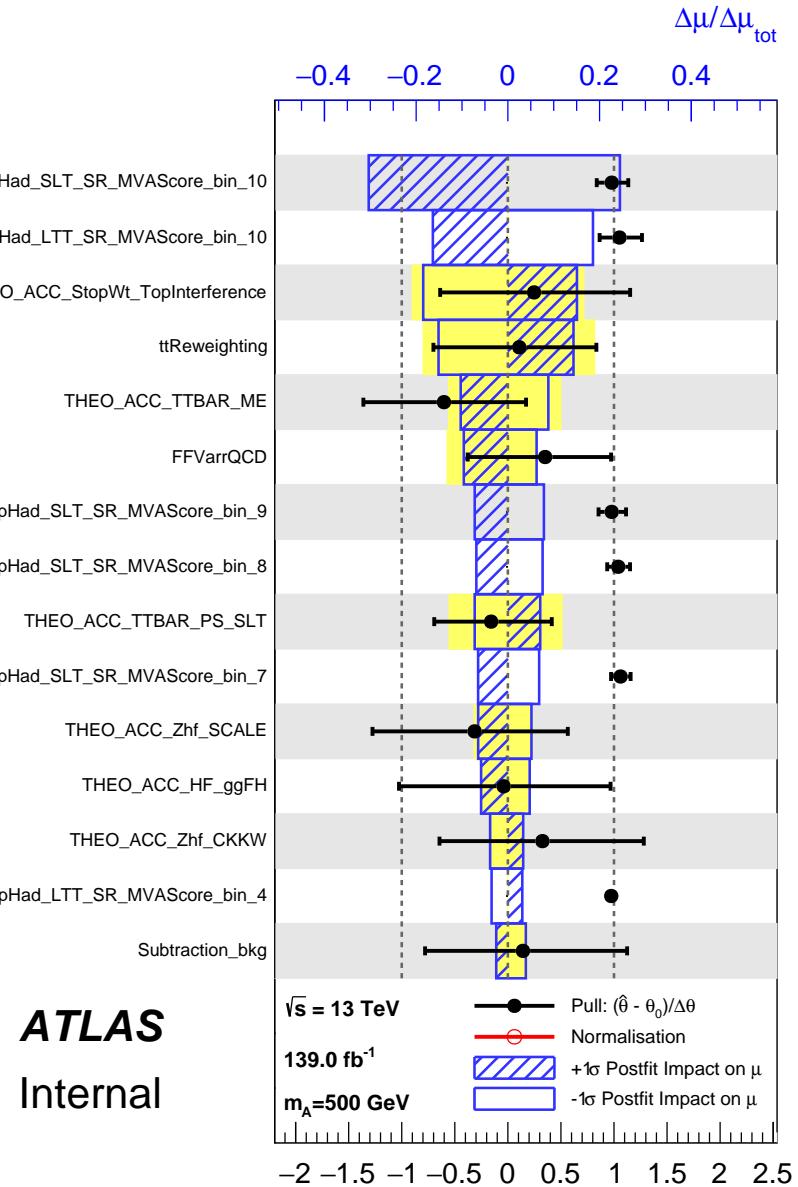


Figure B.46: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit to data for the 500 GeV resonant fit.

