Measurement of the Associated Production of a Single Top Quark and a *W* Boson in Single-Lepton Events with the ATLAS Detector

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Sebastian Mergelmeyer
aus
Düsseldorf

Bonn, 21.03.2016

Dieser Forschungsbericht wurde als Dissertation von der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist auf dem Hochschulschriftenserver der ULB Bonn `http://hss.ulb.uni-bonn.de/diss_online` elektronisch publiziert.

1. Gutachter:	Prof. Dr. Ian C. Brock
2. Gutachter:	Prof. Dr. Klaus Desch
Tag der Promotion:	13.05.2016
Erscheinungsjahr:	2016

Contents

CHAPTER 1

Introduction

In order to understand the physics of the fundamental particles and their interactions at subatomic scales, mankind has built gigantic machines that accelerate the tiniest particles to almost the speed of light, and smash them together. One of these machines is the Large Hadron Collider (LHC), where protons are brought to collision at unprecedented energies. These large energies are partially converted into matter, including kinds of matter that are typically not found on earth. The so-created particles are measured, directly or indirectly, by particle detectors such as the ATLAS detector. From the analysis of the particles conclusions can be drawn about the underlying physics processes, especially how often particular processes occur.

Our current knowledge of the physics at subatomic scales is reflected in the Standard Model of particle physics. Given accurate measurements of about twenty fundamental constants, it allows predictions for the rates at which particular processes occur to be made. These predictions are then confronted with the rates that are measured at the LHC and other experimental facilities. Over more than forty years, the Standard Model has withstood all experimental tests nearly unscathed. However, it is known to be incomplete, motivating the search for yet-undiscovered particles – or at least systematic deviations of the measured from the predicted rates, whose patterns could hint at a more complete theory.

One of the particles in the Standard Model is the top quark, which is the heaviest known elementary particle, with a mass close to that of a gold atom. Top quarks are created abundantly at the LHC: in 2012, when protons were brought to collision with a centre-of-mass energy of 8 TeV, more than ten million top quarks were produced. Studies of singly produced top quarks not only open a window to a potential discovery of new physics phenomena but also allow the direct measurement of one of the fundamental constants of the Standard Model, the Cabibbo-Kobayashi-Maskawa matrix element V_{tb}.

The aim of this thesis is to find evidence for the associated production of a single top quark and a W boson (Wt production) in final states with a single lepton, and to measure the Wt production cross section as well as the CKM matrix element V_{tb}. The identification of singly produced top quarks, in particular in combination with a W boson, is particularly challenging because most top quarks are produced in pairs. It is not the first attempt to measure the Wt production in single-lepton final states. Unfortunately, past studies of Wt production in these final states were strongly limited by systematic uncertainties, leading to measurement errors of nearly 100 %. This thesis seeks to analyse the complications of the single-lepton final states, and develop a strategy to overcome them.

The thesis is organised as follows: Chapter 2 gives an introduction of the basic features of the Standard Model of particle physics. The ATLAS detector at the LHC is described in Chapter 3, followed by an explanation of reconstruction of particles in the ATLAS experiment as well as a discussion of its performance in Chapter 4. Chapter 5 is dedicated to the phenomenology and the experimental status of top-quark physics. It also defines the Wt process, and presents a phenomenological analysis of Wt events, focusing strongly on their separation from top-quark pair production. In Chapter 6, the selection of the observed and simulated candidate events is described in detail. Chapter 7 inspects the features of

the selected events that help to further discriminant between Wt and background events. A discriminant based on an artificial neural network is introduced, and then redesigned after an analysis of the complications arising from the use of this technique in a previous measurement of Wt production. The signal cross section is extracted from the observed distribution of the discriminant using a profile-likelihood fit, whose working principle is portrayed in Chapter 8. The chapter also contains a detailed analysis of the nuisance parameters that incorporate systematic uncertainties into the fit. The measured values for the cross section as well as the nuisance parameters are used to update the signal and background estimates. The validity of these updated estimates is demonstrated repeatedly in comparisons with the observed data in Chapters 6 through 8. Finally, the results are compared to theory predictions as well as other measurements of single top-quark production, and interpreted in terms of the CKM matrix element V_{tb}. The thesis concludes with a summary in Chapter 10.

CHAPTER 2

Phenomenological Aspects

The first half of this chapter gives an introduction to the phenomenology of elementary particles and their interactions. It summarises the key aspects of literature such as Refs. [1–6], to the extent needed to describe the physics that is relevant to the topic of this thesis. In Section 2.1, some of the basic concepts of the Standard Model of particle physics are introduced, followed by a more specific discussion of high-energy proton collisions in Section 2.2.

2.1 Standard Model of particle physics

Our current knowledge of the fundamental building blocks of matter and their interactions is expressed in the Standard Model of particle physics. Elementary particles such as electrons are pointlike entities in spacetime, or at least have no known substructure. They are described in the framework of a relativistic quantum field theory as excitations of quantum fields defined on spacetime.

Everyday matter consists of electrons and the constituents of protons and neutrons, namely u-quarks and d-quarks, all of which fall into the class of fermions. An overview over all fermion fields in the Standard Model is shown in Table 2.1, together with some of their intrinsic charges. The particles are arranged in three generations, each containing an up-type quark, a down-type quark, a neutrino, and a charged lepton. The most striking difference between the particles from different generations is their masses.

Forces between the matter particles are mediated by the gauge bosons. The photon mediates electromagnetic (EM) force between particles with a non-zero electric charge, Q, given in units of the elementary charge. The W and Z bosons mediate the weak force, the force responsible for radioactive β-decays. The W bosons couple to weak isospin, T_3, and carry themselves an electric charge ($Q = \pm 1$) and a weak isospin ($T_3 = Q$). The Z bosons couple to a mixture of electric charge and weak isospin. The gluons are the mediator of the strong force, holding the quarks inside the proton together. For every kind of quark shown in Table 2.1, there actually three different quark fields, labelled *red*, *green* and *blue*. Similarly, there are eight different gluons fields, each carrying a different combination of colours. The gluons couple to colour charge, i.e. to quarks as well as other gluons. Quarks and gluons are not observed freely in nature, but are confined into protons, neutrons and other hadrons, whose net colour charge is white. In addition to these charges, the baryon number, B, and the lepton number, L, are conserved in all known interactions.

For every kind of elementary particle, there is a corresponding antiparticle (but no separate "anti-field"). The quantum numbers for the left-handed (right-handed) antiparticles are of equal magnitude as the right-handed (left-handed) particles but a flipped sign (or contrasting colour). In the later parts of this thesis, statements made about particles, e.g. the b-quark, are generally implied to also hold for the antiparticle unless noted otherwise. The gauge bosons are spin-1 particles, whereas the fermions

Quarks			$B = +1/3$, $L = 0$
Up-type quarks	up quark	charm quark	top quark
$Q = +2/3$	u ($\bar{u}$)	c ($\bar{c}$)	t ($\bar{t}$)
$T_3 = +1/2$	(2.3 ± 0.6) MeV	(1.27 ± 0.03) GeV	(173.2 ± 0.9) GeV
Down-type quarks	down quark	strange quark	bottom quark
$Q = -1/3$	d ($\bar{d}$)	s ($\bar{s}$)	b ($\bar{b}$)
$T_3 = -1/2$	(4.8 ± 0.4) MeV	(95 ± 5) MeV	(4.18 ± 0.03) GeV
Leptons			$B = 0$, $L = +1$
Neutrinos	electron neutrino	muon neutrino	tau neutrino
$Q = 0$	ν_e ($\bar{\nu}_e$)	ν_μ ($\bar{\nu}_\mu$)	ν_τ ($\bar{\nu}_\tau$)
$T_3 = +1/2$			
Charged leptons	electron	muon	tauon
$Q = -1$	e^- (e^+)	μ^- (μ^+)	τ^- (τ^+)
$T_3 = -1/2$	(511.0 ± 0.0) keV	(105.7 ± 0.0) MeV	(1.77 ± 0.02) GeV

Table 2.1: Fermions in the Standard Model. The quantum numbers B, L, Q, and T_3 are given for the left-handed fermion fields; for the right-handed fields, T_3 is always zero. The symbols for the antiparticles are shown in parentheses. The masses of the particles [7] are given in units of electronvolts[1]. The neutrino masses are discussed in the text.

are spin-1/2 particles. The fermion fields can be subdivided into left- and right-handed fields. Only the left-handed fermions and right-handed antifermions carry a non-zero weak isospin.

The behaviour of a system of elementary particles is governed by the Lagrange density of the Standard Model, which is a complex scalar function of the fields.[2] The laws of physics are invariant under rotations and translations of spacetime, and so is the Lagrange density, implying the conservation of energy and momentum. The conservation of the charges is the result of an internal symmetry of the Standard Model:

$$\mathrm{SU(3)_C} \times \mathrm{SU(2)_L} \times \mathrm{U(1)_Y}\,, \tag{2.1}$$

The subscript C refers to the colour. The notation means that the density does not change under when the triplet of the fields labelled red, green and blue is rotated (i.e. acted upon by the SU(3) symmetry group) into each other. The subscript L refers to the weak isospin, where only the left-handed up- and down-type fields are rotated into each other, and Y refers to the hypercharge, which is related to the electric charge and the weak isospin:

$$Y = 2\,(Q - T_3)\,. \tag{2.2}$$

The symmetry $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$ is broken at low energies due to the Higgs mechanism, and only the $\mathrm{U(1)_{EM}}$ subgroup is preserved. Due to the symmetry breaking, the bosons mediating the weak force become massive, and the fermions acquire masses though Yukawa interactions with the scalar Higgs field.

[1] Natural units are used throughout this thesis: $\hbar = c = 1$. As a consequence, masses, energies and momenta are expressed in units of electronvolts (eV). A gigaelectronvolt (GeV) corresponds to 1.602×10^{-10} J, or $1/0.938272$ of the mass of the proton.

[2] Its explicit form is given e.g. in Ref. [1].

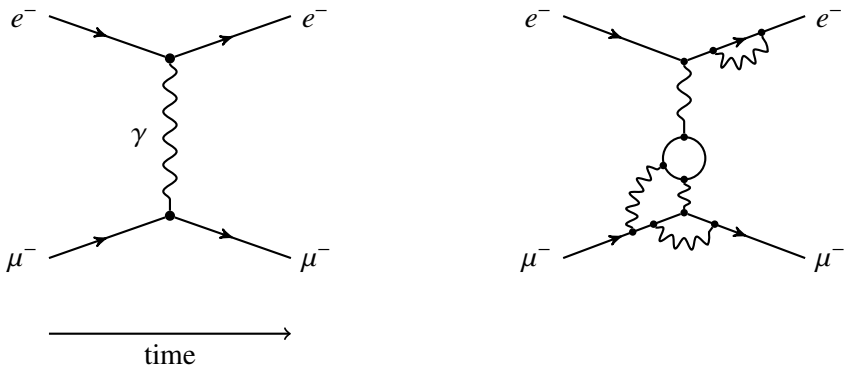

Figure 2.1: The leading-order Feynman diagram for $e^-\mu^- \to e^-\mu^-$ scattering (left), and an example of a higher-order contribution (right).

In the Standard Model, as originally formulated, neutrinos are massless particles. This implies that it does not contain right-handed neutrinos. However, it was established from observation of neutrino oscillations [8–10] that neutrinos must be massive. While it is straightforward to incorporate them into the model, their exact nature is not known. If the mass is generated in the same way as for the charged leptons, the existence of right-handed neutrinos would be implied, but it is also possible that neutrinos are their own antiparticles.

The dynamics of subatomic particles are graphically represented using *Feynman diagrams* [11]. These convey not only an intuitive picture of the subatomic processes but are also powerful tools in the calculation of the dynamics. They are based on the *interaction picture*, which describes the particles as alternating between freely propagating through spacetime and interacting with other particles. For example, Figure 2.1 shows the simplest Feynman diagram that one can draw for the scattering of an electron and a muon. The propagation of a particle is represented by a line, an interaction by a vertex. Every line and every vertex corresponds to a well-defined mathematical expression. Each vertex contributes a factor that is proportional to the coupling strength of the interaction. The particles described by the internal lines do not necessarily propagate forward in time, and may be *off-shell* (or *virtual*), i.e. the magnitude of the four-momentum of the particle is not necessarily equal to its rest mass. The expressions can be multiplied together to obtain the transition amplitude from the initial quantum state on the left side of the diagram to the final quantum state on the right side of the diagram. Technically, a Feynman diagram represents a contribution to the perturbative expansion of the transition amplitude in the coupling strengths.

The diagram shown above is not the only diagram that one can draw for the scattering of an electron and a muon – in fact, there are infinitely many possibilities. Figure 2.1 also shows an example of a more complicated diagram for $e^-\mu^-$ scattering. In practice, one computes Feynman diagrams only up to some order of the perturbative expansion. Assuming that the coupling strength is sufficiently small, this yields a good approximation of the transition amplitude. Section 2.2 elaborates a bit more on this.

The eigenstates of the weak interaction and the eigenstates of free propagation are not necessarily the same. A quark interacting with a W boson does not only convert from an up-type to a down-type quark or vice versa but may also convert between the three generations of quarks. Essentially, a W boson has a different understanding of what the flavour of a b-quark is, and "sees" a superposition of a d'-, a s'-, and b'-quark[3] instead. This is incorporated into the Standard Model as a unitary 3×3 matrix, the *CKM*

[3] The prime is used to denote the eigenstates of the weak interaction.

matrix:

$$\begin{pmatrix} |d'\rangle \\ |s'\rangle \\ |b'\rangle \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \cdot \begin{pmatrix} |d\rangle \\ |s\rangle \\ |b\rangle \end{pmatrix}, \tag{2.3}$$

where:[4]

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97428 & 0.22530 & 0.00347 \\ 0.22520 & 0.97345 & 0.04100 \\ 0.00862 & 0.04030 & 0.999152 \end{pmatrix}. \tag{2.4}$$

The CKM matrix can also be expressed by three angles that describe the rotation of the eigenstates, and a complex phase that is not shown in Equation 2.4. The non-zero off-diagonal elements of the matrix result in a non-zero transition probability for the b-quark into a u- or c-quark.

The Standard Model is an incredibly successful theory. The recently discovered [13, 14] Higgs boson is just one of the particles that had been predicted [15, 16] long before experimental evidence was found. While the description using a truncated perturbative series is not exact, it still leads to extremely accurate results. As an example, the experimentally determined magnetic moment of the electron agrees with the perturbatively calculated value in all eleven significant digits. So far, the Standard Model has been able to explain the observed phenomena at the LHC within the experimental and theoretical uncertainties. However, we know that it cannot be the ultimate "theory of everything":

- Gravity is not described by the Standard Model.[5]
- It is known from cosmological observations that there is another type of matter, *dark matter*, but none of the particles in the Standard Model has the properties of dark matter.
- There is no mechanism that explains why there is so much more matter than antimatter in the observable universe.

In addition, the Standard Model does not offer an explanation for the values of its many parameters. A good example is the vanishing CP-violating phase of the strong interaction: it is highly unlikely that the phase is (extremely close to) zero by pure chance, so it is believed that there is a yet-to-be-discovered mechanism. Another instance of such fine tuning is related to the mass of the Higgs boson. When extending the Standard Model to very high energy scales, the masses of scalar particles receive higher-order corrections of the order of that scale. The mass of the Higgs is however much smaller, and it would take excessive fine tuning of the parameters to achieve such a cancellation – or a more complete theory.

2.2 Predictions for *pp* collisions

The violent scattering of two protons off each other allows a large variety of basic physics processes involving elementary particles to be studied, underpinning our understanding of the Standard Model. The *cross section* measures how frequently a process happens, expressed as an effective area of the colliding particles in units of *barn*, where $1\,\mathrm{b} = 1 \times 10^{-28}\,\mathrm{m}^2$. Cross sections can be calculated as well as measured, connecting theory and experiment.

[4] The 2010 world averages [12] are shown for consistency with what was used for the event generators.

[5] Gravity couples to the energy of the particle, and is always attractive, so unlike electromagnetism, it does not cancel in large bodies. Therefore, it is relevant for our everyday experience of life on earth, but has a completely negligible effect for the interaction between two subatomic particles.

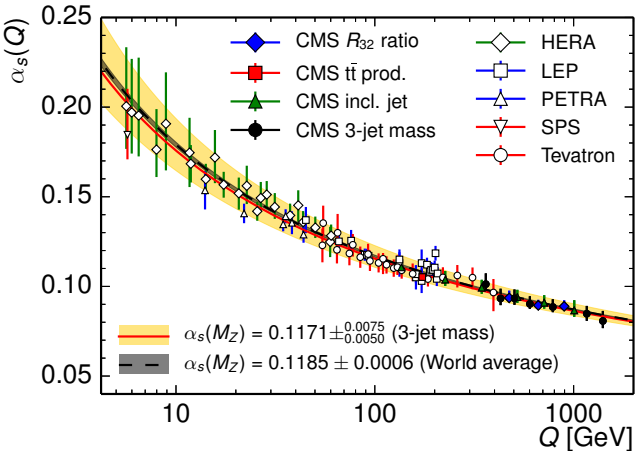

Figure 2.2: Evolution of the strong coupling, α_S, as a function of the energy scale, Q [17]. The curves correspond to the theoretical expectation based on the world average and the measurement of the 3-jet invariant mass by the CMS Collaboration, respectively. They are compared to several measurements of α_S.

