

**Search for Diboson Resonances in the Fully Hadronic Final
State Using Jet Substructure Techniques in 8 and 13 TeV
Proton-Proton Collisions with the ATLAS Detector**

THÈSE

présentée à la Faculté des sciences de l'Université de Genève
pour obtenir le grade de Docteur ès sciences, mention physique

par

Chris Malena Delitzsch

d'Allemagne

Thèse N° 5017

CERN-THESIS-2016-236
02/12/2016





**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES

**Doctorat ès sciences
Mention physique**

Thèse de *Madame Chris Malena DELITZSCH*

intitulée :

**"Search for Diboson Resonances in the Fully Hadronic Final
State Using Jet Substructure Techniques in 8 and 13 TeV
Proton-Proton Collisions with the ATLAS Detector"**

La Faculté des sciences, sur le préavis de Monsieur G. IACOBUCCI, professeur ordinaire et directeur de thèse (Département de physique nucléaire et corpusculaire), Monsieur T. GOLLING, professeur associé (Département de physique nucléaire et corpusculaire), Madame A. SFYRLA, professeure associée (Département de physique nucléaire et corpusculaire) et Monsieur A. ANDREAZZA, professeur (Fisica Sperimentale, Università Degli Studi di Milano, Italia), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 2 décembre 2016

Thèse - 5017 -

Le Doyen

Abstract

In this thesis, searches for narrow resonances decaying into pairs of vector bosons (WW , WZ and ZZ) are presented. The searches are performed using 20.3 fb^{-1} and 3.2 fb^{-1} of proton-proton collision data collected with the ATLAS experiment at the LHC at centre-of-mass energies of 8 TeV and 13 TeV, respectively. Events are selected in the fully hadronic final state, characterised by two large-radius jets. Requirements on the jet mass and substructure variables are imposed to select large-radius jets that are compatible with the hadronic decay of highly boosted W or Z bosons. A variety of techniques are compared to optimise the identification of boosted vector bosons. The invariant mass spectrum of the two boson-tagged jets is used as discriminating variable to search for a resonant structure on top of the smoothly falling distribution of the background processes. In the analysis performed at 8 TeV, an excess of events is observed for an invariant dijet mass of 2 TeV. The largest deviation from the background prediction occurs in the WZ channel with a global significance of 2.5σ . On the contrary, no significant deviations from the background expectations are seen in the analysis using 13 TeV pp collisions. Exclusion limits at 95% confidence level are set on the production cross-section times branching ratio for three benchmark models: a bulk Randall-Sundrum model, an extended gauge model and a heavy vector triplet model. W' bosons as predicted in the extended gauge model are excluded in the mass range from 1.3 to 1.5 TeV at 95% confidence level and W' bosons from the heavy vector triplet model are excluded for masses between 1.38 to 1.6 TeV.

Résumé

Dans cette thèse, les recherches sur des résonances étroites, se désintégrant en paires de bosons vectoriels (WW , WZ et ZZ) sont présentées. Les recherches sont effectuées en utilisant les données de collision proton-proton de 20.3 fb^{-1} et 3.2 fb^{-1} , enregistrées par le détecteur ATLAS au LHC avec une énergie au centre de masse de 8 TeV et 13 TeV, respectivement. Les événements sont sélectionnés dans l'état final entièrement hadronique, caractérisé par deux jets de grand rayon. Des contraintes ont été imposées sur la masse du jet et sur sa structure interne pour sélectionner des jets qui sont compatibles avec la désintégration d'un boson de jauge (W ou Z) avec une grande quantité de mouvement transverse. Plusieurs techniques sont comparées afin d'optimiser l'identification des bosons vectoriels boostés. Le spectre de masse invariante des deux jets est utilisé comme variable de discrimination à la recherche d'une structure de résonance au-dessus de la distribution des bruits de fonds. Dans l'analyse effectuée à 8 TeV, un excès d'événements est observé pour une masse invariante de dijet de 2 TeV. La plus grande déviation est produite dans le canal WZ avec une signifiante globale de 2.5σ . Au contraire, aucune déviation significative par rapport aux attentes de fond n'est vue dans l'analyse en utilisant des collisions de proton de 13 TeV. Des limites supérieures au niveau de confiance de 95% sont établies sur la section efficace fois le rapport de branchement pour trois modèles de référence différents: le modèle bulk Randall-Sundrum, une théorie de jauge prolongée et un modèle de triplet de vecteur lourd. Les masses d'un boson W' dans la théorie de jauge prolongée sont exclues entre 1.3 et 1.5 TeV et les masses d'un boson W' prédit par le modèle de triplet de vecteur lourd sont exclues entre 1.38 et 1.6 TeV.

Contents

1	Introduction	1
2	Standard Model of Particle Physics	5
2.1	Introduction to the Standard Model	5
2.1.1	Fermions	5
2.1.2	Interactions	6
2.1.3	Electroweak Symmetry Breaking	9
2.1.4	Shortcomings of the Standard Model	10
2.2	Looking Beyond	12
2.2.1	Extended Gauge Models	13
2.2.2	Heavy Vector Triplets	13
2.2.3	Bulk Randall–Sundrum Model	14
2.2.4	Previous Diboson Resonance Searches	14
3	The ATLAS Experiment at the Large Hadron Collider	17
3.1	The Large Hadron Collider	17
3.2	The ATLAS Detector	20
3.2.1	The Inner Detector	22
3.2.2	The Calorimeter System	24
3.2.3	Muon Spectrometer	27
3.2.4	The Trigger System	28
3.2.5	Simulation	29
4	Jet Reconstruction, Grooming and Substructure Variables in ATLAS	33
4.1	Inputs to Jet Reconstruction	36
4.1.1	Topological Cluster Formation and Local Cluster Calibration	36
4.1.2	Tracks	38
4.2	Jet Reconstruction Algorithms	38
4.3	Jet Mass	42
4.4	Jet Grooming Algorithms	42
4.4.1	Trimming	43
4.4.2	Split-filtering	45
4.4.3	Pruning	47

4.5	Jet Substructure Moments	49
5	Large-R Jet Calibration in ATLAS	53
5.1	Jet Energy and Mass Calibration	54
5.2	Systematic Uncertainties	56
5.2.1	Track-jet Double Ratio Uncertainties	57
5.2.2	Final Jet Scale Uncertainties in 2012 Data	59
5.2.3	Monte Carlo Based Jet Scale Uncertainties for Early 2015 Analyses	60
5.2.4	Final Jet Scale Uncertainties in 2015 Data	63
5.2.5	Summary of Systematic Uncertainties	66
5.3	Future Improvements	68
5.3.1	Jet Area Pile-up Correction	68
5.3.2	Track-assisted Jet Mass	71
6	Identification of Boosted Vector Bosons	75
6.1	Performance of Boosted Boson Identification at $\sqrt{s} = 8$ TeV	75
6.1.1	MC-based Optimisation	76
6.1.2	Performance in Data	81
6.1.3	Summary	84
6.2	Performance of Boosted Boson Identification at $\sqrt{s} = 13$ TeV	86
6.2.1	Summary	92
6.3	Future Improvements to Boosted Vector Boson Tagging	93
7	Common Aspects for Diboson Resonance Searches	97
7.1	Simulation of Signal and Background Samples	98
7.2	Data Sample	99
7.3	Object Definitions	99
7.3.1	Jet Reconstruction and Boson Identification	99
7.3.2	Leptons	107
7.3.3	Missing Transverse Momentum	108
7.4	Event Selection	108
7.4.1	Preselection	108
7.4.2	Topological Selection	109
7.4.3	Signal Regions	110
7.5	Background Model	115
7.6	Systematic Uncertainties	118
7.7	Statistical Analysis	120
7.8	Summary	122

8	Diboson Resonance Search at $\sqrt{s} = 8$ TeV	125
8.1	Results	125
8.1.1	Cross Checks	127
8.2	Statistical Interpretation	131
8.3	Combination with Other Diboson Resonance Searches	131
9	Diboson Resonance Search at $\sqrt{s} = 13$ TeV	137
9.1	Results	137
9.2	Statistical Interpretation	141
9.3	Combination with Other Diboson Resonance Searches	145
9.4	Improvements for Future Diboson Resonance Searches	148
10	Conclusions	149
A	Definition of Substructure Variables	151
B	Prospect Studies for the Vector Boson Associated Higgs Boson Production for the High-luminosity LHC	157
B.1	Signal and Background Samples	158
B.2	Object Definitions	159
B.3	Event Selection	160
B.4	WH Analysis Validation with $\sqrt{s} = 8$ TeV Analysis	161
B.5	HL-LHC Analysis	165
B.6	Systematic Uncertainties	165
B.6.1	Experimental Uncertainties	170
B.6.2	Background Only Uncertainties	170
B.6.3	Signal Only Uncertainties	171
B.6.4	Reduced Systematic Uncertainty Scenarios	171
B.7	Fit Model	172
B.8	Analysis Improvement Assumptions	172
B.9	Results	173
B.10	Discussion and Outlook	175
	Bibliography	177
	Acknowledgement	193

Chapter 1

Introduction

Our current knowledge about the fundamental particles and their gauge interactions is represented in the Standard Model of Particle Physics (SM). Its predictions are in remarkable agreement with the measurements that have been carried out in many experiments. The quest for the last missing piece of the SM, the Higgs boson, ended after about 50 years with its discovery in July 2012. Despite providing such a successful theoretical framework for particle physics, the Standard Model does not address several experimental observations. For example it cannot explain the origin of dark matter or provide a source for the observed matter-antimatter asymmetry. One of the key signatures in the search for physics beyond the Standard Model is the production of new particles decaying to a pair of electroweak gauge bosons (WW, WZ, ZZ). To enable searches for these particles at new unprecedented energy scales, more and more powerful particle accelerators and detectors are built.

The Large Hadron Collider (LHC) at CERN is the world's largest and most powerful particle accelerator, which collides protons close to the speed of light in four different main interaction points. The first collisions at the LHC were recorded in 2009 at a centre-of-mass energy of 900 GeV. Since then, the LHC has constantly enhanced its performance by not only increasing the centre-of-mass energy up to 13 TeV but also by pushing the instantaneous luminosity beyond its design value. The exceptional performance of the LHC allowed for a significant extension of the mass reach in the search for physics beyond the SM with respect to previous high energy particle colliders. However the high instantaneous luminosity also poses challenges in the reconstruction of particles in the detector due to the large number of proton-proton interactions in the same bunch-crossing.

This thesis presents the results of the search for diboson resonances in the fully hadronic final state using 20.3 fb⁻¹ of 8 TeV pp collisions and 3.2 fb⁻¹ of data collected at $\sqrt{s} = 13$ TeV. The unprecedented centre-of-mass energy of the pp collisions at the LHC enables the production of vector bosons for the first time with a transverse momentum p_T much larger than their rest mass m . Hence the decay products of W and Z bosons are collimated, which is also referred to as *boosted*, and overlap within the detector. The hadronic decay products of the boosted vector bosons are captured within one single large-radius jets with typical radius sizes of $R = 1.0$. Thus, the final state of the fully hadronic diboson resonance search is characterised by two large-radius jets. However, the cross-section of the dominating background, the production of

QCD multi-jet events, is orders of magnitude larger than that of the signal. To distinguish between a large-radius jet that contains the two decay products of a W/Z boson from a quark or gluon-initiated jet, so-called tagging techniques are used to explore the internal *substructure* of the jet. Grooming algorithms are used to improve the jet mass resolution by removing the contamination of soft particles unrelated to the hard scattering and produced in additional collisions in the same bunch crossing. A variety of substructure and grooming techniques have been studied in the ATLAS Collaboration to find the optimal combination to identify boosted hadronically decaying vector bosons and these studies are summarised as well in this thesis.

This thesis is organised as follows. The theoretical framework of particle physics, the Standard Model, is introduced in Chapter 2 and theories beyond the Standard Model are presented that predict the existence of new particles decaying into a pair of vector bosons. The Large Hadron Collider and the ATLAS detector are described in Chapter 3. The concept of jets is introduced in Chapter 4 together with an overview of the available grooming and substructure techniques. The calibration procedure and derivation of systematic uncertainties for large-radius jets are described in Chapter 5, followed by a detailed description of the studies performed to identify boosted hadronically decaying W and Z bosons in Chapter 6. The common aspects for the 8 TeV and 13 TeV diboson resonance searches, such as the object definitions and the event selection as well as the strategy of the analysis are detailed in Chapter 7. The results of the diboson resonance search at $\sqrt{s} = 8$ and 13 TeV are presented in Chapter 8 and 9, respectively and concluding remarks can be found in Chapter 10.

Personal Contribution

The ATLAS Collaboration is comprised of about 3000 scientists who are involved in the construction of detector components, the operation of the detector, the analyses of the collected data and many other tasks. The studies presented in this thesis thus rely on the work of many other people in the collaboration and are not performed by only one individual. The major contributions of the author in this thesis are listed below.

Chapter 5: Large- R Jet Calibration in ATLAS

The author derived the jet energy and mass calibration for several jet collections that were studied in the optimisation of boosted boson tagging identification at $\sqrt{s} = 8$ TeV to allow for a fair comparison between different grooming algorithms. The systematic uncertainties on the jet energy and mass scale as well as various substructure variables were estimated by the author using the track-jet double ratio method for different jet collections. The uncertainties were used by all ATLAS analyses with large- R jets in the final state at $\sqrt{s} = 8$ TeV. Furthermore, the Monte Carlo based systematic uncertainties for anti- k_t $R = 1.0$ trimmed jets, used in early Run-II analyses such as the diboson resonance search, were derived by the author. The jet area

pile-up correction was studied by the author for both small- R jets as well as large- R jets in MC simulation and data. In addition, the effect of the correction on large- R jets was studied in the context of the High-luminosity LHC.

Chapter 6: Identification of Boosted Vector Bosons

As previously mentioned, the author derived the MC based jet calibrations and the in-situ systematic uncertainties for various jet collections. Furthermore, the author measured the background efficiency for two different signal efficiency working points in 8 TeV collision data and compared it to the MC prediction. The author is one of two principal analysers that developed the boosted boson identification algorithm used in all 13 TeV ATLAS analysis with boosted W or Z bosons. This includes (but is not limited to) the identification of the optimal grooming algorithm and substructure variable used in combination with the jet mass but also different signal efficiency working points.

Chapter 8: Diboson Resonance Search at $\sqrt{s} = 8$ TeV

The author was responsible for the derivation of the jet calibration as well as several cross checks. The author assured that events were not accumulated in one particular region of the detector and that the selection efficiency was stable over the whole data-taking period. Furthermore several control regions in data were studied to ensure that the kinematic selection does not cause a distortion of the dijet invariant mass spectrum.

Chapter 9: Diboson Resonance Search at $\sqrt{s} = 13$ TeV

The author developed the algorithm to identify boosted hadronically decaying vector bosons and derived the systematic uncertainties for the variables used in the algorithm. Furthermore, the author participated in the implementation and running of the object definitions and event selection and provided the invariant mass spectra needed for the statistical analysis. Several control regions were studied by the author to ensure that the parametric function describes the invariant dijet mass of the background processes.

Appendix B: Prospect Studies for the $VH \rightarrow b\bar{b}$ Production at the HL-LHC

The author is one of two main analysers of the $WH \rightarrow \ell\nu b\bar{b}$ analysis at the HL-LHC and thus contributed to essentially every aspect of the analysis. The studies were performed in close collaboration with the small team working in parallel on the ZH analysis.

Chapter 2

Standard Model of Particle Physics

The *Standard Model of particle physics* (SM) [1–5] provides the theoretical framework for describing the elementary particles and their fundamental interactions. It was mostly developed in the 1960s-1970s and its predictions have been tested with remarkable precision in various experiments, with no significant deviations observed so far. Approximately 50 years after its prediction [6–11], the last missing piece of the SM model, the Higgs boson, was discovered in 2012 by the ATLAS and CMS Collaboration [12, 13]. In this chapter, the SM is briefly summarised and its limitations will be discussed, giving rise to theories of physics beyond the SM. Different theories that predict the existence of new particles that decay into vector boson pairs will be introduced.

2.1 Introduction to the Standard Model

The SM distinguishes between two types of particles: *fermions* and gauge *bosons*. The gauge bosons are the mediators of the interactions between the fermions. Three of the four fundamental interactions are incorporated in the SM: the electromagnetic, weak and strong forces which are based on relativistic quantum field theories. No quantum theory for the gravitational force exist so far and is not included in the SM, however at the current energy scale of elementary particle physics it can be neglected.

2.1.1 Fermions

The fermions are spin one-half particles and can be divided into two different groups: *leptons* and *quarks*. They are arranged into three generations each consisting of two left-handed doublets and three right-handed singlets. The fermion doublets of the SM are shown in Table 2.1 with their corresponding charges. Each generation contains one charged lepton (electron, muon, tau), the corresponding electromagnetically neutral lepton, referred to as neutrino, (ν_e , ν_μ , ν_τ) as well as an *up-type* (up, charm, top) and *down-type* (down, strange, bottom) quark. The fermions of the first generation are the building blocks of the visible matter and the second and third generations are duplicates of the first generation with the same quantum numbers but larger masses. The number of generations is not predicted in the SM, however measurements of the lifetime of the Z boson, the mediator of the neutral weak interaction, constrain the number of

	Generations			
	I	II	III	Q
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	0 +1
Quarks	$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	+2/3 -1/3

Table 2.1: The fermions in the Standard Model, arranged into three generations with their charge Q given in units of the proton charge.

Lepton	Mass [MeV]	Quark	Mass [MeV]
e	0.511	u	$2.3^{+0.7}_{-0.5}$
ν_e	$< 2 \cdot 10^{-6}$	d	$4.8^{+0.5}_{-0.3}$
μ	105.658	c	$(1.275 \pm 0.025) \cdot 10^3$
ν_μ	< 0.19 (90% CL)	s	95 ± 5
τ	1776.86 ± 0.12	t	$(173.21 \pm 0.51 \pm 0.71) \cdot 10^3$
ν_τ	< 18.2 (95% CL)	b	$(4.18 \pm 0.03) \cdot 10^3$

Table 2.2: Summary of the fermion masses in the SM [15]. The electron and muon masses are measured with a very high precision, thus their uncertainties are neglected here. For the neutrinos masses, the standard hierarchy is assumed.

light neutrinos ($m_\nu < m_Z/2$) to be 2.9840 ± 0.0082 [14]. For each fermion a corresponding anti-fermion exists with the same mass and spin but opposite charge and weak isospin. Quarks carry a further quantum number *colour*, an analogy to the electric charge, which can be either red, blue or green. The masses of the fermions are summarised in Table 2.2. In the SM, neutrinos have zero mass, however they have been measured to be very small. Since quarks do not exist as free particles (as explained in Chapter 2.1.2), their masses are estimated in the $\overline{\text{MS}}$ scheme except for the top quark whose mass is determined from its decay products.

2.1.2 Interactions

The three fundamental interactions of the Standard Model can be derived from the local gauge invariance of the Lagrangian under the transformation of the symmetry group $\text{SU}(N)$. The symmetry group of the Standard Model is $\text{SU}(3)_C \otimes \text{SU}(2)_L \otimes \text{U}(1)_Y$ and combines the strong force ($\text{SU}(3)_C$) with the *electroweak* interaction ($\text{SU}(2)_L \otimes \text{U}(1)_Y$). Each $\text{SU}(N)$ ($\text{U}(N)$) symmetry group has $N^2 - 1$ (N^2) generators which also corresponds to the number of mediators. The indices of the symmetry group indicate the quantum number that is conserved, the colour C for the strong interaction and the weak isospin I and hyper charge Y for the electroweak interaction.

The index L implies that only left-handed fermions are involved in the $SU(2)$ group whereas both chiralities are present in the $U(1)$ group.

Quantum Chromodynamics

The underlying gauge theory of the strong interaction is Quantum ChromoDynamics (QCD) [16–18]. The corresponding symmetry group $SU(3)_C$ is generated by eight three-dimensional Gell-Mann matrices λ_a and the mediators of the strong force are eight colour-charged massless gluons of spin one. Only colour-charged objects, quarks and gluons, interact via the strong interaction. The Lagrangian of the QCD is defined as

$$\mathcal{L}_{\text{QCD}} = \sum_f \bar{q}_f (i\gamma^\mu D_\mu + m_f) q_f - \frac{1}{4} G_a^{\mu\nu} G_{\mu\nu}^a, \quad (2.1)$$

where the first term describes the kinetic term of the quarks and their interaction with the gluons. The second term of the Lagrangian which is quadratic in the gluon field strength $G_a^{\mu\nu}$, defined in Eq. 2.3, gives rise to three- and four-gluon self-interactions. The quark fields are described by Dirac spinors q_f with masses m_f and the sum runs over the number of quark flavours f . The covariant derivative is defined as

$$D^\mu = \partial^\mu - ig_s \frac{\lambda^a}{2} G_a^\mu, \quad (2.2)$$

with the coupling parameter g_s of the strong interaction and the eight gluon fields G_a^μ . Finally, given the QCD structure constants f^{abc} , the gluon field strength can be specified:

$$G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu + g_s f^{abc} G_b^\mu G_c^\nu. \quad (2.3)$$

The coupling parameter of the strong force is often also expressed in terms of the coupling constant $\alpha_s = \sqrt{4\pi g_s}$. Due to finite loop corrections, the strong coupling constant is energy dependent and can be expressed as follows:

$$\alpha_s(Q^2) = \frac{12\pi}{(33 - 2n_f) \log(\frac{Q^2}{\Lambda_{\text{QCD}}^2})}, \quad (2.4)$$

where n_f is the number of quark flavours at a certain energy scale Q and Λ_{QCD} a commonly chosen scale such that $\alpha_s(Q^2)$ depends only on one variable. The constant Λ_{QCD} is about 200 MeV. The behaviour of the coupling constant is determined by the gluon self-interaction of the non-Abelian $SU(3)$ group. At large energy scales ($Q^2 > \Lambda_{\text{QCD}}^2$) the coupling constant α_s decreases. Therefore, quarks can be treated as free particles at short distances which allows the usage of perturbation theory in QCD. This effect is referred to as *asymptotic freedom*. On the other hand, at low energy scales, the coupling constant increases and results in the *confinement* of quarks and gluons into bound states. Therefore quarks cannot be observed as free particles

but form colourless *hadrons* that can be classified as either *mesons* ($q\bar{q}$) or *baryons* (qqq or $\bar{q}\bar{q}\bar{q}$). Gluons cannot exist as colour-singlets as this would otherwise result in an infinite range of the strong interaction. The observation of the Λ^{++} baryon which corresponds to three up quarks with their spin aligned in the same direction lead to the introduction of the additional colour quantum number for quarks to ensure that the Λ^{++} baryon obeys the Fermi-Dirac statistics. The number of different colours for quarks were determined from the ratio of the $e^+e^- \rightarrow q\bar{q}$ cross-section measurement with respect to the $e^+e^- \rightarrow \mu^+\mu^-$ cross-section [19].

Electroweak Interactions

In 1967, Glashow, Weinberg and Salam proposed that the electromagnetic and weak interaction are different manifestations of the same interaction and formulated the *electroweak* theory based on the group $SU(2)_L \otimes U(1)_Y$ [3–5]. The non-Abelian group $SU(2)_L$ is generated by the three Pauli matrices σ_i , with $i = 1, 2, 3$ and consists of an isotriplet of vector bosons $W_\mu^{1,2,3}$. The hypercharge Y is the generator of the Abelian $U(1)$ group which contains only an isosinglet B_μ . The mediators of the electroweak interaction, the $W^\pm$, Z^0 and photon γ are linear combination of the gauge boson fields. The charged gauge bosons $W^\pm$ are defined as

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) , \quad (2.5)$$

whereas the mass eigenstates of the neutral photon field (A_μ) and Z^0 gauge boson field (Z_μ) can be written as

$$A_\mu = W_\mu^3 \sin \theta_W + B_\mu \cos \theta_W \quad (2.6)$$

$$Z_\mu = W_\mu^3 \cos \theta_W - B_\mu \sin \theta_W . \quad (2.7)$$

The mixing of the neutral states depends on the *Weinberg angle* θ_W which can be determined from the coupling constant g and g' of the $SU(2)_L$ and $U(1)_Y$ group:

$$\sin \theta_W = \frac{g'}{\sqrt{g^2 + g'^2}} . \quad (2.8)$$

The $W^\pm$ and Z^0 bosons are the mediators of the weak force and are both spin-one particles. In the SM, gauge bosons are required to be massless to preserve its local gauge invariance. However, the masses of the $W^\pm$ and Z^0 bosons deviate significantly from zero and are 80.376 ± 0.033 GeV and 91.1876 ± 0.0021 GeV [15], respectively. In 1964, the *Brout-Englert-Higgs* mechanism was introduced in which the gauge bosons acquire their masses by the spontaneous breaking of electroweak gauge symmetry, see Chapter 2.1.3. Due to the large masses of the gauge bosons, the weak force has a limited range and dominates only at high energies.

The massless photon is the mediator of the electromagnetic interaction which has an infinite range. As the other gauge bosons, the photon has a spin of one. It only couples to electrically

charged particles and thus no photon self-interactions are possible. Therefore, the electromagnetic coupling constant α_{em} decreases at low energies due to the screening of the electric charge. The weak interaction is the only interaction that allows for flavour changing charged currents, e.g. an up-type quark is turned into a down-type quark via the emission of a W boson. The transition does not only have to occur within the same generation, however the probability for a transition between two generations are much smaller. The unitary 3×3 *Cabibbo-Kobayashi-Maskawa* (CKM) matrix relates the weak eigenstates of the down-type quarks q' to the mass eigenstates q . Its squared matrix elements $|V_{ij}|^2$ represent the probability of a quark of flavour i to decay into a quark of flavour j via the emission of a W boson.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.986 & 0.041 \\ 0.008 & 0.040 & 0.999 \end{pmatrix} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.9)$$

The mixing of mass eigenstates is not restricted to the quark sector, it is also present for leptons. The *Pontecorvo-Maki-Nakagawa-Sakata* (PMNS) matrix, an extension to the SM, describes the mixing between the neutrino flavour eigenstates ν_α ($\alpha = e, \mu, \tau$) and the mass eigenstates ν_i .

2.1.3 Electroweak Symmetry Breaking

The mediators of the weak interaction acquire mass via the *spontaneous symmetry breaking* in which the Lagrangian is invariant under $\text{SU}(2)_L \otimes \text{U}(1)_Y$ but the ground state is not. To accomplish the symmetry breaking, a new self-interacting isospin doublet of complex scalar fields with four degrees of freedom is introduced

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix}, \quad (2.10)$$

and the most general, renormalisable Lagrangian of a scalar field with its kinetic term and the potential V is added to the electroweak Lagrangian:

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi) \quad \text{with} \quad V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \quad (2.11)$$

The covariant derivative D_μ for left-handed particles is defined as

$$D_\mu = \partial_\mu + ig \frac{\sigma_i}{2} \cdot W_\mu^i + i \frac{g'}{2} Y B_\mu. \quad (2.12)$$

The parameter λ is required to be larger than zero to guarantee stable minima and $\mu^2 < 0$ is chosen to obtain a non-vanishing vacuum expectation value $v = \sqrt{\mu^2/\lambda}$ which causes the symmetry breaking of $\text{SU}(2)_L \otimes \text{U}(1)_Y$. The ground state is then chosen such that the gauge group of the electromagnetic interaction is unaffected and the photon remains massless. In the unitary gauge, the Higgs doublet Φ can be obtained by perturbative expansion around the

ground state and becomes

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \quad (2.13)$$

with the real scalar Higgs field H that corresponds to the Higgs boson. Inserting Eq. 2.13 in the scalar field Lagrangian in Eq. 2.11 yields the mass terms for the gauge bosons:

$$\mathcal{L}_{\text{Higgs}} = \underbrace{\frac{1}{2}\partial_\mu H \partial^\mu H}_{\text{kinetic term}} + \underbrace{\frac{1}{4}g^2 v^2 W_\mu^+ W^{-\mu} + \frac{1}{8}(g^2 + g'^2)v^2 Z_\mu Z^\mu - \lambda v^2 H^2}_{\text{mass terms}} - \underbrace{\lambda v H^3 - \frac{1}{4}H^4}_{\text{Higgs self-coupling}}. \quad (2.14)$$

Terms of the order $\mathcal{O}(HW^+W^-, HZZ, HHW^+W^-, HHZZ)$ which represent the coupling of the Higgs field to the massive gauge bosons have been neglected in Eq. 2.14. Three of the four free degrees of freedom of the complex Higgs scalar are absorbed as longitudinal modes of the $W^\pm$ and Z^0 boson and generate their masses which can be purely expressed in terms of the vacuum expectation value $v \approx 246$ GeV and the coupling constants of the electroweak interaction, g, g' :

$$m_{W^\pm} = \frac{1}{2}gv, \quad m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}. \quad (2.15)$$

The remaining degree of freedom gives rise to a new scalar boson, the Higgs boson with $m_H = \sqrt{2\lambda}v$. Its mass is not predicted by the theory since the parameter λ is not known.

The masses of the fermions in the SM are not a priori included in the Brout-Englert-Higgs mechanism, however the fermions can acquire mass via the Yukawa coupling of the Higgs doublet to them. The strength of the coupling g_f is proportional to the mass of the fermion m_f :

$$m_f = \frac{g_f \cdot v}{\sqrt{2}}. \quad (2.16)$$

Almost 50 years after its prediction, the SM Higgs boson was discovered in 2012 by the ATLAS and CMS Collaboration at the Large Hadron Collider using the data taken in 2011 at a centre-of-mass energies of $\sqrt{s} = 7$ TeV and a partial dataset of about $5 - 6 \text{ fb}^{-1}$ taken in 2012 at $\sqrt{s} = 8$ TeV. The mass of the Higgs boson was measured by the ATLAS and CMS Collaboration with high precision, $m_H = 125.09 \pm 0.21 \text{ (stat)} \pm 0.11 \text{ (syst)} \text{ GeV}$ [20], and the spin two hypothesis was excluded with a confidence level above 99.9% by the ATLAS Collaboration [21]. The new particle was discovered through its decay into boson pairs, $\gamma\gamma, WW$ and ZZ , and so far only evidences for the Higgs boson decay into fermions were found. To estimate the sensitivity to the WH and ZH production with the subsequent Higgs decay into $b\bar{b}$, a simulation-based analysis was conducted under the conditions that are expected for the planned high-luminosity upgrade of the Large Hadron Collider. The analysis is described in Appendix B.

2.1.4 Shortcomings of the Standard Model

The SM of particle physics predicts the experimental observations over a broad range with remarkable success as shown in Fig. 2.1. Nonetheless, several experimental observations cannot

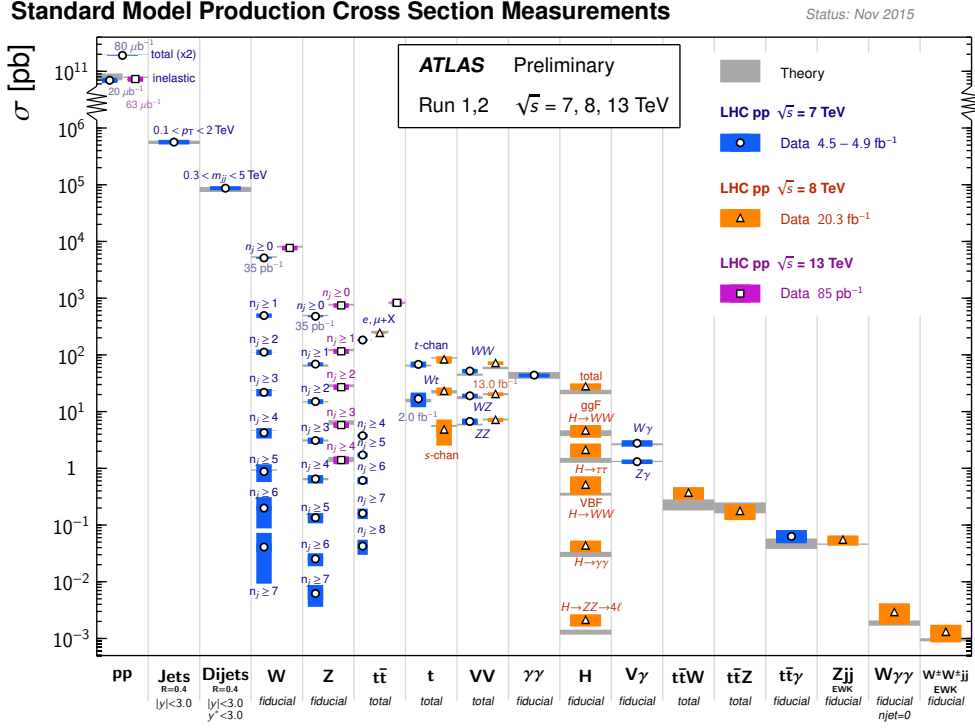


Figure 2.1: Summary of the Standard Model production cross-section measurements at $\sqrt{s} = 7, 8, 13$ TeV and comparison with their theory predictions [22].

be explained by the SM and require an extension of the theoretical framework. A selection of its limitations will be introduced in the following.

The observation of neutrino oscillations [23, 24] conceded that neutrinos are not massless as in the SM and thus as previously mentioned, the neutrino flavour eigenstates differ from their mass eigenstates. To allow for neutrinos with non-zero masses in the SM, two different models of neutrinos exist: Dirac and Majorana neutrinos. In the first case, right handed neutrinos are added to the SM Lagrangian whereas Majorana neutrinos represent as well their own antiparticle. The nature of neutrinos has not been determined yet. However if neutrinos are their own antiparticles, then neutrinoless double β -decays should be observable [25]. While the relative mass differences between generations has been measured, the absolute neutrino masses are beyond the sensitivity of modern experiments.

An unanswered question of the SM is the source of the matter-antimatter asymmetry. During the Big Bang, an equal amount of matter and antimatter was produced. Nowadays, the amount of antimatter in the visible universe is negligible. The CP violation possible through the weak interaction and described by a phase in the CKM matrix is too small to explain this asymmetry. The measurement of the rotation curves of galaxies and the observation of gravitational lensing revealed that only about 4.6% of the energy of the universe is accounted for by baryonic matter [26]. The remaining 95% of the universe energy density is *dark matter* and *dark energy*

which are not included in the SM. Dark matter does not emit electromagnetic radiation and so far only its coupling to the gravitational force has been detected. The only possible dark matter candidate in the SM are neutrinos, however their almost negligible mass and abundance cannot explain the large amount of dark matter.

The unification of the electromagnetic and weak interaction gave rise to *Grand Unified Theories* (GUT) [27, 28], which regard the strong and electroweak interaction as different manifestations of one underlying interaction described by a new gauge group such as SU(5). The coupling constants of the three interactions converge then at the grand unification scale of about 10^{16} GeV and are identical in strength. In the SM, the coupling constants do not converge in one single point. Most grand unified theories predict as well the decay of the proton, which has never been observed.

The tree-level mass of scalars, such as the recently discovered Higgs boson, receives large corrections from virtual loops involving all fermions and gauge bosons to which it couples. Considering only fermion loops, the corrections to the Higgs boson mass are at first order:

$$\Delta m_{\text{H}}^2 = -\frac{|y_f|^2 \Lambda^2}{8\pi^2}, \quad (2.17)$$

where y_f is the Yukawa coupling of the Higgs boson to the fermion and Λ is the cut-off scale up to which the SM is valid. Due to its large mass, the top quark induces the largest loop corrections. Assuming that the SM is the only theory up to the Planck scale with $\Lambda \approx 10^{19}$ GeV, the bare mass of the Higgs boson has to be extremely *fine-tuned* to result in the observed Higgs mass value of about 125 GeV. The enormous difference of the electroweak scale $\mathcal{O}(100 \text{ GeV})$ and the Planck scale is also referred to as *hierarchy problem*. Although this fine-tuning is in principle no problem of the SM, it seems rather unnatural that the cancellation of the loop corrections occur with a precision of the order of 10^{-16} .

2.2 Looking Beyond

Over the past decades, many theories beyond the SM (BSM) have been developed to accommodate the aforementioned limitations of the SM. So far, no BSM theory is able to solve all shortcomings. One of the most elegant theories that solves the hierarchy problem, unifies the gauge couplings and provides a dark matter candidate is *SUPERSymmetry* (SUSY) which introduces a supersymmetric partner to each SM particle [29]. The remainder of this thesis will focus however not on one specific theory but a distinct signature that can be produced by many different BSM theories. One possible signature is the production of vector boson pairs (WW , WZ , ZZ) resulting from the decay of a new particle predicted in extensions of the SM such as in technicolour [30–32], warped extra dimensions [33–35], and Grand Unified Theories [27, 28, 36]. The final state of the diboson production is characterised by the subsequent decay of the vector bosons. The search in this thesis is performed in the fully hadronic mode where both vector bosons decay hadronically making use of the large branching ratio, $BR(W \rightarrow q\bar{q}') \approx 68\%$ and

$BR(Z \rightarrow q\bar{q}) \approx 69\%$ compared to leptonic decays: $BR(W \rightarrow \ell\nu) \approx 10.8\%$, $BR(Z \rightarrow \ell\ell) \approx 3.4\%$ and $BR(Z \rightarrow \nu\nu) \approx 20\%$. The higher branching ratio allows to extend the reach of the search to higher resonance masses with respect to the fully leptonic or semileptonic decays. The invariant diboson mass spectrum is then compared to the predictions of the SM and the results are interpreted in terms of three benchmark BSM models: an extended gauge model [37], a specific heavy vector triplet model [38] and the bulk Randall–Sundrum model [33]. The different models and their theoretical motivation will be introduced in this section.

2.2.1 Extended Gauge Models

The extended gauge model is a general model that does not aim to solve one of the previously described shortcomings of the SM but that predicts the existence of heavier spin-one vector bosons W' and Z' similar to the $W^\pm$ and Z^0 boson of the SM. In the search presented in this thesis, only the W' model will be considered. The couplings of the W' boson to fermions, $W'q\bar{q}$ and $W'\ell\ell$, are assumed to be the same as in the SM. The triple-gauge-boson coupling $W'WZ$ however is suppressed by a factor $c \cdot (m_W^2/m_{W'}^2)$ arising from the mixing of the new gauge boson and the SM vector bosons. The width of the W' thus increases linearly with its mass and is approximately 3.5% of the resonance mass. Without the suppression factor (with $c = 1$), the width would increase with $m_{W'}^5$. The $q\bar{q}$ production of the W' causes a low-mass tail in the W' mass spectrum due to off-shell production. The effect is more pronounced for high W' masses.

2.2.2 Heavy Vector Triplets

The heavy vector triplet (HVT) model that is considered in the diboson search is based on a simplified phenomenological Lagrangian [38]. The simplified model was chosen because resonance searches are only sensitive to parameters that retain the mass as well as the coupling strength of the interaction with other particles. Furthermore this approach allows to describe a large class of models. As opposed to the extended gauge models, only the on-shell production of the new resonance is considered in the simplified model and thus the low-mass tails are removed. A new vector triplet is introduced with one charged (W') and one neutral (Z') heavy spin-one particle which mixes with the SM vector bosons. The coupling of the new vector triplet to the fermions, either via Drell-Yan production or in the decay, is given by $(g^2/g_V^2) \cdot c_F$, where g is the coupling constant of the $SU(2)_L$ gauge group of the weak interaction, g_V the coupling strength to the new vector boson and c_F is a free parameter of the order of one. The coupling to the SM vector bosons and the Higgs boson is controlled by the parameter c_H and is significantly smaller than one in weakly coupled scenarios. The HVT model A [38], an extension of the SM gauge group, was chosen as benchmark model here to interpret the results. The coupling strengths for this explicit model are $c_F \sim 1$ and $c_H \sim -g^2/g_V^2$. The width of the W' and Z' in this model is approximately 2.5% of the resonance mass, while the branching ratio into WW or WZ is about 2% for each channel.

2.2.3 Bulk Randall–Sundrum Model

The Randall–Sundrum (RS) model aims at providing a solution to the hierarchy problem by embedding our four-dimensional spacetime in a larger dimensional bulk with one single warped extra dimension [33]. In the original model, the SM particles were only allowed to be contained in the four-dimensional spacetime whereas the gravitational interaction could propagate through the bulk. However, this model introduced large contributions to flavour-changing neutral current processes and observables in electroweak precision measurements which are in contradiction with their current limits and measurement. The RS model was then extended to allow as well the SM gauge bosons and fermions to propagate in the additional dimension and is referred to as bulk RS model. The spacetime metric of the bulk RS model depends as well on the coordinate of the extra dimension and the hierarchy problem is solved by introducing an exponential warp factor:

$$ds^2 = e^{-2kr_c\phi} \eta_{\mu\nu} dx^\mu dx^\nu + r_c^2 d\phi^2, \quad (2.18)$$

where $\eta_{\mu\nu}$ is the Minkowski metric, x^μ the four-dimensional coordinate, k an energy scale, ϕ the coordinate of the extra dimension ($0 \leq \phi \leq \pi$) and r_c is the radius of the curvature of the warped extra dimension. The Higgs field is constrained to the “TeV brane” with $\phi = 0$ whereas the gravitational interaction is mostly localised at the “Planck brane” with $\phi = \pi$. To generate the large hierarchy between the TeV and Planck branes, no large radius of the warped extra dimension is needed ($kr_c \approx 11 - 12$) due to the exponential warp factor. The benchmark models used in ATLAS require $k/\bar{M}_{\text{Pl}} = 1.0$, where $\bar{M}_{\text{Pl}}$ is the reduced Planck mass. The bulk RS model addresses as well the mass hierarchy of the fermions; the larger the Yukawa coupling of the fermion to the Higgs bosons, the closer its localisation to the TeV brane. Furthermore the bulk RS model would result in the unification of the gauge coupling constants [39].

Each SM particle propagating through the bulk results in Kaluza-Klein (KK) excitations [40, 41] with masses at the TeV scale. For the massless graviton, the corresponding excited spin-2 KK gravitons G^* are close to the TeV scale and thus their decay into fermions is suppressed and is dominated by the decay into top-quark and Higgs boson pairs as well as WW and ZZ . The branching ratio for the WW and ZZ decay are approximately 18.7% to 16% and 9.5% to 8% respectively depending on the resonance mass and the width is about 6%.

2.2.4 Previous Diboson Resonance Searches

Searches for diboson resonances have been performed previously by the CDF [42] and DØ [43] experiments at the Tevatron Collider at Fermilab as well as the ATLAS and CMS Collaboration at the LHC. In this section, only the searches performed at the LHC with a centre-of-mass energy of 7 TeV are presented. The 8 and 13 TeV diboson resonance searches are summarised in Section 8.3 and 9.3, respectively. The previous searches have been mostly explored in the fully leptonic or semileptonic decay modes because leptons are well-understood objects and provide clean signatures. Furthermore the fully hadronic decay channel is dominated by the

overwhelming dijet background and only the higher centre-of-mass energy at the LHC allowed the development of new reconstruction techniques that results in an enormous suppression of the dijet background. Here, the searches where at least one vector boson decayed hadronically will be summarised.

The CDF Collaboration performed a search for WW and WZ resonances in the $e\nu jj$ final state based on an integrated luminosity of 2.8 fb^{-1} collected at a centre-of-mass energy of 1.96 TeV [44]. No deviations from the SM expectations were observed and the result interpreted in terms of the Randal–Sundrum graviton G^* and the extended gauge model W' . The former could be excluded for masses below 606 GeV and the latter is excluded in the mass range from 284 to 515 GeV . Furthermore, the CDF Collaboration searched for ZZ resonances in three different final states $\ell\ell jj$, $\ell\ell\ell\ell$ and $\ell\ell\nu\nu$ using a dataset corresponding to 6 fb^{-1} at $\sqrt{s} = 1.96 \text{ TeV}$ [45]. Only the $\ell\ell\ell\ell$ final state showed an excess of events, whereas the more sensitive searches didn't show any deviations from the expectation. The channels were then combined to achieve the best sensitivity. Upper limits on the production cross-section times branching ratio are set for graviton masses between 300 GeV and 1000 GeV and vary between 0.26 pb and 0.0045 pb . The DØ Collaboration excluded W' masses in the range from 180 to 690 GeV and gravitons with masses between 300 and 754 GeV at 95% confidence level in the $WZ \rightarrow \ell\ell jj$ and $WW \rightarrow \ell\nu jj$ final state respectively, using 5.4 fb^{-1} of $p\bar{p}$ collision data at $\sqrt{s} = 1.96 \text{ TeV}$ [46].

The ATLAS Collaboration explored the $WW/WZ \rightarrow \ell\nu jj$ final state [47] using 4.7 fb^{-1} of pp collisions at $\sqrt{s} = 7 \text{ TeV}$. For the extended gauge model W' boson, masses below 950 GeV could be excluded at 95% confidence level, exceeding the limits from the Tevatron. For the bulk RS graviton masses below 710 GeV were excluded. The CMS Collaboration searched for VZ resonances [48] where the Z boson decays into lepton pairs and the V into two overlapping jets and for the first time as well as for the fully hadronic decay of VV resonances [49]. For the semileptonic search, W' masses between 700 and 940 GeV as well as graviton masses between 750 and 880 GeV were excluded assuming $k/\bar{M}_{\text{Pl}} = 0.05$. The fully hadronic search however was only sensitive to set upper limits on the production cross-section times branching ratio. The novel techniques used for the reconstruction of overlapping hadronic decay products are introduced in Chapter 6.

Chapter 3

The ATLAS Experiment at the Large Hadron Collider

3.1 The Large Hadron Collider

In December 1994, the construction of the *Large Hadron Collider* (LHC) [50–52] was approved. The LHC is a circular proton-proton (pp) collider with a circumference of 27 km. It is the world’s largest and most powerful particle accelerator built by humankind with a design centre-of-mass energy of 14 TeV. The LHC is situated in the previous tunnel of the *Large Electron Positron* collider based at CERN (*Conseil Européen pour la Recherche Nucléaire*) in Geneva, Switzerland. Apart from protons, lead ions can be used in collisions however the remainder of this thesis will focus on proton-proton collisions.

Before being injected in the LHC and brought to collision, the protons are pre-accelerated in a chain of different accelerators. The full accelerator complex at CERN is depicted in Fig. 3.1. The protons are extracted from a simple bottle of hydrogen gas by stripping off their electrons in an electric field and are accelerated to an energy of 50 MeV in the linear accelerator LINAC2. They are then brought to an energy of 1.4 GeV in the Proton Synchrotron Booster before being transferred to the Proton Synchrotron (PS) which accelerates the protons to an energy of 25 GeV. In the Super Proton Synchrotron (SPS) the protons reach an energy of 450 GeV and are then injected into the two beam pipes of the LHC. Eight radio frequency cavities accelerate the beam, consisting of several hundreds of *bunches* then to its final energy of 3.5 TeV (2010-2011), 4.0 TeV (2012), 6.5 TeV (2015) or the design energy of 7 TeV. 1232 superconducting dipole magnets with a maximum magnetic field strength of 8.33 T assure that the protons are kept on their circular path.

The circulating hadrons are brought to collision at four different interaction points where the main LHC experiments ALICE, ATLAS, CMS and LHCb are located. Superconducting quadrupole magnets, cooled down to a temperature of 1.9 K with liquid helium, are used to focus the beams before the interaction points. The two general-purpose detectors ATLAS [54] and CMS (*Compact Muon Solenoid*) [55] investigate a wide range of physics topics, including precision measurements of the Standard Model and searches for physics beyond the SM. ALICE (*A Large Ion Collider Experiment*) [56] was built to investigate the quark-gluon plasma created

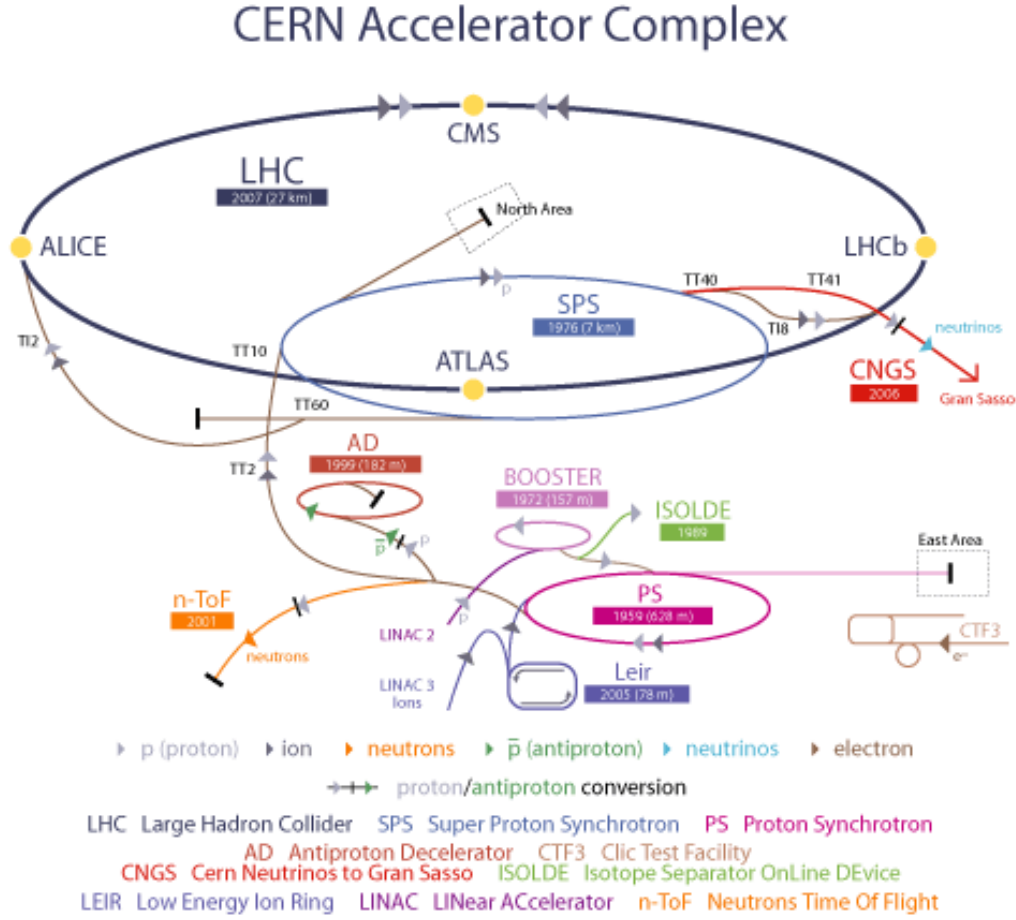


Figure 3.1: Graphical illustration of the accelerator complex at CERN with its flagship, the Large Hadron Collider and the different pre-accelerators used for the proton and heavy ion injection [53].

in heavy ion collisions (proton-lead or lead-lead) and LHCb [57] studies amongst others the matter-antimatter asymmetry in the b -physics sector.

One of the main goals of the LHC was the discovery of the SM Higgs boson. The predicted number of events per second N with a Higgs boson is directly proportional to its cross-section σ and is relatively small compared to other SM processes:

$$\frac{dN}{dt} = \sigma \cdot L, \quad (3.1)$$

where L is the instantaneous luminosity determined by the beam parameters. The production rate of interesting events can then be enhanced by increasing the luminosity which depends on the number of bunches n_b , the number of protons per bunch N_1, N_2 , the revolution frequency f

Parameter	2010	2011	2012	2015	Design
Beam energy [TeV]	3.5	3.5	4.0	6.5	7.0
Bunch spacing [ns]	150	75, 50	50	50, 25	25
Max. number of bunches	368	1380	1380	2232	2808
Max. number of protons per bunch [$\times 10^{11}$]	1.2	1.45	1.7	1.21	1.15
Peak luminosity [$\text{cm}^{-2}\text{s}^{-1}$]	2.1×10^{32}	3.7×10^{33}	7.7×10^{33}	5.02×10^{33}	1×10^{34}

Table 3.1: Summary of the beam condition parameters during data-taking from 2010 to 2012 and in 2015 compared to the design values.

and the overlapping area of the colliding bunches A :

$$\begin{aligned}
 L &= \frac{n_b N_1 N_2 f}{A} \\
 &= \frac{n_b N_1 N_2 f}{4\pi\sigma_x\sigma_y}.
 \end{aligned} \tag{3.2}$$

The second equation in Eq. 3.2 holds only for Gaussian-shaped beams with equal vertical and horizontal beam sizes, σ_x and σ_y , for the two colliding bunches. To increase the luminosity, the number of bunches and the maximum number of protons per bunch have been increased whereas the bunch spacing and the bunch size have been reduced sequentially since the beginning of data-taking in 2010. The significant beam parameters to determine the luminosity for data-taking in 2010 to 2015 are summarised in Tab. 3.1. Furthermore, the design peak luminosity is given which has been surpassed for the first time in June 2016. The peak luminosity is obtained at the beginning of the fill since the instantaneous luminosity decreases during the run due to the beam losses caused by the collisions.

The accumulation of more protons in the bunches to increase the instantaneous luminosity results in the occurrence of multiple pp interactions in the same bunch crossing and is referred to as *in-time pile-up*. These additional interactions are uncorrelated with the hard-scatter interaction and are considered as background. The in-time pile-up is characterised by the number of reconstructed vertices N_{vtx} as measured in the tracking detectors.

Out-of-time pile-up describes the impact of the signal of adjacent bunch crossings on the signal of the current bunch crossing. It occurs mostly due to the short bunch spacing compared to the readout time of the detector systems such as the ATLAS liquid-argon calorimeter (LAr), described in Section 3.2.2. The average number of interactions per bunch crossing $\langle\mu\rangle$ is used to parameterise the amount of out-of-time pile-up and is calculated from the luminosity. The distribution of $\langle\mu\rangle$ for 2012 and 2015 data-taking is shown in Fig. 3.2. Both in-time and out-of-time pile-up will be referred to as pile-up in the remainder of this thesis. The suppression of pile-up plays an important role in the reconstruction of final states including hadrons and will be discussed in Chapter 5.3.1.

The integrated luminosity is the amount of data collected over a certain period of time, $\mathcal{L} = \int L dt$. For the data-taking in 2012 and 2015, the integrated luminosity as accumulated over

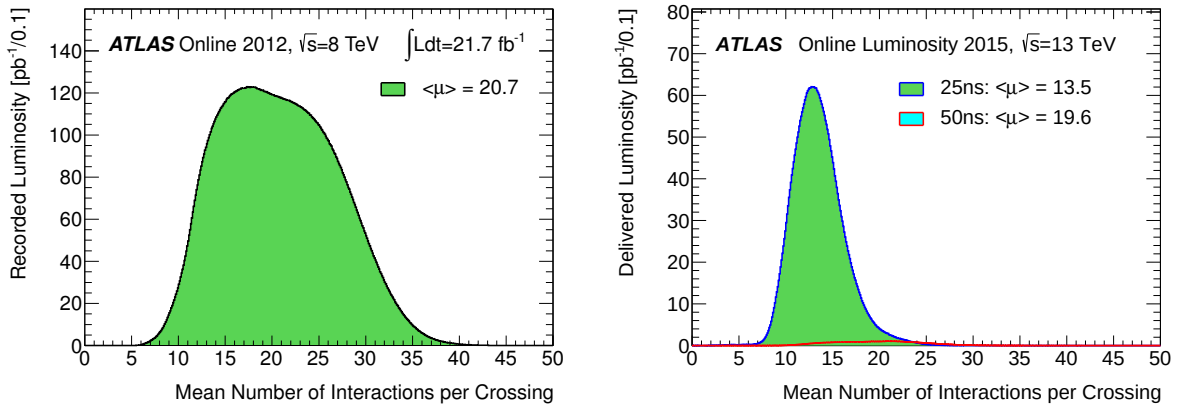


Figure 3.2: Average number of interactions per bunch crossing $\langle\mu\rangle$ in 2012 (left) and 2015 (right) [58, 59].

time is shown in Fig. 3.3. In 2012, the LHC delivered 22.8 fb^{-1} of 8 TeV proton-proton collisions and 4.2 fb^{-1} at a centre-of-mass energy of 13 TeV in 2015. The expected number of events for a certain process in the given dataset can then be calculated as $N = \sigma \cdot \mathcal{L}$. In the remainder of the thesis, the data-taking period before the long shutdown in 2013 will be referred to as Run-1, whereas the data-taking after the upgrade will be denoted as Run-2.

3.2 The ATLAS Detector

The ATLAS detector is one of the multi-purpose detectors at the LHC and enables precision measurements of SM parameters and searches for new physics beyond the SM. It consists of four major components which are arranged in the typical onion-shell-like structure: the *Inner Detector*, the *calorimeter system*, the *muon spectrometer* and two magnet systems consisting of a *solenoid* and *toroid* magnet. The layout of the ATLAS detector with its subcomponents is shown in Fig. 3.4. It is forward-backward symmetric around the interaction point with a length of 44 m and a diameter of 25 m. The total weight of the detector is approximately 7000 tons. The design of the ATLAS detector was driven by the requirement of identifying a large spectrum of particles and their tracks and energies with high precision. Furthermore the detector has to withstand high radiation doses and energies produced by large particle multiplicities due to the high luminosity of the LHC.

Coordinate System

The ATLAS detector uses a right-handed coordinate system whose origin is placed in the interaction point. The x -axis points towards the centre of the LHC, the y -axis upwards and the z -axis along the beam pipe. Positions within the detector are well-defined by the coordinates

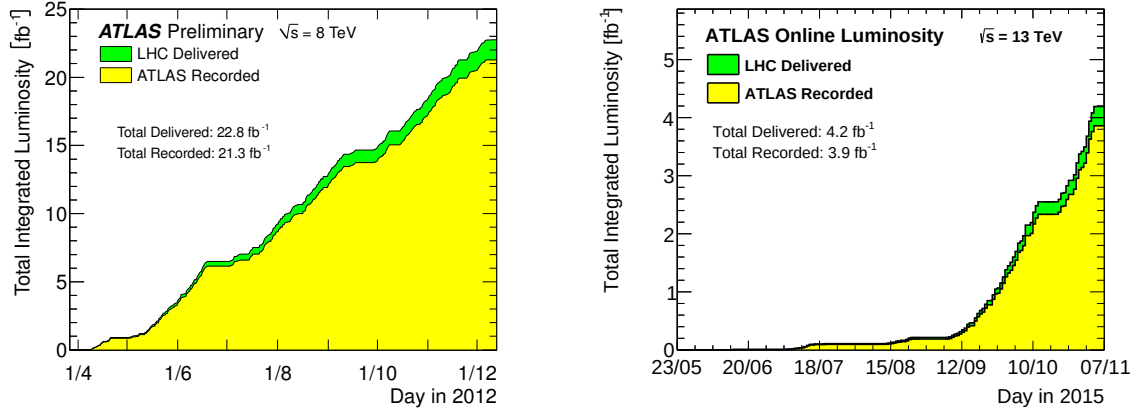


Figure 3.3: The delivered integrated luminosity by the LHC, recorded with the ATLAS detector as a function of data-taking time in 2012 (left) and 2015 (right) [58, 59].

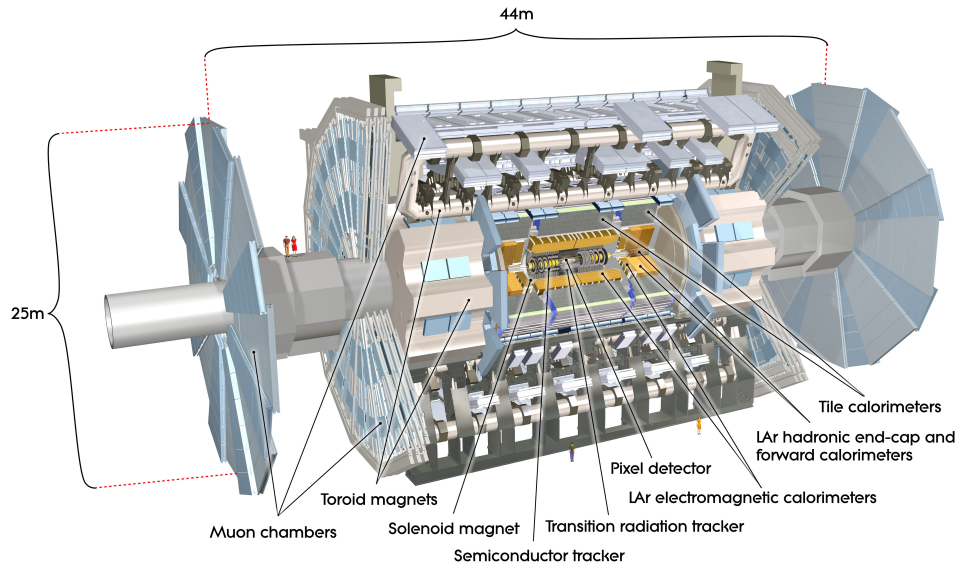


Figure 3.4: Schematic view of the ATLAS detector and its subcomponents [54].

(ϕ, η, z) , where ϕ is the azimuthal angle in the xy -plane. The *pseudorapidity* η is defined as

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

where the polar angle θ is measured between the momentum of the particle and the beam-axis. The pseudorapidity, an approximation of the rapidity y for particles with small masses, is used as ydifferences $\Delta\eta$ are invariant under Lorentz transformations. The momentum of particles in the xy plane, the so-called *transverse momentum* p_T is given by

$$p_T = \sqrt{p_x^2 + p_y^2} ,$$

where p_x and p_y are the momenta in the x and y direction, respectively. Distances between particles are defined in the η - ϕ -plane according to

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} .$$

3.2.1 The Inner Detector

The Inner Detector is the innermost component of the ATLAS detector, only a few centimetres away from the beam-pipe with a length of 6.2 m and a diameter of 2.1 m. During each collision, about 1000 particles emerge from the interaction point. To allow for a precise measurement of the trajectories of charged particles and to provide an excellent momentum resolution, the Inner Detector is required to have a fine granularity and as little material as possible in order to minimally affect the energy measurement in the calorimeter. Furthermore the Inner Detector was designed to both reconstruct primary vertices and secondary vertices. Their precise measurement is of utmost importance to distinguish the hard-scatter vertex from pile-up vertices and for the identification of b -quarks respectively. To achieve these requirements, the Inner Detector consists of three subsystems: the *pixel detector*, the *semiconductor tracker* (SCT) and the *transition radiation tracker* (TRT). The layout of the Inner Detector is illustrated in Fig. 3.5. The Inner Detector is surrounded by a thin superconducting solenoidal magnet with a field strength of 2 T which bends the trajectory of charged particles to measure their momenta.

The pixel detector is the innermost part of the Inner Detector and is comprised of three cylindrical barrel layers and three perpendicular oriented disc layers in the end-cap region. The high resolution is achieved by 1744 silicon sensors consisting of 47232 pixels each with a size of $50 \times 400 \mu\text{m}^2$ and a thickness of $250 \mu\text{m}$ that are readout individually. The active area of the pixel detector is thus approximately 1.7 m^2 . The pixel detector measures the trajectory of charged particles, denoted as *tracks* in the pseudorapidity range of $|\eta| < 2.5$ and for $p_T > 500 \text{ MeV}$. When a charged particle traverses the detector, electron-hole pairs are created and an electric field is used to collect the charge which is then transformed into a signal if the pulse-height exceeds a certain threshold.

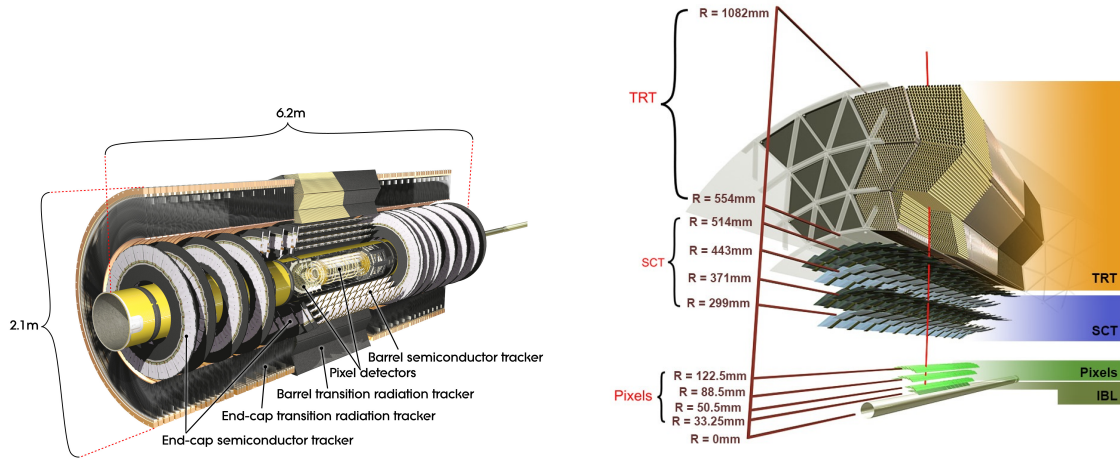


Figure 3.5: Schematic view of the ATLAS Inner Detector and its three subdetectors (left). The transverse view of the Inner Detector (right) shows already the updated layout of the pixel detector as it was used for Run-2 of the LHC [54].

Due to its proximity to the beam-pipe, the pixel detector is exposed to a very high radiation dose resulting in a decrease of its performance. During the long shutdown starting in 2013, the size of the beam-pipe was reduced to allow for an insertion of a fourth barrel layer at a distance of 33.25 mm from the interaction point, the *insertable b-layer* (IBL) [60]. After the insertion of the IBL, the impact parameter resolution, a parameter needed for the identification of b -quarks, improved by about 40% for tracks with $p_T < 1$ GeV.

The semiconductor tracker (SCT) consists of four radial layers in the barrel covering the pseudorapidity range $|\eta| \lesssim 1.4$ and nine end-cap discs on either side covering $1.4 < |\eta| < 2.5$. Each layer consists of single-sided silicon microstrip detectors which are mounted back-to-back with an angle of 40 mrad to improve their spatial resolution. The detection technique of the SCT relies on the same principle as for the pixel detector, however long microstrips are used compared to pixels because of the smaller particle density. The SCT comprises of 4088 modules, adding to a total of 15912 microstrip sensors with a length of 6.4 cm and a strip pitch of 80 μm .

The transition radiation tracker is the outermost part of the Inner Detector and consists of about 300000 polyimide straw tubes, 4 mm in diameter, which are filled with a gas mixture of 70 % Xe, 27 % CO₂ and 3% O₂. In the barrel, the straws are 144 cm long and aligned parallel to the beam-axis, thus they don't provide information on the z -position of the traversing particle. In the end-caps, the straws are arranged radially in wheels with a length of 37 cm. With about 36 hits per track, the TRT allows for continuous tracking within $|\eta| < 1.0$ for the barrel region and $1.0 < |\eta| < 2.0$ in the end-caps. The TRT does not only provide a measurement of the trajectory of traversing particle but can be also used to distinguish between electrons and pions. When a charged particle traverses regions with different dielectric constants, photons are emitted. The rate of emitted photons depends on the characteristics of the traversing particle. Due to their

relatively small mass, electrons tend to produce a larger amount of emitted photons compared to heavier particles such as pions.

3.2.2 The Calorimeter System

The energy of particles, except for muons and neutrinos, is measured in the *electromagnetic* and *hadronic* calorimeters. Muons are minimally ionising particles and deposit only a small amount of their energy in the calorimeter and thus their momentum is measured in the outermost part of the ATLAS detector, the muon spectrometer, as is introduced in Section 3.2.3. Neutrinos on the other hand interact only so weakly with the detector material such that they are reconstructed from the momentum imbalance in the transverse plane.

The ATLAS calorimeters are sampling calorimeters with alternating layers of absorbing and active material. When particles traverse the calorimeter, the dense absorbing material induces the particle to interact and create a cascade of secondary less energetic particles, referred to as showers. The type of interaction depends on the initial particle. In the electromagnetic calorimeter, mostly electrons and photons induce electromagnetic showers via Bremsstrahlung or e^+e^- pair-production at high energies and via ionisation and the photoelectric effect at low energies. Hadronic showers, initiated by hadrons from e.g. hadronic τ -lepton decays are usually longer than electromagnetic showers and more complex. Hadronic showers can be broken down into four different components: escaped energy mostly due to neutrinos produced in the cascade, *invisible energy* and hadronic and electromagnetic energy deposits. The term *invisible energy* refers to the loss of energy needed for the excitations of an atom or corresponds to the binding energy that is required to fragment an atom. A large fraction of the deposited energy is of electromagnetic origin when neutral mesons such as π^0 and η^0 are produced in the hadronic shower which decay into photon-pairs, $\pi^0, \eta^0 \rightarrow \gamma\gamma$. The energy of the secondary particles is measured in the active material, while the energy deposits in the absorber material are missed. To allow for a precise energy measurement, it is important that the hadronic showers are fully contained in the calorimeter and the amount of energy escaping the calorimeter (punch-through) is minimised.

The calorimeters are placed between the solenoid and toroid magnets, with the hadronic calorimeter surrounding the electromagnetic calorimeter. Both calorimeters consist of several independent sub-systems covering the barrel, end-cap and forward regions covering $|\eta| < 4.9$. The layout of the ATLAS calorimeter system is illustrated in Fig. 3.6. The electromagnetic calorimeter was designed to measure the energy of electrons and photons with high precision. A figure of merit is the energy resolution which is parameterised as

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C . \quad (3.3)$$

The first term denoted S , the stochastic term, accounts for the intrinsic fluctuations of the number of particles in the shower evolution. The term N in Eq. 3.3 describes the effect of

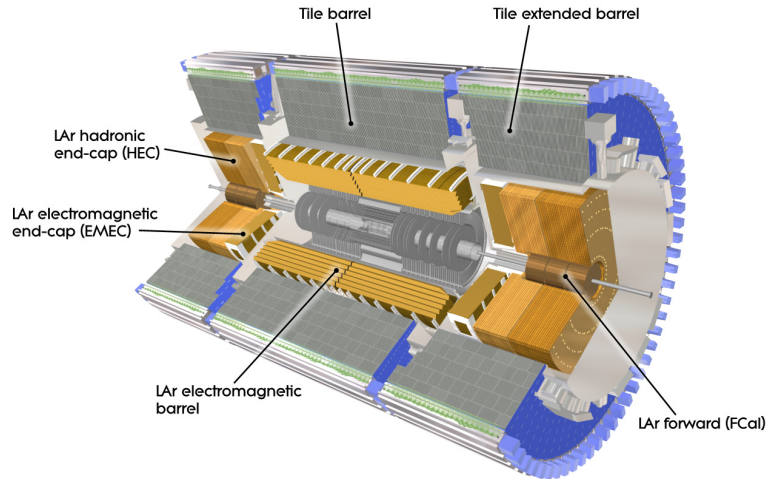


Figure 3.6: Schematic overview of the components of the electromagnetic and hadronic calorimeter in ATLAS [54].

pile-up noise and noise from readout electronics on the energy resolution and is independent of the jet energy. The constant term C accounts for systematic effects due to miscalibration of the detector as well as dead detector material. At low energies, the energy resolution is limited by the pile-up noise whereas at high energy, the constant term is limiting the performance. The electromagnetic (hadronic) calorimeter has been designed to achieve an energy resolution with $N = 10\%$ (50%) and $C = 0.7\%$ (3%).