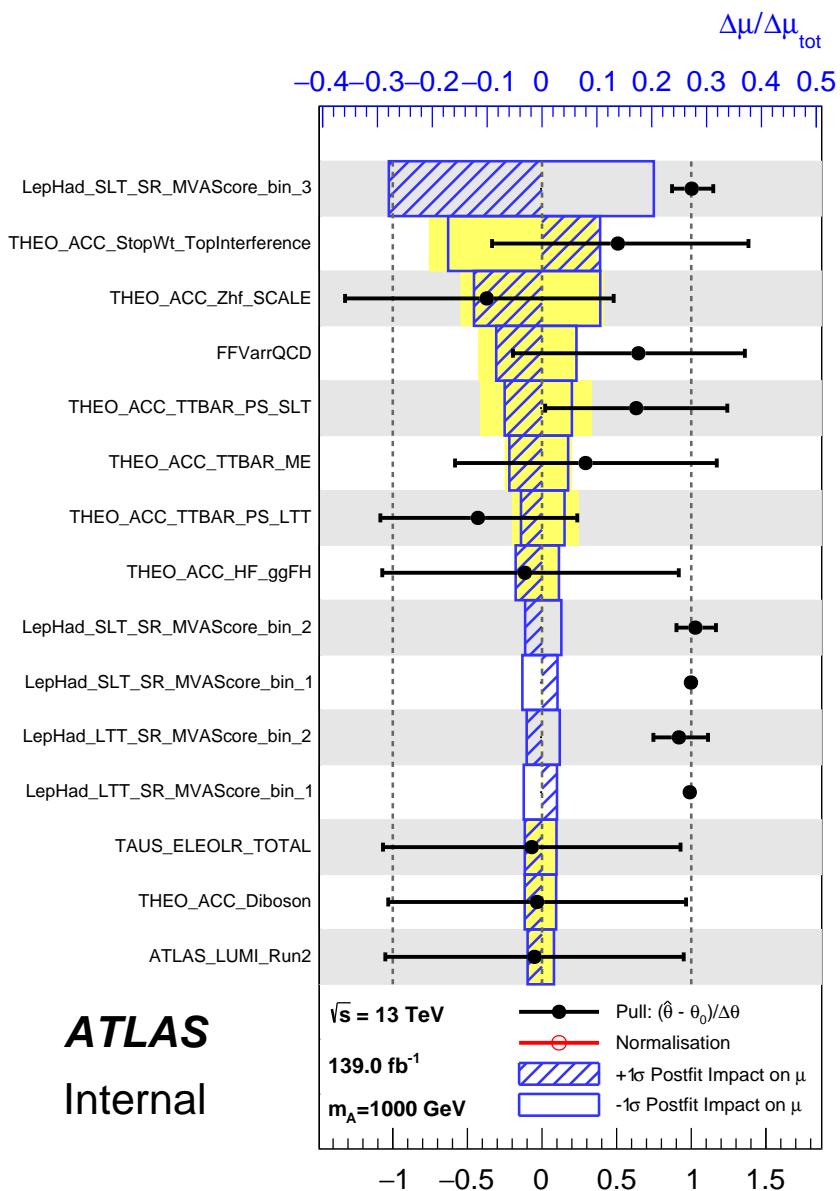


Figure B.47: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit to data for the 1000 GeV resonant fit.

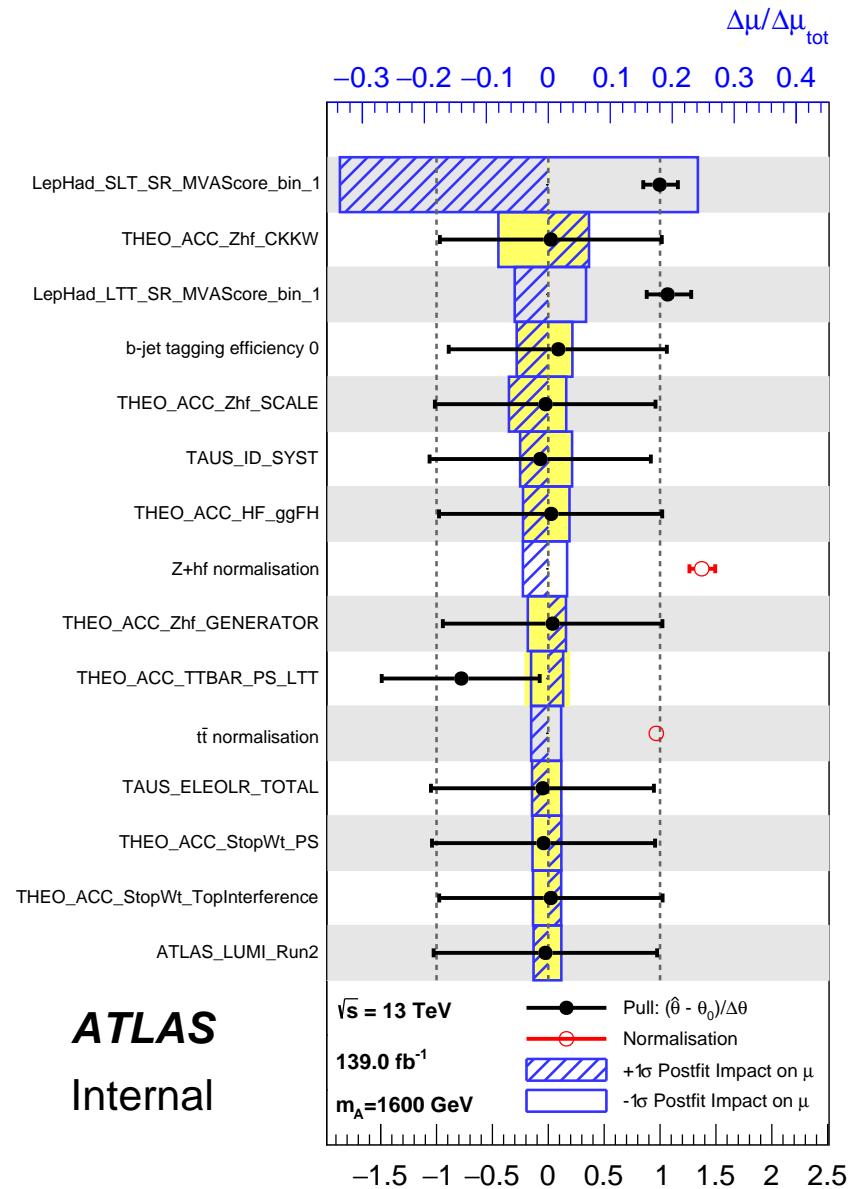
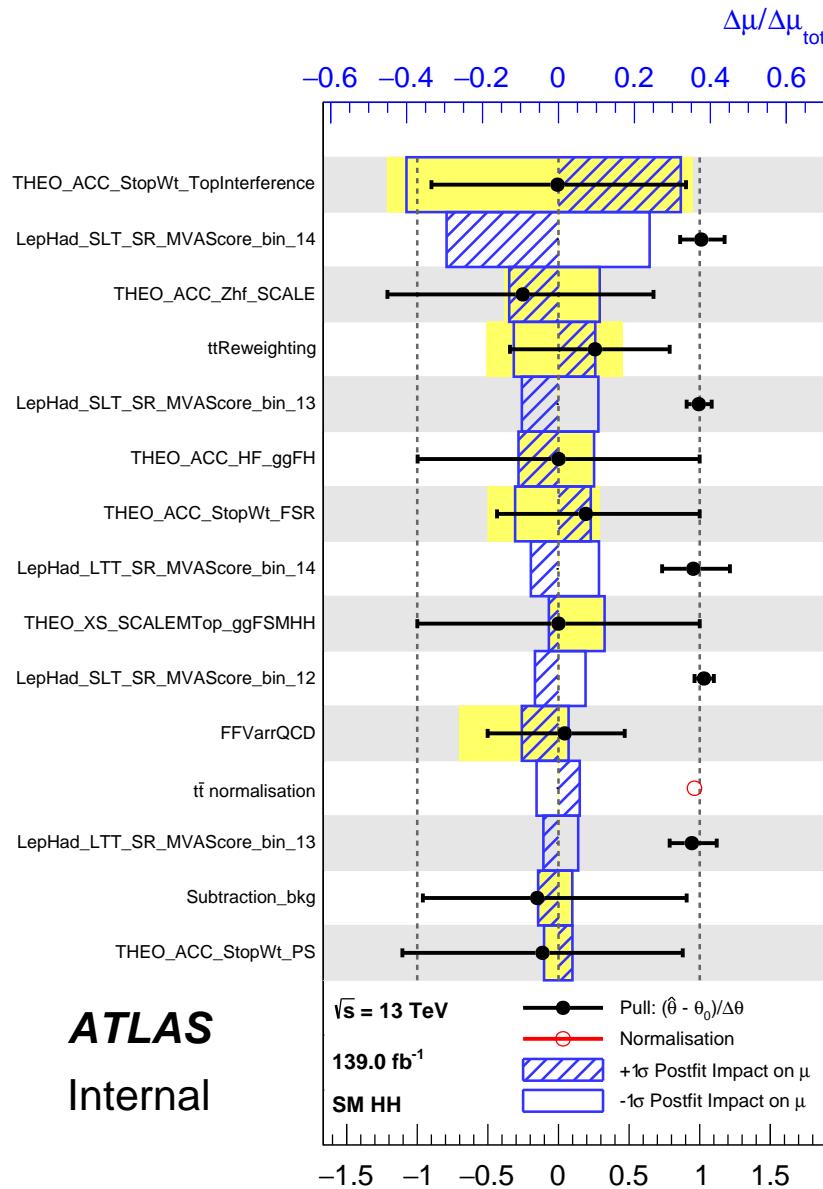