When two protons collide, they see each other in a superposition of the proton as a whole and its constituents. The former leads to elastic scattering, whereas the latter leads to inelastic scattering. Although the proton is most simply described as the composition of two u- and one d-quark (valence quarks), one can draw Feynman diagrams where quarks exchange gluons, which can in turn produce more gluons or quark–antiquark pairs, and so on. Hence, a proton can be viewed as a "sea" of quarks and gluons (partons), with the same net flavour as the valence quarks. This sea is responsible for the large mass of the proton, $m_p = 938\,\text{MeV}$, which cannot be explained adding up the masses of the valence quarks (cf. Table 2.1). The relevance of the proton as a whole vs. the valence quarks vs. the sea depends on the energy scale at which the proton is probed. A key point is the *running* of the coupling strength, α_S, of Quantum Chromodynamics (QCD), i.e. the part of the Standard Model that describes the interactions between quarks and gluons:

$$\alpha_S(Q^2) \propto \frac{1}{\log Q^2/\Lambda^2_{\text{QCD}}}\,, \tag{2.5}$$

where Q denotes the energy scale. The coupling strength diverges as the energy scale approaches the confinement scale of the strong interaction, $\Lambda_{\text{QCD}} \approx 200\,\text{MeV}$. As illustrated in Figure 2.2, the running of α_S is experimentally well-established. As a consequence of the divergence, the inner structure of the proton cannot be calculated by means of a perturbative expansion in α_S.

However, if one is interested in physics processes at an energy scale much larger than Λ_{QCD}, the dynamics of the high-energy (hard) process and the dynamics of the sea decouple to a large extent. The sea practically becomes a static pool of partons, and the cross section for pp scattering can be factorised into an effective description of this pool, and the perturbatively calculable hard scattering process:

$$\sigma(pp \to X) = \sum_{ij} \int \mathrm{d}x_1 \mathrm{d}x_2\, \hat{\sigma}_{ij}(x_1, x_2, \alpha_S(\mu_R), \mu_R, \mu_F) f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2)\,. \tag{2.6}$$

The splitting into the two parts is not uniquely defined, but requires the choice of a factorisation scheme and a factorisation scale, μ_F. The former defines e.g. whether the b-quarks are considered to be part of the proton, while the latter determines at which energy scale the structure of the proton is resolved. The choice is in principle arbitrary, but after the truncation of the perturbative expansion, a residual dependence on μ_F appears.

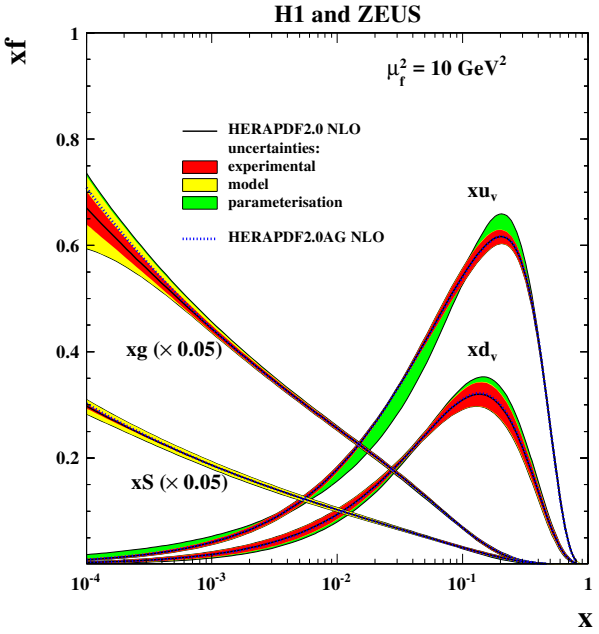

Figure 2.3: The HERAPDF set of proton PDFs, determined from a combination of measurements at HERA [18], shown as a function of the longitudinal momentum transfer at a fixed factorisation scale. The gluon and sea-quark distributions are scaled down by a factor of 20. The uncertainties on the PDFs are shown as well, broken into the experimental, model and parametrisation uncertainties.

Parton density functions

The structure of the proton is described by parton density functions (PDFs), $f_i(x, \mu_F^2)$. At leading order, they can be interpreted as the probability of finding a quark or gluon with the flavour i carrying a fraction x of the momentum of the proton. They are determined from measurements in several experiments, e.g. at the HERA collider, where an $e^{\pm}$ beam was used to probe the electrically charged constituents of the proton [18]. Figure 2.3 shows the PDFs determined from the measurements at HERA at a fixed value of μ_F. At large x, the contributions from the valence quarks dominate, whereas gluons are predominantly found in the low-x regime. The PDFs are universal in the sense that they do not depend on the details of the hard scattering process, and can be extrapolated to different energy scales. Beside the HERA PDFs, there are several other sets of PDFs, which are based on different measurements and different parametrisations of the data.

Hard scattering

The transition amplitude for the hard scattering process can be calculated as a perturbation series, up to some order in α_S.

$$S = c_0 \alpha_S^l(\mu_R) + \sum_{i=1}^{n} c_i(\mu_R, \mu_F) \alpha_S^{l+i}(\mu_R) + O(\alpha_S^{l+n}) \tag{2.7}$$

The c_i are calculated based on the Feynman diagrams for the given initial and final state particles and number of QCD vertices. The number l identifies the lowest order at which a Feynman diagram for the transition can be drawn. For $n = 0$, one speaks of a calculation at *leading order* (LO) in QCD, the corresponding term for $n = 1$ is *next-to-leading order* (NLO). Similar to the factorisation scale, the dependencies of α_S and the c_i on the renormalisation scale, μ_R, cancel each other out when summing up the contributions from all orders. However for a finite n, a residual dependence of calculation on μ_R remains. It is therefore desirable to include higher-order contributions as far as possible (which in practice often means NLO, sometimes NNLO), and have an appropriate choice of μ_R that reflects the

energy scales occurring in the hard scattering process. The systematic uncertainty due to the truncation of the series is commonly estimated by varying μ_R.

Parton shower and hadronisation

Like the valence quarks of the proton, the partons entering and exiting the hard scattering process are accompanied by a cloud of other partons: each of them radiates gluons, typically collinearly or with a small momentum transfer, $p_\perp$. These can in turn radiate even more gluons, split into pairs of a quark and an antiquark, which can again emit gluons, and so on.[6] Eventually, after $O(1 \times 10^{-23}\,\text{s})$, the partons condense into hadrons. The parton showering and hadronisation process are typically described using phenomenological models. In addition, these models also address additional soft parton interactions that happen simultaneously with the hard scattering process as well as the treatment of the proton remnant.

Particle decay

Many of the particles produced in the hard scattering process and during hadronisation are unstable, and decay spontaneously. The probability of the spontaneous decay of a particle into two or more lighter particles can be calculated analogously to the probability of the scattering of two particles. For every possible final state, the *partial (decay) width* Γ_i is defined in analogy to the cross section for two colliding particles. The *branching fraction* of the particle into a specific final state i is given by Γ_i/Γ. The mean lifetime, τ, of the particle is given by the reciprocal of the sum, Γ, of all partial widths.

Kinematics

A few definitions that are commonly used at the LHC and (at least in similar form) other hadron colliders need to be introduced. The coordinate system is defined as follows: the x-axis points towards the centre of the accelerator ring, the y-axis towards the sky, and the z-axis along the beams of the colliding protons. Depending on context, the origin is either in the centre of the detector or at the best estimate of the pp interaction point. The symmetry of the system of the colliding protons makes the use of cylindrical coordinates advantageous. The *transverse momentum*, p_T, the azimuthal angle, ϕ, the rapidity, y, and the invariant mass, M, of a particle with a four-momentum vector of (p_x, p_y, p_z, E) are defined as:

$$p_T = \sqrt{p_x^2 + p_y^2}, \tag{2.8}$$

$$\phi = \begin{cases} \arctan\frac{p_y}{p_x} & : p_x \neq 0 \\ \frac{\pi}{2}\,\text{sign}\,p_y & : p_x = 0, p_y \neq 0 \end{cases}, \tag{2.9}$$

$$y = \operatorname{arctanh}\frac{p_z}{E}, \tag{2.10}$$

$$M = \sqrt{E^2 - p_x^2 - p_y^2 - p_z^2}. \tag{2.11}$$

As the momentum fractions x_1 and x_2 in general not equal, the hard scattering process is not at rest in the laboratory frame but boosted along the z-axis. The polar angle, θ, depends non-trivially on such a boost, which motivates the use of the rapidity. The boost just adds a constant offset to the rapidities, leaving differences of the rapidities of two particles unchanged. For $E \gg M$, the *pseudorapidity*, η, is a

[6] The radiation of a gluon may also be included in the higher-order contributions of the hard scattering process, and *matching* procedures have to be used to avoid double-counting of such contributions with a hard scattering process.

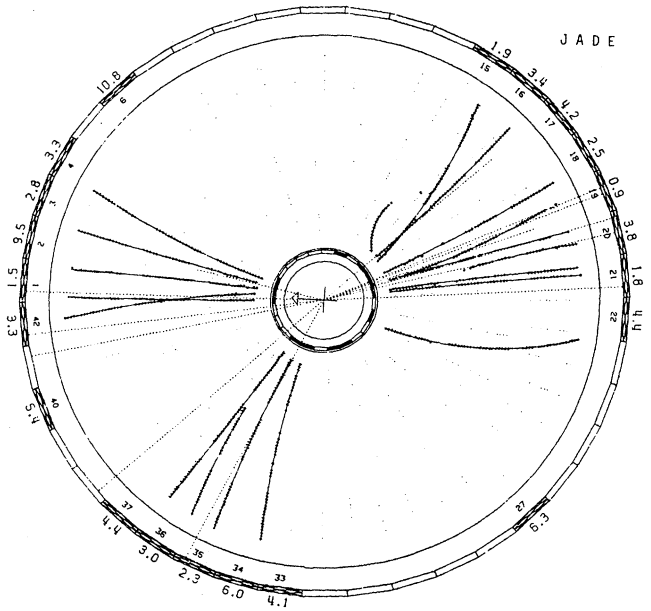

Figure 2.4: Display of a 3-jet ($e^+e^- \to q\bar{q}g$) event recorded by the JADE detector at the e^+e^- collider PETRA, as seen in looking down the beam pipe. [19] The solid (dashed) lines indicate the reconstructed trajectories of charged (neutral) particles.

good approximation of the rapidity:

$$\eta = \operatorname{arctanh} \frac{p_z}{|\vec{p}|} = -\ln \tan \frac{\theta}{2} \,. \tag{2.12}$$

It has the advantage that it can be calculated directly from the polar angle, without the use of the energy. Together with the azimuthal angle, it is used to express the angular separation between two (approximately massless) objects in a way that is invariant under boosts along the z-axis:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \,, \tag{2.13}$$

where $\Delta\phi$ denotes the azimuthal angle between the two objects, and $\Delta\eta$ their difference in pseudorapidity.

Jets

The parton shower typically develops in a small angle around the highly energetic partons exiting the hard interaction, giving rise to the collimated sprays of particles, *jets*, which can even be seen by eye in event displays (Figure 2.4). The jets can be identified using *jet algorithms*, and then used to determine the properties of the original partons. Contemporary jet algorithms fulfil the criterion of *infrared safety*, i.e. their outcome remains unchanged when adding a collinear splitting or a soft emission, ensuring the applicability of fixed-order perturbative calculations to the hard scattering process. An excellent lecture on jets can be found in Ref. [20].

Jet algorithms start off with a list of objects, in particular the particles in the final state. For the purpose of jet finding, particles with a mean lifetime above 3×10^{-9} s are typically considered stable. The generalised k_t *jet algorithm* introduces a distance measures for every pair (i, j) of objects:

$$d_{ij} = \min \left\{ p_{\mathrm{T},i}^{2\rho}, p_{\mathrm{T},i}^{2\rho} \right\} \frac{\Delta R_{ij}^2}{R^2} \,, \tag{2.14}$$

and between every object, i, and the beam axis:

$$d_{i\mathrm{B}} = p_{\mathrm{T},i}^{2\rho}. \tag{2.15}$$

The algorithm identifies the smallest of all d_{ij} and $d_{i\mathrm{B}}$. If it is a d_{ij}, the objects i and j are combined (by adding their four-momentum vectors) into a single object, while if it is a $d_{i\mathrm{B}}$, the object i is identified with a jet, and is removed from the list of objects. The finding proceeds iteratively until no objects remain. The parameter R gives control over the resolution in the η–ϕ plane, while the parameter ρ influences the order in which objects are recombined. For $\rho = 1$, the original k_t algorithm [21, 22] is obtained, which tends to cluster pairs of closeby soft objects first. It is a very attractive choice in principle because it reverses the QCD splitting in the parton shower. For $\rho = -1$, the algorithm is called *anti-k_t algorithm* [23], and prefers to combine a soft object with a nearby hard object before combining soft objects among themselves. This makes the boundary of the jet in the η–ϕ plane more robust to the effects of additional soft interactions that happen simultaneously with the hard scattering process. For this practical reason, the anti-k_t algorithm is typically preferred for physics at the LHC. The resolution parameter R is set to 0.4 for most ATLAS physics analyses of the top quark.[7]

[7] The implementation of the anti-k_t algorithm in FastJet [24] is used. The particular choice of R reflects a trade-off between the sensitivity to the soft interactions and initial-state radiation, and losses due to final-state radiation.

CHAPTER 3

Experimental Setup

The analysis presented in this thesis was performed using proton–proton collisions recorded with the ATLAS detector, which is one of the major experiments at the hadron collider LHC. This chapter portrays the relevant aspects of the design of the LHC and the ATLAS detector.

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) [25] is a circular particle accelerator located underground at the Swiss–French border near the city Geneva. It is one of the facilities of the European Organisation for Nuclear Research (CERN). With its circumference of 26.7 km, the LHC is designed to accelerate two counter-rotating beams of protons[1] to energies of up to 7 TeV per proton.

The proton beams are prepared by extracting protons from hydrogen gas atoms, and then accelerated to higher and higher energies in multiple stages. The entire accelerator complex is schematically depicted in Figure 3.1. The first stage is the linear accelerator (LINAC), from which the protons are successively injected into the PSB (PS Booster), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). The proton beams enter the LHC with an energy of 450 GeV per proton. In the LHC, the beams are kept on a circular path using dipole magnets, which generate magnetic fields up to 8.3 T. The protons are accumulated in bunches of about 10^{11} protons, and accelerated by a system of 2×16 RF cavities, until they reach their nominal energy after about 20 minutes. Quadrupole magnets act as cylindrical lenses that focus the beams on four interaction points. Magnets and cavities are operating in a superconducting state, which requires the LHC to be cooled at a temperature of 1.9 K. The beams are kept circulating for up to 24 hours.

In the year 2012, there were 200 days of operation dedicated to the *pp* experiments [26]. A peak instantaneous luminosity of $7.7 \times 10^{33}\ \mathrm{cm^{-2}\,s^{-1}}$ was achieved, close to its design value of $1 \times 10^{34}\ \mathrm{cm^{-2}\,s^{-1}}$. At that time, the LHC operated with a beam energy of 4 TeV, around 1 380 bunches per beam, and a bunch collision frequency of 20 MHz.

The LHC houses several experiments at the interaction points. ATLAS [28] and CMS [29] are two general-purpose particle detectors with a forward–backward symmetric cylindrical geometry. Both detectors are designed to register particles produced in the interactions over nearly the full solid angle. The names are also used to refer to the experiments in the LHC tunnel as well as the collaborations performing these experiments. While their physics programmes are very similar, they often employ complementary technologies and ideas, which constitutes a valuable cross-check between the experiments. The other experiments investigate more specific aspects of hadron collisions. At LHCb [30], the CP-violation and rare decays of B hadrons are studied, seeking for clues why there is more matter

[1] There is also a physics programme built around collisions with heavy ions (lead nuclei), which shall not be of further concern here.