The Electromagnetic Calorimeter

The electromagnetic calorimeter is divided into a barrel region (EMB) and two end-caps (EMEC) divided into two coaxial wheels, covering $|\eta| < 1.475$ and $1.375 < |\eta| < 3.2$ respectively. One additional electromagnetic calorimeter layer in the forward calorimeter extends the coverage up to $|\eta| < 4.9$. The transition region between the EMB and EMEC ($1.375 < |\eta| < 1.52$), known as the crack region, contains dead material in the form of services for the Inner Detector and are thus removed from most physics analyses using electrons or photons. Lead is chosen as absorber due to its high density and liquid Argon (LAr) as active material based on its radiation-hardness and intrinsic linear behaviour. They are arranged in an accordion geometry to provide full ϕ coverage and to avoid cracks at the same time. The electromagnetic calorimeter has a total thickness of at least 22 and 24 radiation lengths (X_0) in the barrel and end-caps respectively depending on the η range. The EMB is segmented into three longitudinal layers as shown in Fig. 3.7. The first layer has an excellent granularity in the η dimension to provide a precise measurement of the η position of the electromagnetic shower and to differentiate photons from $\pi^0 \rightarrow \gamma\gamma$ decays. The third layer is supposed to measure only the tails of the electromagnetic shower and therefore a coarser granularity was chosen. The inner wheel of the EMEC is segmented in two layers

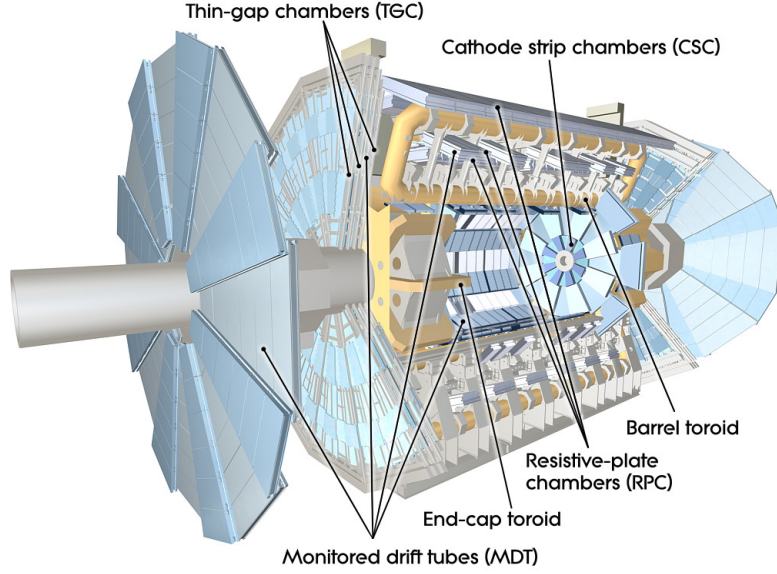


Figure 3.8: Overview of the ATLAS muon spectrometer with its four components used for high-precision tracking and triggering of muon events. The toroid magnet system is also shown [54].

The Forward Calorimeters

The forward calorimeter (FCAL), covering $3.1 < |\eta| < 4.9$ and at a distance of approximately 4.7 m from the interaction point, is exposed to a very high particle flux which dictates its design. It consists of three layers in each of the end-caps, the first used as electromagnetic calorimeter and the remaining as hadronic calorimeters. All three layers use LAr as active material however different absorber material. Each layer consists of a matrix of concentric rods and tubes using copper as passive material for the innermost layer (electromagnetic forward calorimeter) and tungsten for the two outmost layers (hadronic forward calorimeter). Tungsten was chosen to minimise the lateral spread of the hadronic showers.

3.2.3 Muon Spectrometer

Muons are the only directly detectable SM particles that traverse the calorimeter depositing only a tiny fraction of their energy. The momentum of muons is measured in the outermost subdetector of the ATLAS detector: the muon spectrometer (MS). The design of the MS is depicted in Fig. 3.8. The trajectories of the muons are bent by a superconducting toroid magnet system with a peak field strength of about 4 T. Each of the toroids in the barrel region and the two end-caps consists of eight coils which are assembled radially and symmetrically around the beam axis. Their layout with the open structure was designed to minimise the interaction of muons with the magnet material and avoid energy losses. The magnetic field of the toroids is orthogonal to that of the solenoid and thus provides an independent measurement of the momentum with respect to the ID. Furthermore the orthogonality to the muon trajectories reduces

the amount of multiple scattering. The momentum of muons is measured in two high precision tracking chambers, the *monitored drift tubes* (MDTs) in the barrel and the *cathode strip chambers* (CSCs) in the end-caps. Furthermore, *resistive plate chambers* (RPCs) and *thin gap chambers* (TGCs) are used in the barrel and end-caps respectively to allow for fast triggering of muon events with $|\eta| < 2.4$. The trigger chambers are also used to measure the ϕ coordinate of the track to complement measurements by the MDTs.

The 1150 MDTs are arranged in three cylindrical layers in the barrel region and four end-cap wheels, covering a pseudorapidity range of $|\eta| < 2.7$. The chambers consist of multiple layers of aluminium tubes of about 3 cm in diameter and filled with a Ar/CO₂ gas mixture.

Due to the higher particle flux at higher $|\eta|$, the innermost layer of the end-caps ($2.0 < |\eta| < 2.7$) uses the CSCs which are radially oriented multi-wire proportional chambers, 32 chambers in total. Cathode strips in orthogonal directions are used for fast readout to cope with the higher particle flux.

The RPCs are used in the barrel region to trigger muon events due to their good spatial and time resolution. They cover the pseudorapidity range $|\eta| < 1.05$ and are arranged in three cylindrical layers. A total of 606 RPCs is used, consisting of two resistive electrode-plates which are oriented parallel to each other in a distance of 2 mm and filled with a gas mix.

In the end-cap region, $1.05 < |\eta| < 2.4$, the good time resolution and high rate capability is provided by 3588 TGC chambers which rely on the same principle as multi-wire proportional chambers. TGCs consists of two cathode plates with an anode wire in between. Both the RPCs as well as TGCs have a readout time ranging between 15 and 25 ns.

3.2.4 The Trigger System

The high collision frequency of the LHC with up to 40 MHz for the design bunch spacing of 25 ns or up to 20 MHz during 2012 operations precludes the ATLAS detector from reading out each event and recording them. A trigger system has been developed to record only interesting events, such as high- p_T leptons, photons, jets or a large amount of missing transverse momentum, and to discard minimum bias events. That ATLAS trigger system is comprised of a purely hard-ware based *level-1* (L1) trigger and a software based *high-level trigger* (HLT). In Run-1, the HLT was further divided into two trigger stages, the *level-2 trigger* (L2) and the *event filter* (EF). During the first long shutdown of the LHC, the ATLAS trigger system has undergone major changes to cope with the approximately 5 times larger event rates due to the decrease in bunch spacing and the increase of the peak luminosity expected for Run-2 data-taking conditions. The initial collision rates of 20 MHz and 40 MHz in Run-1 and Run-2 are reduced to about 400-600 Hz and 1kHz respectively.

Only a subset of the detector information from the calorimeters with coarser granularity and the muon chambers is used to reject events at the L1 stage within a latency of 25 μ s. No information from the Inner Detector is considered due to the long readout time. In Run-1 (Run-2), the L1

reduced the event rate to 75 (100) kHz. Regions of interest (RoI) are then defined in the η - ϕ plane which are further analysed by the HLT.

In the following, the HLT as used in Run-1 will be introduced. For Run-2, only one trigger system exists that combines the steps of the L2 and EF. The L2 investigates the RoI using the full granularity information of all sub-detectors. Now also information from the tracking detectors is used to allow for a more precise identification of particles and their p_T . The readout time is reduced because only information pointing to the RoI is requested. Within 40 ms a decision is made, reducing the rate to about 3 kHz. The full event is then reconstructed in the EF with algorithms similar to what is being used offline. Furthermore, the objects are calibrated and alignment corrections are applied. Events that pass all trigger levels are written to streams which are stored permanently. To improve the usage of tracks in the trigger and avoid the limitation to RoI, a hardware-based tracker, the *Fast TracKer* (FTK) will be incorporated in 2017 to provide full tracking information for the HLT.

3.2.5 Simulation

Monte Carlo (MC) generators play an important role in high energy physics and are used for different purposes, e.g. for the prediction of experimental observables which discriminate between signal and background processes, for the estimation of SM backgrounds, for signal yield predictions of new physics processes as well as derivation of calibration and systematic uncertainties. The simulation of MC samples is divided into three main steps: the event generation, the simulation of the detector and the digitisation which will be explained in the following. The factorisation theorem [61] plays an important role in the simulation of events as it enables the short distance components of the hard scattering process to be treated independently from the non-perturbative formation of final state particles.

An example of the simulation of a pp collision at the LHC with its different steps is schematically shown in Fig. 3.9. The leading-order cross-section of the *hard interaction* of two partons with high transverse momentum transfer, $ab \rightarrow X$, can be calculated through

$$\sigma_{pp \rightarrow X} = \sum_{a,b} \int dx_a dx_b \int f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \times d\hat{\sigma}_{ab \rightarrow X}(x_a, x_b, \mu_F, \mu_R) . \quad (3.4)$$

The parton-level cross-section $\hat{\sigma}_{ab \rightarrow X}$ is convoluted with the parton distribution functions (PDFs) $f_a(x_a, \mu_F^2)$ which depend on the Bjorken momentum fraction x_a of the parton with respect to the proton momentum and the factorisation scale μ_F . The PDF $f_a(x_a, \mu_F^2)$ describes the probability of finding a parton a with momentum x_a and $\mu_F = Q$ where Q is the characteristic scale of the given process. The parton-level cross-section depends also on the renormalisation scale μ_R and is calculated from the matrix element with an integration over the phase space of all possible outgoing particles. In the second step, the *parton shower* is generated. Colour-charged partons with a high transverse momentum can emit gluons which can radiate further gluons or split into quark-antiquark pairs. Particle showers can be initiated by the partons before the collision by

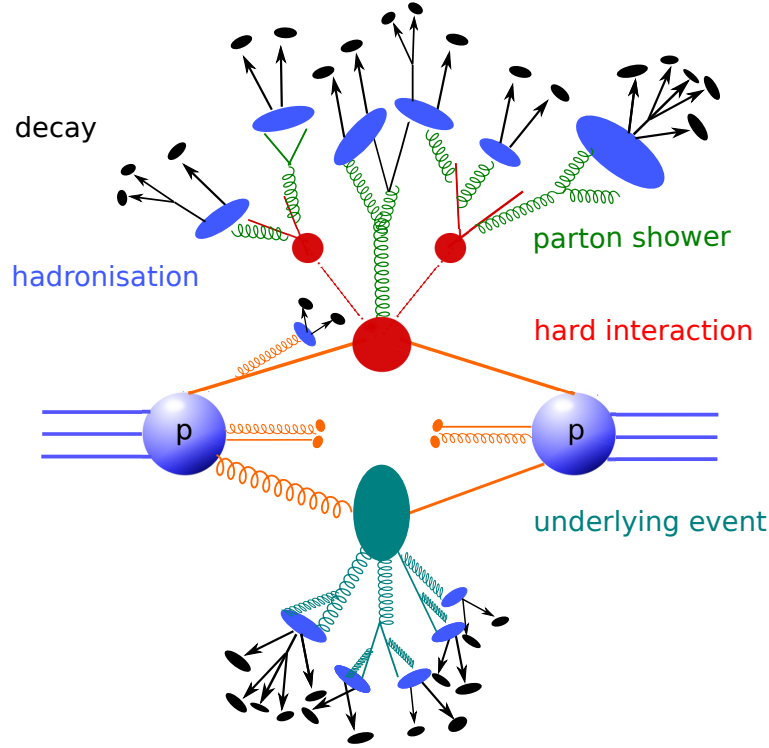


Figure 3.9: Schematic view of a Monte Carlo simulated event including the hard scattering process, parton shower, hadronisation and underlying event.

initial state radiation or by one of the particles produced in the collision, referred to as *final state radiation*. Particles are emitted and produced until a certain energy scale is reached at which the *hadronisation* process starts and colourless hadrons are formed. These hadrons are not necessarily stable and decay into the final state particles. In the final step of the event generation, the *underlying event* is modelled which consists of multiple parton interactions, beam remnants and pile-up interactions which are mostly characterised by low p_T transfers. Non-perturbative phenomenological models are needed to describe these processes. Two different types of MC generators exist: multi-purpose MC generators which generate all previously described processes and MC generators which are only used for one specific purpose. ALPGEN [62] and MADGRAPH [63] are amongst other tree-level matrix element generators only specialised on the simulation of the hard scattering process and then interfaced to full event simulation generators such as PYTHIA [64], HERWIG [65] and SHERPA [66]. Two different approaches exist for the simulation of the hadronisation: the Lund string fragmentation model which is implemented in PYTHIA while HERWIG and SHERPA use the cluster fragmentation model. Some variables are not well modelled by the simulation and special tunes are derived to improve the agreement in data and MC simulation.

The generated stable truth particles are then passed through a GEANT4 [67] model of the ATLAS detector [68] which simulates the interaction of the stable particles with the detector material. The energy deposits in the calorimeter from the different particles is then converted

into electronic signals. Since the same format is used as for the ATLAS detector read-out system, the same reconstruction algorithms are used for MC generated samples as for the collected data.

Chapter 4

Jet Reconstruction, Grooming and Substructure Variables in ATLAS

Quarks and gluons cannot be observed as individual particles in the detector, as opposed to other particles of the SM like electrons and muons, because of the asymptotic freedom of the strong force. The coupling g_s increases toward larger distances as described in Section 2.1.2 and therefore quarks and gluons fragment and hadronise into colourless bound states due to the colour confinement. The resulting collimated spray of charged and neutral hadrons is known as a *jet*. Depending on the algorithm that is used to reconstruct this collimated spray of hadrons to gain information about the initial partons, the resulting jets can be completely different objects. Jets are commonly produced in the hard scattering process as shown in Fig. 4.1, the hadronic decay of particles and initial or final state radiation of gluons. Hence the understanding of jets is of key importance to be able to reconstruct events at the LHC.

In this thesis, a search for hadronically decaying vector boson pairs is presented. Typically each of the quarks, produced in the decay of the vector bosons, would be reconstructed as an individual jet. However because of the high centre-of-mass energy of the LHC, vector bosons are often produced with a transverse momentum much larger than their mass. Thus, the decay products of the W and Z boson are boosted and collimated along the bosons momentum. The angular separation $\Delta R_{q\bar{q}}$ of the hadronic decay products depends on the mass $m_{W/Z}$ of the vector boson and its transverse momentum p_T

$$\Delta R_{q\bar{q}} \simeq \frac{1}{\sqrt{x(1-x)}} \frac{m_{W/Z}}{p_T}, \quad (4.1)$$

where x and $(1-x)$ are the momentum fraction carried by the two quarks, respectively. For the decay of a W or Z boson, the two quarks are expected to have approximately the same momentum and thus Equation 4.1 simplifies to $\Delta R_{q\bar{q}} \simeq \frac{2 \cdot m_{W/Z}}{p_T}$. The angular separation of the W boson decay products as a function of its momentum is depicted in Fig. 4.2. For a W boson with a p_T of 200 GeV, the $\Delta R_{q\bar{q}}$ between the quarks for the majority of events is approximately 0.8 whereas for a p_T of 500 GeV, the average angular separation is less than 0.5. Using a typical jet size of $R = 0.4$ or $R = 0.6$, the parton showers of the two quarks start to overlap and cannot be resolved anymore. Therefore, the decay products of boosted hadronically decaying vector

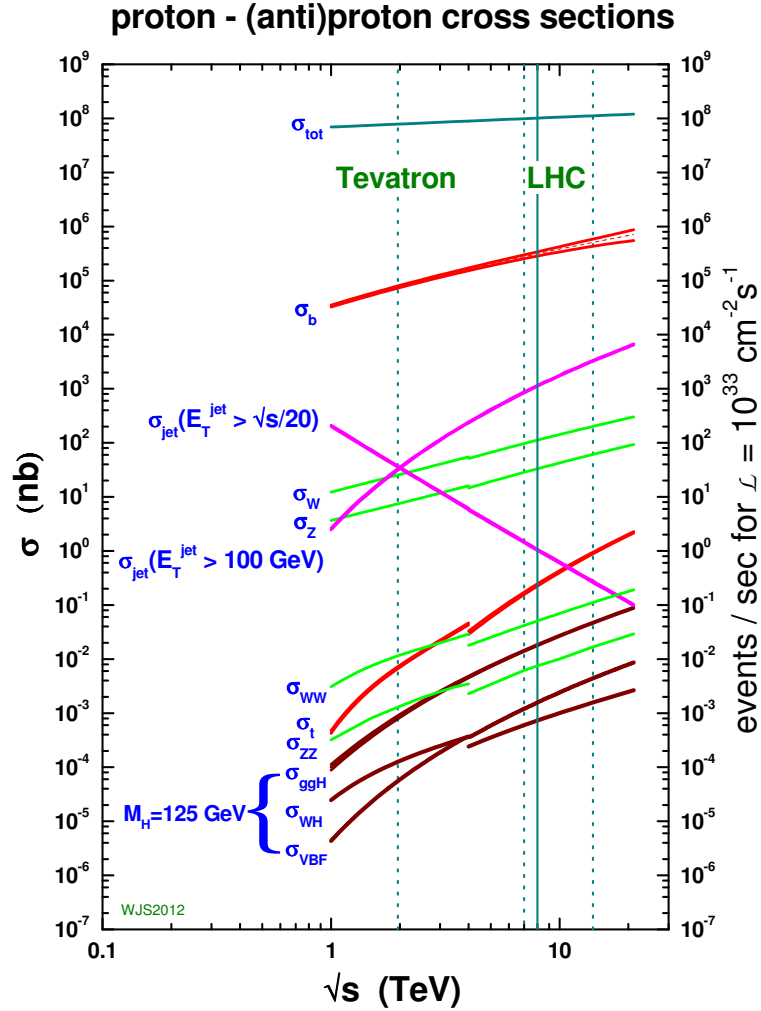


Figure 4.1: Cross-section for different Standard Model processes at pp and $p\bar{p}$ colliders as a function of the centre-of-mass energy [69]. The vertical dashed lines represent the centre-of-mass energies of the Tevatron and the LHC. The discontinuity in the cross-section prediction is due to the Tevatron being a $p\bar{p}$ collider while the LHC collides protons. The dominating process at hadron colliders is the production of dijet events. About 80 million $b\bar{b}$ events are produced per second, indicated by the σ_b curve, at a centre-of-mass energy of 13 TeV and assuming the design luminosity of the LHC $\mathcal{L} = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

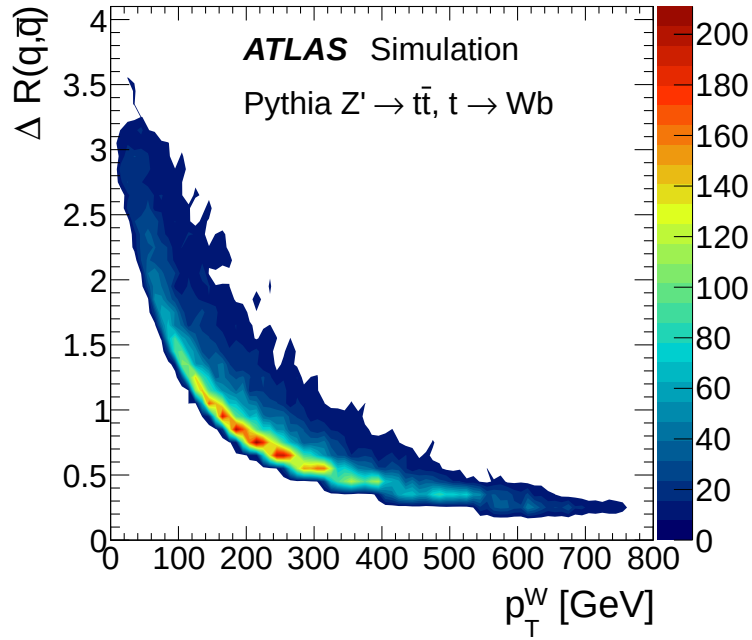


Figure 4.2: Right: Angular separation $\Delta R(q, \bar{q})$ of the hadronic decay products of a W boson produced in the decay chain of a new heavy gauge boson Z' into pairs of top quarks at $\sqrt{s} = 7$ TeV [70].

bosons are usually reconstructed within a *large-radius* jet of the size of $R \approx 1.0$ to ensure that all energy deposits from the hadronic decay are captured. The choice of jet radius parameter size was extensively optimised in Run-I in ATLAS and is described in more detail in Chapter 6. Jets originating from the vector boson decay will be denoted in the following as W/Z -jets and light quark or gluon initiated jets as QCD-jets.

Due to the high-luminosity conditions, soft particles unrelated to the hard scattering can contaminate these jets resulting in a diminished mass resolution. To mitigate the influence of pile-up effects on large- R jets, jet *grooming* techniques were introduced to remove jet constituents based on certain criteria.

The production cross-section of new heavy particles with vector bosons in the final state is several orders of magnitude smaller than for QCD-jets. To enhance the sensitivity to the production of these new particles, grooming algorithms are further used to reveal the hard structure within a jet in the case of additional pile-up vertices. Large- R jets coming from a vector boson decay are characterised by two dense regions with high energy which share approximately the same p_T whereas for a QCD-jet only one narrow region with high energy is expected assuming no hard final state radiation. So called *substructure* variables are calculated from the constituents of the jet. These variables allow to define criteria to distinguish between large- R jets originating from the hadronic decay of W/Z bosons and large- R jets from QCD multi-jet production.

This chapter is organised as follows. The inputs to the jet reconstruction algorithm in ATLAS

are introduced in Section 4.1. The most commonly used jet reconstruction algorithms are described in Section 4.2 and their different features and applications are discussed. Afterwards, the three main ingredients to suppress the enormous QCD background to searches such as the diboson resonance search presented in Chapter 7 are defined in Section 4.3–4.5: the jet mass, a selection of different grooming algorithms and substructure variables. The vector boson identification algorithm that imposes criteria on these three ingredients is described in detail in Chapter 6.

4.1 Inputs to Jet Reconstruction

The inputs to the jet reconstruction algorithms used in ATLAS analyses are *topo-clusters* which are formed from topologically connected calorimeter cells [71]. For the reconstruction of large- R jets, a *Local Calibration Weighting* (LCW) scheme is applied to the topo-clusters to calibrate them to the hadronic scale [71]. For the derivation of the jet energy and mass scale uncertainties of large- R jets, *track-jets* are used as reference object, which are built from charged-particle tracks in the Inner Detector from the primary vertex and as such are insensitive to pile-up effects. Furthermore stable particles with a lifetime of at last 10 ps, excluding muons and neutrinos, can be used as input to the reconstruction algorithm, resulting in *truth jets* which are needed as baseline for the derivation of jet energy and mass calibrations as discussed in Section 5.1

4.1.1 Topological Cluster Formation and Local Cluster Calibration

Topo-cluster Formation

To extract the signal and to suppress the noise background, topologically connected calorimeter cells are grouped together in topo-clusters. The grouping of cells is based on the cell significance $\zeta_{\text{cell}}^{\text{EM}}$, defined as the ratio of the cell signal $E_{\text{cell}}^{\text{EM}}$ measured at the electromagnetic scale and the noise σ_{noise}

$$\zeta_{\text{cell}}^{\text{EM}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sigma_{\text{noise, cell}}^{\text{EM}}} = \frac{E_{\text{cell}}^{\text{EM}}}{\sqrt{(\sigma_{\text{noise}}^{\text{electronic}})^2 + (\sigma_{\text{noise}}^{\text{pile-up}})^2}}, \quad (4.2)$$

where $\sigma_{\text{noise}}^{\text{electronic}}$ and $\sigma_{\text{noise}}^{\text{pile-up}}$ are the noise due to electronics and pile-up, respectively. In 2012, the noise due to in- and out-of-time pile-up was dominant and determined in MC simulations with an average number of interactions per bunch-crossing of $\langle\mu\rangle = 30$. As described in Section 3.1, the actual $\langle\mu\rangle$ in 2012 data taking was smaller than the value used for the noise determination, resulting in a lower sensitivity of the calorimeter to small signals.

In the first step of the topo-clustering algorithm, all cells are categorised based on their cell significance and three algorithm specific parameters S , N , P .

1. Seed cells: $|\zeta_{\text{cell}}^{\text{EM}}| > S$, default $S = 4$

2. $|\varsigma_{\text{cell}}^{\text{EM}}| > N$, default $N = 2$
3. $|\varsigma_{\text{cell}}^{\text{EM}}| > P$, default $P = 0$

It should be stressed here that the criteria are based on the absolute value of the cell significance. Therefore the random positive and negative noise fluctuations caused by pile-up are suppressed by taking into account cells with negative cell significance. The algorithm starts then with the seed cell with the highest significance which defines the first proto-cluster. All neighbouring cells with $|\varsigma_{\text{cell}}^{\text{EM}}| > N$ are then collected in the proto-cluster where neighbouring either means adjacent cells in the same layer or with partial overlap in the η - ϕ plane in case they are in different layers. In case an adjacent cell to one of the neighbouring cells passes as well the second requirement it is also added to the proto-cluster. Therefore, the parameter N defined the growth of the proto-clusters. If no further cell with $|\varsigma_{\text{cell}}^{\text{EM}}| > N$ can be added to the proto-cluster, all neighbouring cells with $|\varsigma_{\text{cell}}^{\text{EM}}| > P$ are collected as well in the proto-cluster. The algorithm proceeds then with the next seed cell (ordered according to the cell significance) until no cells remain. The choice of parameters S , N , P results in clusters with different cell multiplicity which have a core of high-significance signal cells, surrounded by cells with low signal significance.

Clusters do not necessarily contain the calorimeter response to only one single particle. Due to large shower fluctuations, a topo-cluster can also contain the response to several particles such as from the hadronic decay of a W boson with a very high transverse momentum. If the responses of the two quarks are collected within one single topo-cluster, the substructure techniques in Chapter 4.5 cannot be used anymore, as the inner structure of the large- R jet cannot be resolved. Therefore in case a topo-cluster has at least two local energy maxima with $E_{\text{cell}}^{\text{EM}} > 500$ MeV, it is split between the maxima. The energy of cells at the boundary of the two newly formed topo-clusters is then shared based on the distance of the cell to the centre of the clusters and their energies.

The resulting topo-clusters are defined to be massless and their energies are the weighted sum of the energies of the building cells. To avoid distortions in the direction of the topo-clusters due to cells with negative energies, the η and ϕ of the clusters are calculated as the sum of the individual cell positions weighted by their absolute energy. Only topo-clusters with positive energy are used as inputs to jet reconstruction algorithms. For the reconstruction of large- R jets, the topo-clusters are calibrated beforehand.

Local Cluster Calibration

Topo-clusters are formed at the electromagnetic scale and then calibrated to the LCW-scale to account for the non-compensating character of the calorimeter which causes a smaller signal for hadrons than for electrons or photons of the same incident energy. Furthermore the LCW procedure corrects for signal losses on cluster boundaries caused by the choice of $\sigma_{\text{noise}}^{\text{pile-up}}$ in the topo-clustering algorithm and for signal losses caused by inactive detector material. To improve the linearity of the calorimeter response to hadronic and electromagnetic signals and thus the

energy measurement, correction factors are applied depending on whether the cluster is caused by an electromagnetic or hadronic signal. The correction factors are derived in single particle MC simulations using neutral pion decays into photon pairs for the electromagnetic calibration and charged pions for the hadronic calibration.

4.1.2 Tracks

When charged particles pass through the Inner Detector, their trajectories (denoted as *tracks*) and their kinematic properties are reconstructed from hits in the individual Inner Detector layers. To ensure the quality of tracks originating from the primary vertex, defined as the vertex with the highest $\sum (p_T^{\text{track}})^2$, and to reject fake tracks, the following track selection criteria have to be satisfied [72]

- $p_T^{\text{track}} > 500 \text{ MeV}$;
- Transverse impact parameter: $|d_0| < 1.5 \text{ mm}$;
- Longitudinal impact parameter: $|z_0| \times \sin(\theta) < 1.5 \text{ mm}$;
- At least one hit in the pixel detector: $N_{\text{Pixel}} \geq 1$;
- At least six hits in the SCT: $N_{\text{SCT}} \geq 6$.

The transverse and longitudinal impact parameters are measured with respect to the primary vertex and the angle θ is the angle between the beam and the track.

4.2 Jet Reconstruction Algorithms

Particles are grouped together to build jets according to a certain set of rules specified by the *jet algorithm*. Two close-by particles are reconstructed in one single jet or two separate jets based on the definition of the *distance parameter*, which is also often referred to as radius parameter (R) of the jet algorithm. An algorithm should also be able to identify a single isolated particle as a jet with size R . Furthermore to complete the *jet definition* [73], the *recombination method* needs to be specified which determines how the particles are combined. The most commonly used recombination scheme adds the 4-vector of the individual particles together (E -scheme). As different jet algorithms result in different jets, it is important to define certain quality criteria that need to be met by each jet definition. In 1990 [74], a group of theorists and experimentalists agreed on a list of necessary properties for jet algorithms which are as follows:

1. Simple to implement in an experimental analysis;
2. Simple to implement in the theoretical calculation;
3. Defined at any order of perturbation theory;

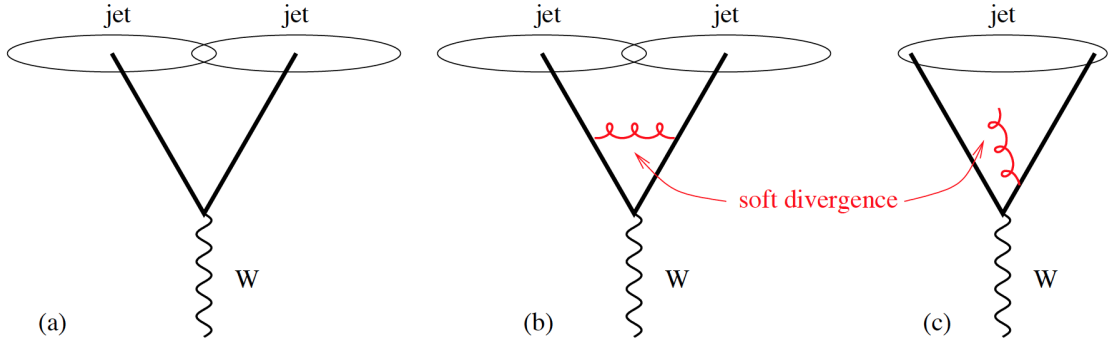


Figure 4.3: Illustration of an infrared unsafe jet algorithm applied to an event with a hadronically decaying W boson where (a) the two decay products are reconstructed as two separate jets and (c) the additional soft gluon emission induces the two partons to be reconstructed in one single jet as opposed to an infrared safe algorithm in (b) [73].

4. Yields finite cross-section at any order of perturbation theory;
5. Yields a cross-section that is relatively insensitive to hadronisation;

Whereas the first two points are more of technical importance, the last three criteria are today incorporated in the definition of *infrared* and *collinear safety*. In an infrared safe algorithm, the set of jets remain unchanged if the event has been modified by an additional low- p_T gluon emission, whereas an additional collinear (small angle) splitting does not affect the set of jets in a collinear safe algorithm. A jet algorithm is required to be insensitive to these two effects as soft gluon emissions as well as collinear splittings occur randomly through perturbative and non-perturbative effects during the hadronisation of a hard parton and are therefore difficult to predict. Fig. 4.3 illustrates how the radiation of a soft gluon can change the number of jets such that the two decay products of a W boson are merged in one single jet rather than two separate jets that are obtained without gluon emission.

If the jet algorithm is infrared and collinear safe, the divergent tree-level matrix elements that are associated to collinear splittings and soft gluon emissions cancel the divergent loop matrix elements resulting in a finite cross-section. This effect is schematically demonstrated in Fig. 4.4 where the soft gluon emission for the collinear unsafe algorithm splits the partons in two separate jets compared to the loop matrix element process where one parton undergoes a virtual gluon correction. This results in a different number of jets for the tree-level matrix element with collinear splitting and therefore the divergencies do not cancel anymore.

An example of a collinear unsafe algorithm uses the highest energetic parton in the event as a seed and groups all partons together in a jet that are within a certain angular distance around the seed. If however, the hardest parton undergoes a collinear splitting, the seed may change and with it the cone around the seed. Particles previously reconstructed within one jet can then fall in the cone of two different seeds if their angular separation is larger than the cone size of

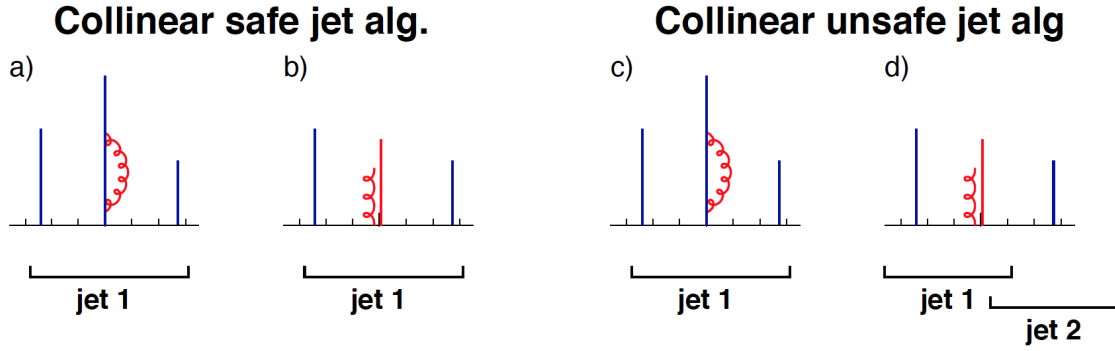


Figure 4.4: Illustration of a collinear safe (left) and unsafe (right) jet algorithm [73]. The vertical lines represent partons whose transverse momentum and rapidity are indicated by the length and horizontal position of the lines, respectively. The collinear splitting in (d) results in creating two jets for the collinear unsafe algorithm rather than one jet as in (b). Hence, the collinear unsafe algorithm changes the jet multiplicity compared to the loop matrix element configuration in (a) and (c).

the new seed.

Jet algorithms that are currently being used and that have been used in the past in high-energy physics can be divided into two main categories: cone algorithms and sequential reconstruction algorithms. As only special kinds of cone algorithms are infrared and collinear safe, the remainder of this chapter will focus on sequential reconstruction algorithms.

Sequential Reconstruction Algorithms

The idea of sequential reconstruction algorithms is to iterate over all inputs to the algorithm, which could be either topo-clusters, tracks or truth particles and to combine them based on their distance to each other and the chosen recombination scheme. Two distance measures are introduced: the distance between two entities d_{ij} and the distance between the entity and the beam d_{iB} . Two entities i and j are combined if the distance d_{ij} is smaller than the beam distance d_{iB} . The new object is then added to the list of inputs to the algorithm, replacing the two objects i and j and all distances are recalculated. Otherwise if $d_{iB} < d_{ij}$ for all entities j in the input list, the entity i is designated as jet and removed from the list of inputs. This procedure is repeated until no objects remain in the event.

The first sequential reconstruction algorithm was introduced for the electron-positron collider PETRA by the JADE Collaboration in the 1980's [75]. However, the algorithm needed to be modified to be used as well at hadron colliders. The distance measures used at hadron colliders are defined as follows:

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad (4.3)$$

$$d_{iB} = p_{T,i}^{2p} \quad (4.4)$$

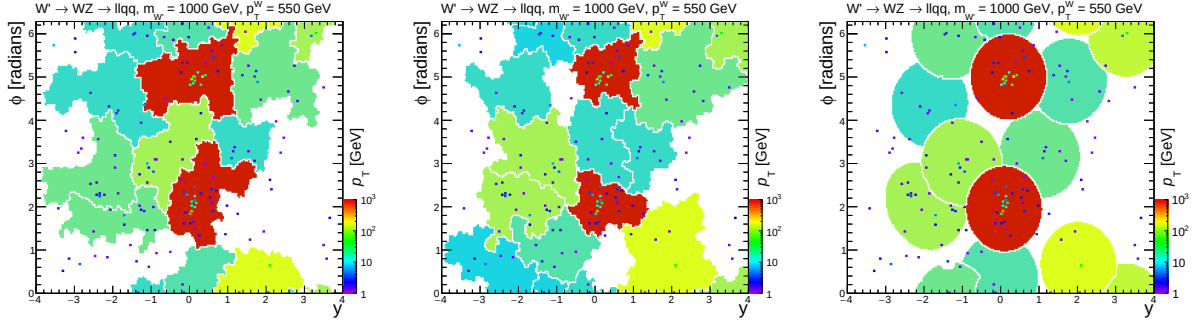


Figure 4.5: An example of the active area of jets reconstructed with a radius parameter of $R = 1.0$ for a simulated $W' \rightarrow WZ \rightarrow \ell\ell qq$ event with $m_{W'} = 1000$ GeV at $\sqrt{s} = 8$ TeV using the k_t (left), the C/A (middle) and the anti- k_t algorithm (right). The colours indicate the p_T of the jets and the points show the energy deposits in the calorimeter.

where ΔR_{ij} is the angular distance between the two objects i and j and R is the radius parameter of the jet algorithm which regulates the size of the jets. As a consequence of the limiting jet radius, an arbitrarily soft entity without any other objects within $\Delta R_{ij} < R$, will be reconstructed as a jet. To avoid this feature of the jet algorithm, a lower limit on the transverse momentum should be set. The free parameter of the algorithm, p , defines how the input particles are combined.

The k_t algorithm [76, 77] corresponds to $p = 1$, the Cambridge-Aachen (C/A) algorithm [78] to $p = 0$ and the anti- k_t algorithm to $p = -1$ [79]. Only three different values are considered for p as for $p > 1$ and $p < -1$ the features are relatively similar to the k_t and anti- k_t algorithm respectively. For the C/A algorithm, entities are combined purely based on their angular separation. For the k_t algorithm, soft particles are favoured to be merged first in contrast to the anti- k_t algorithm where the jet starts growing outwards from a high- p_T object. Despite starting from the highest p_T object, the anti- k_t algorithm is collinear safe as the distance measure includes a combination of an energy and angular dependent merging criterion such that collinear splittings are merged with the high- p_T object in the beginning before merging with each other. The growth of the jet from the inside to the outside results in conical jets compared to the k_t and C/A algorithm which exhibit irregular shapes caused by the handling soft radiation. The shape of jets, reconstructed with each of the three different jet algorithms, is compared in Fig. 4.5 using a simulated W' event. The event simulates the decay of a W' with $m_{W'} = 1000$ GeV into a W and Z boson where the Z boson decays into a lepton pair and the W boson hadronically. In principle conical jets are favoured since they are not only less susceptible to the underlying event and pile-up but also easier to calibrate (the correction for energy losses is explained in Chapter 5.1). Therefore the anti- k_t algorithm is the default algorithm used in ATLAS with a typical radius parameter of $R = 0.4$. However, the irregular shapes of the k_t and C/A algorithm are useful in terms of studying the parton shower of the initial particle as the clustering process of the two algorithms reverses this process.

4.3 Jet Mass

The most powerful variable to distinguish boosted hadronically decaying vector bosons from QCD-jets is the jet *mass* m , which is calculated from the energy E and momentum $\vec{p}$ of its constituents (e.g. topo-clusters in the case of calorimeter jets):

$$m^2 = \left(\sum_i E_i \right)^2 - \left(\sum_i \vec{p}_i \right)^2. \quad (4.5)$$

For a W/Z -jet, the jet mass is close to the mass of its originating boson. QCD-jets originate from approximately massless partons, however a large fraction of jets has a significantly larger mass than zero. At leading-order, the differential jet mass distribution is approximately given by

$$\frac{1}{n} \frac{dn}{dm^2} \sim \frac{1}{m^2} \frac{\alpha_s C_i}{\pi} \left(\ln \frac{R^2 p_T^2}{m^2} + \mathcal{O}(1) \right), \quad (4.6)$$

where α_s is the coupling constant of the strong force and C_i the colour factor for either a quark- or gluon-jet [73, 80]. The mass distribution diverges for $m = 0$ because of soft emissions. At higher orders in perturbation theory, taking e.g. wide-angle radiation into account, the low mass divergences are removed by the Sudakov double-logarithms [81] causing the peak of the mass distribution to shift to higher values (Sudakov peak) resulting in non-zero masses of the jet. The position of the Sudakov peak can not be predicted precisely as the jet mass depends strongly on the underlying event, pile-up activity and non-perturbative effects including the hadronisation. A QCD dijet event, reconstructed with the anti- k_t algorithm and a radius parameter of $R = 1.0$ is depicted in Fig. 4.6. Two high- p_T jets are shown but with significantly different jet masses. Whereas the leading jet (jet1) is dominated by one very high- p_T cluster and gains its mass by two additional soft clusters, the subleading jet (jet2) possesses several soft and wide-angle clusters causing the jet mass to be large. The chosen event was simulated with only a small number of reconstructed vertices ($N_{\text{vtx}} = 2$) and thus the comparatively different jet masses are mostly caused by the different hadronisation of the initial quarks.

4.4 Jet Grooming Algorithms

As previously shown, the mass distribution of jets at calorimeter level is significantly different from the distribution at truth-level where effects such as additional pile-up are not considered. To restore the mass resolution at reconstruction level, three different grooming techniques are considered: *trimming* [82], *split-filtering* [83] and *pruning* [84]. Further techniques that have not been studied in full detail yet by the ATLAS Collaboration can be found in [80, 85].

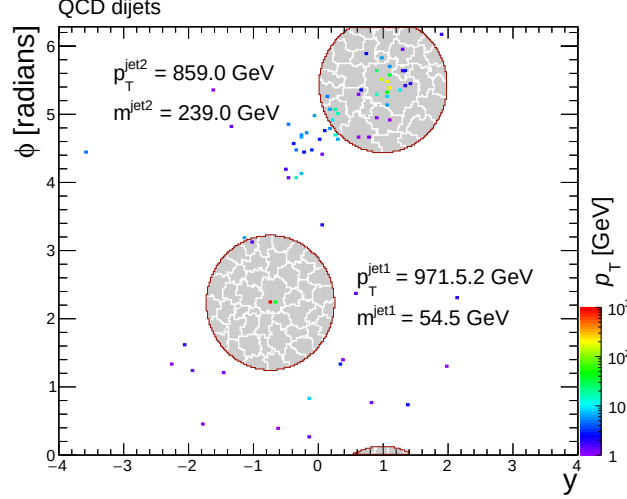


Figure 4.6: An example of jets reconstructed with the anti- k_t algorithm with $R = 1.0$ for a simulated dijet event at $\sqrt{s} = 8$ TeV. The grey areas within the jets represent subjets with $R_{\text{sub}} = 0.2$. Only jets with $p_T > 100$ GeV are shown and the points indicate the energy deposits in the calorimeter.

4.4.1 Trimming

In a large- R jet containing the decay products of a boosted, hadronically decaying vector boson, two dense cores of high energy are expected. To identify these cores, the constituents of the large- R jet are clustered into subjets with a typical size of $R_{\text{sub}} = 0.2$ or 0.3 . The k_t algorithm is preferably used to reconstruct subjets as the energy share between the subjets is equitable [82] compared to the C/A and anti- k_t algorithms. A simple criterion to remove the contamination from pile-up effects from a two-prong W/Z -jet would be to keep only the three hardest subjets. Three subjets are chosen instead of two subjets to account for the occurrence of final-state radiation. However, a grooming algorithm should not only be applicable to jets from a heavy particle decay but also to QCD-jets where at leading order only one hard subjet is expected. Hence keeping a certain number of subjets would not be efficient in removing pile-up effects for a QCD-jet. Instead, one uses the fact that the subjets built from constituents corresponding to initial-state radiation and pile-up are relatively soft with respect to the subjets from the hard scattering process. The trimming procedure removes subjets whose transverse momentum (p_T^{sub}) is below a certain fraction of the p_T of the large- R jet: $p_T^{\text{sub}} < f_{\text{cut}} \cdot p_T$. The remaining subjets are then added to form the trimmed large- R jet. The relative p_T fraction f_{cut} is a parameter of the trimming algorithm and needs to be optimised. Typical values chosen for f_{cut} are of the order of 5%, however the value strongly depends on the subjet radius, as larger radii necessitate an increased f_{cut} as more radiation can contaminate the subjets.

To illustrate the trimming procedure, the same event as in Fig. 4.5 is shown in Fig. 4.7(a) using the anti- k_t algorithm but only jets with $p_T > 100$ GeV are depicted and the transverse momentum and mass of the two remaining jets are indicated. Furthermore the subjets within

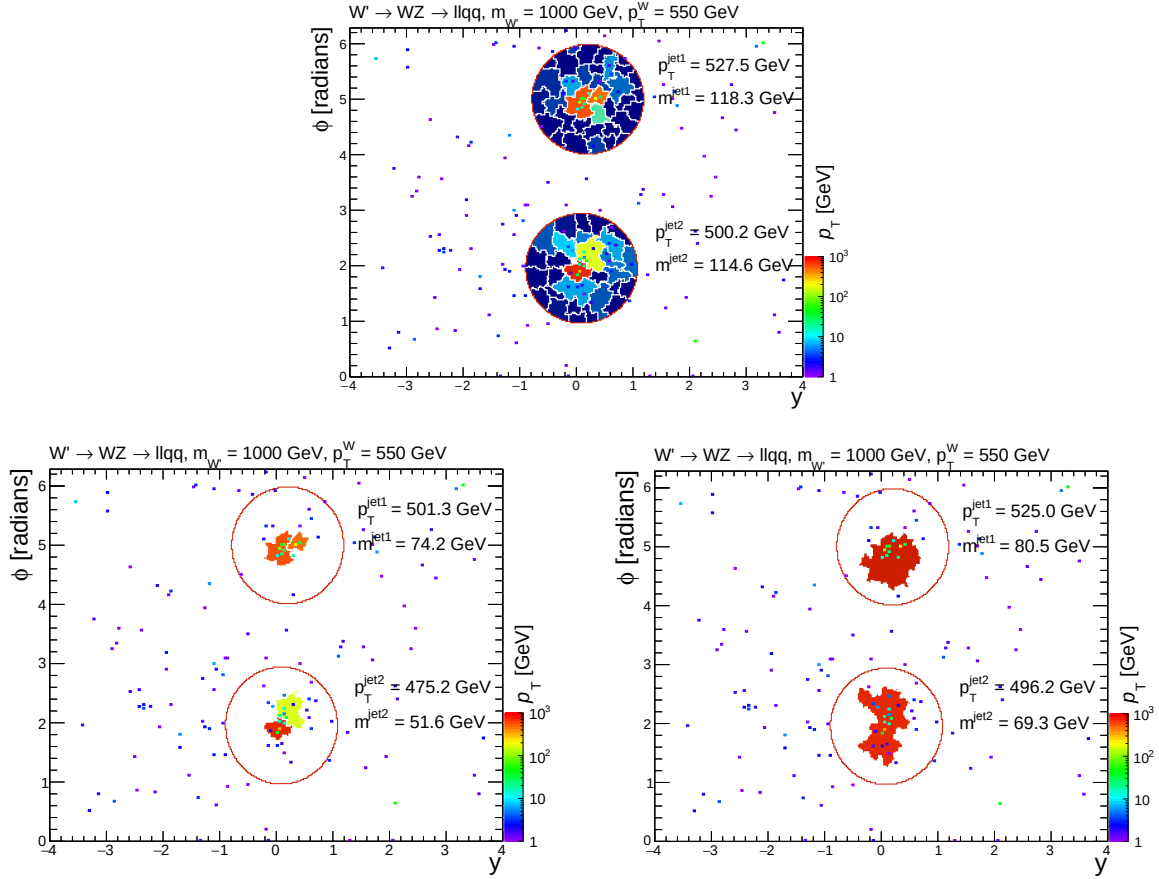


Figure 4.7: An example of jets reconstructed with the anti- k_t algorithm with $R = 1.0$ before (top) and after (bottom) applying trimming for a simulated $W' \rightarrow WZ \rightarrow \ell\ell qq$ event with $m_{W'} = 1000$ GeV at $\sqrt{s} = 8$ TeV. The coloured areas within the jets represent the subjects of size $R_{\text{sub}} = 0.2$ (top, bottom left) and 0.3 (bottom right) and the p_T of the subject is indicated by its colour. After trimming (bottom), only subjects with a p_T fraction larger than 5% of the jets p_T are kept. Only jets with $p_T > 100$ GeV are shown and the points show the energy deposits in the calorimeter and their colours indicate their p_T .

the large- R jet are shown, built with the k_t algorithm and a radius size of $R_{\text{sub}} = 0.3$. In the left figure, no jet trimming is applied to the jets. One of the two jets (denoted as jet1) is supposed to contain the decay products of the W boson decay and thus its jet mass is expected to be close to the mass of the W boson. However the mass of the jet is increased due to the contamination of pile-up and ISR. The second jet (jet2) is initiated by a light quark or gluon. After trimming the jets in Fig. 4.7(c), only one subject remains for each jet resulting in a significant decrease of the jet mass such that the W -jet mass is close to the mass of the W boson.

The comparison of the mass distribution of W/Z -jets and QCD-jets, reconstructed with the anti- k_t algorithm with $R = 1.0$, before and after trimming is depicted in Fig. 4.8. For the ungroomed jets, both the signal and background like jets peak at a jet mass around 140 GeV. After trimming the jets, the discrimination of the two jet topologies as well as the mass resolution

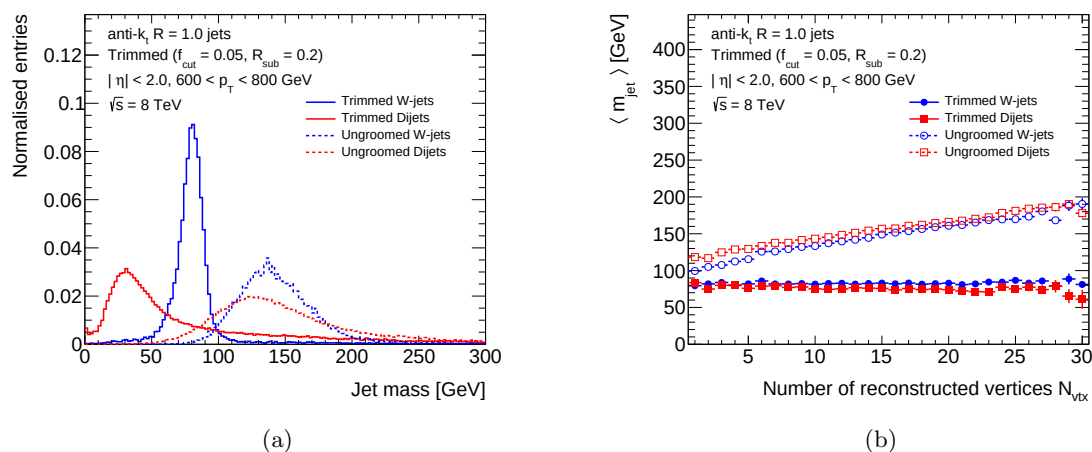


Figure 4.8: Comparison of the mass distribution for anti- k_t $R = 1.0$ ungroomed and trimmed jets with $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.2$ for simulated W signal and multi-jet background events (left) and its average dependence on the number of reconstructed vertices N_{vtx} (right). The choice of the algorithm parameters are explained in Chapter 6.1.

of the W/Z -jets is significantly improved. The majority of W/Z -jets have a mass close to the W boson peak. Furthermore the average jet mass as a function of the number of reconstructed vertices is shown. Whereas the ungroomed jet mass depends strongly on the in-time pile-up, the average trimmed jet mass does not show any dependence. Although the jet mass distribution shows a clear discrimination of the QCD-jet and W -jet mass, the average jet mass for the two topologies is very similar. This is due to the fact that the QCD-jet mass distribution exhibits long tails to high mass values compared to W/Z -jets. With increasing p_T of the jet, the tails are more pronounced causing the average jet mass of QCD-jets to be larger than for W/Z -jets.

4.4.2 Split-filtering

In 2008, Butterworth, Davison, Rubin and Salam introduced the split-filtering algorithm, also referred to as the BDRS algorithm, to improve the sensitivity of the vector boson associated production of the at that time undiscovered Higgs boson [83]. The decay of the Higgs boson into a pair of b -tagged jets was studied due to the large branching ratio. However the associated Higgs production suffered from the large $t\bar{t}$ background and the detector acceptance in reconstructing the b -jets. Both effects could be reduced by considering the boosted topology. Despite being proposed for the Higgs boson search, the BDRS algorithm is used nowadays as well for searches beyond the SM including the hadronic decay of vector bosons. The BDRS algorithm comprises two stages: the splitting stage uses the substructure of the jet and in the filtering stage, the pile-up and soft radiation contamination is removed. The algorithm was optimised for C/A jets because of the angular-ordered jet reconstruction. This implies that reversing the last step of the clustering algorithm reveals the two proto-jets that have the widest angular separation

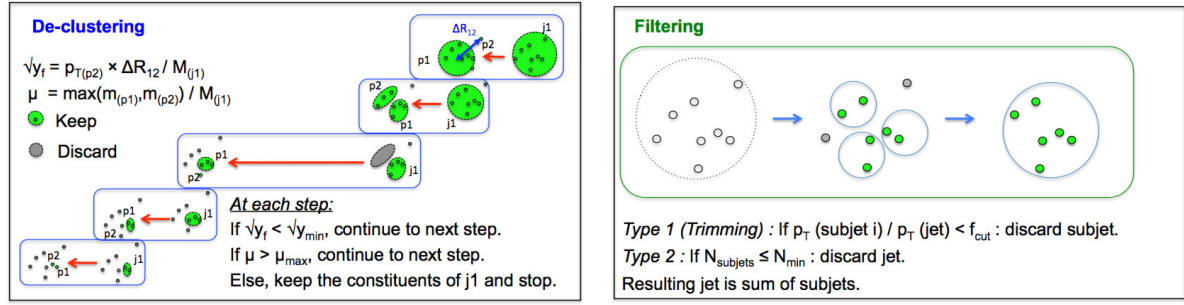


Figure 4.9: Illustration of the two stages of the split-filtering algorithm: The splitting stage de-clusters the large- R jet and requires a significant mass-drop and p_T balanced subjects (left). In the filtering stage, the C/A jet is reclustered and any radiation outside the three hardest subjects is discarded (right) [86].

and allows to study the hard splitting of a heavy particle decay. The split-filtering algorithm is schematically illustrated in Fig. 4.9 and proceeds as follows:

1. The last step of the C/A clustering of the jet j is undone and thus it is split into two subjects j_1 and j_2 where $m_{j_1} > m_{j_2}$.
2. If there is a significant mass drop $\mu_{\max}$ such that $m_{j_1} < \mu_{\max} \cdot m_j$ and the subjects are balanced in momentum:

$$\sqrt{y_{12}} = \frac{\min(p_{T1}, p_{T2})}{m_{12}} \Delta R_{12} > \sqrt{y_{\text{cut}}} , \quad (4.7)$$

where ΔR_{12} is the angular separation of the two subjects and m_{12} the invariant mass of j_1 and j_2 , the jet j is considered as a hard object.

3. If the jet j does not fulfil one of the above criteria, define j_1 as j and go back to the first step, continuing to de-cluster the more massive subject.

If the mass-drop and symmetry criteria are not fulfilled at any stage of the de-clustering procedure, the jet is discarded. The mass-drop, $\mu_{\max}$, and the momentum balance, $\sqrt{y_{\text{cut}}}$, are parameters of the jet algorithm. The original paper suggested to use $\mu_{\max} = 0.67$ and $y_{\text{cut}} = 0.09$, however it has been shown that the mass-drop criteria is not needed for a better performance of the algorithm [80].

To remove the contribution from the underlying event, the jet is reclustered in the filtering stage using subjects of the size $R_{\text{sub}} = \min(0.3, \Delta R_{12})$ using the angular separation ΔR_{12} of the two subjects at the last stage of the splitting procedure. Only the three hardest subjects are taken to capture as well possible final state radiation.

A comparison of the jet mass for ungroomed and split-filtered C/A $R = 1.2$ signal and background like jets is shown in Fig. 4.10. The momentum balance was chosen to be $y_{\text{cut}} = 0.04$ (also denoted as y_{filt}) as it has been used in several diboson resonance searches performed in

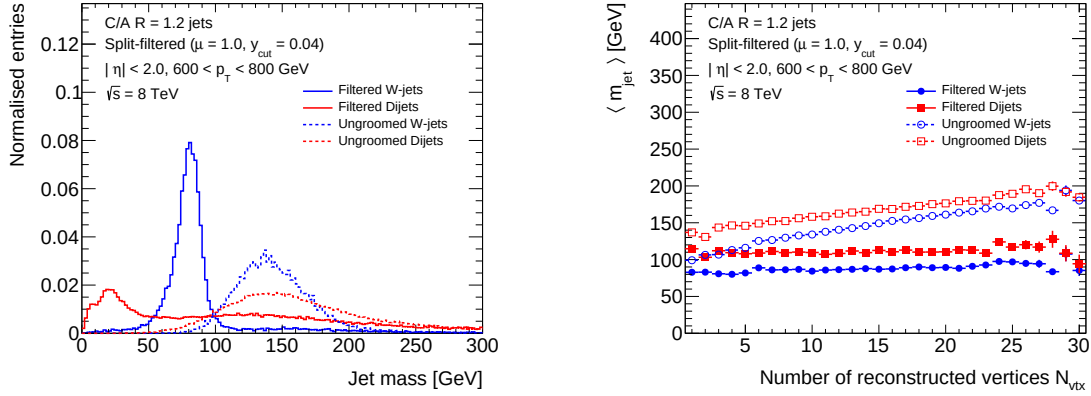


Figure 4.10: Comparison of the mass distribution for $C/A R = 1.2$ ungroomed and split-filtered jets with $y_{\text{filt}} = 0.04$ and $R_{\text{sub}} = 0.3$ for simulated W signal and multi-jet background events (left) and its average dependence on the number of reconstructed vertices N_{vtx} (right). The choice of the algorithm parameters are explained in Chapter 6.1.

ATLAS [87–89] of which the all-hadronic search will be described in Chapter 8. After the split-filtering procedure, no pile-up dependence can be observed. However, the average jet mass of QCD-jets is larger than for W/Z -jets due to the high-mass tail as opposed to the trimming algorithm in Fig. 4.8. Using tighter y_{cut} values reduces the high-mass tail of the background jets and hence the contamination of QCD-jets in the region of the W -jet mass distribution peak.

4.4.3 Pruning

The pruning algorithm utilises the fact that soft and wide-angle emissions do not occur in the fragmentation and hadronisation of a heavy particle decay but are only present inside a large- R jet because of pile-up and the underlying event. The algorithm proceeds as follows:

1. The constituents of a large- R jet, reconstructed with any jet algorithm, are reclustered with either the C/A or k_t algorithm.
2. At each step of the pairwise recombination of two constituents j_1 and j_2 , the softer constituent is discarded if it is either too soft: $\frac{\min(p_T^{j_1}, p_T^{j_2})}{p_T^{j_1+j_2}} < z_{\text{cut}}$ or the angular separation $\Delta R_{j_1, j_2}$ between the two constituents is too wide: $\Delta R_{12} > R_{\text{cut}} \cdot \frac{2 \cdot m_{\text{jet}}^{\text{jet}}}{p_T^{\text{jet}}}$.
3. Otherwise the two constituents are merged and the pruning algorithm proceeds with the first step.

The algorithm has two free parameters, z_{cut} and R_{cut} , however the results of the algorithm are relatively insensitive to the exact parameter values. The criterion on the angular separation changes from jet-to-jet taking into account the opening angle of the jet [84]. A schematic illustration of the pruning algorithm is depicted in Fig. 4.11.

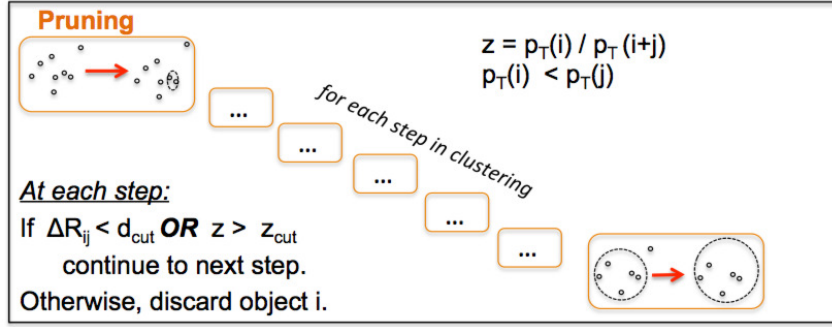


Figure 4.11: Illustration of the pruning algorithm removing wide angle and soft radiation at each step of the recombination algorithm [86].

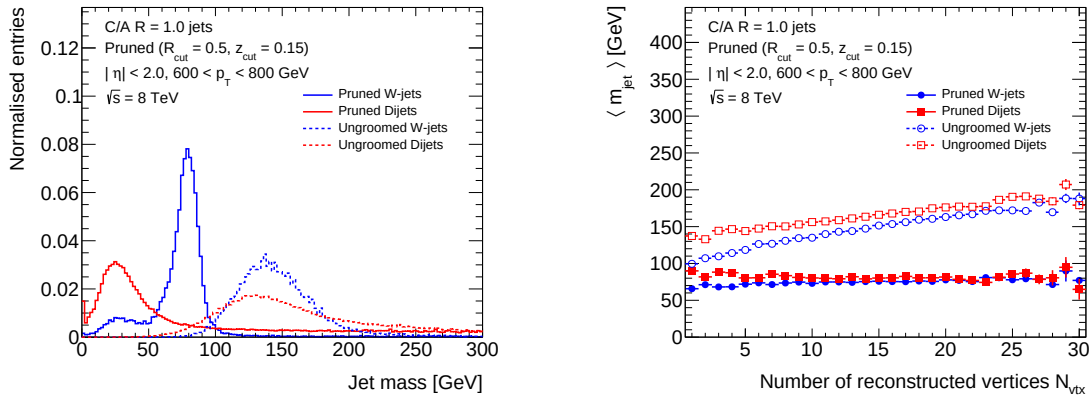


Figure 4.12: Comparison of the mass distribution for C/A $R = 1.0$ ungroomed and pruned jets with $R_{\text{cut}} = 0.5$ and $z_{\text{cut}} = 0.15$ for simulated W signal and multi-jet background events (left) and its average dependence on the number of reconstructed vertices N_{vtx} (right). The choice of the algorithm parameters are explained in Chapter 6.1.

The C/A $R = 1.0$ jet mass distributions for QCD-jets and W/Z -jets are compared in Fig. 4.12 for ungroomed and pruned jets. Compared to the other grooming algorithms, a small plateau in the pruned W -jet mass distribution below the W boson peak can be observed. For these jets, either the irregular jet shape of the C/A algorithm does not cover the full decay of the boosted W boson, which could be recovered by using a larger jet radius or the pruning algorithm is too tight and removes as well particles associated to the parton shower of the decay products.

The pruning algorithm is favoured by the CMS Collaboration [90, 91] however in ATLAS, it still exhibits a dependence on the number of primary vertices at small jet p_T despite optimising the free parameters of the algorithm. This difference can most likely be explained by the usage of *particle flow* inputs, which combine information from the calorimeter with that of the Inner Detector, to the jet reconstruction algorithm in the CMS Collaboration [92, 93].

4.5 Jet Substructure Moments

In this section, only the substructure variables that have a large discriminating power to distinguish the one-prong structure of QCD-jets from the two-prong structure of W/Z -jets are described. Further variables that have been studied by ATLAS [86] are defined in Appendix A. The variables are calculated from the constituent clusters of the groomed large- R jet to reduce their sensitivity to the pile-up conditions.

N -subjettiness

The N -subjettiness [94] variable, τ_N , describes to what degree the inner structure of a jet is compatible with the hypothesis of it consisting of N or fewer subjects. First, the subjects are reconstructed within the large- R jet using the exclusive k_t algorithm [77] requiring exactly N subjects. If a large fraction of the jets energy is not aligned along the subjet directions but rather away from them, the hypothesis of N subjects seems unlikely. To quantify this, the N -subjettiness variable τ_N is defined

$$\tau_N = \frac{\sum_i p_{T,i} \min(\Delta R_{a_1,i}, \dots, \Delta R_{a_N,i})}{\sum_i p_{T,i} \cdot R}, \quad (4.8)$$

where the sum runs over the constituents i of the given jet J , $p_{T,i}$ are their transverse momenta and $\Delta R_{a_k,i}$ is the angular distance of the constituent i to the axis of subjet k .

A value of $\tau_N \approx 0$ means that most of the energy within the jet is aligned with the subjects whereas for $\tau_N \gg 0$, a significant fraction of the energy has a large angular separation from the subjects and hence more than N subjects are expected. The τ_N variable itself does not provide enough separation power and instead the ratios τ_N/τ_{N-1} are used. To identify the two-prong structure of boosted vector boson, the variable $\tau_{21} = \tau_2/\tau_1$ provides good discrimination power with respect to QCD-jets whereas for a three-prong structure such as the top-quark decay, the ratio $\tau_{32} = \tau_3/\tau_2$ is used.

The discrimination of τ_N can be further improved with an alternative definition of the subjet axis used in Equation 4.8 [95]. The *winner-takes-all* (WTA) axis uses the hardest constituent inside the subjet k to calculate the angular separation $\Delta R_{a_k,i}$ instead of the subjet axis. To indicate the axis used in the τ_N definition, the τ_{21} ratio calculated with the winner-takes-all axis will be denoted as τ_{21}^{wta} in the following.

The discrimination power of the τ_{21}^{wta} variable is displayed in Fig. 4.13 for anti- k_t $R = 1.0$ trimmed W -jets and QCD-jets. A large fraction of QCD-jets have a τ_{21}^{wta} value of zero. For these jets, the exclusive k_t algorithm failed to reconstruct two subjects. The same effect can be observed for W -jets but less pronounced than for QCD-jets. It is important to point out that the usage of the exclusive k_t algorithm can create biases for QCD-jets with a hard one-prong structure and a soft and wide-angle emission which is the most dominant way a QCD-jet gains its mass if $p_T \gg m$. To calculate τ_2 the algorithm forces the jet to build two subjects, one around the hard parton and the other will contain the wide-angle emission and thus give low τ_{21} values as for W/Z -jets. After restricting the mass of the trimmed jet to a mass close to the W boson,

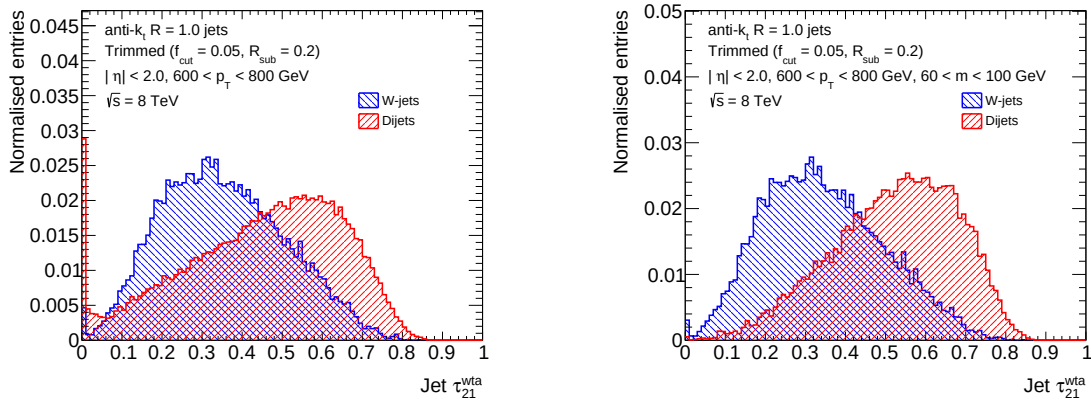


Figure 4.13: N -subjettiness τ_{21}^{wta} distribution for the leading anti- k_t $R = 1.0$ trimmed jet in W signal and multi-jet background events without mass window requirement (left) and for jets with $60 < m < 100$ GeV (right).

$60 < m_{\text{jet}} < 100$ GeV, and thus imposing some kind of inner structure of the jets, the number of jets with $\tau_{21}^{wta} = 0$ is significantly reduced. Moreover the separation of the boson jet and QCD-jet distribution is enhanced with the mass window requirement.

Energy Correlation Variables

The idea behind energy correlation variables [96] is very similar to that of N -subjettiness but has two main advantages: (a) the variable does not rely on finding subjets and consequently (b) has a better treatment of soft emissions at large angles. The starting point are N -point energy correlation functions (ECF) that run over all constituents of any given jet J and are based purely on the p_T of the constituents and their pair-wise angular separation ΔR_{ij} :

$$\text{ECF}(0, \beta) = 1, \quad (4.9a)$$

$$\text{ECF}(1, \beta) = \sum_i p_{T_i}, \quad (4.9b)$$

$$\text{ECF}(2, \beta) = \sum_{i < j \in J} p_{T_i} p_{T_j} (\Delta R_{ij})^\beta, \quad (4.9c)$$

$$\text{ECF}(3, \beta) = \sum_{i < j < k \in J} p_{T_i} p_{T_j} p_{T_k} (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^\beta, \quad (4.9d)$$

where the angular exponent β is a free parameter that can be optimised based on the mass of the resonance under study. For low mass resonances such as the W and Z boson, values of $\beta \simeq 2$ are preferable whereas for high mass resonances, small values of $\beta \simeq 0.5$ give an optimal separation [96]. The N -point correlation functions can be used for quark-gluon discrimination and boosted top tagging, however the focus here is the identification of boosted vector bosons. Two dimensionless ratios, C_2^β [96] and D_2^β [97, 98], are defined to identify $N = 2$ dense cores of

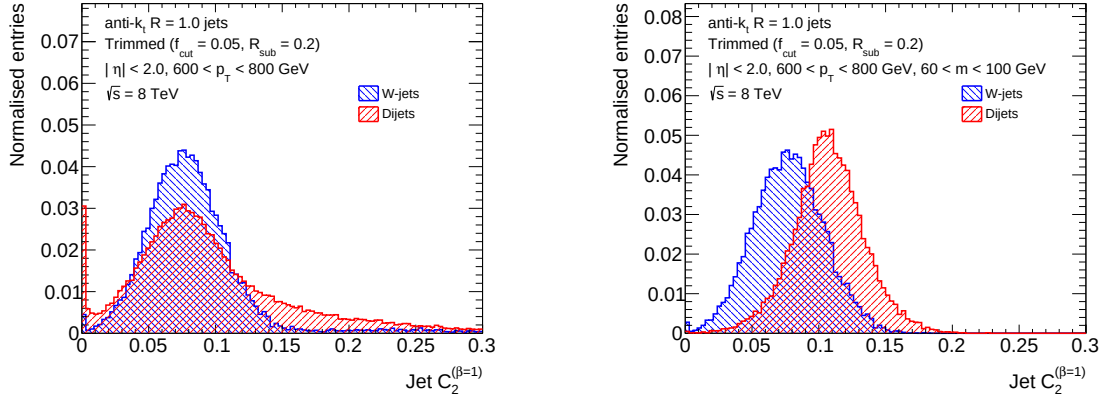


Figure 4.14: Energy correlation variable $C_2^{\beta=1}$ for the leading anti- k_t $R = 1.0$ trimmed jet in W signal and multi-jet background events without mass window requirement (left) and for jets with $60 < m < 100$ GeV (right).

energies within the jet:

$$C_2^\beta = \frac{\text{ECF}(3, \beta) \text{ECF}(1, \beta)}{\text{ECF}(2, \beta)^2}, \quad (4.10a)$$

$$D_2^\beta = \frac{\text{ECF}(3, \beta) \text{ECF}(1, \beta)^3}{\text{ECF}(2, \beta)^3}. \quad (4.10b)$$

If a jet contains N subjets, the $N+1$ -point correlation function $\text{ECF}(N+1, \beta)$ will be significantly smaller than $\text{ECF}(N, \beta)$. Both C_2^β and D_2^β are not infrared and collinear safe unless a mass window requirement is imposed on the jet for example around the boson mass.