Figure B.48: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit to data for the 1600 GeV resonant fit.

Figure B.49: NP rankings for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit to data for the SM fit.

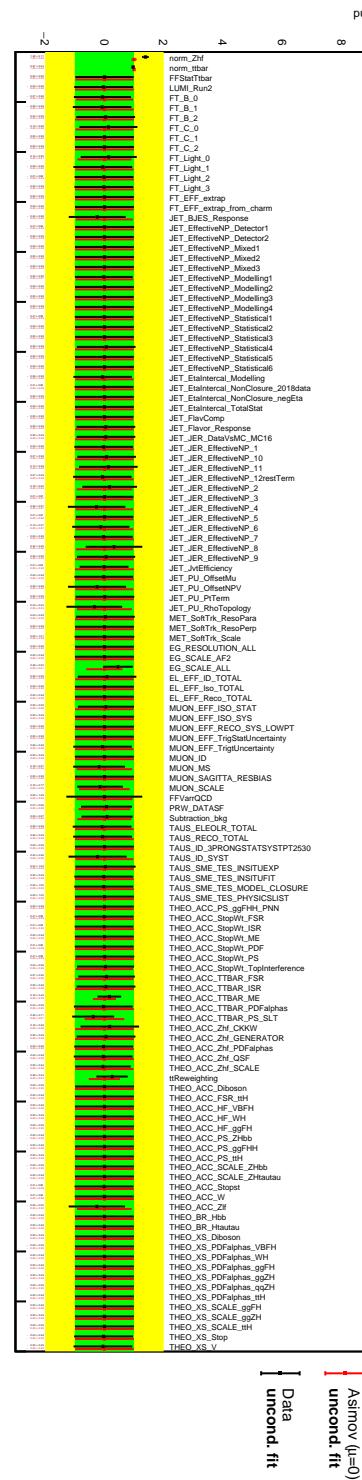


Figure B.50: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 300 GeV resonant fit.