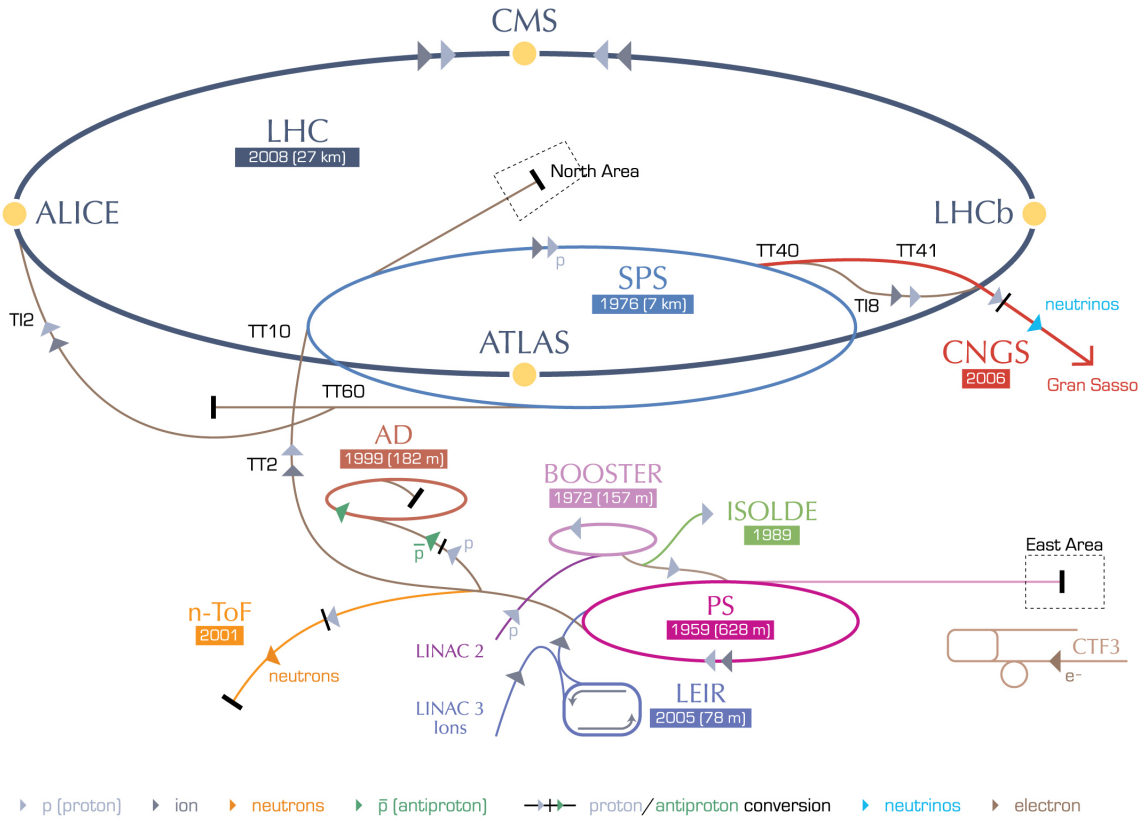

Figure 3.1: Schematic of the CERN accelerator complex [27].

than antimatter in the observable universe as well as hints of physics beyond the Standard Model. ALICE [31] is optimised to study heavy-ion collisions, in particular the formation and properties of the quark–gluon plasma that is believed to have existed shortly after the Big Bang. TOTEM [32] is located near the CMS detector, sharing its interaction point. At TOTEM, the total pp cross section as well as the contributions from elastic scattering and diffractive processes are measured.

3.2 ATLAS Detector

The ATLAS[2] detector is an almost hermetic general-purpose detector operated as part of the ATLAS experiment. It is designed to capture the particles emerging from pp collisions over a solid angle as large as possible. As depicted in Figure 3.2, it is roughly cylindrical, with components placed in different layers either orthogonal to the beam axis (endcap parts), or concentric (barrel parts). Several subdetector groups are arranged around the beam pipe: the *tracking detectors*, which measure charged particles a few centimetres away from the interaction point; the *calorimeter*, which absorb and measure the energy of charged and neutral particles a few metres away from the interaction point; and the *muon spectrometers*, which are tracking detectors in the outermost part of the detector, specialised at registering muons. Each group consists of several subdetectors that are specialised into measuring some properties better than the others. In addition, there are many components to support the operation of the subdetectors, such as magnets, cooling, electronics, cabling, and beam monitors.

[2] The acronym "ATLAS" expands to "A Toroidal LHC Apparatus".

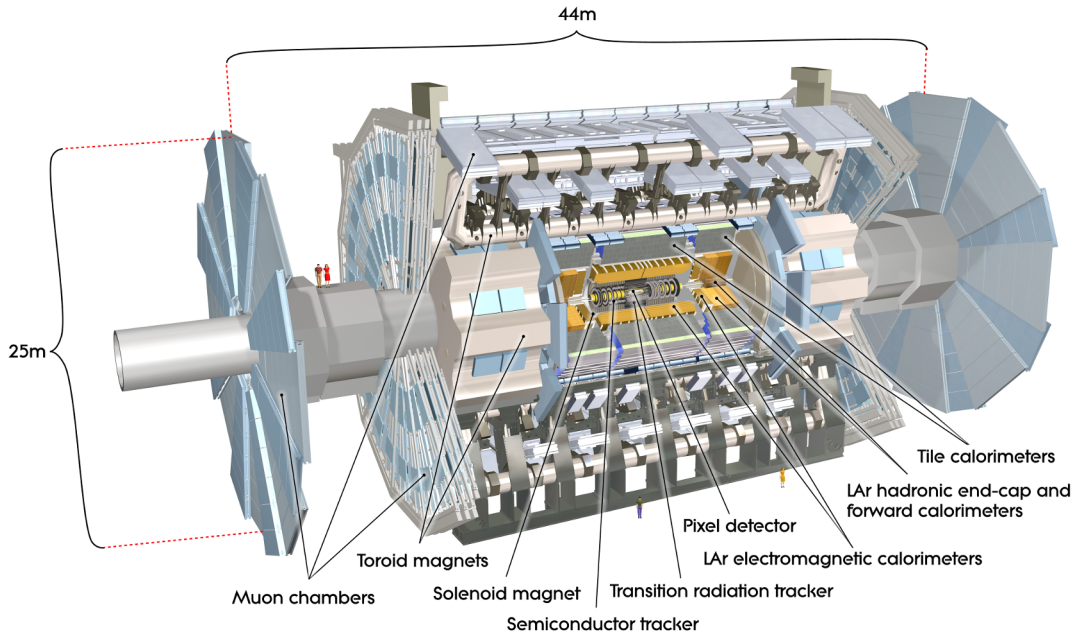

Figure 3.2: Illustration of the ATLAS detector.[33]

In the following, the geometry and workings of the subdetectors are described, with a slight emphasis on the central part ($|\eta| < 2.5$) of the detector, which is more relevant to the analysis later on. Details are taken from Ref. [28]. An overview of how particles are reconstructed from the information provided by the subdetectors, and which resolutions and efficiencies are achieved will be given in Chapter 4.

3.2.1 Inner Detector

The Inner Detector (ID) detects the passage of charged particles, covering a polar angle corresponding to $|\eta| < 2.5$. It is schematically depicted in Figure 3.3. The ID is encompassed by a solenoid magnet, which provides a 2 T axial magnetic field that bends the trajectories of charged particles proportionally to their transverse momentum. As a charged particle moves outwards, its trajectory is measured in multiple space points, allowing the reconstruction of its direction and momentum. Being close to the interaction point, distortions due to interactions of the particles with detector material prior to their detection in the ID are minimised. The ID comprises three subdetectors: the Pixel detector, the Semiconductor Tracker (SCT), and the Transition-Radiation Tracker (TRT).

Two types of sensors are in use. The Pixel and the SCT are silicon-based detectors, in which charged particles repeatedly deposit a small energy. Every energy transfer knocks an electron out of its place in the lattice, generating a measurable electric current. These detectors have a very good time[3] and energy resolution. The TRT is a gaseous ionisation detector. It is filled with gas that is ionised when a charged particle traverses the TRT. The now-free electrons move in an electrostatic field towards an electrode, in particular a wire. The electrons create a cascade of electrons on the way, resulting in noiseless amplification of the signal, which can then be read out at the wire.

The Pixel consists of three barrel layers, and three layers in each of the two endcaps of the detector. The inmost layer in the barrel is just 50.5 mm away from the beam axis. A particle with a pseudorapidity of $|\eta| < 2.5$ crosses at least three layers of the Pixel. The sensors on each layer are segmented in pixels

[3] Such that out-of-time pileup is not an issue.

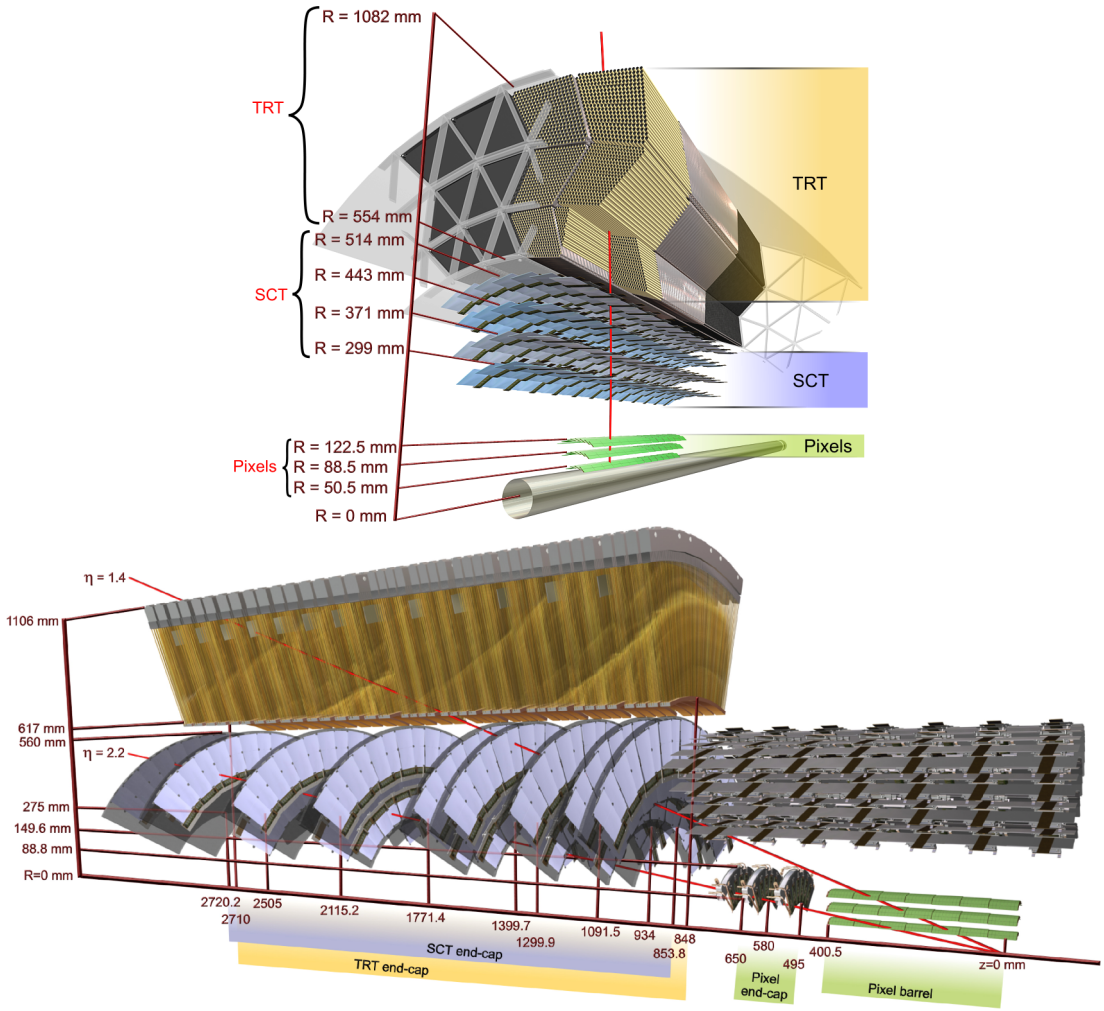

Figure 3.3: Illustrations of the sensors and structural elements focusing on the barrel (top) and endcap (bottom) of the Inner Detector.[33]

of 50 µm × 400 µm, which translates into 80 million readout channels in total – more than any other part of the ATLAS detector. With its fine segmentation and its proximity to the interaction point, it is highly suited to determine the position at which particles are produced, but it needs to be complemented by tracking detectors with that have a longer lever arm for a better measurement of the curvature of the track.

The SCT comprises four layers in the barrel, and nine layers in each endcap. It is located between 299 mm and 560 mm away from the beam axis. Arrays of 120 mm long silicon strips are mounted on both sides of each layer with a pitch of 80 µm. The arrays on each side are tilted by 2.29° with respect to each other. In the endcaps, the strips are arranged radially with the same average pitch as in the barrel. A particle with a pseudorapidity of $|\eta| < 2.5$ crosses at least four layers of the SCT.

The TRT covers the outermost part of the inner detector (563 mm $< \rho <$ 1 066 mm). It is made of straw tubes that are arranged in layers, and filled with a xenon–CO_2 gas mixture. Interleaved between the straw layers are radiators. Electrons and photons emit a characteristic pattern of X-rays (transition radiation) when crossing the boundary between the two materials. The readout electronics operate at two thresholds in order to distinguish the simple passage of a particle from the absorption of the more

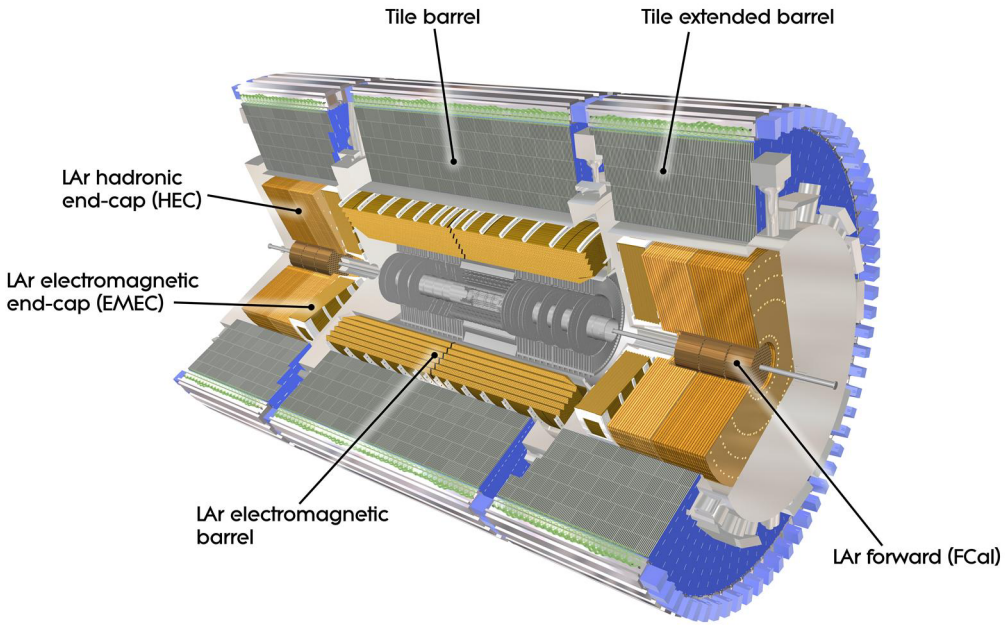

Figure 3.4: Illustration of the ATLAS calorimeter.[33]

energetic transition radiation. In the barrel, the tubes are all parallel to the beam axis, so the longitudinal (z) position of a hit cannot be measured. Likewise, the radial position cannot be determined in the endcaps. The TRT provides about 30 measurements for a charged particle with a pseudorapidity of $|\eta| < 2.1$, with an intrinsic single-hit resolution of about 130 µm.

3.2.2 Calorimeter

Calorimeters are designed to absorb the energy of incident particles as much as possible, and measure the energy deposit. They are crucial for the reconstruction of photons and jets, about 30 % of whose energy is carried by neutral particles. ATLAS employs sampling calorimeters, where passive material is interleaved with active material that is used for the measurement. The calorimeters are subdivided into an electromagnetic calorimeter, targeting the reconstruction and identification of electrons and photons, and a hadron calorimeter, both depicted together in Figure 3.4. Particles that interact primarily via the electromagnetic force start to develop showers in the EM calorimeter, whose energies can be measured very precisely. Hadrons on the other hand deposit the bulk of their energy in the hadron calorimeter. Various kinds of secondary particles are produced, in proportions that are affected by large event-by-event fluctuations. As the calorimeter is *non-compensating*, i.e. its response depends on the kind of particle, the fluctuations translate into a less accurate measurement of the energies of hadrons.

Electromagnetic calorimeter. The EM calorimeter consists of a barrel ($|\eta| < 1.475$) and two coaxial endcap ($1.375 < |\eta| < 3.2$) parts. It uses liquid argon (LAr), which is ionised by the secondary particles, as active material, and lead as passive material. The calorimeter comprises three layer, and is complemented by a presampler that measures the energy loss in front of the calorimeter in the pseudorapidity range $|\eta| < 1.8$. The inner layer, together with the dead material and the presampler in front

of the calorimeter, covers 6 radiation lengths. The middle layer forms the bulk of the EM calorimeter, collecting about 80 % of the energy of an incident electromagnetically interacting particle on average. The outermost layer collects the tail of the shower and does not extend into the forward region, covering only the central pseudorapidity range, $|\eta| < 2.5$. In total 22 to 45 radiation lengths are covered, depending on the polar angle. Every layer is structured as a grid of cells. In the barrel, the inner layer has a granularity of about 0.003×0.1 in the η–ϕ plane, and the middle layer approximately 0.025×0.025. The granularity allows a precise measurement of the shower shapes of the electrons and photons, and helps in the matching of energy deposits to the tracks found in the ID.

Hadron calorimeters. The hadron calorimeter is subdivided into three parts: the Tile calorimeter, the LAr hadronic endcap, and the LAr forward calorimeter. Together, their depth is about ten interaction lengths.

The Tile calorimeter spans the pseudorapidity range $|\eta| < 1.7$, and constitutes the bulk of the calorimeter. It uses steel as absorber and scintillating tiles as active material. It has a granularity of 0.1×0.1 in the η–ϕ plane. It is further subdivided into the barrel ($|\eta| < 1.0$) and two extended barrels ($0.8 < |\eta| < 1.7$).

The LAr hadronic endcap spans $1.5 < |\eta| < 3.2$. It consists of four layers, using copper as absorber. For $\eta < 2.5$, it has the same granularity as the Tile calorimeter. The LAr forward calorimeter spans $3.1 < |\eta| < 4.9$. Although considered part of the hadron calorimeter, its first layer is actually more geared towards EM showers, and uses copper as absorber. The other two layers use tungsten.