The $C_2^{\beta=1}$ and $D_2^{\beta=1}$ distributions with and without requiring a mass window around the W boson mass are shown for signal and background jets in Fig. 4.14 and Fig. 4.15 respectively. Two different β values have been studied by ATLAS [86], $\beta = 1$ and $\beta = 2$, however better separation of W/Z and QCD-jets was observed for $\beta = 1$. After the mass window requirement, the signal and background discrimination is improved, especially for the C_2^β variable as it's strongly correlated to the jet mass.

Momentum Balance

The momentum balance variable $\sqrt{y_{12}}$

$$\sqrt{y_{12}} = \frac{\min(p_{T1}, p_{T2})}{m_{12}} \Delta R_{12},$$

as previously defined in Equation 4.7 is only meaningful for jets groomed with the split-filtering algorithm and calculated at the first step of declustering where the momentum balance (and mass-drop if specified by the algorithm) criterion is met. The authors of the split-filtering al-

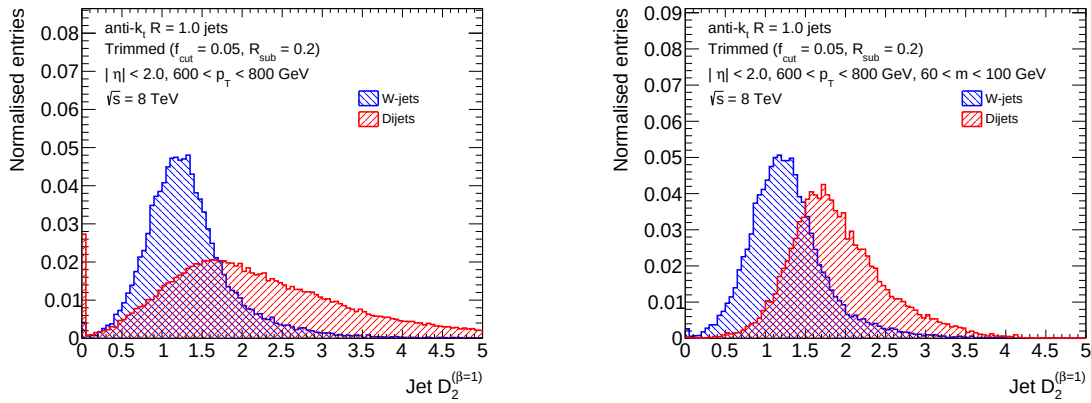


Figure 4.15: Energy correlation variable $D_2^{\beta=1}$ for the leading anti- k_t $R = 1.0$ trimmed jet in W signal and multi-jet background events without mass window requirement (left) and for jets with $60 < m < 100$ GeV (right).

gorithm proposed the momentum balance variable to be larger than $\sqrt{y_{\text{cut}}} = 0.3$ to efficiently remove pile-up, initial state radiation and underlying event contributions. In the diboson resonance searches however, a looser criteria of $\sqrt{y_{\text{cut}}} = 0.2$ was used since a robust pile-up removal was still guaranteed and at the same time $\sqrt{y_{12}}$ could be used to discriminate signal from background jets. The momentum balance variable as used in the all-hadronic diboson search performed at $\sqrt{s} = 8$ TeV is shown in Fig. 7.3.

Chapter 5

Large- R Jet Calibration in ATLAS

The concept of jets as proxies for the hadronisation of quarks and gluons has been previously introduced in Chapter 4. Jets are one of the key ingredients of measurements and searches for physics beyond the Standard Model at the LHC. However, jets reconstructed with the algorithms described in Section 4.2 cannot be directly used in an analysis and need to be calibrated to the truth jet energy and mass scale using Monte Carlo simulations to correct for different detector effects such as the non-compensating character of the calorimeter. The calibration procedure for large- R jets is described in Section 5.1. Systematic uncertainties on the jet energy, mass and substructure scales are then derived to account for differences observed between data and simulation using well-understood reference objects such as track-jets. The systematic uncertainties for both 2012 and 2015 data-taking are summarised in Section 5.2. Future improvements to the reconstruction of large- R jets are discussed in Section 5.3.

The calibration procedure and systematic uncertainties based on 2012 data will be presented for anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 0.05$ and $R_{\text{sub}} = 0.3$. This jet collection was the default reconstruction algorithm for 2012 analyses, besides split-filtered C/A jets with $R = 1.2$ that were used in diboson resonance searches as described in Section 7.3.1. For the diboson resonance search in 2015, the boosted vector boson identification algorithm, including the jet reconstruction and grooming algorithm, was optimised to account for the higher centre-of-mass energy and changed pile-up conditions as described in Section 6.2. Therefore, the diboson resonance search in 2015 data-taking used anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 0.05$ and a slightly smaller subjet radius size of $R_{\text{sub}} = 0.2$ and the systematic uncertainties were derived for this jet collection.

The decrease in subjet radius size for the trimming algorithm under different pile-up and centre-of-mass energy conditions demonstrates one limitation of the usage of the algorithm in ATLAS. If the average number of interactions per bunch crossing increases, the trimming parameters of the algorithm, i.e. f_{cut} and R_{sub} need to be reoptimised. In order to avoid this, the subjects within the large- R jet can be first cleaned from pile-up contamination by applying the jet area pile-up correction, introduced in Section 5.3.1. The pile-up corrected subjects are then removed or kept based on their p_T fraction with respect to the ungroomed large- R jet as in the default trimming algorithm.

5.1 Jet Energy and Mass Calibration

The jet energy (mass) response $\mathcal{R}_E$ ($\mathcal{R}_m$) of reconstructed jets is defined as the ratio between the calorimeter jet energy (mass) and its corresponding truth jet energy (mass)

$$\mathcal{R}_E = \frac{E_{\text{reco}}}{E_{\text{true}}}, \quad \mathcal{R}_m = \frac{m_{\text{reco}}}{m_{\text{true}}} . \quad (5.1)$$

A truth jet is matched to a reconstructed jet if their distance in the η - ϕ space is less than 0.75 times the large- R jet radius: $\Delta R = \sqrt{(\eta_{\text{reco}} - \eta_{\text{truth}})^2 + (\phi_{\text{reco}} - \phi_{\text{truth}})^2} < 0.75 \cdot R$. The jet energy (mass) response is smaller than one due to several detector and reconstruction effects such as the non-compensation of the calorimeter, inactive material, signal losses because of the chosen noise thresholds, energy losses due to particles outside of the calorimeter (leakage) and out-of-cone particles that are not inside the calorimeter jet cone but are in the truth jet cone. Furthermore the jet mass is heavily influenced by the topo-clustering algorithm and how topo-clusters are split [72]. The jet energy and mass calibration accounts for these effects and calibrates the *jet energy and mass scales* (JES, JMS) to that of truth jets. The calibration factors are derived in QCD dijet MC samples but are applied to the data as well. It is not guaranteed that the JES and JMS in Monte Carlo simulations and data are the same and *in-situ* techniques can be used to calibrate the scales in both data and MC to the same footing and to define uncertainties for the observed differences in data and MC. This procedure also decreases the jet energy and mass scale related uncertainties. However this was not part of the calibration procedure of large- R jets in Run-I as opposed to small- R jets. Several in-situ calibration techniques where a large- R jet is balancing a photon, Z boson or another jet depending on the p_T range are currently being studied to be included in the large- R jet calibration chain for data-taking in 2016.

The jet energy (mass) of calorimeter jets is on average corrected to the truth energy (mass) scale using the “numerical inversion” technique. The calibration factors are derived in bins of η and the truth energy E_{true} as the jet mass and energy response strongly depend on these quantities. The binning in terms of E_{true} instead of E_{reco} was chosen such that the energy and mass response do not depend on the underlying p_T distribution and thus result in Gaussian distributions.

In the first step of the calibration, the energy (mass) responses are fitted with a Gaussian distribution and the fitted mean is extracted. An example of the jet mass response for anti- k_t $R = 1.0$ trimmed jets and the corresponding Gaussian fit is depicted in Figure 5.1(a). The reference truth jets are also trimmed even though the inputs are only stable particles from the hard scattering vertex. Nevertheless, the mass resolution is diminished because of the underlying event. The following steps of the numerical inversion technique will be first explained based on the jet energy, while the differences with respect to the jet mass calibration are discussed afterwards. The energy calibration factors are the inverse of the fitted average jet energy response, however they are only parameterised in bins of E_{true} . The numerical inversion transforms then the average E_{true} into the corresponding E_{reco} value with the fitted average response for each η

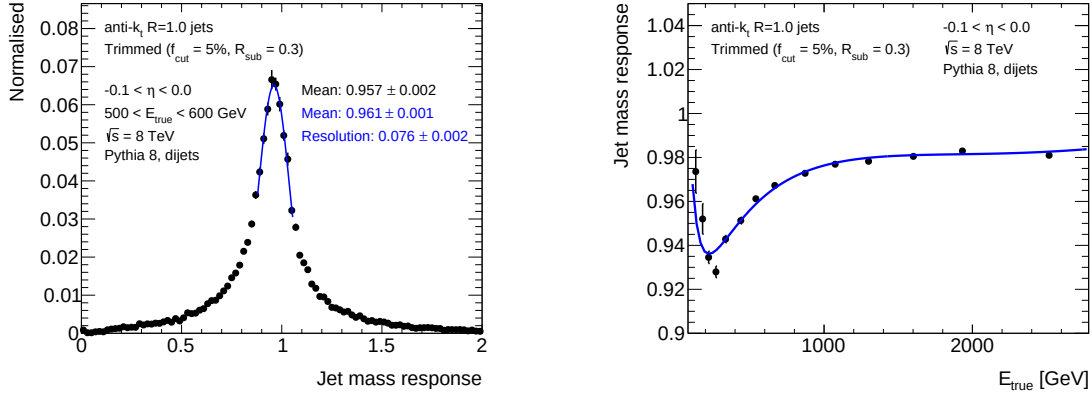


Figure 5.1: Example of the Gaussian fit to the jet mass response (left) for anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.3$ with $500 < E_{\text{true}} < 600$ GeV and example of the parameterisation of the jet mass response as a function of the truth energy (right).

and truth energy bin. For each η bin k , the resulting energy response curve as a function of the reconstructed jet energy is parameterised as

$$\mathcal{F}_{\text{calib}, k}(E) = \sum_{i=0}^{N_{\text{max}}} a_i \left(\ln E_{\text{LCW}}^{\text{jet}} \right)^i, \quad (5.2)$$

where a_i are the free parameters of the fit and N_{max} ranges between 1 and 6 and is determined by minimising $\chi^2/\text{d.o.f}$ for each of the fits. The subscript LCW indicates that the inputs to the jet algorithm are calibrated with the LCW calibration. The calibrated jet energy, denoted at the LCW+JES scale, is then:

$$E_{\text{LCW+JES}}^{\text{jet}} = \frac{E_{\text{LCW}}^{\text{jet}}}{\mathcal{F}_{\text{calib}, k}^E(E_{\text{LCW}}^{\text{jet}})}. \quad (5.3)$$

Since the jet mass depends directly on the jet energy, the calibration of the energy and mass are not independent. Therefore, the average jet mass response is not parameterised as a function of the reconstructed jet energy but the truth energy. In this way, the energy calibration is taken into account as the calibrated jet energy corresponds on average to the truth jet energy. An example fit of the jet mass response as a function of the truth energy with the function in Equation 5.2 is shown in Fig. 5.1(b). The corrected jet mass at the LCW+JES scale can then be calculated from the calibrated jet energy $E_{\text{LCW}}^{\text{jet}} \cdot \text{JES}$, with JES being the inverse of the energy response, and the uncalibrated jet mass:

$$m_{\text{LCW+JES}}^{\text{jet}} = \frac{m_{\text{LCW}}^{\text{jet}}}{\mathcal{F}_{\text{calib}, k}^m(E_{\text{LCW}}^{\text{jet}} \cdot \text{JES})}. \quad (5.4)$$

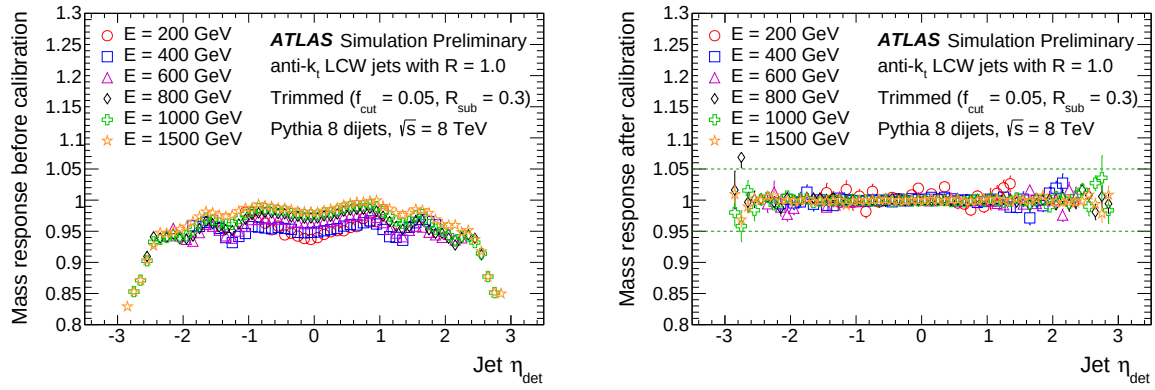


Figure 5.2: The jet mass response as a function of the jet η for anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.3$ for different jet energies before (left) and after (right) the jet mass calibration [99].

The jet mass response as a function of the jet η is depicted in Fig. 5.2 for anti- k_t $R = 1.0$ trimmed jets before and after the calibration procedure. After the calibration, the energy and η dependence of the jet mass response is removed and is approximately restored to one. In addition to the jet energy and mass, the jet η is also corrected with the numerical inversion technique. The jet p_T is then calculated accordingly from the calibrated variables.

To further improve the jet mass calibration, it is not only parameterised in bins of η and E_{truth} for Run-II but also in bins of the truth jet mass. However for high- p_T jets, which have not been probed extensively in Run-I, and for low truth jet masses, the numerical inversion technique breaks down. After the calibration, the mass responses exhibits non-Gaussian behaviour with a double-peak structure. In this phase-space, the jets with a small area and a small number of constituents cause the double-peak structure in the mass response which is most likely caused by the non-linearity of the correlation of the reconstructed and truth jet mass. To overcome the difficulties with the mass calibration and to improve the jet mass resolution at high p_T , the *track-assisted* jet mass is introduced in Section 5.3.2.

5.2 Systematic Uncertainties

Uncertainties for large- R jets accounting for differences in the calorimeter jet response in data and Monte Carlo simulations are derived using in-situ methods by comparing the measurement in the calorimeter to that of a well calibrated reference object:

$$\langle X^{\text{jet}}/X^{\text{ref}} \rangle_{\text{data}} / \langle X^{\text{jet}}/X^{\text{ref}} \rangle_{\text{MC}}. \quad (5.5)$$

For Run-I, the jet energy scale uncertainties were derived in γ +jet events using photons as reference. Due to the lack of events in which a photon is balancing a large- R jet at high transverse

momentum, the uncertainties are combined with uncertainties that use track-jets as reference objects. For analyses using the data collected in 2015, no γ +jet derived uncertainties are available yet. Uncertainties for the jet mass scale and different substructure variables are derived only using track-jets as reference objects as the γ +jet method does not provide a handle on these scales. The γ +jet based uncertainties are discussed in detail in Ref. [100]. The remainder of this chapter will focus on the *track-jet double ratio* method which has been widely used since 2011 [70]. The double-ratio technique will be explained with the help of the jet mass, but the procedure is the same for the jet energy and substructure variables.

5.2.1 Track-jet Double Ratio Uncertainties

To probe the jet mass measurement in the calorimeter, track-jets are used as reference objects since charged-particle tracks are both well measured and independent of the calorimeter. Furthermore track-jets are not susceptible to pile-up because the tracks are required to be associated to the hard-scatter vertex. The ratio of the calorimeter mass m^{jet} to track-jet mass $m^{\text{track-jet}}$, defined as

$$r_{\text{track-jet}}^m = \frac{m^{\text{jet}}}{m^{\text{track-jet}}}, \quad (5.6)$$

is used as a proxy to estimate the systematic uncertainties on the jet mass scale. The track-jets are matched to calorimeter jets using a geometrical matching in the η - ϕ space with $\Delta R < 0.75 \cdot R$ where R is the radius parameter of the jet reconstruction algorithm. Previous studies [101] have shown that there is excellent agreement between the measured positions of clusters and tracks in data, indicating no systematic misalignment between the calorimeter and the Inner Detector. The uncertainties on the jet mass calibration are then estimated by comparing the ratio defined in Eq. 5.6 in data and simulation, giving the double-ratio method its name:

$$R_{\text{track-jet}}^m = \frac{r_{\text{track-jet}}^{m,\text{data}}}{r_{\text{track-jet}}^{m,\text{MC}}}. \quad (5.7)$$

The ratio in Eq. 5.6 allows for the separation of physics and detector effects. That means that even if a physics effect might not be well-modelled in simulation, the effect will cancel in Eq. 5.6 if it affects both the Inner Detector as well as calorimeter measurement. The double-ratio in Eq. 5.7 is then well described in case detector effects are well modelled in simulation. Two different Monte Carlo generators are considered, PYTHIA8 [102] and HERWIG++ [103] to account for uncertainties due to different fragmentation and hadronisation models.

An example of the distribution of the calorimeter to track-jet mass ratio, $r_{\text{track-jet}}^m$ for the two leading anti- k_t $R = 1.0$ trimmed jets in PYTHIA8, HERWIG++ and the full dataset collected in 2012 corresponding to an integrated luminosity of 20.3 fb^{-1} is shown in Figure 5.3(a). Events are required to pass the lowest un-prescaled large- R jet trigger that has a minimum E_T of 360 GeV. The leading jet in the event must fulfil $p_T > 500 \text{ GeV}$ to ensure that the trigger is fully efficient. The shape of the $r_{\text{track-jet}}^m$ distribution is well described by both MC simulations. The $r_{\text{track-jet}}^m$

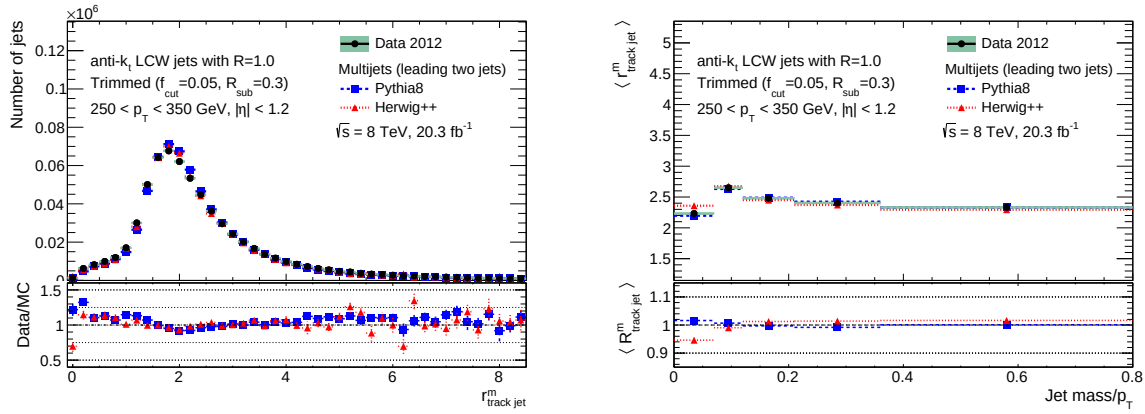


Figure 5.3: Distribution of the calorimeter to track-jet mass ratio $r_{\text{track-jet}}^m$ in data and Monte Carlo simulations for anti- k_t $R = 1.0$ trimmed jets (left) and the mean calorimeter to track-jet mass ratio as a function of m/p_T .

distribution is expected to be mostly larger than unity because the track-jet mass does not account for the neutral components. A small, non-negligible fraction of jets is however measured with $r_{\text{track-jet}}^m < 1$. This effect is caused by the trimming procedure as the fraction of ungroomed anti- k_t $R = 1.0$ jets with $r_{\text{track-jet}}^m < 1$ is significantly smaller. The trimming procedure removes subjects in case they fall below a certain p_T threshold defined by the ungroomed jet p_T . In case a subject fails the trimming criteria, it is removed even though it might be matched to a charged-particle track from the hard-scatter vertex and thus gives $r_{\text{track-jet}}^m < 1$ for this particular jet. To avoid the removal of hard-scatter energy deposits in the calorimeter, charged-particle track based grooming techniques [104] are currently being studied. The peak near $r_{\text{track-jet}}^m \approx 2$ is caused by relatively soft particles that constitute the calorimeter jet mass. These particles are either not included in the track-jet because they result from additional pp collisions in the same bunch crossing or their tracks are bent by the magnetic field and thus result in a smaller track-jet mass. High $r_{\text{track-jet}}^m$ values are suppressed by the trimming algorithm that removes most of the soft constituents.

To estimate the systematic uncertainty, the mean value of $r_{\text{track-jet}}^m$ is used as it is primarily sensitive to the particle composition of the jet and should be well modelled in any well-tuned Monte Carlo generator whereas the full distribution is more susceptible to fluctuations. The uncertainties are derived in different kinematic p_T and η regions of the detector and are only limited by the acceptance of the Inner Detector at $|\eta| = 2.5$. Therefore, uncertainties are only derived for jets with $|\eta| < 2.0$ to ensure that the core of the large- R jet is within the detector acceptance.

The mean calorimeter to track-jet mass ratio, $\langle r_{\text{track-jet}}^m \rangle$ for the two leading anti- k_t $R = 1.0$ trimmed jets as a function of the jet m/p_T is depicted in Fig. 5.3(b). Furthermore the data and Monte Carlo agreement, $\langle R_{\text{track-jet}}^m \rangle$ is shown. The double-ratio uncertainties are binned in m/p_T

to capture the dependence of the uncertainty on the mass of the jet, but to remain independent of the actual scale of the mass. The calorimeter does not directly measure mass, rather it sees energy collimation, where m/p_T provides a measure of that collimation for large- R jets within the calorimeter. The variable bin sizes were chosen to avoid statistical fluctuations which would increase the systematic uncertainties artificially. For 2011 and 2012 data-taking, the relative mass scale uncertainty was taken as the larger deviation of the PYTHIA8 or HERWIG++ double-ratio $R_{\text{track-jet}}^m$ from the unity. For jets in the p_T range $250 < p_T < 350$ GeV and $|\eta| < 1.2$, the Monte Carlo to data deviations are of the order of 2% but can increase to approximately 6% for very small m/p_T .

Track Reconstruction Uncertainties

The measurement of the track-jet mass depends strongly on the track reconstruction efficiency and the associated systematic uncertainties. For 2011 and 2012 data-taking, only the relative uncertainties on the track reconstruction efficiency due to the Inner Detector material modelling were considered. These uncertainties vary between 2% for charged-particle tracks in the central region of the detector $|\eta| < 1.3$ and increase to 7% for $2.3 < |\eta| < 2.5$ and $p_T > 500$ MeV [105]. The impact of the track reconstruction efficiency systematics on $r_{\text{track-jet}}^m$ is evaluated in PYTHIA8 by randomly removing the constituents of the track-jet according to the uncertainties. The resulting $r_{\text{track-jet}}^m$ distribution is then compared to the nominal distribution. The systematic effect is to good approximation flat in η and p_T and is approximately 2.7%.

5.2.2 Final Jet Scale Uncertainties in 2012 Data

The final systematic uncertainties on the jet mass scale using the full 2012 dataset is obtained by adding the uncertainties derived with the double-ratio track-jet method in quadrature to the track reconstruction efficiency uncertainties under the assumption that the uncertainties are independent. The relative mass scale uncertainties are shown in Fig. 5.4(a) for anti- k_t $R = 1.0$ trimmed jets as a function of the jet p_T for three different detector regions and at $m/p_T = 0.2$. The uncertainties are of the order of 4% and increase to approximately 7.5% at high transverse momentum and η due to the limited statistics in the 2012 dataset in this region of phase space. As previously mentioned, the double-ratio method as well as γ +jet events are used to derive the p_T scale uncertainties. The uncertainties derived from γ +jet events are however limited in p_T reach in the 2012 data and are only statistically significant for $p_T < 800$ GeV. The p_T scale uncertainties derived with the photon-balance method are of the order of 1% at low transverse momentum and approximately 2% at high p_T ($p_T \approx 800$ GeV) compared to 4% and 6% respectively for the double-ratio track-jet method. To benefit from the drastically reduced p_T scale uncertainties derived with γ +jet events, a linear interpolation is performed around $p_T = 900$ GeV between the two methods. A further component arising from the topological composition of the jet and accounting for the different energy distribution and hard substructure

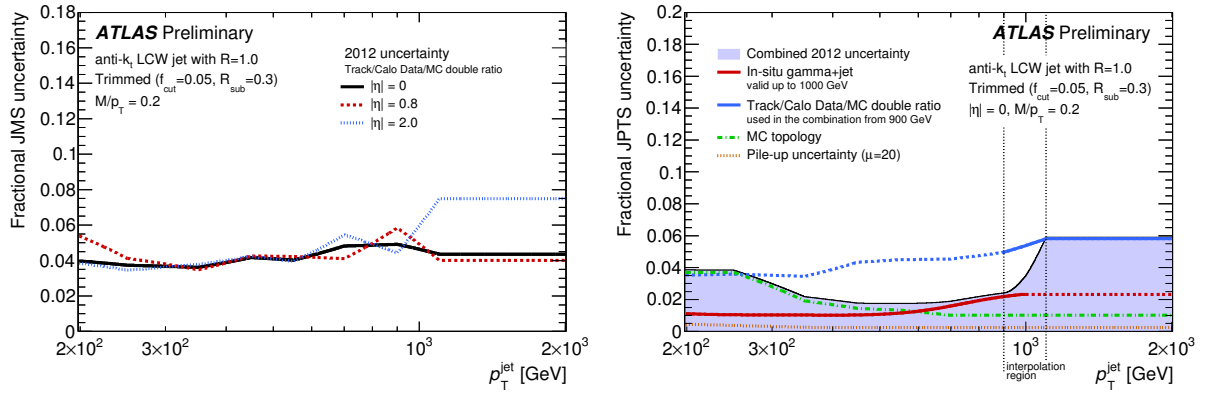


Figure 5.4: Summary of the systematic uncertainties for anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.3$ on the jet mass scale for different η ranges (left) determined with the double-ratio track-jet method and the jet energy scale derived from the track-jet double-ratio method, γ +jet events and topology differences in MC as a function of the jet p_T (right). The uncertainties are shown for $m/p_T = 0.2$ and were derived with data taken at $\sqrt{s} = 8$ TeV [106].

within a jet formed from a W boson or top-quark decay with respect to a quark/gluon jet [99] is added in quadrature to form the total uncertainty. The total uncertainty and its components are shown as a function of p_T , and for $m/p_T = 0.2$, in Fig. 5.4(b).

The dominant uncertainties in analyses including large- R jets are the jet energy, mass, and substructure scale uncertainties. With the large dataset of 8 TeV collision data available, the uncertainties are no longer statistically limited and more advanced techniques are needed to further reduce them. One possibility is *in-situ* calibrations that correct the average response in MC simulation to the average response in data. So far, only energy and mass calibration factors are derived in simulations and applied to data and MC. Furthermore, the double-ratio method uses the mean of the $r_{\text{track-jet}}^m$ distribution instead of a fit to the core of the distribution which is expected to be much better described than the tails of the distribution.

5.2.3 Monte Carlo Based Jet Scale Uncertainties for Early 2015 Analyses

The derivation of in-situ uncertainties, such as was done at 8 TeV, takes a considerable amount of time and effort. In order to have results with the new 13 TeV data as quickly as possible, a set of uncertainties were derived that exploit the knowledge acquired during Run-I and add MC-based uncertainties that account for differences between Run-I and Run-II [107, 108]. These changes include (but are not limited to) differences in the detector, the beam conditions and the reconstruction of topo-clusters. The derived uncertainties were used for the all-hadronic diboson resonance search, described in Section 9, which was one of the first public ATLAS analyses with the full 2015 dataset. The changes between Run-I and Run-II that are expected to have a non-negligible effect on the energy and mass scale of large- R jets are summarised in the following section. The installation of the IBL has shown to have only a negligible effect on the jet energy

scale of small-radius jets [107] and is therefore not considered here.

Changes from Run-I to Run-II

The physics of electromagnetic and hadronic interactions within the detector are described in specific physics lists. In Run-II, these interactions are simulated with the `FTFP_BERT` [109] physics list which uses the Bertini Intra-nuclear Cascade model (BERT) [110] for hadrons below 5 GeV and the Fritiof string model [111] with the Precompound model for hadrons with energies above 4 GeV. During Run-I, the `QGSP_BERT` [109] physics list was used that models the interactions of hadrons with the detector below 9.9 GeV with the Bertini intra-nuclear model, in the range $9.5 < E < 25$ GeV with the `GEISHA` model above 12 GeV with the Quark-Gluon-String model [112]. The noise thresholds in the topo-clustering algorithm have been adjusted for Run-II to match the noise measurements performed in 2012. Furthermore, the topo-clustering algorithm has been modified to mitigate the impact of pile-up on the energy of topo-clusters by preventing them from growing from the pre-sampler layers.

To convert the analogue signal of the LAr detector into an energy measurement, four samples per channel are read-out in Run-II instead of five as for Run-I [113]. The change in number of samples being read-out also influences the calculation of the timing which is important to distinguish between energy deposits from the bunch crossing of interest and adjacent bunches. The main modification of the beam conditions that is expected to have an impact on the jet calibration is the reduction of the bunch crossing from 50 ns to 25 ns in 2015. This results in an increase of the amount of out-of-time pile-up by a factor of two.

Monte Carlo Simulations

To study the effect of each of the changes between Run-I and Run-II on the jet energy and mass scale, dedicated QCD dijet MC samples were generated which are compared to a *reference* sample. The differences between the reference and the variation sample are accounted for as systematic uncertainty.

The reference sample is simulated using `PYTHIA8`, with the AU2 [114] underlying event tune and the CT10 [115] parton distribution function set. The generated events are propagated through a full `GEANT4`-based simulation of the ATLAS detector. The hadronic shower is described by the `FTFP_BERT` physics list and a bunch-spacing of 25 ns is simulated. Minimum bias events, generated with `PYTHIA8` using the A14 tunes [116] and the NNPDF2.3LO [117] PDFs, are overlaid onto the hard-scatter events to simulate the effects of pile-up.

The different variation samples are simulated with the same MC generator, underlying event tune, PDF set and minimum bias events as the reference sample unless otherwise stated. The differences with respect to the reference sample are:

- *Noise threshold variation*: Simulated as the reference sample but with the noise thresholds shifted upwards and downwards by a fraction that reproduces the threshold size difference between Run-I and Run-II.

- *Run-I topo-clustering*: Simulated as the reference sample but allowing the topo-clusters to grow from the pre-sampler layer.
- *50 ns with 5 sampling points in the LAr calorimeter*: Simulated as the reference sample but with 50 ns bunch-spacing and using 5 LAr sampling points for the energy reconstruction as in Run-I.
- *50 ns with 4 sampling points in the LAr calorimeter*: Simulated as the reference sample but with varied bunch-spacing.

Due to the lack of jets with $p_T > 1000$ GeV in the Run-I data, several additional variations are generated that consider different simulations of the intra-nuclear and low energy processes within the detector material to obtain a better understanding of the jet energy and mass scale in this phase-space. To quantify the effect of the high- p_T variations, a second reference sample referred to as the *nominal* sample is generated using PYTHIA8 with the A14 underlying event tune and the NNPDF2.3LO PDFs. The material interactions are described by the FTFP_BERT physics list. The nominal sample is compared to various high- p_T variation samples with different physics lists:

- *FTFP_BIC*: Simulated as the nominal sample but using the Binary Intra-nuclear Cascade (BIC) model for hadrons with $E < 9.9$ GeV.
- *QGSP_BIC*: Simulated as the nominal sample but using the Quark-Gluon-String and Pre-compound model and BIC model in energy ranges depending on the hadron type.
- *Rescattering*: Simulated as the nominal sample but with high-energy re-scattering simulated with the BIC model.

Further variations of the physics list have been studied in [108] but resulted in negligible effects on the jet energy and mass and are thus not discussed here.

The in-situ determination of the systematic uncertainties always includes a modelling component that accounts for differences between two MC generators, these are usually PYTHIA8 and HERWIG++ for large- R jets. Therefore QCD dijet samples are generated as well with HERWIG++ for the high p_T region where Run-I in-situ scale uncertainties are not available due to the limited statistics. The HERWIG++ sample uses the UE-EE-4 [118] underlying event tune and the CTEQ6 [119] PDFs.

Jet p_T , Mass and $D_2^{\beta=1}$ Scale Uncertainties

For the derivation of the systematic uncertainties, anti- k_t $R = 1.0$ trimmed jets are matched to truth jets from the corresponding jet collection (reconstructed from stable particles) with $\Delta R < 0.75$ and their response is calculated. The jet energy, mass and $D_2^{\beta=1}$ response distributions are fitted with a Gaussian function and the mean value of the different variation samples is compared to that of the corresponding reference samples. Systematic uncertainties are derived

for jets with $200 < p_T^{\text{truth}} < 2000$ and $|\eta^{\text{truth}}| < 2.0$ in ranges of m/p_T .

For jets with $p_T^{\text{truth}} < 1000$ GeV, the baseline systematic uncertainties are the track-jet double-ratio uncertainties derived in Run-I. The MC-based derived systematic uncertainties for each of the changes between Run-I and Run-II on the jet p_T , mass and $D_2^{\beta=1}$ scale are shown in Fig. 5.5 for jets with $500 < p_T^{\text{truth}} < 1000$ GeV. The systematic uncertainties originating from the changes in the topo-clustering algorithm and the choice of sampling points for the LAr energy reconstruction are not shown in Fig. 5.5 as their impact on the different jet scales are negligible. The total MC-based uncertainty is obtained from the quadratic sum of three non-negligible components. For the boosted vector boson identification, the range $0.0 < m/p_T < 0.2$ is of particular interest for the p_T region shown in Fig. 5.5. The MC-based uncertainties in this phase-space are below 2% for the jet p_T scale, below 5% for the jet mass scale and less than 3% for the $D_2^{\beta=1}$ scale.

For jets with $p_T^{\text{truth}} > 1000$ GeV, no baseline uncertainties from the track-jet double-ratio method are available. Instead, the three physics list variations are considered in this high- p_T region as well as an additional modelling uncertainty using the generated HERWIG++ MC sample. The variation samples describing the changes in the topo-clustering algorithm as well as the choice of sampling points for the LAr energy reconstruction are not considered in the high- p_T region as they are only relevant for $p_T^{\text{truth}} < 1000$ GeV where the track-jet double-ratio method was used in Run-I. The MC-based derived systematic uncertainties on the jet p_T , mass and $D_2^{\beta=1}$ scale for each of the changes between Run-I and Run-II and the different physics lists are shown in Fig. 5.6 for jets with $1000 < p_T^{\text{truth}} < 1500$ GeV. The total MC-based uncertainty is obtained from quadrature sum of the seven non-negligible components. For the boosted vector boson identification, the range $0.0 < m/p_T < 0.1$ is of particular interest for the p_T region shown in Fig. 5.6. The uncertainties in this phase-space are below 3% for the jet p_T scale, below 6% for the jet mass scale and less than 8% for the $D_2^{\beta=1}$ scale and are dominated by the uncertainties due to the choice of physics list.

5.2.4 Final Jet Scale Uncertainties in 2015 Data

For the combination of the diboson resonance searches performed on the 2015 dataset (see Section 9.3), in-situ systematic uncertainties on the jet p_T and mass scale were derived using the track-jet double-ratio method with small modifications compared to what is described in Section 5.2.1. In 2015, the deviations of $\langle R_{\text{track-jet}}^m \rangle$ from unity for both MC generators, PYTHIA8 and HERWIG++, are taken into account and are used as two nuisance parameters to be propagated through the individual analyses. In addition to the uncertainties related to the track reconstruction efficiency, two further sources of systematic uncertainties on the track reconstruction were considered that were propagated $\langle r_{\text{track-jet}}^m \rangle$: a relative 50% uncertainty on the reconstruction rate of fake tracks and a possible momentum bias of the reconstructed jets [120, 121]. The resulting jet p_T and mass scale uncertainties are shown in Fig. 5.7 for jets with $p_T > 150$ GeV and $m/p_T = 0.1$. The in-situ uncertainties on the jet p_T and mass scale derived with 3.2 fb^{-1} of

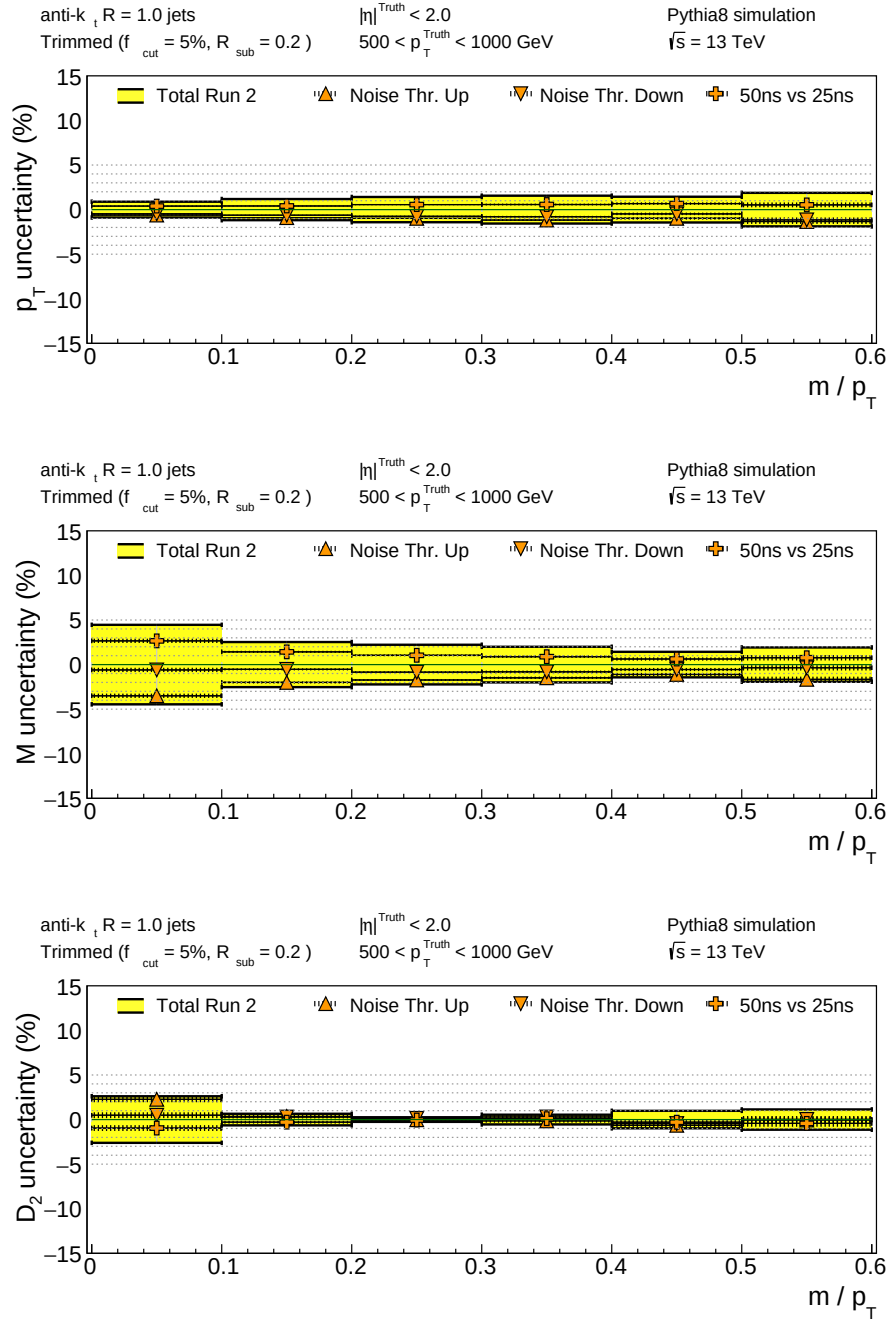


Figure 5.5: Systematic uncertainty on the p_T (top), mass (middle) and $D_2^{\beta=1}$ (bottom) scale for anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$ jets. The total uncertainty as well as the individual components are shown as a function of m/p_T for jets with $500 < p_T^{\text{truth}} < 1000$ GeV and $|\eta^{\text{truth}}| < 2.0$. The uncertainty components are derived from the disagreement of the jet response in the reference MC QCD dijet simulation and the different variation samples. The variations with a negligible uncertainty are not shown.

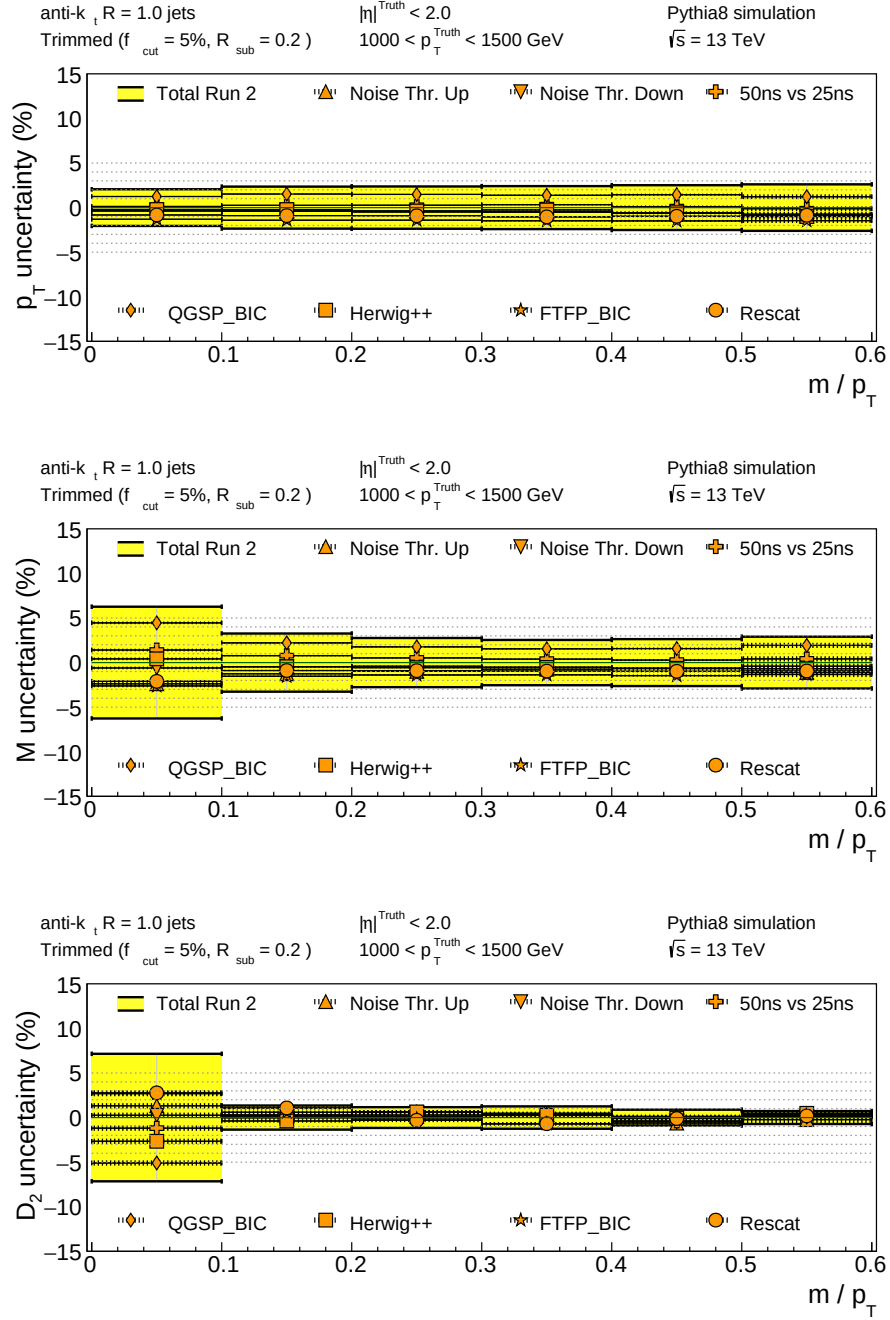


Figure 5.6: Systematic uncertainty on the p_T (top), mass (middle) and $D_2^{\beta=1}$ (bottom) scale for anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$ jets. The total uncertainty as well as the individual components are shown as a function of m/p_T for jets with $1000 < p_T^{\text{truth}} < 1500$ GeV and $|\eta^{\text{truth}}| < 2.0$. The uncertainty components are derived from the disagreement of the jet response in the reference MC QCD dijet simulation and the different variation samples. The variations with a negligible uncertainty are not shown.

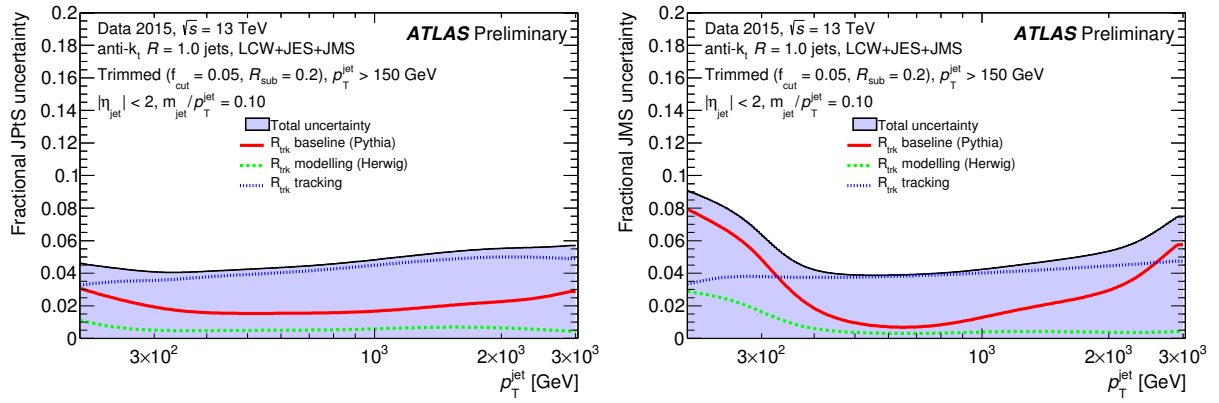


Figure 5.7: Fractional jet p_T (left) and mass (right) scale uncertainties for anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$ [121]. The total uncertainty and the individual components are shown as a function of the jet p_T for jets with $m/p_T = 0.1$ and were derived using data collected in 2015 with $\sqrt{s} = 13$ TeV. The individual components are described in Section 5.2.4.

$\sqrt{s} = 13$ TeV collisions are compared in Fig. 5.8 to the uncertainties, described in Section 5.2.3. Even though the MC-based uncertainties were derived for $p_T > 1500$ GeV, the usage of jets in analyses was restricted to $p_T < 1500$ GeV due to limitations in the jet mass calibration procedure until in-situ systematic uncertainties were derived. Therefore, the MC-based systematic uncertainties were set to zero for $p_T > 1500$ GeV. With the derivation of in-situ systematic uncertainties, the jet p_T range could be extended up to 3 TeV. The in-situ derived jet p_T scale uncertainties are slightly smaller than the MC-based systematics, but the differences are negligible. For the jet mass scale however, the systematic uncertainties could be reduced by about 2% in the interesting p_T range.

5.2.5 Summary of Systematic Uncertainties

In this section, the systematic uncertainties on the jet energy, mass and substructure scale of large- R jets were presented. In both Run-I and Run-II, the standard technique to derive systematic uncertainties is the track-jet double ratio method that uses track-jets, reconstructed from charged-particle tracks in the Inner Detector, as reference object. The systematic uncertainties are estimated by comparing the ratio of the calorimeter jet moment (e.g. mass) to that of the track-jet in data and simulation. Two Monte Carlo generators are used: PYTHIA8 and HERWIG++. Furthermore, the uncertainty on the reconstruction of charged-particle tracks is considered.

The systematic uncertainties for the diboson resonance search performed at 8 TeV, summarised in Chapter 8, were obtained with a dataset corresponding to 20.3 fb^{-1} of 8 TeV pp collisions. A different approach was used to derive systematic uncertainties for analyses that published results with 3.2 fb^{-1} of 13 TeV data at the end of 2015 before in-situ derived systematic uncer-

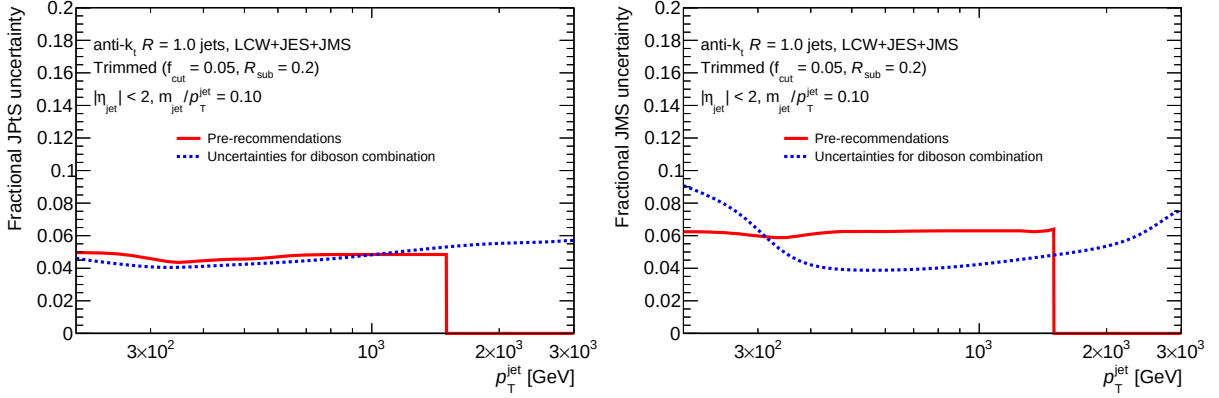


Figure 5.8: Comparison of the MC based (see Section 5.2.3) and data-driven (see Section 5.2.4) systematic uncertainties on the jet p_T (left) and jet mass (right) scale uncertainties for anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$ jets. The uncertainties are shown as a function of the jet p_T for jets with $m/p_T = 0.1$.

tainties were available as they take a considerable amount of time and effort. These uncertainties exploit the knowledge acquired during Run-I and add MC-based uncertainties that account for differences in the detector and jet reconstruction between Run-I and Run-II as described in Section 5.2.3. The systematic uncertainties considered in the diboson resonance search in Chapter 9 are based on this method.

The procedure for the derivation of in-situ based systematic uncertainties was slightly changed for Run-II with respect to Run-I. In Run-I, only the larger deviation of $\langle R_{\text{track-jet}}^m \rangle$ from unity for the two MC generators was used to assess the systematic uncertainties whereas both deviations are used as nuisance parameters in Run-II. In addition, two further sources of systematic uncertainties on the track reconstruction were considered in Run-II: a relative 50% uncertainty on the reconstruction rate of fake tracks and a possible momentum bias of the reconstructed jets. The systematic uncertainties, described in Section 5.2.4, are used for the combination of diboson resonance searches based on 3.2 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ pp collision data collected in 2015 (see Section 9.3).

In Run-I, systematic uncertainties on the jet energy scale were derived as well for anti- k_t $R = 1.0$ trimmed jets with $R_{\text{sub}} = 0.3$ and $f_{\text{cut}} = 5\%$ using photons as reference objects for jets with $p_T < 800 \text{ GeV}$. The resulting uncertainties are significantly smaller than those obtained with the track-jet double ratio method. In Run-II, these uncertainties are not yet available, however with the higher centre-of-mass energy, the reach of this method can be extended to higher transverse momenta. The γ +jet method provides however only a handle on the jet energy scale and not the jet mass scale. To decrease the uncertainties on the jet mass scale, a fit can be performed to the W boson resonance peak in $t\bar{t}$ events in the lepton+jet channel [70, 122, 123].

5.3 Future Improvements

Two possible improvements to the reconstruction of large- R jets are discussed in this section. Currently, the trimming algorithm is the standard technique in ATLAS to remove pile-up contaminations from large- R jets. However this algorithm removes only low- p_T subjets with respect to the ungroomed jet p_T and thus energy deposits from pile-up vertices can still contaminate the subjets that are kept in the trimming procedure. The jet area pile-up correction can be used to subtract pile-up contributions from the subjets within the large- R jet before the trimming algorithm is applied. In addition, the track-assisted jet mass is introduced to overcome the limitations of the mass reconstruction of highly boosted vector bosons due to the calorimeter granularity and thus improve the mass resolution at high transverse momentum.

5.3.1 Jet Area Pile-up Correction

Pile-up can have two different effects on the reconstruction of jets. The jet multiplicity increases due to additional pile-up vertices and further energy is added to the topo-clusters from the hard-scatter vertex causing an increase of the jet energy. As previously described, the topo-clustering algorithm suppresses the creation of additional low-energy topo-clusters from pile-up if the noise threshold is adjusted accordingly [71]. However if a topo-cluster is seeded by energy deposits from the hard-scatter vertex, soft contributions to this topo-cluster from pile-up vertices cannot be removed. Therefore, further techniques to suppress the contamination of pile-up are needed. An obvious choice would be to associate charged-particle tracks originating from the primary vertex to the topo-clusters. However this method cannot account for additional topo-clusters from neutral particles from pile-up vertices. Instead, each jet is corrected based on its susceptibility to pile-up, defined by the *jet area* A and the event pile-up p_T density ρ [104, 124]. The jet-by-jet corrected four-vector is then given by:

$$p_\mu^{\text{corrected}} = p_\mu - \rho \cdot A_\mu \quad (5.8)$$

This pile-up suppression technique only corrects the jet four-vector but since no energy is subtracted from the topo-clusters, the jet substructure variables remain unchanged. Several different techniques to correct jet shape variables or to remove pile-up contributions from topo-clusters have been proposed and can be found in Ref. [125–128] but will not be discussed here.

The pile-up event p_T density ρ can be interpreted as the transverse momentum per unit area added to the event by pile-up. It is calculated on an event-by-event basis as the median of the p_T/A distribution of jets reconstructed with the k_t algorithm using a radius parameter of $R = 0.4$:

$$\rho = \text{median} \left\{ \frac{p_{T,i}^{\text{jet}}}{A_i^{\text{jet}}} \right\} . \quad (5.9)$$

The k_t algorithm was chosen due to its sensitivity to soft radiation and the formation of regular pile-up jets under the assumption of a uniform soft particle background. No minimal p_T criterion

is imposed on the jets used for the ρ calculation, however only topo-clusters with $|\eta| < 2.0$ are considered as inputs to the reconstruction of k_t jets. The calculation of ρ in the forward region is more challenging as the calorimeter cell size becomes coarser. Therefore, the probability of clusters in the forward region to be seeded by pile-up noise decreases and thus the ρ distribution which is approximately flat in the central region drops significantly for $|\eta| > 2.5$ and the p_T/A distribution is no longer distributed around the expected ρ . The average amount of p_T added to the jet due to the underlying event and pile-up effects is approximately 5 and 20 GeV per unit area for an average number of one and 26 interactions per bunch crossing respectively [104].

The jet *catchment area* defines the region in the η - ϕ space in which additional energy due to the underlying event or pile-up effects the jet kinematics. Since the inputs to the jet clustering algorithm are point-like particles and thus cannot span an area, more sophisticated techniques are needed to determine the jet area [129]. The ATLAS Collaboration uses the *active area* method [104]. In addition to the actual topo-clusters, a grid of randomly distributed particles in the η - ϕ space of the detector with infinitesimal transverse momentum, so-called *ghosts*, are added to the list of inputs to the jet reconstruction algorithm. If the jet reconstruction algorithm is infrared safe, the four-vector of the resulting jets does not change despite adding the ghost particles. The active area of a jet is then determined as the ratio of the number of ghosts clustered inside the jet to the number of ghosts per unit area ν_g . The jet area depends on the initial number of ghosts added to the jet reconstruction algorithm. To define a stable version of the jet area, a dense coverage is needed (i.e. $\nu_g \rightarrow \infty$) and the actual area is determined from the average of many measurements with different sets of ghosts. The four-vector version of the jet area A_μ is then defined as

$$A_\mu = \lim_{\nu_g \rightarrow \infty} \left\langle \frac{1}{\nu_g \langle g_t \rangle} \sum_{g_i \in J} g_{\mu i} \right\rangle, \quad (5.10)$$

where the sum runs over all ghosts g_i clustered within the jet J and $\nu_g \langle g_t \rangle$ is the ghost transverse momentum density. Assuming that additional energy due to pile-up or the underlying event is uniformly distributed in the detector, the jet area provides a measure of its susceptibility to pile-up because the corresponding pile-up topo-clusters will be clustered inside the jet analogously to the ghosts.

The jet mass distribution of anti- k_t $R = 1.0$ jets in a QCD dijet sample is depicted in Fig. 5.9 for three different pile-up conditions as they are expected for the high-luminosity LHC: $\langle \mu \rangle = 80, 140$ and 200. For jets that are only trimmed with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.3$, the mass distributions for the different pile-up scenarios are significantly different. With increasing $\langle \mu \rangle$, the background mass distribution does not show the expected behaviour for trimmed jets (Sudakov peak with falling mass distribution) but exhibits a second peak around the top-quark mass which would increase the background efficiency for a top tagging algorithm and result in a worse performance. If the subjects are however corrected beforehand with the jet area subtraction technique, the background mass distributions for the three pile-up scenarios are almost identical and much

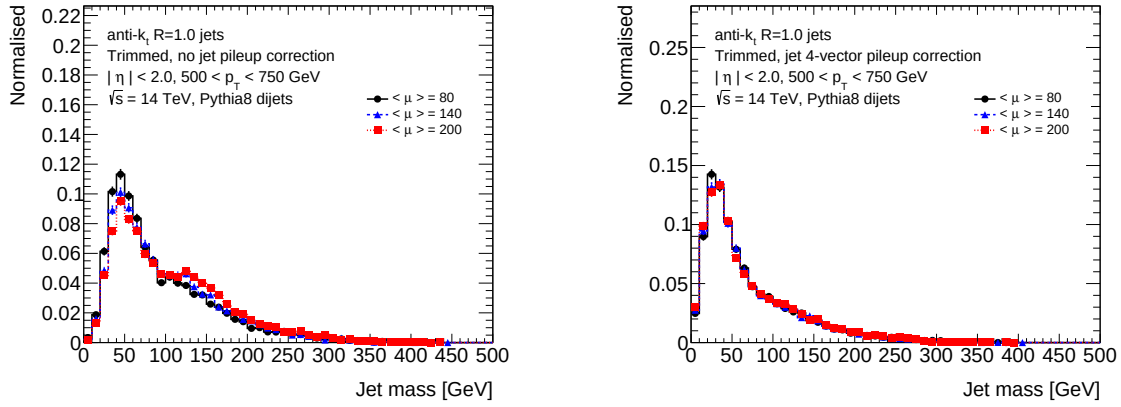


Figure 5.9: Jet mass distribution for anti- k_t $R = 1.0$ jets in a QCD dijet sample, trimmed with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.3$ (left) and with an additional jet area correction of the subjets (right) for three different pile-up scenarios for the high-luminosity LHC conditions.

narrower compared to the trimmed distributions.

Although the focus of this thesis is large- R jets, the jet area subtraction can also be applied to small- R jets. So far, only the pile-up correction of the jet p_T is part of the standard small- R ($R \approx 0.4$) jet calibration scheme [130]. With the increasing boost, vector bosons and other particles can be reconstructed as small- R jets at high- p_T and hence the stability of the jet mass to pile-up contamination is becoming more important. No grooming techniques have been studied yet for small- R jets and the trimming algorithm for example requires the reconstruction of significantly smaller subjets which will be close to the calorimeter granularity. Figure 5.10 shows the mean jet mass response for anti- k_t $R = 0.4$ jets as a function of the number of reconstructed vertices in a QCD dijet sample. The mass response is defined as the ratio of the calorimeter jet mass to the matched truth jet mass where the matching is simply geometrical in the η - ϕ space as explained in Section 5.1. Three different types of jets are shown: without any pile-up correction, after scaling the full four-vector by a factor such that the corrected p_T of the jet corresponds to $p_T^{\text{SF}} = p_T - \rho \cdot A_T$ and for jets with the full four-vector jet area subtraction. For each jet, independent of the correction applied to it, the denominator is the same because the generator-level (truth) information does not include any pile-up. The small- R jet response increases approximately linearly with the number of reconstructed vertices for uncorrected jets. The slope decreases slightly after the simple scale factor correction is applied to the jet's four-vector based on the simple jet p_T correction. After the full jet area correction is applied to the jet four-vector, the mean jet mass response does not show any dependence on the pile-up conditions. Furthermore, the correction reduces the tails in the Gaussian-distributed jet mass response caused by increased reconstructed jet mass due to pile-up contamination.

The jet area subtraction can result in unphysical negative jet masses. This effect occurs in events with a large pile-up density ρ and mostly low- p_T jets with relatively large jet areas but

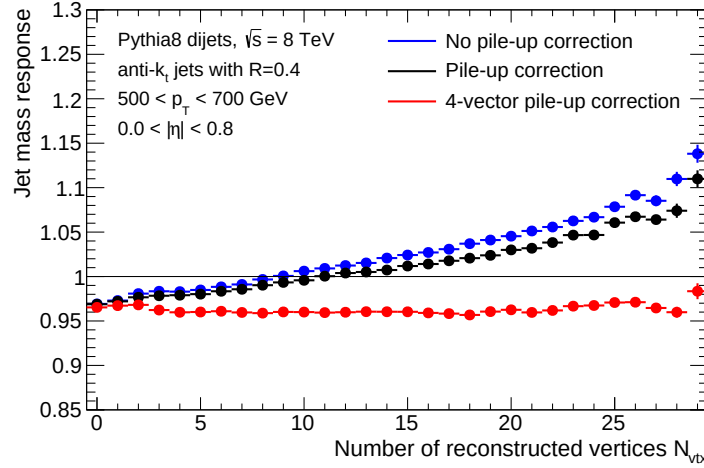


Figure 5.10: Comparison of the jet mass response as a function of the number of reconstructed vertices for anti- k_t $R = 0.4$ jets without any pile-up correction, after correcting the p_T of the jet, and after the full 4-vector jet area correction.

small masses. It is not particularly obvious how to treat these jets with negative jet masses. One possibility would be to set the mass explicitly to zero. These jets would be discarded then in analyses that use the jet mass as variable to distinguish between boosted vector bosons and QCD-jets. A second possibility would be to not correct the full four-vector in case this would result in a negative jet mass but to only apply the p_T scale factor correction. It is not obvious though if this technique would still be able to remove any pile-up contamination from the jet. Further studies are needed to understand why the jets mass is overcorrected in some cases to be included as standard procedure in that ATLAS calibration chain for small- R jets.

5.3.2 Track-assisted Jet Mass

The track-assisted jet mass takes advantage of the good angular resolution of charged-particle tracks to improve the mass resolution. To define the track-assisted jet mass, tracks need to be associated to the calorimeter jet via *ghost association*. The procedure is similar to the determination of the active area in Section 5.3.1. Instead of a dense grid of particles, the reconstructed tracks are added with infinitesimal small p_T to the inputs of the jet reconstruction algorithm. A track is then associated to a jet in case it is clustered into the jet. The jet mass calculated solely from the ghost-associated tracks, m_{trk} cannot provide an accurate definition of the jet mass because it only incorporates the charged-particle fraction and neglects the neutral components. As it has been previously seen however, the jet mass scales with the transverse momentum. Therefore, the track-jet mass m_{trk} is scaled by the ratio between the calorimeter jet p_T and the track-jet p_T (p_T^{trk}) to account for the missing neutral component. The track-assisted

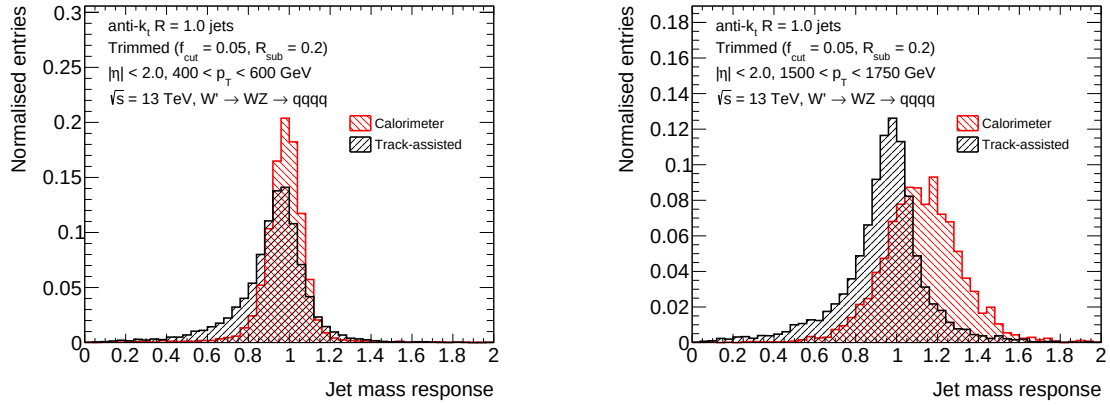


Figure 5.11: Comparison of the jet mass response calculated with the calorimeter and the track-assisted mass for anti- k_t $R = 1.0$ trimmed jets in a $W' \rightarrow WZ$ sample with $400 < p_T < 600$ GeV (left) and $1500 < p_T < 1750$ GeV (right).

mass m^{TA} is then defined as

$$m^{\text{TA}} = \frac{p_T^{\text{reco}}}{p_T^{\text{trk}}} \times m_{\text{trk}} . \quad (5.11)$$

The jet mass response is compared in Fig. 5.11 for anti- k_t $R = 1.0$ trimmed jets using the calorimeter and track-assisted mass definition in the numerator and the same matched truth jet mass in the denominator. For low- p_T jets, the resolution of calorimeter jet mass is better than that of the track-assisted jet mass. At high p_T however, the improved angular resolution of the Inner Detector improves the jet mass resolution significantly compared to the calorimeter jet mass. Furthermore, the track-assisted jet mass is already well calibrated and its response is much closer to unity than for the calorimeter jet mass.

A comparison of the jet mass resolution as a function of the jet p_T is shown in Fig. 5.12 for the calorimeter jet mass and the track-assisted mass definition. The mass resolution is defined here as the ratio of the interquartile range to the median of the mass response as it is more stable with respect to the RMS for non-Gaussian distributions. In addition another curve is depicted that combines the calorimeter and the track-assisted jet mass to take advantage of the good calorimeter resolution at low p_T and the good angular resolution of the Inner Detector at high p_T . The two mass definitions are combined linearly weighted with their corresponding inverse mass resolution squared:

$$m^{\text{comb}} = \frac{\sigma_{\text{calo}}^{-2}}{\sigma_{\text{calo}}^{-2} + \sigma_{\text{TA}}^{-2}} \times m_{\text{calo}} + \frac{\sigma_{\text{TA}}^{-2}}{\sigma_{\text{calo}}^{-2} + \sigma_{\text{TA}}^{-2}} \times m_{\text{TA}} . \quad (5.12)$$

The combination of the two mass definitions results in a better mass resolution over the full p_T range than for the calorimeter or track-assisted jet mass. The idea of using information from the calorimeter as well as the Inner Detector and to combine them has already been used previously

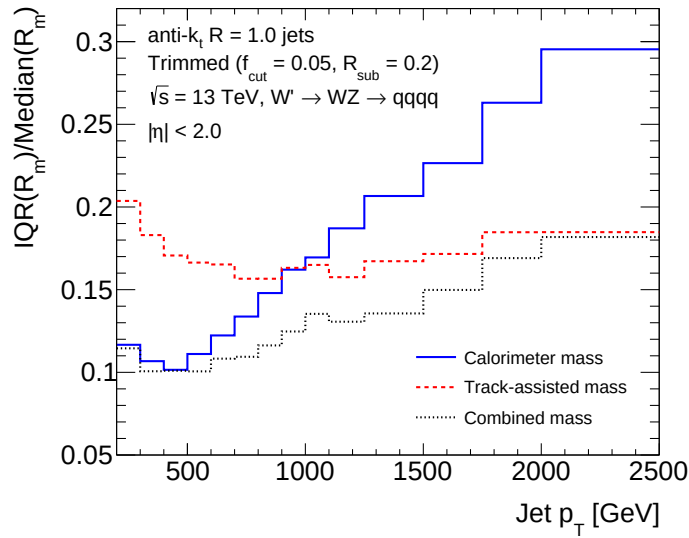


Figure 5.12: Jet mass resolution as a function of the jet p_T for three different definitions of the jet mass: the reconstructed jet mass, track-assisted jet mass and a linear combination of the two masses each weighted by the square of the inverse mass resolution. The mass resolution is defined as the interquartile range to the median of the mass response.

in the *particle flow* algorithm that is the standard reconstruction technique for jets in the CMS Collaboration [92, 93] and is currently under study in the ATLAS Collaboration. The particle flow algorithm could also improve the resolution of substructure variables at high p_T .

Chapter 6

Identification of Boosted Vector Bosons

The successful identification of boosted hadronically decaying vector bosons is of utmost importance to enhance the sensitivity of analyses such as the diboson resonance searches, presented in Chapter 8 and 9, which are dominated by the production of multi-jet events. An algorithm to effectively identify boosted W/Z bosons has three main ingredients in the ATLAS Collaboration: a grooming algorithm to remove pile-up contaminations from the large- R jet, a requirement on the jet mass and the imposition of criteria on substructure variables that allows for the discrimination of a QCD-jet from a W/Z -jet. Over the past years a large number of grooming algorithms and substructure variables have been presented. To determine the most useful techniques, a MC-based optimisation has been carried out by the ATLAS Collaboration at $\sqrt{s} = 8$ TeV to provide a detailed comparison of the majority of these techniques [86]. For one benchmark grooming algorithm, a selection of tagging techniques was validated in data using 20.3 fb^{-1} of pp collisions collected at a centre-of-mass energy of 8 TeV. The results of this large body of work are summarised in Section 6.1.

The boosted vector boson identification algorithm was then optimised for Run-II with $\sqrt{s} = 13$ TeV based on the results obtained in Section 6.1. As detailed in Section 6.2, the vector boson tagging algorithm defines two signal efficiency working points (50% and 25%) based on anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$ using the substructure variable $D_2^{\beta=1}$. This algorithm is the default algorithm for analyses in 2015 with boosted vector bosons in the final state such as the all-hadronic diboson resonance search in Chapter 9.