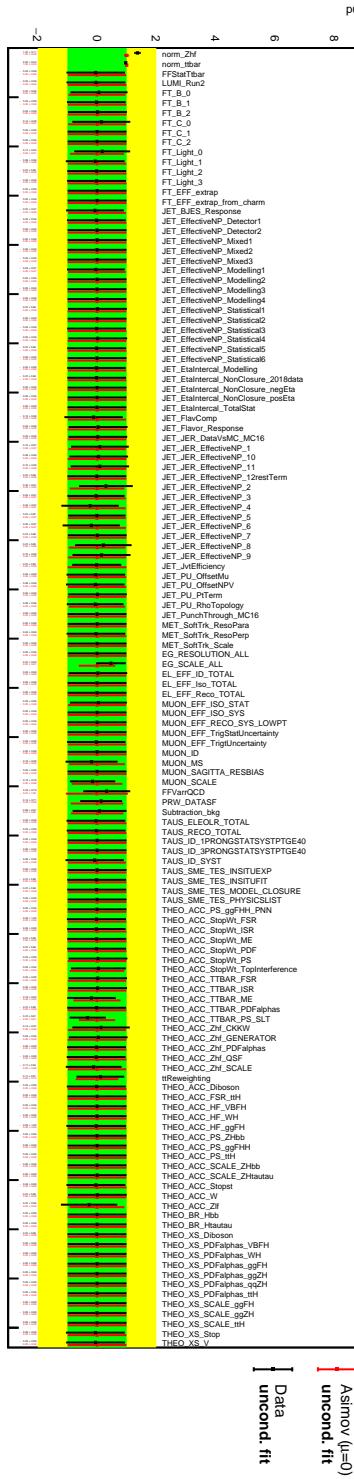


Figure B.51: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 500 GeV resonant fit.

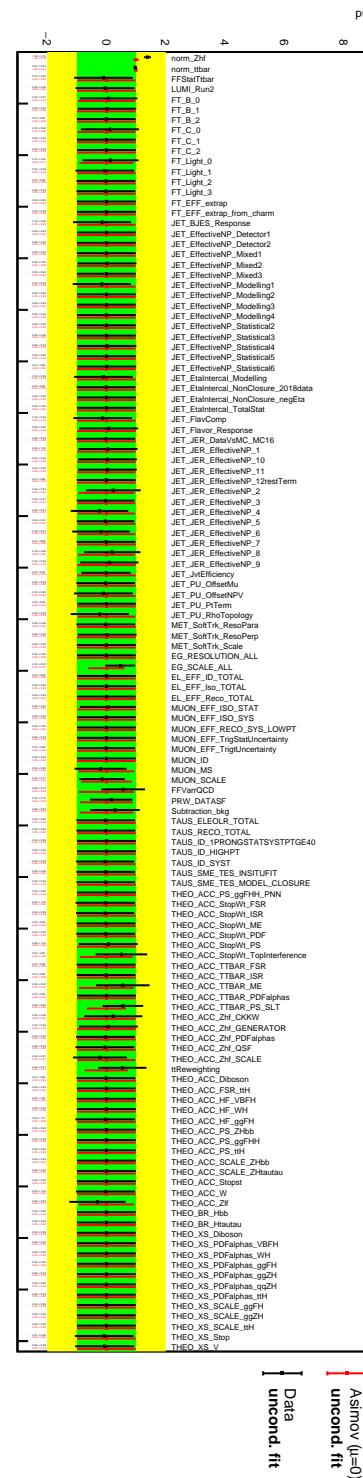


Figure B.52: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 1000 GeV resonant fit.

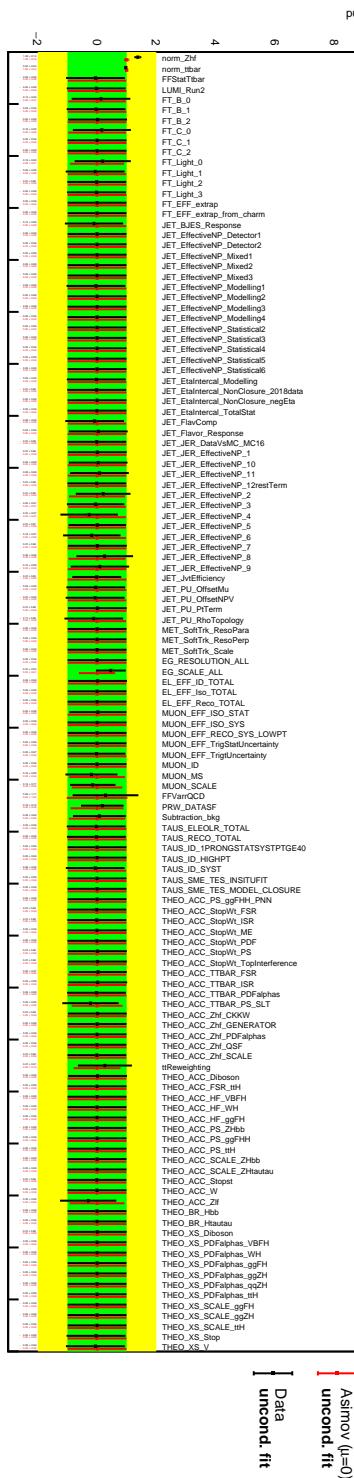


Figure B.53: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 1600 GeV resonant fit.