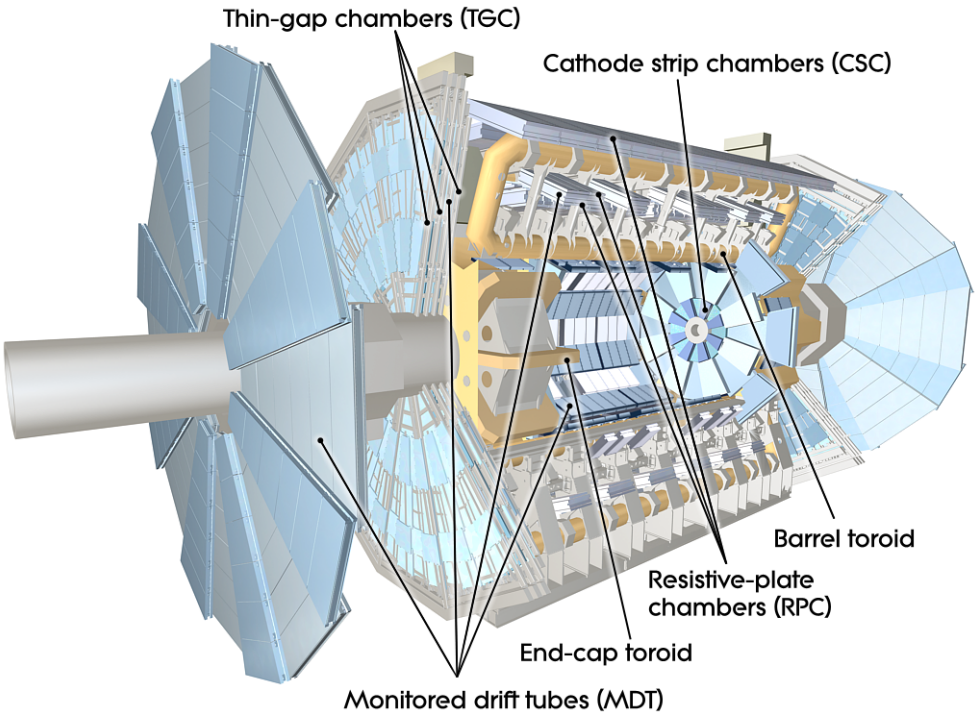

Figure 3.5: Computer-generated view of the Muon Spectrometer its supporting magnet system.[33]

3.2.3 Muon Spectrometer

The Muon Spectrometer in the outermost radial part of the ATLAS detector is a tracking detector dedicated to the reconstruction and identification of muons, which are the only known kind of particle passing the calorimeter nearly unimpeded but still interacting in the muon spectrometer. As illustrated in Figure 3.5, the spectrometer comprises Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs) for the momentum measurement, as well as Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs), which are mainly used for triggering. The momentum measurements are enabled by a toroidal magnet system that generates a non-uniform magnetic field with a bending power of 1 T m to 7.5 T m. The system consists of a barrel ($|\eta| < 1.4$) and an endcap ($1.6 < |\eta| < 2.7$) part. The design of the magnetic field results in a nearly constant momentum resolution as a function of the pseudorapidity.

The MDTs are gaseous detectors using an argon–CO_2 gas mixture. They are arranged in three coaxial barrel layers located 5 to 10 meters away from beam axis, and four endcap disks located at $|z| = 7.4\,\text{m} \ldots 21.5\,\text{m}$ on each side of the detector. A muon passing the MDTs is measured in about 20 tubes. The achieved single-hit resolution depends on the drift radius in the tubes, and is 80 µm on average. As for the TRT, no measurement can be made along the wire, so the measurement is only performed in the z (barrel) or ρ (endcaps) direction.

The CSCs replace the first endcap disk of the MDTs for $\eta > 2.0$. They are multiwire proportional chambers, and can withstand the high rates present in this detector region. Segmented cathode strips are mounted on two sides of the chamber. The CSCs provide typically 3 to 4 measurements of a passing muon, with a single-hit resolution in ρ that is comparable with the MDTs. Since the strips on each side are tilted by 90°, the ϕ component can be measured as well, albeit with a lower single-hit resolution of about 10 mm.

The RPCs cover the outer barrel part of the muon system ($|\eta| < 1.05$). They are filled with gas between two parallel plates. They are designed to quickly signal the detection of a muon to the trigger system, and provide basic position information, which is determined to about 10 mm in both directions. This role is taken over by multiwire proportional chambers, the TGCs, in the endcaps ($1.05 < |\eta| < 2.7$). The TGCs also complement the MDTs in the reconstruction, providing a measurement of the ϕ component. Owing to their small sizes, the RPCs and TGCs can inform the trigger system about the detection of a muon within 15 ns to 25 ns.

3.3 Trigger and data acquisition

In each of the 20 million bunch crossings per second, a collision event of interest can happen. However, it is neither sensible nor possible to record (or even read out) all detector signals, which amount to 1 MB to 2 MB of data, at such a high rate. Hence, a trigger system is employed to quickly identify events that are of potential interest for physics analysis or performance measurements.

The ATLAS trigger system has three consecutive levels. At every level a large fraction of events is discarded, which allows more time to be spent on a potentially interesting event on the successive levels. The level-1 trigger (L1) is integrated into the hardware of the ATLAS detector. It relies on the trigger chambers of the muons, as well as low-granularity information from the calorimeter for the identification of jet, electron, and photon candidates, and reaches a decision after about 2.5 µs. Interesting events are passed to the software-based higher-level triggers at a rate of up to 70 kHz. The level-2 trigger (L2) refines the decision, using detailed information in the relevant regions of the detector, as indicated by hints provided by the L1 trigger. Unlike the L1 trigger, it also has access to the data from the Inner Detector. After about 10 ms, the events are passed to the Event Filter (EF) at a rate of up to 6.5 kHz. The EF fetches all information about the event, and reconstructs it with the same computer programmes

that are also used to reconstruct the event for physics analyses later on. The decision is made within a few seconds, reducing the rate to 600 Hz.[4]

Events that pass any electron or photon trigger are saved in the `Egamma` data stream. Likewise the `Muons` data stream is made of events that pass any muon trigger. During the data taking in 2012, about 100 events per second were written to each of the two streams, with a very small overlap (1 %) between the streams.

3.4 Determination of pileup and luminosity

Due to the high instantaneous luminosity, a single bunch crossing typically results in multiple pp collisions. As a consequence, contributions from additional pp collisions, *(in-time) pileup events*, are typically superimposed on the hard pp collision of interest. In addition, the integration times in the calorimeter are $O(100\,\text{ns})$, so the response of the calorimeter is susceptible to *out-of-time pileup* caused by pp interactions in adjacent bunch crossings.

Given that the bunch spacing is constant, the average number of pp interactions per bunch crossing, $\langle\mu\rangle$, can be used as a measure of both in-time and out-of-time pileup. It can be calculated from the luminosity per bunch crossing and the total inelastic pp cross section:

$$\langle\mu\rangle = \sigma_{\text{inel}}(pp \to X) \times \frac{\text{instantaneous luminosity}}{\text{\# number of bunch crossings per second}} \,. \tag{3.1}$$

The in-time component of the pileup can be estimated from the number of primary vertices[5] in an event. The two numbers are used to correct the reconstructed event for pileup. The total inelastic cross section is assumed to be 73 mb, as calculated with the Pythia event generator. This is consistent with the measurement from TOTEM [35]: $(74.7 \pm 1.7)\,\text{mb}$. The consistency was also confirmed for a centre-of-mass energy of 7 TeV by ATLAS [36] and TOTEM [37].

The remainder of this section summarises the key aspects the luminosity measurement [38]. Two specialised detectors, the Beam Conditions Monitor (BCM) and the LUCID detector, measure the visible inelastic pp collision rate at the granularity of bunch crossings. The rate is proportional to the luminosity up to an almost constant factor, which is determined using an absolute measurement from time to time.

The BCM comprises two stations located 1.8 m away from the interaction point, one station for each side of the detector. In each station, four small sensors made from $1\,\text{cm}^2$ of synthetic diamond are mounted at an angle of 45° with respect to the beam axis, capturing activity at $\eta = \pm 4.2$. LUCID is a segmented Cherenkov detector that also comprises two stations that register activity in the range $5.6 < |\eta| < 6.0$, about 17 m away from the interaction point. It is a gaseous detector consisting of 16 aluminium tubes that are wrapped around the beam pipe. Photomultipliers detect the Cherenkov light emitted by charged particles that pass through the tubes.

The absolute instantaneous luminosity is determined using the relation:

$$\text{instantaneous luminosity} = \frac{\text{\# bunch crossings per time} \times \text{\# protons per bunch}}{2\pi\,\sigma_x\sigma_y} \,, \tag{3.2}$$

where σ_x and σ_y characterise the horizontal and vertical widths of the *luminous region* at the interaction

[4] The rates quoted here refer to the end of the 2012 data-taking [34], and depend roughly linearly on the instantaneous luminosity. The rate after the EF averaged over the full data-taking period was about 400 Hz.

[5] With the help of the tracking system, space points are reconstructed where pp collisions have presumably happened. These are called *primary vertices* (see also Section 4.1).

point, where the two beams intersect. The widths were measured using beam separation scans (*van-der-Meer scans*) [39, 40] in several dedicated runs during the data-taking periods. The uncertainty on the integrated luminosity is 2.8 % [38], affecting the overall normalisation of all physics processes estimated from the MC simulation (see Section 6.2).

3.5 Detector simulation

With an ideal measurement apparatus, the predicted (see Section 2.2) and the observed rates of the physics processes could be directly compared. In practice, detector effects such as resolution and efficiency need to be taken into account. This is achieved by a simulation of the ATLAS detector that runs on top of the Monte Carlo (MC) simulation of the physics processes. Afterwards, experimental data and theoretical predictions can be compared at the level of visible rates. Conversely, one can use the simulation to derive a correction that is applied to the observed rate, and then confront theory and experiment at the level of the cross section.

The interactions of the particles with the detector material are simulated using the GEANT4 toolkit [41] with a detailed description of the ATLAS detector. The effects of in-time and out-of-time pileup are simulated with PYTHIA, and mixed in, together with samples of noise from electronics and cavern backgrounds. The response of the individual detector components to these hits is simulated, including the digital output of each component. Variations of the run conditions over time, e.g. faulty modules and changes in the amount of pileup, are taken into account.[6] The output of the simulation closely resembles the data recorded with the ATLAS detector. It is accompanied by a record that contains information about the particles in the hard scattering process, the parton shower and the decay products, information about hits in the detector, and so on.

The precise simulation of all interactions takes $\mathcal{O}(10 \times 10^3\,\mathrm{s})$ to compute on commodity hardware, making the available computing power a limiting factor for the simulation. Several approaches have been developed to simplify and speed up the simulation by replacing some details of the simulation with effective parametrisations, in particular for the otherwise time-consuming modelling of the calorimeter responses. One of these approaches, referred to as ATLFAST2 [42], is relevant in the context of this thesis, and is about 20 times faster than the full simulation with GEANT. The fast simulation provides a less accurate description of jet substructures and electron isolation, which is accounted for by corrections that are applied in the reconstruction procedures (see next chapter).

[6] Note that the simulation must often be run before the data-taking finishes, and thus relies on projections of the LHC and detector conditions into the future. This is addressed later on by a reweighting that matches the conditions in the simulation to the data (see Section 6.2). The same applies to minor corrections of the detector modelling.

CHAPTER 4

Reconstruction of Physics Objects

Particles produced in the *pp* collision are observed experimentally by means of the traces they leave when they interact with the detector. Some of them are directly detectable when they traverse the detector, while others, including the top quark, decay before they can reach any subdetectors, so they are seen only through their decay products. Yet others, to our current knowledge only neutrinos, are stable but do not undergo any significant interaction with the detector. This chapter introduces the methods used to identify and measure the types, the trajectories and other properties of particles that reach the detector, to the extent relevant for the present analysis. At the beginning of the chapter, general concerns of the particle reconstruction are discussed, followed by the reconstruction of tracks (Section 4.1). Since each reconstruction algorithm is specialised in the detection of different kinds of particles, the rest of the chapter is subdivided accordingly into sections for jets and photons (Section 4.2), charged leptons (Section 4.3), and neutrinos (Section 4.4).

The reconstruction and calibration of physics objects fall under the responsibilities of the ATLAS Combined Performance working groups, who provide recipes and software tools for the use of the reconstructed objects in all physics analyses. A comprehensive overview is given in Ref. [43], upon which the ATLAS Top Physics working group has built additional recipes for analyses of top-quark physics [44].

Detector geometry and conditions. A realistic description of the geometry of the entire detector and the conditions of the sub-detectors is required not only for the detector simulation but also for the optimal performance of the reconstruction. The exact details vary depending on the reconstruction algorithm. As a general example, faulty sensors are masked, and the algorithms then rely more strongly on the information from the neighbouring sensors, for example by interpolating between them. The positions of all sensors should be known significantly better than the intrinsic single-hit resolution. Since this precision cannot always be achieved by the precise assembly of the components, an *alignment* procedure was performed in order to measure the positions of the sensors [45]. In addition, systematic deformations were corrected in order to avoid biases in the track reconstruction [46]. The description of the detector also includes dead material and the magnetic field, as particles traversing them can undergo significant scattering processes.

Momentum calibration. As with most measurement devices, the measured momenta of the particles must be calibrated. At the lowest level, the digitised responses of the subdetectors, such as voltages and time differences, are converted into measurements of the energy losses or the positions of the incident particles. A concrete example is the translation of the drift time in the TRT into a drift radius [48]. A calibration of these measurements is derived, usually from the MC simulation, where the true momentum of the object is known. Afterwards, one or more well-understood physics processes are used to validate the calibration using the observed data, as for example the resonances shown in Figure 4.1.

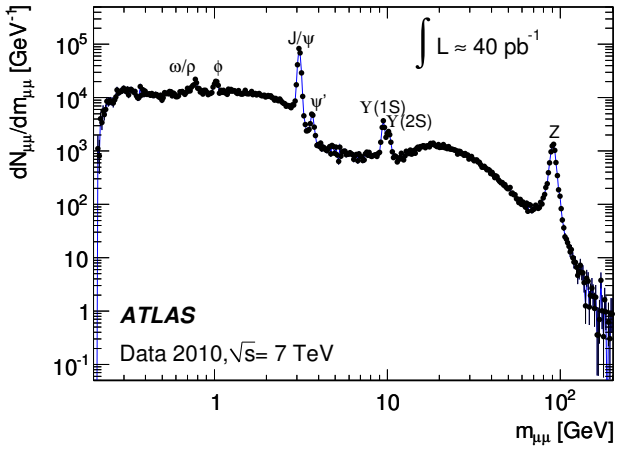

Figure 4.1: Invariant mass spectrum of reconstructed muon pairs.[47] The J/ψ, Υ and Z resonances are clearly visible as peaks which serve as "standard candles" against which the muon reconstruction is calibrated.

There can be small residual differences between data and the simulation due to small imperfections, for which additional corrections and systematic uncertainties can be derived from observed data.

Efficiencies. Each reconstruction procedure is designed to identify objects of a certain kind with high efficiency and purity. The *efficiency* is the probability that an object, say an electron, is actually reconstructed and identified as such. The *purity* is the fraction of the reconstructed electron candidates that truly are electrons rather than other kinds of objects or noise. Closely related to this is the *rejection*, defined as the inverse efficiency for a non-electron to be misidentified as an electron. The more distinguishable the signature of objects of a given type is, the higher the efficiencies and purities are that can be achieved. Usually a trade-off between efficiency and purity has to be made, which often depends on the specific needs of the analysis. As with the momentum calibration, the efficiencies can be slightly different in the observed data and the MC simulation, and if so, the simulation is reweighted. Here one faces the problem how to obtain a sufficiently pure and unbiased sample of the objects whose the efficiency can be measured in data. A common technique to solve this is the so-called *tag-and-probe* method. It works by selecting events with a very clean topology and a pair of objects in the final state, such as $Z \to \ell^+\ell^-$ events. A very pure sample of these pairs can be created using cuts on the invariant mass of the pair. If one of the two objects passes the usual strict set of identification criteria, $A \cap B$, it is used as *tag*. The other object is only required to meet some very basic criteria, A, and is then used to *probe* the efficiency of the identification cuts, B.

4.1 Tracks and Vertices

Charged particles produce a series of hits when traversing the tracking detectors (see Section 3.2.1). Although these tracks are not used directly in this analysis, they are an indispensable input for the reconstruction algorithms outlined in the rest of this chapter as well as the suppression of backgrounds.