Finally, future improvements for the identification of boosted hadronically decaying vector bosons are discussed in Section 6.3.

6.1 Performance of Boosted Boson Identification at $\sqrt{s} = 8$ TeV

A detailed comparison of the majority of boosted vector boson identification techniques has been carried out in a two-stage MC-based optimisation procedure [86]. In the first step, the optimal grooming configurations are determined based on the groomed jet mass distribution and its pile-up dependence. In the second stage, the different substructure variables are investigated within a mass window containing 68% of the signal around the W boson mass and two different signal efficiency working points (50% and 25%) are defined. Only the optimisation of

W boson identification has been studied, however the techniques can be as well applied to identify Z bosons. One grooming algorithm is then selected in combination with three substructure variables that provide a large background rejection (inverse of the background efficiency) for the 50% signal efficiency working point. The signal and background efficiencies determined in MC for the chosen W -tagger (the combination of a grooming algorithm along with selection criteria on the mass and one further substructure variable) are then validated in data using 20.3 fb^{-1} of pp collisions collected at $\sqrt{s} = 8 \text{ TeV}$.

6.1.1 MC-based Optimisation

The aim of this section is to find a W -tagger that provides a high background rejection while retaining a high signal efficiency based on MC simulations. The process $W' \rightarrow WZ \rightarrow qq\ell\ell$ is used to provide a clean set of hadronically decaying W bosons whose substructure properties are compared to background jets produced in QCD dijet events. The signal samples were generated with PYTHIA8 with the AU2 tune and MSTW20080LO [131] PDF set. To cover a broad kinematic range of boosted W bosons, the signal samples were produced with W' masses ranging from 400 GeV to 2000 GeV. The leading p_T ungroomed truth jet (using stable particles as input to the jet algorithm) reconstructed with the C/A algorithm and a radius parameter of $R = 1.2$ is used as a proxy for the W boson. Its p_T distribution is reweighted to the corresponding p_T distribution in the QCD dijet background sample. The background samples were simulated with PYTHIA8 using the AU2 tune and the CT10 PDFs. As various grooming algorithms affect the large- R jet in different ways and can thus result in different transverse momenta, a change of the substructure properties is expected as well due to their correlation with p_T . Therefore the events are categorised in p_T ranges of the leading ungroomed truth jet p_T^{Truth} to allow for a fair comparison of the various grooming techniques. Only the leading jet in each event is used for these studies as otherwise the signal would be comprised of background jets as well. Three different p_T^{Truth} ranges are considered: $[200, 350)$, $[350, 500)$ and $[500, 1000)$ GeV and jets are restricted to $|\eta^{\text{Truth}}| < 1.2$ to be fully contained in the acceptance of the inner tracking detector, needed for the derivation of systematic uncertainties. The jet energy and mass are not calibrated in the MC-based optimisation studies.

Jet Grooming Comparison

In the first stage, more than 500 different grooming configurations were studied. The definition of a grooming configuration comprises the jet reconstruction algorithm, the radius parameter R , the grooming technique and the chosen parameters of the grooming algorithm. To rank the various grooming configurations, the mass distribution of the leading groomed jet is examined for signal and background jets. A good grooming algorithm is required to have a low background efficiency $\varepsilon_{\text{QCD}}^G$ within a mass window defined around the W boson mass. This mass window is defined as the smallest interval that contains 68% of the signal events, $\varepsilon_W^G = 68\%$. The superscript G is used here to indicate that only grooming techniques are applied to the jet. The

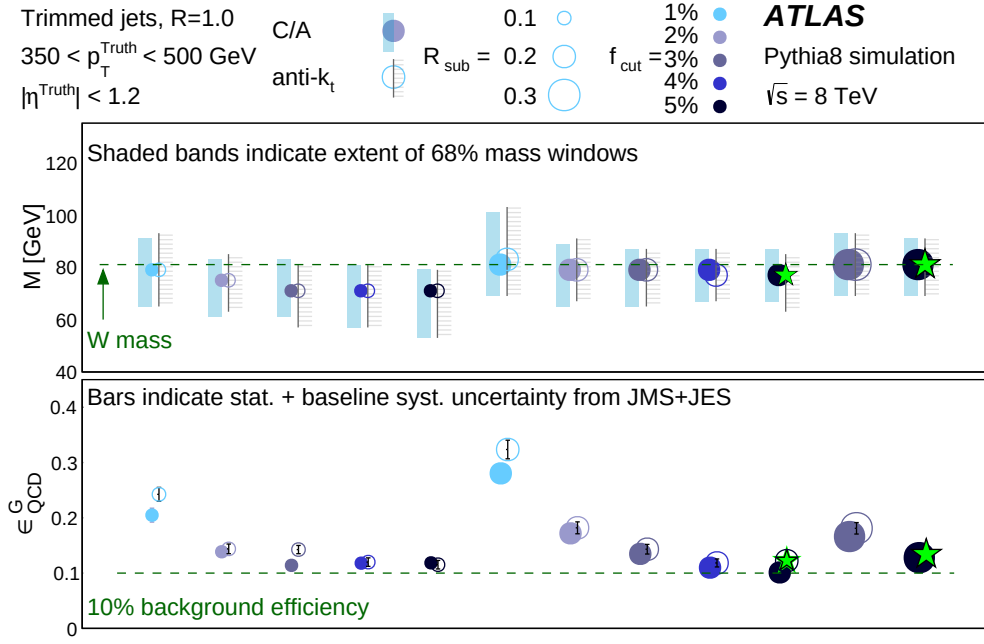


Figure 6.1: Summary of the signal mass window size and background efficiency for different jet algorithms and trimming configurations for jets in the range $350 < p_T < 500$ GeV [86]. The points in the upper panel show the most probable value of the W -jet mass distribution and the shaded bands indicate the size of the smallest mass window containing 68% of the signal. The lower panel displays the background efficiency within the 68% mass window. The uncertainty on the background efficiency is obtained for each jet collection by varying the jet mass and energy by their respective uncertainty. Two different jet reconstruction algorithms are shown, the C/A algorithm (filled circles) and the anti- k_t algorithm (empty circles). The size of the points indicate the subjet radius used in the trimming procedure and the colours represent different f_{cut} criteria. The stars illustrate trimming configurations with a high background rejection and good stability versus pile-up.

denominator in both the signal and background efficiency definitions is the number of events that have an ungroomed C/A $R = 1.2$ truth jet with $p_T^{\text{Truth}} > 200$ GeV and $|\eta^{\text{Truth}}| < 1.2$. Besides the low background efficiency for a fixed signal efficiency, the mass distribution should not be susceptible to pile-up contamination and should also fulfil certain quality requirements. First, the signal jet mass distribution should peak around the W boson mass as much lower values would indicate that the W boson decay is not fully contained within the large- R jets. Second, the W -jet mass distribution is required to be relatively Gaussian without pronounced tails to lower jet masses and thus resulting in a symmetric mass window. Furthermore grooming algorithms for which the QCD-jet mass distribution exhibit a local maxima within the W boson mass window should be avoided.

The average of the signal jet mass distribution and the size of the 68% mass window as well as $\varepsilon_{\text{QCD}}^G$ are shown for various trimming, split-filtering and pruning configurations in Fig. 6.1, 6.2 and Fig. 6.3, respectively for jets with $350 < p_T^{\text{Truth}} < 500$ GeV and $|\eta^{\text{Truth}}| < 1.2$. Although

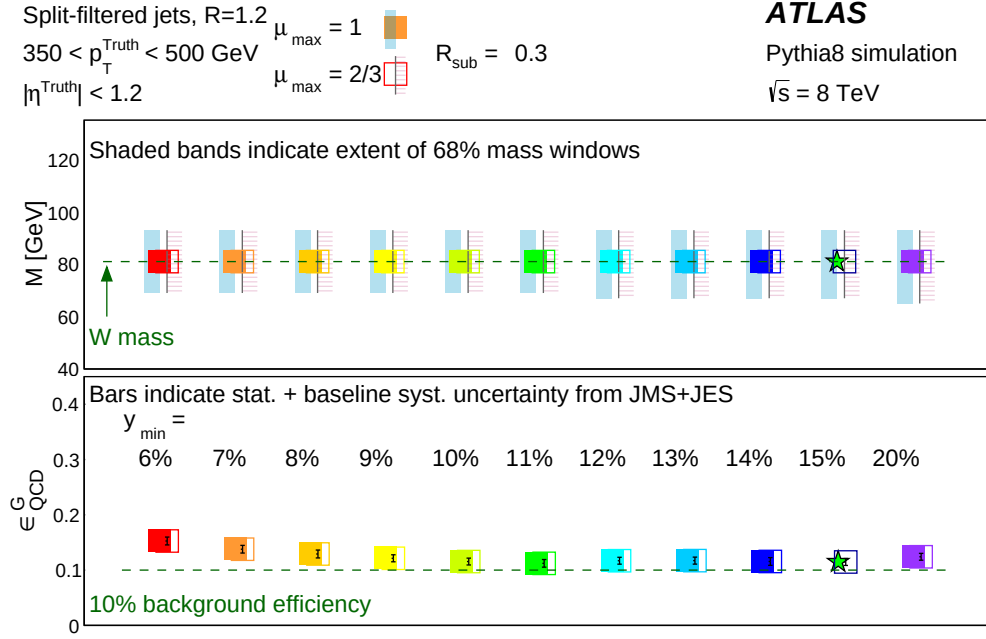


Figure 6.2: Summary of the signal mass window size and background efficiency for C/A $R = 1.2$ jets with different split-filtering configurations in the range $350 < p_T < 500 \text{ GeV}$ [86]. The points in the upper panel show the most probable value of the W -jet mass distribution and the shaded bands indicate the size of the smallest mass window containing 68% of the signal. The lower panel displays the background efficiency within the 68% mass window. The uncertainty on the background efficiency is obtained for each jet collection by varying the jet mass and energy by their respective uncertainty. Two different mass drop parameters μ_{max} are shown, $\mu_{\text{max}} = 1$ (filled squares) and $\mu_{\text{max}} = 2/3$ (empty squares). The colours represent different y_{min} criteria (denoted as y_{cut} in Section 4.4.2), ranging from 6% (left) to 20% (right). The stars indicate split-filtering configurations with a high background rejection and good stability versus pile-up.

several different jet radius parameters have been studied, ranging from 0.6 to 1.2, only one specific jet radius parameter per grooming technique is shown here. The stars in the different figures indicate the four jet grooming configurations with a low background efficiency as well as a low pile-up dependence of the average jet mass that increases by significantly less than one GeV per reconstructed vertex (only for the pruned configuration for jets with $200 < p_T^{\text{Truth}} < 350 \text{ GeV}$, the pile-up dependence of the jet mass is of the order of 1 GeV per reconstructed vertex).

For trimmed jets, the anti- k_t and C/A algorithms have relatively similar background efficiencies and signal mass window sizes. The background efficiency can be significantly reduced using larger f_{cut} values. For anti- k_t $R = 1.0$ trimmed jets with $R_{\text{sub}} = 0.2$, the background efficiency reduces from approximately 30% for $f_{\text{cut}} = 1\%$ to 10% using $f_{\text{cut}} = 5\%$. For fixed f_{cut} values, a larger subjet radius size R_{sub} results in a larger background efficiency.

The peak of the W -jet mass distribution as well as the background efficiency does not depend on the mass-drop criteria μ_{max} for split-filtered jets. The background efficiency decreases for increasing y_{cut} values, denoted as y_{min} in Fig. 6.2 but for $y_{\text{cut}} \geq 0.09$ the background rejection

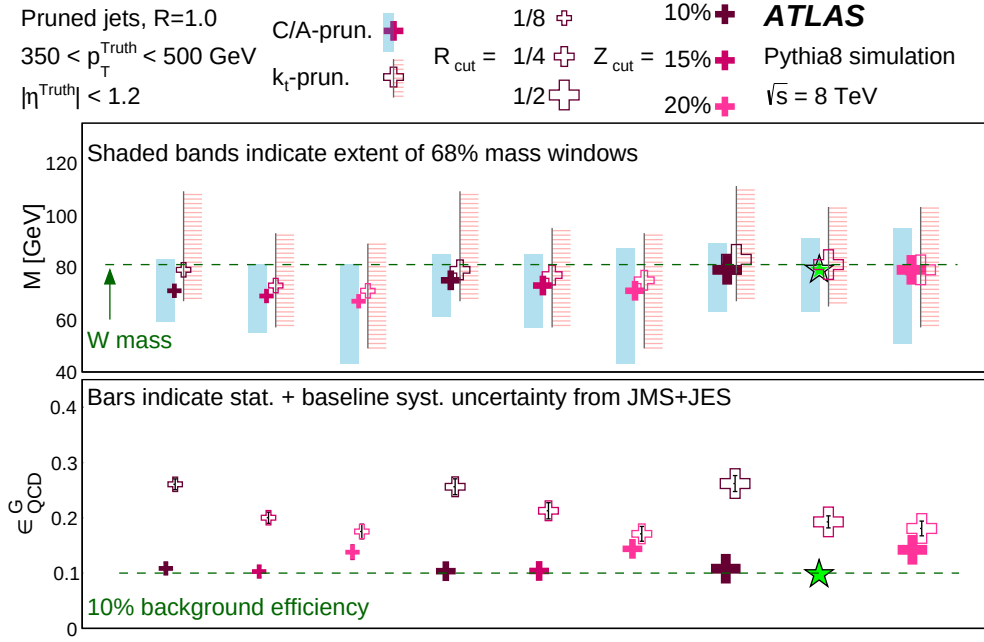


Figure 6.3: Summary of the signal mass window size and background efficiency for different jet algorithms and pruning configurations for jets in the range $350 < p_T < 500$ GeV [86]. The points in the upper panel show the most probable value of the W -jet mass distribution and the shaded bands indicate the size of the smallest mass window containing 68% of the signal. The lower panel displays the background efficiency within the 68% mass window. The uncertainty on the background efficiency is obtained for each jet collection by varying the jet mass and energy by their respective uncertainty. Two different jet reconstruction algorithms are shown, the C/A algorithm (filled crosses) and the k_t algorithm (empty crosses). The size of the points indicate the subjet radius used in the pruning procedure and the colours represent different z_{cut} criteria. The stars illustrate pruning configurations with a high background rejection and good stability versus pile-up.

is stable and does not improve significantly.

For pruned jets, the initial jets have been reconstructed with the C/A algorithm, however both the k_t and C/A algorithms have been studied for the reclustering of the input jet. Using the C/A algorithm for reclustering results in a significantly smaller background efficiency than for the k_t algorithm for the same R_{cut} and Z_{cut} values. Furthermore a R_{cut} value of 50% results in signal jet masses that are the closest to the W boson mass compared to other R_{cut} values. For the C/A algorithm, $Z_{\text{cut}} = 20\%$ yields a larger background efficiency than smaller values.

The same conclusions on the parameters of the grooming algorithms hold as well for the other p_T^{Truth} ranges that have been studied. Overall, the size of the 68% signal mass window as well as the background efficiency decrease with increasing p_T^{Truth} . This effect is more pronounced for pruned jets reconstructed with the C/A algorithm than with the k_t algorithm. The mean of the signal jet mass distribution gets closer to the true W boson mass at high- p_T^{Truth} which would result overall in smaller calibration factors. Furthermore, the dependence of the average

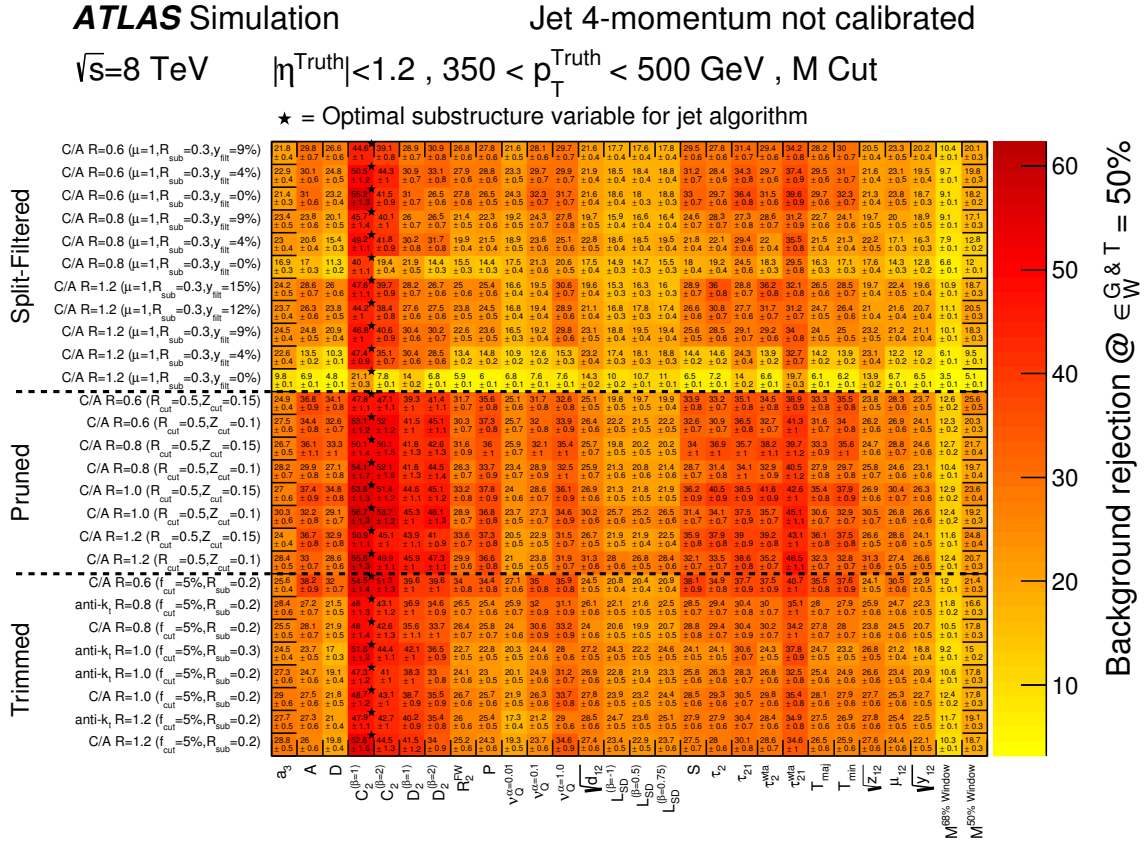


Figure 6.4: Background rejection corresponding to a signal efficiency of $\epsilon_W^{\text{G\&T}} = 50\%$ for different combinations of substructure variables and grooming algorithms for jets with $350 < p_T^{\text{Truth}} < 500$ GeV after applying the mass window corresponding to a signal efficiency of $\epsilon_W^{\text{G}} = 68\%$ [86].

jet mass on the number of reconstructed vertices decreases for higher p_T^{Truth} ranges.

Jet Substructure Comparison

Out of the initial 500 different grooming configurations, only 27 configurations are selected based on the previously discussed criteria to proceed to the second stage of the optimisation in which the mass window criteria is combined with another substructure variable. Overall, 26 different substructure variables have been studied in combination with the 27 grooming configurations to find the pairwise combinations that result in the smallest background efficiency $\epsilon_{\text{QCD}}^{\text{G\&T}}$ for a defined working point. he superscript G&T is used to indicate that both grooming techniques and tagging selections are applied to the jet. Single-sided selection criteria on the substructure variable are imposed, selecting 73.5% of the signal events, to define an overall *medium* working point of $\epsilon_W^{\text{G\&T}} = 68\% \cdot 73.5\% = 50\%$. The inverse of the background efficiencies $\epsilon_{\text{QCD}}^{\text{G\&T}}$ for the different pairwise combinations of a grooming configuration and a substructure variable for the medium working point are shown in Fig. 6.4 for jets with $350 < p_T^{\text{Truth}} < 500$ GeV. In addition

the background rejection values corresponding to $\varepsilon_W^G = 68\%$ are included for each of the grooming configurations to show the improvement in background rejection that can be achieved by using a substructure variable that explores the inner structure of the jet. The mass window size can also be reduced to result in a signal efficiency of 50% but the background rejections achieved with the narrower mass window size are significantly smaller than the ones which also use tagging techniques. The substructure variables that have not been previously defined in Chapter 4.5 but are shown in Fig. 6.4 can be found in Appendix A. For most of the grooming configurations, the substructure variables $C_2^{\beta=1}$, $D_2^{\beta=1}$ and τ_{21}^{wta} have the highest background rejection. The variable resulting in the highest background rejection for each grooming configuration, $C_2^{\beta=1}$, is indicated by the star in Fig. 6.4.

Based on the pile-up dependence of the grooming configurations and the background rejection for the 50% signal efficiency working point, four different grooming configurations in combination with three different substructure variables, $C_2^{\beta=1}$, $D_2^{\beta=1}$ and τ_{21}^{wta} , were found to be particularly good in identifying boosted hadronically decaying W bosons. The selected grooming configurations are as follows:

- anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$, $R_{\text{sub}} = 0.2$,
- anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$, $R_{\text{sub}} = 0.3$,
- C/A $R = 1.0$ pruned jets with $Z_{\text{cut}} = 0.15$, $R_{\text{cut}} = 0.5$,
- C/A $R = 1.2$ split-filtered jet with $y_{\text{cut}} = 0.15$, $R_{\text{sub}} = 0.3$.

Although only results for jets with $350 < p_T^{\text{Truth}} < 500$ GeV were shown here, the same conclusion holds for the two other p_T^{Truth} ranges considered in the studies performed in Reference [86].

6.1.2 Performance in Data

A benchmark grooming configuration of anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$ is studied in combination with the three substructure variables ($C_2^{\beta=1}$, $D_2^{\beta=1}$ and τ_{21}^{wta}). This benchmark is used to measure the performance in data and to compare it to the predictions from MC simulations. Hadronically decaying W bosons are reconstructed from an enriched sample of top anti-top quark pair ($t\bar{t}$) events in the lepton-plus-jets channel whereas the background is measured in a multi-jet enriched data sample. Details about the exact object definition and event selections can be found in Reference [86].

The energy and mass of the jets are calibrated using MC simulations as described in Chapter 5.1. To allow for a rigorous data and MC simulation comparison, systematic uncertainties are derived for the energy, mass, and the three substructure variables using track-jets as reference objects as detailed in Section 5.2.

The comparison of the $C_2^{\beta=1}$, $D_2^{\beta=1}$ and τ_{21}^{wta} variables for jets with $p_T > 200$ GeV in data and MC simulations for the $t\bar{t}$ enriched data samples is depicted in Fig. 6.5 after applying the mass window selection criteria corresponding to $\varepsilon_W^G = 68\%$ in MC simulations. The dominant

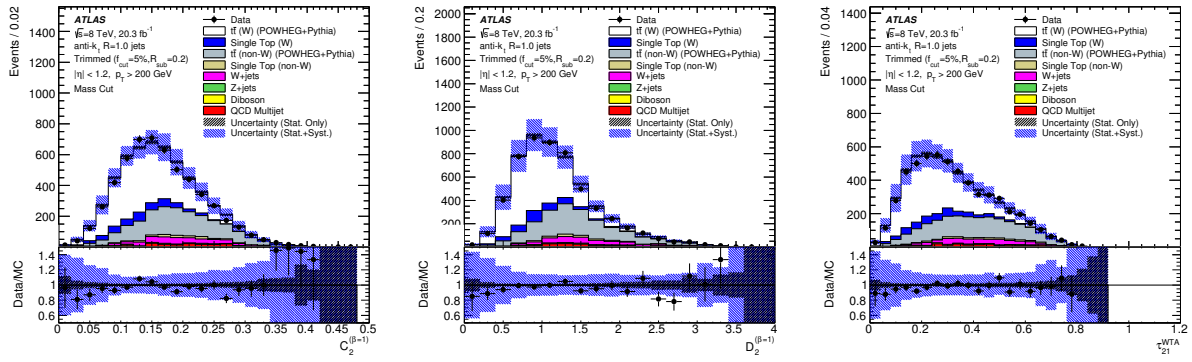


Figure 6.5: Comparison of $C_2^{\beta=1}$ (left), $D_2^{\beta=1}$ (middle) and τ_{21}^{wta} (right) in data and MC simulation for anti- k_t $R = 1.0$ trimmed jets in a lepton-plus-jets enriched $t\bar{t}$ sample for the combined electron and muon channel [86]. A mass window corresponding to a signal efficiency of $\varepsilon_W^G = 68\%$ is applied.

background comes from $t\bar{t}$ events, where the W boson is not fully contained within the radius size of $R = 1.0$ (labelled as non- W background). Good agreement between the data and MC simulation is observed for all three variables except for the tails of the $C_2^{\beta=1}$ distribution.

The same substructure variables for the background enriched data set are shown in Fig. 6.6 using PYTHIA8 and HERWIG++ for the MC comparison. Whereas the $D_2^{\beta=1}$ and τ_{21}^{wta} are well modelled by PYTHIA8, larger discrepancies can be observed for the HERWIG++ generator. For the $C_2^{\beta=1}$ variable, both MC generators show large discrepancies with respect to the data. In particular, the $C_2^{\beta=1}$ distribution in PYTHIA8 seems to be shifted, causing the slope in the data over MC ratio. However, the discrepancies are mostly covered within the large systematic uncertainties.

The signal and background efficiency are measured in data for the medium working point and also for a further *tight* working point corresponding to $\varepsilon_W^{G\&T} = 25\%$. These working points are based on MC simulations using a mass window around the W boson mass and a single-sided substructure criteria as previously described. The signal efficiencies in data are extracted using template fits to the jet mass distribution for jets that passed and failed the substructure criteria for the working point under study. Two templates are constructed from MC simulations for the signal and the background processes whose normalisations are allowed to float in the fit. The signal efficiency is only measured up to $p_T = 500$ GeV due to the lack of produced top quark pairs in data at high- p_T .

The measured signal and background efficiencies in data are shown in Fig. 6.7 as a function of the jet p_T for the 50% working point with the $D_2^{\beta=1}$ variable. The $D_2^{\beta=1}$ variable was chosen here as the tagging variable as it provides the highest background rejection for jets with $p_T > 350$ GeV. The data measurements are compared to the MC predictions for multi-jet and top quark pair events, where each process is simulated with two different MC generators. Multi-jet events are

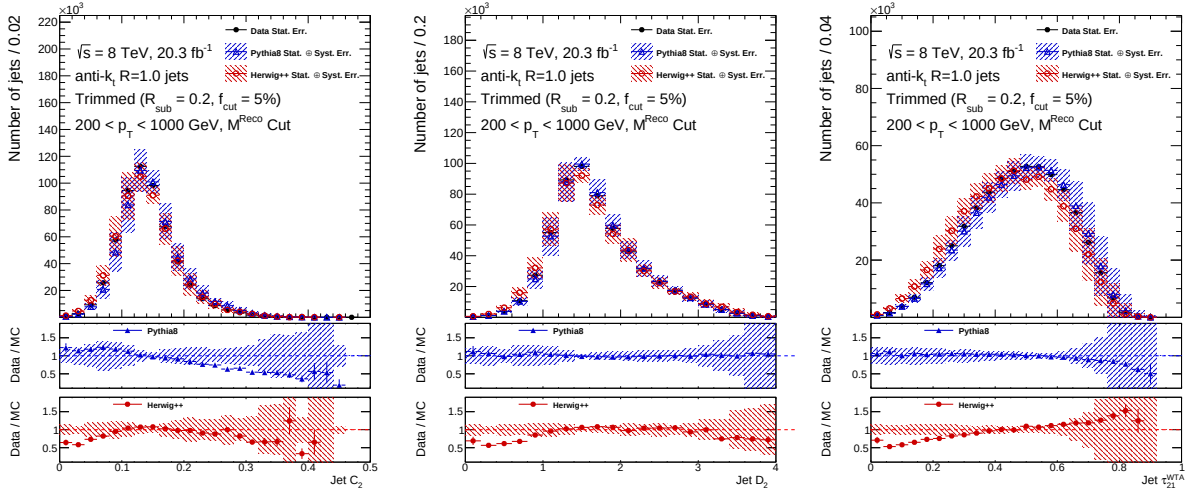


Figure 6.6: Comparison of $C_2^{\beta=1}$ (left), $D_2^{\beta=1}$ (middle) and τ_{21}^{wta} (right) in data and MC simulation for anti- k_t $R = 1.0$ trimmed jets in a multi-jet enriched sample. The mass window corresponding to a signal efficiency of $\varepsilon_W^G = 68\%$ is applied.

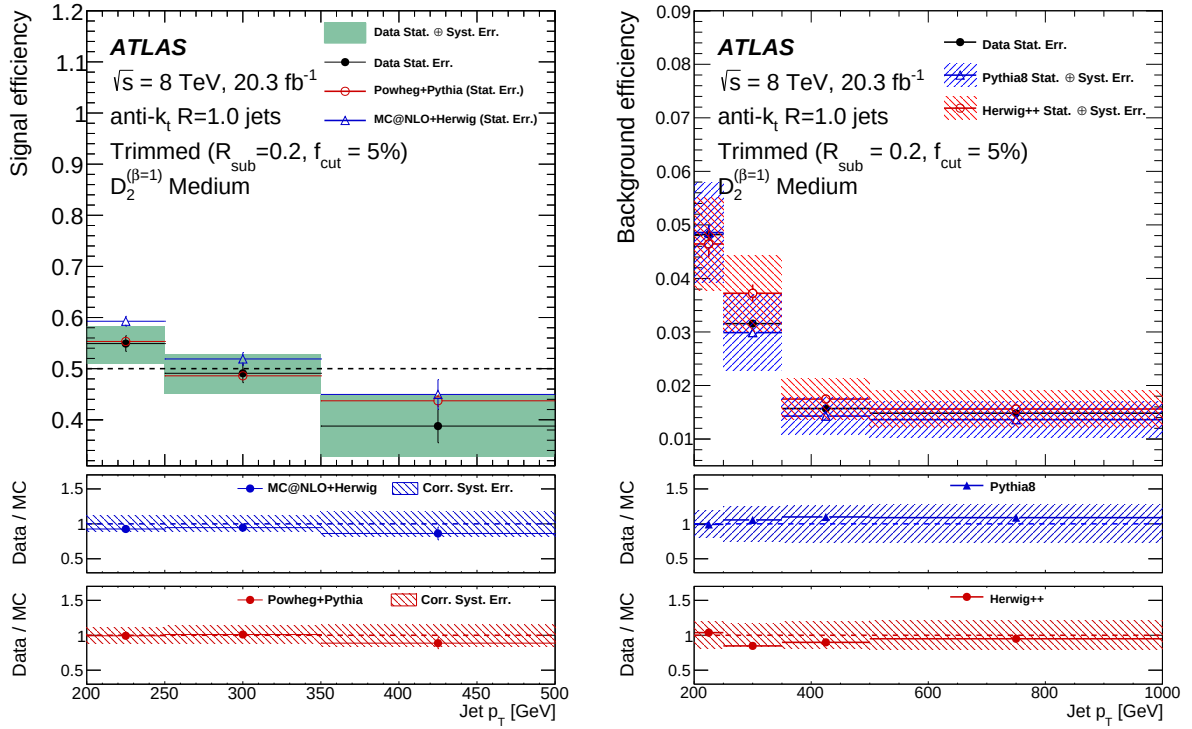


Figure 6.7: W boson tagging efficiency measurement in data and MC for signal W -jets in $t\bar{t}$ events (left) and multi-jet background events (right) in different p_T ranges for the 50% signal efficiency working point obtained using a mass window criteria and a requirement on the energy correlation variable $D_2^{\beta=1}$ [86].

simulated using PYTHIA8 and HERWIG++ whereas POWHEG [132] interfaced with PYTHIA and MC@NLO [133] interfaced with Herwig are used to generate top quark pair events. The measured and predicted signal efficiency for W -jets in $t\bar{t}$ events differs from the 50% defined signal efficiency working point for jets with $200 < p_T < 250$ and $350 < p_T < 500$ GeV. The W -jets used in the optimisation stage to define the working points are the decay products of a new heavy resonance W' and are exclusively longitudinal polarised. W bosons from the decay of a top quark can however have a longitudinal as well as transverse polarisation. The W boson polarisation is correlated to the substructure of the jet and thus a W -tagger should be optimised for individual searches depending on the initial particle that decayed into the W boson.

The background efficiencies in data and MC simulation are decreasing with increasing jet p_T . For $\varepsilon_W^{\text{G\&T}} = 50\%$, only approximately 1.5% of background jets with $p_T > 350$ GeV pass the selection criteria for jets. It must be emphasised that the background rejections obtained in Fig. 6.4 consider only the leading jet whereas the measurement in data in Fig. 6.7 is performed using the leading and subleading jet due to trigger restrictions. Therefore differences in the background rejection are expected because the subleading jet tends to have a larger jet mass and thus is more likely to be selected in the mass window around the W bosons mass. The difference in the jet mass distribution of leading and subleading jets is caused by their different quark and gluon composition. Gluon-initiated jets tend to have a larger mass than quark-initiated jets and tend to be subleading jets instead of leading jets.

The signal efficiency versus background rejection curves from $t\bar{t}$ and multi-jet events as predicted by MC simulation for $C_2^{\beta=1}$, $D_2^{\beta=1}$, τ_{21}^{wta} are summarised in Fig. 6.8. Furthermore, the data measurements for the medium and tight working point are included. The $D_2^{\beta=1}$ variable outperforms $C_2^{\beta=1}$ and τ_{21}^{wta} and gives the largest background rejection for a fixed signal efficiency. The background rejection for τ_{21}^{wta} and $C_2^{\beta=1}$ is relatively comparable. For the highest signal efficiencies, the τ_{21}^{wta} variable performs slightly better than $D_2^{\beta=1}$ for jets with $350 < p_T < 500$ GeV. Discrepancies between the MC prediction and the data measurement can be observed for all three variables in all p_T ranges.

6.1.3 Summary

Various combinations of jet grooming algorithms and substructure variables have been studied for the tagging of boosted W bosons. A selection of grooming algorithms were identified that result in a minimal pile-up dependence of the average groomed jet mass and provide a large background rejection in a mass window corresponding to a signal efficiency of $\varepsilon_W^{\text{G}} = 68\%$. The pairwise combination of the groomed jet mass and several substructure variables was then studied for each selected grooming algorithm to define $\varepsilon_W^{\text{G\&T}} = 50\%$ signal efficiency working points. The energy correlation variables $C_2^{\beta=1}$, $D_2^{\beta=1}$ and the N -subjettiness τ_{21}^{wta} were found to be particularly good variables to identify boosted W bosons and to suppress the multi-jet background. The performance of the best few configurations of the trimming, pruning and split-filtering algorithm are very similar for the defined working points and one algorithm,

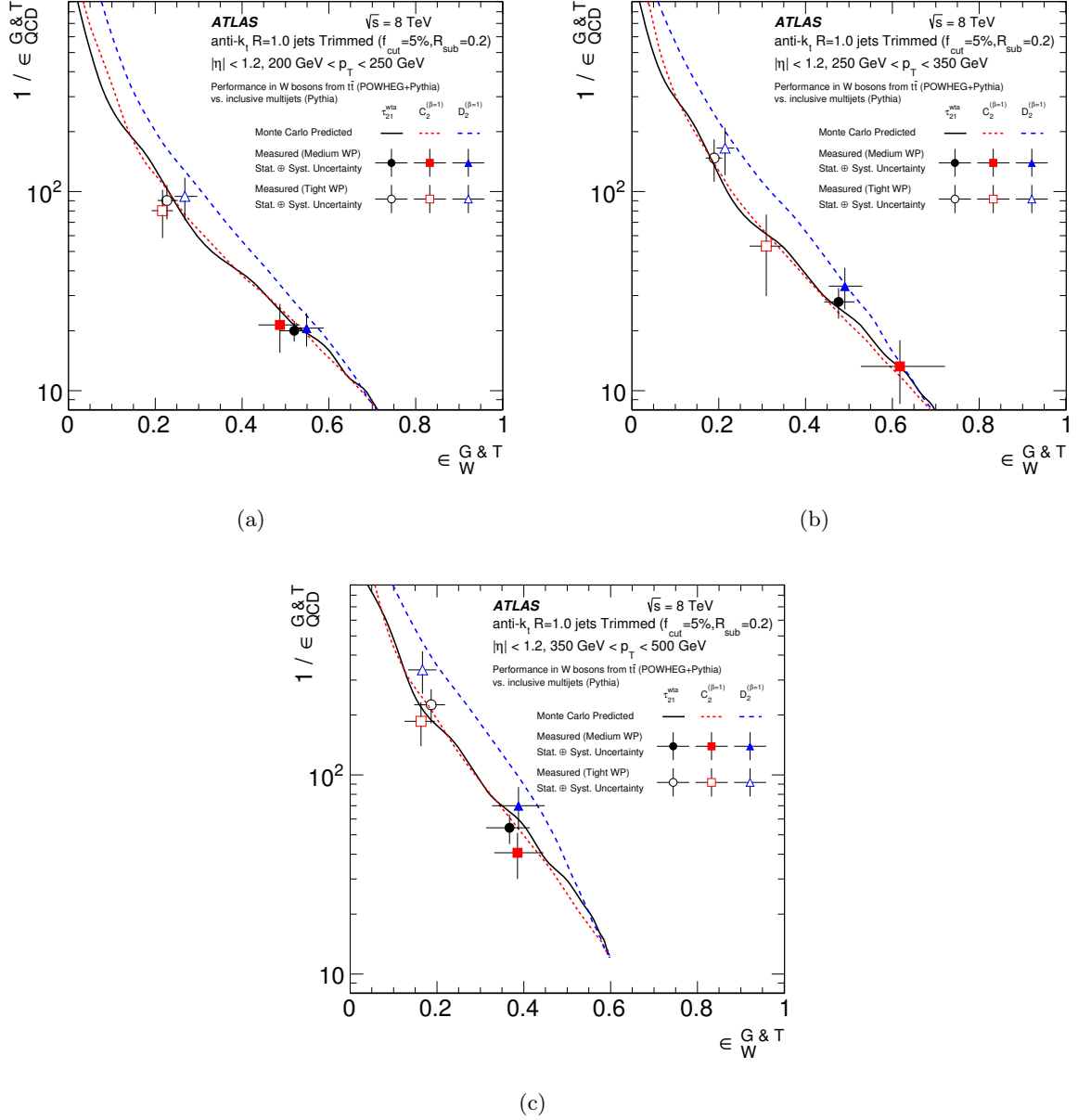


Figure 6.8: W boson tagging efficiency versus background rejection for jets with (a) $200 < p_T < 250$ GeV (b) $250 < p_T < 350$ GeV and (c) $350 < p_T < 500$ GeV for $C_2^{\beta=1}$, $D_2^{\beta=1}$ and τ_{21}^{wta} after applying the 68% signal efficiency mass window [86]. The curves are obtained using MC simulation and compared to the 50% and 25% signal efficiency working point measurements in data using a lepton-plus-jet enriched $t\bar{t}$ sample and multi-jet events.

anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$, was chosen as benchmark for further measurements in 8 TeV collision data. Discrepancies between the measured signal and background efficiency in data and MC simulation can be observed for all three substructure variables.

6.2 Performance of Boosted Boson Identification at $\sqrt{s} = 13$ TeV

Based on the extensive comparison of grooming and substructure techniques at $\sqrt{s} = 8$ TeV, presented in Section 6.1, the performance of the best few configurations and substructure variables to identify boosted vector bosons was re-evaluated in MC simulation at $\sqrt{s} = 13$ TeV. The best-performing algorithm, in terms of pile-up removal and background rejection was then determined, for jets over a broad p_T range ($200 < p_T < 2000$ GeV). Two working points corresponding to a signal efficiency of 50% (medium) and 25% (tight) were defined for the most promising algorithm. As opposed to the 8 TeV studies, both the identification of W and Z bosons are optimised in this section. Therefore, signal samples, simulating the process $W' \rightarrow WZ \rightarrow qqqq$, are generated with PYTHIA8 using the NNPDF2.3LO PDF set and the AU2 underlying event and parton shower tune. To allow for the extension of the jet p_T range, signal samples are produced with W' masses ranging from 400 to 5000 GeV. The QCD dijet background is generated with exactly the same setup as the signal samples.

Optimisation

The optimisation of the boosted vector boson identification at $\sqrt{s} = 13$ TeV closely follows the strategy in Section 6.1.1. Events are categorised based on the leading ungroomed C/A $R = 1.2$ truth jet with $|\eta^{\text{Truth}}| < 2.0$ in five different p_T^{Truth} ranges: $[200, 350)$, $[350, 500)$, $[500, 1000)$, $[1000, 1500)$ and $[1500, 2000)$ GeV. In the studies presented here, four grooming configurations and three substructure variables ($C_2^{\beta=1}$, $D_2^{\beta=1}$, τ_{21}^{wta}), that were deemed promising based on the 8 TeV studies, are selected:

- anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$, $R_{\text{sub}} = 0.2$,
- C/A $R = 1.0$ pruned jets with $Z_{\text{cut}} = 0.15$, $R_{\text{cut}} = 0.5$,
- C/A $R = 1.2$ split-filtered jets with $y_{\text{cut}} = 0.15$, $R_{\text{sub}} = 0.3$,
- C/A $R = 1.2$ split-filtered jets with $y_{\text{cut}} = 0.04$, $R_{\text{sub}} = 0.3$,

where the last jet collection corresponds to the one widely used in diboson resonance searches at $\sqrt{s} = 8$ TeV. For each grooming configuration, requirements on the groomed jet mass are imposed, yielding a signal efficiency of $\varepsilon_{W,Z}^G = 68\%$. The pairwise combination of the mass windows with each of the three considered substructure variables is then studied to define a medium ($\varepsilon_{W,Z}^{G\&T} = 50\%$) and tight ($\varepsilon_{W,Z}^{G\&T} = 25\%$) signal efficiency working point. The selection criteria to obtain these signal efficiency working points are then applied to the QCD dijet background

to determine the background efficiencies $\varepsilon_{\text{QCD}}^{\text{G\&T}}$ and to compare them. The combination of the grooming algorithm and the substructure variable yielding the smallest background rejection over a broad range of p_{T} is then further optimised.

The leading jet mass distributions for the four grooming configurations are depicted in Fig. 6.9 for the ranges $200 < p_{\text{T}}^{\text{Truth}} < 350$ GeV and $1500 < p_{\text{T}}^{\text{Truth}} < 2000$ GeV. No jet energy or mass calibrations are applied at this stage. Even though the distributions are truncated at 200 GeV, the background mass distributions exhibit high mass tails beyond 1 TeV for jets with $1500 < p_{\text{T}}^{\text{Truth}} < 2000$ GeV. By studying the average jet mass dependence on the number of reconstructed vertices, it can be shown that these high-mass tails are not caused by the contamination of energy deposits from pile-up vertices but are instead from initial-state radiation captured in the large- R jet. On the other hand, the jet mass distribution for the pruned and split-filtered collection with $y_{\text{cut}} = 0.04$ still exhibits a strong pile-up dependence in the lowest $p_{\text{T}}^{\text{Truth}}$ range considered. For both the trimmed and pruned jet collection, the shape of the background mass distribution strongly depends on the jet p_{T} . For the split-filtered jet collections, the background mass distributions are more stable with p_{T} , but the grooming configuration with $y_{\text{cut}} = 0.04$ has a maximum close to the signal mass distribution which can be removed by increasing the y_{cut} criteria.

The smallest mass window that contains 68% of the signal events is then calculated for W and Z bosons for the different grooming configurations. Signal efficiency versus background rejection curves are determined for the pairwise combinations of the groomed mass window with the three considered substructure variables for each of the grooming algorithms. To compare their performance, two figures of merit are defined: a 50% and 25% signal efficiency working point. The background rejection factors for the medium signal efficiency working point are shown in Fig. 6.10 and Fig. 6.11 separately for W and Z boson tagging with $200 < p_{\text{T}}^{\text{Truth}} < 350$ GeV and $1500 < p_{\text{T}}^{\text{Truth}} < 2000$ GeV, respectively. Furthermore, the background rejection factors are indicated for a narrower mass window size corresponding to $\varepsilon_{W,Z}^{\text{G}} = 50\%$. The rejection factors are however significantly smaller than for the combination of selection criteria on the jet mass and an additional substructure variable. In the case of the narrower mass window, the better separation of the Z -jet mass distribution from the background compared to W -jets, results in larger background rejection factors for Z -jets than W -jets. The same behaviour can be observed for the pairwise combination of the jet mass and a substructure variable for high- $p_{\text{T}}^{\text{Truth}}$ jets in Fig. 6.11. Conversely, for low- $p_{\text{T}}^{\text{Truth}}$ jets in Fig. 6.10, the pairwise combinations result in higher background rejection power for W -jets than for Z -jets. As it can be seen in Fig. 6.9, the mass distribution of Z -jets is broader than that of W -jets at low $p_{\text{T}}^{\text{Truth}}$. Therefore, the mass window containing 68% of the Z boson signal has to be wider than that for W bosons and thus resulting in a smaller background rejection power compared to W -jets. At high $p_{\text{T}}^{\text{Truth}}$, the width of the W -jet and Z -jet mass distributions are about the same. The substructure selection criteria are expected to have approximately the same effect on W -jets and Z -jets.

Comparing the background rejection power corresponding to the medium and tight signal efficiency working point for all considered $p_{\text{T}}^{\text{Truth}}$ ranges, both the trimmed jet collection in combi-

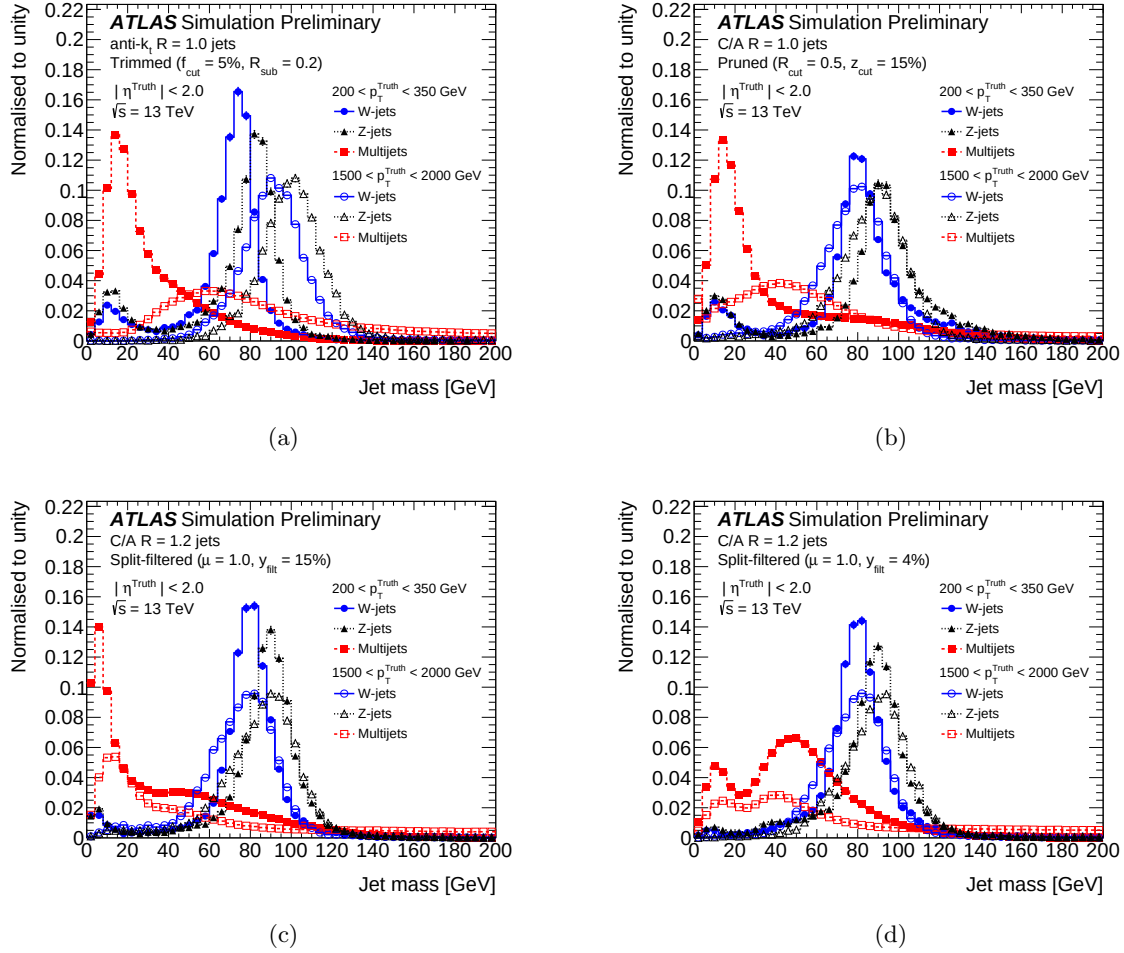


Figure 6.9: Leading uncalibrated jet mass distribution for W and Z signal samples and QCD dijet background in two different p_T ranges: $200 < p_T^{\text{Truth}} < 350$ GeV and $1500 < p_T^{\text{Truth}} < 2000$ GeV [108]. The distributions are shown for (a) anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$, (b) C/A $R = 1.0$ pruned jets with $Z_{\text{cut}} = 0.15$, $R_{\text{cut}} = 0.5$, (c) C/A $R = 1.2$ split-filtered jet with $y_{\text{cut}} = 0.15$, $R_{\text{sub}} = 0.3$ and (d) C/A $R = 1.2$ split-filtered jet with $y_{\text{cut}} = 0.04$, $R_{\text{sub}} = 0.3$.

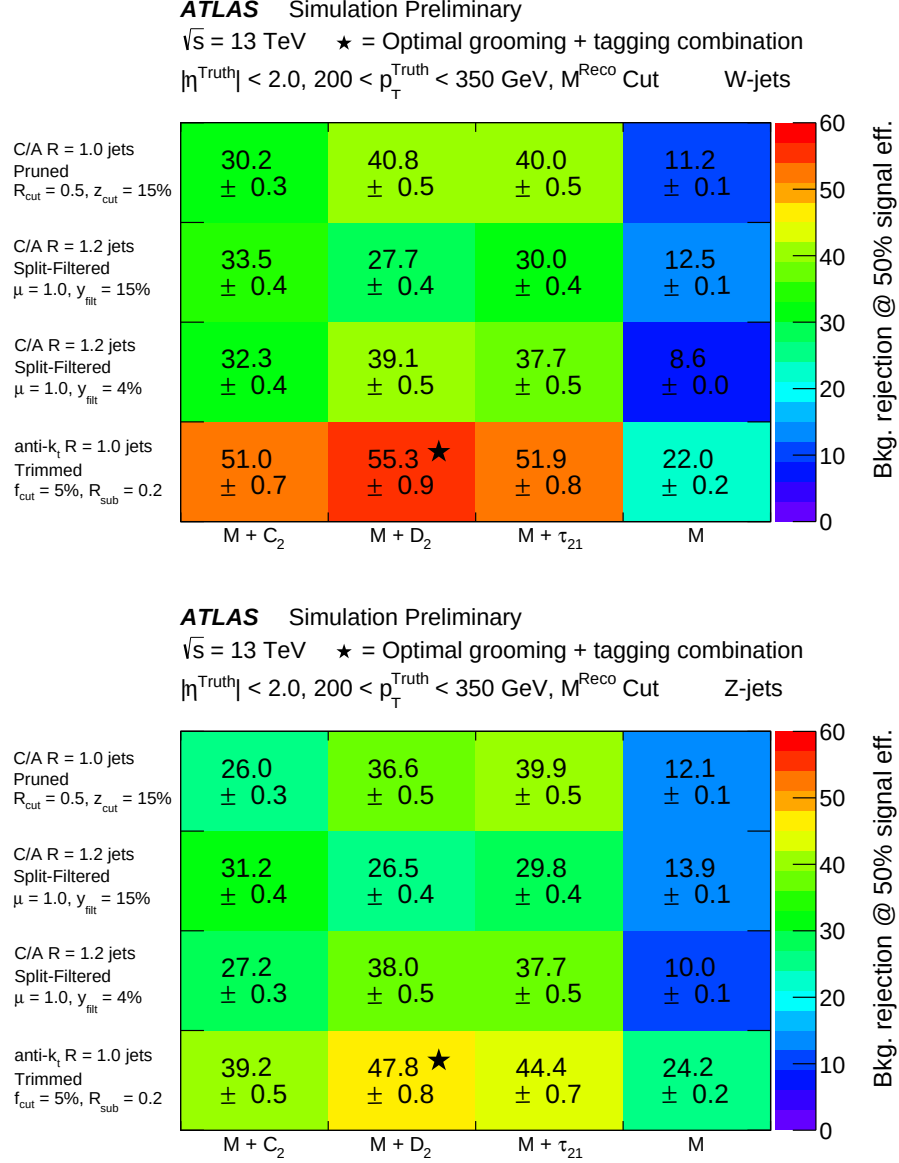


Figure 6.10: Background rejection corresponding to a W (top) and Z (bottom) boson signal efficiency of $\varepsilon_{W,Z}^{\text{G\&T}} = 50\%$ for different combinations of substructure variables and grooming algorithms for jets with $200 < p_T^{\text{Truth}} < 350$ GeV. The uncertainties are statistical errors only.

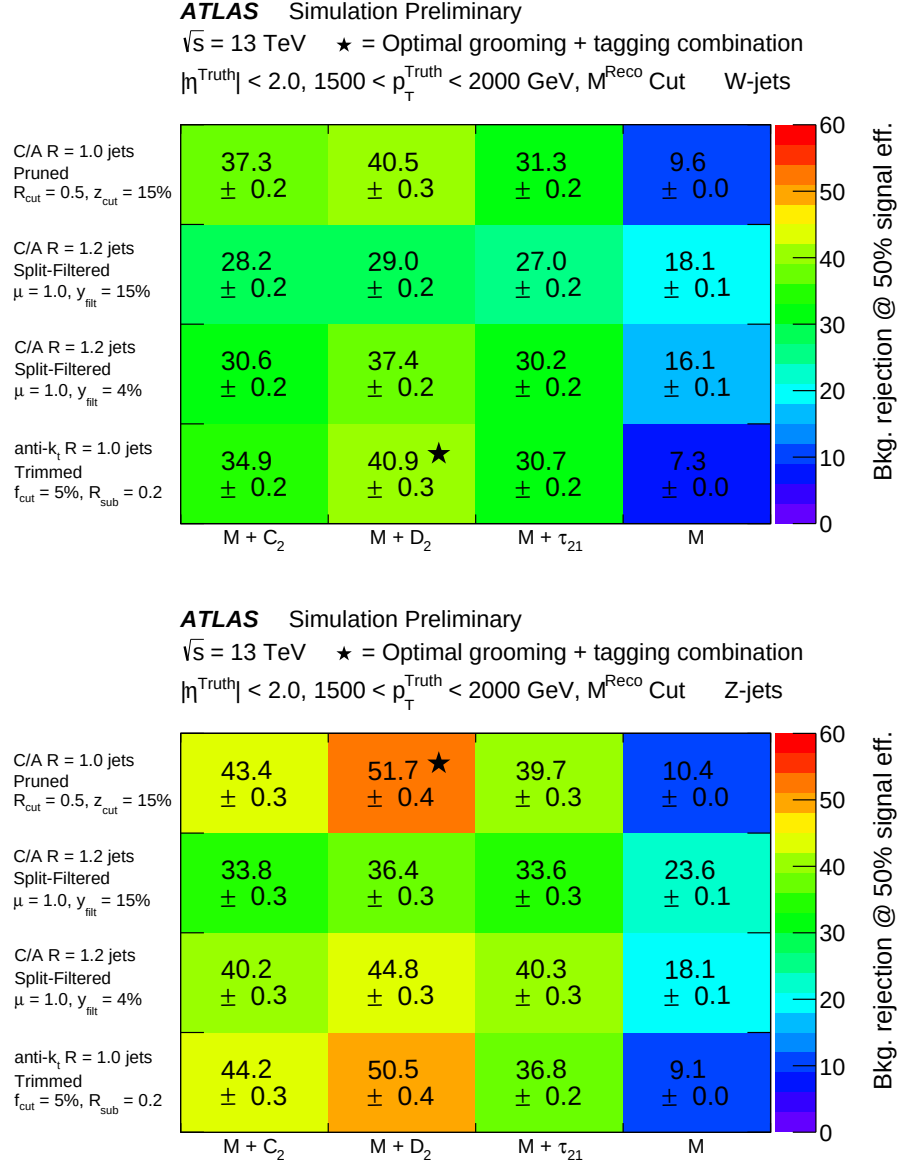


Figure 6.11: Background rejection corresponding to a W (top) and Z (bottom) boson signal efficiency of $\varepsilon_{W,Z}^{\text{G\&T}} = 50\%$ for different combinations of substructure variables and grooming algorithms for jets with $1500 < p_T^{\text{Truth}} < 2000$ GeV. The uncertainties are statistical errors only.

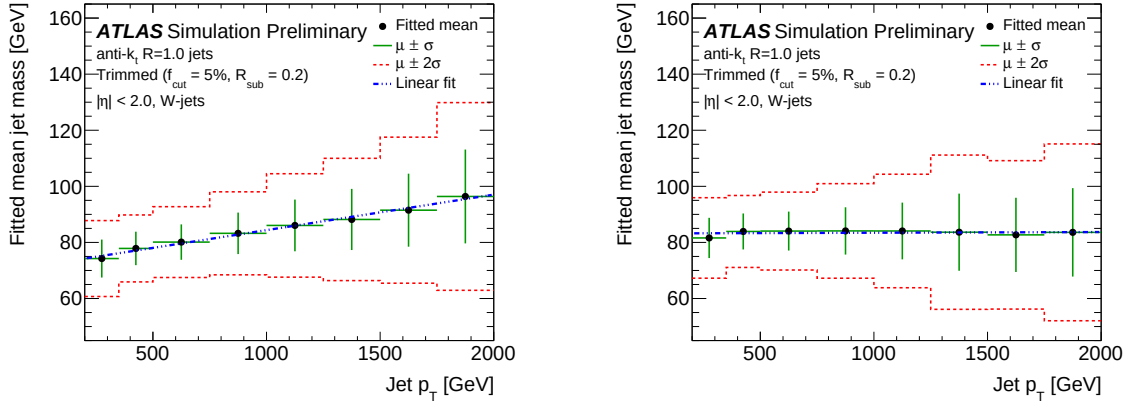


Figure 6.12: Average W -jet signal mass of anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$ as a function of the jet p_T before (left) and after (right) the mass calibration [108].

nation with $D_2^{\beta=1}$ and the pruned jet collection in combination with $C_2^{\beta=1}$ seem to perform well in terms of W - and Z boson tagging. Taking into account that the pruned jet mass exhibit a non-negligible pile-up dependence at low- p_T , anti- k_t $R = 1.0$ trimmed jets with $f_{\text{cut}} = 5\%$, $R_{\text{sub}} = 0.2$ were chosen as baseline grooming algorithm for boosted vector boson identification in early Run-II.

Working Point Derivation

The baseline vector boson identification algorithm is then further optimised to be used in physics analyses. Dedicated jet energy and mass calibrations are derived, using the procedure described in Section 5.1. To demonstrate the effect of the calibration on the jet mass, the W -jet mass distribution is fitted with a Gaussian distribution and the mean value is plotted as a function of the jet p_T in Fig. 6.12 before and after the mass calibration was applied. The striking p_T -dependence of the jet mass is removed by the jet mass calibration. Furthermore, the width of the mass distribution is indicated by the 1σ and 2σ ranges. The optimisation procedure that was previously used to identify the best-performing algorithm requires the definition of signal mass windows that contain 68% of the signal. As the width of the mass distribution increases with p_T , also the width of the mass window would increase. For analyses that rely on a smoothly falling distribution of the invariant dijet mass spectrum, the changing mass window sizes in the different p_T ranges could distort the spectrum and result in difficulties to parameterise the background. Therefore, instead of defining the mass window size based on a fixed signal efficiency, a ± 15 GeV mass window around the mean of the W/Z -jet mass distribution is chosen across the p_T range, resulting in $\varepsilon_{W,Z}^G = 55 - 80\%$. The 15 GeV mass window size provides a compromise between the good mass resolution at low p_T and its degradation at high p_T .

The medium and tight working point are then derived by imposing selection criteria on the $D_2^{\beta=1}$ variable after requiring jets to fall within the 15 GeV mass window. The $D_2^{\beta=1}$ selection

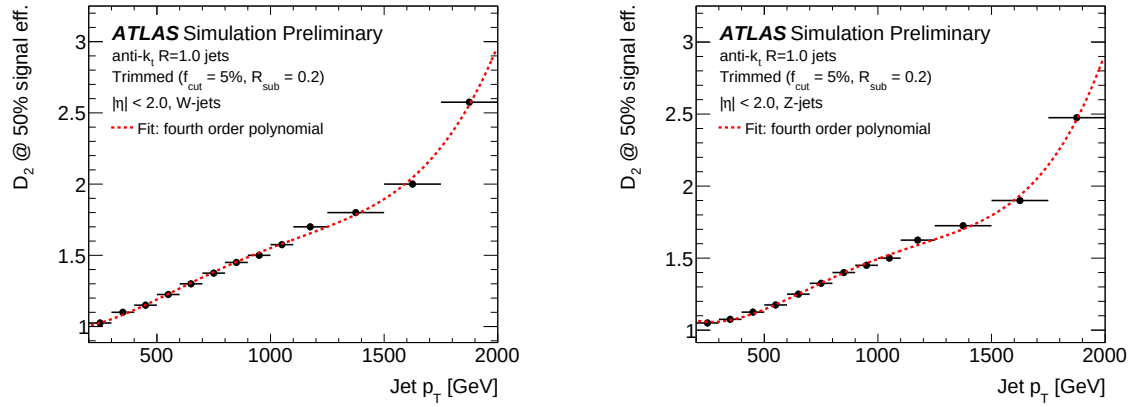


Figure 6.13: Requirement on the $D_2^{\beta=1}$ variable as a function of the calibrated jet p_T for the medium working point corresponding to a W (left) and Z (right) signal efficiency of 50% [108].

requirements are one-sided, with the lower boundary of zero and the upper boundary as shown in Fig. 6.13 for the medium working point as a function of the jet p_T for W -jets and Z -jets. The maximum $D_2^{\beta=1}$ selection criteria increases approximately linear with the jet p_T . A change in the slope can be observed for jets with $p_T > 1750$ GeV where the jet mass resolution degrades significantly. Only a slightly larger fraction of signal events, compared to the medium working point, are thus selected and result in the loose criteria on $D_2^{\beta=1}$. To avoid bin-edge effects that may result from the use of discrete selection criteria, the p_T dependence of the maximum $D_2^{\beta=1}$ selection criteria is fitted with a fourth-order (second-order) polynomial for the medium (tight) signal efficiency working point.

The resulting signal efficiencies and background rejection factors for W -tagged jets are shown in Fig. 6.14. The efficiencies for Z -tagged jets are not shown here as they are almost identical to those of W -tagged jets. The uncertainty bands include uncertainties on the jet p_T , mass and $D_2^{\beta=1}$ scale, derived in Section 5.2.3 as well the corresponding resolution uncertainties. The scale uncertainties are treated as fully correlated whereas the resolution uncertainties are treated as uncorrelated. The large uncertainties on the background rejection for the tight working point are dominated by the $D_2^{\beta=1}$ scale uncertainties. The derived upper boundaries on the $D_2^{\beta=1}$ variable, corresponding to the tight working point, are close to the maximum of the $D_2^{\beta=1}$ distribution in the background sample. Therefore, even small variations of the $D_2^{\beta=1}$ value have a large impact on the background efficiency.

6.2.1 Summary

In this section, an algorithm was developed for Run-II to identify boosted hadronically decaying vector bosons based on the optimisation studies performed in Run-I.

To mitigate the influence of pile-up effects on anti- k_t jets with a radius parameter of $R = 1.0$, jets are trimmed with $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 5\%$. Furthermore, selection criteria are imposed

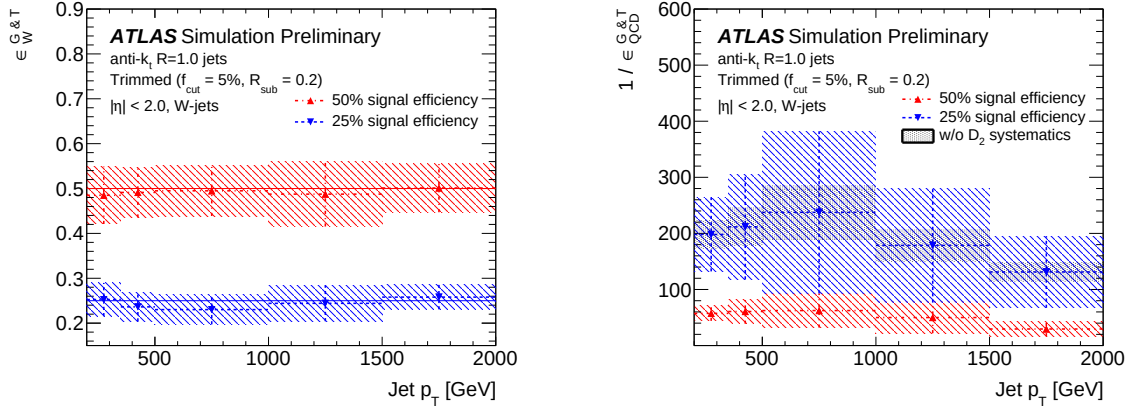


Figure 6.14: Signal efficiency for W -jets (left) and background rejection in QCD dijet samples (right) for two different working points with $\epsilon_{\text{sig}}^{G\&T} = 50\%$ and 25% [108].

on the jet mass and the $D_2^{\beta=1}$ variable to discriminate jets containing the decay products of a W or Z boson from a quark or gluon-initiated jet. The selection criteria are p_T dependent to define two working points of constant signal efficiency, corresponding to 50% and 25%, for jets with a p_T between 200 and 2000 GeV. For the jet mass, a 15 GeV wide window around the mean of the W/Z -jet mass distribution was chosen whereas the p_T dependence of the $D_2^{\beta=1}$ criteria was parameterised with a functional form. The fixed mass window and the parameterisation of the $D_2^{\beta=1}$ criteria was motivated by the fully hadronic diboson resonance search, see Section 7, that relies on the smoothly falling distribution of the invariant dijet mass spectrum. Varying selection criteria in different p_T ranges could distort the spectrum and would complicate the data-driven background parameterisation.

For jets with $p_T < 1500$ GeV, a background rejection of about 60 can be achieved for the 50% signal efficiency working point. The background rejection decreases to 25 for jets with a very high transverse momentum, $p_T > 1500$ GeV, due to the diminished resolution of the $D_2^{\beta=1}$ variable. Possible improvements to achieve an approximately constant background rejection over the full p_T range are discussed in the following section.

6.3 Future Improvements to Boosted Vector Boson Tagging

In this section, the drawbacks of the algorithm developed in Section 6.2 to identify boosted vector bosons at $\sqrt{s} = 13$ TeV as well as possible future improvements, are discussed.

In the next years of data-taking at the LHC, vector bosons will be frequently produced with $p_T > 1500$ GeV. Thus a detailed understanding of the substructure techniques at high transverse momenta will be of utmost importance and techniques are needed that overcome the limitations due to the granularity of the hadronic calorimeter. It was shown already in Section 5.3.2 that

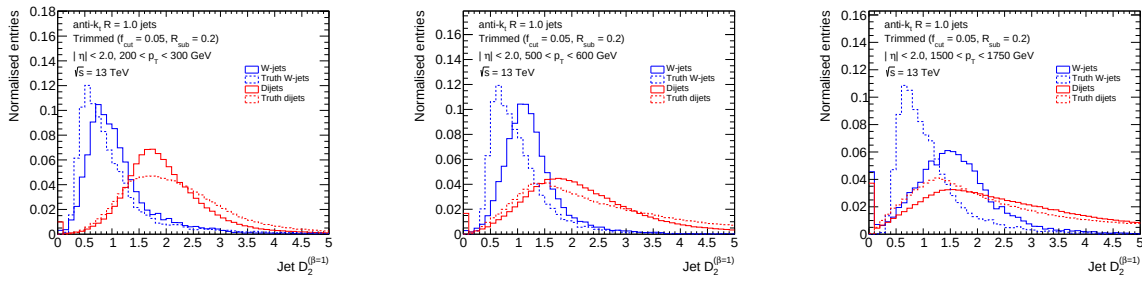


Figure 6.15: Jet D_2 distribution for calorimeter and truth-level W -jets and jets in the QCD dijet background sample with $200 < p_T < 300$ GeV (left), $500 < p_T < 600$ GeV (middle) and $1500 < p_T < 1750$ GeV (right).

the mass resolution diminishes significantly at high p_T and the same conclusions can be drawn as well for substructure variables as illustrated for $D_2^{\beta=1}$ in Fig. 6.15. The differences between the $D_2^{\beta=1}$ distribution of calorimeter jets and their matched truth-particle level jets increase with p_T , especially for W -jets. At high p_T , the discrimination of the $D_2^{\beta=1}$ variable between signal and background jets is significantly reduced compared to the truth-level $D_2^{\beta=1}$ distribution due to the shift of the W -jet distribution and its broadening. This diminished resolution of the $D_2^{\beta=1}$ variable is responsible for the decrease in background rejection at high p_T in Fig. 6.14. The comparison of the inputs for the $D_2^{\beta=1}$ calculation, the energy correlation functions defined in Eq. 4.9b-4.9d, are shown in Fig. 6.16 for low- p_T ($500 < p_T < 600$ GeV) and high- p_T ($1500 < p_T < 1750$ GeV) truth-level and calorimeter jets. The energy correlation functions ECF2 and ECF3 are in general larger for calorimeter jets than for truth-particle level jets at high p_T because of the limitations of topo-clustering algorithm due to the ATLAS calorimeter granularity. If the energy deposits of the decay products of a highly boosted W boson are collected within one single topo-cluster, the algorithm will split the topo-cluster between the two local energy maxima. Hence, the separation (ΔR) between the two dense cores of energy from the W boson decay is enforced to be larger than the calorimeter granularity, resulting also in larger energy correlation values. On the other hand, the separation between two particles at truth level can have any value and is not limited by the granularity of the calorimeter. For QCD-jets, the difference between calorimeter and truth-particle level jets is less pronounced because the calculation of the energy correlation functions is dominated by the p_T of the initial quark or gluon and soft wide-angle radiation don't have a large impact on the calculation. As for the jet mass resolution, the resolution of the $D_2^{\beta=1}$ variable can be improved by exploiting the good angular resolution of charged-particle tracks from the Inner Detector at high p_T . Either track-assisted substructure variables, similar to the definition of the track-assisted jet mass in Section 5.3.2, could be used or variables could be calculated from particle flow inputs.

The improved mass resolution, achieved for example by combining the calorimeter jet mass and track-assisted jet mass, will also result in a better discrimination of W bosons from Z bosons.