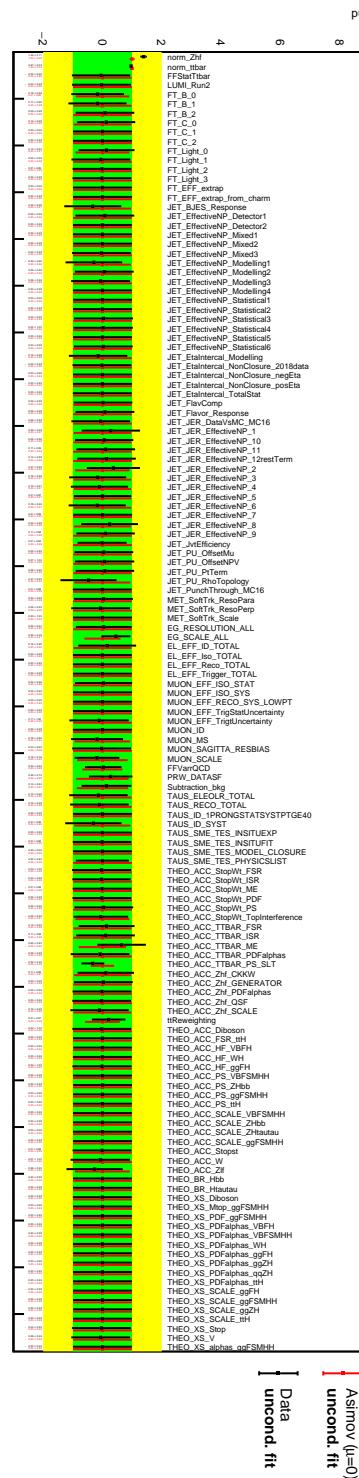


Figure B.54: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  SLT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the SM fit.

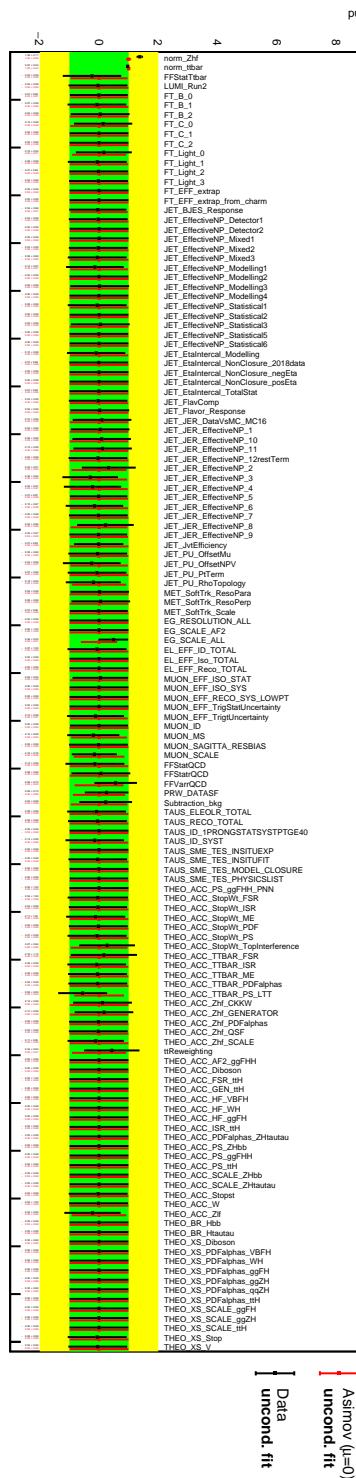


Figure B.55: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 300 GeV resonant fit.

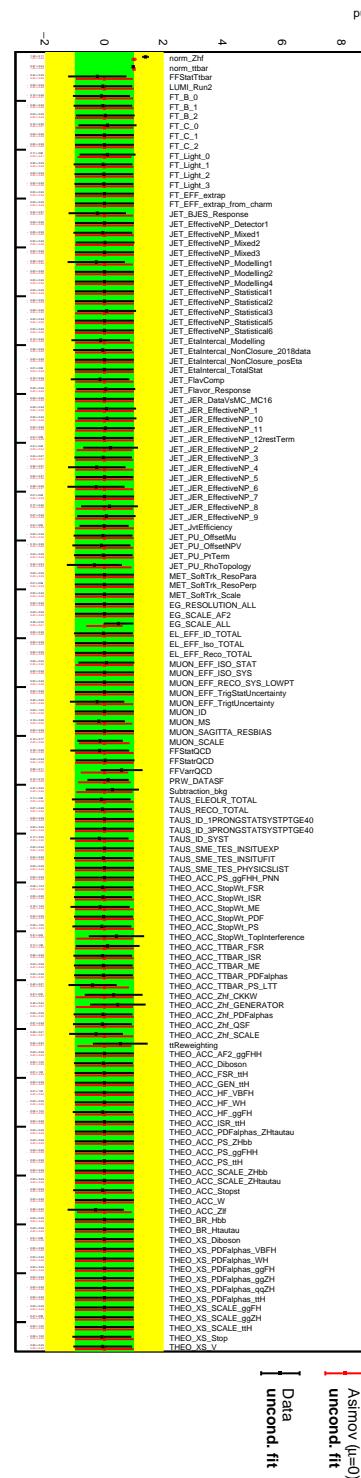


Figure B.56: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 500 GeV resonant fit.