The track reconstruction algorithm starts from seeds made of three space points in the silicon detectors. This is sufficient to define a corridor in which further hits are gathered using a *Kalman filter* [49]. Whenever a hit is added to the track, the parameters of the reconstructed trajectory of the particle are updated using a fit. Hits that would worsen the quality of fit are removed again. Ambiguities in the association of hits to tracks are resolved by preferring long tracks with more precise hits and fewer gaps. Afterwards, a similar procedure is performed starting from the TRT and adding hits inwards,

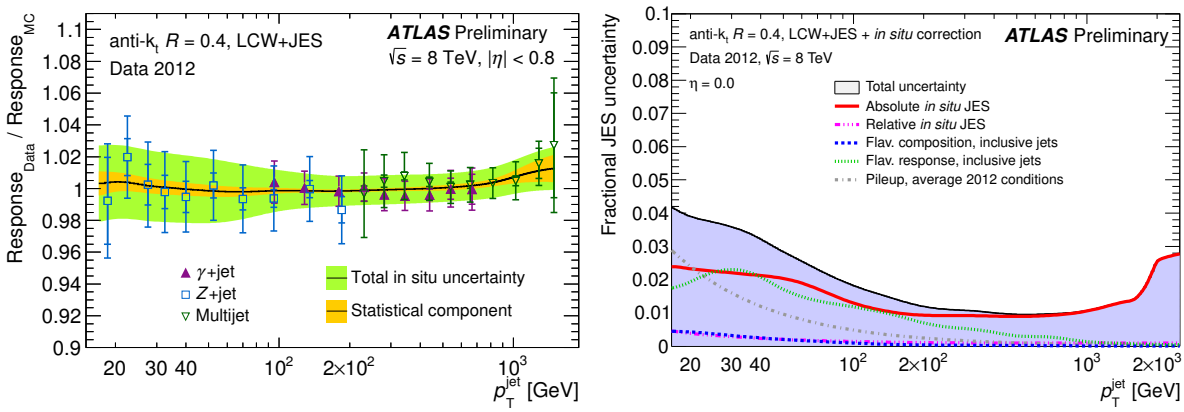

Figure 4.2: Ratio of the jet energy scale in the observed data before the in-situ correction divided by that in the MC simulation (left), and breakdown of the uncertainties on the jet energy scale (right), both as function of p_T for jets in the central part of the detector. The figures originally appeared in [54], and have been updated by the ATLAS Collaboration during the 2012 data taking.

which helps to reconstruct tracks that originate from interactions of a primary particle with the detector. Details of the algorithm can be found in Ref. [50, 51]. The extension of the tracks to the Muon Spectrometer is discussed in the context of the muon reconstruction (Section 4.3.2).

Tracks are reconstructed down to transverse momenta of $p_T = 0.4\,\text{GeV}$ and an efficiency of 70 % to 85 % depending on the momentum and the polar angle of the track [52]. The relative measurement error on p_T increases with p_T because it is measured using the inverse of the curvature of the track. The overall performance of this procedure is limited by the accuracy of the single-hit measurements, scattering processes, noise, and the alignment.

Closely related to tracking is the reconstruction of *vertices*, points from which a set of tracks originate. After using a pattern recognition technique to gather the tracks for a vertex candidate, a fit of the vertex in three-dimensional space is performed [53]. The vertices where a pp interaction has presumably happened are called *primary vertices*. An event can have more than one such vertex due to pileup interactions. The primary vertex of the hard interaction will be referred to as the *hard collision primary vertex*. It is distinguished from the pileup vertices by a much larger quadratic sum of its associated track transverse momenta. *Secondary vertices*, where a particle has presumably decayed into two or more charged particles, are reconstructed as well. This includes the decays of unstable particles as well as photon conversions ($\gamma \to e^+e^-$) and hadronic interactions in the detector material.

As the trajectory of a charged particle is bent in the magnetic field, it can be represented locally as a helix[1]. In order to quantify the parameters of the helix, a reference point needs to be chosen. An advantageous choice of reference point is the reconstructed origin of the particle, i.e. its production vertex. Defining the perigee point of the track as the point of closest approach of the reconstructed track to the reference point in the transverse plane, the helix is parametrised by five quantities: d_0 and z_0 are the radial and longitudinal positions of the perigee point with respect to the reference point, ϕ_0 is the azimuthal direction at the perigee point, θ is the polar angle, and q/p the electric charge over the momentum. As all Standard Model particles that are visible in the tracking detectors have a charge of $q = \pm 1$, the momentum can be inferred directly from the curvature of the helix. If the track has

[1] The helix approximation holds very well inside the luminous region, but not for the full trajectory, for various reasons: the Inner Detector is longer than the solenoid, the toroids do not produce a homogeneous magnetic field, and the particle undergoes scattering processes on the way.

actually been produced at the chosen reference point, and the finite resolution of the track resolution is ignored, then d_0 and z_0 essentially vanish. Tracks that do not actually stem from the reference point can be distinguished based on their larger values of $|d_0|$ and $|z_0|$.

4.2 Jets

Quarks and gluons produced in the hard collision cannot be observed directly, but are accessible via an appropriate jet definition as portrayed in Section 2.2. These jets are bunches of particles, predominantly charged pions (amounting to 40 % of the energy of the jet), photons from the decays of neutral pions (25 %), and kaons[2] (20 %) that deposit large amounts of energy in the calorimeters. High-energetic photons are not reconstructed explicitly in this analysis, because the production of such photons constitutes a negligible source of background.

Before running the jet algorithm, it is advantageous to combine the energy deposits into clusters, each of which corresponds approximately to a stable particle. So-called *topological clusters* are formed, taking cells with a signal of at least 4 times the noise level[3] as starting points [55]. The cluster is grown by repeatedly adding the direct neighbours of all cells with a signal-to-noise ratio above 2. Afterwards, it is split into two or more clusters if multiple energy maxima are found inside the cluster. A fraction of the energy of the incident particle may escape detection due to dead material in front of the calorimeter and excitations of nuclei in hadronic showers. An MC-based calibration is used to correct the measured energy of the cluster for such energy losses. The correction distinguishes electromagnetic and hadronic showers based on the shape and the energy density of the cluster, partially addressing the non-compensating nature of the calorimeter [56].

The jets are obtained by running the jet algorithm with the topological clusters as input. Afterwards, a sequence of corrections is applied to the jets [54, 57]:

- The direction of the jet is computed from the position of its energy deposits in the detector, under the assumption that the jet originates from the hard collision primary vertex.
- The expected contribution from pileup is subtracted.
- The energy of the jet is calibrated, compensating for energy losses that are not recovered by the calibration of the clusters (e.g. due to the noise threshold, or bending in the magnetic field). The calibration corrects the energy of the jet to the energy of its true counterpart in the MC simulation. The true jets are built from stable particles, excluding neutrinos and muons.
- The energy measurement is refined based on information such as shower shapes, tracks, and activity in the muon chambers [58]. These corrections do not change the overall calibration, but help to improve the resolution.
- The residual *in-situ* correction is applied to each jet in observed data. It was derived from multiple measurements. The energy scales of jets within and outside the central pseudorapidity range $|\eta| < 0.8$ were calibrated against each other, exploiting the momentum balance of the two jets in dijet events. Using samples of Z + jet and γ + jet events, the jets were then calibrated against electrons and photons, taking advantage of their precisely measured energies. Finally the energy scale was extrapolated into the TeV regime by balancing the momentum of high-p_{T} jets against multiple jets with smaller momenta.

[2] The short-lived neutral kaon, K^0_{S}, however decays into pions before it can reach the calorimeter.
[3] The noise level is known from events captured at random bunch crossings, i.e. without requiring a pp interaction.

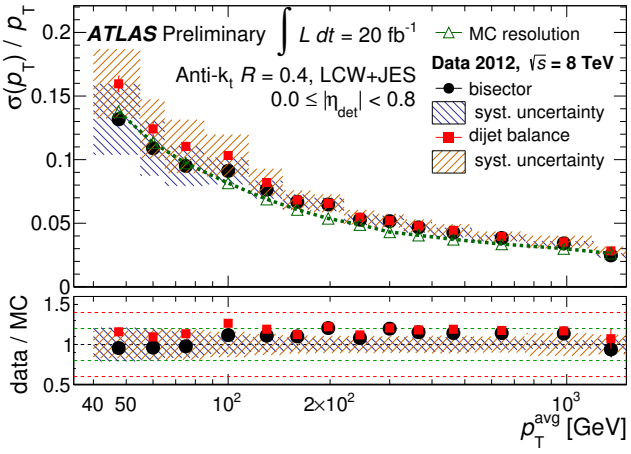

Figure 4.3: Relative resolution of the reconstructed transverse momentum of the jet, measured in dijet events with two different methods, and compared to the MC simulation.[59]

Jets with a calibrated transverse momentum below 20 GeV are discarded. Figure 4.2 (left) shows the performance of the reconstruction for jets in the central part of the detector. The jet energy scale in the observed data is modelled well by the MC simulation. The total uncertainty for central jets with a transverse momentum of 40 GeV is about 3 %, and decreases to about 1.5 % for 100 GeV, while increasing to 4 % to 7 % beyond $|\eta| > 2.5$. Despite this remarkable precision given the difficulties of the energy measurement, it is one of the dominant systematic uncertainties in top-quark measurements.

Figure 4.3 shows the relative resolution of the jet energy measurement. It is about 15 % for jets with a transverse momentum up to 30 GeV. The energy resolution in the observed data is known from dijet events with a precision of about 20 %. The MC simulation reflects the energy resolution within this uncertainty.

The jet reconstruction is fully efficient over the phase space considered in this analysis. There is however a small uncertainty of 0.23 % on the efficiency of the reconstruction for jets with a transverse momentum below 30 GeV.

Cleaning

Quarks and gluons are not the only particles that manifest as jets. Electrons deposit an even larger fraction of their energy visibly in the calorimeter, hence electron candidates are also reconstructed as jets. This is addressed by removing the jet that is closest to an identified electron within a distance of 0.2 in the η–ϕ plane[4]. If an electron is not identified as such, it counts as a jet, whose energy tends to be overestimated due to the hadronic corrections.

Noise, cosmic rays, and beam-induced backgrounds[5] can give rise to jets as well. These so-called *bad jets* are identified based on their energy distributions over the different calorimeter layers, pulse shapes, tracks associated with the jet, and timing information. Events containing a bad jet are rejected [61]. Moreover, there was a hot spot in the second layer of the Tile Calorimeter around $(\eta, \phi) \approx (-0.15, 2.7)$ in the first week of May 2012, which was not masked in the reconstruction. Therefore, jets with a large fractional energy deposit in the region of the hot spot also count as bad jets.

[4] There are ongoing efforts in the ATLAS collaboration to address this overlap more elegantly and efficiently using so-called *particle flow* algorithms. The energy deposits of the electron would then be identified at the level of calorimeter clusters, and no longer be available as inputs for the jet finding. Other experiments (ALEPH, CMS, ZEUS) successfully used this strategy before.

[5] This includes collisions of the proton beam with gas atoms in the beam pipe as well as the beam halo, which consists of muons and pions accompanying the beam [60].

Finally, a significant amount of jets with low transverse momentum originates from pileup interactions. A quantity called *jet vertex fraction* [62] (JVF) is used to suppress such jets, and is defined as:

$$\frac{\sum_k^{J\cap K_1} p_{\mathrm{T}}(\text{track } k)}{\sum_k^{J\cap K} p_{\mathrm{T}}(\text{track } k)} \,. \tag{4.1}$$

The set J includes all tracks inside the jet cone, and K contains all tracks associated with any primary vertex, whereas K_1 is restricted to the tracks associated with the hard collision primary vertex. This quantity is only used for jets inside a pseudorapidity range of $|\eta| < 2.4$ as coverage by the Inner Detector is required, and is undefined if there are no tracks inside the jet cone. A jet with a transverse momentum below 50 GeV is identified as pileup jet if its JVF is defined and below 0.5. As jets with $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ are very rarely the result of pileup interactions, they are never classified as pileup jets regardless of their JVF. The JVF is assumed to have an uncertainty of ±0.03 near the cut value of 0.5.

4.2.1 Identification of *b*-quark jets

The presence of a *b*-quark jet from a top-quark decay distinguishes top-quark events from most other background events. Algorithms that identify such *b*-quark jets, so-called *b-taggers*, are therefore an essential device for most top-quark analyses, including the present one. The ATLAS Collaboration developed several algorithms, each of which assigns a score[6] to a given jet that reflects how likely the jet is to contain a *b*-quark – or more precisely, a *B* hadron. The MV1 tagger [63] is based on an artificial neural network that combines the scores calculated by three other *b*-taggers, namely IP3D, SV1 and JetFitterCombNN, whose details can be found in [64, 65]. For the present analysis, the MV1c tagger, a version of the MV1 tagger with an improved rejection of *c*-quark jets is used[7], which provides additional suppression of background events from the production of *W* bosons in association with *c*-quark jets.

B hadrons have a mean lifetime of ∼1.5 ps, and given their typical momenta, they travel a distance of ∼1 mm on average before their decay. The tracks produced in the decay often appear significantly displaced with respect to the primary vertex, and point to a common secondary vertex. One or more of the following topological features are exploited by the three taggers:

- the decay length, i.e. the distance between the primary and the secondary vertex, taking the error estimates from the track and vertex fits are taken into account;
- the invariant mass of the tracks belonging the secondary vertex;
- the fraction of the tracks in the jet associated with the secondary vertex;
- the presence and properties of another secondary vertex that appears in *B*-to-*D* decay cascades;
- the impact parameter, avoiding the requirement to fully reconstruct a secondary vertex.

The decay length is measured in the direction of flight of the jet and thus presumably the *B* hadron. This projection defines a sign for the decay length. Negative values can only be reached due to resolution effects. The projection also cancels residual displacements orthogonal to the direction of flight.

[6] This score is called *tag weight* in the literature published by the ATLAS Collaboration, a terminology that is avoided in this thesis in order to prevent confusion with the scale factors used to re*weight* the *b*-tagging efficiency in the MC simulation to the one observed in data.

[7] The improved rejection of *c*-quark jets is achieved by training the neural networks against a mixture of light-flavour and *c*-quark jets instead of a pure sample of light-flavour jets.

Based on the score of the MV1c tagger, a binary classification of each jet is performed: if the score is above a cut value, the jet is said to be *b-tagged*, otherwise it is said to be *b-untagged*. The cut value depends on the *b*-tagging *working point*, and its choice represents a trade-off between the efficiency and rejection. The relationship between the cut value and the efficiency is established during the calibration of the *b*-tagger. For the calibration used in this analysis, a tag-and-probe method was applied to obtain a very pure sample of *b*-quarks from $t\bar{t}$ events where both top quarks decay semileptonically [66]. The rejection factors of *c*-quark jets and light-flavour jets were measured separately [63]. A sample of events with identified D^* mesons was used to determine the rejection of *c*-quark jets. The misidentification of light jets as *b*-tagged jets is driven by resolution effects, and the decay length and impact parameter distributions are essentially symmetric about zero[8]. This allowed the rejection of light-flavour jets to be estimated from the negative tails of these distributions.

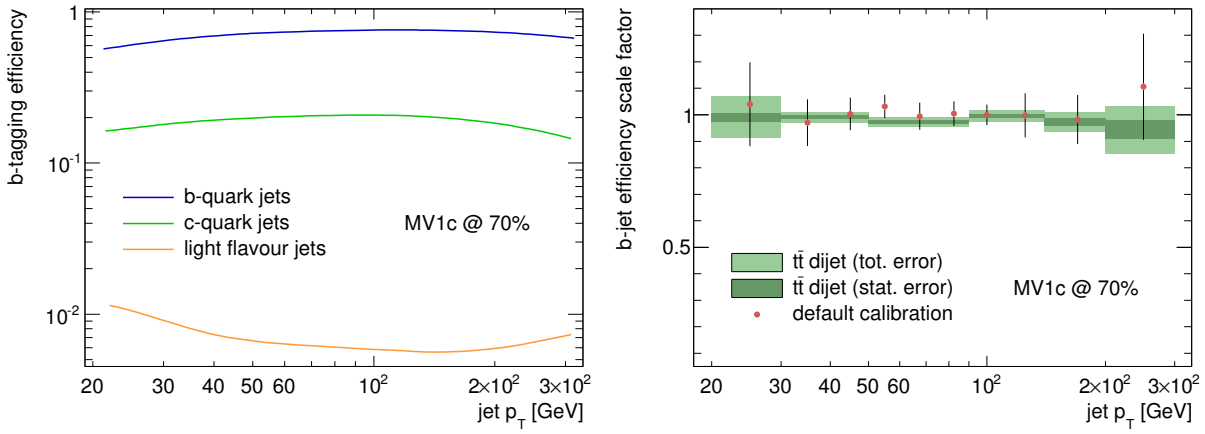

Figure 4.4: *b*-tagging efficiency for *b*-quark, *c*-quark and light-flavour jets in the Pythia simulation (left), and scale factors for *b*-quark jets (right), both for $|\eta(\text{jet})| < 1.2$ and as function of $p_{\text{T}}(\text{jet})$. In addition to the calibration using the dijet balance in $t\bar{t}$ events, the default calibration provided by the ATLAS Flavour Tagging group is shown. The two methods agree within the uncertainties. This figure does not show official ATLAS plots; the plots were created using the numbers of the MV1c calibration data that were distributed internally in the ATLAS Collaboration.

Following the optimisation of the working point for the selection of *Wt* events in the single-lepton channel in Ref. [67], the working point with a nominal efficiency of 70 % has been selected for the present analysis. The actual efficiency depends on the topology of the events, primarily the kinematics of the jets because the efficiency is not constant as function of η and p_{T}, as can be seen in Figure 4.4 (left). Simulated events are reweighted in order to correct the efficiencies for *b*-quark jets to the observed efficiencies. The weights are close to unity with an uncertainty at the percent level, which is illustrated in Figure 4.4 (right).

Jets with heavy-flavour decays naturally produce more muons and neutrinos compared to light-flavour jets. The energy carried by the muons and neutrinos escapes the jet finding, reducing the reconstructed energy of the jet by 5 % to 10 % [54]. As this happens equally in data and simulation, it is tolerable for physics analyses, and should just be kept in mind when reading plots that show the momentum of *b*-tagged jets or dependent quantities.