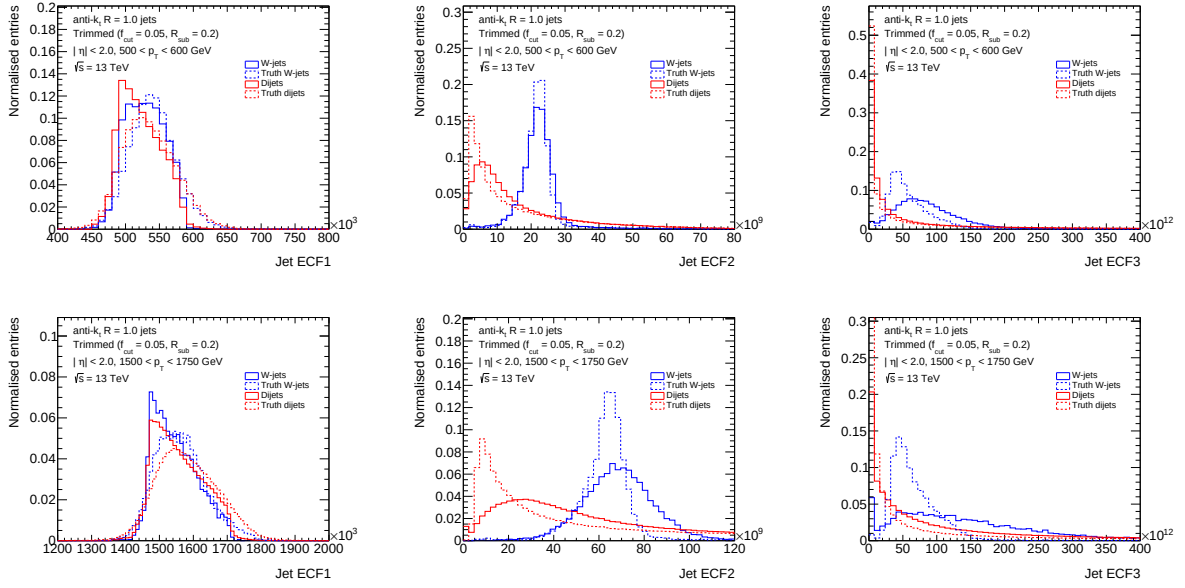


Figure 6.16: Distribution of the energy-correlation functions ECF1 (left), ECF2 (middle) and ECF3 (right) for calorimeter and truth-level W -jets and jets in the QCD dijet background sample with $500 < p_T < 600$ GeV (top) and $1500 < p_T < 1750$ GeV (bottom).

Even though the identification of W bosons and Z bosons have been optimised separately, the 30 GeV broad mass window criteria results in a large overlap between the vector bosons. Thus many W -tagged jets also pass the selection criteria of the Z boson tagger and vice-versa. A smaller mass window criteria would thus also reduce the overlap between W -tagged and Z -tagged jets. To further enhance the differentiation of Z bosons from W bosons in Run-II, a likelihood tagger can be constructed from jet-related variables as it was done in Run-I [134].

The performance of boosted vector boson identification can be improved in the future by using more advanced techniques such as *variable jet size* (variable- R) [135, 136] algorithms, *shower deconstruction* [137, 138] or *multivariate techniques*. Various substructure variables are combined in multivariate techniques to exploit their correlation and to gain sensitivity. However, the gain has to be weighted against the systematic uncertainties associated to each of the variables.

Apart from the grooming techniques, introduced in Section 4.4, variable- R jet algorithms were found to be beneficial in terms of pile-up removal. According to the rule of thumb, the angular separation of the bosons decay products decreases with increasing p_T and thus a jet algorithm whose radius parameter shrinks as a function of the jet p_T reduces the area susceptible to soft particle contamination. Nevertheless, grooming techniques are still needed to remove the remaining pile-up contribution from the jet, although their effect on a variable- R jet is much smaller. It was shown in Ref. [136] that variable- R jets have a better background rejection with respect to fixed-size trimming algorithms for jets with $p_T > 1000$ GeV using a combination of

the jet mass and $D_2^{\beta=1}$ variable.

A very promising algorithm to enhance the background rejection is Shower Deconstruction (SD) which assigns to each event a probability to arise either from the signal or background process by exploiting the substructure information of the hadronically-decaying jets. For top-tagging, SD outperforms other tagging techniques for low signal efficiencies. These low signal efficiencies correspond to high background rejection factors and are of utmost importance for searches like the diboson resonance search to suppress the QCD background as much as possible.

Chapter 7

Common Aspects for Diboson Resonance Searches

The existence of resonances with couplings to vector boson pairs is predicted in many extensions to the SM such as technicolour, warped extra dimensions, and Grand Unified Theories. As described in Section 2.2, the analysis presented in this thesis focuses on the search for diboson resonances (WW , WZ and ZZ) in the fully hadronic final state due to the large branching ratio of vector bosons to quark pairs. However, the analysis suffers from the enormous background of SM QCD dijet production. Further contributions from SM processes, all of which are much smaller in magnitude at high p_T , are W/Z +jets, diboson, top-quark pair ($t\bar{t}$) and single-top production. The background can be significantly suppressed by reconstructing the decay products of each vector boson as one single large- R jet and then applying the grooming and tagging techniques introduced in Section 4 and 6. The signature of the decay of a heavy particle is thus a narrow resonant structure on top of the smoothly falling invariant dijet mass spectrum of the background processes. The background in the analysis is not estimated from MC simulation but rather from the data using a parametric function to avoid statistical limitations of the background MC samples especially in the high mass region. The fully hadronic diboson resonance search was performed both at $\sqrt{s} = 8$ TeV and 13 TeV and the common aspects of the two analyses are summarised in this chapter. The results of the two searches are presented in Section 8 and Section 9.

This chapter is organised as follows. The simulation of the signal and background processes is presented in Section 7.1. The 2012 and 2015 data samples are described in Section 7.2. The objects used in the diboson resonance search are introduced in Section 7.3 with an emphasis on the differences in the jet reconstruction and boosted vector boson identification between the 8 and 13 TeV analysis. The event selection is defined in Section 7.4. The parameterisation of the background is described in Section 7.5, followed by the presentation of the systematic uncertainties in Section 7.6. The statistical techniques, used to interpret the results of the searches, are discussed in Section 7.7.

7.1 Simulation of Signal and Background Samples

Three specific benchmark models, introduced in Section 2.2, are used in the diboson resonance searches at $\sqrt{s} = 8$ and 13 TeV to assess the sensitivity of the analyses and to optimise the event selection. Both searches used a bulk Randall–Sundrum graviton model, predicting a spin-2 graviton G_{RS} decaying into WW or ZZ . In addition, an extended gauge model $W' \rightarrow WZ$ was used in the Run-I analysis, whereas a Heavy Vector Triplet Model with $W' \rightarrow WZ$ or $Z' \rightarrow WW$ was studied in Run-II. The HVT model was chosen for the Run-II analysis over the extended gauge model as it allows to describe a large class of models (see Section 2.2.2) and does not exhibit loss mass tails in the invariant dijet mass distribution due to off-shell production. The event selection is optimised with respect to the dominant background: dijet events produced in QCD interactions. The QCD dijet sample is further used to characterise the expected invariant dijet mass spectrum.

$W' \rightarrow WZ \rightarrow qq\bar{q}\bar{q}$ signal events, predicted in an extended gauge model (EGM) are generated with PYTHIA8 using the MSTW2008 PDFs and the AU2 underlying event tune. The signal samples are generated with different W' masses, covering the range $1.3 < m_{W'} < 3.0$ TeV in 100 GeV intervals.

The MADGRAPH generator with the NNPDF2.3LO PDF set and the A14 tune, interfaced to PYTHIA8 for the simulation of the parton shower and hadronisation, is used to simulate the production of $Z' \rightarrow WW$ and $W' \rightarrow WZ$ as predicted in the HVT model A. W' and Z' masses are generated between 1.2 and 2.0 TeV in 100 GeV intervals and between 2.0 and 3.0 TeV in 200 GeV spacings. Only the hadronic decay of the vector bosons is simulated.

The Randall–Sundrum graviton signal samples are generated in Run-I with CALCHEP [139] using the CTEQ6L1 parton distribution functions and PYTHIA8 for the simulation of the hadronisation and the parton shower. For Run-II, MADGRAPH interfaced to PYTHIA8 with the NNPDF2.3LO PDF set and the A14 tune is used. The RS model is characterised by $k/\bar{M}_{Pl} = 1.0$ and the graviton can either decay into WW or ZZ whereupon only the hadronic vector boson decay is simulated. Samples with graviton mass in the range $1.2(1.3) < G_{RS} < 3.0$ TeV, using the same intervals as for the HVT (EGM W') samples, are generated for Run-II (Run-I).

The QCD dijet background events are produced with PYTHIA8 and the CT10 parton distribution functions in Run-I whereas the NNPDF2.3LO PDFs and the A14 tunes are used for the 2015 analysis.

The expected contribution from additional pp interactions is accounted for by overlaying minimum-bias events on the events of interest. These additional events are generated with PYTHIA8 and the A2 tune. For the 8 TeV analysis, the distribution of the average number of pp interactions per bunch crossing in MC simulation was adjusted to the distribution in data, shown in Fig. 3.2, with an average of 20.7. For Run-II however, the $\langle\mu\rangle$ distribution was estimated prior to data-

taking and slightly overestimated the amount of pile-up in collision data. No reweighting of the $\langle\mu\rangle$ distribution in MC simulation to that in data was performed in this analysis as it is not sensitive to the $\langle\mu\rangle$ profile and the reweighting would have caused a loss in statistics.

7.2 Data Sample

The searches presented in this thesis use pp collision data collected in 2012 and 2015 with a centre-of-mass energy of $\sqrt{s} = 8$ TeV and 13 TeV, respectively. Events are selected with the lowest un-prescaled single jet trigger for which the leading- p_T ungroomed anti- k_t $R = 1.0$ jet has a minimum E_T of 360 GeV. After requiring good beam and detector conditions, the integrated luminosity of the 8 TeV and 13 TeV dataset correspond to 20.3 fb^{-1} and 3.2 fb^{-1} with a relative uncertainty of 2.8% and 5%, respectively. The uncertainties on the integrated luminosities are determined from beam-separation scans [140].

7.3 Object Definitions

The fully hadronic diboson resonance search relies on the reconstruction of the vector bosons as two large- R jets. Events containing either charged leptons or a large amount of missing transverse energy are vetoed. To allow for a later combination with other diboson resonance searches, sensitive in the high mass range, decaying into $\ell\nu qq$ [89, 141], $\ell\ell qq$ [88, 142] or $\nu\nu qq$ [143], the same object definitions are used and selection criteria are imposed to obtain an orthogonal data sample with respect to the other channels.

7.3.1 Jet Reconstruction and Boson Identification

Two different approaches have been pursued in the reconstruction of jets and boson identification for the $\sqrt{s} = 8$ TeV and 13 TeV diboson resonance searches. After the 8 TeV analysis was published, the boson tagging algorithm was optimised for $\sqrt{s} = 13$ TeV to account for the higher boost and changed pile-up conditions as described in Section 6.2. In both analyses, large- R jets are reconstructed from topo-clusters and boosted vector bosons are identified by imposing criteria on the groomed jet mass and the number of tracks, ghost-associated to the ungroomed jet, as well as one further substructure variable, $\sqrt{y_{12}}$ or $D_2^{\beta=1}$, depending on the jet collection.

In addition to the large- R jet collections, which are needed to reconstruct the hadronic decay of boosted vector bosons, jets are also built from topo-clusters with the anti- k_t algorithm and a distance parameter of $R = 0.4$. These small-radius jets are only used to reject events that contain fake jets from noise in the calorimeter or non-collision backgrounds.

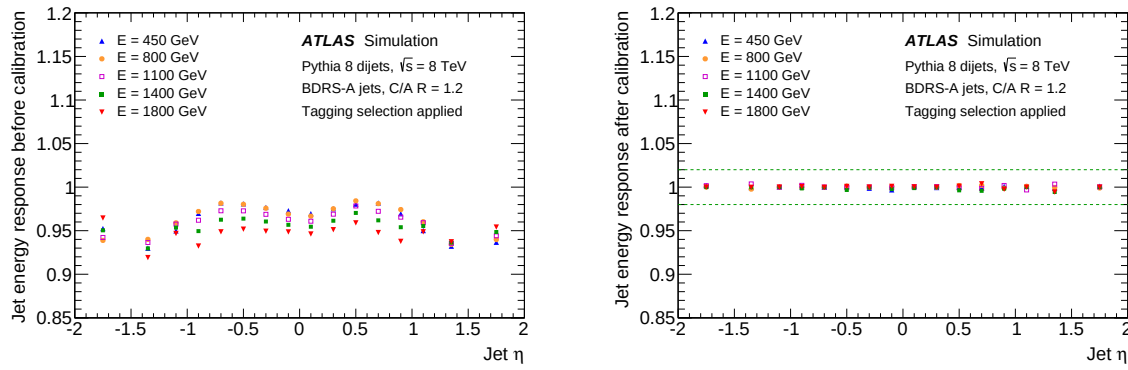


Figure 7.1: Jet energy response as a function of the pseudorapidity for C/A $R = 1.2$ BDRS-A jets in a PYTHIA8 dijet sample for different jet energies before (left) and after (right) the calibration to the particle-level. The selection as described in Section 7.4 is applied.

$\sqrt{s} = 8$ TeV Analysis

Jets are reconstructed with the C/A algorithm using a radius parameter of $R = 1.2$. To remove contamination from pile-up and the underlying event, the BDRS algorithm as described in Section 4.4.2 is used with the parameters $\sqrt{y_{\text{cut}}} = 0.2$ and $\mu_{\text{max}} = 1.0$. In the filtering stage, the jet constituents at the last step of de-clustering procedure are reclustered using a fixed subjet radius size of $R_{\text{sub}} = 0.3$ and only the three highest- p_T subjets are kept.

The energy and mass of the BDRS-jets are then calibrated to the particle-level with the numerical inversion technique. The jet energy and mass response before and after the calibration is applied are shown in Fig. 7.1 and Fig. 7.2, respectively.

To identify jets that are compatible with the hadronic decay of boosted vector bosons, only jets within a 26 GeV wide mass window, centred around the peak of the W -jet and Z -jet mass distribution in MC simulation, are selected. The peaks of the W -jet and Z -jet mass distributions correspond to 82.4 GeV and 91.8 GeV, respectively. The mass distribution for signal and background events is depicted in Fig. 7.3. The applied mass window requirement reduces the QCD dijet background by approximately a factor of 10.

The background jets that fulfil the requirements of the BDRS algorithm are mostly composed of one hard core of energy from the initial quark or gluon and a soft gluon emission. Consequently, the subjets within a QCD-jet are less balanced in momentum compared to a W -jet or Z -jet as shown in Fig. 7.3. To further suppress the selection of QCD-jets, a more stringent momentum balance criterion is applied: $\sqrt{y_{12}} > 0.45$.

The third variable that is used to improve the discrimination of signal and background jets accounts for the significant fraction of gluon-initiated background jets. Due to the larger colour factor of gluons ($C_A = 3$) compared to quarks ($C_F = 4/3$), a larger number of particles is

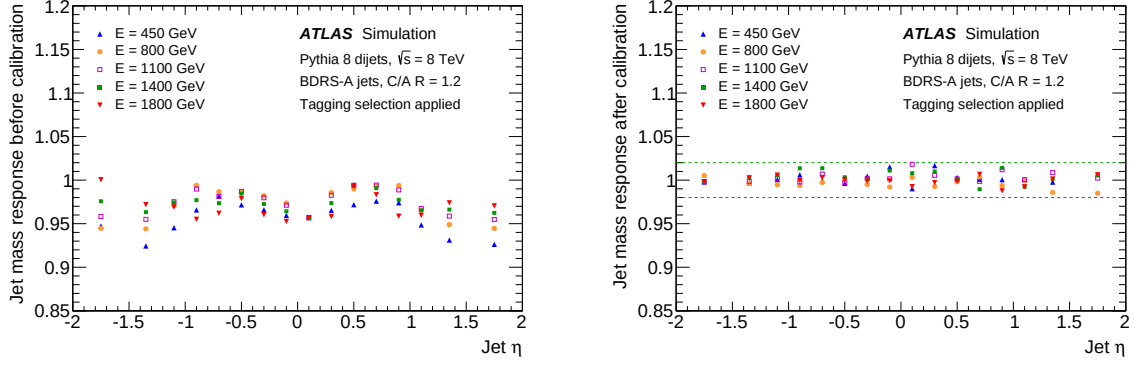


Figure 7.2: Jet mass response as a function of the pseudorapidity for C/A $R = 1.2$ BDRS-A jets in a PYTHIA8 dijet sample for different jet energies before (left) and after (right) the calibration to the particle-level. The selection as described in Section 7.4 is applied.

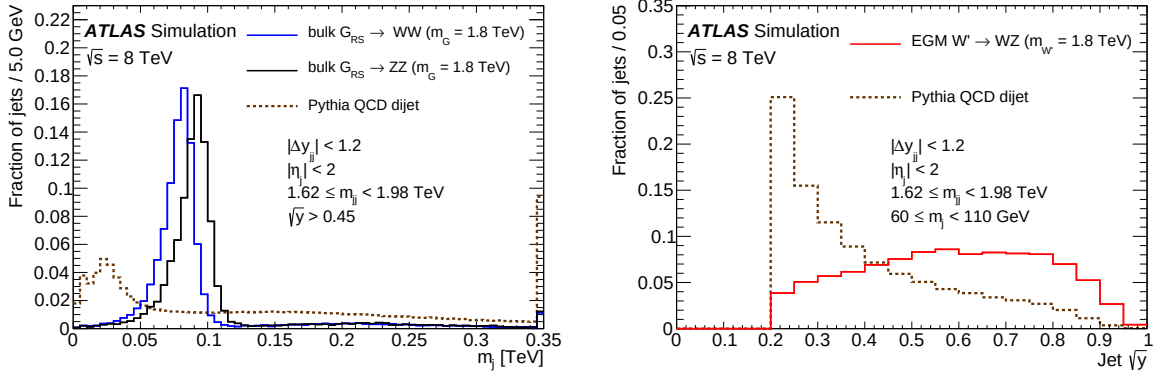


Figure 7.3: Comparison of the variables used to identify high- p_T vector bosons in the 8 TeV diboson analysis for signal jets and QCD dijet background: jet mass (left) and $\sqrt{y_{12}}$ (right).

produced in the fragmentation of a gluon. To differentiate between a gluon-initiated and quark-initiated jet, the number of charged-particle tracks in the Inner Detector, associated to the ungroomed large- R jet, is used [144]. The ungroomed large- R jet is chosen for the ghost-association of charged-particle tracks as the active area is reduced in the grooming procedure and thus less tracks would be associated to the active area of the groomed jet compared to that of the ungroomed jet. This choice does not introduce any pile-up dependence as the charged-particle tracks are associated to the primary vertex. Furthermore the charged-particle tracks have to fulfil the quality criteria described in Section 4.1.2. The number of tracks associated to the ungroomed large- R jet, n_{trk} , is depicted in Fig. 7.4 for signal and background events. The agreement of the n_{trk} distribution for anti- k_t $R = 0.4$ jets in data with the MC predictions is also shown in Fig. 7.4. Deviations of about 20% (50%) can be observed in the core (tails) of the n_{trk} distribution simulated with PYTHIA8. The core of the n_{trk} distribution in data is

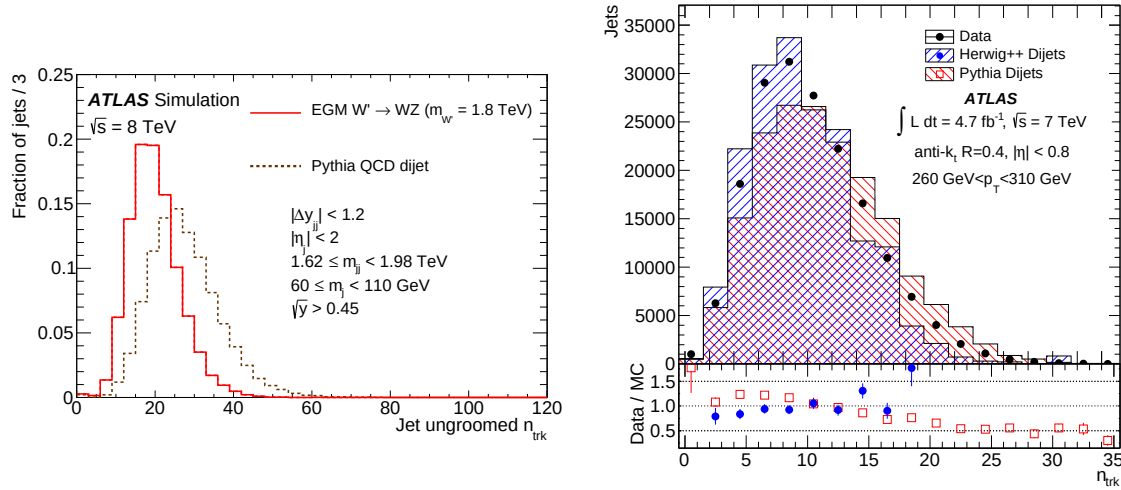


Figure 7.4: Left: Comparison of the n_{trk} distribution for C/A $R = 1.2$ split-filtered jets for QCD dijet background and $W' \rightarrow WZ$ signal events. Right: The n_{trk} distribution for anti- k_t $R = 0.4$ jets in data is compared to two MC generators: PYTHIA8 and HERWIG++ [144].

better modelled by HERWIG++ than PYTHIA8 as opposed to the tails of the distribution with deviations of more than 70% for HERWIG++. Therefore, the selection criteria on n_{trk} as well as its efficiency is measured in data and MC simulation and a scale factor is applied to the simulated signal to correct for this difference.

To determine the optimal n_{trk} requirement, events are pre-selected based on the event selection in Section 7.4. In order to select a V +jet enriched sample in data, the leading jet in the event is required to have a transverse momentum in the range $580 < p_T < 680$ GeV and to satisfy the boson tagging criteria $\sqrt{y_{12}} > 0.45$. Four different upper selection boundaries on the n_{trk} variable are considered: 20, 25, 30 and 35. The mass distribution is then fitted for each considered n_{trk} requirement to extract the signal and background composition. A polynomial function is used to parameterise the QCD dijet background and the W and Z contribution is described by a pair of crystal ball functions. The shape parameters of the crystal ball function are determined by a fit to the W/Z +jet spectrum in a $W' \rightarrow WZ \rightarrow qq\bar{q}\bar{q}$ sample in the same p_T range as considered in data. An example fit to the mass distribution of the leading jet after a requirement of $n_{\text{trk}} < 30$ is shown in Fig. 7.5. Two polynomial functions of fourth and third order are considered for the parameterisation of the background mass distribution. For each n_{trk} requirement, the signal-to-background ratio and $S\sqrt{B}$ sensitivity are determined, resulting in an optimal criteria of $n_{\text{trk}} < 30$. The signal efficiency of the $n_{\text{trk}} < 30$ requirement in data is then calculated as the average of the signal efficiencies obtained by the two polynomial fits and is about 83%.

The scale factor S to correct the signal efficiency in MC simulation, ϵ_{MC} , to that in data, $\epsilon_{\text{data}} = S \cdot \epsilon_{\text{MC}}$, is then derived with V +jets events generated with PYTHIA. As for the optimisa-

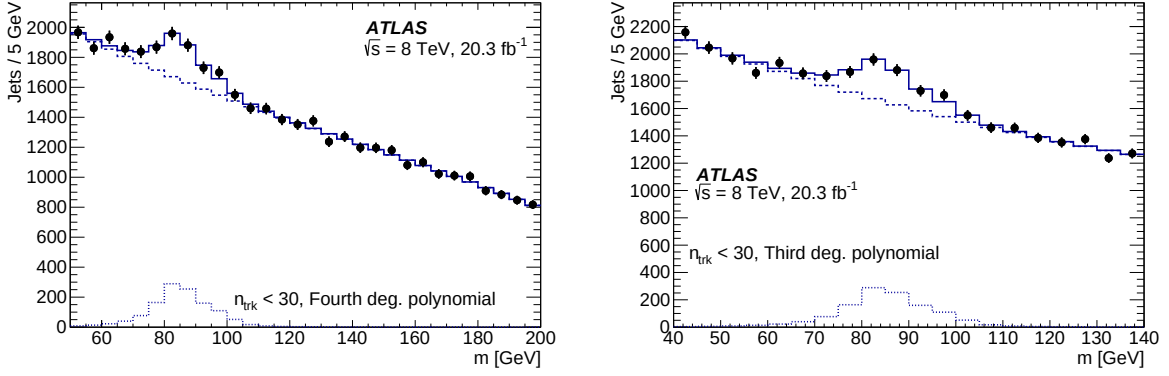


Figure 7.5: Jet mass distribution in a V +jet enriched sample in $\sqrt{s} = 8$ TeV data for jets passing the $n_{\text{trk}} < 30$ requirement using a fourth order (left) and third order (right) polynomial for the QCD dijet background parameterisation.

tion of the n_{trk} requirement, the mass distribution in MC simulation is fitted again with a fourth and third degree polynomial and a crystal ball function before and after requiring $n_{\text{trk}} < 30$. The resulting scale factor per jet is 0.90 ± 0.08 . The uncertainty on the scale factor is dominated by the mismodelling of the jet mass spectrum of the background and is derived by comparing the yields obtained with both polynomial functions.

$\sqrt{s} = 13$ TeV Analysis

Jets used in the 2015 diboson resonance search are reconstructed with the anti- k_t $R = 1.0$ algorithm and trimmed with $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.05$ based on the optimisation studies described in Section 6.2. Dedicated energy and mass calibrations are then applied. The requirements to identify hadronically decaying boosted vector bosons were described in detail in Section 6.2 and include p_T dependent cuts on the jet mass and the $D_2^{\beta=1}$ variable to obtain a signal efficiency of approximately 50% and a background efficiency of 1.5 – 2.0%. The leading jet mass and $D_2^{\beta=1}$ distributions for a $\text{HVT} \rightarrow WW$ signal sample are compared in Fig. 7.6 to the distributions in the QCD dijet background sample and data collected at a centre-of-mass energy of 13 TeV.

Furthermore, only jets with less than 30 charged-particle tracks of $p_T > 500$ MeV associated to the ungroomed anti- k_t $R = 1.0$ jet are selected. The n_{trk} selection criterion was re-optimised for the 13 TeV analysis due to the change in jet reconstruction algorithm, however the same requirement was found to be optimal. An improvement of about 30% in sensitivity of the analysis for HVT signals is expected by requiring $n_{\text{trk}} < 30$. The n_{trk} distribution is shown in Fig. 7.7 separately for the leading and subleading jet in the $\text{HVT} \rightarrow WW$ signal and QCD dijet background sample. The agreement between the n_{trk} distribution in PYTHIA8 MC simulation and data is significantly improved with respect to Run-I due to the A14 generator tune.

As for the Run-I analysis, the efficiency of the $n_{\text{trk}} < 30$ selection is measured in an enriched V +jet data sample by requiring the leading jet to have a p_T between 500 and 700 GeV and

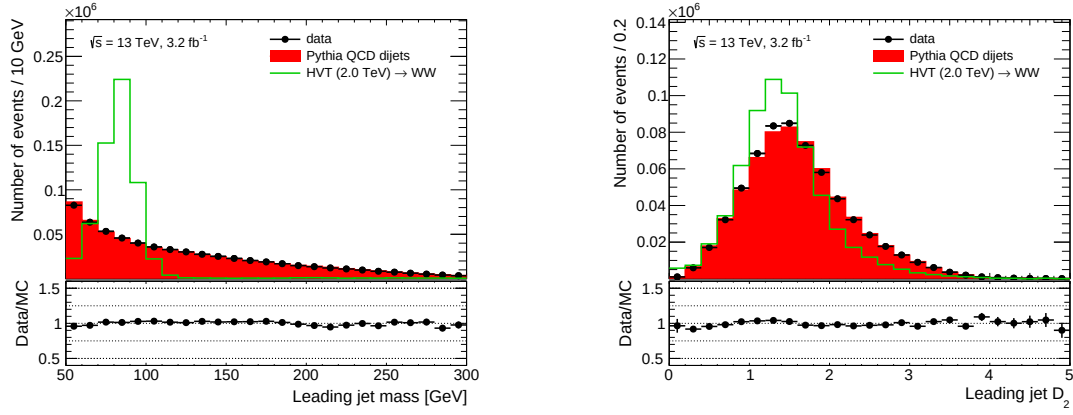


Figure 7.6: Comparison of the variables used to identify high- p_T vector bosons in the 13 TeV diboson analysis for signal jets, QCD dijet background and data: jet mass (left) and $D_2^{\beta=1}$ (right). The dijet background and signal sample are normalised to data.

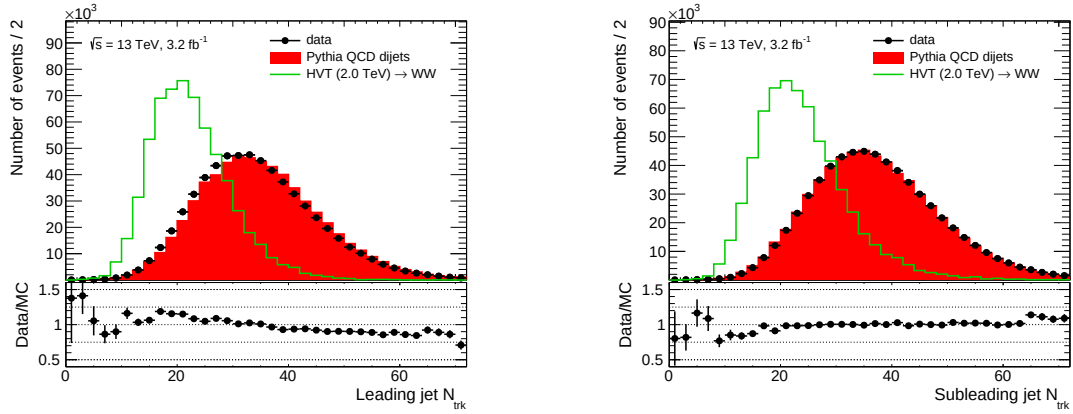


Figure 7.7: Comparison of the n_{trk} distribution for the leading jet (left) and subleading jet (right) in a HVT signal and QCD dijet background sample and for $\sqrt{s} = 13$ TeV data. The dijet background and signal sample are normalised to data.

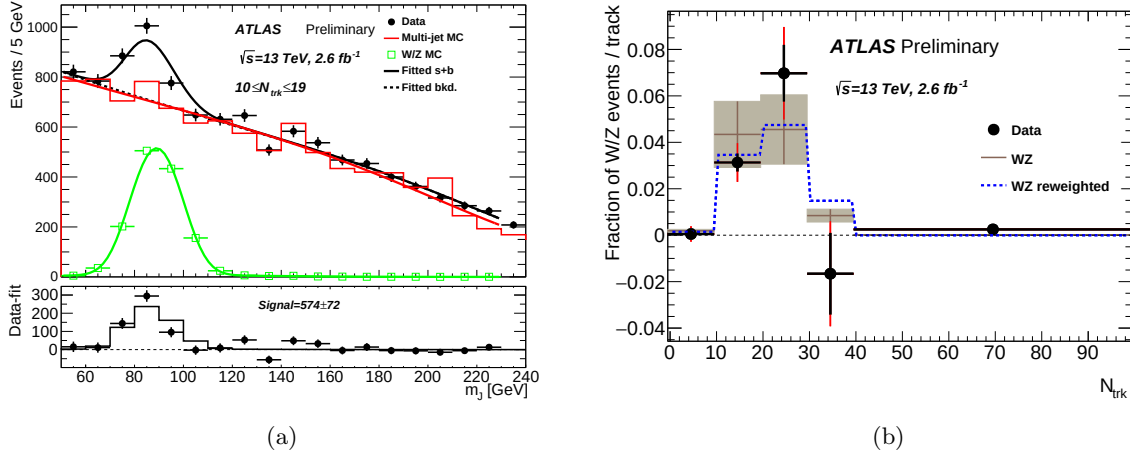


Figure 7.8: Left: Fit to the leading jet mass distribution in a V +jet enriched sample in $\sqrt{s} = 13$ TeV data for jets with $10 \leq n_{\text{trk}} \leq 19$ to extract the number of W/Z events. Right: Comparison of the fraction of W/Z events in data and $W' \rightarrow WZ$ MC simulation for the different considered n_{trk} bins.

to pass either the W -jet or Z -jet $D_2^{\beta=1}$ criteria. The number of W and Z boson candidates is extracted from fits to the leading jet mass distribution m_J . The background mass distribution is parameterised by an exponential function multiplied with a sigmoid turn-on and the signal mass distribution is described by a double Gaussian. Only the signal strength is allowed to float in the fit and the signal parameters are obtained by fits to the W and Z mass distribution in the $W' \rightarrow WZ$ sample with $m_{W'} = 1200$ GeV. The leading jet mass distribution is fitted in five distinct bins of n_{trk} : $n_{\text{trk}} \leq 9$, $10 \leq n_{\text{trk}} \leq 19$, $20 \leq n_{\text{trk}} \leq 29$, $30 \leq n_{\text{trk}} \leq 39$ and $n_{\text{trk}} \geq 40$. The mass distribution for jets with $10 \leq n_{\text{trk}} \leq 19$ in 2.6 fb^{-1} of 13 TeV data and the corresponding fits are shown in Fig. 7.8(a). The parameterisations of the signal and background mass distributions are represented by the continuous green and red line. A background only and signal plus background fit is then performed to the mass distribution in data. The number of W/Z events in data is then obtained by subtracting the background only fit from the data and is shown in the lower panel of Fig. 7.8(a). Furthermore the lower panel contains the number of W/Z events as expected in MC simulation (black continuous line), obtained by subtracting the background only fit from the total fit. For each considered n_{trk} bin, the fraction of W/Z events is extracted in data and compared in Fig. 7.8(b) to that in $W' \rightarrow WZ \rightarrow qqqq$ MC simulation. To derive a scale factor that corrects the n_{trk} distribution in MC simulation to that in data, a template fit is performed, yielding a scale factor of about 1.07. The larger number of tracks in data results in a lower efficiency for the $n_{\text{trk}} < 30$ selection of $5\% \pm 6\%$ compared to MC simulation. This 6% uncertainty on the difference in efficiency is taken as a systematic uncertainty on the $n_{\text{trk}} < 30$ requirement but no shift in the central value is applied.

To be able to compare the performance of the boosted vector boson identification used in the

	Run-I	Run-II
Jet algorithm	C/A $R = 1.2$	anti- k_t $R = 1.0$
Jet grooming	Split-Filtering: $R_{\text{sub}} = 0.3$, $y_{\text{cut}} = 0.04$, $\mu = 100\%$	Trimming: $f_{\text{cut}} = 5\%$, $R_{\text{sub}} = 0.2$
Jet tagging	Mass window ± 13 GeV $\sqrt{y_{12}} \geq 0.45$ $n_{\text{trk}} < 30$	Mass window ± 15 GeV D_2 p_T -dependent cut $n_{\text{trk}} < 30$

Table 7.1: Summary of the Run-I and Run-II boosted boson tagging algorithms.

8 TeV analysis to that of the 13 TeV analysis, the C/A $R = 1.2$ jet collection is also studied at $\sqrt{s} = 13$ TeV. The default anti- k_t $R = 1.0$ trimmed jets and the corresponding boson tagging criteria will be denoted in the following as Run-II tagger whereas the split-filtered C/A $R = 1.2$ jets with boson tagging criteria on the jet mass and $\sqrt{y_{12}}$ will be denoted as Run-I tagger. Dedicated jet energy and mass calibrations at 13 TeV were derived for the Run-I tagger. The differences between the Run-I and Run-II tagger are summarised in Tab. 7.1.

In the Run-II analysis, jets are furthermore reconstructed from charged-particle tracks using the anti- k_t algorithm and a radius parameter of $R = 0.2$. These track-jets are ghost-associated to large- R jets to identify b -quarks [145]. The identification of b -jets is important for this analysis to avoid possible contaminations from $H \rightarrow b\bar{b}$ decays in the high-mass validation region, defined in Section 7.5. Therefore, large- R jets with masses between 110 and 140 GeV and two or more ghost-associated b -tagged track-jets are rejected. The high-mass validation region is used amongst others to study the background parameterisation.

Three main b -tagging algorithms have been developed in the ATLAS Collaboration to identify jets to originate from the hadronisation of a b -quark: IP3D, SV1 and JetFitter [146, 147]. These algorithms exploit the long lifetime of b -hadrons ($\tau \approx 1.5$ ps), produced in the hadronisation of b -quarks and their kinematic properties. Due to their long lifetime, b -hadrons travel about 0.5 mm from their production vertex before they decay, resulting in a displaced secondary vertex (SV) and a large impact parameter (IP) significance for at least one track associated to the jet. The transverse and longitudinal impact parameter are defined as the minimum distance between the track and the primary vertex in the $x - y$ plane or z -direction, respectively. The JetFitter algorithm seeks to reconstruct the full decay chain of a b -hadron, including tertiary vertices from c -hadron decays. The input variables used in the three b -tagging algorithms are then combined in a boosted decision tree (BDT) to enhance the discrimination of b -jets from light-flavour and c -jets. The output of the BDT is referred to as MV2c20 algorithm as the training is performed with b -jets as signal and a c -fraction of 20% and light-flavour fraction of 80% as background.

7.3.2 Leptons

Electrons

Electrons are reconstructed from isolated energy clusters in the electromagnetic calorimeter that are matched to an Inner Detector track [148, 149]. To ensure that the tracks originate from the primary vertex, selection criteria on the longitudinal and transverse impact parameter are applied. Three identification criteria are defined with increasing background rejection power, labelled as *loose*, *medium* and *tight* to suppress fake electrons. The quality criteria are based on shower shape variables, the fraction of energy deposited in the hadronic calorimeter and quality requirements on the associated tracks. In Run-II, the three identification criteria are based on likelihood method that simultaneously evaluates several properties of the electron whereas only a simple cut-based algorithm was used in Run-I.

Electrons are required to satisfy the *medium* quality requirements and to have $p_T > 20$ GeV and $|\eta| < 2.47$, excluding the transition region between the barrel and endcaps in the LAr calorimeter ($1.37 < |\eta| < 1.52$). The transverse momentum requirement was increased to 25 GeV for the Run-II analysis.

Muons

Muons are reconstructed by combining two independent measurements of charged particles in the Inner Detector and muon spectrometer [150, 151]. To a lesser extent, information from the calorimeter is used as well in regions where measurements from the muon spectrometer cannot be provided. The measured momenta of muons p in the Inner Detector and Muon spectrometer are required to be consistent with each other, $|(q/p)_{\text{ID}} - (q/p)_{\text{MS}}| < 5\sigma$, where q is the charge of the track. Furthermore the tracks are required to originate from the primary vertex. As for electrons, three levels of identification purity are defined for muons, *loose*, *medium* and *tight*, based on the number of hits in the different subsystems.

Muons are required to satisfy the *loose* quality requirements and have $p_T > 20(25)$ GeV and $|\eta| < 2.5$ in the Run-I (Run-II) analysis.

Isolation

To suppress the background from hadronic decays, calorimeter-based (E_T^{cone}) and track-based (p_T^{cone}) isolation criteria are imposed. The scalar sum of the transverse energy within a cone of size $\Delta R = 0.2$ around the lepton, subtracting the transverse energy of the lepton and the contribution from pile-up, has to be less than 30% of the lepton's p_T . The track-based isolation requires the sum of the transverse momenta in a cone of $\Delta R = 0.2$ (0.3) around the electron (muon) candidate to be less than 15% of the lepton's p_T . For the Run-II analysis, the isolation criteria are p_T and η dependent to achieve a flat efficiency of 99% for reconstructed leptons in $Z \rightarrow \ell\ell$ decays [149, 151].

7.3.3 Missing Transverse Momentum

The momentum conservation in the $x - y$ plane transverse to the beam axis requires that the vectorial sum of the transverse momenta of all particles sums up to zero. Any imbalance is quantified by the missing transverse momentum ($\vec{E}_T^{\text{miss}}$) and may indicate the presence of undetectable particles such as neutrinos. The missing transverse momentum is calculated as the negative vector sum of the calibrated transverse momenta of all electrons (e), photons (γ), tau leptons (τ), jets and muons (μ) in the event. To avoid double counting, the energy deposits in the calorimeter are only associated to one object. The components along the x and y axis are:

$$E_{x,y}^{\text{miss}} = E_{x,y}^{\text{miss}, e} + E_{x,y}^{\text{miss}, \gamma} + E_{x,y}^{\text{miss}, \tau} + E_{x,y}^{\text{miss}, \text{jets}} + E_{x,y}^{\text{miss}, \mu} + E_{x,y}^{\text{miss}, \text{soft}}. \quad (7.1)$$

The *soft-term*, $E_T^{\text{miss}, \text{soft}}$, accounts for signals in the detector that are not associated to any hard physics object. In the Run-I analysis, a calorimeter-based soft-term was used [152], calculated from topo-clusters with $|\eta| < 4.9$, whereas a track-based soft-term built from tracks in the ID was used in Run-II [153]. Even though the track-based soft-term misses components from soft neutral particles, it was found to be beneficial, as it removes any pile-up contributions compared to the calorimeter-based variant which only uses noise suppression in the topo-clustering algorithm to avoid fake signals. The magnitude of the missing transverse momentum can then be calculated as:

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}. \quad (7.2)$$

7.4 Event Selection

Events aiming at reconstructing the hadronic decay of vector boson pairs are characterised by two large- R jets with high transverse momentum in the centre of the detector. To enhance the sensitivity of the analysis over a broad range of resonances masses, the event selection criteria were optimised based on MC simulations of the signal processes and the QCD dijet background. Even though the event selection criteria were re-optimised for the Run-II analysis, almost identical requirements were found to be optimal for the Run-I and Run-II analyses. The main differences are the identification of boosted vector bosons, discussed in Section 7.3. Due to the similarities in the event selection, the focus of this section will be on the re-optimisation of the Run-II analysis.

7.4.1 Preselection

Events are required to pass the single jet trigger, introduced in Section 7.2, and to have a vertex with at least two tracks with $p_T > 500$ MeV. A Good Runs List is applied to ensure that events are only selected when all sub-detectors were fully operational. Events with any lepton or a large amount of missing transverse momentum, $E_T^{\text{miss}} \geq 350$ GeV, are vetoed to be orthogonal to other diboson resonance searches.

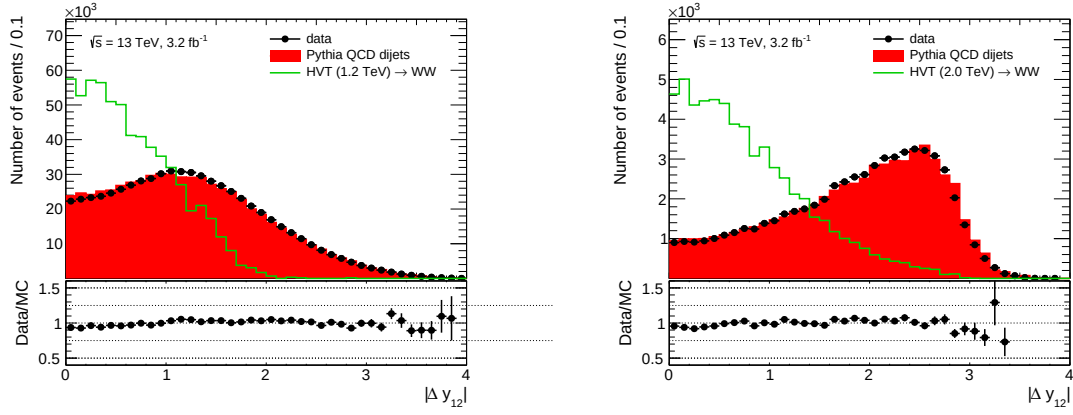


Figure 7.9: Rapidity difference $|\Delta y_{12}|$ between the two leading jets in data, QCD dijet background and for two $Z' \rightarrow WW$ signal samples, $m_{Z'} = 1.2$ TeV (left) and $m_{Z'} = 2.0$ TeV (right). Events are required to pass the preselection and $|m_{JJ} - m_{Z'}| < 0.1 \times m_{Z'}$.

Two large- R groomed jets with $|\eta| < 2.0$ and $p_T > 200$ GeV are required. Furthermore, due to difficulties in the calibration procedure for low jet masses, only jets with $m > 50$ GeV are selected in the Run-II analysis. To select only events in the region where the trigger is 99.9% efficient, the leading- p_T jet is required to fulfil $p_T > 450$ (540) GeV in the Run-II (Run-I) analysis. A larger p_T threshold is needed for the Run-I analysis due to the larger jet radius size that was used and due to changes in the trigger. To avoid distortions of the invariant mass spectrum of the two leading jets (m_{JJ}) due to the p_T requirements, only events with $m_{JJ} > 1000$ GeV and $m_{JJ} > 1050$ GeV are selected for the Run-II and Run-I analysis, respectively.

7.4.2 Topological Selection

Two further selection criteria on the kinematic properties of the large- R jets are defined to suppress the QCD dijet background and to enhance the sensitivity of the analysis.

The QCD dijet background contamination can be reduced by taking into account that the background is dominantly produced in the t -channel compared to the more central s -channel production of the signal processes. Therefore, the rapidity difference $|\Delta y_{12}| = |y_2 - y_1|$ distribution for signal events is expected to peak close to zero whereas the background peaks at higher values. For events passing the preselection, the rapidity difference of the two leading anti- k_t $R = 1.0$ trimmed jets is shown in Fig. 7.9 for a HVT ($Z' \rightarrow WW$) signal sample and the QCD dijet background. The distribution are compared to that in data corresponding to an integrated luminosity of 3.2 fb^{-1} , collected at $\sqrt{s} = 13$ TeV. Two different signal resonance masses, $m_{Z'} = 1.2$ and 2.0 TeV, are shown and each event has to fulfil $|m_{JJ} - m_{Z'}| < 0.1 \times m_{Z'}$. As the discrimination between the signal and QCD dijet background increases for higher resonance masses, the selection criteria is optimised considering two different signal samples ($Z' \rightarrow WW$

and $G_{\text{RS}} \rightarrow ZZ$) and three different resonance masses (1.2, 2.0 and 3.0 TeV). The signal ε_S and background ε_B efficiency as well as the ratio $\varepsilon_S/\sqrt{\varepsilon_B}$ are shown in Fig. 7.10 for the aforementioned signal samples. For low resonance masses, the $\varepsilon_S/\sqrt{\varepsilon_B}$ distribution has a relatively flat maximum over a broader $|\Delta y_{12}|$ range whereas a much larger gain in sensitivity can be achieved for high resonance masses. Based on the distributions in Fig. 7.10, the two leading jets in the event are required to pass $|\Delta y_{12}| < 1.2$.

For the two-body decay of a resonance, the decay products are expected to be balanced in momentum. To reject events with a potentially badly reconstructed jet, the p_T asymmetry A is defined as $A = \frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}$, where $p_{T,1}$ and $p_{T,2}$ are the transverse momenta of the leading and subleading jet, respectively. The same optimisation procedure as for the rapidity gap requirement is adopted, however events are required to pass in addition the $|\Delta y_{12}| < 1.2$ criteria. The corresponding efficiency curves are depicted in Fig. 7.11. As shown in Fig. 7.11, the p_T asymmetry provides only a small discrimination between the signal and the background. Its main aim is to reduce the contamination of mis-measured signal jets and can thus be understood as a jet cleaning criteria. Therefore, the selection criteria on the p_T asymmetry A is not directly motivated by the procedure in Fig. 7.11. Based on the Run-I analysis, only events with $A < 0.15$ are selected. This requirement is very conservative for the $Z' \rightarrow WW$ sample, however a tighter selection criterion would result in a significant reduction of the signal with only minimal gain in sensitive for the graviton signal.

7.4.3 Signal Regions

Each of the two leading jets in events that pass the preselection and topological selection are required to pass either the W boson or Z boson tagging criteria. Events are then categorised in three signal regions based on the fulfilled boson tagging requirements: WW , WZ or ZZ . It should be stressed at this point that the signal regions are not distinct due to the overlapping mass window sizes of W -jets and Z -jets.

The signal efficiencies, obtained in the Run-I analysis, are shown in Fig. 7.12 as a function of the resonance mass after the topological selection and for the full event selection, requiring both leading large- R jets to be boson-tagged. For the $W' \rightarrow WZ$ signal, the total signal efficiency ranges from about 17% for $m_{W'} = 1.2$ TeV to 11% in the $m_{W'} = 3.0$ TeV sample. For the graviton, the signal efficiency decreases from about 14% for $m_{G_{\text{RS}}} = 1.2$ TeV down to 7.5% for graviton masses of 3 TeV.

For the Run-II analysis, signal efficiencies are obtained in Fig. 7.13 as a function of the resonance mass for both the default Run-II boson tagging algorithm as well as for the Run-I boson tagging algorithm. The signal efficiencies are shown after the topological selection, after requiring the two leading large- R jets to be boson-tagged except for the requirement on n_{trk} and after the full

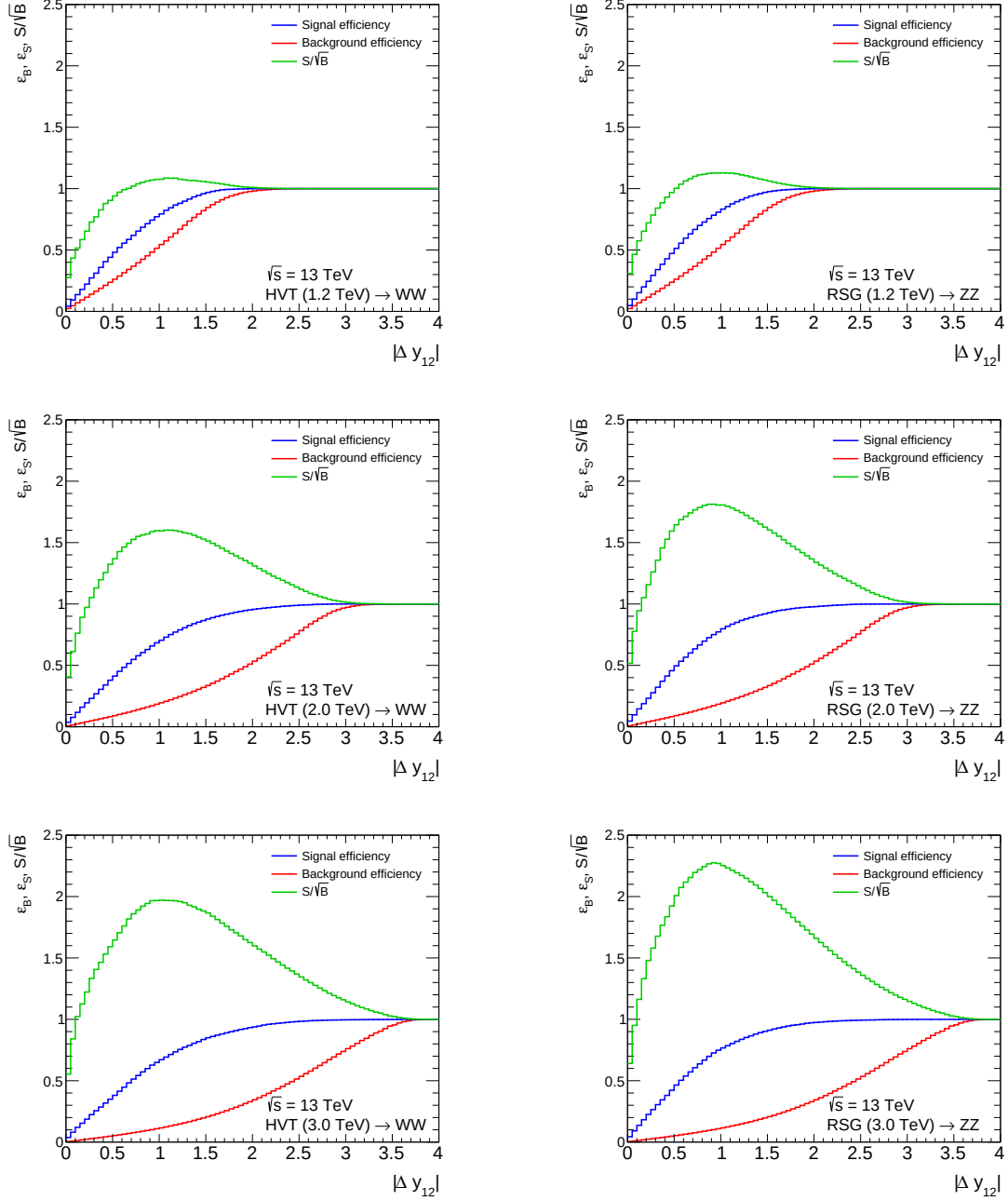


Figure 7.10: Signal and QCD dijet background efficiency (ε_S , ε_B) and the ratio $\varepsilon_S/\sqrt{\varepsilon_B}$ are shown as a function of the requirement on the rapidity gap $|\Delta y_{12}|$ for events passing the pre-selection. The distributions are shown for two signal samples, $Z' \rightarrow WW$ (left) and $G_{RS} \rightarrow ZZ$ (right) and three different signal resonance masses: 1.2 TeV (top), 2.0 TeV (middle) and 3.0 TeV (bottom).

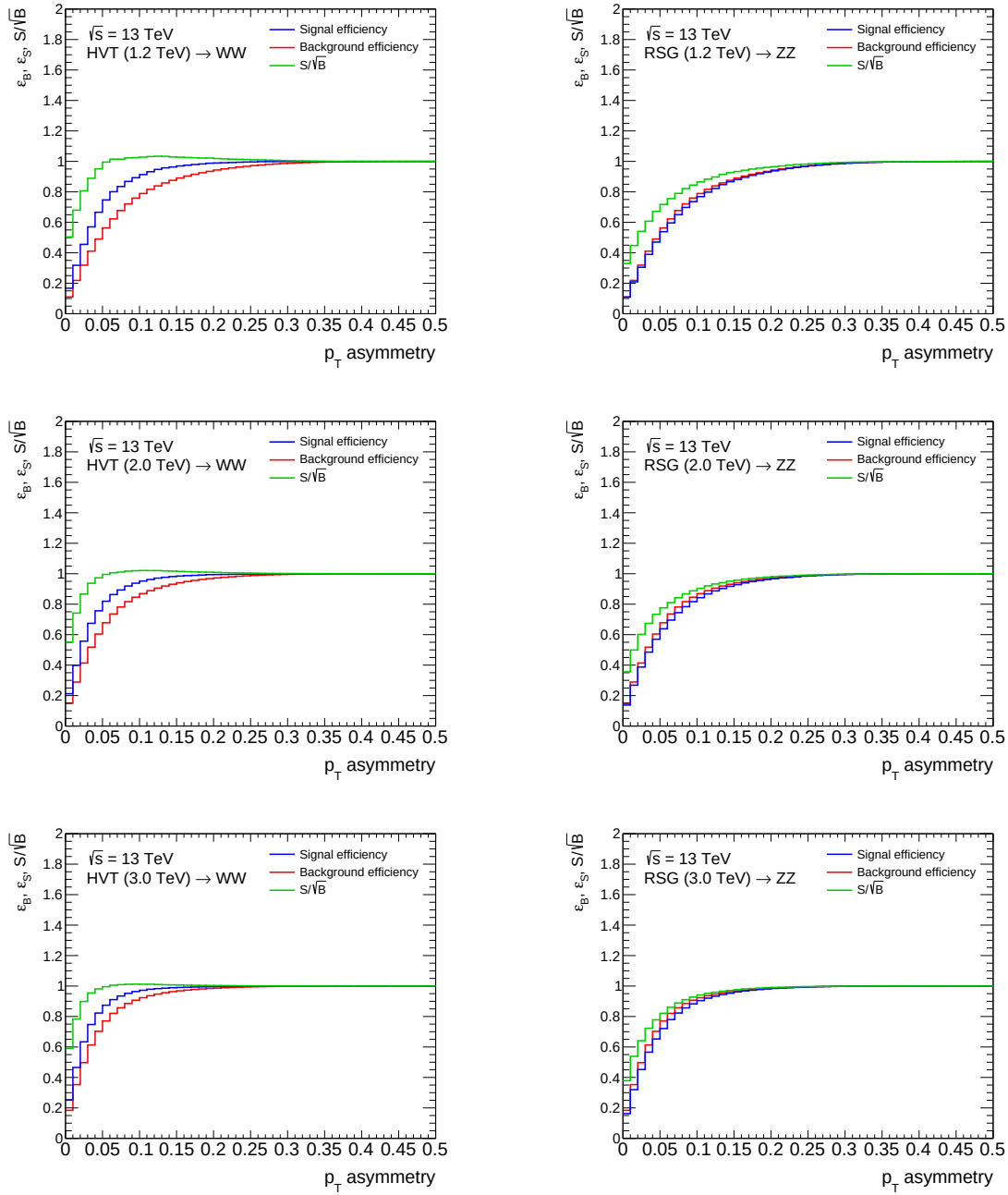


Figure 7.11: Signal and QCD dijet background efficiency (ε_S , ε_B) and the ratio $\varepsilon_S/\sqrt{\varepsilon_B}$ are shown as a function of the requirement on the p_T asymmetry A for events passing the preselection and $|\Delta y_{12}| < 1.2$. The distributions are shown for two signal samples, $Z' \rightarrow WW$ (left) and $G_{RS} \rightarrow ZZ$ (right) and three different signal resonance masses: 1.2 TeV (top), 2.0 TeV (middle) and 3.0 TeV (bottom).

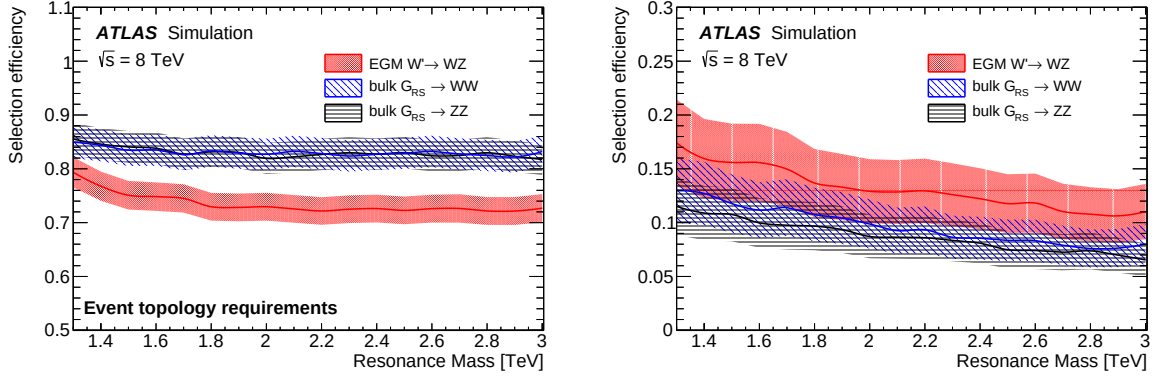


Figure 7.12: Signal selection efficiencies for the Run-I analysis as a function of the resonance mass for the $W' \rightarrow WZ$ and the bulk graviton models $G_{\text{RS}} \rightarrow WW, ZZ$ after requiring the topological selection criteria (defined in Section 7.4.2) only (left) and after the full event selection including the corresponding boson tagging requirements (right).

event selection, including the $n_{\text{trk}} < 30$ selection.

The efficiency of the topological selection is almost identical for both jet collections. For high resonance masses, a higher efficiency is obtained for the graviton signal samples than for the HVT samples. This topological difference between the two samples is related to the spin-1 and spin-2 nature of the HVT and graviton, respectively.

After the requirements on the jet mass and the substructure variable ($D_2^{\beta=1}$ for the Run-II tagger and $\sqrt{y_{12}}$ for the Run-I tagger), the signal efficiencies of the Run-II tagger do not depend on the resonance mass. This behaviour was expected as the boson tagging algorithm was constructed with a flat 50% signal efficiency. For the Run-I tagger, the fixed $\sqrt{y_{12}} > 0.45$ requirement results in a decrease of the signal efficiency with increasing resonance mass. For low resonance masses, higher signal efficiencies are obtained with the Run-I tagger than the Run-II tagger and vice-versa for high resonance masses. The signal efficiency is however only one side of the coin and the performance of the two taggers can only be equitably compared if also the background efficiencies are known.

The signal efficiency after the full event selection, using the Run-II tagger, ranges between approximately 18% and 16% for the HVT samples and 14% to 12% for the graviton sample decaying into Z bosons.

The background efficiency after the topological selection is shown in Fig. 7.14 for both the Run-II and Run-I tagging algorithms. For both tagging algorithms, the predicted background efficiency in MC simulation is compared to the efficiency in data and the distributions are in very good agreement. It was shown in Fig. 7.9 that the rapidity gap of the QCD dijet background depends significantly on the invariant mass of the dijet system, resulting in the steep decrease of the background efficiency with higher m_{JJ} .

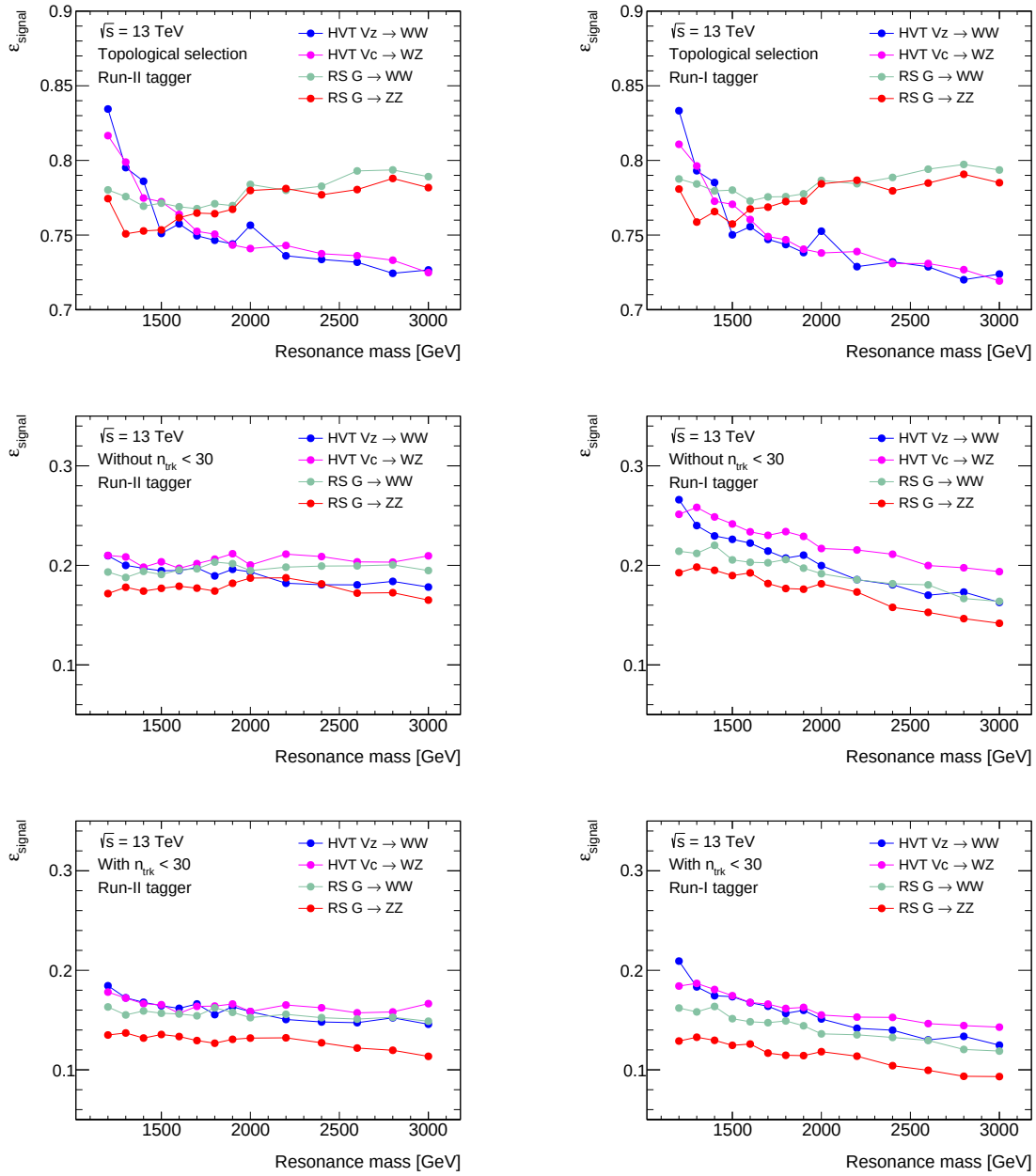


Figure 7.13: Signal selection efficiencies as a function of the resonance mass for the HVT $\rightarrow WW$, WZ and $G_{RS} \rightarrow WW$, ZZ samples after requiring the topological selection criteria only (top), after the boson tagging requirements on the mass and substructure variable (middle) and after the full event selection including the requirement on n_{trk} (bottom). The efficiencies are compared for the Run-II tagging algorithm (left) and the Run-I tagging algorithm (right). The signal efficiencies are computed with respect to the preselection and the invariant mass of the dijet system is required to be within a 10% window around the resonance pole mass.

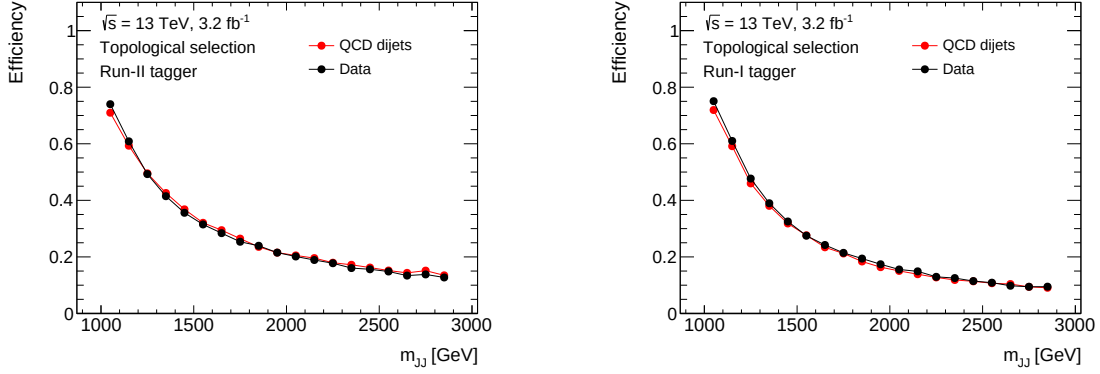


Figure 7.14: Comparison of the efficiency as a function of the invariant mass of the dijet system in QCD dijet MC simulation and $\sqrt{s} = 13$ TeV data after the topological selection is applied. The efficiencies are computed with respect to the preselection and shown for Run-II jet collection (left) and the Run-I jet collection (right).

Furthermore the efficiency of the boson tagging criteria (without the n_{trk} cut) on QCD dijet background events are compared for the two tagging algorithms in Fig. 7.15. Whereas the background rejection is similar between the Run-I and Run-II tagger for high invariant dijet masses, the Run-II tagger outperforms the Run-I tagger in terms of background rejection for low values of m_{JJ} . The background efficiency after the full event selection, using the Run-II tagger, ranges between 0.1% for $m_{JJ} = 1.2$ TeV and 0.01% for $m_{JJ} = 3.0$ TeV. The background efficiencies including the n_{trk} requirement are not shown here as the results are statistically limited.

7.5 Background Model

The fully hadronic diboson resonance search relies on the smoothly falling invariant dijet mass distribution from the background processes and searches for narrow resonances on top of it. The background is dominated by the production of QCD dijets events and the contributions from other SM processes such as diboson production, W/Z +jets and $t\bar{t}$ events are either small or negligible and not expected to distort the shape of the background distribution. The shape of the background is determined by performing a maximum-likelihood fit to the observed m_{JJ} spectrum of the following functional form:

$$\frac{dn}{dx} = p_1(1-x)^{p_2+\xi p_3}x^{p_3}, \quad x = m_{JJ}/\sqrt{s}, \quad (7.3)$$

where p_1 is a normalisation factor, p_2 and p_3 are dimensionless shape parameters and ξ is a dimensionless constant. To estimate the background, the fit is performed in the range $1.05 < m_{JJ} < 3.55$ TeV for the Run-I analysis and $1.0 < m_{JJ} < 2.5$ TeV for the Run-II analysis in 100 GeV wide m_{JJ} bins. The restriction of the Run-II analysis to $m_{JJ} < 2.5$ TeV

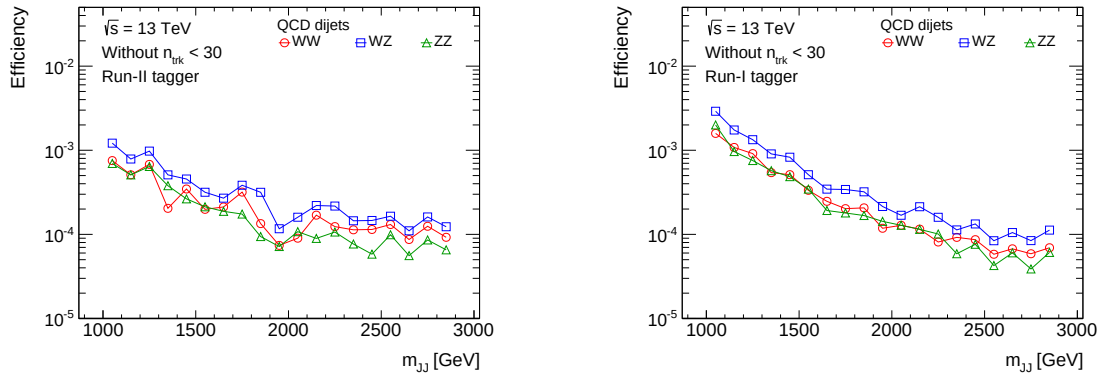


Figure 7.15: Selection efficiency in QCD dijet background events as a function of the invariant mass of the dijet system for the topological selection and the requirements on the mass window and substructure variable. The efficiencies are computed with respect to the preselection and shown for Run-II jet collection (left) and the Run-I jet collection (right).

was induced by the breakdown of the numerical inversion technique when applied to the jet mass calibration for jets with high transverse momentum. The upper limit for the Run-I analysis was chosen based on the fact that no background events were expected to pass the full event selection in the collected dataset beyond $m_{JJ} = 3.55$ TeV. Only the parameters p_1 , p_2 and p_3 are free to float in the fit and the parameter ξ was determined from fits to the m_{JJ} distribution in MC simulation to minimise the correlation between the maximum likelihood estimates of p_2 and p_3 . The functional form in Eq. 7.3 is tested for compatibility to describe distributions with similar characteristics such as the invariant dijet mass in MC generated QCD dijet events or in collision data events in a phase-space close to the signal region. The background model is compared to the m_{JJ} distribution of WZ tagged events in PYTHIA8 dijet simulation in Fig. 7.16 for two different centre-of-mass energies. The lower insets in Fig. 7.16 display the significance, defined as the z -value as introduced in Ref. [154]. The z -value can be directly translated into the p -value, the probability to obtain the observed value or larger values under a given hypothesis. A significance larger than 5σ is required for the discovery of new physics beyond the SM.