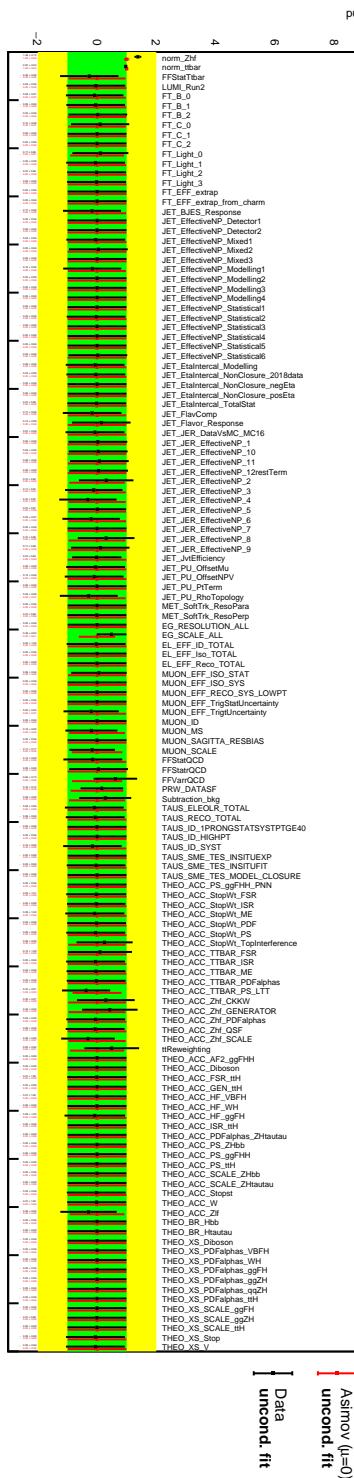


Figure B.57: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 1000 GeV resonant fit.

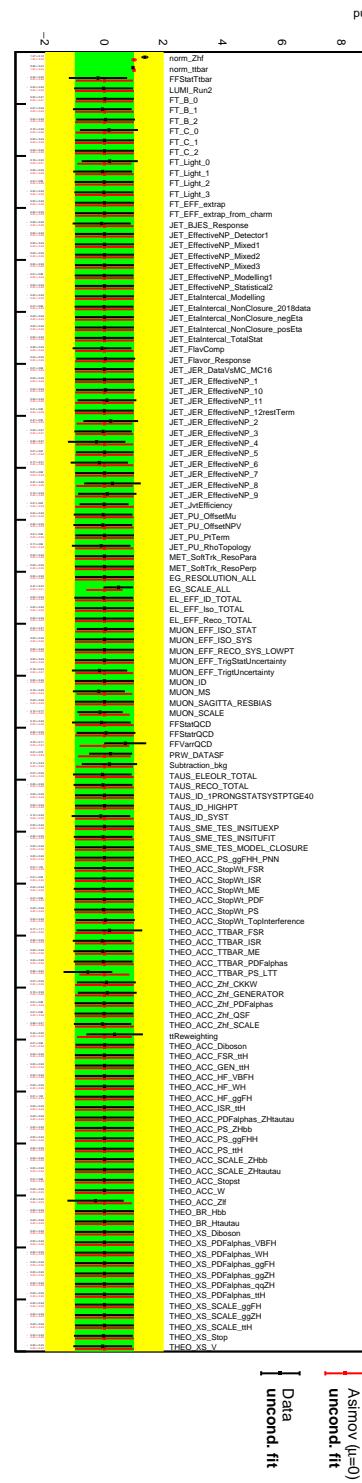


Figure B.58: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 1600 GeV resonant fit.