[8] Any asymmetries due to decays of long-lived particles like K^0_{S} or material interactions were corrected for.

4.3 Charged leptons

For the reconstruction of electrons and muons, the tracks from the Inner Detector are combined with the information from the Calorimeter and the Muon Spectrometer, respectively. The following discussion is focused on so-called *prompt leptons*. This is the term used for the charged leptons that have been created in the hard collision, including the decay chains of unstable[9] particles. All other charged lepton candidates, including real leptons produced in hadronic decays and material interactions, are collectively referred to as *fake leptons* (see Section 6.2.5). Tauons are not explicitly reconstructed, but their feed-down is included in the analysis (see Section 6.3).

4.3.1 Electrons

Electrons show up as tracks in the Inner Detector before dispensing almost all of their energy in the electromagnetic calorimeter. The reconstruction of electron starts in the calorimeter. After projecting the deposits onto an η–ϕ grid, with cell sizes according to the granularity of the calorimeter, isolated clusters[10] with $E_{\mathrm{T}} > 2.5\,\mathrm{GeV}$ are identified and used as seeds. The algorithm then tries to find a track that matches to the cluster within a region of $\Delta\eta \lesssim 0.005$ and $\Delta\phi \lesssim 0.05$ in the middle layer of the EM calorimeter. In order to account for bremsstrahlung, matches are allowed up to $\Delta\phi \lesssim 0.15$ in the bending direction of the track. The cluster energy is recalculated, taking adjacent energy deposits in the calorimeter into account, and an MC-based correction is applied. Further information on the reconstruction algorithm is given in [68, 69].

Tight electron identification cuts are applied in order to attain a high-purity sample of electron candidates. The barycentre of the cluster must be found within a pseudorapidity range of $|\eta| < 2.47$. The track is required to have at least 7 hits in the SCT and at least 2 hits in the Pixel detector, one of which is expected to be in the innermost layer. In order to suppress fake electron candidates from misidentified hadrons, p_{T}-dependent cuts on the shower shapes, leakage into the hadronic calorimeter, and $E(\mathrm{cluster})/p(\mathrm{track})$ are applied. At least ~10 % of the hits in the TRT must be high-threshold hits, because a significant amount of transition radiation is expected from electrons (see Section 3.2.1). The track must be compatible with the hard collision primary vertex, satisfying requirements on the transverse and longitudinal impact parameters, $|d_0| < 1\,\mathrm{mm}$ and $|z_0| < 2\,\mathrm{mm}$. If the track carries the signature of a photon conversion, the candidate is rejected.

Apart from these *tight* identification cuts, so-called *medium* identification cuts play an important role in the assessment of the background resulting from fake electrons. Apart from fake electrons caused by the misidentification of hadrons, principal sources of non-prompt electrons are semileptonic decays in b/c-quark jets, and photon conversions. The medium quality cuts are very similar to the tight cuts in the kinematic region of this analysis. The most notable difference is a looser requirement on the transverse impact parameter, $|d_0| < 5\,\mathrm{mm}$.

The direction of the electron candidate is very precisely determined by the tracker but the calorimeter provides a more accurate measurement of its energy. The transverse momentum[11] of the candidate is therefore constructed from a combination of the two measurements:

$$p_{\mathrm{T}}(e) = \frac{E(e\text{−cluster})}{\cosh\eta(e\text{−track})} \tag{4.2}$$

[9] Particles with a mean lifetime of less than 3×10^{-9} s are considered unstable, see Section 2.2.

[10] These clusters are to be distinguished from the three-dimensional topological clusters that are used by the jet reconstruction.

[11] Given the smallness of the electron mass compared to the energies considered here, the difference between p_{T} and E_{T} is safely neglected for electrons.

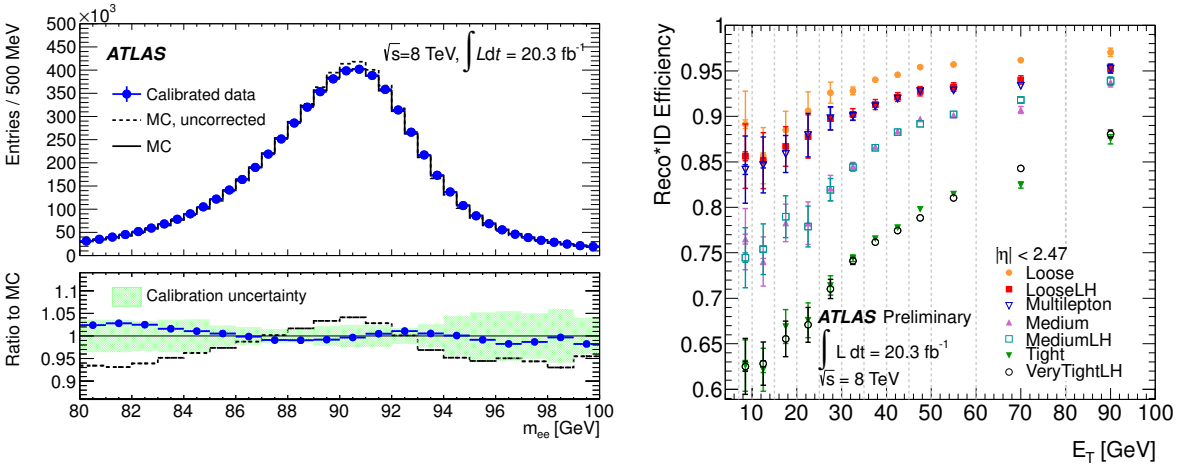

Figure 4.5: Invariant mass of electron pairs near the Z boson mass in data and the MC simulation before and after data-driven corrections (left) [70], and measured reconstruction efficiency for electrons as a function of $E_{\mathrm{T}}(e)$ (right) [71]. The efficiency includes the quality cuts, but does not include the isolation. The "tight" quality is the one used in the analysis.

The performance of the electron identification and energy measurement was evaluated by reconstructing the Z resonance in $Z \to e^+e^-$ events (Figure 4.5, left). The energy scale agrees between data and the MC simulation at the permille level. The energy resolution is about 1 % to 2 % with a 10 % uncertainty [70]. Small corrections of the energy scale and resolution are applied to the electron candidates in the MC simulation. The electron reconstruction and identification efficiency ranges from 72.5 % for transverse momenta of 30 GeV to 87.5 % for 100 GeV, with an uncertainty below 1 % (Figure 4.5, right).

In order to further enrich prompt electrons, the electron candidate must satisfy additional isolation requirements [44]. The energy measured in EM calorimeter cells within a radius of 0.2 around the electron in the η–ϕ plane must not exceed a certain p_{T}- and η-dependent threshold. A similar requirement is imposed on the summed transverse momenta of the tracks in a radius of 0.3 around the electron. Together these requirements have a nearly uniform efficiency of 85 % in p_{T} and η by design. The largest uncertainty on the efficiency is 2 % due to topological differences between top-quark events and the $Z \to e^+e^-$ events, in which the efficiency has been measured. After all identification cuts, the non-electron backgrounds are rejected by a factor of about 1×10^5.

4.3.2 Muons

Hits in the Muon Spectrometer are a strong indicator for the presence of a muon in an event, and so serve as starting point for the muon reconstruction. A pattern recognition algorithm looks for hits in the outermost part of the spectrometer, and works its way inwards. The track of the muon candidate is then extrapolated into the Inner Detector, where it is expected to match an Inner Detector track, reconstructed as outlined in Section 4.1. The two tracks are combined, and a global refit of the track is executed in order to improve the parameter estimates. Afterwards, an inside-out pass is performed, extrapolating the remaining Inner Detector tracks into the Muon Spectrometer, and adding the hits found there. The procedure is elaborated in Ref. [72].

Muons are reconstructed up to $|\eta| = 2.5$, and down to $p_{\mathrm{T}} \approx 10\,\mathrm{GeV}$, and with some limitations even below that. Due to the requirement of an Inner Detector track, the reconstruction efficiency drops

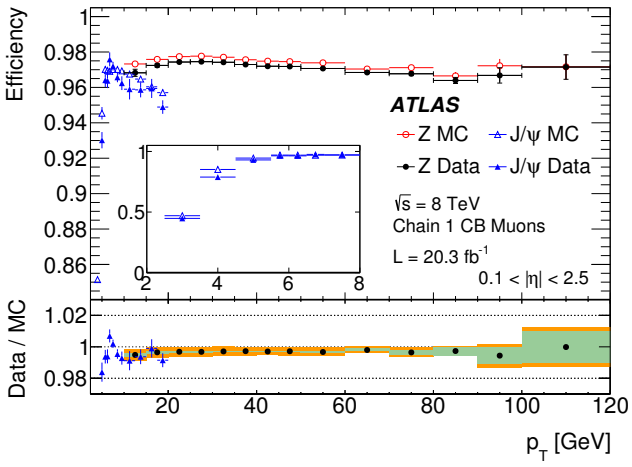

Figure 4.6: Reconstruction efficiency for muons as a function of $p_{\mathrm{T}}(\mu)$ for $0.1 < |\eta|(\mu) < 2.5$ (top), and its ratio between data and the simulation (bottom).[72] For the kinematic region that is relevant to this analysis, the efficiency is accessible through $Z \to \mu\mu$ events. Its statistical uncertainty is indicated by the error bars and the green areas, while the orange areas also include systematic uncertainties.

sharply for the pseudorapidities $|\eta| > 2.4$. The track is required to have at least 1 hit[12] in the Pixel detector, 5 hits in the SCT, and in addition 6 TRT hits if the muon falls in a region covered by the TRT ($0.1 < |\eta| < 1.9$). The muon must be compatible with the hard collision primary vertex.

The momentum scale and resolution were validated in $Z \to \mu\mu$ events[13]. A tiny shift and a smearing of the momentum are applied in the MC simulation in order to match the momentum scale and resolution found in observed data. The momentum scale is known with excellent precision, with an uncertainty at the permille level for reconstructed masses of the Z boson candidates, and a resolution of about $(2.0 \pm 0.1)\,\%$ [72][14]. In the phase space relevant to the present analysis, the accuracy of the momentum is dominated by the measurement in the Inner Detector. The reconstruction efficiency has been measured in $Z \to \mu\mu$ events, too, using a tag-and-probe method. It is close to 99 % with an uncertainty of about 0.25 % (Figure 4.6). Apart from an inefficiency for $\eta \approx 0$, which will be explained in Section 6.4, it approximately flat over the phase space relevant for this analysis.

Muon candidates selected in this way are very likely to be real muons, but still include a considerable fraction of muons from non-prompt sources, predominantly from heavy-flavour decays. In addition hadronic interactions in the detector material and punch-through from the calorimeter produce (mostly softer) muons which can accidentally match a track in the Inner Detector. In order to further enrich prompt muons, an isolation requirement is imposed:

$$\frac{1}{p_{\mathrm{T}}(\mu)} \sum_{t}^{T} p_{\mathrm{T}}(\text{track } t) < 0.05\,, \tag{4.3}$$

where the set T includes each track, t, that satisfies:

$$\Delta R(t,\mu) < 10\,\mathrm{GeV}/p_{\mathrm{T}}(\mu) \tag{4.4}$$

unless it is the track associated with the muon itself. The efficiency of the cut has been measured to be about 97 % in $Z \to \mu\mu$ events, with a 2 % uncertainty [72]. Afterwards, muons within the radius of 0.4

[12] In this context dead Pixel or SCT sensors crossed by the fitted muon track also count as hits. However, no more than two of such *holes* are allowed.

[13] $\Upsilon \to \mu\mu$ and $J/\psi \to \mu\mu$ events were also used for low momentum muons, but the muons produced via $Z \to \mu\mu$ mirror the kinematics of the muons in this analysis better.

[14] A dimuon mass resolution of 2 % implies a momentum resolution of $2\,\%/\sqrt{2}$ for a single muon, assuming that the dimuon system is at rest.

around a jet in the η–ϕ plane are removed. Here, all jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ are considered except those classified as pileup jets based on their JVF.

4.4 Neutrinos

Neutrinos constitute the only known kind of stable particle that have a completely negligible cross section for the interaction with the detector material. While they cannot be observed directly in the detector, their presence in an event can still be inferred from the conservation of the total energy–momentum. All particles that escape the detector unseen create a difference in the 4-momentum vectors of the colliding and the outgoing particles. While the boost along the beam axis is unknown, the initial transverse momenta of the colliding particles are small enough to enable the detection of missing momentum in the transverse plane, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, and therefore neutrinos.

In practice, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is calculated as [73][15]:

$$\left(p^{\mathrm{miss}} \approx 0 - \sum_{\ell \in \mathrm{leptons}} p(\ell) - \sum_{j \in \mathrm{jets}} p(j) - \sum_{c \in \mathrm{soft}} p(c) \right)_{x,y} \tag{4.5}$$

The last term of Equation 4.5, called the *soft term*, is a sum over all topological clusters that are not associated with any lepton or jet.

It is essential to capture the energy of jets and leptons as effectively as possible because any lost energy is attributed to the neutrinos. Therefore, the calibrated momenta of all jets and leptons are used with a comparatively loose kinematic selection. Jets are accepted over the full pseudorapidity range ($|\eta| < 4.9$), including those identified as pileup jets. Electrons are identified with the tight criteria given in Section 4.3.1, and failing that, appear as jets or clusters. The acceptance for muons is inherently limited by the coverage of the Muon Spectrometer to $|\eta| \lesssim 2.7$. The inclusion of leptons with low transverse momenta helps to cure losses from heavy-flavour decays inside jets (as discussed at the end of Section 4.2.1).

The systematic uncertainties on the calibrated momenta of jets and leptons are propagated into the missing transverse momentum. In addition the soft term comes with uncertainties of about 3.6 % on its scale, and about 2.3 % on its resolution. Apart from the difficulty to measure all objects in the event as precisely as possible over the full solid angle, the presence of pileup interactions adds a large amount of background noise to the calorimeter, making the reconstruction of missing momentum even more challenging. The relative resolution of $p_{\mathrm{T}}^{\mathrm{miss}}$ behaves as the inverse square root of the scalar sum over the transverse momenta entering Equation 4.5, and hence becomes worse as the number of pileup interaction increases.

In events where a single neutrino is expected from the decay of a W boson, it is possible to improve the estimate of the neutrino momentum by using the known mass of the W boson, m_W, as a constraint. The method exploits that m_W should be recovered when summing the 4-momenta of the charged lepton and the neutrino:

$$|p(\ell) + p(\nu)| = |p(W)| = m_W = 80.399\,\mathrm{GeV}\,. \tag{4.6}$$

This can be written as quadratic equation, and solved for the z component of the neutrino momentum, $p_z(\nu)$. Due to the nature of quadratic equations, there are two solutions if $M_{\mathrm{T}}(\ell\nu) < m_W$, and no real-

[15] Technically a weighted sum over all objects is used, which allows object momenta to contribute partially to the sum. These weights primarily encode the object selection used for the reconstruction of missing momentum, and are almost always either 0 or close to 1, so the formula given here is a good approximation.

valued solutions if $M_{\mathrm{T}}(\ell\nu) > m_W$. In the first case, the ambiguity is resolved by picking the solution with the smaller $|p_z(\nu)|$. The latter case occurs if the measured $p_{\mathrm{T}}^{\mathrm{miss}}$ is too large due to resolution effects or other imperfections of the reconstruction, and is resolved by adjusting $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ until a real-valued solution is found, under the constraint that the adjustment be minimal in the x, y plane. Details are given in Ref. [74, 75]. The procedure was also used for the measurement of singly produced top quarks via the t-channel exchange of a W boson with the ATLAS detector.[76]

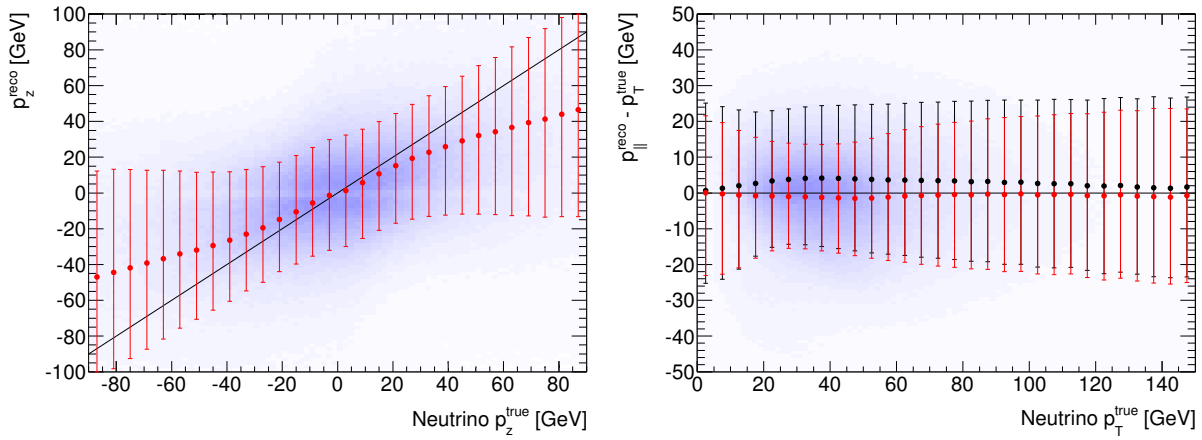

Figure 4.7: Distribution of the reconstructed momentum of the neutrino plotted against the true momentum. The *right* plot shows the momentum component along the true momentum in the transverse plane, while the *left* plot compares the component along the z axis. Superimposed are red markers representing the median and $\pm 1\sigma$ quantiles in slices of the true momentum, which indicate the bias and resolution of the reconstruction. For the transverse component, black markers indicate the bias and resolution of the reconstruction without the m_W constraint.