The background fit to the m_{JJ} distribution in two validation regions in 2015 collision data is shown in Fig. 7.17. Events in the validation regions are required to pass the topological selection described in Sec. 7.4.2. The mixed mass sideband region requires in addition that one of the two leading- p_T jets has a mass between 50 to 65 GeV and the other jet a mass between 110 and 140 GeV whereas both leading- p_T jets are required to have masses between 110 and 140 GeV in the high mass sideband. To avoid possible contaminations from jets capturing the $H \rightarrow b\bar{b}$ decay, jets with masses between 110 and 140 GeV are required to have less than two ghost-associated b -tagged anti- k_t $R = 0.2$ track-jets. The invariant dijet mass distribution for boson-tagged events in MC simulation and the two validation regions is well-described by the functional form in Eq. 7.3.

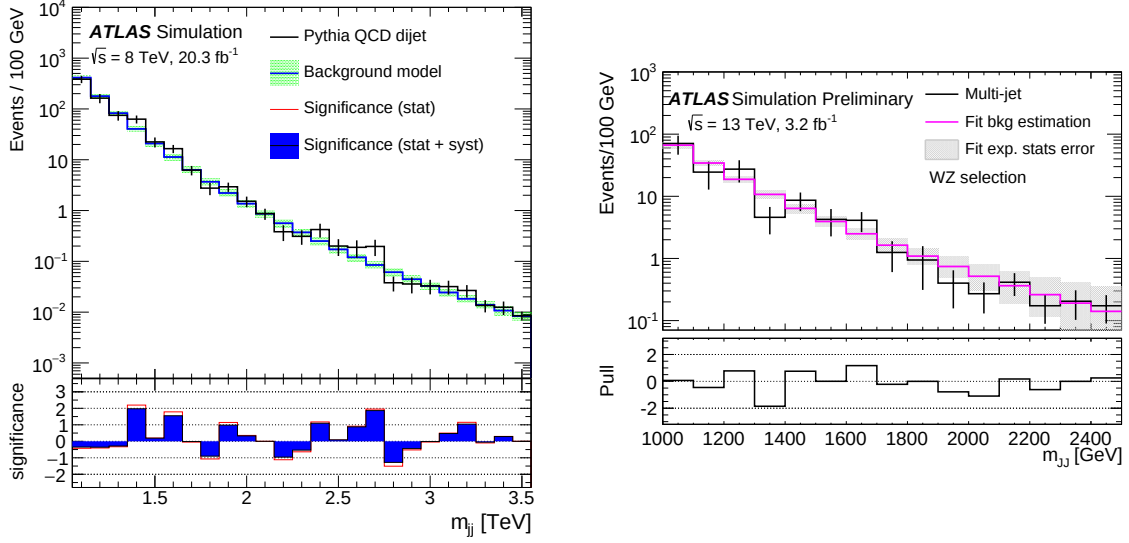


Figure 7.16: Invariant dijet mass spectrum in PYTHIA8 dijet background events, simulated at $\sqrt{s} = 8$ TeV (left) and for $\sqrt{s} = 13$ TeV (right). Events are required to pass the full event selection and WZ tagging requirements and are scaled to the corresponding integrated luminosities of 20.3 fb^{-1} and 3.2 fb^{-1} , respectively. The m_{JJ} spectrum is fitted with the function in Eq. 7.3 and the significance of the deviation from the expectation in each bin is shown in the lower panel.

Further studies were performed to determine if the background determination could be significantly improved by adding additional degrees of freedom to the functional form in Eq. 7.3. Two functional forms were tested with one and two additional parameters:

$$\frac{dn}{dx} = p_1(1-x)^{p_2+\xi p_3} x^{p_3+p_4 \log(x)}, \quad (7.4)$$

$$\frac{dn}{dx} = p_1(1-x)^{p_2+\xi p_3} x^{p_3+p_4 \log(x)+p_5 \log(x)^2}. \quad (7.5)$$

The untagged invariant dijet mass distribution was fitted in pseudo-data, extracted from MC simulation and corresponding to an integrated luminosity 4 fb^{-1} , with each of the functional forms. No significant improvements were observed using higher-order functions and the functional form with the lowest number of parameters (Eq. 7.3) was used for the diboson resonance searches.

For larger data samples however, higher-order functional forms might improve the background estimation. Therefore, the goodness of fit for the different functional forms is tested as a function of the collected data sample size with a log-likelihood ratio statistics employing Wilk's theorem [155]. The test statistics is defined as $-2 \log(\lambda)$, where λ is the ratio of the likelihood of the tested functional form with respect to the likelihood of the functional form with additional parameters. According to Wilk's theorem, the test statistics is a χ^2 distribution with $n - m$

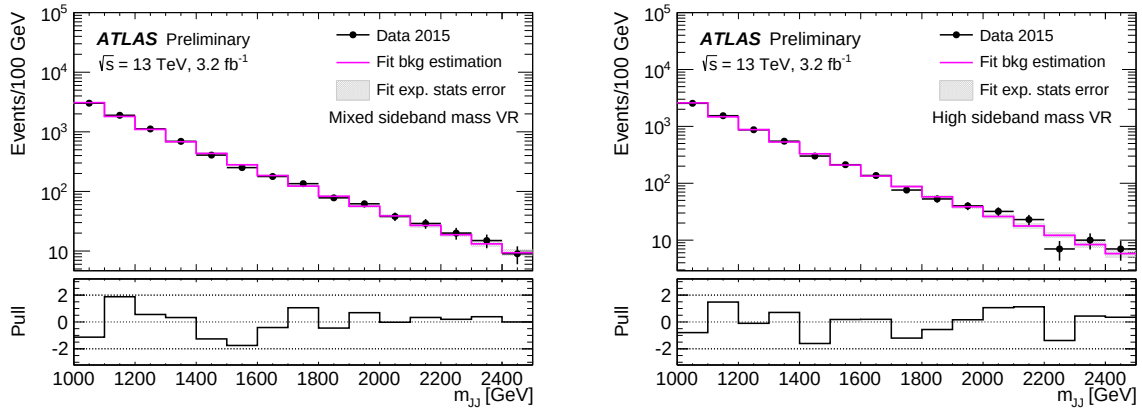


Figure 7.17: Comparison of the background model to the dijet invariant mass spectrum in two validation regions in 2015 data. Left: one jet is required to have $50 < m_J < 65$ GeV and the other jet has to fulfil $110 < m_J < 140$ GeV. Right: both jets are required to have masses between 110 and 140 GeV.

degrees of freedom, where n and m are the dimensions of the tested functional form and the function with additional parameters, respectively. For instance, when the three parameter function is compared to that with four parameters, the test statistic will follow a χ^2 distribution with one degree of freedom. The p -value can then be calculated from the probability density function of the χ^2 distribution and values of less than 0.05 are used as an indication that the background estimate models the data reasonably well. For $p > 0.05$, the fitting function under study is rejected and the procedure starts with the next higher order function of dimension. It was found that no additional parameters, with respect to Eq. 7.3, are necessary for integrated luminosities up to 50 fb^{-1} at $\sqrt{s} = 13 \text{ TeV}$.

7.6 Systematic Uncertainties

Systematic uncertainties, affecting the shape and the normalisation of the m_{JJ} distribution for the different signals, are taken into account as nuisance parameters in the statistical analysis. For the background estimate, only the statistical uncertainties on the fit parameters are considered. Uncertainties arising from the possibility of the fitting function to fake a signal, were estimated by fitting the m_{JJ} distribution in validation regions in data and by considering signal plus background fits using the background model in Eq. 7.3. These effects were estimated to be less than 25% (4%) of the statistical uncertainties on the background model at any mass in the Run-I (Run-II) analysis and thus neglected in the statistical analysis.

The systematic uncertainties are dominated by the uncertainties on the scale and resolution of the large- R jet observables: the jet p_T , mass and substructure variable ($\sqrt{y_{12}}$ or $D_2^{\beta=1}$). The invariant dijet mass is dominated by the sum of the p_T of the two large- R jets and their angular

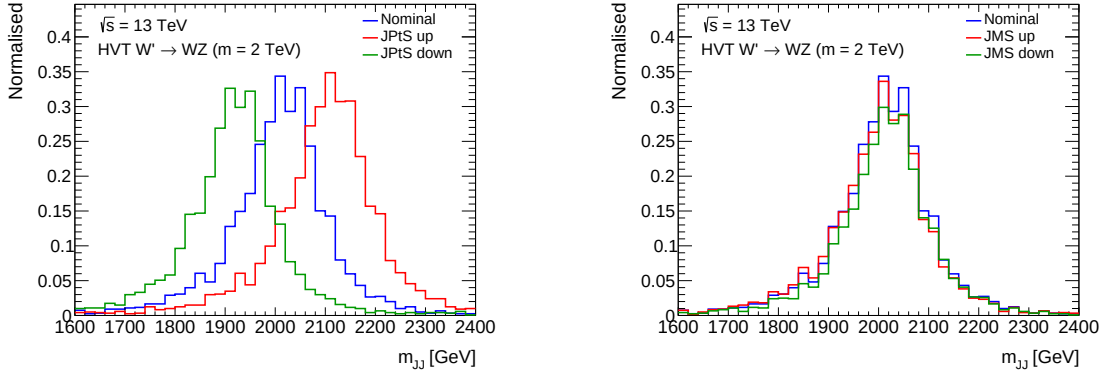


Figure 7.18: The effect of the jet p_T (left) and mass (right) scale uncertainty on the invariant dijet mass spectrum for the HVT signal sample ($W' \rightarrow WZ$) with a mass of 2 TeV. The distributions are normalised to unity to demonstrate only the impact of the uncertainties on the shape of the m_{JJ} distribution.

separation. The uncertainty on the p_T scale shifts the position of the signal peak of the m_{JJ} distribution and the uncertainty on the p_T resolution affects the width of the distribution. The uncertainties on the jet mass and the substructure variables affect mostly the normalisation of the signal m_{JJ} distribution as these variables play only a role in the selection of events. The effect of the jet p_T and mass scale uncertainties on the invariant dijet mass distribution is shown in Fig. 7.18 for the 2 TeV HVT signal. For the Run-I analysis, the systematic uncertainties on the jet variables are derived with the double-ratio method in Section 5.2.1, whereas the Run-II analysis uses the systematic uncertainties described in Section 5.2.3. The scale uncertainties used in the Run-II analysis are in general expected to be larger than the ones of the Run-I analysis since they are based on the Run-I uncertainties with additional MC-based uncertainties to account for differences in the detector and jet reconstruction between Run-I and Run-II as described in Section 5.2.3.

For the Run-I analysis, the uncertainty on the jet p_T scale is 2%, for the jet mass scale is 3% and for the momentum balance $\sqrt{y_{12}}$ is 2%. A relative uncertainty of 20% on the resolution of the jet observables is taken into account in the Run-I analysis. The resolution uncertainties are conservatively estimated by comparing the resolution of the jet mass and substructure variables obtained from variations of the detector geometry, the MC generator and the physics list [156]. The resolutions of the different jet observables are determined from the width of the corresponding responses and are 5% for the jet energy, 7.5% for the mass and 10% for the momentum balance $\sqrt{y_{12}}$.

For the Run-II analysis, the uncertainty on the jet p_T scale is 5%, on the jet mass scale is 6% and for the $D_2^{\beta=1}$ variable is 4–5%. A relative uncertainty of 20% on the jet energy and mass resolution are considered and a relative uncertainty of 10% on the $D_2^{\beta=1}$ resolution.

The systematic uncertainties associated to the selection criteria on the number of charged-particle tracks (n_{trk}) are 20% and 6% for the Run-I and Run-II analyses, respectively. The systematic uncertainties could be significantly reduced because of the better modelling of the n_{trk} distribution in Run-II as discussed in Section 7.3.1.

Furthermore, an uncertainty on the determination of the luminosity of 2.8% and 5% that affects the signal normalisation is considered for the Run-I and Run-II analyses, respectively.

In addition, the Run-I analysis takes into account uncertainties on the parton shower and hadronisation model as well as uncertainties due to the chosen PDF. The uncertainty on the parton shower model is 5% and is estimated by comparing the selection efficiency in signal samples, generated with PYTHIA8 to the selection efficiencies obtained for signal samples generated with PYTHIA8 but using HERWIG++ for the parton shower. The uncertainty on the choice of the PDF is 3.5%. Both uncertainties affect the normalisation of the m_{JJ} distribution. These uncertainties were not included in the Run-II analysis due to time constraints however their impact on the cross-section times branching ratio limits in the Run-I analysis were negligible.

7.7 Statistical Analysis

The observed m_{JJ} spectra, obtained in Chapter 8 and 9, are compared to the background-only and signal-plus-background hypotheses. In the absence of a significant deviation from the background-only hypothesis, quantified in terms of the local p_0 value, upper limits on the production cross-section times branching ratio of the different benchmark models can be set using the CL_s method [157].

A test statistic, based on the profile-likelihood ratio [158], is used to test the compatibility of the observed m_{JJ} spectra with a given hypothesis and to extract the *signal strength*, μ , defined as a scale factor on the cross-section times branching ratio predicted by the signal hypothesis. The background-only hypothesis corresponds to $\mu = 0$ and the signal-plus-background hypothesis corresponds to $\mu = 1$. Assuming that the search is performed in m bins in the invariant dijet mass spectrum, which are statistically independent, the likelihood to observe n_i events under the assumption of the signal-plus-background hypothesis, is given by

$$L(\mu) = \prod_{i=1}^m \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} , \quad (7.6)$$

where s_i and b_i are the expected signal and background events, respectively. The number of expected signal events is evaluated based on MC simulations, assuming the cross-section of the signal model. The number of expected background events in one particular m_{JJ} bin is obtained by integrating the function in Eq. 7.3 over the considered range. The effect of systematic uncer-

tainties is taken into account through nuisance parameters, θ , in the likelihood function which are constrained with a corresponding probability density function $f(\theta)$. Systematic uncertainties that affect the shape of the invariant dijet mass are modelled by a Gaussian prior whereas uncertainties affecting the normalisation of the m_{JJ} spectrum are assumed to have a log-normal probability density function. For instance, a systematic uncertainty of 2% on the p_T scale is modelled by a Gaussian probability density function with mean value of one and a standard deviation of 0.02. To account for systematic uncertainties, the likelihood function is modified:

$$L(\mu) = \prod_{i=1}^m \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \prod_k f_k(\theta_k) . \quad (7.7)$$

The test statistic t is then defined through the profile-likelihood ratio $\lambda(\mu)$:

$$t_\mu = -2 \ln \lambda(\mu) = -2 \ln \left(\frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\hat{\theta}})} \right) , \quad (7.8)$$

where $\hat{\mu}$ and $\hat{\hat{\theta}}$ maximise the likelihood function and $\hat{\hat{\theta}}(\mu)$ maximises the likelihood function in the numerator for a certain value of μ .

To quantify the compatibility of the data with the background-only model, the p_0 value is calculated from Eq. 7.8:

$$p_0 = \int_{t_{0,\text{obs}}}^{\infty} f(t_0|0) dt_0 , \quad (7.9)$$

where $t_{0,\text{obs}}$ is the value of the test static observed from the data and $f(t_0|0)$ denotes the probability density function of the test statistic t_0 under the assumption of the background-only hypothesis. The local p_0 value is defined as the probability of the background-only model to produce a signal-like fluctuation at least as large as observed in the data. Taking into account that an excess could occur anywhere in the invariant dijet mass range, a global p value of 1.3×10^{-3} , corresponding to a significance of 3σ , indicates the evidence of new physics whereas a global p value of 2.9×10^{-7} , corresponding to a significance of 5σ , is needed to claim a discovery of a new particle.

To establish upper limits on the signal strength parameter, the agreement between the observed data and the signal-plus-background hypothesis with a given μ is quantified by the p -value:

$$p_\mu = \int_{t_{\mu,\text{obs}}}^{\infty} f(t_\mu|\mu) dt_\mu , \quad (7.10)$$

with the probability density function $f(t_\mu|\mu)$ of the test statistic for a given μ . The test statistic t_μ is set to zero for $\mu > \hat{\mu}$ to ensure that signals larger than expected are not considered as evidence against a model. For $p_\mu < \alpha$, the signal-plus-background hypothesis is excluded at

$1 - \alpha$ confidence level (CL). The disadvantage of using p_μ is that if the number of observed events is less than the number of predicted background events, the signal-plus-background hypothesis could be excluded even though the experiment is not sensitive to the signal because of small expected signal yields. To account for this effect, the *Modified Frequentist confidence level* CL_s was introduced:

$$\text{CL}_s = \frac{p_\mu}{1 - p_0} .$$

The upper limits from CL_s are weaker than the ones from p_μ since $\text{CL}_s > p_\mu$. As for p_μ , the signal-plus-background hypothesis is excluded at 95% confidence level if $\text{CL}_s < 0.05$. For each benchmark model and signal mass point, the signal strength corresponding to the 95% CL is determined and then multiplied by the initial cross-section times branching ratio to determine the upper limits.

To calculate the p -values, the probability density functions of the test statistic in Eq. 7.9 and Eq. 7.11 need to be known. These distributions can be determined using MC simulation, however a large number of pseudo-experiments is needed which is computationally expensive. Instead, the asymptotic distribution of the test statistics, described in detail in Ref. [158], is used to determine the upper limits on the cross-section times branching ratio. This approximation of the profile-likelihood ratio becomes exact in the large sample limit. For the diboson resonance search, a few mass points were tested to ensure that the upper cross-section limits and the including uncertainty bands of the approximation are in good agreement with the exact results.

7.8 Summary

In this chapter, the common aspects of the 8 and 13 TeV diboson resonance searches in the fully hadronic final state were presented. The final state is characterised by two large-radius jets with high transverse momentum. The analysis thus searches for a resonant structure on top of the smoothly falling invariant dijet mass spectrum of the background processes. To avoid statistical limitations of the background MC samples in the high mass region, the background is parameterised with a parametric function. A statistical analysis is used to either quantify an excess or in the absence of a significant deviation to set limits on the production cross-section times branching ratio with the CL_s method. Systematic uncertainties, affecting the shape or the normalisation of the m_{JJ} distribution for different signal processes, are taken into account as nuisance parameters in the statistical analysis. The sources of systematic uncertainties considered in the Run-I and Run-II analyses and their magnitude are summarised in Tab. 7.2. In the following, the differences between the 8 and 13 TeV analysis are highlighted.

For the Run-I analysis, jets are reconstructed with the C/A algorithm with a radius parameter of $R = 1.2$. The split-filtering algorithm with $R_{\text{sub}} = 0.3$ and $y_{\text{cut}} = 0.04$ was used to

Systematic uncertainty	Run-I	Run-II
Jet p_T scale	2%	4%
Jet mass scale	3%	6%
Jet $\sqrt{y_{12}}$ scale	2%	—
Jet $D_2^{\beta=1}$ scale	—	4-5%
Jet p_T resolution	20%	20%
Jet mass resolution	20%	20%
Jet $\sqrt{y_{12}}$ resolution	20%	—
Jet $D_2^{\beta=1}$ resolution	—	10%
$n_{\text{trk}} < 30$	20%	6%
Luminosity	2.8%	5%
Parton shower model	5%	—
PDF	3.5%	—

Table 7.2: Summary of the systematic uncertainties considered in the Run-I and Run-II analyses.

remove pile-up effects from the jet. To identify boosted vector bosons, jets are required to have a mass within a 26 GeV wide window around the peak of the W/Z -jet mass distribution and to have a momentum balance of $\sqrt{y_{12}} \geq 0.45$. Furthermore, the number of charged-particle tracks, ghost-associated to the large- R jet, has to be less than 30. The signal efficiency of the boson tagging criteria decreases with p_T .

For the Run-II analysis, jets are reconstructed with the anti- k_t algorithm with $R = 1.0$. The trimming algorithm with $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 5\%$ was chosen as baseline grooming algorithm. The boosted vector boson algorithm requires jets to have a mass within a 30 GeV wide window around the mean of the W/Z -jet mass distribution and imposes a p_T dependent criteria on the $D_2^{\beta=1}$ variable to achieve a constant signal efficiencies of 50% per jet. As for the Run-I analysis, only jets with $n_{\text{trk}} < 30$ are selected.

Due to the smaller jet radius used in the Run-II analysis, the requirement on the transverse momentum of the leading- p_T jet could be reduced with respect to the Run-I analysis. In the Run-II analysis, the leading- p_T jet is required to fulfil $p_T > 450$ GeV, whereas a threshold of 540 GeV was used in the Run-I analysis. This criterion is needed to select only events in the region where the trigger is 99.9% efficient. The invariant mass range could furthermore be extended down to $m_{JJ} > 1000$ GeV in the Run-II analysis as opposed to $m_{JJ} > 1050$ GeV in the Run-I analysis because of the lower transverse momentum threshold.

Events with leptons in the final state are vetoed in the fully hadronic diboson resonance search to ensure that an orthogonal data sample to other diboson searches with leptons in the final state is obtained. The different calibration schemes applied to leptons in Run-I with respect to Run-II have thus only a negligible impact on the analysis and are therefore not further summarised here.

Chapter 8

Diboson Resonance Search at $\sqrt{s} = 8$ TeV

The results of the search for narrow resonances decaying into pairs of vector bosons (WW , WZ and ZZ) using 20.3 fb^{-1} of pp collisions at $\sqrt{s} = 8$ TeV are presented in this chapter [87]. The background-only hypothesis, defined in Section 7.5, is fitted to the invariant dijet mass spectra in the three different signal regions in Section 8.1 to search for an excess in data. A possible narrow excess of data events over the background parameterisation could indicate the presence of a BSM diboson resonance. In the absence of a significant excess, upper limits on the production cross-section times branching ratio of the different benchmark models are set. The statistical interpretation of the analysis is described in Section 8.2. To improve the sensitivity to new diboson resonances, the results of the fully hadronic search are combined in Section 8.3 with other diboson resonance searches [159], where at least one vector boson decays into leptons. The fully leptonic final state is only sensitive to low resonance masses whereas the semi-leptonic searches also improve the sensitivity to diboson resonances with high masses ($m_{JJ} > 1$ TeV).

8.1 Results

The background-only fits to the invariant dijet mass distributions for events passing the event selection described Section 7.4 are shown in Fig. 8.1 for the three signal regions. Furthermore the expected m_{JJ} distributions for a RS graviton and a W' resonance with different masses are displayed. The m_{JJ} distributions are well-described by the background parameterisations over the entire mass range explored, except for a slight excess of observed events over the background model at the end of the accessible mass range in the 2012 data of about 2.0 TeV. A total of 425, 333 and 604 events are observed in the explored mass range in the WW , ZZ and WZ signal region, respectively. Due to the size of the chosen mass window around the W boson and Z boson mass, the three signal regions are not orthogonal and thus about 20% of the selected events are shared among all three signal regions. The fraction of events selected by both the WZ and WW selection are 49% whereas 43% of the events are common to the WZ and ZZ selection. The compatibility of the data with the background-only model is quantified in terms of the local p_0 value, defined in Section 7.7, and is shown in Fig. 8.2 for the three signal regions. The largest excess of events is observed in the WZ signal region in the mass range $1.95 < m_{JJ} < 2.05$ TeV, where $2.08^{+0.39}_{-0.33}$ events are expected from the background model and eight events are observed

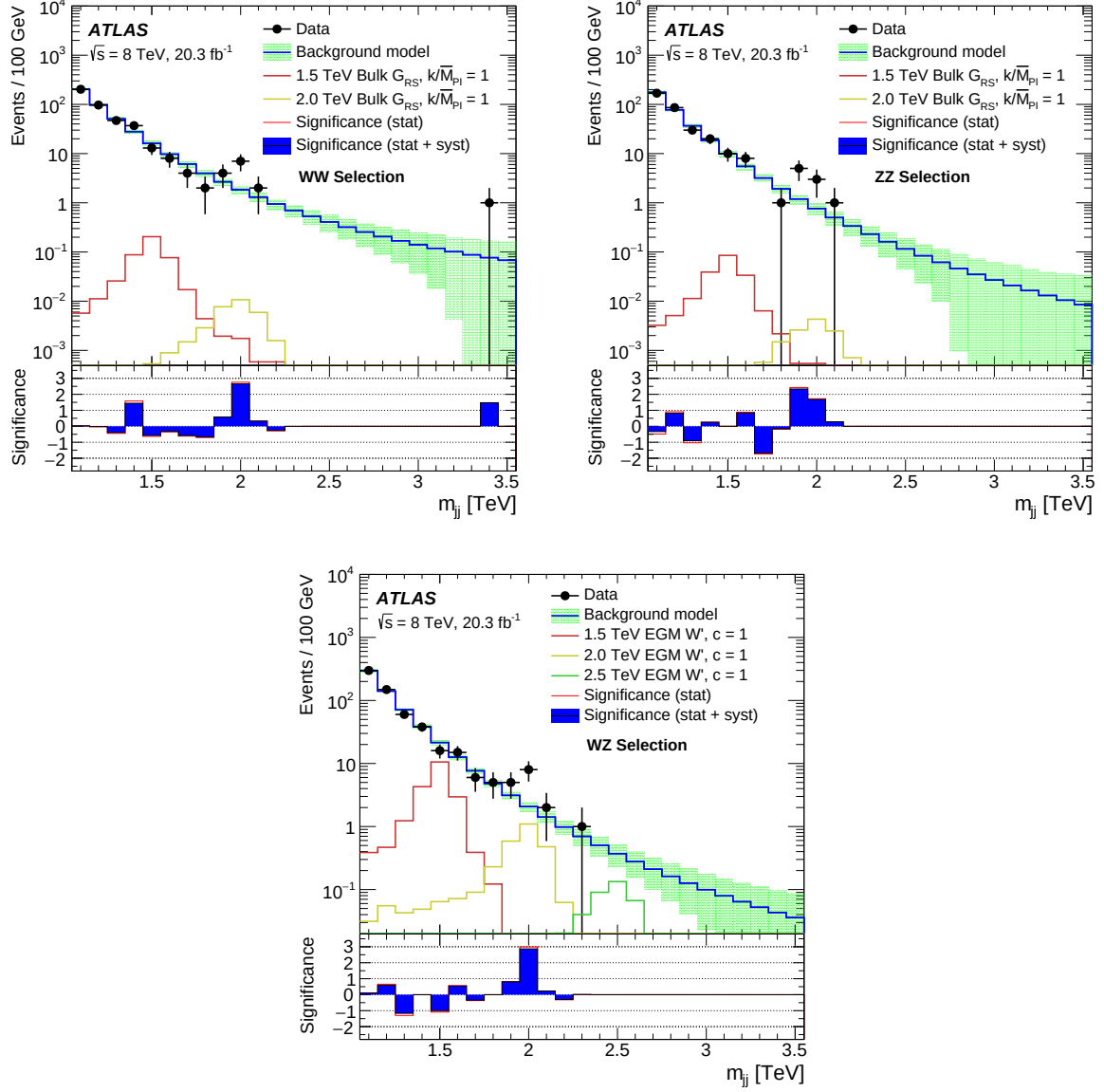


Figure 8.1: Background-only fit to the invariant dijet mass (m_{JJ}) distribution in $\sqrt{s} = 8$ TeV data for boson-tagged jets in the three different signal regions: WW (top left), ZZ (top right) and WZ (bottom). The m_{JJ} distribution in data is compared to the expected signal distributions of a W' with $m_{W'} = 1.5, 2.0$ or 2.5 TeV in the WZ signal region and a RS graviton with $m_{GRS} = 1.5$ or 2.0 TeV in the WW and ZZ signal region.

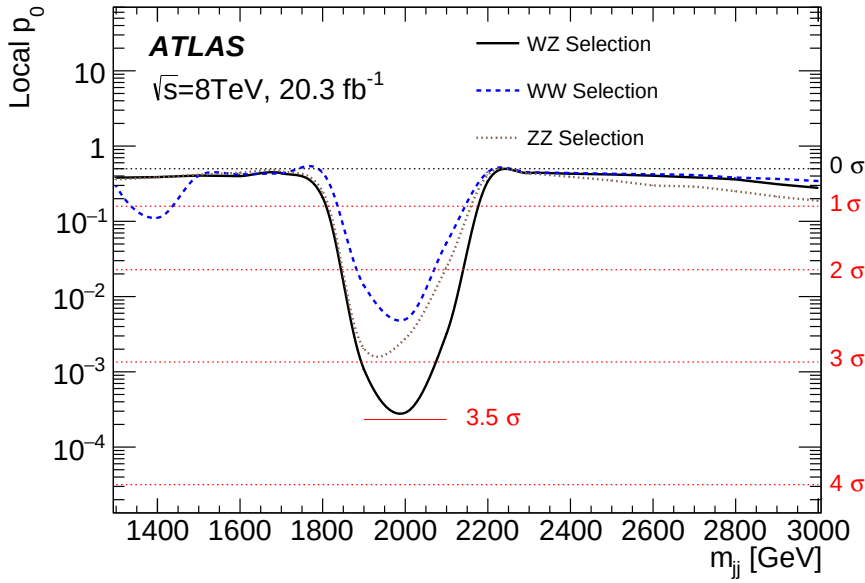


Figure 8.2: The local p_0 value and local significance for the WZ , WW and ZZ signal regions.

in data, corresponding to a local significance of 3.4σ . Taking into account that an excess can occur anywhere in the explored mass range, the global significance of the discrepancy in the WZ channel is 2.5σ . The smallest local p_0 values for the WW and ZZ signal regions correspond to a significance of 2.6σ and 2.9σ , respectively. The calculation of the local p_0 values does not take into account that the three signal regions are not independent whereas this effect is considered in the calculation of the global significance.

8.1.1 Cross Checks

Many studies were performed to ensure that no systematic effect caused a shaping of the background in the high invariant dijet mass region to produce the excess of events observed around $m_{JJ} = 2$ TeV. A small selection of these cross-checks will be summarised here.

To assure that the events selected in the high mass region are not accumulated in one particular region of the detector, which might suffer from dead modules or miscalibration, the pseudorapidity of the leading and subleading jet is depicted separately in Fig. 8.3 as a function of the invariant dijet mass. All events that satisfy the boson tagging criteria of any of the three signal regions are displayed. Furthermore it was shown that the event selection efficiency was stable over the whole data-taking period and that events in the high mass region were not selected in one particular run.

To exclude the possibility that one of the boson tagging criteria caused the distortion of the invariant dijet mass spectrum, the m_{JJ} distribution was analysed by varying the applied boson tagging criteria. Two separate tests were performed to study the effects of the boson tagging

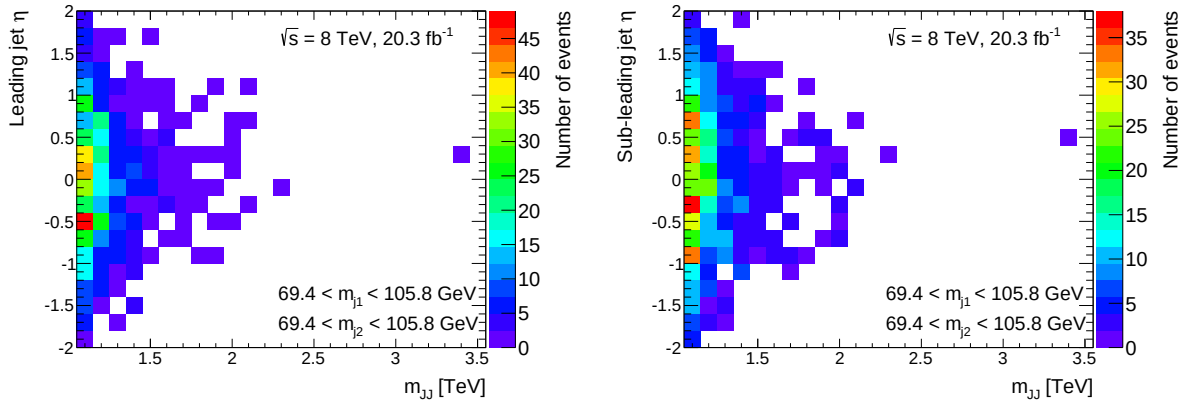


Figure 8.3: Distribution of the pseudorapidity of the leading (left) and subleading (right) jet as a function of the invariant dijet mass in a region in $\sqrt{s} = 8$ TeV data covering the three signal regions.

criteria on the m_{JJ} distributions in data: firstly one of the three boson tagging requirements was removed and secondly only one boson tagging requirement was applied at a time. The resulting invariant dijet mass spectra are shown in Fig. 8.4 for the three signal regions and compared to the untagged m_{JJ} distributions and to the distributions with all boson tagging requirements applied. If either the criterion on the jet mass window or on $\sqrt{y_f}$ is removed, the invariant dijet mass still follows a smoothly falling distribution. Only the combination of the mass window and the $\sqrt{y_f} > 0.45$ cut results in the excess of events around 2 TeV. If on the other hand, only one boson tagging criterion is applied at a time, each of the m_{JJ} spectra is a smoothly falling distribution and none of the cuts produce a distortion in the spectrum.

Furthermore, the choice of jet reconstruction and grooming algorithm was considered as one possible source of creating a distortion in the invariant mass spectrum as the usage of substructure techniques was never tested in this high mass region before and the MC statistics in this regime was limited. However, at the time of the analysis, no other algorithm than the BDRS algorithm was optimised to efficiently identify high- p_T vector bosons. When using a non-optimal W/Z identification algorithm, no significant excess was observed in the 8 TeV data. After Run-I, the W/Z boson identification was re-optimised for 2015 analyses. A new algorithm was developed, as described in Section 6.2, based on anti- k_t $R = 1.0$ trimmed jets and with criteria on the jet mass and the substructure variable $D_2^{\beta=1}$. Dedicated energy and mass calibrations were derived for this algorithm at $\sqrt{s} = 8$ TeV to be able to analyse the 2012 data again with the new vector boson identification algorithm. Apart from the change of jet collection, the leading jet p_T threshold was lowered to 450 GeV, allowing to select events with $m_{JJ} > 1.0$ TeV instead of $m_{JJ} > 1.05$ TeV as for the default 8 TeV analysis. The resulting m_{JJ} distributions for the three signal regions (WW , WZ and ZZ) are depicted in Fig. 8.5 with and without the requirement on n_{trk} . Comparing Fig. 8.5 with the results obtained with the BDRS algorithm in Fig. 8.1, it

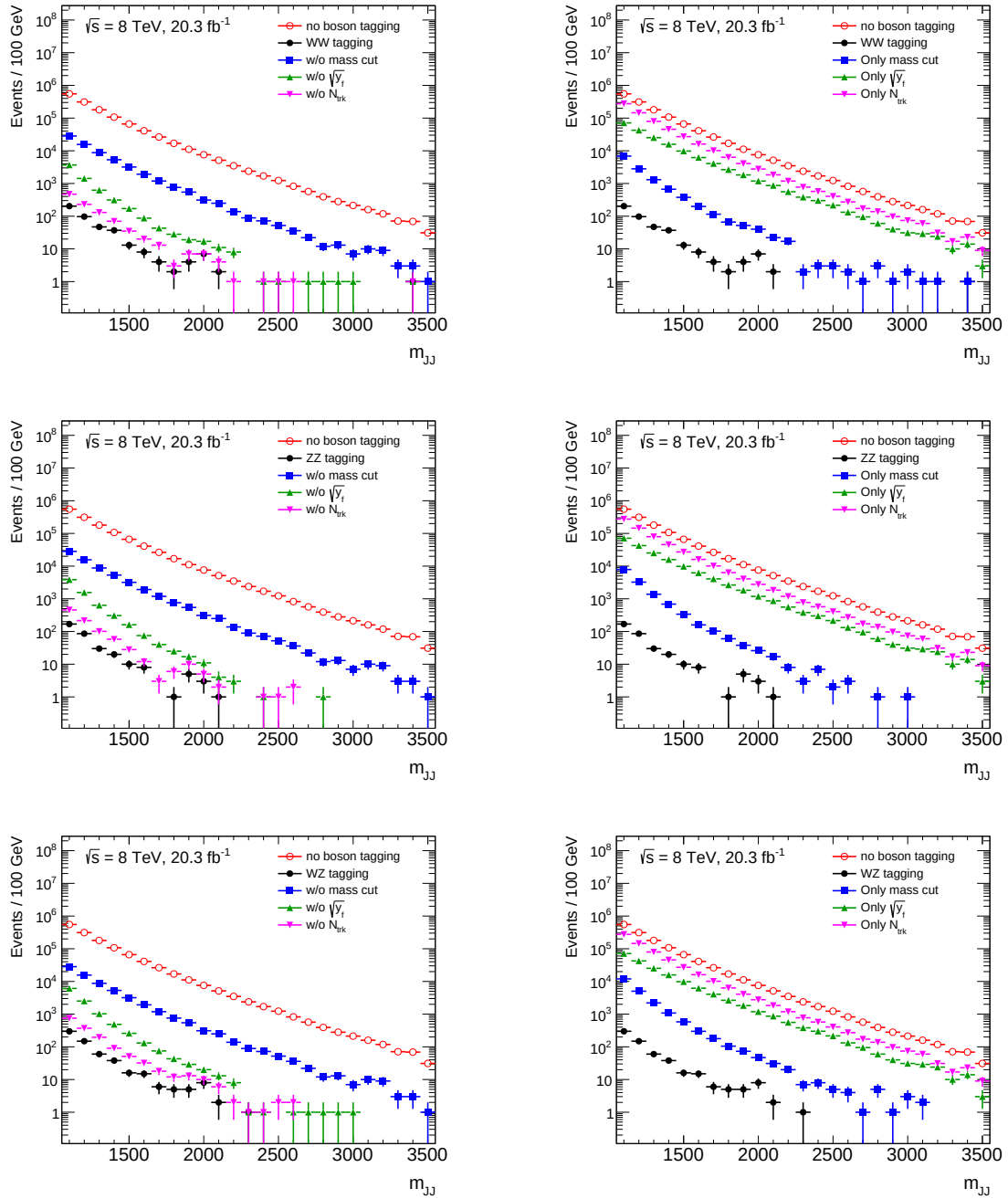


Figure 8.4: Effect on the invariant dijet mass spectrum in $\sqrt{s} = 8$ TeV data when varying the boson tagging criteria in the WW (top), ZZ (middle) and WZ (bottom) signal region. Left: the untagged m_{JJ} distribution and the spectrum with all boson tagging criteria is compared to the m_{JJ} distributions with only one boson tagging requirement applied to each jet at a time. Right: the untagged m_{JJ} distribution and the spectrum with all boson tagging criteria is compared to the m_{JJ} distributions when one boson tagging requirement is removed.

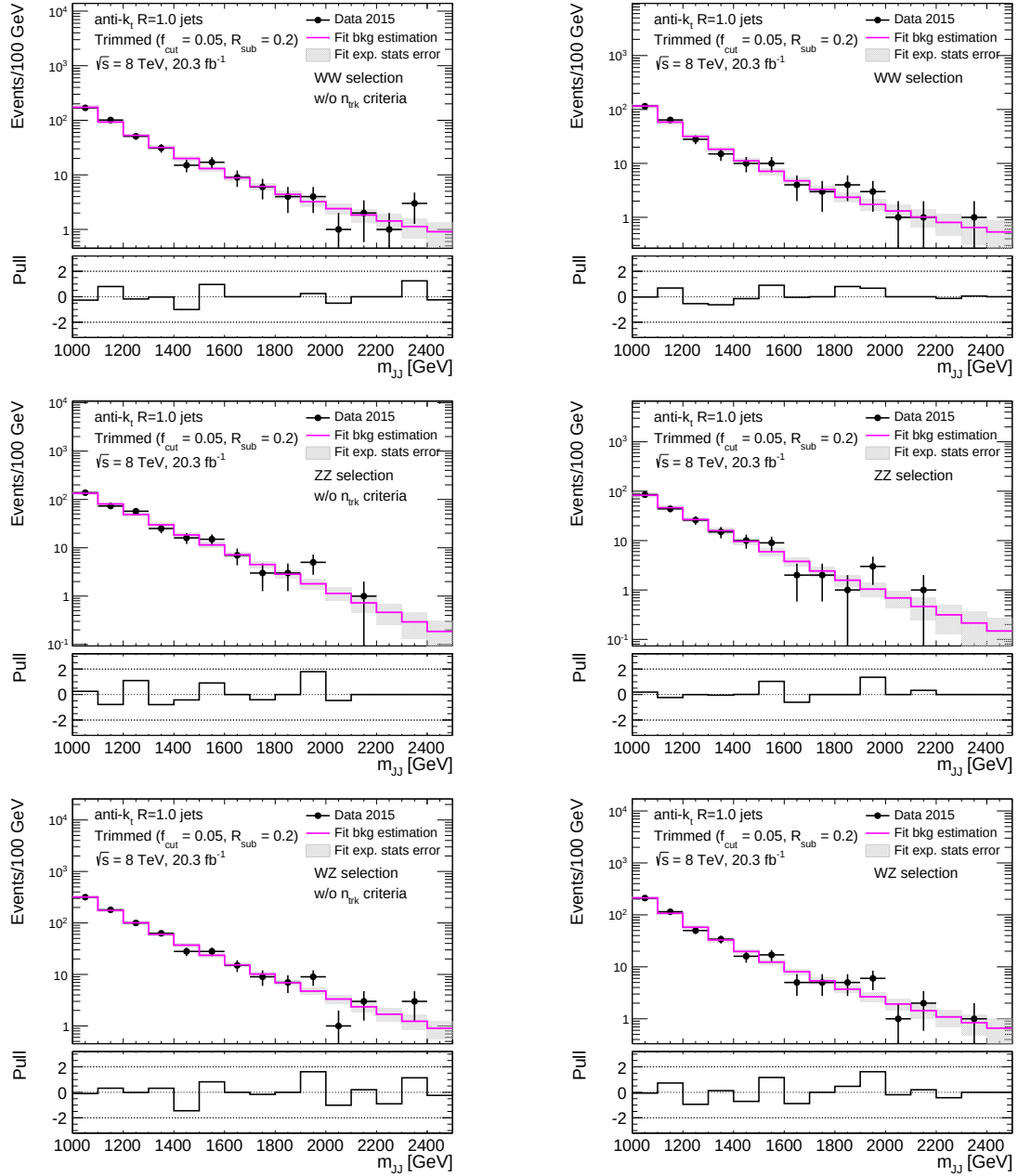


Figure 8.5: Background only-fit to the invariant dijet mass (m_{JJ}) distribution in $\sqrt{s} = 8$ TeV data for boson-tagged jets using the Run-II boson tagging algorithm described in Section 7.3.1 in the three channels WW (top), ZZ (middle) and WZ (bottom) without (left) and with (right) the requirement $n_{\text{trk}} < 30$.

can be seen that the Run-II tagging algorithm has a higher background suppression than the algorithm used in Run-I. A slight excess of events can be observed for the WZ and ZZ selection with the Run-II tagging algorithm, however the significance is significantly smaller than the one obtained with the Run-I boson tagging algorithm.

8.2 Statistical Interpretation

Upper limits at 95% confidence level are set on the production cross-section times branching ratio of the EGM W' benchmark model in the WZ channel and the RS graviton model in the WW and ZZ channels, following the description of the CL_s technique in Section 7.7. The expected and observed upper limits on the cross-section times branching ratio of the two benchmark models as a function of the resonance mass are shown in Fig. 8.6. Even though events are selected with $m_{JJ} > 1.05$ TeV, limits are only set for resonances masses greater than 1.3 TeV to ensure that the invariant dijet mass distribution of the signal samples is fully contained in the considered mass range. The excess of observed events in data around $m_{JJ} = 2$ TeV results in weaker observed upper limits than expected as opposed to most of the remaining mass range in which the observed limits are more stringent than the expected limits. EGM $W' \rightarrow WZ$ resonances with masses between 1.3 and 1.5 TeV are excluded at 95% confidence level, whereas no exclusion was expected in the considered mass range. The analysis is not sensitive enough to exclude gravitons in the RS bulk model with $k/\bar{M}_{\text{Pl}} = 1.0$ as the predicted cross-sections times branching ratios are too low.

8.3 Combination with Other Diboson Resonance Searches

To improve the sensitivity to new diboson resonances and to provide more stringent constraints on the production cross-section times branching ratio of the two considered benchmark models, four diboson searches, differing in the final state, are combined. All analyses use the full 2012 dataset, corresponding to an integrated luminosity of 20.3 fb^{-1} . In the search for charged diboson resonances, $W' \rightarrow WZ$, the $\ell\nu\ell'\ell'$ [160], $\ell\ell qq$ [88], $\ell\nu qq$ [89] ($\ell = e, \mu$) and the fully hadronic final state [87] are considered whereas only the last three final states are considered in the search for neutral RS gravitons decaying to WW or ZZ . The four different analyses are statistically independent due to the orthogonal event selection but the correlation of the systematic uncertainties in the different channels are considered in the statistical interpretation. The results are interpreted using the statistical techniques described in Section 7.7. To optimise the sensitivity to new diboson resonances over the full mass range, the $\ell\nu\ell'\ell'$, $\ell\ell qq$, $\ell\nu qq$ searches define different signal regions with optimised selection criteria, referred to as low- p_T resolved (low mass) and high- p_T resolved (high mass). For the $\ell\ell qq$ and $\ell\nu qq$ analyses an additional high- p_T merged signal region is defined in which the hadronic decay of the vector boson is reconstructed as one large- R jet as opposed to the resolved regions in which two small- R jets with a radius parameter of 0.4 are used to reconstruct the vector boson. The background in the diboson resonance

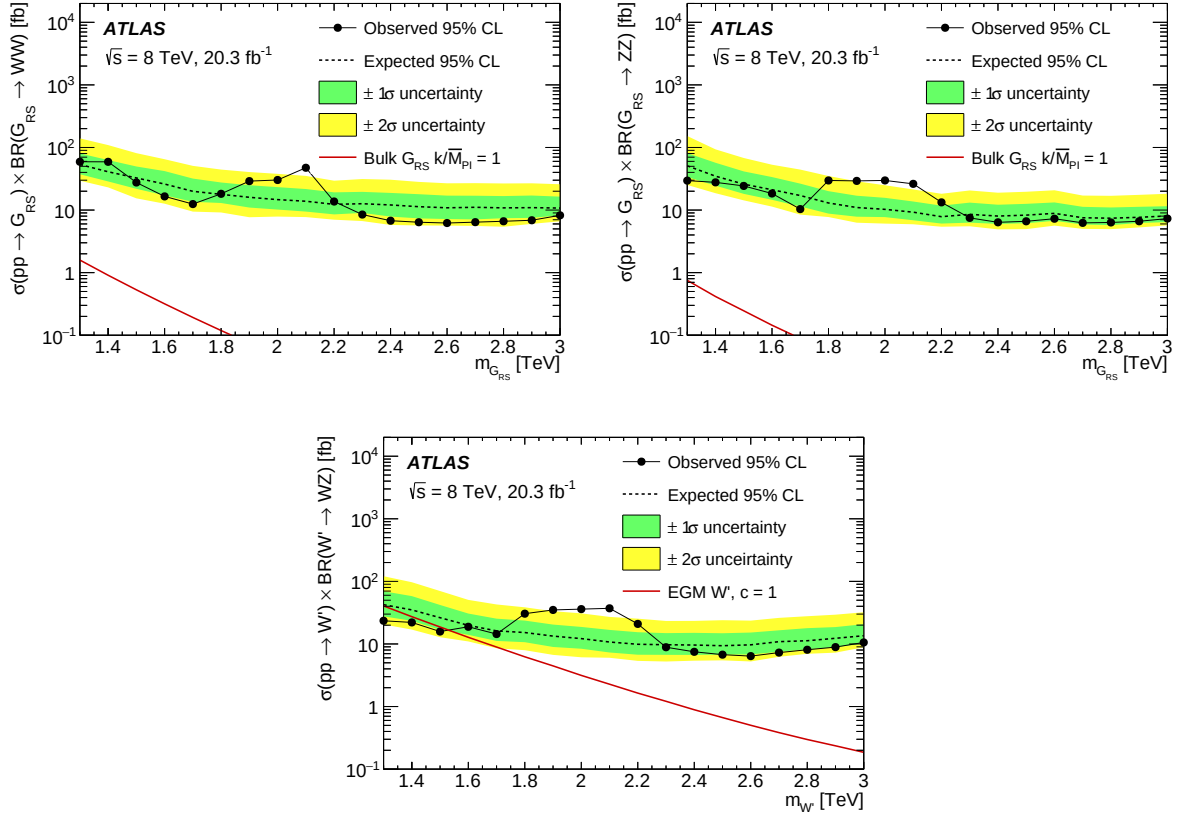


Figure 8.6: The 95% CL upper limits on the cross-section times branching ratio of the RS graviton model in the WW (top left) and ZZ (top right) signal region and on the EGM W' model (bottom) in the WZ signal region. The green and yellow bands represent the 1σ and 2σ interval of the expected limit, respectively and the red line displays the predicted cross-section of the considered models.

searches with at least one leptonically decaying vector boson is estimated from MC simulation but in the case of the $\ell\ell qq$ and $\ell\nu qq$ analyses, control regions in data are used to correct the normalisation and shape of the dominating Z +jets and W/Z +jets background, respectively.

The acceptance times efficiency of the analyses contributing to the search for neutral and charged diboson resonance are shown in Fig. 8.7 as a function of the resonance mass. In the search for a charged diboson resonance W' , the $\ell\nu qq$ analysis has the highest acceptance times efficiency for resonance masses below 1.3 TeV, whereas for $m_{W'} > 1.3$ TeV, the fully hadronic search has the best sensitivity. In the search for $G_{RS} \rightarrow WW$, the $\ell\nu qq$ analysis has the highest acceptance times efficiency over the full considered mass range. The combined acceptance times efficiency is up to 11% for the $G_{RS} \rightarrow ZZ$ search, and up to 17% for the $G_{RS} \rightarrow WW$ and $W' \rightarrow WZ$ search.

The p_0 value, obtained in the search for the G_{RS} and EGM W' are shown separately in Fig. 8.8

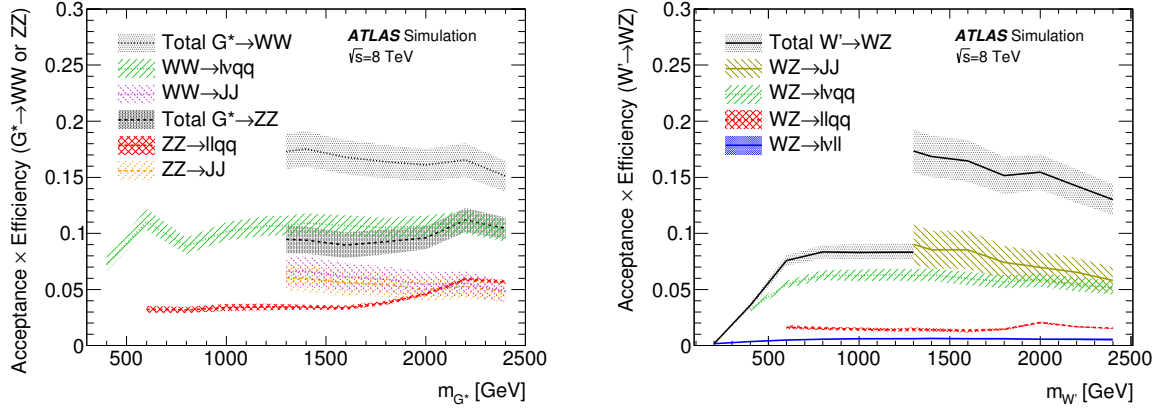


Figure 8.7: Signal acceptance times efficiency as a function of the resonance mass for the RS graviton (left) and EGM W' (right) model for the different analyses used for the combination. The error bands represent the statistical and systematic uncertainty.

as a function of the resonance mass for the individual analysis as well as for the combined diboson resonance search. The smallest local p_0 value is observed in the search for the EGM W' at a mass of $m_{W'} = 2.0$ TeV, corresponding to a significance of 2.5σ , compared to a significance of 3.4σ that was obtained in the fully hadronic search. The significance at $m_{W'} = 2.0$ TeV decreased for the combination as the data is well-described by the MC background prediction in the other diboson resonance searches.

The 95% CL upper limits on the production cross-section times branching ratio of the RS graviton model in the WW and ZZ channel are shown in Fig. 8.9 for the individual channels as well as the combined diboson resonance search. The combined upper cross-section limit is furthermore compared to the theoretical predictions of the graviton model in Fig. 8.9. The combined diboson resonance search excludes gravitons with masses below 810 GeV at 95% CL in a model with $k/\bar{M}_{\text{Pl}} = 1.0$, where a lower limit of 790 GeV was expected. The combination thus exceeds the lower mass limit of the most stringent individual analysis in the $\ell\nu qq$ final state by 50 GeV.

The 95% CL upper limits on the production cross-section times branching ratio of the individual channels contributing to the $W' \rightarrow WZ$ search as predicted in the EGM benchmark model are compared to the combined diboson resonance search in Fig. 8.10. Furthermore the comparison of the upper cross-section limit of the combined search to the theoretical prediction is shown. In the WZ channel, the combined search excludes EGM W' with masses below 1.81 TeV at 95% CL, with an expected limit of 1.81 TeV. The combination of the different diboson resonance searches enabled to expand the conclusion of W' masses, compared to the $\ell\nu qq$ analysis which set the most stringent individual lower mass limit of 1.59 TeV.

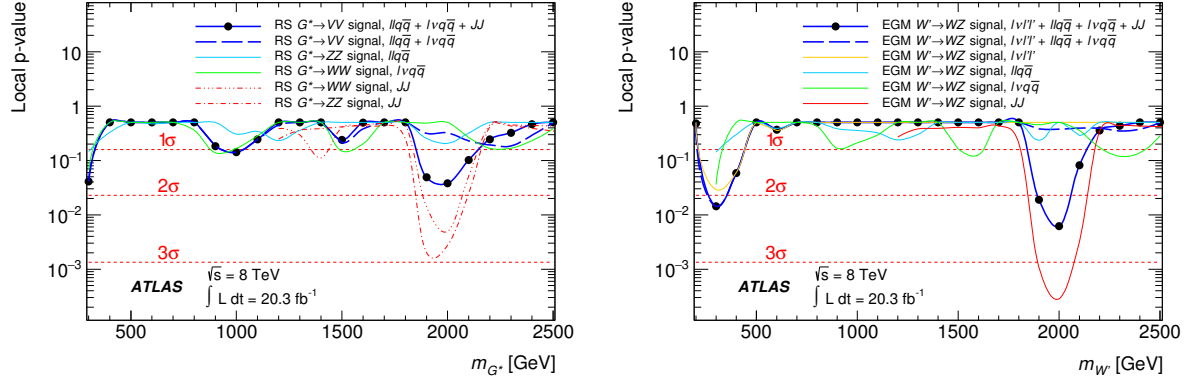


Figure 8.8: The local p_0 value, obtained in the search for the G_{RS} (left) and EGM W' (right), as a function of the resonance mass for the individual searches and the combined diboson resonance search. Furthermore, the local significance is indicated by the red dashed horizontal lines.

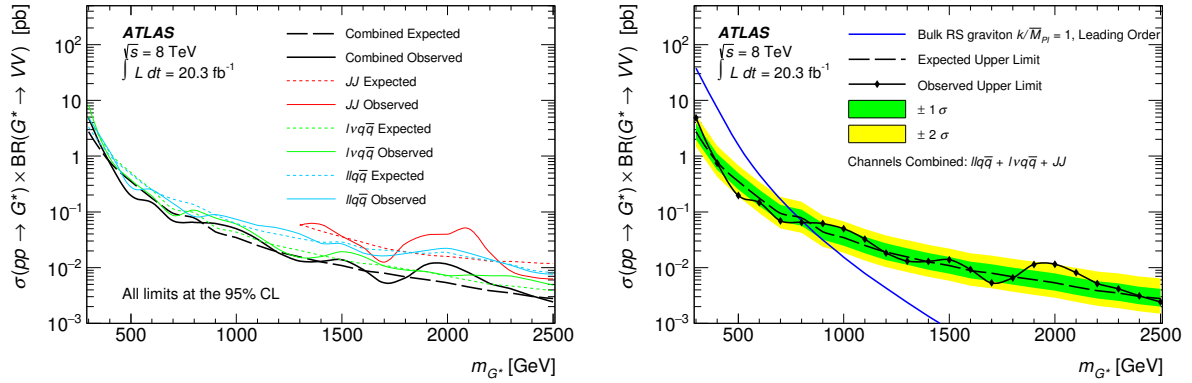


Figure 8.9: The expected and observed upper limits on the cross-section times branching ratio of the RS graviton model at 95% CL for the individual $\ell\ell qq$, $\ell\nu qq$ and $qqqq$ channels and their combination (left), and the comparison of the theory prediction to the 95% CL upper cross-section limits of the combined search (right). The green and yellow bands on the right represent the 1σ and 2σ interval of the expected limit, respectively and the blue line displays the predicted cross-section of the graviton model.

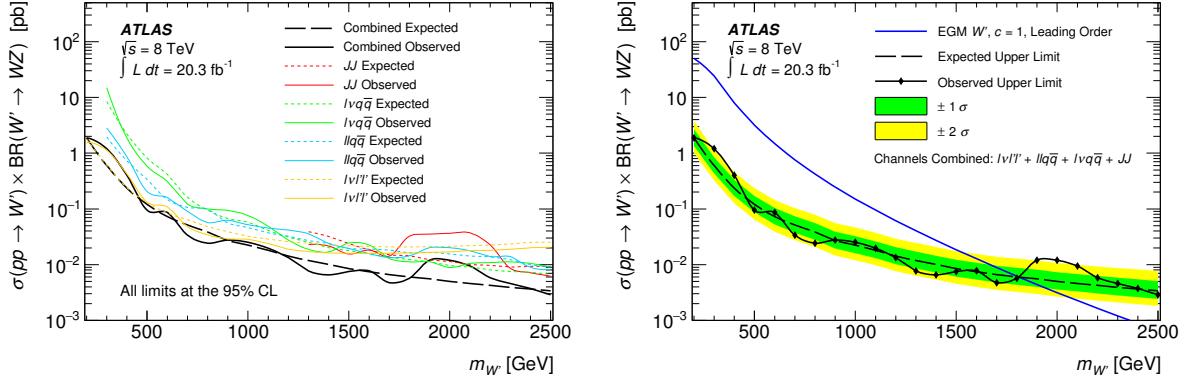


Figure 8.10: The expected and observed upper limits on the cross-section times branching ratio of the EGM W' model at 95% CL for the individual $l\nu l l$, $llqq$, $l\nu qq$ and $qqqq$ channels and their combination (left), and the comparison of the theory prediction to the 95% CL upper cross-section limits of the combined search (right). The green and yellow bands on the right represent the 1σ and 2σ interval of the expected limit, respectively and the blue line displays the predicted cross-section of the EGM model.

The CMS Collaboration published a similar analysis in the fully hadronic final state [161] at $\sqrt{s} = 8$ TeV with comparable sensitivity to the ATLAS analysis. Jets are reconstructed with the C/A algorithm and a radius parameter of $R = 0.8$. The pruning algorithm was used to remove pile-up contributions and criteria on the jet mass and the N -subjettiness variable τ_{21} are imposed to identify boosted hadronically decaying W and Z bosons. The largest deviation from the background model was found in the $G_{RS} \rightarrow WW$ selection at a mass of 1.9 TeV with a local significance of 1.3σ . In the absence of a significant deviation from the background hypothesis, the CMS Collaboration excluded Randall-Sundrum gravitons in a model with $k/\bar{M}_{\text{Pl}} = 0.1$ decaying into WW for masses up to 1.2 TeV at 95% CL. Furthermore, W' bosons decaying into WZ are excluded at 95% CL for masses below 1.7 TeV.

Chapter 9

Diboson Resonance Search at $\sqrt{s} = 13$ TeV

After the observation of an excess of events over the background prediction with a global significance of 2.5σ in the diboson resonance search at 8 TeV, it is of utmost importance to repeat the analysis with the first collected data at $\sqrt{s} = 13$ TeV to either enhance the significance of the excess or to verify that the excess was only a statistical fluctuation. The excess occurred at an invariant dijet mass of 2.0 TeV which was at the edge of the region that was statistically accessible in the 8 TeV analysis. Due to the increase of parton luminosity from 8 to 13 TeV, new resonances with high masses, initiated by gg , qg or $q\bar{q}$ production, are produced more abundantly. For instance, 3 fb^{-1} of 13 TeV data are needed to achieve the same sensitivity as with the 8 TeV dataset for a $q\bar{q}$ initiated production of a new particle with a mass of 2000 GeV.

This chapter is organised as follows. The results of the search for resonances decaying into pairs of vector bosons, using 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV [162] is presented in Section 9.1 and their statistical interpretation is discussed in Section 9.2. Furthermore, the combination with other diboson resonance searches [163] is presented in Section 9.3.

9.1 Results

The background-only fits to the invariant dijet mass distribution for events passing the event selection described Section 7.4 are shown in Fig. 9.1 for the WW , ZZ and WZ signal regions. The m_{JJ} distributions are well-described by the background parameterisations over the entire mass range explored, $1.0 < m_{JJ} < 2.5$ TeV, and no significant deviations are observed. The background-only fit yields a χ^2 over the degrees of freedom of 2.0/6 for the WW selection, 2.8/5 for the ZZ selection and 4.9/6 for the WZ selection. The analysis is restricted to $m_{JJ} < 2.5$ TeV due to limitations in the jet mass calibration procedure, however no events are observed beyond 2.5 TeV in any of the signal regions. A total of 95, 77 and 128 events are observed in the WW , ZZ and WZ signal region, respectively. Due to the overlapping size of the mass window around the W boson and Z boson mass, 38 events are shared among all channels. Among these 38 events is the event with the highest invariant dijet mass of $m_{JJ} = 2170$ GeV. The fraction of events selected by both the WZ and WW selection are 60% whereas 56% of the events are common to the WZ and ZZ selection

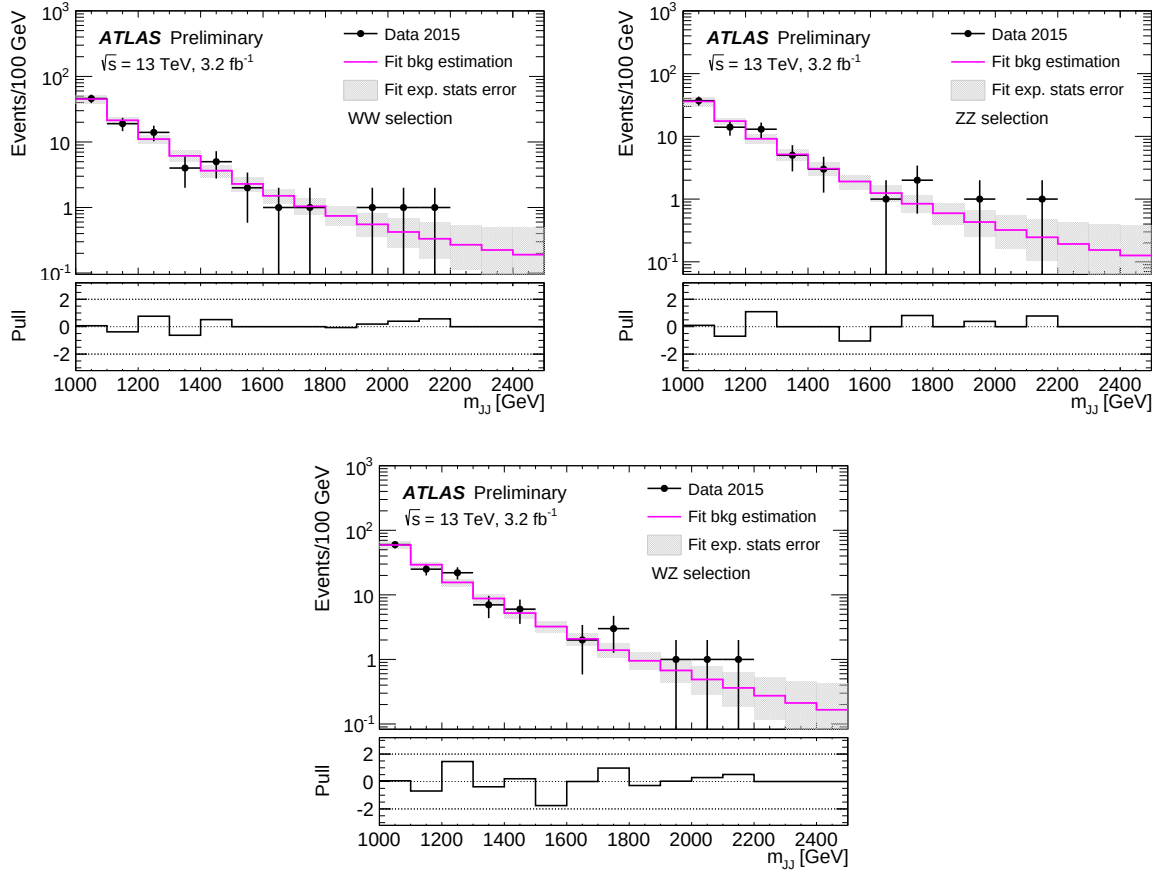


Figure 9.1: Background-only fit to the invariant dijet mass (m_{JJ}) distribution in $\sqrt{s} = 13$ TeV data for boson-tagged jets in the three different signal regions: WW (top left), ZZ (top right) and WZ (bottom).

An event display of an event with $m_{JJ} = 1275$ GeV, passing the WW , ZZ and WZ selection, is shown in Fig. 9.2. Both jets are balanced in p_T , with a leading jet p_T of 578 GeV and the subleading jet with $p_T = 564$ GeV. The masses of both jets are close to the Z boson and W boson mass with $m = 91.4$ GeV and $m = 81.7$ GeV for the leading and subleading jet, respectively. Furthermore, a display of an event with an invariant dijet mass of $m_{JJ} = 1760$ GeV is depicted in Fig. 9.3. The leading jet has $p_T = 769$ GeV and a mass of 72.6 GeV and the subleading jet is selected with $p_T = 778$ GeV and $m = 69.4$ GeV. Due to the relatively low jet masses compared to the mass of the Z boson, the jets do not fall in the 15 GeV mass window around the Z boson mass and therefore the event passes only the WW selection. Two dense cores of energy deposits, approximately balanced in p_T , are discernible within the large radius of the leading and subleading jet as expected from the hadronic decay of a vector boson.

As for the Run-I analysis, the effect of the different boson tagging criteria on the invariant dijet mass spectrum was studied in data to ensure that no requirement causes a distortion of the m_{JJ}

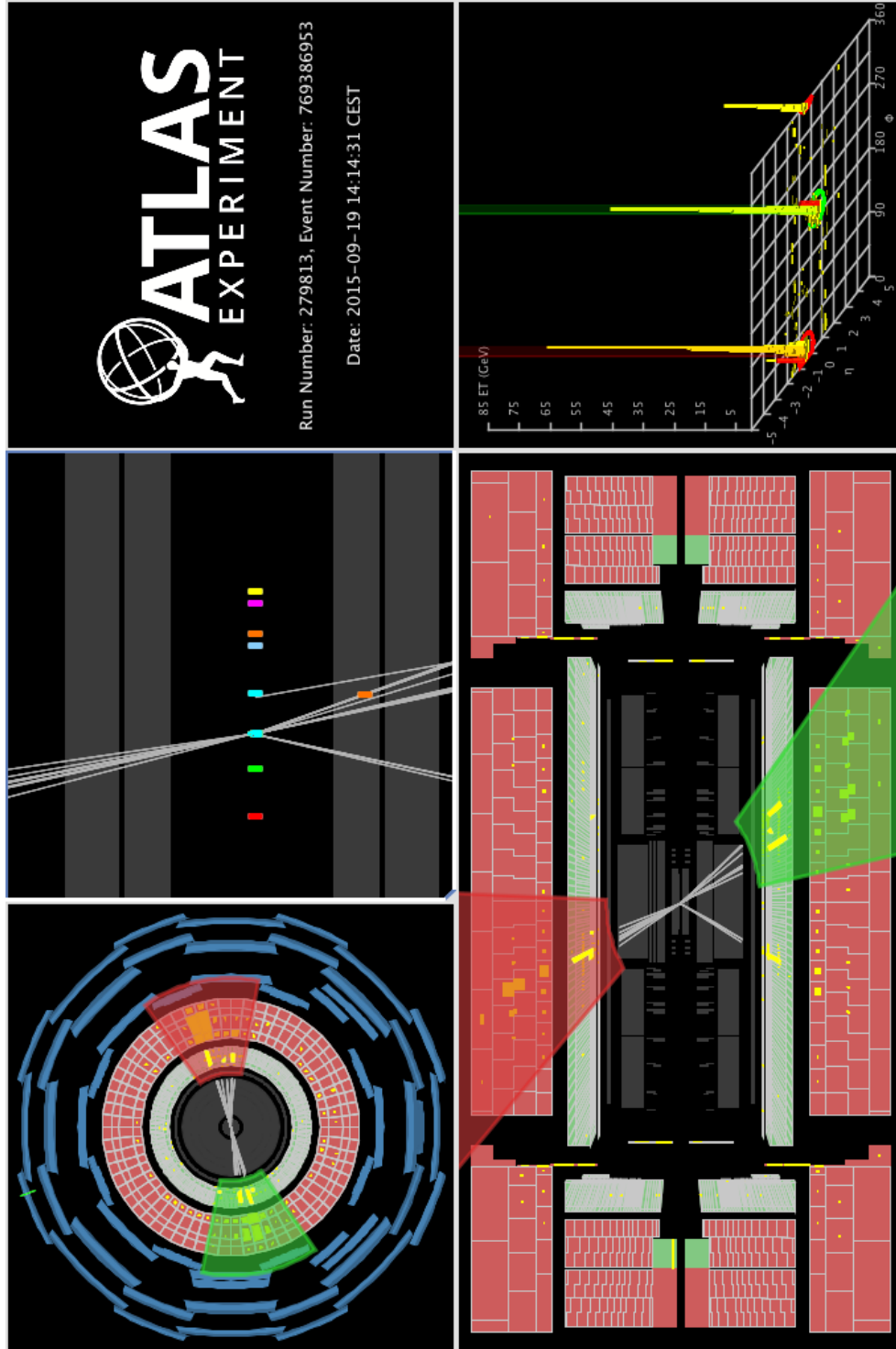


Figure 9.2: Event display of an event passing the WW , ZZ and WZ boson tagging requirements with an invariant dijet mass of $m_{JJ} = 1275$ GeV. The leading jet (red cone) has $p_T = 578$ GeV and a mass of 91.4 GeV and the subleading jet (green cone) has $p_T = 564$ GeV and $m = 81.7$ GeV. The top (bottom) left panel shows the transverse (longitudinal) view of the ATLAS detector. The bottom right panel shows the deposited energy in the calorimeter in the $\eta \times \phi$ space while the coloured points in the top middle panel represent the reconstructed hard-scatter and pile-up vertices in the event. Charged-particle tracks in the Inner Detector are displayed as grey lines and energy deposits in the calorimeters are displayed in yellow.