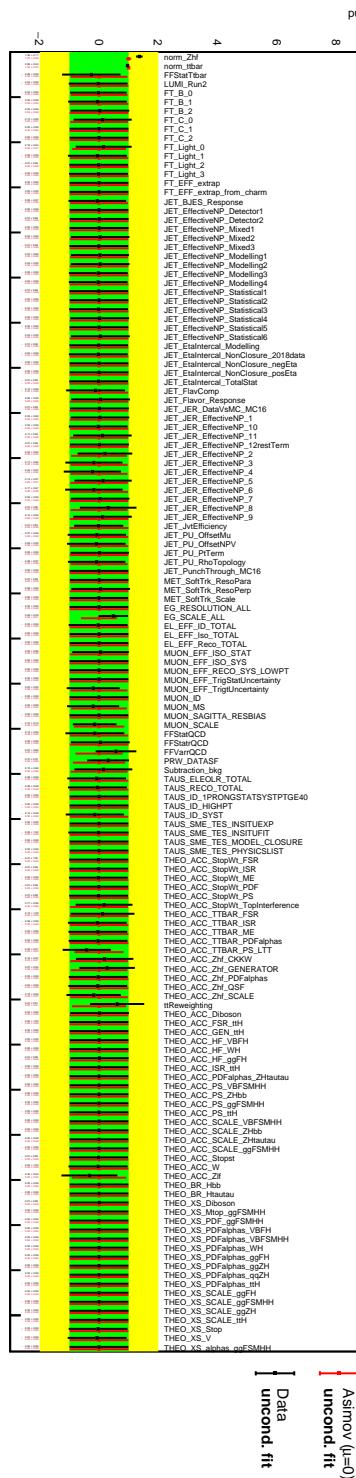


Figure B.59: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  LTT fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the SM fit.

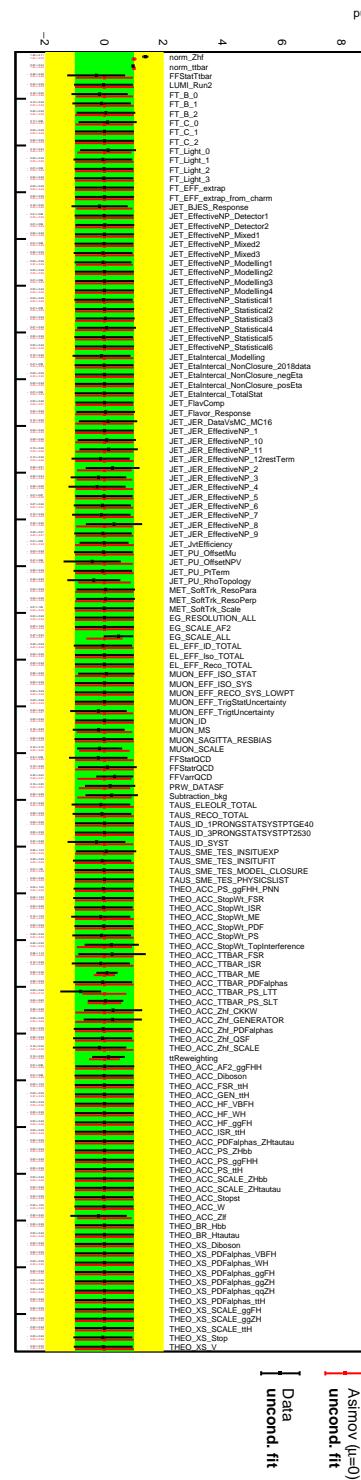


Figure B.60: NP pulls for the di-Higgs  $\tau_{\text{lep}} \tau_{\text{had}}$  fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 300 GeV resonant fit.

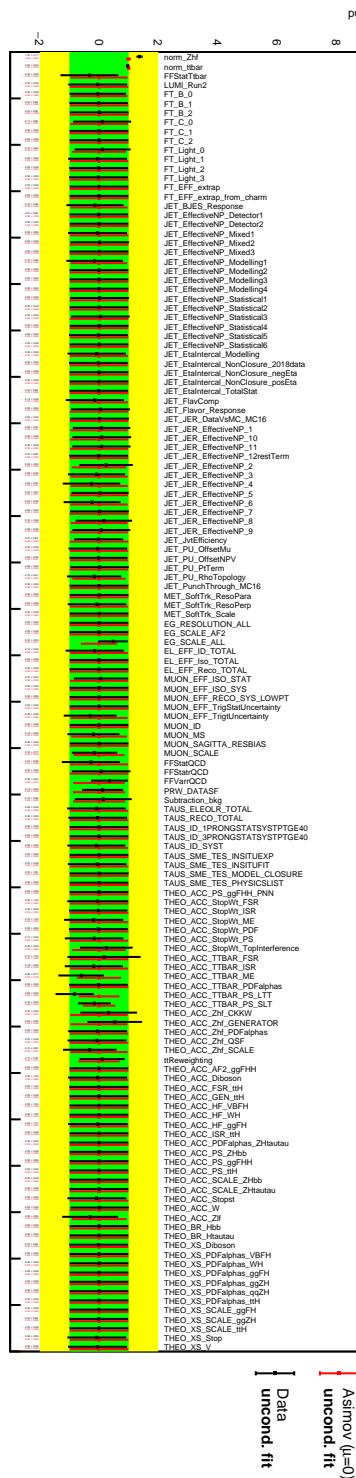


Figure B.61: NP pulls for the di-Higgs  $\tau_{\text{lep}} \tau_{\text{had}}$  fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 500 GeV resonant fit.