Figure 4.7 shows the performance of this method in $t\bar{t} \to \ell\nu q\bar{q}' b\bar{b}$ events after the analysis selection, which is discussed in Section 6.3. The correlation between the true and reconstructed z components is clearly visible, though it is considerably smeared out. Because the smaller solution is picked in the case of an ambiguity, the reconstructed p_z is biased towards zero. In order to work out the resolution of $\vec{p}_{\mathrm{T}}(\nu)$, the reconstructed $\vec{p}_{\mathrm{T}}(\nu)$ is split[16] into the components parallel, $p_{\parallel}(\nu)$, and perpendicular to the true neutrino direction. The parallel component is then compared to the true $p_{\mathrm{T}}(\nu)$, and the perpendicular component to 0. Compared to the default approach without the m_W constraint, the resolution of $p_{\parallel}(\nu)$ improves from more than 20 GeV to about 18 GeV. The apparent bias in the reconstructed momentum is due to the event selection, specifically the requirement that $M_{\mathrm{T}}(\ell\nu) > 50$ GeV, and is nearly avoided with the improved reconstruction.[17] The lateral component, not shown in the figure, has a similar resolution as $p_{\parallel}$, and is unbiased in both approaches.

[16] In direct comparisons of the reconstructed and true $p_{\mathrm{T}}(\nu)$, one finds that $p_{\mathrm{T}}(\nu)$ is biased towards larger values. This can be understood when considering that events with a vanishing true $p_{\mathrm{T}}(\nu)$ are reconstructed with a larger $p_{\mathrm{T}}(\nu)$ whose average magnitude is determined by the resolution. The spurious bias (also noted in Ref. [77]) is avoided by the splitting.

[17] Without a cut on $M_{\mathrm{T}}(\ell\nu)$, the improved neutrino reconstruction would be biased, whereas the $p_{\mathrm{T}}^{\mathrm{miss}}$ would be unbiased with an approximately constant resolution in $\vec{p}_{\mathrm{T}}(\nu)$.

CHAPTER 5

Top-quark physics

The existence of the top quark was predicted in 1973, when it became necessary to add a third quark generation to the predecessor of the CKM matrix for an explanation of CP violation in neutral kaon decays. The prediction was strongly supported by the discovery of the b-quark in 1977 [78]. In 1995, the top quark was finally discovered in $p\bar{p}$ collisions at the Tevatron [79, 80], once again showcasing the predictive power of the Standard Model. About 15 years later, studies of the top quark started at the LHC, which is justifiably referred to as a "top-quark factory".

This chapter starts with an introduction of the properties of the top quark in Section 5.1, and explains how top quarks are produced at the LHC, before moving over to more specific aspects of the production of single top-quarks. Section 5.2 is devoted to the definition of the Wt process, and closes with a phenomenological study of observables that can be used to identify the production of single top-quarks in association with a W boson.

5.1 Top-quark physics at the LHC

Figure 5.1 gives an overview over recent measurements of the cross sections for several physics processes involving the production of top quarks. It also shows the theoretical predictions, which are in very good agreement with the measurements. A review of the measurements and the theoretical predictions for the production of top quarks at the LHC can be found in Ref. [82].

As the Yukawa coupling to the Higgs boson is remarkably large compared to the other fermions, it is very possible that the top quark plays a key role in the deeper understanding of the electroweak symmetry breaking. Also, several New Physics models predict the existence of new, heavy particles which decay predominantly into top quarks. Even if these new resonances too heavy to be produced directly at the LHC, they can give rise to higher-order corrections that would manifest themselves as systematic deviations between the measured and the theoretically predicted Standard Model cross sections.

The top quark is the heaviest known elementary particle. Its mass is experimentally determined at the Tevatron as well as the LHC. The combination of all measurements [83] yields the world average for the top-quark mass:

$$m_t = 173.34 \pm 0.27(\text{stat}) \pm 0.71(\text{syst})\,\text{GeV}\,. \tag{5.1}$$

Due to its large mass and the option to decay into an on-shell W boson and a quark, the top quark decays after about 5×10^{-25} s, before hadronisation sets in. An interesting consequence is that its mass and its polarisation can be reconstructed from its decay products. The top quark decays almost exclusively[1] into a W boson and a b-quark. The decay is often incorporated in theoretical predictions by using the narrow-width approximation: the production and the decay of top quarks are factorised, i.e. the hard

[1] with a probability of 99.8 %, as can be directly deduced from Equation 2.4, neglecting masses.

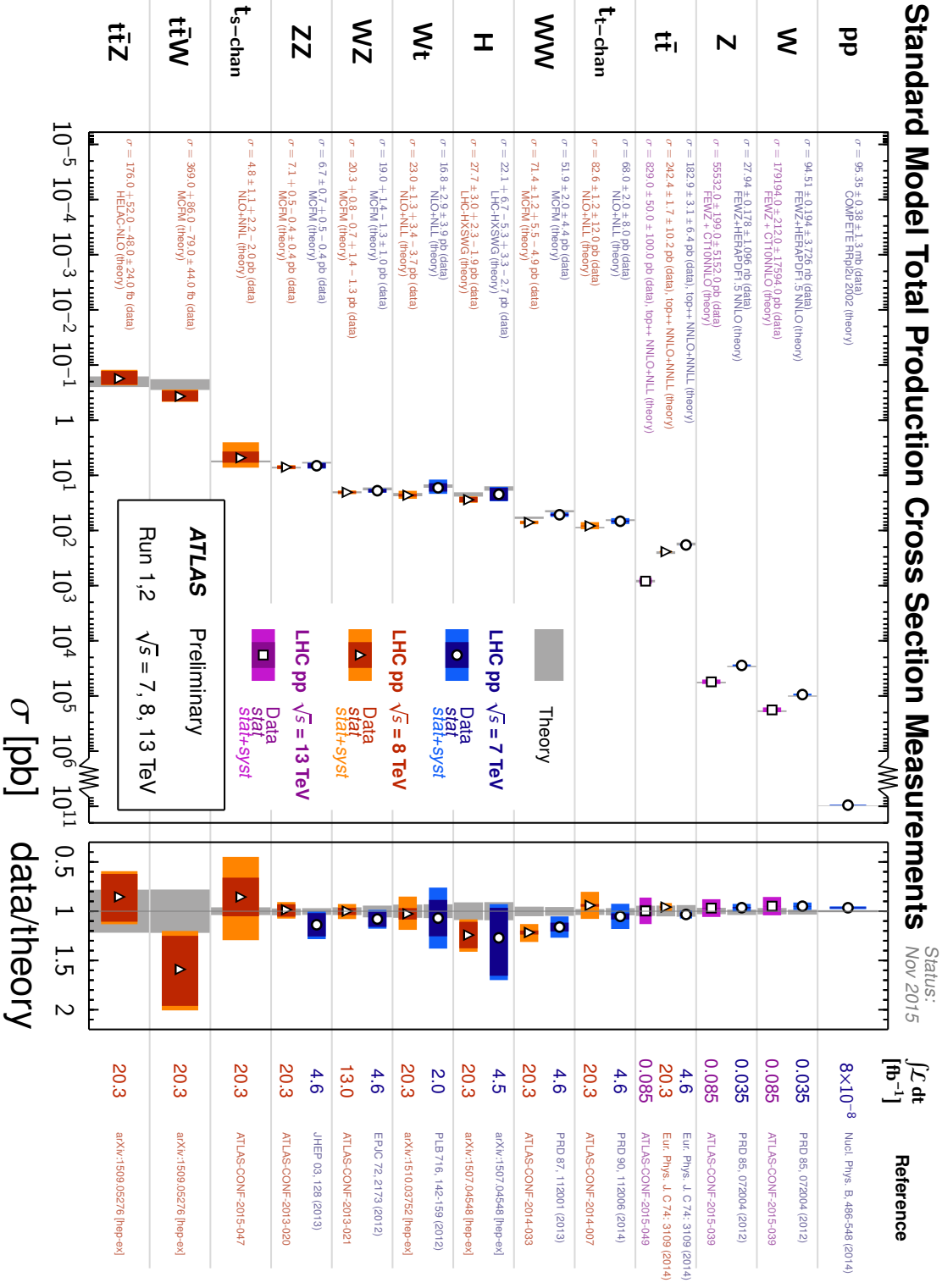

Figure 5.1: Measured inclusive cross sections for various Standard Model processes with the ATLAS detector, compared to theory predictions.[81]

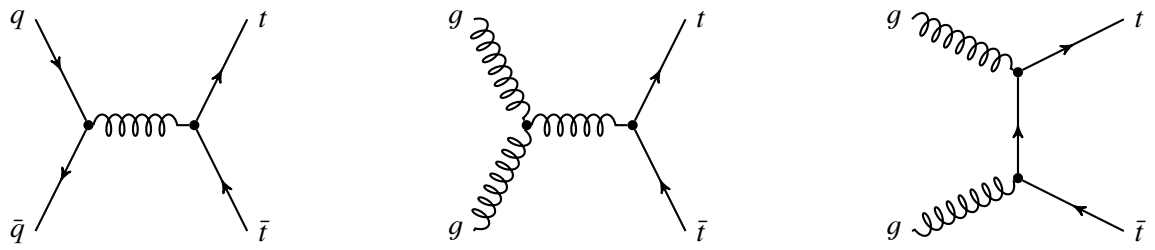

Figure 5.2: Feynman diagrams for the production of top-quark pairs at LO QCD: $q\bar{q} \to t\bar{t}$ (left), and $gg \to t\bar{t}$ (middle and right).

scatter is calculated assuming $\Gamma_t = 0$, and the finite Γ_t of the top quark is restored afterwards by a reshuffling of the momenta.

Figure 5.2 shows the representative diagrams for the production of top-quark pairs. At the LHC, the dominating contribution is from gluon fusion (90 %). The Tevatron operated at only $\sqrt{s} \approx 2\,\text{TeV}$, and since the two colliding partons need to have a centre-of-mass energy of at least $2m_t$, the threshold for top-quark pair production could be reached only through large x_1 and x_2. Taking the shape of the PDFs (Figure 2.3) into account, it is evident that the process $gg \to t\bar{t}$ was disfavoured at the Tevatron, which made $q\bar{q} \to t\bar{t}$ the primary (85 %) production mechanism. The theoretically predicted inclusive cross section for $t\bar{t}$ production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ is:

$$\sigma_{\text{th}}(pp \to t\bar{t}X) = 253^{+13}_{-15}\text{pb}\,, \tag{5.2}$$

calculated at NLO QCD including next-to-leading logarithmic (NLL) soft-gluon resummation with $m_t = 172.5\,\text{GeV}$ [84]. The uncertainty includes PDF and α_S variations according to the PDF4LHC prescription [85] as well as the scale uncertainty. The calculation is consistent with the experimentally determined cross section:

$$\sigma_{\text{obs}}(pp \to t\bar{t}X) = 241.5 \pm 1.4\,(\text{stat}) \pm 5.7\,(\text{syst}) \pm 6.2\,(\text{lumi})\,\text{pb}\,, \tag{5.3}$$

which is based on a combination of measurements by the ATLAS and the CMS collaborations [86].

The production of single top quarks proceeds through the weak interaction. It is subdivided according to the virtuality of the W boson that is mediating the interaction into: s-channel exchange of a W boson, W-associated production (Wt), and t-channel exchange of a W boson. The relevant Feynman diagrams are depicted in Figure 5.3. The theoretically predicted cross sections are $5.6 \pm 0.2\text{pb}$ for the s-channel exchange, $87.8^{+3.4}_{-1.9}\text{pb}$ for the t-channel exchange, and $22.4 \pm 1.5\text{pb}$ for the W-associated production [87–89].

The production of top-quark pairs proceeds through electroweak interactions as well, but its contribution of only about 0.2 pb can be neglected. Note that the distinction between the different production mechanisms is only an approximation. For the s- and t-channel production, it is valid up to NLO QCD; at NNLO, identical diagrams appear for the two processes, leading to interference. The interference of the W-associated production with top-quark pair production is addressed in Section 5.2.

While the s- and t-channel processes were already observed at the Tevatron [90, 91], the Wt mode eluded discovery until 2012, when it was finally discovered in pp collisions at $\sqrt{s} = 7\,\text{TeV}$ at the LHC [92, 93]. Recently published measurements based on pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with two leptons in the final state agree with the theoretical predictions within their uncertainties of $O(10\,\%)$ [94, 95]. Under some assumptions, V_{tb} can be directly measured in single top-quark production without relying

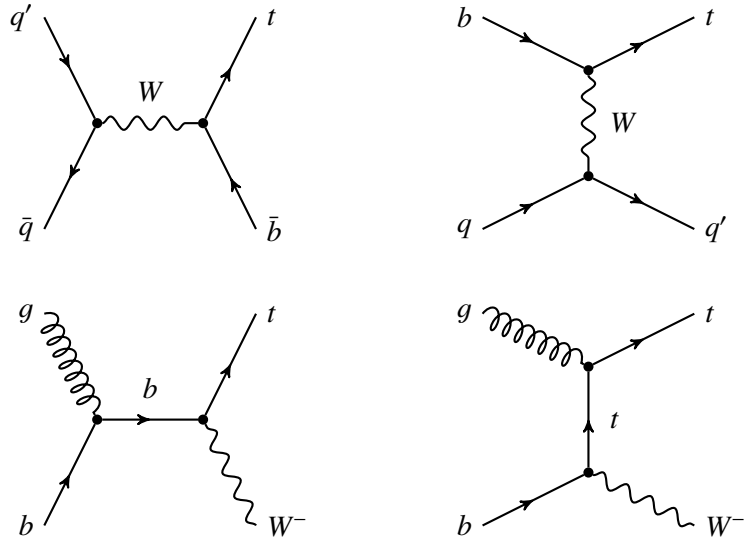

Figure 5.3: Feynman diagrams for the production of a single top quark at LO QCD: *s*-channel (top left) and *t*-channel (top right) exchange of a *W* boson, and *W*-associated production (bottom).

on the unitarity of the CKM matrix (see Chapter 9). Single top-quark production also plays an important role in the studies of physics beyond the Standard Model. As mentioned above, single top quarks can be produced in the decays of unknown heavy particles, such as vector-like or excited quarks [96] or superpartners of the top quark predicted by supersymmetry [97]. The Standard-Model production of a single top quark in association with a *W* boson is an important background to direct searches for such particles. Moreover, several extensions of the Standard Model predict effects to which the production of single top quarks is particularly sensitive [98]. Measurements of the different single-top-quark production processes can be used to systematically probe physics beyond the Standard Model in the context of Effective Field Theory [99].

5.2 Signatures of the *Wt* process

The analysis presented in this thesis targets event topologies where one *W* boson decays hadronically into two quarks, and another *W* boson decays leptonically into a charged lepton and a neutrino, as exemplified in Figure 5.4. The branching fraction of the leptonic decay of a *W* boson is 10.8 % per lepton flavour, so the total branching fraction for $WW \to \ell\nu q'\bar{q}$ is 43.8 %. There are several advantages and disadvantages compared to the dilepton topology (i.e. both *W* bosons decay leptonically) that was used in previous measurements of the *Wt* process. The total branching fraction for two leptonic decays is just 10.5 %, which translates into fewer observable events and an increased statistical error of the measurement. On the other hand, the dilepton topology is relatively free of backgrounds other than $t\bar{t}$ production. Also, charged leptons can be detected very well whereas measurements of jets tend to have to larger systematic uncertainties. The additional charged lepton comes at the price of an additional neutrino: only the sum of all neutrinos are measured, so with the dilepton topology one has no handle on the individual neutrinos and the four-momenta of the individual *W* bosons. The $t\bar{t}$ production poses a significant challenge for measurements in every topology. This will be discussed in detail in Chapter 7, but for the time being, Figure 5.4 already gives an impression how similar final states of the two processes look. On top of that, as shown in the previous section, the $t\bar{t}$ background is expected to have

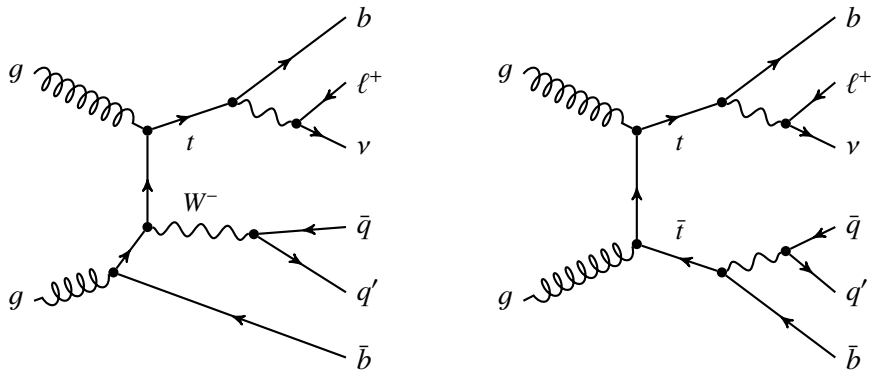

Figure 5.4: Example Feynman diagrams of Wt (left) and $t\bar{t}$ (right) production at LO QCD, including the decay of the top quark.

a more than 10 times larger cross section.