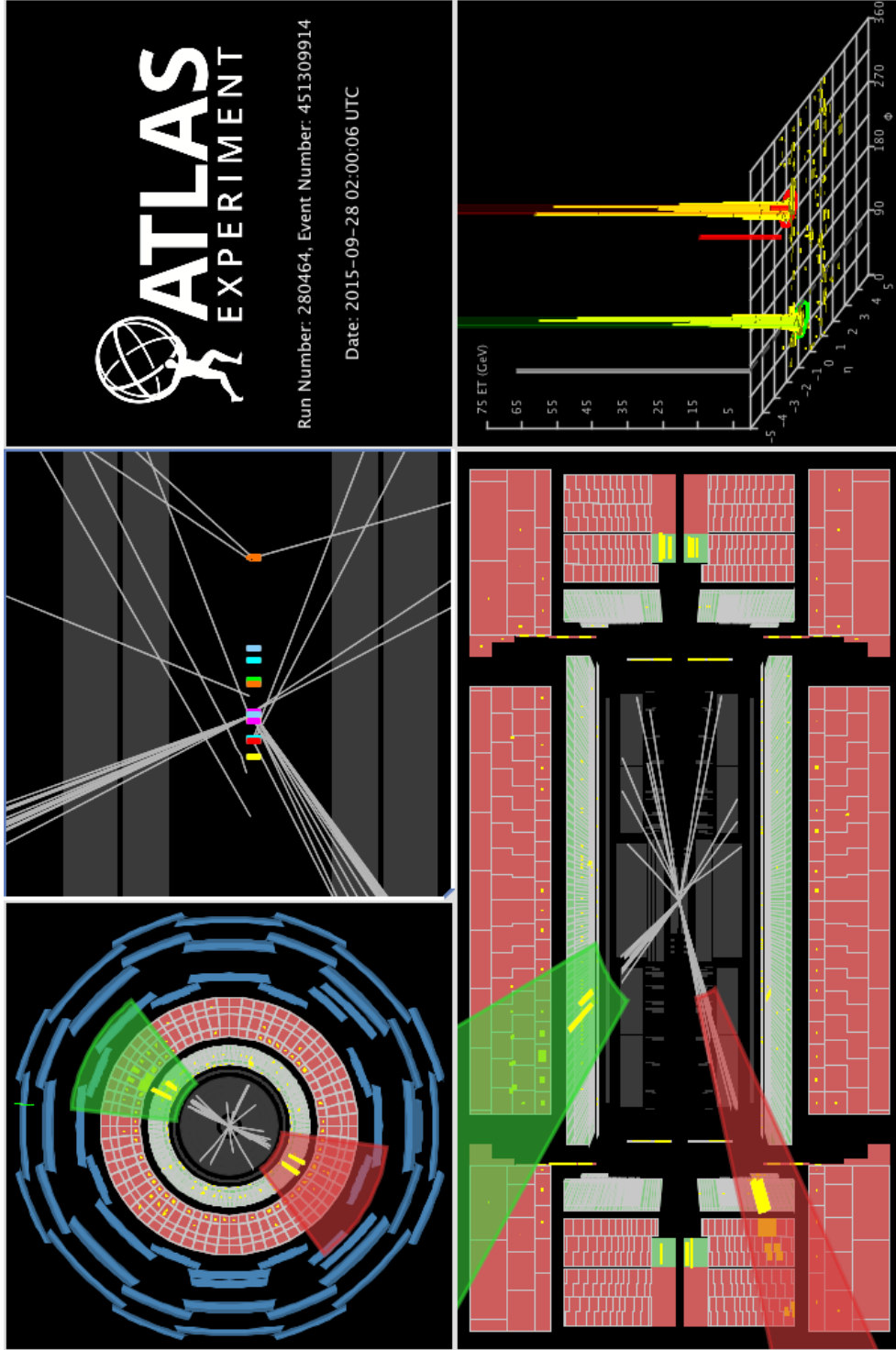


Figure 9.3: Event display of an event passing the WW boson tagging requirements with an invariant dijet mass of $m_{JJ} = 1760$ GeV. The leading jet (red cone) has $p_T = 769$ GeV and a mass of 72.6 GeV and the subleading jet (green cone) has $p_T = 778$ GeV and $m = 69.4$ GeV. The top (bottom) left panel shows the transverse (longitudinal) view of the ATLAS detector. The bottom right panel shows the deposited energy in the calorimeter in the $\eta \times \phi$ space while the coloured points in the top middle panel represent the reconstructed hard-scatter and pile-up vertices in the event. Charged-particle tracks in the Inner Detector are displayed as grey lines and energy deposits in the calorimeters are displayed in yellow.

spectra. The invariant dijet mass spectra after removing one boson tagging requirement and with only one boson tagging requirement applied are shown in Fig. 9.4 for the three signal regions and are compared to the untagged m_{JJ} distributions and to the distributions with all boson tagging requirements applied. In the case where only one boson tagging criteria is applied at a time, the invariant dijet mass distribution is smoothly falling without any distortions and the requirement on the jet mass window results in the largest suppression of the QCD background. On the other hand, if one boson tagging criteria is removed, the m_{JJ} distribution exhibits small deviations from the smoothly falling background model at high invariant dijet masses around 1.9 TeV, especially visible for the distributions without the n_{trk} requirement in the ZZ and WZ channel. However, these deviations from the background-only hypothesis are not statistically significant.

The analysis at $\sqrt{s} = 13$ TeV was also performed with the boson tagging algorithm used in the Run-I search for which dedicated energy and mass calibrations were derived. As it can be seen in Fig. 7.15, the Run-I tagger has a lower background suppression at low invariant dijet masses, however in the high m_{JJ} region, the background suppression is approximately the same for the Run-I and Run-II tagger. Even though the background suppression is about the same in the high m_{JJ} region for both taggers, this cross check was performed to ensure that the Run-II tagger does not suppress a possible signal that would be selected by a different boson tagging algorithm. The background-only fits to the invariant dijet mass distribution obtained with the Run-I tagger are shown in Fig. 9.5 for the three signal regions. A total of 259, 199 and 470 events are selected in the WW , ZZ and WZ signal region, respectively. As for the Run-II tagger, the m_{JJ} distribution in data, obtained with the Run-I tagger, is in good agreement with the fitted background parameterisation. The largest deviation from background parameterisation is observed in the ZZ selection at a mass of 1.9 TeV, where one event is observed in data, however this deviation is not statistically significant.

9.2 Statistical Interpretation

Given that there are no significant deviations from the background-only hypothesis, upper limits are set on the production cross-section times branching ratio of the HVT benchmark model in the WZ and WW selection and the RS graviton model in the WW and ZZ channel. The statistical interpretation of the results follows the description in Section 7.7 and upper limits are set at 95% CL with the CL_s method.

The expected and observed upper limits on the cross-section times branching ratio of the HVT benchmark model in the WW and WZ selection are shown in Fig. 9.6 as a function of the resonance mass. The observed limits on the HVT W' and Z' model are weaker than the expected limits for $m_{W'} > 1.7$ TeV and $m_{Z'} > 1.9$ TeV, respectively, because of the three observed events with $m_{JJ} > 1.9$ TeV. The HVT W' model can be excluded at 95% CL for W' masses between 1.38 and 1.6 TeV.

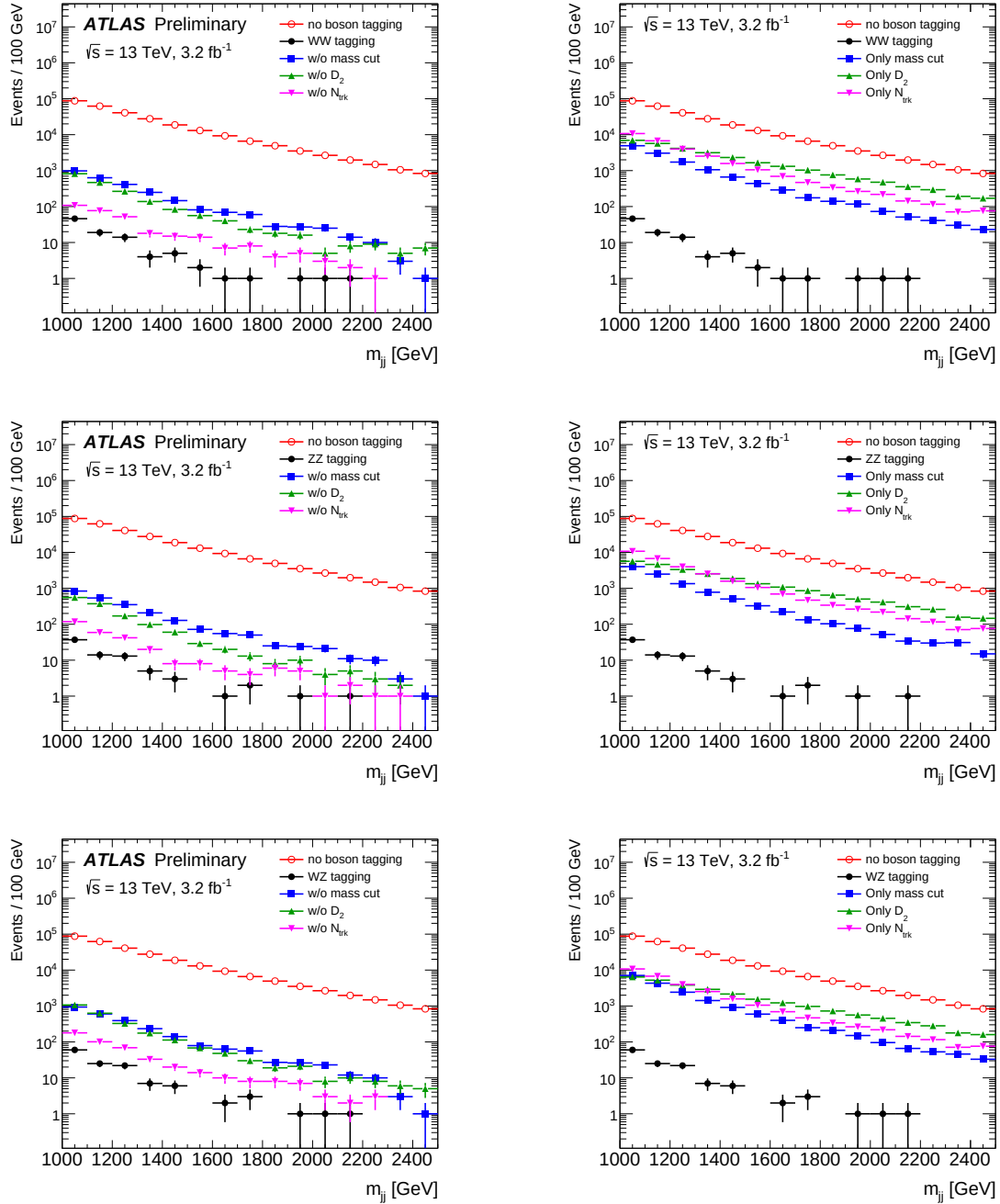


Figure 9.4: Effect on the invariant dijet mass spectrum in $\sqrt{s} = 13$ TeV data when varying the boson tagging criteria in the WW (top), ZZ (middle) and WZ (bottom) signal region. Left: the untagged m_{JJ} distribution and the spectrum with all boson tagging criteria is compared to the m_{JJ} distributions with only one boson tagging requirement applied to each jet at a time. Right: the untagged m_{JJ} distribution and the spectrum with all boson tagging criteria is compared to the m_{JJ} distributions when one boson tagging requirement is removed.

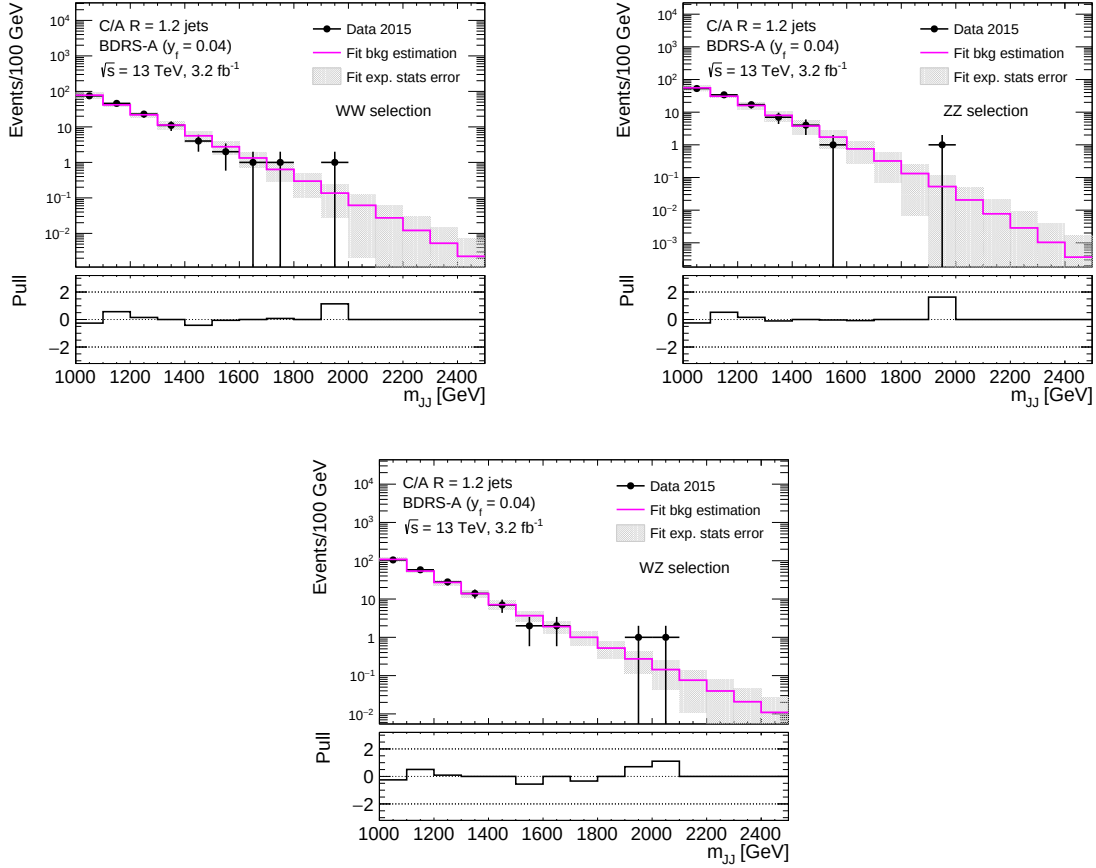


Figure 9.5: Background-only fit to the invariant dijet mass (m_{JJ}) distribution in $\sqrt{s} = 13$ TeV data for boson-tagged jets using the Run-I tagging algorithm in the three different signal regions: WW (top left), ZZ (top right) and WZ (bottom).

For the RS graviton, the expected and observed 95% CL upper limits on the cross-section times branching ratio are displayed in Fig. 9.7 as a function of the resonance mass for the WW and ZZ selection. The expected and observed limits on $\sigma(pp \rightarrow G_{\text{RS}} + X) \times \text{BR}(G_{\text{RS}} \rightarrow WW)$ are in good agreement up to $m_{G_{\text{RS}}}$ masses of 1.8 TeV and for higher masses, the observed limits are weaker than the expected limits. As for the Run-I analysis, no graviton masses can be excluded for the RS bulk model with $k/\bar{M}_{\text{Pl}} = 1.0$.

The Run-II analysis is not sensitive enough to exclude the excess of events observed at an invariant dijet mass of 2 TeV in the Run-I analysis. More statistics is needed to conclude on the observed excess. The combination of the fully hadronic diboson resonance search with analyses in the semi-leptonic final state will however exceed the sensitivity of the 8 TeV analysis especially at high resonance masses.

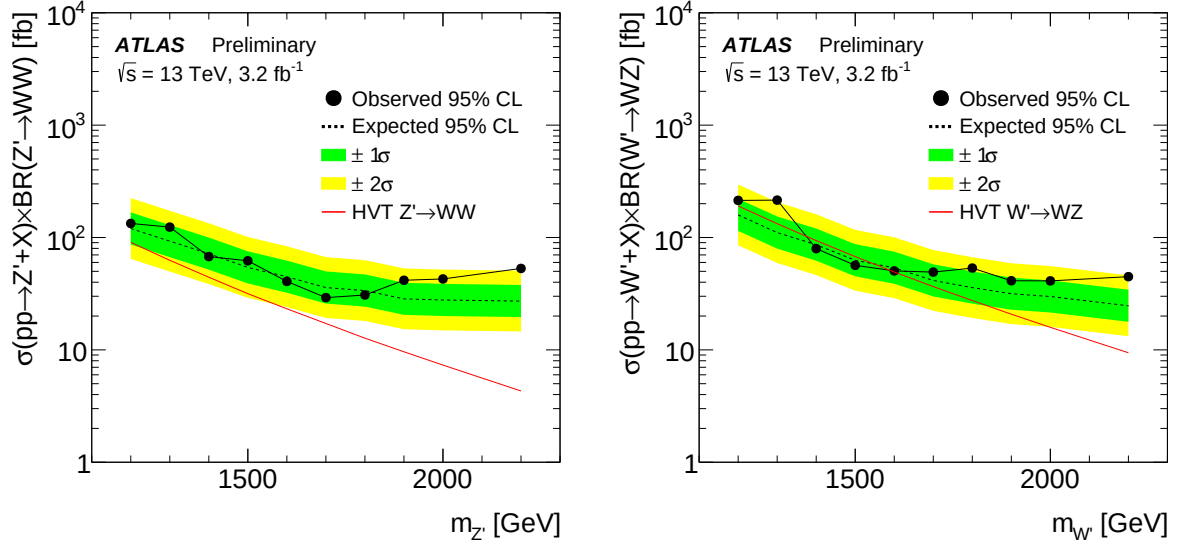


Figure 9.6: The 95% CL upper limits on the cross-section times branching ratio of the HVT model in the WW (left) and WZ (right). The green and yellow bands represent the 1σ and 2σ interval of the expected limit, respectively and the red line displays the predicted cross-section of the considered models.

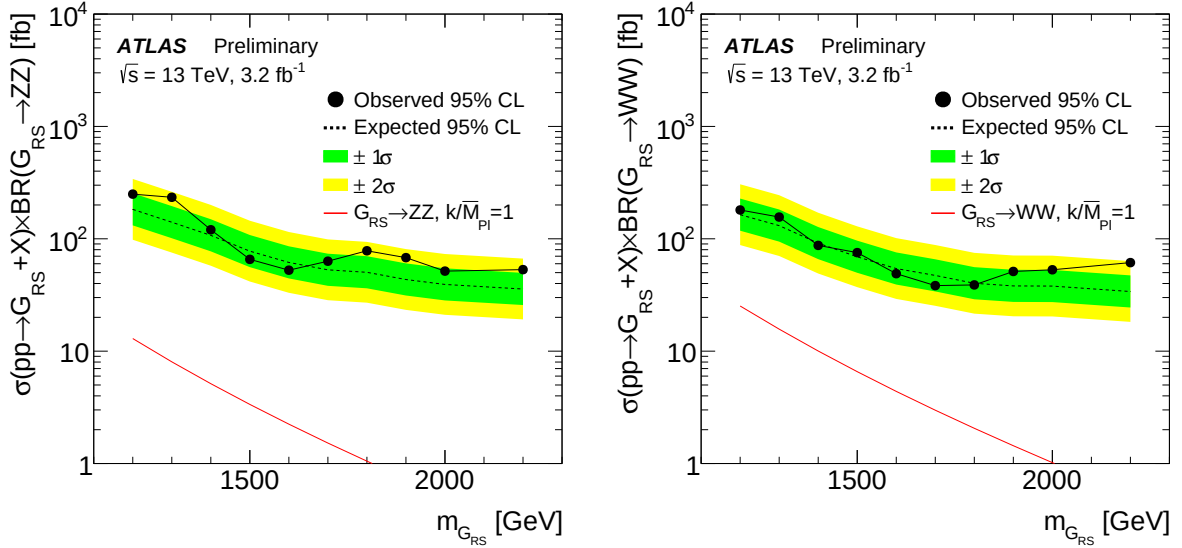


Figure 9.7: The 95% CL upper limits on the cross-section times branching ratio of the RS graviton model in the ZZ (left) and WW (right). The green and yellow bands represent the 1σ and 2σ interval of the expected limit, respectively and the red line displays the predicted cross-section of the considered models.

9.3 Combination with Other Diboson Resonance Searches

The results, presented in this chapter, are combined with three other analyses that search for diboson resonances in the semi-leptonic final state: $\nu\nu qq$ [143], $\ell\nu qq$ [141] and $\ell\ell qq$ [142]. The analyses are statistically independent due to the chosen event selection criteria. All analyses use the dataset collected in 2015, corresponding to an integrated luminosity of 3.2 fb^{-1} . Two benchmark models are considered: a simplified model predicting the existence of a heavy vector boson triplet and a bulk RS graviton model. They are used to search for a narrow resonance in data on top of the background prediction in either the invariant mass spectrum for the $\ell\nu qq$, $\ell\ell qq$ and fully hadronic analysis or the transverse mass spectrum for the $\nu\nu qq$ analysis. The considered mass range was extended with respect to the results presented in Section 9.1 due to the availability of data-driven systematic uncertainties as described in Section 5.2.4. The combined search for diboson resonances thus covers the range from 500 GeV up to 3000 GeV, in which the fully hadronic search contributes between 1.2 and 3.0 TeV. An upper limit of 3.0 TeV was chosen for the diboson combination because of the lack of Monte Carlo simulations of boosted vector bosons with $p_T > 2.0 \text{ TeV}$.

The acceptance times efficiency of the analyses contributing to the search for the RS graviton and the HVT model are shown in Fig. 9.8 as a function of the resonance mass. For RS graviton masses below 1.3 TeV, the $\nu\nu qq$ final state has the highest sensitivity whereas for $m_{G_{\text{RS}}} > 1.3 \text{ TeV}$, the $\ell\nu qq$ final state has the highest acceptance times efficiency. By including the fully hadronic diboson resonance search at $m_{G_{\text{RS}}} = 1.2 \text{ TeV}$, the combined acceptance times efficiency increases from 16% to 22%. In the search for $\text{HVT} \rightarrow WW/WZ$, the $\ell\nu qq$ final state has the highest acceptance times efficiency over the full considered mass range, however the sensitivity of the fully hadronic search is approximately the same for $m_{W'/Z'} > 1.2 \text{ TeV}$. The combined acceptance times efficiency for W'/Z' is up to 19% in the high mass region.

For all analyses contributing to the combination, the data is in very good agreement with the background prediction and no significant deviations are observed. The results are therefore used to set upper limits on the production cross-section times branching ratio of the two benchmark models, following the statistical analysis in Section 7.7. The dominating backgrounds in the semi-leptonic diboson resonance search are the production of Z +jets, W +jets and $t\bar{t}$ events whose modelling is constrained in dedicated control regions in data. These control regions are included in the combined profile likelihood fit to constrain the normalisation of the Z +jets, W +jets and $t\bar{t}$ background processes in the signal regions. The 95% CL upper limits on the production cross-section times branching ratio of the RS graviton model and the HVT benchmark model are depicted in Fig. 9.9 for the combined diboson resonance search and compared to the theoretical prediction. The combined diboson resonance search excludes RS gravitons, decaying into WW or ZZ , with masses below 1100 GeV at 95% CL in a model with $k/\bar{M}_{\text{Pl}} = 1.0$. In the WW or WZ channel, the combined search excludes heavy vector triplets ($g_V = 3$) with

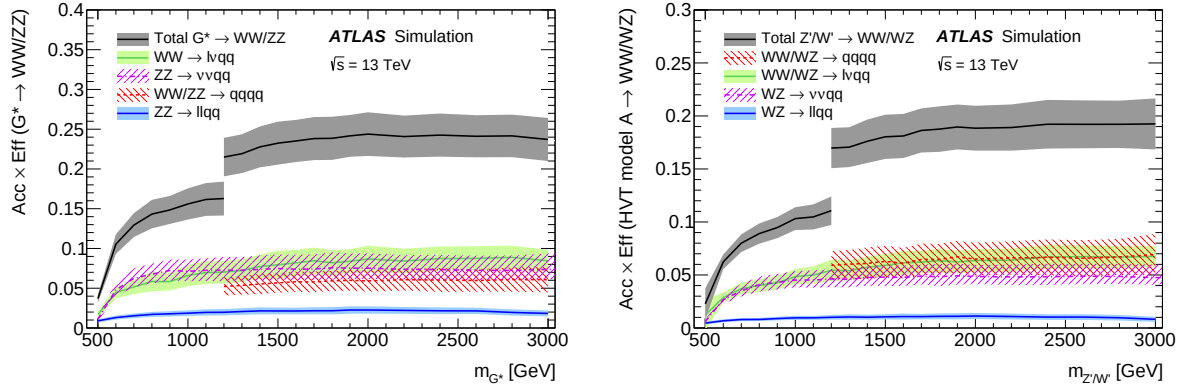


Figure 9.8: Signal acceptance times efficiency as a function of the resonance mass for the RS graviton (left) and HVT W' and Z' (right) model for the different analyses used for the combination. The error bands represent the statistical and systematic uncertainty.

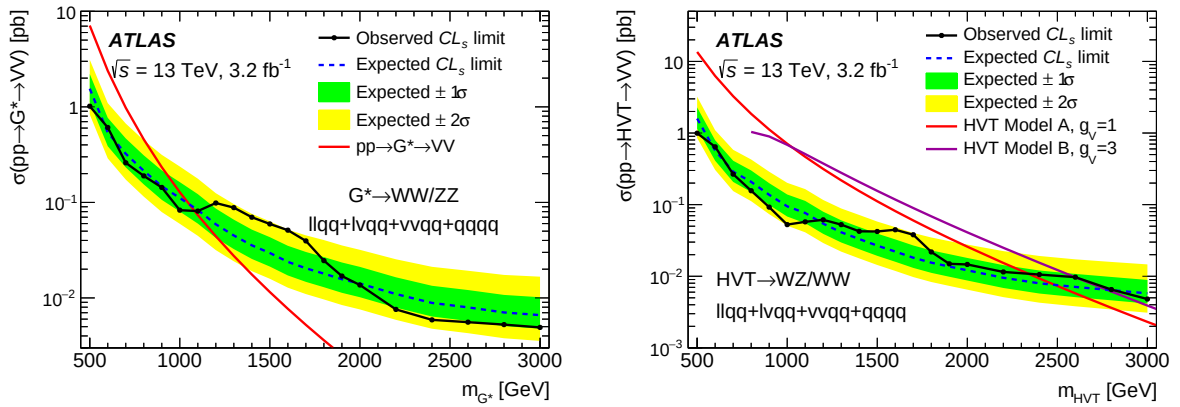


Figure 9.9: The 95% CL upper limits on the cross-section times branching ratio of the RS graviton model (left) using the combination of the $\ell\ell q\bar{q}$, $\ell\nu q\bar{q}$, $\nu\nu q\bar{q}$ and $qq\bar{q}\bar{q}$ channels and on the HVT W' and Z' model (right). The green and yellow bands represent the 1σ and 2σ interval of the expected limit, respectively. Two HVT models with different coupling strengths to the new vector boson, $g_V = 1$ and $g_V = 3$, are considered.

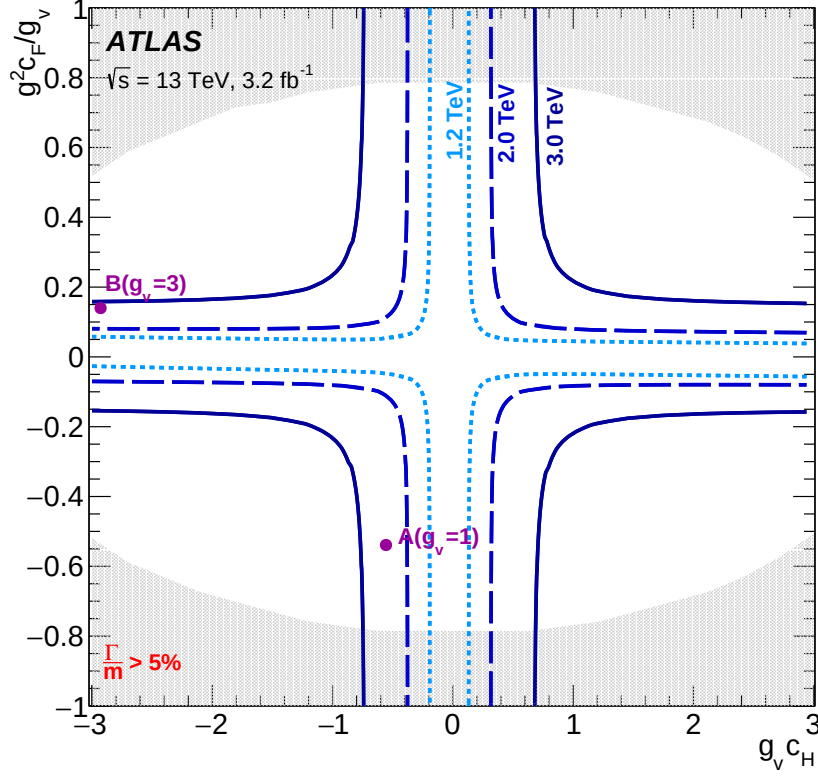


Figure 9.10: Observed 95% CL exclusion contours in the HVT parameter space ($g^2 c_F / g_V$, $g_V c_H$) for three different resonance masses: 1.2 TeV (short-dashed lines), 2.0 TeV (long-dashed lines) and 3.0 TeV (solid line). The parameters of the two HVT benchmark models A and B are also displayed in purple. The grey area indicates the parameter space in which the ratio of the resonance's total width Γ to its mass m is larger than 5%, which is comparable to the experimental mass resolution. Points inside the ellipse formed by the grey area but with $g^2 c_F / g_V$ and $g_V c_H$ values larger than the exclusion contours are excluded at 95% CL.

masses below 2.6 TeV at 95% CL. Furthermore, upper cross-section limits are set on the HVT model A with $g_V = 1$ and masses below 2.45 TeV are excluded. This model was chosen due to its similarities with the EGM W' model, used in the Run-I analyses, but does not exhibit the low mass tails of the EGM W' model. The mass limits obtained with 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13 \text{ TeV}$ exceed the limits obtained in the Run-I combination significantly. The cross-section limits for three different resonance masses (1.2, 2.0 and 3.0 TeV) are also interpreted in terms of exclusion contours in the HVT parameter space ($g^2 c_F / g_V$, $g_V c_H$) as shown in Fig. 9.10. For instance, if the parameters ($g^2 c_F / g_V$, $g_V c_H$) of a certain model, lie outside the contours of the 3.0 TeV resonance mass but inside the grey ellipse, then a new heavy vector triplet with masses below 3.0 TeV is excluded at 95% CL for this specific model.

9.4 Improvements for Future Diboson Resonance Searches

In the summer of 2016, the fully hadronic diboson resonance search was performed with a dataset corresponding to 15.5 fb^{-1} of pp collisions collected at $\sqrt{s} = 13$ TeV between August 2015 and July 2016 [164]. The obtained limits on the HVT model significantly extend the previous limits obtained with the 2015 dataset. Z' (W') bosons, predicted in the heavy vector triplet model A, are excluded with masses in the range $1.2 - 1.8$ TeV ($1.2 - 1.9$ TeV) at 95% CL, with an expected mass range exclusion of $1.2 - 1.9$ TeV ($1.2 - 2.3$ TeV).

Until the next scheduled extended break in the LHC data-taking program at the end of 2018, the LHC is expected to collect data corresponding to a total integrated luminosity of about 100 fb^{-1} which will allow for a significant extension of the mass reach of the fully hadronic diboson resonance search. However the analysis cannot only enhance its sensitivity to new resonances by using larger datasets but also by improving the identification of boosted W and Z bosons.

The importance of using information from the Inner Detector for the reconstruction of jets with $p_T > 1500$ GeV, which will be frequently produced in the next years, was already highlighted in Section 6.3. Track-assisted calculations of the jet mass and substructure variables will not only increase the discrimination of QCD-jets from W/Z -jets but also enable to distinguish W -jets from Z -jets due to the improvement in the jet mass resolution.

To increase the sensitivity to diboson resonances with masses in the tail of the invariant dijet mass spectrum, where only a small number of background events is expected ($\lesssim 5$), new working points with higher W/Z signal identification efficiencies ($\approx 80\%$) could be introduced. These working points with high signal efficiencies would allow for the selection of almost all signal events in the high invariant dijet mass region which are produced at very low rates.

For the optimisation of the algorithm to identify boosted hadronically vector bosons, only longitudinal polarised W and Z bosons were used. However, vector bosons from the decay of a new resonance can also be produced with transverse polarisation. As the polarisation of the W/Z boson affects the substructure variables, the derived algorithms are not ideal for the identification of transversely polarised W/Z bosons. Therefore, dedicated algorithms could improve the sensitivity to resonances with transversely polarised vector bosons in the final state.

The combination of the improvements mentioned above together with the large incoming dataset are set to significantly advance the sensitivity to new resonances decaying to vector bosons.

Chapter 10

Conclusions

Two main topics were discussed in this thesis: the optimisation of an algorithm to identify boosted hadronically decaying vector bosons and the application thereof to the search for diboson resonances.

Because of the high centre-of-mass energy at the LHC, vector bosons are often produced with a transverse momentum much larger than their mass, resulting in a collimation of their decay products. Innovative approaches were needed to successfully reconstruct boosted hadronically decaying vector bosons at the LHC and their performance was extensively studied in MC simulation and data at 8 and 13 TeV in the context of this thesis. Three main ingredients are needed to effectively identify vector bosons: a grooming algorithm to mitigate the influence of pile-up effects on the large-radius jet, a requirement on the jet mass and the imposition of criteria on substructure variables to discriminate a W/Z -jet from a QCD-jet. Various combinations of jet grooming algorithms and substructure variables have been studied and their performance was evaluated in terms of the multi-jet background rejection for fixed signal efficiencies. Jets reconstructed with the anti- k_t algorithm ($R = 1.0$) and trimmed with $f_{\text{cut}} = 5\%$ and $R_{\text{sub}} = 0.2$ were found to perform particularly well in combination with a mass window requirement and a cut on the energy correlation variable $D_2^{\beta=1}$. For a W -jet signal efficiency of 50%, background efficiencies of about 2% were found in data which are in good agreement with the MC predictions within the uncertainties. To be able to use the algorithm to identify boosted W and Z bosons in an analysis, MC-based energy and mass calibrations were derived and systematic uncertainties on the jet energy, mass and $D_2^{\beta=1}$ scale were determined using in-situ techniques.

Two searches for narrow resonances decaying into pairs of vector bosons (WW , WZ and ZZ) were presented based on datasets corresponding to an integrated luminosity of 20.3 fb^{-1} and 3.2 fb^{-1} of proton-proton collisions, collected with the ATLAS detector at centre-of-mass energies of 8 TeV and 13 TeV respectively. The hadronic decay of both vector bosons was studied due to its large branching ratio and the final state is thus characterised by two large-radius jets that are compatible with boosted W and Z bosons. The analysis searches for a resonant structure in the invariant dijet mass on top of the smoothly falling background distribution. The background is estimated from a fit to the invariant dijet mass in data. An excess of events was observed in

all three signal regions for an invariant dijet mass of 2 TeV in the 8 TeV analysis. The largest deviation was observed in the WZ channel corresponding to a global significance of 2.5σ . The analysis was repeated using the data recorded in 2015 and no significant deviations from the background hypothesis were found. The results obtained at 8 and 13 TeV were interpreted in terms of exclusion limits at 95% CL on the production cross-section times branching ratio for three different benchmark models: a bulk Randall-Sundrum model, an extended gauge model and a heavy vector triplet model. W' bosons with couplings as predicted by the extended gauge model are excluded at 95% CL for masses between 1.3 and 1.5 TeV using the 8 TeV data. The 13 TeV data was used to exclude W' bosons, predicted in the heavy vector triplet model A, in the mass range from 1.38 to 1.6 TeV. The fully hadronic diboson searches were not sensitive enough to exclude RS gravitons with in models with $k/\bar{M}_{\text{Pl}} = 1.0$.

Appendix A

Definition of Substructure Variables

Angularity

The *angularity* [165] examines the energy flow within the large- R jet and is defined as

$$a_3 = \frac{1}{M} \sum_i E_i \sin^\alpha \left(\frac{\pi \theta_i}{2R} \right) \left[1 - \cos \left(\frac{\pi \theta_i}{2R} \right) \right]^{1-\alpha} \quad (\text{A.1})$$

where the sum runs over the constituents i of the jet and θ_i is the angle between the constituent i and the jet-axis. The parameter α can be chosen such that either more or less weight is given to constituents at the edge of the cone of the jet with respect to the ones close to the jet-axis. In Reference [165], $\alpha = -2$ was used. The angularity distribution for W -jets peaks at small values close to zero and falls off sharply whereas the peak for QCD-jets is slight shifted to higher values and exhibits much longer tails than the W -jet distribution.

Aplanarity and Sphericity

Two variables, the *aplanarity* A and *sphericity* S are defined from the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of the sphericity tensor [166]:

$$S^{\alpha,\beta} = \frac{\sum_i p_i^\alpha p_i^\beta}{\sum_i |\vec{p}_i|^2}. \quad (\text{A.2})$$

The sphericity tensor is calculated from the momenta of the constituents $\vec{p}_i$ where α and β refer to the x, y and z component of the momentum. The eigenvalues are ordered, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, and normalised $\lambda_1 + \lambda_2 + \lambda_3 = 1$. The aplanarity is then defined as

$$A = \frac{3\lambda_3}{2}, \quad (\text{A.3})$$

and the sphericity can be calculated as follows:

$$S = \frac{3}{2}(\lambda_2 + \lambda_3). \quad (\text{A.4})$$

The aplanarity and sphericity are restricted to $0 < A < 0.5$ and $0 < S < 1.0$ respectively. For W -jets, the aplanarity and sphericity distributions peak at values close to zero whereas a value of one corresponds to an isotropic distribution of the energy of the clusters within the

jet. As for the angularity, the peak of the QCD-jet distributions are shifted and exhibit tails in the direction of an isotropic energy distribution. Both the angularity and sphericity have been used in the cross-section measurement of boosted vector bosons at $\sqrt{s} = 7$ TeV by the ATLAS Collaboration [167].

Dipolarity

The dipolarity $\mathcal{D}$ [168] was introduced to identify boosted hadronically decaying top quarks. It can also be used to discriminate between W -jets and QCD-jets by exploring the colour flow within the jet. For a colour singlet as the W boson, most of the radiation of the two quarks will be clustered between the two subjets compared to QCD-jets with soft and wide-angled radiation patterns. The bipolarity is defined as

$$\mathcal{D} = \frac{1}{\Delta R_{12}^2} \sum_i \frac{p_{Ti}}{p_{TJ}} R_i^2 \quad (\text{A.5})$$

where ΔR_{12} is the angular separation of the two subjets and the sum runs over the constituents of the jet. For a QCD-jet, the dipolarity will be large because of the wide-angle radiation controlled by R_i in Equation A.5 whereas W -jets have a small dipolarity.

Fox-Wolfram Moments

The Fox-Wolfram Moments [169–171] have previously been used to explore the geometrical structure of QCD events. However it has been shown that the ratio of the second to zeroth Fox-Wolfram moment R_2^{FW} can be used to identify the two-prong structure of W boson decays in the centre-of-mass frame:

$$R_2^{\text{FW}} = \frac{\sum_i \sum_{j=i+1} |\vec{p}_i| |\vec{p}_j| (3 \cos^2(\theta_{ij}) - 1)}{2 \cdot (\sum_i E_i)^2} . \quad (\text{A.6})$$

The two sums run over the constituents of the large- R jet and θ_{ij} is the angle between two constituents i and j . For a W boson in the rest frame, the two subjets are expected to be back-to-back resulting in $R_2^{\text{FW}} \approx 1$ compared to $R_2^{\text{FW}} \approx 0$ for QCD-jets.

Planar Flow

The *planar flow* P [165] measures how the energy of the jet's constituents is distributed with respect to the jet axis. The two dimensional matrix I_{jk} is defined

$$I_{jk} = \frac{1}{M} \sum_i \frac{p_{i,j} p_{i,k}}{E_i} \quad (\text{A.7})$$

that described the momentum correlation of the constituents. M is the mass of the large- R jet and the sum runs over the constituents of the jet with $p_{i,j}$ being the j^{th} component of the

transverse momentum of the i^{th} constituent. The planar flow can then be calculated from the momentum correlation matrix:

$$P = 4 \times \det(I) / \text{tr}(I)^2 . \quad (\text{A.8})$$

For an isotropic distribution of the jet energy as dominantly for QCD-jets, a planar flow close to one is expected whereas for the two-prong decay of a W boson, the planar flow tends to values close to zero.

Volatility

The purpose of jet reclustering algorithms is to invert the parton showering process and to combine the 4-vector of particles or pseudo-jets into one final state jet. The clustering of the final state particles is however not unique such that one jet can be reclustered in many different ways. The set of jets produced in the many different clustering choices for one single jet are referred to as Q -jets [172]. For each of the clustering choices a unique observable can be calculated. Therefore, an observable, e.g. the mass of a jet is not given by a single value anymore but characterised by a distribution. To improve the discrimination of W -jets from QCD-jets, the width of the mass distribution for an individual jet, the so-called *volatility* ν_{QJets} can be used:

$$\nu_{\text{QJets}} = \frac{\sqrt{\langle m^2 \rangle - \langle m \rangle^2}}{\langle m \rangle} , \quad (\text{A.9})$$

The mass distribution is much wider for QCD-jets compared to W -jets due to their intrinsic mass scale. The Q -jet algorithm has been studied in the ATLAS Collaboration and shown to have better background rejection than the N -subjettiness τ_{21} for high W tagging efficiencies [173].

Soft Drop Variable

The soft-drop algorithm [85] is a generalisation of the modified mass drop tagger [80] which takes any given large- R jet as input and removes wide-angle and soft radiation from the jet. In the first step, the jet is reclustered with the C/A algorithm and afterwards declustered following the branch of the highest p_{T} . At each step of the clustering history, the softer proto-jet is removed unless the following condition is fulfilled:

$$\frac{\min(p_{\text{T}1}, p_{\text{T}2})}{p_{\text{T}1} + p_{\text{T}2}} > z_{\text{cut}} \times \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta} , \quad (\text{A.10})$$

where R_0 is the radius parameter of the initial jet algorithm, ΔR_{12} the angular distance of the two proto-jets and $p_{\text{T}1}, p_{\text{T}2}$ their transverse momenta. Nine different values have been studied for the z_{cut} parameter ranging from 4% to 20% and three for β : -1.0 , -0.75 , and -0.5 .

To define the *soft-drop-level* variable L_{SD} , the algorithm starts with $z_{\text{cut}} = 4\%$. If the requirement in Equation A.10 is not fulfilled, the algorithm proceeds to the next step in the jet's clustering history. In case the condition is not satisfied at any point of the algorithm, L_{SD}

L_{SD}	1	2	3	4	5	6	7	8	9
z_{cut}	4%	6%	8%	10%	12%	14%	16%	18%	20%

Table A.1: The soft-drop level variable L_{SD} corresponding to the highest z_{cut} value that are fulfilled in the soft-drop algorithm.

is defined as zero and otherwise $L_{\text{SD}} = 1$. The algorithm then further tests if the condition in Equation A.10 is also fulfilled for larger z_{cut} values and assigns L_{SD} variables to it as indicated in Table A.1. The resulting soft-drop-level is thus discrete but could be also defined as continuous variable if not only concrete z_{cut} values would be explored.

Splitting Scales

For jets reclustered with the k_t algorithm, the k_t *splitting scale* $\sqrt{d_{12}}$ [70] is defined at the last step of the reclustering where two proto-jets are clustered into the large- R jet:

$$\sqrt{d_{12}} = \min(p_{\text{T1}}, p_{\text{T2}}) \times \Delta R_{12} . \quad (\text{A.11})$$

ΔR_{12} is the distance in the η - ϕ plane of the two proto-jets at the last step of the reclustering and p_{T1} and p_{T2} their transverse momenta. The splitting scale variable is the square root of the distance measure in Equation 4.3 multiplied by the jet radius. As the k_t algorithm is used for the reclustering, the two hardest proto-jets are combined last and thus the p_{T1} and p_{T2} are expected to be relatively symmetric for W -jets and thus $\sqrt{d_{12}} \approx m_{\text{jet}}/2$. On the other hand, the splitting of QCD-jets is asymmetric and results in a steeply falling spectrum.

Thrust Minor and Thrust Major

The *thrust axis* $\vec{n}$ [170] is defined as the normalised axis that maximises the following equation:

$$T = \max \frac{\sum_i |\vec{p}_i \cdot \vec{n}|}{\sum_i |\vec{p}_i|} , \quad (\text{A.12})$$

where the sum runs over the constituents of the jet. Two further variables, *thrust major* and *thrust minor* [170] can be constructed:

$$T_{\text{maj}} = \max \frac{\sum_i |\vec{p}_i \cdot \vec{n}_1|}{\sum_i |\vec{p}_i|} , \quad (\text{A.13a})$$

$$T_{\text{min}} = \frac{\sum_i |\vec{p}_i \cdot \vec{n}_2|}{\sum_i |\vec{p}_i|} . \quad (\text{A.13b})$$

The normalised axes $\vec{n}_1$ and $\vec{n}_2$ are perpendicular to the thrust axis and to each other. For the thrust minor variable, values of $T_{\text{min}} \rightarrow 0$ correspond to jets where the energy is aligned along the jet direction compared to an isotropic energy distribution with $T_{\text{min}} \rightarrow 1$. Therefore W -jets

tend to have smaller thrust minor values compared to QCD-jets.

z Variable

The z_{12} variable [174] was introduced to identify boosted top quarks but can also be used for W boson identification. The variable uses the differences in the early shower history because the first splitting for a QCD-jet is expected to be soft whereas the splitting of a top jet or W -jet splitting should be hard as the subjects are revealed. Based on the definition of the k_t splitting scale, the z_{12} variable is modified as follows:

$$z_{12} = \frac{d_{12}}{d_{12} + M^2} . \tag{A.14}$$

Appendix B

Prospect Studies for the Vector Boson Associated Higgs Boson Production for the High-luminosity LHC

The measurements of the decay into boson pairs ($\gamma\gamma$, WW , ZZ) of the new particle discovered by the ATLAS and CMS Collaborations in 2012 have shown that it is compatible with the predictions of the Standard Model Higgs boson [12, 13]. On the other hand, only evidences for the direct Higgs decay into fermions were found so far [175, 176] and the Yukawa coupling of the Higgs boson to fermions remains to be thoroughly studied. The precise measurements of the Higgs sector to understand the electroweak symmetry breaking mechanism are thus a priority in the running of the LHC and the planned high-luminosity (HL) upgrade of the LHC.

During the long shutdown (LS) 2 in 2018, the injector chain and the LHC will be upgraded to deliver an increased peak luminosity of about $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ at $\sqrt{s} = 14 \text{ TeV}$ [177]. Approximately 300 fb^{-1} of pp collision data with an average of 60 interactions per bunch cross-section are expected to be collected by 2022, the start of the LS3. In LS3, the LHC will be upgraded to the HL-LHC to a new design luminosity of $10^{35} \text{ cm}^{-2}\text{s}^{-1}$ [177]. To achieve this goal, the beam will constitute of more populated and denser bunches and the beam size at the collision point will be reduced. Furthermore, the detectors will undergo major changes to cope with the high radiation dose close to the interaction points and the much higher data rate that needs to be recorded. Starting from 2024, the HL-LHC is expected to deliver an integrated luminosity of about $250 - 300 \text{ fb}^{-1}$ per year with $\langle\mu\rangle = 140$, resulting in a total dataset of about 3000 fb^{-1} by 2035.

The large dataset, expected to be collected during the running of the HL-LHC, does not only allow for the measurement of rare decays of the Higgs boson, such as $t\bar{t}H$ with $H \rightarrow \mu\mu$ and the Higgs self-coupling, but also to thoroughly measure the vector boson scattering $VV \rightarrow VV$ with $V = W/Z$ [178]. In the SM, the Higgs boson is responsible to preserve the unitarity of the $VV \rightarrow VV$ cross-section which would otherwise grow linearly due to the quartic gauge boson self-coupling. Deviations from the predictions would indicate the presence of a new particle that couples to gauge bosons via the electroweak interaction. Furthermore, the large dataset enables the search for rare exotic processes with small production cross-sections and to extend

the covered mass range in Run-II.

The study presented here investigates the sensitivity to the Higgs boson production ($H \rightarrow b\bar{b}$) in association with either a W or Z boson at the high-luminosity (HL) running of the ATLAS experiment at the LHC [179]. At a Higgs mass of approximately 125 GeV, the Higgs boson decays with approximately 58% into $b\bar{b}$. However this channel can only be studied in the associated production with either a vector boson (W , Z) or top quark due to the overwhelming multijet background. Even though the production cross-section is more than an order smaller compared to the gluon-gluon fusion, the leptonic decay of the vector boson in the VH channel can be used to suppress the background and to trigger the event. The search for WH and ZH production is therefore performed in the $\ell\nu b\bar{b}$ or $\ell\ell b\bar{b}$ final state, respectively. The emphasis of this chapter will be on the WH channel, however it will be shown that the combination of the WH and ZH channels improves the sensitivity of the search.

The HL-LHC analysis was performed in 2014 and follows closely the general strategy of the VH , $H \rightarrow b\bar{b}$ search performed at $\sqrt{s} = 8$ TeV [180] which did not observe any significant excess of events over the SM background. The 8 TeV analysis will be denoted in the following as benchmark analysis. The main difference is that the benchmark analysis uses full event simulation as described in Chapter 3.2.5 whereas the HL analyses uses particle-level simulation with applied parameterisations of the expected ATLAS detector response at the HL conditions. The invariant mass distribution of the two b -tagged jets, $m_{b\bar{b}}$ is used as input to a binned likelihood fit that evaluates the significance and expected error on Higgs production at $m_H = 125$ GeV. Systematic uncertainties are taken into account as nuisance parameters.

B.1 Signal and Background Samples

Monte Carlo events for the $qq \rightarrow VH$ signal with $m_H = 125$ GeV are generated using PYTHIA8 with the CTEQ6L1 parton distribution functions and the AU2 tune for parton showering and hadronisation. The ZH production also has significant contributions from gluon-initiated processes [181, 182] and is simulated with POWHEG using the CT10 PDFs and PYTHIA8 for the parton showering. Furthermore the AU2 tunes are used.

The main background processes are W +jets and single-top quark production for WH and Z +jets for ZH . Both channels have furthermore large contributions from $t\bar{t}$ and diboson production. The $t\bar{t}$, single top s-channel and Wt production processes are simulated with POWHEG interfaced with PYTHIA and the CTEQ6L1 PDFs. For the t-channel exchange process, the ACERMC [183] generator is used with the CTEQ6L1 PDFs. Diboson processes are simulated with HERWIG using CTEQ6L1 PDFs and the AUET2 tune [184]. The V +jets background samples, generated with SHERPA and interfaced with CT10 PDFs, are categorised based on the flavour of the two leading jets in the event. The flavour identification of jets is described in Section B.2. The multijet background estimate from the 8 TeV analysis is used to predict the multijet contami-

nation at the HL-LHC and scaled by the increase in gluon-gluon parton luminosity between 8 and 14 TeV, which corresponds to a factor of ~ 2.3 [185] .

B.2 Object Definitions

Due to the absence of full simulation of the upgraded HL-LHC ATLAS detector at the time of these studies, the analyses presented here relies on the usage of particle-level information and parameterisations of the detector response of objects such as leptons, jets and missing transverse energy. The parameterisations of the identification efficiency and energy smearing of the different objects are described in detail in [186, 187]. The MC samples used here do not include any additional interactions from pile-up but the effects of pile-up on the selected physics objects are included in the efficiency and resolution parameterisations

Jets

Particle-level jets are reconstructed with the anti- k_t algorithm and a radius parameter of $R = 0.4$ using stable truth particles with a lifetime greater than 10 ps as inputs (excluding muons and neutrinos). The energy of the jets is then smeared to simulate the expected degradation of the energy resolution under the HL-LHC conditions taking into account that the noise term in Eq. 3.3 increases linearly with the average number of interactions per bunch-crossing. To suppress jets coming from vertices others than the hard-scatter vertex, a track-jet confirmation is applied that requires the jet to be reconstructed close to a track originating from the hard-scatter vertex. The efficiency of the track-jet confirmation is 75% for jets with $p_T < 25$ GeV, 80% for jets with $25 < p_T < 35$ GeV and above 95% for jets with $p_T > 35$ GeV. Jets are required to have $p_T > 30$ GeV compared to 20 GeV that was used in the benchmark analysis and are limited to the central detector region in this analysis with $\eta < 2.5$. The jet p_T threshold was increased in order to cope with the higher pile-up contamination at the HL-LHC. The jet p_T threshold was chosen such that for $\langle \mu \rangle = 140$, the contamination of pile-up jets corresponds to 10% in the central detector region.

The flavour of the jet is then determined by matching hadrons to the jet within a cone of $R = 0.4$. If a b -hadron is found, the jet is labelled as b -jet, if no b -hadron is found but a c -hadron, the jet is labelled a c -jet otherwise the jet is classified as light jet. The b -jet, c -jet and light-jet efficiencies using the multivariate MV1 [187] tagger are parameterised in p_T and η and correspond to working point with an average b -tagging efficiency of 70% as measured in a semi-leptonic $t\bar{t}$ sample.

Leptons

Two types of leptons (*loose* and *tight*) are defined based on the purity of the identification criteria. Loose electrons and muons are selected with transverse energy $E_T > 10$ GeV and required to have $|\eta| < 2.47$ and $|\eta| < 2.7$, respectively. Leptons are identified as tight leptons

p_T^V bins		Exclusive number	Exclusive number
Bin 1	$p_T^V < 90$ GeV	of b -jets	of total jets
Bin 2	$90 \text{ GeV} < p_T^V < 120 \text{ GeV}$	0	2
Bin 3	$120 \text{ GeV} < p_T^V < 160 \text{ GeV}$	1	3
Bin 4	$160 \text{ GeV} < p_T^V < 200 \text{ GeV}$	2	
Bin 5	$p_T^V > 200 \text{ GeV}$		

Table B.1: Definition of the bins used in the analysis. The one-lepton analysis uses only p_T^V bins 2 - 5; bin 1 is not used to avoid the large multijet contamination.

if they have transverse momentum $E_T > 25$ GeV and $|\eta| < 2.47$ for electrons and $|\eta| < 2.5$ for muons. Furthermore, both loose and tight leptons must pass isolation requirements to reject jets misidentified as electrons. The identification efficiency of loose (tight) electrons ranges from 82 (66)% to 96 (85)% for electrons with low and high transverse energy, respectively. The efficiency of reconstructing a jet as an electron is less than 8% at low transverse energy and decreases with increasing E_T . A track isolation criteria is imposed on the ratio of the sum of the p_T of all charged particles in a cone of $\Delta R = 0.4$ around the lepton to lepton- p_T . The track isolation has to be $< 10\%$ and $< 4\%$ for loose and tight leptons, respectively. For tight leptons an additional calorimeter isolation requirement, defined as the ratio of the sum of E_T of all particles in a cone of $\Delta R = 0.3$ around the lepton to the lepton- E_T , is imposed and has to be less than 7%.

Missing Transverse Energy

The missing transverse energy E_T^{miss} is calculated from the vector sum of particles within the detector acceptance and smeared according to the expected degradation with increasing pile-up conditions.

B.3 Event Selection

The selection criteria used in this analyses follow closely the ones of the benchmark analysis. The analysis is performed in bins of the transverse momentum of the vector boson, p_T^V , the exclusive number of b -tagged jets and the exclusive number of jets in the event, to increase the sensitivity of the search. The definitions of the bins used in the analysis are summarised in Table B.1. Thus the signal region, that requires two b -tagged jets consists of 8 different bins for the WH channel and ten for the ZH channel. In the one-lepton channel (WH), the lowest p_T^V bin is not considered due to the large multijet contamination.

Events are required to pass a parameterised single-lepton trigger requirement, as detailed in Ref. [187]. The single electron trigger requires one electron with $p_T > 18$ GeV within $|\eta| < 2.5$ and corresponds to an efficiency of 88%. The single muon trigger requires one muon with $p_T > 20$ GeV with an efficiency of 64% and 86% for muons with $|\eta| < 1.0$ and $1.0 < |\eta| < 2.4$,

	p_{T}^V (GeV)	0 – 90	90 – 120	120 – 160	160 – 200	> 200
All channels	$\Delta R(b, b)$	0.7 – 3.4	0.7 – 3.0	0.7 – 2.3	0.7 – 1.8	< 1.4
One-lepton	$E_{\text{T}}^{\text{miss}}$	> 25 GeV				> 50 GeV
	m_{T}^W	40 – 120 GeV			< 120 GeV	
Two-lepton	$E_{\text{T}}^{\text{miss}}$	< 60 GeV				

Table B.2: Event selection criteria for the one and two-lepton channel in the different p_T^V bins.

respectively.

In the $WH \rightarrow \ell \nu b \bar{b}$ decay channel (one-lepton channel), one tight lepton is required and no additional loose lepton. Furthermore a large amount of missing transverse energy is expected and thus $E_T^{\text{miss}} > 25$ GeV and $E_T^{\text{miss}} > 50$ GeV is required for $p_T^V < 200$ GeV and $p_T^V > 200$ GeV, respectively. The transverse mass, m_T^W , is defined from the transverse momenta and the azimuthal angle of the charged lepton (p_T^ℓ and ϕ^ℓ) and the missing transverse energy and its direction ϕ^{miss} : $m_T^W = \sqrt{2p_T^\ell E_T^{\text{miss}}(1 - \cos(\phi^\ell - \phi^{\text{miss}}))}$. The selection criteria are summarised in Table B.2. Events consistent with the $ZH \rightarrow \ell \ell b \bar{b}$ decay (two-lepton channel) need to contain two loose leptons of the same flavour. To be consistent with the Z boson decay, the invariant dilepton mass $m_{\ell\ell}$ is restricted to $83 < m_{\ell\ell} < 99$ GeV. In addition, the amount of missing transverse energy is constrained to $E_T^{\text{miss}} < 60$ GeV.

For both channels, at least two b -tagged jets are required in the event with $p_T > 40$ and $|\eta| < 2.5$ and the leading jet must fulfil $p_T > 60$ GeV for the 14 TeV analysis. At most one additional jet with $p_T > 20$ (30) GeV and $|\eta| < 2.5$ is allowed in this two (one)-lepton analysis, otherwise the event is rejected. The p_T criteria for the third jet was increased in the one-lepton channel to reduce the $t\bar{t}$ background. Even though a lower veto jet p_T criteria, e.g. 20 GeV, results in the largest signal-over-background ratio, a criteria of 30 GeV was imposed to avoid an increase in the pile-up jet contamination. The two b -tagged jets in the event, used to build the Higgs candidate are required to be the two leading jets in the event. Furthermore requirements on the angular separation of the two b -tagged jets $\Delta R(b\bar{b})$ in dependence of the transverse momentum of the vector boson transverse are summarised in Table B.2

B.4 WH Analysis Validation with $\sqrt{s} = 8$ TeV Analysis

Before studying the VH production at the HL-LHC, the event selection was applied to the MC samples that were used for the 8 TeV benchmark analysis (and the cross-sections were adjusted accordingly) to not only validate the setup framework but also to study the the impact of the usage of particle-level information and the corresponding smearing and efficiency functions compared to full detector simulation.

The same event selection as previously described was used, however the leading jet p_T requirement was lowered to 45 GeV. New smearing and efficiency functions were derived in some cases

based on full ATLAS detector simulation samples as the ones derived for the HL-LHC assumed different pile-up conditions. For the lepton identification and reconstruction efficiencies, the HL-LHC smearing functions were used as these efficiencies assumed 2012 pile-up conditions. The trigger efficiencies for the single lepton trigger are taken from [188, 189] and correspond to the efficiencies used in the 8 TeV analysis. The same jet energy resolution parameterisation as derived for $\sqrt{s} = 14$ TeV, are used as well for the 8 TeV validation as the parameterisation depends on $\langle\mu\rangle$. For a central jet with $p_T = 60$ GeV at $\langle\mu\rangle = 20$, the contribution to the jet energy resolution due to pile-up is 25%. Furthermore new b -tagging efficiency maps were derived for 8 TeV based on the MV1 algorithm [146]. In the benchmark analysis, jets from pile-up vertices are suppressed by requiring that the scalar sum of the p_T of tracks from the hard-scatter vertex matched to the jet are at least 50% of the scalar sum of the p_T of all tracks matched to the jet [104]. The efficiency of this criterion corresponds to 90% for jets with $p_T = 25$ GeV and 95% for jets with $p_T = 50$ GeV. The missing transverse energy resolution is parameterised in $\langle\mu\rangle$ and thus the HL-LHC parameterisation can be used.

Non-negligible differences can be observed between the HL-LHC VH and the benchmark analysis, however only the WH differences will be discussed here. For the ZH analysis details can be found in [179]. The event yields differ between the particle-level and the benchmark analysis especially for the signal and the two dominating backgrounds $t\bar{t}$ and $W + bb$ and the discrepancy is increasing with p_T^V . Differences in the WH signal shape distributions can be also observed. Furthermore, the flavour composition of the W +jets background shows discrepancies with respect to the benchmark analysis and the $W + bb$, $W + bl$ and $W + cc$ yields are mostly larger in the particle-level analysis whereas the $W + cl$ and $W + l$ contributions are smaller. The $m_{b\bar{b}}$ distributions of the benchmark analysis are compared in Fig. B.1 and Fig. B.2 to the particle-level analysis in different p_T^V categories and for the two and three jet category respectively. Furthermore, the $m_{b\bar{b}}$ distributions are shown for the smeared particle-level analysis where each of the backgrounds is scaled to the event yields of the benchmark analysis to allow for a better shape comparison.

To understand the overall larger event yields of the smeared analysis, a simplified full reconstruction analysis was performed for the WH signal and the dominating $t\bar{t}$ background where simplified implies here that no object calibrations and no scale factors were used. The event yields were then compared at each step of the event selection. About 20% more low- p_T electrons are selected in the particle-level analysis whereas the number of selected muons agrees well within the uncertainties with the full simulation analysis. The comparison of the E_T^{miss} distribution showed that the missing transverse energy in the particle-level analysis is over-smeared by approximately 5% compared to the full simulation analysis. This does not only result in larger E_T^{miss} and more events that pass the selection criteria on it but it also propagates to the p_T^V distribution and thus causes events to migrate to higher p_T^V categories in which the discrepancy between the particle-level and benchmark analysis is more evident. Furthermore the lower event

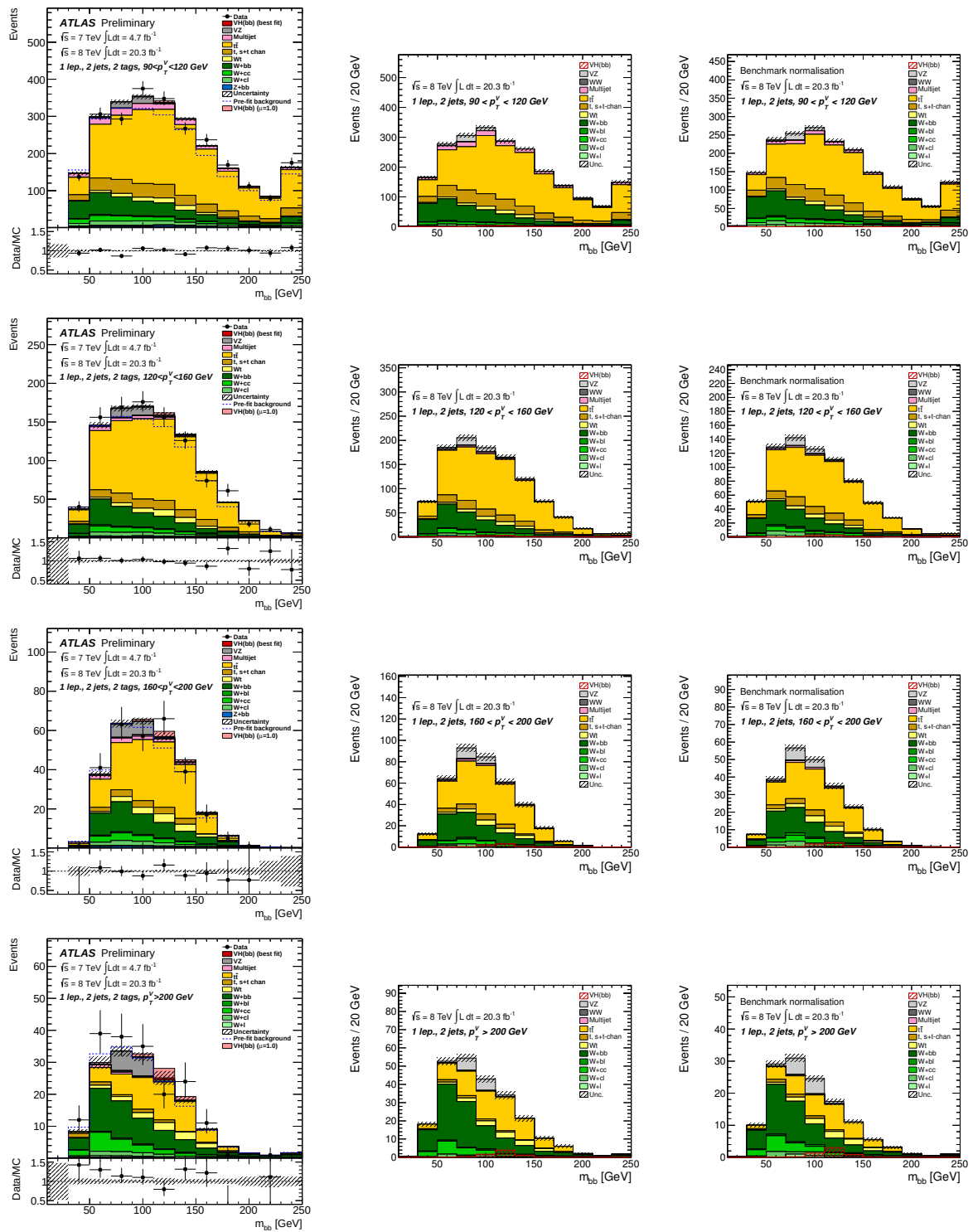


Figure B.1: m_{bb} distribution for the 2-jet signal regions of the one-lepton channel for the full 8 TeV benchmark analysis (left), the particle level plus smearing (centre) and the particle level plus smearing with event yields of each background normalised to the yields of benchmark analysis (right). The entries in overflow are included in the last bin

yields of the full simulation analysis can be understood due to the presence of pile-up jets. Even though pile-up suppression techniques are applied and thus the number of pile-up jets is reduced, more events are rejected especially for the $t\bar{t}$ background as they fail the veto on the number of additional jets.

The $m_{b\bar{b}}$ shape differences between the benchmark and particle-level analysis for the WH signal are due to the muon-in-jet correction applied in the former one. The jet reconstruction algorithm used in ATLAS does not include muons or neutrinos however they can be produced in the semileptonic decay of b -jets and thus creates a bias in the jet energy scale and a diminished energy resolution. After adding the energy of the closest muon to a b -jet (within a $R = 0.4$ cone) in the particle-level analysis, the energy resolution is improved and results in a better agreement with the benchmark analysis.

The comparison of the particle-level analysis with the benchmark analysis helped in understanding the event selection and object definitions as well as the impact of the usage of truth-level information and corresponding smearing functions instead of full detector simulation. Based on these studies, the object definition of b -tagged jets was modified for the HL-LHC analysis to include the muon-in-jet correction.

B.5 HL-LHC Analysis

The event yields in the two-jet and three-jet signal regions for the one-lepton channel are summarised in Tab. B.3 and Tab. B.4 assuming an integrated luminosity of 300 fb^{-1} and $\langle\mu\rangle = 60$ and $\mathcal{L} = 3000 \text{ fb}^{-1}$ and $\langle\mu\rangle = 140$ respectively. Furthermore the signal-over-background ratios $S/\sqrt{B}$ are indicated and show that the highest p_{T}^V has the highest sensitivity. However the $t\bar{t}$ background will have to be significantly reduced to improve the sensitivity of the analysis. Some improvements that were beyond of the scope of this analysis are discussed in Section B.8. The $m_{b\bar{b}}$ distributions for the two-jet and three-jet signal regions are shown in Fig. B.3 and B.4 for an integrated luminosity of 300 fb^{-1} and $\langle\mu\rangle = 60$ and $\mathcal{L} = 3000 \text{ fb}^{-1}$ and $\langle\mu\rangle = 140$, respectively.

B.6 Systematic Uncertainties

Most of the systematic uncertainties, considered in the HL-LHC analysis, are based on the uncertainties used in the benchmark analysis. However only the uncertainties with a non-negligible impact on the benchmark analysis are taken into account here. Systematic uncertainties that affect the normalisation of the signal and background samples as well as the shape of the $m_{b\bar{b}}$ distribution are taken into account and divided into three main categories: experimental uncertainties, background only and signal only uncertainties. Each of the systematic uncertainties is considered as a nuisance parameter in the statistical treatment.

One-lepton, 2-jet signal region				
Process	$p_T^W [\text{GeV}]$			
	90 – 120	120 – 160	160 – 200	> 200
WH	114.1 ± 3.1	118.6 ± 3.1	130.8 ± 3.2	216 ± 4
ZH	7.11 ± 0.18	3.79 ± 0.13	2.89 ± 0.11	2.28 ± 0.11
VH total	121.2 ± 3.1	122.4 ± 3.1	133.7 ± 3.2	218 ± 4
$W + l$	60 ± 60	57 ± 32	25 ± 19	67 ± 30
$W + cl$	870 ± 130	480 ± 80	360 ± 70	400 ± 60
$W + cc$	1930 ± 190	1200 ± 130	960 ± 110	2090 ± 130
$W + bl$	250 ± 30	182 ± 28	114 ± 18	149 ± 18
$W + bb$	3670 ± 130	3270 ± 120	2900 ± 100	5290 ± 110
$Z + bb$	1130 ± 60	470 ± 40	224 ± 29	164 ± 25
$Wt\text{-ch}$	3070 ± 60	2160 ± 50	1610 ± 40	1790 ± 50
$s - /t\text{-ch}$	6110 ± 60	2170 ± 40	683 ± 19	365 ± 14
$t\bar{t}$	62700 ± 400	32570 ± 290	16470 ± 210	10180 ± 160
Multijet	1181 ± 15	146.5 ± 3.4	71.1 ± 2.3	35.7 ± 1.8
WW	46 ± 13	25 ± 10	18 ± 8	60 ± 15
VZ	369 ± 19	413 ± 19	464 ± 20	597 ± 22
Total bkg	81400 ± 500	43200 ± 400	23930 ± 270	21190 ± 250
$S/B (\times 10^{-3})$	1.49 ± 0.04	2.84 ± 0.08	5.59 ± 0.15	10.28 ± 0.23
$S/\sqrt{B}$	0.43 ± 0.01	0.59 ± 0.02	0.86 ± 0.02	1.50 ± 0.03
One-lepton 3-jet signal region				
Process	$p_T^W [\text{GeV}]$			
	90 – 120	120 – 160	160 – 200	> 200
WH	24.9 ± 1.4	27.0 ± 1.5	33.1 ± 1.6	72.2 ± 2.3
ZH	2.79 ± 0.11	1.80 ± 0.09	2.00 ± 0.09	2.26 ± 0.11
VH total	27.7 ± 1.4	28.8 ± 1.5	35.1 ± 1.6	74.4 ± 2.3
$W + l$	0 ± 0	38 ± 21	0 ± 0	31 ± 23
$W + cl$	320 ± 80	130 ± 40	115 ± 33	146 ± 32
$W + cc$	960 ± 140	370 ± 70	500 ± 80	900 ± 80
$W + bl$	126 ± 25	73 ± 18	65 ± 14	72 ± 13
$W + bb$	1600 ± 90	1200 ± 70	1200 ± 60	3000 ± 80
$Z + bb$	490 ± 40	206 ± 28	110 ± 20	48 ± 14
$Wt\text{-ch}$	3180 ± 60	1910 ± 50	1570 ± 40	2350 ± 50
$s - /t\text{-ch}$	3360 ± 40	1062 ± 25	331 ± 14	297 ± 12
$t\bar{t}$	101000 ± 500	42810 ± 340	21430 ± 240	17200 ± 200
Multijet	370 ± 9	49.3 ± 1.9	26.4 ± 1.3	13.7 ± 1.1
WW	11 ± 6	11 ± 6	18 ± 8	25 ± 9
VZ	73 ± 9	85 ± 9	92 ± 8	177 ± 1
Total bkg	111500 ± 500	48000 ± 400	25500 ± 270	24260 ± 250
$S/B (\times 10^{-3})$	0.25 ± 0.01	0.60 ± 0.03	1.38 ± 0.07	3.07 ± 0.11
$S/\sqrt{B}$	0.083 ± 0.004	0.131 ± 0.007	0.22 ± 0.01	0.48 ± 0.02

Table B.3: Estimated signal and background event yields, for the one-lepton channel. The yields are for $\sqrt{s} = 14$ TeV, $\langle\mu\rangle = 60$ and an integrated luminosity of 300 fb^{-1} , before the fit model has been applied. The uncertainties correspond to limited Monte Carlo statistics.