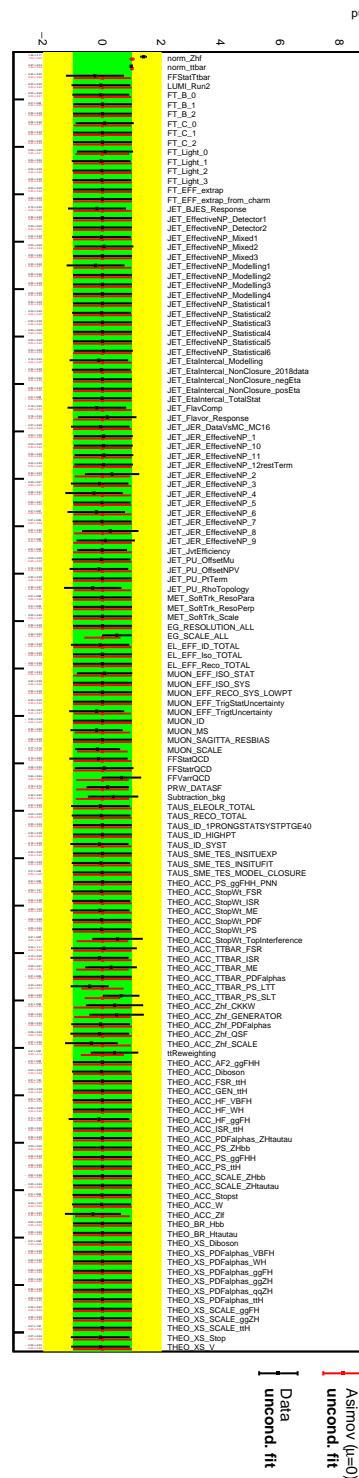


Figure B.62: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 1000 GeV resonant fit.

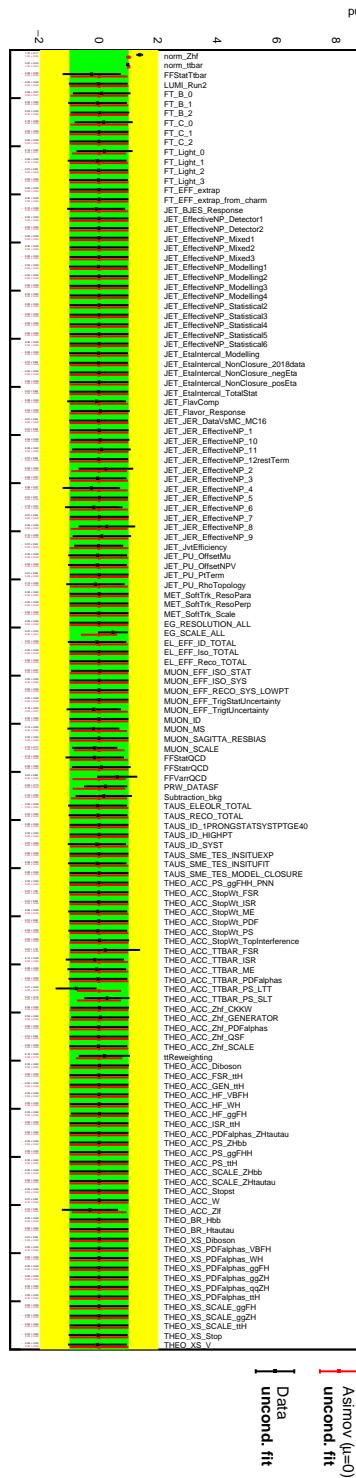


Figure B.63: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the 1600 GeV resonant fit.

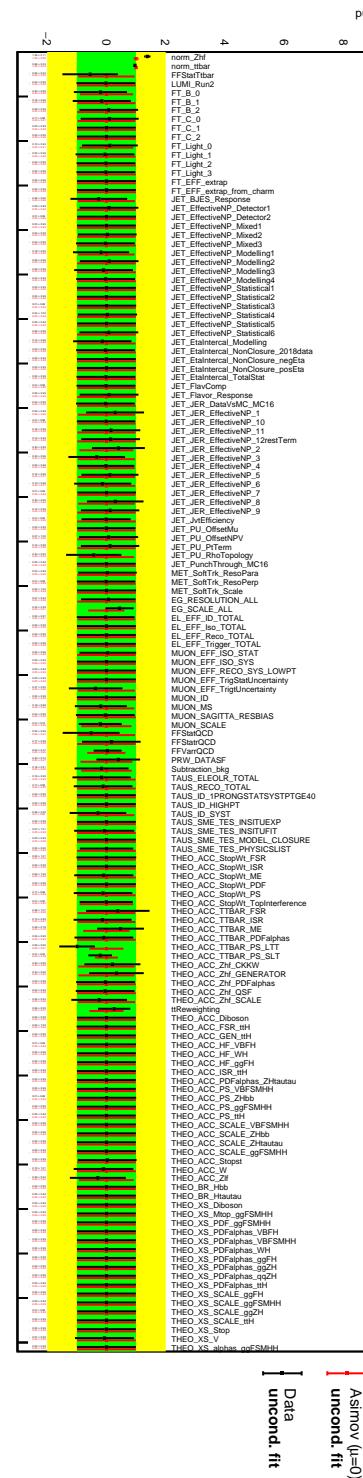


Figure B.64: NP pulls for the di-Higgs  $\tau_{\text{lep}}\tau_{\text{had}}$  fit for the fit to the Asimov dataset built with  $\mu = 0$  (red) and for the fit to data (black) for the SM fit.