One-lepton 2-jet signal region				
Process	p_T^W [GeV]			
	90 – 120	120 – 160	160 – 200	> 200
WH	928 ± 28	950 ± 28	986 ± 28	2030 ± 40
ZH	71.1 ± 1.8	39.7 ± 1.3	28.1 ± 1.1	26.5 ± 1.1
VH total	1000 ± 28	990 ± 28	1014 ± 28	2060 ± 40
$W + l$	1200 ± 800	1700 ± 900	22 ± 22	1200 ± 600
$W + cl$	5400 ± 1000	4000 ± 800	4100 ± 800	4900 ± 700
$W + cc$	14000 ± 1600	10000 ± 1300	7600 ± 900	23100 ± 1500
$W + bl$	2400 ± 300	1020 ± 210	1380 ± 230	1350 ± 170
$W + bb$	23800 ± 1100	24200 ± 1000	23900 ± 900	59400 ± 1300
$Z + bb$	10900 ± 600	4700 ± 400	3110 ± 340	3200 ± 400
Wt -ch	23300 ± 500	16600 ± 500	13800 ± 400	18200 ± 500
$s - /t$ -ch	50200 ± 500	19940 ± 340	8710 ± 220	5100 ± 170
$t\bar{t}$	483000 ± 4000	260000 ± 2600	157300 ± 2000	122500 ± 1800
Multijet	11800 ± 150	1465 ± 34	711 ± 23	356 ± 18
WW	390 ± 120	280 ± 100	210 ± 90	600 ± 150
VZ	2820 ± 160	3200 ± 170	3940 ± 190	6630 ± 240
Total bkg	629400 ± 4400	348100 ± 3400	224800 ± 2600	246500 ± 2900
S/B ($\times 10^{-3}$)	1.59 ± 0.05	2.84 ± 0.09	4.51 ± 0.14	8.34 ± 0.19
$S/\sqrt{B}$	1.26 ± 0.04	1.68 ± 0.05	2.14 ± 0.06	4.14 ± 0.08
One-lepton 3-jet signal region				
Process	p_T^W [GeV]			
	90 – 120	120 – 160	160 – 200	> 200
WH	213 ± 13	212 ± 13	288 ± 15	720 ± 23
ZH	28.1 ± 1.1	21.6 ± 0.9	20.0 ± 0.9	26.5 ± 0.9
VH total	241 ± 13	234 ± 13	308 ± 15	746 ± 23
$W + l$	70 ± 70	230 ± 180	0 ± 0	180 ± 180
$W + cl$	2000 ± 600	1400 ± 400	920 ± 260	1800 ± 400
$W + cc$	4600 ± 900	3300 ± 700	4400 ± 800	11000 ± 1000
$W + bl$	800 ± 190	540 ± 150	470 ± 120	950 ± 150
$W + bb$	11000 ± 700	11000 ± 700	10300 ± 600	31500 ± 900
$Z + bb$	5200 ± 400	1730 ± 250	1230 ± 210	780 ± 180
Wt -ch	25600 ± 600	16200 ± 500	13600 ± 400	23200 ± 500
$s - /t$ -ch	29800 ± 400	11240 ± 270	4470 ± 160	3700 ± 140
$t\bar{t}$	825000 ± 5000	379900 ± 3200	221800 ± 2400	199400 ± 2300
Multijet	3710 ± 80	490 ± 20	264 ± 13	137 ± 11
WW	140 ± 70	70 ± 50	70 ± 50	280 ± 100
VZ	700 ± 70	810 ± 90	890 ± 80	1840 ± 120
Total bkg	908600 ± 4900	426900 ± 3400	258500 ± 2700	274800 ± 2800
S/B ($\times 10^{-3}$)	0.27 ± 0.02	0.55 ± 0.03	1.20 ± 0.06	2.72 ± 0.09
$S/\sqrt{B}$	0.25 ± 0.01	0.36 ± 0.02	0.61 ± 0.03	1.42 ± 0.05

Table B.4: Estimated signal and background event yields, for the one-lepton channel. The yields are for $\sqrt{s} = 14$ TeV, $\langle\mu\rangle = 140$ and an integrated luminosity of 3000 fb^{-1} , before the fit model has been applied. The uncertainties correspond to limited Monte Carlo statistics.

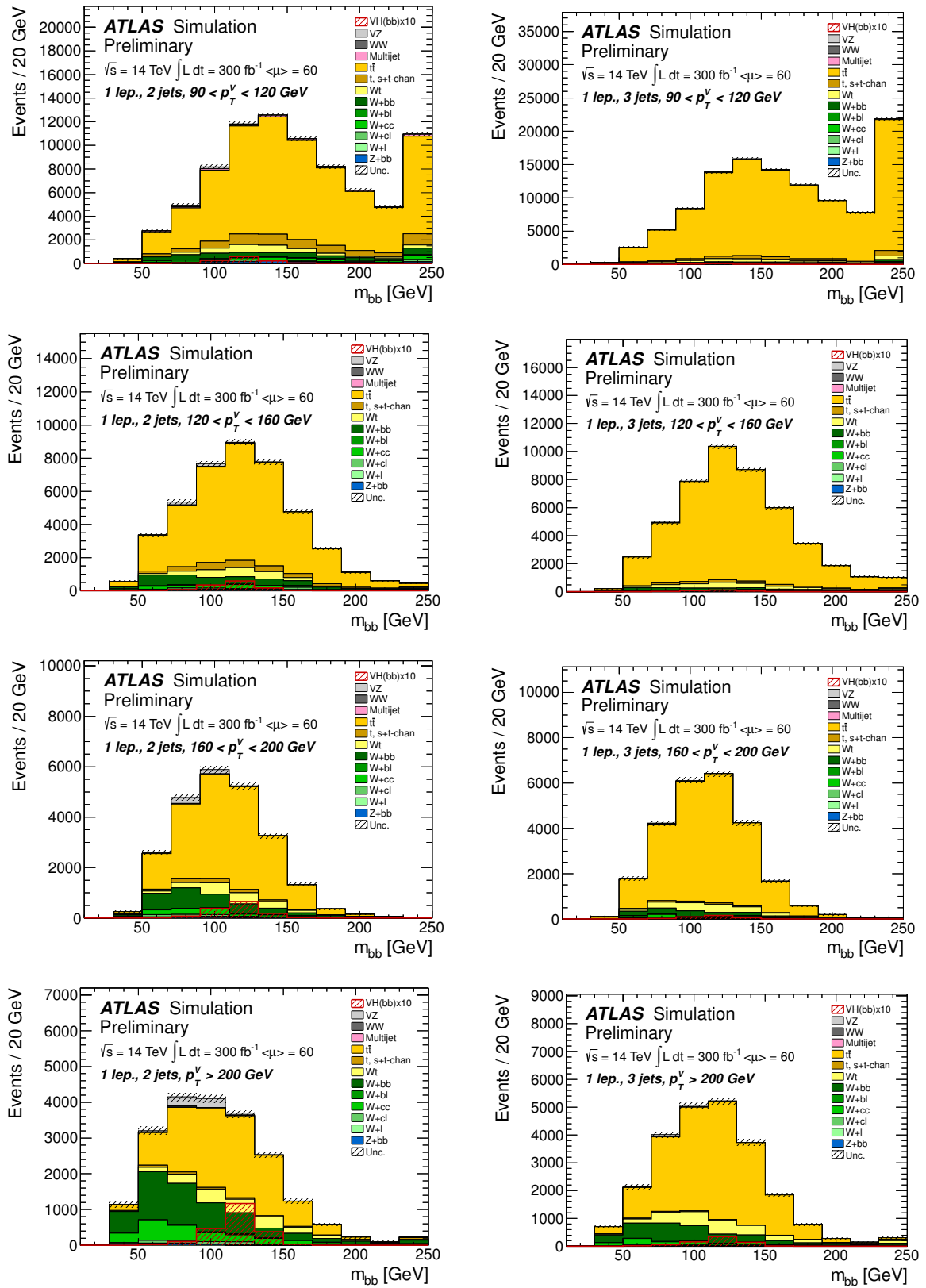


Figure B.3: m_{bb} distribution for the 2-jet (left) and 3-jet (right) bins of the one-lepton channel for $\sqrt{s} = 14$ TeV, $\langle\mu\rangle_{pu} = 60$ and an integrated luminosity of 300 fb^{-1} . The entries in overflow are included in the last bin. The Higgs boson signal cross section has been multiplied by a factor of 10. The dashed band corresponds to the statistical uncertainties only.

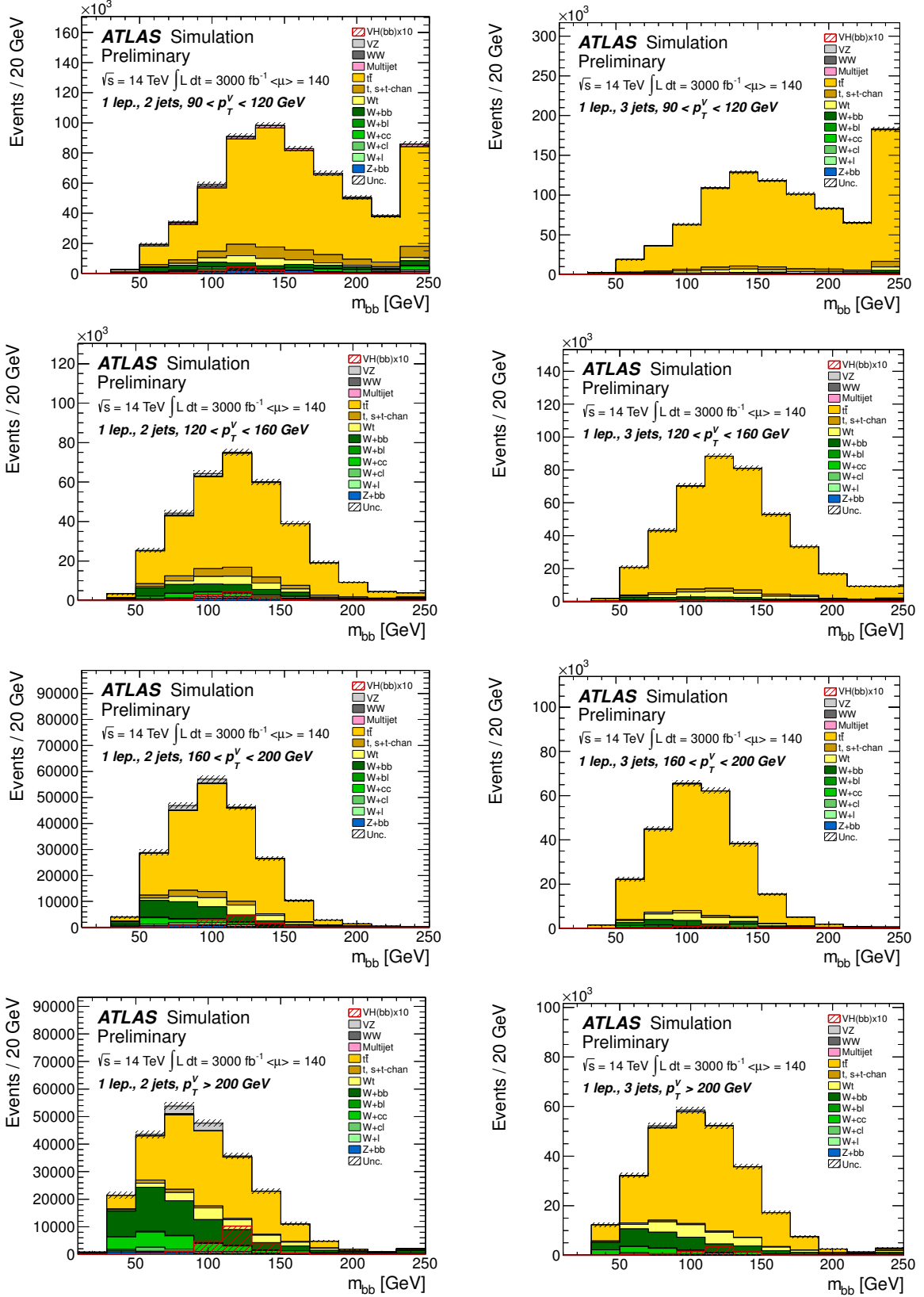


Figure B.4: m_{bb} distribution for the 2-jet (left) and 3-jet (right) bins of the one-lepton channel for $\sqrt{s} = 14$ TeV, $\langle\mu\rangle_{pu} = 140$ and an integrated luminosity of 3000 fb^{-1} . The entries in overflow are included in the last bin. The Higgs boson signal cross section has been multiplied by a factor of 10. The dashed band corresponds to the statistical uncertainties only.

B.6.1 Experimental Uncertainties

The uncertainty on the identification efficiency of electrons and muons, measured in Z +jets using the tag and probe method are of the order of 1%. Uncertainties on the isolation and energy scale of leptons are neglected due to their negligible impact on the analysis.

The uncertainties on the jet energy scale for the HL-LHC studies are based on the 8 TeV uncertainties provided for Moriond 2013 with a significantly reduced number of eight components. The Moriond uncertainties are parameterised in η and p_T and the HL-LHC uncertainties are derived by scaling the p_T range of the Moriond uncertainties by a factor of $14/8 = 1.75$ for a rough estimation of the systematic uncertainties that can be expected after multiple years of performance studies and a large amount of statistics. Due to the statistical limitation of many MC samples, the JES uncertainty is only considered in the statistical analysis for the main backgrounds such as $t\bar{t}$, $W + b\bar{b}$ and $W + c\bar{c}$ for the one-lepton channel. The component accounting for the difference in quark and gluon flavour composition has the largest impact on the fit. The uncertainties are about 5% for very central jets with $p_T = 20$ GeV and decrease down to 2% for jets with a transverse momentum of 1 TeV.

Three sources of uncertainties are considered for the resolution of the missing transverse energy and treated as uncorrelated in the statistical analysis. An uncertainty of 5% corresponds to the dependence of E_T^{miss} on the used MC generators, the possible variation of the pile-up noise threshold can be accounted for by a fixed 5% uncertainty and the uncertainty on the calibration of the energy clusters which are not associated to any reconstructed object is considered as third nuisance parameter. The systematic uncertainties related to the missing transverse energy are directly taken from the benchmark analysis.

Systematic uncertainties on the b -tagging efficiency of b -jets and on the efficiency of mistakenly identifying a jet from a c -quark or light-flavour parton as a b -jet, referred to as c -tagging efficiency and mistag rate, respectively, are considered in this analysis and correspond to those of the benchmark analysis. No dedicated systematic uncertainties on the flavour tagging efficiencies were derived for the HL-LHC analysis. Seven (six) nuisance parameters are used to model the p_T dependence of the b -tagging (c -tagging) efficiency uncertainties whereas only one nuisance parameter is used for the mistag rate uncertainty. To simplify the fit model, only the dominant flavour-tagging systematic uncertainty components are considered.

Furthermore, an uncertainty on the luminosity measurement of 2.8% is used, obtained from van der Meer scans at $\sqrt{s} = 8$ TeV [140].

B.6.2 Background Only Uncertainties

The systematic uncertainties on the overall normalisation of the cross-section as well as the relative normalisation of the backgrounds in the different analysis bins are taken from the benchmark analysis. The detailed summary of the uncertainties on the background normalisations can be found in Tab B.5. For instance, the uncertainty on the cross-section of the dominating $t\bar{t}$ background is 10% and the ratio of the number of events in the 2-jet bin with respect to that of

Background		Normalisation		
		cross-section	2-jet to 3-jet ratio	p_T^V
Diboson	WZ , WW and ZZ	$\pm 10\%$	-	-
Multijet		$\pm 10\%$	-	-
Single-top	t -channel	$\pm 4\%$	$\pm 9\%$	-
Single-top	s -channel	$\pm 4\%$	$\pm 9\%$	-
Single-top	Wt -production	$\pm 7\%$	$\pm 15\%$	-
V +jets	$V + b\bar{b}$ production	$\pm 30\%$	$\pm 20\%$	Y
V +jets	$V + c\bar{c}$ production	$\pm 30\%$	$\pm 10\%$	Y
V +jets	V +light production	$\pm 10\%$	$\pm 10\%$	Y
$t\bar{t}$		$\pm 10\%$	$\pm 5\%$	Y

Table B.5: Summary of the systematic uncertainties on the overall normalisation of the cross-sections of the various background processes. Furthermore, the relatively normalisation in the 2-jet to 3-jet analysis bins are shown and whether the number of events is allowed to vary in successive the p_T^V bins.

the 3-jet bin is allowed to vary by up to 5% compared to the MC prediction. In addition, the number of events in the various p_T^V bins are allowed to vary. Shape uncertainties on the $m_{b\bar{b}}$ are taken into account for all background sources except for the diboson and multijet background.

B.6.3 Signal Only Uncertainties

Systematic uncertainties affecting only the signal include uncertainties on the normalisation, the renormalisation and factorisation scales (0.5 – 1.5% for qq and 50% for gg induced processes) and on the PDFs (3.8% for qq and 17% for gg induced processes) [190] and are considered to be correlated qq initial states and decorrelated for $gg \rightarrow ZH$. Furthermore a relative uncertainty of 3.3% on the $H \rightarrow b\bar{b}$ branching ratio for $m_H = 125$ GeV [191] is considered.

An additional uncertainty of about 2 – 3% accounts for the p_T^V -dependent NLO electroweak corrections on the qq -initiated WH and ZH signals. The uncertainties are treated as correlated for the two channels and across the p_T^V bins.

B.6.4 Reduced Systematic Uncertainty Scenarios

Many of the systematic uncertainties considered in the HL-LHC analysis are based on the uncertainties used in the benchmark analysis. However the reduction of some sources of experimental systematic uncertainties can be expected due to the large amount of data to be collected and analysed which result in a better understanding of the uncertainties. Two different scenarios are considered based on the JES uncertainty. For the one-lepton channel, the uncertainty on the jet energy scale is scaled down to 10 (5) % of the HL-LHC JES uncertainties in scenario I (II). The JES uncertainty reduction has only little impact on the two-lepton channel due to the

lower $t\bar{t}$ background yields and is thus kept constant.

Furthermore improved analysis techniques will be used at the HL-LHC which will be discussed in Section B.8 that can result in an improved sensitivity of the analysis. Their impact is also explored in this analysis.

B.7 Fit Model

The statistical analysis employs a binned likelihood function, $\mathcal{L}(\mu, \theta)$, constructed as the product of Poisson probability terms. The inputs are 18 $m_{b\bar{b}}$ distributions for the combined one and two-lepton channel and the impact of the systematic uncertainties on the signal and background expectations is described by nuisance parameters θ . These nuisance parameters are parameterised by Gaussian or log-normal priors. The expected numbers of signal and background events in each bin are functions of θ . A signal strength parameter, μ , multiplies the expected SM Higgs boson production cross section. The test statistic q_μ is then constructed according to the profile likelihood ratio: $q_\mu = -2 \ln \left(\frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right)$, where $\hat{\mu}$ and $\hat{\theta}$ are the parameters that maximise the likelihood (with the constraint $0 \leq \hat{\mu} \leq \mu$), and $\hat{\theta}_\mu$ are the nuisance parameter values that maximise the likelihood for a given μ . This test statistic is used to obtain the median experimental sensitivity of the search using an *Asimov dataset* as defined in Ref. [158] and for exclusion intervals derived with the CLs method, defined in Ref. [157].

B.8 Analysis Improvement Assumptions

As previously discussed, two scenarios are assumed about possible reductions of the JES uncertainties. Further improvements were already used during Run-1 to improve the sensitivity of the VH search such as multivariate techniques (MVA) or jet calibrations that result in a better jet energy resolution such as global sequential calibration [192]. The global sequential calibration exploits information from the inner detector, the energy deposits in the calorimeter and the muon spectrometer to correct for fluctuations in the particle content of the hadronic shower development. These corrections do not only improve the jet energy resolution but also reduce the flavour dependence of the jet energy scale. Studies have been performed to test different discriminating kinematic variables between the WH signal and the dominating $t\bar{t}$ background however in a cut-based analysis no significant improvement could be achieved and a MVA analysis was beyond the scope of this analysis. However improvements of up to 10% in sensitivity are expected in the one-lepton channel using a MVA analysis. On the other hand, parameterisations of the jet energy resolution using the global sequential calibration were not available for the HL-LHC. Furthermore a large increase in signal sensitivity of about 35% in the one-lepton channel could be achieved using continuous b -tagging calibrations [146]. Instead of defining fixed b -tagging efficiency working points based on the output of the boosted decision tree, continuous b -tagging uses the full weight distribution of the BDT. Including control regions into the

		One-lepton	Two-lepton	One+Two-lepton
Stat-only	Significance	2.7	3.0	4.1
	$\hat{\mu}_{\text{Stats}}$ error	+0.37 - 0.37	+0.33 - 0.33	+0.25 - 0.25
Theory-only	$\hat{\mu}_{\text{Theory}}$ error	+0.08 - 0.05	+0.08 - 0.05	+0.09 - 0.06
	Significance	1.2	2.4	2.6
Scenario I	$\hat{\mu}_{\text{w/Theory}}$ error	+0.86 - 0.85	+0.44 - 0.43	+0.39 - 0.38
	$\hat{\mu}_{\text{wo/Theory}}$ error	+0.85 - 0.85	+0.43 - 0.43	+0.38 - 0.38
Scenario II	Significance	1.4	-	2.8
	$\hat{\mu}_{\text{w/Theory}}$ error	+0.71 - 0.70	-	+0.38 - 0.37
	$\hat{\mu}_{\text{wo/Theory}}$ error	+0.70 - 0.70	-	+0.37 - 0.36

Table B.6: Expected signal sensitivity as well as the precision on the signal strength measurement for $m_H = 125$ GeV for the one-lepton, two-lepton and combined searches with 300 fb^{-1} and $\langle\mu\rangle = 60$.

fit also resulted in small improvements. Based on these improvements, uncertainty reduction factors were considered and included in the HL-LHC analysis. The resulting expected signal sensitivities are compared to the nominal HL-LHC analysis in Section B.9.

B.9 Results

The expected sensitivity and the precision on the signal strength μ for a Higgs boson with $m_H = 125$ GeV and production rates as predicted in the SM, i.e. $\mu = 1$, were obtained by performing a profile likelihood fit with an Asimov dataset. The results are summarised in Tab. B.6 and B.7 for a dataset corresponding to 300 fb^{-1} and 3000 fb^{-1} at $\sqrt{s} = 14$ TeV respectively. The dominating systematic uncertainties for the combined analysis for 300 fb^{-1} correspond to the jet energy scale systematic and the modelling of the dominating backgrounds in the WH analysis ($t\bar{t}$ and $W + b\bar{b}$), while for 3000 fb^{-1} they correspond to theory signal uncertainties (PDF and signal acceptance due to variations of scale) and $t\bar{t}$ modelling. Furthermore, the expected sensitivity and signal strength precision are indicated for the two scenarios in which the JES systematic uncertainties are scaled down to 10% and 5% of their nominal values. Assuming that the systematic uncertainties can be reduced down to 10% of what was expected for the HL-LHC analysis, the expected sensitivity for the VH , $H \rightarrow b\bar{b}$ process is 2.6σ (5.9σ) for the combined one- and two-lepton channel and the signal strength with a precision of $\Delta\hat{\mu}_{-0.38}^{+0.39}$ ($\Delta\hat{\mu}_{-0.19}^{+0.19}$) for an integrated luminosity of 300 fb^{-1} (3000 fb^{-1}) and an average interaction per bunch-crossing of 60 (140). The expected significances assuming in addition the improved analysis techniques discussed in Section B.8 are given in Tab. B.8 and B.9 for the two luminosity scenarios, resulting in a sensitivity of 3.9σ and 8.8σ for an integrated luminosity of 300 fb^{-1} and 3000 fb^{-1} , respectively.

		One-lepton	Two-lepton	One+Two-lepton
Stat-only	Significance	7.7	7.5	10.7
	$\hat{\mu}_{\text{Stats}}$ error	+0.13 – 0.13	+0.14 – 0.13	+0.09 – 0.09
Theory-only	$\hat{\mu}_{\text{Theory}}$ error	+0.09 – 0.07	+0.07 – 0.08	+0.07 – 0.07
Scenario I	Significance	1.8	5.6	5.9
	$\hat{\mu}_{\text{w/Theory}}$ error	+0.56 – 0.54	+0.20 – 0.19	+0.19 – 0.19
	$\hat{\mu}_{\text{wo/Theory}}$ error	+0.54 – 0.54	+0.18 – 0.18	+0.18 – 0.17
Scenario II	Significance	3.2	-	6.4
	$\hat{\mu}_{\text{w/Theory}}$ error	+0.33 – 0.32	-	+0.18 – 0.17
	$\hat{\mu}_{\text{wo/Theory}}$ error	+0.32 – 0.32	-	+0.16 – 0.16

Table B.7: Expected signal sensitivity as well as the precision on the signal strength measurement for $m_H = 125$ GeV for the one-lepton, two-lepton and combined searches with 3000 fb^{-1} with $\langle\mu\rangle = 140$.

		One-lepton	Two-lepton	One+Two-lepton
Stat-only	Significance	5.5	4.6	7.1
	$\hat{\mu}_{\text{Stats}}$ error	+0.18 – 0.18	+0.23 – 0.22	+0.14 – 0.14
Theory-only	$\hat{\mu}_{\text{Theory}}$ error	+0.08 – 0.05	+0.08 – 0.06	+0.09 – 0.06
Scenario I	Significance	1.8	3.5	3.9
	$\hat{\mu}_{\text{w/Theory}}$ error	+0.57 – 0.57	+0.30 – 0.29	+0.27 – 0.26
	$\hat{\mu}_{\text{wo/Theory}}$ error	+0.56 – 0.57	+0.29 – 0.29	+0.26 – 0.26
Scenario II	Significance	2.1	-	4.1
	$\hat{\mu}_{\text{w/Theory}}$ error	+0.48 – 0.47	-	+0.26 – 0.25
	$\hat{\mu}_{\text{wo/Theory}}$ error	+0.46 – 0.46	-	+0.25 – 0.24

Table B.8: Expected signal sensitivity as well as the precision on the signal strength measurement for $m_H = 125$ GeV for the one-lepton, two-lepton and combined searches with 300 fb^{-1} and $\langle\mu\rangle = 60$ after including the perspective of a more performant analysis.

		One-lepton	Two-lepton	One+Two-lepton
Stat-only	Significance	15.4	11.3	19.1
	$\hat{\mu}_{\text{Stats}}$ error	+0.07 – 0.06	+0.09 – 0.09	+0.05 – 0.05
Theory-only	$\hat{\mu}_{\text{Theory}}$ error	+0.09 – 0.07	+0.07 – 0.08	+0.07 – 0.07
	Significance	2.7	8.4	8.8
Scenario I	$\hat{\mu}_{\text{w/Theory}}$ error	+0.37 – 0.36	+0.15 – 0.15	+0.14 – 0.14
	$\hat{\mu}_{\text{wo/Theory}}$ error	+0.36 – 0.36	+0.14 – 0.12	+0.12 – 0.12
Scenario II	Significance	4.7	-	9.6
	$\hat{\mu}_{\text{w/Theory}}$ error	+0.23 – 0.22	-	+0.13 – 0.13
	$\hat{\mu}_{\text{wo/Theory}}$ error	+0.21 – 0.21	-	+0.11 – 0.11

Table B.9: Expected signal sensitivity as well as the precision on the signal strength measurement for $m_H = 125$ GeV for the one-lepton, two-lepton and combined searches with 3000 fb^{-1} with $\langle\mu\rangle = 140$ after including the perspective of a more performant analysis.

B.10 Discussion and Outlook

A study of Higgs boson production in association with leptonically decaying W and Z bosons was presented in this chapter based on particle-level information with applied parameterisations of the expected detector response under the HL conditions. The lack of full event simulation samples is one of the main drawbacks of this analysis as the used MC samples do not include any additional pile-up interactions. The effects of pile-up on the selected physics objects are included in the efficiency and resolution parameterisations, however other pile-up effects such as jets originating from pile-up vertices are not included in this analysis and thus a realistic estimation of the sensitivity of this study is not possible.

A selection of improvements to the analysis has been already discussed in Section B.8 but the large dataset that will be collected in Run-II will allow for further improvements in the reconstruction of physics objects and will enable the reduction of systematic uncertainties. Especially the identification of b -tagged jets plays a major role in this analysis and a significantly better performance of b -tagging algorithms could be achieved already in Run-II with respect to Run-I due to the addition of the IBL. Depending on the final design of the inner tracking detector of the ATLAS detector at the HL-LHC, the b -jet identification can be further improved.

The expected improvements of a multivariate analysis over a cut-based analysis were only considered here as a reduction factor on the total systematic uncertainty. However a MVA-based analysis would also allow for a better discrimination of the signal process from the dominating background by exploiting the correlations of the different input variables and result in a better sensitivity.

Further improvements that might increase the sensitivity of the VH analysis rely on the choice of the bins, used in this analysis. The highest bin in p_{T}^V have the highest sensitivity, therefore the

higher centre-of-mass energy of the HL-LHC compared to the 8 TeV of the benchmark analysis would allow for the inclusion of further bins in p_{T}^V in the analysis.

In the high p_{T}^V regime, $p_{\text{T}} > 300$ GeV, the decay products of the Higgs-boson will be more collimated and could be reconstructed in one single large- R jets. Substructure information could then be used to identify whether the jet is compatible with the Higgs-boson and to suppress the $t\bar{t}$ background. The identification of boosted Higgs-bosons, decaying into $b\bar{b}$, was already studied in Run-II [193]. For jets with a transverse momentum between 400 and 600 GeV, a Higgs-jet signal efficiency of 45% can be obtained with a rejection factor of 60 for hadronic top quark jets using a loose mass window around the Higgs-boson mass and by requiring that the two subjets within the large- R jets are identified as b -jets (77% b -tagging working point).

Also, the inclusion of the $ZH \rightarrow \nu\bar{\nu}b\bar{b}$ channel will contribute to an enhanced sensitivity.

The described improvements should enable an observation of the $VH, H \rightarrow b\bar{b}$ process in the dataset corresponding to 300 fb^{-1} to gain further insights in the electroweak symmetry breaking mechanism.

Bibliography

- [1] D. Griffiths, *Introduction to elementary particles*, Weinheim, Germany: Wiley-VCH (2008) .
- [2] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course In Modern Particle Physics*, John Wiley and Sons (1984) .
- [3] S. L. Glashow, *Partial Symmetries of Weak Interactions*, [Nucl. Phys. **22** \(1961\) 579–588](#).
- [4] S. Weinberg, *A Model of Leptons*, [Phys. Rev. Lett. **19** \(1967\) 1264–1266](#).
- [5] A. Salam and J. C. Ward, *Weak and electromagnetic interactions*, [Nuovo Cim. **11** \(1959\) 568–577](#).
- [6] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, [Phys. Rev. Lett. **13** \(1964\) 321–323](#).
- [7] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, [Phys. Lett. **12** \(1964\) 132–133](#).
- [8] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, [Phys. Rev. Lett. **13** \(1964\) 508–509](#).
- [9] G. Guralnik, C. Hagen, and T. Kibble, *Global conservation laws and massless particles*, [Phys. Rev. Lett. **13** \(1964\) 585–587](#).
- [10] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, [Phys. Rev. **145** \(1966\) 1156–1163](#).
- [11] T. Kibble, *Symmetry breaking in non-Abelian gauge theories*, [Phys. Rev. **155** \(1967\) 1554–1561](#).
- [12] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, [Phys. Lett. **B716** \(2012\) 1–29](#), [arXiv:1207.7214 \[hep-ex\]](#).
- [13] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, [Phys. Lett. **B716** \(2012\) 30–61](#), [arXiv:1207.7235 \[hep-ex\]](#).

- [14] SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3 Collaboration, S. Schael et al., *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006) 257–454, [arXiv:hep-ex/0509008 \[hep-ex\]](#).
- [15] Particle Data Group Collaboration, K. A. Olive et al., *Review of Particle Physics*, *Chin. Phys.* **C38** (2014) 090001.
- [16] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, *Phys. Rev. Lett.* **30** (1973) 1346–1349.
- [17] D. J. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. 1*, *Phys. Rev.* **D8** (1973) 3633–3652.
- [18] D. J. Gross and F. Wilczek, *ASYMPTOTICALLY FREE GAUGE THEORIES. 2.*, *Phys. Rev.* **D9** (1974) 980–993.
- [19] F. Ceradini et al., *Multihadron production in e^+e^- collisions up to 3 GeV total center-of-mass energy*, *Phys. Lett.* **B47** (1973) 80–84.
- [20] ATLAS and CMS Collaboration, *Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments*, *Phys. Rev. Lett.* **114** (2015) 191803, [arXiv:1503.07589 \[hep-ex\]](#).
- [21] ATLAS Collaboration, *Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector*, *Eur. Phys. J.* **C75** (2015) no. 10, 476, [arXiv:1506.05669 \[hep-ex\]](#). [Erratum: *Eur. Phys. J.*C76,no.3,152(2016)].
- [22] ATLAS Collaboration, *Summary plots from the ATLAS Standard Model physics group*, <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/index.html>.
- [23] Super-Kamiokande Collaboration, Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, *Phys. Rev. Lett.* **81** (1998) 1562–1567, [arXiv:hep-ex/9807003 \[hep-ex\]](#).
- [24] SNO Collaboration, Q. R. Ahmad et al., *Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory*, *Phys. Rev. Lett.* **87** (2001) 071301, [arXiv:nuc1-ex/0106015 \[nuc1-ex\]](#).
- [25] W. H. Furry, *On transition probabilities in double beta-disintegration*, *Phys. Rev.* **56** (1939) 1184–1193.
- [26] G. Bertone, D. Hooper, and J. Silk, *Particle dark matter: Evidence, candidates and constraints*, *Phys. Rept.* **405** (2005) 279–390, [arXiv:hep-ph/0404175 \[hep-ph\]](#).

-
- [27] H. Georgi and S. L. Glashow, *Unity of All Elementary Particle Forces*, [Phys. Rev. Lett. **32** \(1974\) 438–441](#).
- [28] H. Fritzsch and P. Minkowski, *Unified Interactions of Leptons and Hadrons*, [Annals Phys. **93** \(1975\) 193–266](#).
- [29] S. P. Martin, *A Supersymmetry primer*, [arXiv:hep-ph/9709356 \[hep-ph\]](#). [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [30] E. Eichten and K. Lane, *Low-scale technicolor at the Tevatron and LHC*, [Phys. Lett. **B669** \(2008\) 235–238](#), [arXiv:0706.2339 \[hep-ph\]](#).
- [31] S. Catterall, L. Del Debbio, J. Giedt, and L. Keegan, *MCRG Minimal Walking Technicolor*, [Phys. Rev. **D85** \(2012\) 094501](#), [arXiv:1108.3794 \[hep-ph\]](#).
- [32] J. R. Andersen et al., *Discovering Technicolor*, [Eur. Phys. J. Plus **126** \(2011\) 81](#), [arXiv:1104.1255 \[hep-ph\]](#).
- [33] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, [Phys. Rev. Lett. **83** \(1999\) 3370–3373](#), [arXiv:hep-ph/9905221 \[hep-ph\]](#).
- [34] L. Randall and R. Sundrum, *An Alternative to compactification*, [Phys. Rev. Lett. **83** \(1999\) 4690–4693](#), [arXiv:hep-th/9906064 \[hep-th\]](#).
- [35] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo, *Experimental probes of localized gravity: On and off the wall*, [Phys. Rev. **D63** \(2001\) 075004](#), [arXiv:hep-ph/0006041 \[hep-ph\]](#).
- [36] J. C. Pati and A. Salam, *Lepton Number as the Fourth Color*, [Phys. Rev. **D10** \(1974\) 275–289](#). [Erratum: Phys. Rev.D11,703(1975)].
- [37] G. Altarelli, B. Mele, and M. Ruiz-Altaba, *Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders*, [Z. Phys. **C45** \(1989\) 109](#). [Erratum: Z. Phys.C47,676(1990)].
- [38] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, *Heavy Vector Triplets: Bridging Theory and Data*, [JHEP **09** \(2014\) 060](#), [arXiv:1402.4431 \[hep-ph\]](#).
- [39] L. Randall and M. D. Schwartz, *Quantum field theory and unification in AdS_5* , [JHEP **11** \(2001\) 003](#), [arXiv:hep-th/0108114 \[hep-th\]](#).
- [40] T. Kaluza, *On the Problem of Unity in Physics*, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) **1921** (1921) 966–972.
- [41] O. Klein, *Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English)*, [Z. Phys. **37** \(1926\) 895–906](#). [Surveys High Energ. Phys.5,241(1986)].
- [42] CDF Collaboration, R. Blair et al., *The CDF-II detector: Technical design report*, FERMILAB-DESIGN-1996-01, 1996. <https://cds.cern.ch/record/1478626>.
-

- [43] D0 Collaboration, B. Pifer et al., *An Experiment at D0 to Study anti-Proton - Proton Collisions at 2-TeV: Design Report*, . <https://cds.cern.ch/record/1478625>.
- [44] CDF Collaboration, T. Aaltonen et al., *Search for diboson (W^+W^- or $W^\pm Z^0$) resonances in electron, missing E_T and two jets final state*, CDF note 9730 (2009) CDF note 9730 (2009).
http://www-cdf.fnal.gov/physics/exotic/r2a/20100114.WW_WZ_resonance/.
- [45] CDF Collaboration, T. Aaltonen et al., *Search for high-mass resonances decaying into ZZ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev.* **D85** (2012) 012008, [arXiv:1111.3432](https://arxiv.org/abs/1111.3432) [hep-ex].
- [46] D0 Collaboration, V. M. Abazov et al., *Search for resonant WW and WZ production in $p\bar{p}$ collisions at $s = 1.96$ TeV*, *Phys. Rev. Lett.* **107** (2011) 011801, [arXiv:1011.6278](https://arxiv.org/abs/1011.6278) [hep-ex].
- [47] ATLAS Collaboration, *Search for resonant diboson production in the $WW/WZ \rightarrow \ell\nu jj$ decay channels with the ATLAS detector at $\sqrt{s} = 7$ TeV*, *Phys. Rev.* **D87** (2013) no. 11, 112006, [arXiv:1305.0125](https://arxiv.org/abs/1305.0125) [hep-ex].
- [48] CMS Collaboration, *Search for exotic resonances decaying into WZ/ZZ in pp collisions at $\sqrt{s} = 7$ TeV*, *JHEP* **02** (2013) 036, [arXiv:1211.5779](https://arxiv.org/abs/1211.5779) [hep-ex].
- [49] CMS Collaboration, *Search for heavy resonances in the W/Z -tagged dijet mass spectrum in pp collisions at 7 TeV*, *Phys. Lett.* **B723** (2013) 280–301, [arXiv:1212.1910](https://arxiv.org/abs/1212.1910) [hep-ex].
- [50] O. S. Bruning et al., *LHC Design Report. 1. The LHC Main Ring*, CERN-2004-003 (2004) .
- [51] O. S. Buning et al., *LHC Design Report. 2. The LHC infrastructure and general services*, CERN-2004-003 (2004) .
- [52] M. Benedikt et al., *LHC Design Report. 3. The LHC injector chain*, CERN-2004-003 (2004) .
- [53] C. Lefevre, *LHC: the guide*, CERN-Brochure-2009-003-Eng. March 2009.
- [54] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [55] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.
- [56] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002.

-
- [57] LHCb Collaboration, *The LHCb Detector at the LHC*, **JINST** **3** (2008) S08005.
- [58] ATLAS Collaboration, *Luminosity Public Results*,
<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>.
- [59] ATLAS Collaboration, *Luminosity Public Results Run2*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [60] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013, 2010. <https://cds.cern.ch/record/1291633>.
- [61] J. C. Collins, D. E. Soper, and G. F. Sterman, *Factorization of Hard Processes in QCD*, *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, [arXiv:hep-ph/0409313 \[hep-ph\]](#).
- [62] M. L. Mangano et al., *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, **JHEP** **07** (2003) 001, [arXiv:hep-ph/0206293 \[hep-ph\]](#).
- [63] J. Alwall et al., *MadGraph 5 : Going Beyond*, **JHEP** **06** (2011) 128, [arXiv:1106.0522 \[hep-ph\]](#).
- [64] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, **JHEP** **05** (2006) 026, [arXiv:hep-ph/0603175 \[hep-ph\]](#).
- [65] G. Corcella et al., *HERWIG 6.5 release note*, [arXiv:hep-ph/0210213 \[hep-ph\]](#).
- [66] T. Gleisberg et al., *Event generation with SHERPA 1.1*, **JHEP** **02** (2009) 007, [arXiv:0811.4622 \[hep-ph\]](#).
- [67] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*, **Nucl. Instrum. Meth.** **A506** (2003) 250–303.
- [68] ATLAS Collaboration, *The ATLAS simulation infrastructure*, **Eur. Phys. J. C** **70** (2010) 823, [arXiv:1005.4568 \[physics.ins-det\]](#).
- [69] W. J. Stirling, *PARTON LUMINOSITY AND CROSS SECTION PLOTS*,
<http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>.
- [70] ATLAS Collaboration, *Performance of jet substructure techniques for large- R jets in proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, **JHEP** **09** (2013) 076, [arXiv:1306.4945 \[hep-ex\]](#).
- [71] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, [arXiv:1603.02934 \[hep-ex\]](#).
- [72] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV*, **Eur. Phys. J. C** **73** (2013) 2304, [arXiv:1112.6426 \[hep-ex\]](#).
-

- [73] G. P. Salam, *Towards Jetography*, *Eur. Phys. J.* **C67** (2010) 637–686, [arXiv:0906.1833 \[hep-ph\]](#).
- [74] J. E. Huth et al., *Toward a standardization of jet definitions*, FERMILAB-CONF-90-249-E (1990) .
http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-90-249.
- [75] JADE Collaboration, W. Bartel et al., *Experimental Studies on Multi-Jet Production in $e^+ e^-$ Annihilation at PETRA Energies*, *Z. Phys.* **C33** (1986) 23.
- [76] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D 48** (1993) 3160, [arXiv:hep-ph/9305266 \[hep-ph\]](#).
- [77] S. Catani et al., *Longitudinally invariant $k_\perp$ clustering algorithms for hadron hadron collisions*, *Nucl. Phys.* **B 406** (1993) 187.
- [78] Y. L. Dokshitzer et al., *Better jet clustering algorithms*, *JHEP* **08** (1997) 001, [arXiv:hep-ph/9707323 \[hep-ph\]](#).
- [79] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_t jet clustering algorithm*, *JHEP* **04** (2008) 063, [arXiv:0802.1189 \[hep-ph\]](#).
- [80] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, *Towards an understanding of jet substructure*, *JHEP* **09** (2013) 029, [arXiv:1307.0007 \[hep-ph\]](#).
- [81] M. Dasgupta, *Jet masses for high p_T QCD jets at all orders*, *PoS DIS2010* (2010) 104, [arXiv:1006.5679 \[hep-ph\]](#).
- [82] D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, *JHEP* **02** (2010) 084, [arXiv:0912.1342 \[hep-ph\]](#).
- [83] J. M. Butterworth et al., *Jet substructure as a new Higgs search channel at the LHC*, *Phys. Rev. Lett.* **100** (2008) 242001, [arXiv:0802.2470 \[hep-ph\]](#).
- [84] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys. Rev.* **D 81** (2010) 094023, [arXiv:0912.0033 \[hep-ph\]](#).
- [85] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *Soft drop*, *JHEP* **05** (2014) 146, [arXiv:1402.2657 \[hep-ph\]](#).
- [86] ATLAS Collaboration, *Identification of boosted, hadronically decaying W bosons and comparisons with ATLAS data taken at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J.* **C76** (2016) no. 3, 154, [arXiv:1510.05821 \[hep-ex\]](#).

-
- [87] ATLAS Collaboration, *Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, **JHEP** **12** (2015) 055, [arXiv:1506.00962 \[hep-ex\]](#).
- [88] ATLAS Collaboration, *Search for resonant diboson production in the $\ell\ell q\bar{q}$ final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, **Eur. Phys. J.** **C75** (2015) 69, [arXiv:1409.6190 \[hep-ex\]](#).
- [89] ATLAS Collaboration, *Search for production of WW/WZ resonances decaying to a lepton, neutrino and jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, **Eur. Phys. J.** **C75** (2015) no. 5, 209, [arXiv:1503.04677 \[hep-ex\]](#). [Erratum: Eur. Phys. J.C75,370(2015)].
- [90] CMS Collaboration, *Identification techniques for highly boosted W bosons that decay into hadrons*, **JHEP** **12** (2014) 017, [arXiv:1410.4227 \[hep-ex\]](#).
- [91] CMS Collaboration, *Search for BSM $t\bar{t}$ Production in the Boosted All-Hadronic Final State*, CMS-PAS-EXO-11-006, 2011. <https://cds.cern.ch/record/1370237>.
- [92] CMS Collaboration, *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*, CMS-PAS-PFT-09-001, 2009. <https://cds.cern.ch/record/1194487>.
- [93] CMS Collaboration, *Jet Performance in pp Collisions at 7 TeV*, CMS-PAS-JME-10-003, 2010. <https://cds.cern.ch/record/1279362>.
- [94] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, **JHEP** **03** (2011) 015, [arXiv:1011.2268 \[hep-ph\]](#).
- [95] A. J. Larkoski, D. Neill, and J. Thaler, *Jet Shapes with the Broadening Axis*, **JHEP** **04** (2014) 017, [arXiv:1401.2158 \[hep-ph\]](#).
- [96] A. J. Larkoski, G. P. Salam, and J. Thaler, *Energy Correlation Functions for Jet Substructure*, **JHEP** **06** (2013) 108, [arXiv:1305.0007 \[hep-ph\]](#).
- [97] A. J. Larkoski, I. Mount, and D. Neill, *Power Counting to Better Jet Observables*, **JHEP** **12** (2014) 009, [arXiv:1409.6298 \[hep-ph\]](#).
- [98] A. J. Larkoski, I. Mout, and D. Neill, *Analytic Boosted Boson Discrimination*, [arXiv:1507.03018 \[hep-ph\]](#).
- [99] ATLAS Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*, ATLAS-CONF-2015-037, 2015. <http://cdsweb.cern.ch/record/2044941>.
-

- [100] ATLAS Collaboration, *Determination of the jet energy scale and resolution at ATLAS using Z/γ -jet events in data at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2015-057, 2015.
<http://cdsweb.cern.ch/record/2059846>.
- [101] ATLAS Collaboration, *Measurement of dijet azimuthal decorrelations in pp collisions at $\sqrt{s} = 7$ TeV*, *Phys. Rev. Lett.* **106** (2011) 172002, [arXiv:1102.2696 \[hep-ex\]](#).
- [102] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852–867, [arXiv:0710.3820 \[hep-ph\]](#).
- [103] M. Bahr et al., *Herwig++ physics and manual*, *Eur. Phys. J. C* **58** (2008) 639, [arXiv:0803.0883 \[hep-ph\]](#).
- [104] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, [arXiv:1510.03823 \[hep-ex\]](#).
- [105] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in High Pile-Up LHC Environment*, ATLAS-CONF-2012-042, 2012.
<http://cdsweb.cern.ch/record/1435196>.
- [106] ATLAS Collaboration, *Large- R (anti- k_t $R=1.0$) JES and JMS for 2012 data*,
<http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2016-001/>.
- [107] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-015, 2015.
<https://cds.cern.ch/record/2037613>.
- [108] ATLAS Collaboration, *Identification of Boosted, Hadronically-Decaying W and Z Bosons in $\sqrt{s} = 13$ TeV Monte Carlo Simulations for ATLAS*, ATL-PHYS-PUB-2015-033, 2015. <http://cdsweb.cern.ch/record/2041461>.
- [109] A. Ribon et al., *Status of Geant4 hadronic physics for the simulation of LHC experiments at the start of the LHC physics program*, CERN-LCGAPP-2010-02, 2010.
<http://lcgapp.cern.ch/project/docs/noteStatusHadronic2010.pdf>.
- [110] H. W. Bertini, *Intranuclear-cascade calculation of the secondary nucleon spectra from nucleon-nucleus interactions in the energy range 340 to 2900 MeV and comparisons with experiment*, *Phys. Rev. A* **188** (1969) 1711–1730.
- [111] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist, *A model for low- p_T hadronic reactions with generalizations to hadron-nucleus and nucleus-nucleus collisions*, *Nucl. Phys. B* **281** (1987) 289–309.
- [112] G. Folger and J. Wellisch, *String parton models in GEANT4*, [arXiv:0306007 \[nucl-th\]](#).

-
- [113] H. Abreu et al., *Performance of the electronic readout of the ATLAS liquid argon calorimeters*, *JINST* **5** (2010) P09003.
- [114] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003, 2012. <http://cds.cern.ch/record/1474107>.
- [115] H.-L. Lai et al., *New parton distributions for collider physics*, *Phys. Rev. D* **82** (2010) 074024, [arXiv:1007.2241](https://arxiv.org/abs/1007.2241) [hep-ph].
- [116] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014. <http://cdsweb.cern.ch/record/1966419>.
- [117] R. D. Ball et al., *Parton distributions with LHC data*, *Nucl. Phys. B* **867** (2013) 244–289, [arXiv:1207.1303](https://arxiv.org/abs/1207.1303) [hep-ph].
- [118] S. Gieseke, C. Rohr, and A. Siodmok, *Colour reconnections in Herwig++*, *Eur. Phys. J. C* **72** (2012) 2225, [arXiv:1206.0041](https://arxiv.org/abs/1206.0041) [hep-ph].
- [119] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, *Phys. Rev. D* **78** (2008) 013004, [arXiv:0802.0007](https://arxiv.org/abs/0802.0007) [hep-ph].
- [120] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-051, 2015. <http://cdsweb.cern.ch/record/2110140>.
- [121] ATLAS Collaboration, *Large- R jet p_T and mass scale uncertainties using in-situ track-based measurements*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissApproved2014LargeRJES>.
- [122] ATLAS Collaboration, *Measurement of large radius jet mass reconstruction performance at $\sqrt{s} = 8$ TeV using the ATLAS detector*, ATLAS-CONF-2016-008, 2016. <https://cds.cern.ch/record/2139642>.
- [123] ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*, ATLAS-CONF-2016-035, 2016. <https://cds.cern.ch/record/2200211>.
- [124] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett. B* **659** (2008) 119, [arXiv:0707.1378](https://arxiv.org/abs/0707.1378) [hep-ph].
- [125] G. Soyez et al., *Pileup subtraction for jet shapes*, *Phys. Rev. Lett.* **110** (2013) 162001, [arXiv:1211.2811](https://arxiv.org/abs/1211.2811) [hep-ph].
- [126] M. Cacciari, G. P. Salam, and G. Soyez, *SoftKiller, a particle-level pileup removal method*, *Eur. Phys. J. C* **75** (2015) 59, [arXiv:1407.0408](https://arxiv.org/abs/1407.0408) [hep-ph].
- [127] P. Berta, M. Spouta, D. W. Miller, and R. Leitner, *Particle-level pileup subtraction for jets and jet shapes*, *JHEP* **06** (2014) 092, [arXiv:1403.3108](https://arxiv.org/abs/1403.3108) [hep-ex].
-

- [128] D. Bertolini, P. Harris, M. Low, and N. Tran, *Pileup Per Particle Identification*, *JHEP* **10** (2014) 59, [arXiv:1407.6013 \[hep-ph\]](#).
- [129] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) 005, [arXiv:0802.1188 \[hep-ph\]](#).
- [130] ATLAS Collaboration, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J. C* **75** (2015) 17, [arXiv:1406.0076 \[hep-ex\]](#).
- [131] G. Watt and R. Thorne, *Study of Monte Carlo approach to experimental uncertainty propagation with MSTW 2008 PDFs*, *JHEP* **08** (2012) 052, [arXiv:1205.4024 \[hep-ph\]](#).
- [132] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004) 040, [arXiv:hep-ph/0409146 \[hep-ph\]](#).
- [133] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, [arXiv:0709.2092 \[hep-ph\]](#).
- [134] ATLAS Collaboration, *A new method to distinguish hadronically decaying boosted Z bosons from W bosons using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) no. 5, 238, [arXiv:1509.04939 \[hep-ex\]](#).
- [135] D. Krohn, J. Thaler, and L.-T. Wang, *Jets with Variable R* , *JHEP* **06** (2009) 059, [arXiv:0903.0392 \[hep-ph\]](#).
- [136] ATLAS Collaboration, *Boosted Object Tagging with Variable- R Jets in the ATLAS Detector*, ATL-PHYS-PUB-2016-013, 2016. <https://cds.cern.ch/record/2199360>.
- [137] D. E. Soper and M. Spannowsky, *Finding physics signals with shower deconstruction*, *Phys. Rev. D* **84** (2011) 074002, [arXiv:1102.3480 \[hep-ph\]](#).
- [138] ATLAS Collaboration, *Performance of shower deconstruction in ATLAS*, ATLAS-CONF-2014-003, 2014. <http://cdsweb.cern.ch/record/1648661>.
- [139] A. Belyaev, N. D. Christensen, and A. Pukhov, *CalcHEP 3.4 for collider physics within and beyond the Standard Model*, *Comput.Phys.Commun.* **184** (2013) 1729–1769, [arXiv:1207.6082 \[hep-ph\]](#).
- [140] ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, *Eur. Phys. J. C* **73** (2013) 2518, [arXiv:1302.4393 \[hep-ex\]](#).
- [141] ATLAS Collaboration, *Search for WW/WZ resonance production in the $\ell\nu qq$ final state at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC*, ATLAS-CONF-2015-075, 2015. <http://cdsweb.cern.ch/record/2114847>.

-
- [142] ATLAS Collaboration, *Search for diboson resonances in the $\ell\ell q\bar{q}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2015-071, 2015. <http://cdsweb.cern.ch/record/2114843>.
- [143] ATLAS Collaboration, *Search for diboson resonances in the $\nu\nu q\bar{q}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2015-068, 2015. <http://cdsweb.cern.ch/record/2114840>.
- [144] ATLAS Collaboration, *Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Eur. Phys. J.* **C74** (2014) no. 8, 3023, [arXiv:1405.6583](https://arxiv.org/abs/1405.6583) [hep-ex].
- [145] ATLAS Collaboration, *Expected Performance of Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-035, 2015. <http://cdsweb.cern.ch/record/2042155>.
- [146] ATLAS Collaboration, *Performance of b -Jet Identification in the ATLAS Experiment*, *JINST* **11** (2016) no. 04, P04008, [arXiv:1512.01094](https://arxiv.org/abs/1512.01094) [hep-ex].
- [147] ATLAS Collaboration, *Expected performance of the ATLAS b -tagging algorithms in Run-2*, ATL-PHYS-PUB-2015-022, 2015. <http://cdsweb.cern.ch/record/2037697>.
- [148] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, *Eur. Phys. J.* **C74** (2014) no. 10, 3071, [arXiv:1407.5063](https://arxiv.org/abs/1407.5063) [hep-ex].
- [149] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data*, ATLAS-CONF-2016-024, 2016. <https://cds.cern.ch/record/2157687>.
- [150] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*, *Eur. Phys. J.* **C74** (2014) no. 11, 3130, [arXiv:1407.3935](https://arxiv.org/abs/1407.3935) [hep-ex].
- [151] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV*, *Eur. Phys. J.* **C76** (2016) no. 5, 292, [arXiv:1603.05598](https://arxiv.org/abs/1603.05598) [hep-ex].
- [152] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in Proton-Proton Collisions at 7 TeV with ATLAS*, *Eur. Phys. J.* **C72** (2012) 1844, [arXiv:1108.5602](https://arxiv.org/abs/1108.5602) [hep-ex].
- [153] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton-proton collisions at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-027, 2015. <http://cdsweb.cern.ch/record/2037904>.
-

- [154] G. Choudalakis and D. Casadei, *Plotting the differences between data and expectation*, *The European Physical Journal Plus* **127** (2012) no. 2, 1–11.
- [155] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Annals Math. Statist.* **9** (1938) no. 1, 60–62.
- [156] ATLAS Collaboration, *Jet mass and substructure of inclusive jets in $\sqrt{s} = 7$ TeV pp collisions with the ATLAS experiment*, *JHEP* **05** (2012) 128, [arXiv:1203.4606 \[hep-ex\]](#).
- [157] A. L. Read, *Presentation of search results: The $CL(s)$ technique*, *J.Phys.* **G28** (2002) 2693–2704.
- [158] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, *Eur.Phys.J.* **C71** (2011) 1554, [arXiv:1007.1727 \[physics.data-an\]](#).
- [159] ATLAS Collaboration, *Combination of searches for WW, WZ, and ZZ resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Phys. Lett.* **B755** (2016) 285–305, [arXiv:1512.05099 \[hep-ex\]](#).
- [160] ATLAS Collaboration, *Search for WZ resonances in the fully leptonic channel using pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, *Phys. Lett.* **B737** (2014) 223–243, [arXiv:1406.4456 \[hep-ex\]](#).
- [161] CMS Collaboration, *Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **08** (2014) 173, [arXiv:1405.1994 \[hep-ex\]](#).
- [162] ATLAS Collaboration, *Search for resonances with boson-tagged jets in 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, ATLAS-CONF-2015-073, 2015. <http://cdsweb.cern.ch/record/2114845>.
- [163] ATLAS Collaboration, *Searches for heavy diboson resonances in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, [arXiv:1606.04833 \[hep-ex\]](#).
- [164] ATLAS Collaboration, *Search for resonances with boson-tagged jets in 15.5 fb^{-1} of pp collisions at $\sqrt{s}=13$ TeV collected with the ATLAS detector*, TLAS-CONF-2016-055, 2014. <http://cds.cern.ch/record/2206137>.
- [165] L. G. Almeida et al., *Substructure of high- p_T Jets at the LHC*, *Phys. Rev.* **D 79** (2009) 074017, [arXiv:0807.0234 \[hep-ph\]](#).
- [166] J. D. Bjorken and S. J. Brodsky, *Statistical Model for electron-Positron Annihilation Into Hadrons*, *Phys. Rev.* **D1** (1970) 1416–1420.

-
- [167] ATLAS Collaboration, *Measurement of the cross-section of high transverse momentum vector bosons reconstructed as single jets and studies of jet substructure in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *New J. Phys.* **16** (2014) 113013, [arXiv:1407.0800 \[hep-ex\]](#).
- [168] A. Hook, M. Jankowiak, and J. G. Wacker, *Jet Dipolarity: Top Tagging with Color Flow*, *JHEP* **04** (2012) 007, [arXiv:1102.1012 \[hep-ph\]](#).
- [169] C. Bernaciak, M. S. A. Buschmann, A. Butter, and T. Plehn, *Fox-Wolfram Moments in Higgs Physics*, *Phys. Rev.* **D87** (2013) 073014, [arXiv:1212.4436 \[hep-ph\]](#).
- [170] C. Chen, *New approach to identifying boosted hadronically-decaying particle using jet substructure in its center-of-mass frame*, *Phys. Rev. D* **85** (2012) 034007, [arXiv:1112.2567 \[hep-ph\]](#).
- [171] G. C. Fox and S. Wolfram, *Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes*, *Phys. Rev. Lett.* **41** (1978) 1581.
- [172] S. D. Ellis et al., *Q-jets: A Non-Deterministic Approach to Tree-Based Jet Substructure*, *Phys. Rev. Lett.* **108** (2012) 182003, [arXiv:1201.1914 \[hep-ph\]](#).
- [173] ATLAS Collaboration, *Performance and Validation of Q-jets at the ATLAS Detector in pp Collisions at $\sqrt{s} = 8$ TeV in 2012*, ATLAS-CONF-2013-087, 2013. <http://cdsweb.cern.ch/record/1572981>.
- [174] J. Thaler and L.-T. Wang, *Strategies to Identify Boosted Tops*, *JHEP* **07** (2008) 092, [arXiv:0806.0023 \[hep-ph\]](#).
- [175] ATLAS Collaboration, *Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb^{-1} of proton-proton collision data*, ATLAS-CONF-2014-009, 2014. <http://cdsweb.cern.ch/record/1670012>.
- [176] ATLAS Collaboration, *Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment*, *Eur. Phys. J.* **C76** (2016) no. 1, 6, [arXiv:1507.04548 \[hep-ex\]](#).
- [177] ATLAS Collaboration, *ATLAS Phase-II Upgrade Scoping Document*, Tech. Rep. CERN-LHCC-2015-020. LHCC-G-166, CERN, Geneva, Sep, 2015. <https://cds.cern.ch/record/2055248>.
- [178] ATLAS Collaboration, *Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment*, Tech. Rep. CERN-LHCC-2012-022. LHCC-I-023, CERN, Geneva, Dec, 2012. <https://cds.cern.ch/record/1502664>. Draft version for comments.
-

- [179] ATLAS Collaboration, *A study of Standard Model Higgs boson production in the decay mode $H \rightarrow b\bar{b}$ in association with a W or Z boson for High Luminosity LHC Running*, ATL-PHYS-PUB-2014-011, 2014. <https://cds.cern.ch/record/1740962>.
- [180] ATLAS Collaboration, *Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector*, ATLAS-CONF-2013-079, 2013. <http://cdsweb.cern.ch/record/1563235>.
- [181] C. Englert, M. McCullough, and M. Spannowsky, *Gluon-initiated associated production boosts Higgs physics*, *Phys. Rev. D* **89** (2014) 013013, [arXiv:1310.4828](https://arxiv.org/abs/1310.4828) [[hep-ph](#)].
- [182] L. Altenkamp et al., *Gluon-induced Higgs-strahlung at next-to-leading order QCD*, *JHEP* **1302** (2013) 078, [arXiv:1211.5015](https://arxiv.org/abs/1211.5015) [[hep-ph](#)].
- [183] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, *Comp. Phys. Comm.* **184** (2013) 919, [arXiv:hep-ph/0405247](https://arxiv.org/abs/hep-ph/0405247) [[hep-ph](#)].
- [184] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11*, ATL-PHYS-PUB-2011-009, 2011. <http://cds.cern.ch/record/1363300>.
- [185] C. Quigg, *LHC Physics Potential versus Energy*, [arXiv:0908.3660](https://arxiv.org/abs/0908.3660) [[hep-ph](#)].
- [186] ATLAS Collaboration, *Performance assumptions based on full simulation for an upgraded ATLAS detector at a High-Luminosity LHC*, ATL-PHYS-PUB-2013-009, 2013. <http://cdsweb.cern.ch/record/1604420>.
- [187] ATLAS Collaboration, *Performance assumptions for an upgraded ATLAS detector at a High-Luminosity LHC*, ATL-PHYS-PUB-2013-004, 2013. <http://cdsweb.cern.ch/record/1527529>.
- [188] ATLAS Collaboration, *Performance of the Electron and Photon Trigger in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector at the LHC*, ATLAS-CONF-2011-114, 2011. <http://cdsweb.cern.ch/record/1375551>.
- [189] ATLAS Collaboration, *Preliminary results on the muon reconstruction efficiency, momentum resolution, and momentum scale in ATLAS 2012 pp collision data*, ATLAS-CONF-2013-088, 2013. <http://cdsweb.cern.ch/record/1580207>.
- [190] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, *CERN-2011-002* (2011) , [arXiv:1101.0593](https://arxiv.org/abs/1101.0593).
- [191] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A Program for Higgs boson decays in the standard model and its supersymmetric extension*, *Comput. Phys. Commun.* **108** (1998) 56–74, [arXiv:hep-ph/9704448](https://arxiv.org/abs/hep-ph/9704448) [[hep-ph](#)].

- [192] ATLAS Collaboration, *Jet global sequential corrections with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2015-002, 2015.
<http://cdsweb.cern.ch/record/2001682>.
- [193] ATLAS Collaboration, *Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATLAS-CONF-2016-039, 2016.
<http://cds.cern.ch/record/2206038>.

Acknowledgement

I think writing the acknowledgements might be actually the hardest part of my Ph.D thesis. During the last four years, I have met so many amazing people who walked this long path with me and helped making this chapter of my life a huge success. Unfortunately it's impossible to thank all of you so apologies to whoever I might have forgotten.

First of all I would like to thank my thesis advisor Giuseppe Iacobucci for giving me the opportunity to come back to Geneva for my Ph.D after spending two month already as a “summer” student in the group. Thank you for your support and giving me a lot of freedom in what I was doing but also for continuously being available for discussions.

I am very grateful to the members of my thesis committee, Attilio Andreazza, Tobias Golling and Anna Sfyrla for their willingness to read my thesis (I have tried to keep it short) and their very useful comments that improved the quality of the thesis.

None of this would have been possible without my two great supervisors Reina and Steven. I cannot have wished for better supervisors although I am still shocked that I am older than both of you :) When I came to Geneva, Reina had just finished her Ph.D but she immediately took care of me as if she has never done anything else. I cannot thank Reina enough for introducing me to the world of jets and for her continuous encouragement. Without her, I would not be where I am right now. Steven had the awful task to be my supervisor after Reina left the group but as expected he did an amazing job. I could always bother him with my questions and he knew all the answers (even late at night). I always felt lazy around Steven given that he is one of the hardest working people in ATLAS but at the same time he is so down to earth and took the time to listen to my complaints. Thanks to both of you as well for correcting my thesis!

And then there is my unofficial supervisor Andrea. Thanks to Andrea, I joined the particle physics group in Goettingen in 2010 where she was my Bachelor and Master thesis supervisor. Her way of working and the way she handles everything inspired me from the beginning and thanks to her I cannot imagine a world without particle physics anymore. Even though we are not at the same institute anymore, she is always available on skype to help me, cheers me up during personal downs and celebrates the ups with me. My Ph.D would have been definitely less fun without her! Thanks for correcting my thesis, despite being incredibly busy.

I also would like to thank all the current and past members of the UniGe group for the Friday coffee meetings, for believing in my baking skills, the skiing days and many other things we did together. I should especially point out my wonderful office mates Ece and Francesco. Special thanks to Luis, Szymon and Yann for helping with any computing issues and especially for keeping the cluster running (luckily I only crashed it once). Also many thanks to Catherine for

all the administration work.

There are three people that made my time in Geneva incredibly awesome: Akshay, Gaetano and Javier. Thanks so much for the many dinners together (especially the ones you prepared for me), for bearing my various complicated moods and for the countless nights in different bars (since Gaetano does not appreciate chilling at someones place). Special credits to Akshay for staying awake for so many nights to talk to me until the sunrise. Even though you will probably never admit it but I know you guys enjoyed the nights at Palais Mascotte or other clubs as much as I did :) I also spent amazing holidays with you on Crete, making my birthday very special. You also organised the best thesis gift for me that I could have ever imagined. I will miss you guys a lot when I move to the US, which according to you is my country of birth, but I am sure that we will keep in touch and will share many more amazing memories together.

Thanks so much to the fully hadronic diboson resonance search team, it was a pleasure to work with all of you on this exciting analysis. Even though there was a lot of pressure, there was always a great atmosphere among the analysts and everyone was happy to help the others. I am deeply grateful to Attilio and Enrique who brought me up to speed when I joined the analysis. Even at the very end of my thesis, both of them still answered all of my questions and I always felt that I could ask them whatever I want, no matter how stupid it was without being judged. Special thanks to Valerio for his explanations of the statistical framework.

My Ph.D would have been only half as much fun without the Jet/Etmiss group. I am incredibly happy that I joined the best performance group in ATLAS. I will definitely not forget the fun HCW dinners and how a bunch of physicists sang "hooked on a feeling" in a small karaoke bar in Bratislava. I would like to thank especially Lily for enduring the process of publishing the W -tagging paper with me. It was a pleasure working with you.

I would also like to say a special thank you to Johannes who encouraged me to join the Geneva group and also for the many runs we did together in 2014. Lunch on Mondays would definitely not be the same without Stefan. I was always scared to run into him in R1 after a probably disappointing weekend of my favourite football team. I definitely had a hard time around the Bayern and Dortmund fans, especially Olaf who loves making jokes about the HSV but does not really appreciate jokes about Dortmund :) Thanks as well to Tamara for always smiling and cheering me up immediately. There is one very important person missing: Steffen. He is such an amazing human being that no matter what I am going to write now, it would not give him enough credit. I loved our coffee meetings, the trips we did together, the great conversations Please stay as you are!

I owe a very big thank you to Nedaa, my first office mate in Geneva :) Without you, my French summary would be a mess! Thanks so much for all the amazing dinners you cooked for me, later also expanded to cocktail mixing sessions with Jeff. Fridays will never be the same anymore without our lovely tradition to go to Ole Ole after work. I will truly miss that!

Josefina, I am so incredibly happy that we met during the summer student program in 2011. Even though we are usually separated by more than 10000 km, I know you are always there for me! I will never forget our great holidays in Sardinia. Te quiero mucho!

I am deeply grateful to my friends from Goettingen Filip, Jenni, Melinda, Micha and Sara as well as Stefania and Timo. Despite always making fun of me for working in the morning after a long party night, I could not imagine any better friends. Most of you luckily followed my advise and moved close to an airport making weekend visits very easy. Thanks to you, I am a frequent traveller now :) We shared so many great memories together, especially our festival trips were always a highlight for me!!!

Thanks to Marissa, Karsten, Mandus, Markus and Thies. Almost ten years ago, we left Buchholz and even though we all have barely time to keep in touch, whenever we meet it feels like we just finished school. I love the fact that you are all so interested in my work and you have the great ability to make me laugh, either because you tell me how you envision my Nobel Prize ceremony or when you teach me why $t\bar{t}$ sounds extremely funny.

Last but definitely not least I would like to thank the most important people in my life, my mother and my sister. Thank your for your continuous support and for encouraging me to follow my dreams even though that means that I move further away with every new chapter in my life. I know you always have my back whatever happens. The last words my dad told me were that I should finish my thesis soon. It still took me three years but I finally did it :) This chapter of my life is written now.