To make matters worse, the Wt and the $t\bar{t}$ production interfere with each other starting at the order of $\alpha_{\mathrm{EW}}\alpha_{\mathrm{S}}^2$, see Figure 5.4. It is not unusual that the distinctions between processes blur at higher orders, and interferences are not an uncommon phenomenon for analyses at the LHC. However, the $t\bar{t}$ production has a so much larger cross section than the Wt production, and the interference occurs already at a relatively low order in perturbation theory, so the question arises how well Wt production can actually be treated as a separate process. This was discussed and addressed in several ways by the theory community, e.g. in Refs. [100–103]. The prevailing view is that it is possible under certain conditions. In particular, the interference tends to be small in the phase space region that is most important for the measurement of Wt production. When both b-quarks have large transverse momenta, the effects of interference become large in the calculation of the Wt process, but these regions are anyway dominated by $t\bar{t}$ events, so one would not attempt to measure the Wt production there. In addition, two treatments of the interference term in the calculation of the Wt cross sections were developed:

1. In the Diagram Removal (DR) scheme, diagrams containing a second top-quark propagator that can be on-shell are removed from the amplitude.

2. In the Diagram Subtraction (DS) scheme, the $t\bar{t}$ contribution to the cross section is cancelled when the top-quark propagator is on-shell.

Physics analyses sensitive to Wt production were suggested to evaluate their results using both schemes. If the results are similar, the Wt production can be treated as a separate process, and the remaining difference can be taken into account as a systematic uncertainty. Recently, a calculation of the cross section for the $W^+W^-b\bar{b}$ final state at NLO QCD underpinned that the separation of Wt and $t\bar{t}$ as well as the use of the narrow-width approximation can be justified [104].

The rest of this section explores the characteristics of the events that can help to distinguish the two, giving an impression how well their $WWb(b)$ final states can be separated from each other. Particles after the parton showering are studied using Monte Carlo samples (see Section 6.2) for the Wt and the $t\bar{t}$ production. The W bosons are treated as stable[2], as their decays decouple from the dynamics of the hard scatter to a large extent.[3] The quarks and gluons are clustered into jets, upon which an η-dependent

[2] Except for the calculation of the top-quark polarisation.

[3] This statement cannot be generalised to other analysis, in particular precision analyses with vector bosons, where interference and colour reconnection effects in the decays of W and Z bosons can be very relevant!

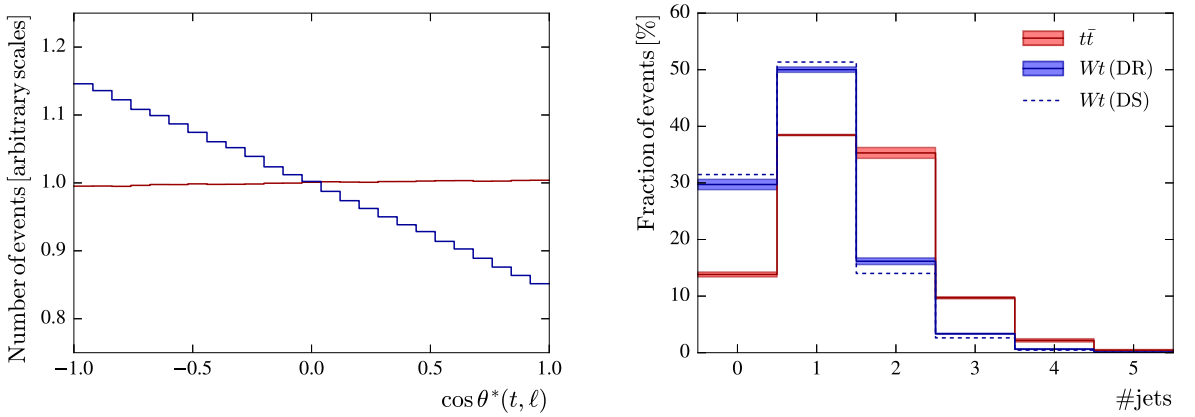

Figure 5.5: Distributions of the opening angle between the top quark and the charged lepton in the rest frame of the top quark (left), and in the number of jets (right), for Wt and $t\bar{t}$ events. The semitransparent band in the right plot indicates the uncertainty due to the renormalisation and factorisation scales (see Section 8.2).

cut (modelled after the rapidity-dependent jet veto reported in Ref. [105]) is imposed:

$$p_{\mathrm{T}}(\mathrm{jet}) \geq 30\,\mathrm{GeV} \cdot \cosh\eta(\mathrm{jet})\,. \tag{5.4}$$

Jets originating from b-quarks are identified by their net bottom quantum number, $n_b - n_{\bar{b}}$, summed over their constituent quarks.[4] Although the study was performed as part of the measurement presented in this thesis, no assumptions on the decay mode of the W bosons are made, so the conclusions hold equally for the single-lepton and the dilepton channels.

Due to the chiral nature of the weak interaction, singly produced top quarks are strongly polarised, as opposed to the top quarks produced in pairs. This is reflected in the opening angle, $\theta^*(t, \ell^+)$, between the top quark and the charged lepton (or the down-type quark) produced in its decay, viewed in the rest frame of the top quark:

$$\frac{\mathrm{d}N}{\mathrm{d}\cos\theta^*(t,\ell^+)} = (1 + P)\,\cos\theta^*(t,\ell^+)\,, \tag{5.5}$$

where N denotes the number of events, and P the polarisation. For completely unpolarised top quarks, the number of events would be constant as a function of $\theta^*(t, \ell^+)$. The actual distributions of $\theta^*(t, \ell^+)$ for Wt and $t\bar{t}$ events are shown in Figure 5.5. Unfortunately, the practical value of the polarisation for the analysis later on is somewhat limited due to the reconstruction of jets and neutrinos. Also, the combination $W^+\bar{b}$ often appears as a valid top-quark candidate at the experimental level (see Section 6.4), and since $\theta^*(W^+\bar{b}, \ell^+)$ peaks strongly at $-\pi$, the measured spectrum of $\theta^*(t, \ell^+)$ gets distorted. Moreover, for hadronic decays $W^+ \to u\bar{d}$, it is very difficult to reconstruct the sign of $\theta^*(W^+b, u)$ because the u- and the $\bar{d}$-quark jets have very similar experimental signatures.

Figure 5.5 also shows the number of jets, which tends to be smaller on average for the Wt process. Recalling its LO QCD diagram (Figure 5.3), where the splitting of the gluon into a $b\bar{b}$ pair happens inside the proton, one can anticipate that the b-quark from the splitting tends to have a smaller transverse momentum, and therefore does not lead to a formation of a jet. As suggested earlier in the discussion of the Wt–$t\bar{t}$ interference, events that do not have exactly one b-quark jet are rejected. Without loss of general-

[4] Initially, the flavour-k_t jet algorithm was considered for the studies presented here, but it was found to lead to a relatively poor correspondence between b-quarks and b-quark jets. This is probably due to an insufficient robustness against contributions from the underlying event, as was also noted recently in Ref. [106].

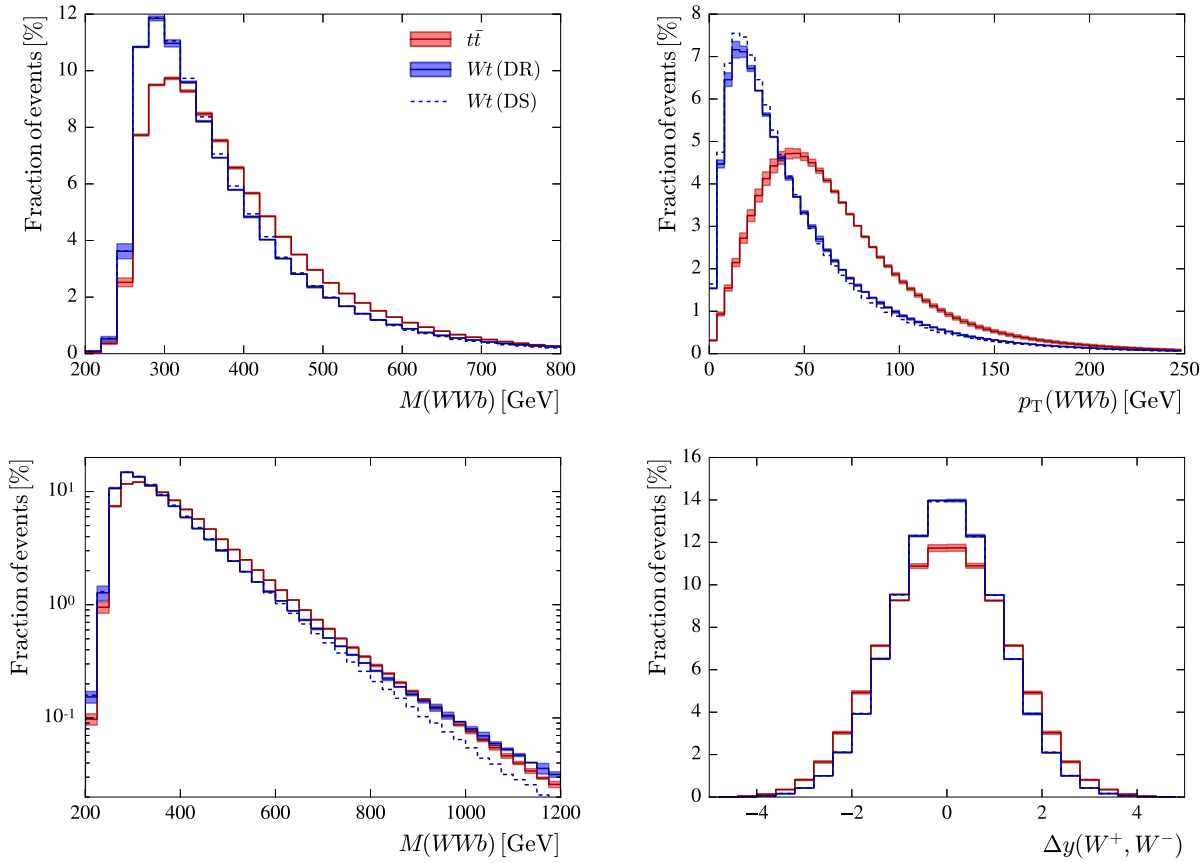

Figure 5.6: Distributions of the invariant mass (left) and transverse momentum (top right) of the *WWb* system, as well as in the rapidity difference between the two *W* bosons (bottom right), for *Wt* and $t\bar{t}$ events. The histograms for the invariant mass are shown on linear and logarithmic scale. The semitransparent band indicates the uncertainty due to the renormalisation and factorisation scales (see Section 8.2).

ity, the notation in the following assumes that the jet originates from a quark, not an antiquark. There are two observables that are distributed rather differently for *Wt* and $t\bar{t}$ events: the mass, $M(WWb)$, and the transverse momentum, $p_{\mathrm{T}}(WWb)$, of the system of two *W* bosons and the *b*-quark. The *Wt* events have softer spectra in both observables, as displayed in Figure 5.6. There is a notable discrepancy between the Diagram Removal and the Diagram Subtraction schemes in the tail of the $M(WWb)$ distribution: for DS, the spectrum continues to fall off at about the same rate as for $t\bar{t}$ production, but for DR, there is a crossover with the prediction for $t\bar{t}$ production. It is important not to rely on this behaviour in the analysis later on. Lesser but not still notable differences between *Wt* and $t\bar{t}$ production are found in the spectra of $\Delta y(W^+, W^-)$: for *Wt* events, the two *W* bosons are closer to each other in rapidity.

The projections of the distributions of *Wt* and $t\bar{t}$ events on one-dimensional axes shown so far do not tell the full story. In particular, the events exhibit a strong correlation between $M(WWb)$ and $p_{\mathrm{T}}(WWb)$, as displayed in Figure 5.7. Such correlations will play in important role for the suppression of the $t\bar{t}$ background in the analysis later on. As can be seen, with a suitable cut in the 2-D plane, a large fraction of the $t\bar{t}$ background could be removed. A systematic approach to the construction of a cut separating the *Wt*-enriched from the $t\bar{t}$-enriched regions will be discussed in Section 7.2.

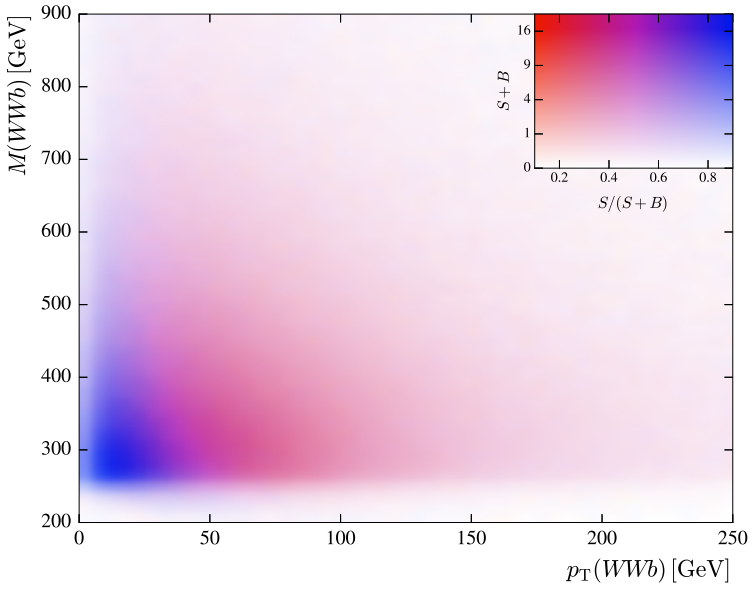

Figure 5.7: Normalised distributions for Wt and $t\bar{t}$ events as a function of the invariant mass and the transverse momentum of the W bosons and the b-quark jet. S and B are proportional to the number of signal respectively background events in each bin of the underlying 50×50 grid. The signal-to-background ratio and the number of events are translated into colours as indicated in the box in the upper right of the plot: the hue of each point indicates the signal-to-background ratio, using reddish colours for $t\bar{t}$- and bluish colours for Wt-enriched regions; the saturation scales with the square root of the number of events.

CHAPTER 6

Selection of Data and Simulated Samples

As shown in Section 5.2, the topology of *Wt* signal events can be characterised by:

- one charged lepton,
- missing transverse energy from the neutrino produced in the decay $W \to \ell\nu$,
- two light-flavour jets from the decay $W \to qq'$,
- one b-quark jet from the decay of $t \to Wb$,

each of which carries large transverse momentum. These characteristics are defining for the event selection, that is the set of requirements that the reconstructed physics objects must fulfill in order to be taken into account in the analysis. Other physics processes that can mimic the characteristic final state of the *Wt* signal need to be considered as backgrounds in the measurement.

This chapter starts with the details of the pp-collision data used for the analysis (Section 6.1). The measurement of the *Wt* cross section requires accurate predictions of the detector acceptance and efficiency for the *Wt* signal and the irreducible backgrounds, which will be discussed in the subsequent Section 6.2. The remaining sections are dedicated to studies performed specifically for this thesis: after the specification and study of the event selection (Section 6.3), the chapter concludes with an explanation of the control plots that are used to check the modelling of observed data (Section 6.4). Further details of the analysis chain are presented in Appendix A.

Since the *Wt* signal and the top-quark pair production background exhibit very similar topologies, the selection shares many aspects with those used in measurements of top-quark pair production with one charged lepton and jets in the final state. Consequently many well-established procedures, for example the estimation of the fake-lepton background and the evaluation of systematic uncertainties, could be transferred to this analysis.

6.1 Observed data

This analysis is based on pp-collision data at $\sqrt{s} = 8$ TeV recorded with the ATLAS detector in the year 2012.

The analysed data were collected when the proton beams were stable, and all detector components relevant for the reconstruction of physics objects (see Chapter 4) were fully operational and without any significant problems. A team of experts monitored the beam conditions and the detector performance around the clock, and based on their assessment problematic data were identified and marked[1]. A good-

[1] The decision made during data-taking is not final, and parts of the data can usually be recovered when a problem has been understood.

runs list [107] was assembled accordingly, consisting of luminosity blocks from 265 runs that were deemed *good for physics*.

In addition, a few individual events need to be rejected. Occasionally a large fraction of cells in the LAr calorimeter gave large signals with distorted shapes for a very short time (5 ns) [108]. Events affected by these *noise bursts* were identified based on the shapes of the electronic signals from the calorimeter, and are vetoed in the analysis. The loss of luminosity is taken into account in the luminosity calculation. A minuscule number of events with incomplete information or corruption in the data from the Tile calorimeter is excluded from the analysis. Minor problems, such as a power trips in an isolated module, are tolerated as they were corrected for during reconstruction.

Due to the excellent performance of the detector in 2012, more than 95 % of the recorded data are also good for physics, corresponding to an integrated luminosity of $\mathcal{L} = 20.3\,\text{fb}^{-1}$ with a relative uncertainty of 2.8 %. This translates to a total of about 1.5×10^{15} inelastic pp collisions. Figure 6.1 shows how the good-for-physics data were accumulated in the course of the year. The dataset is subdivided into ten periods, during each of which no modifications with a possibly substantial effect on the detector performance were made. Averaged over all periods, the number of interactions per bunch crossing is about 20 (see Figure 6.3). The time dependence the LHC conditions (such as the average number of interactions per second, the longitudinal position of the pp interaction point, and the efficiencies of the triggers), was taken into account in the MC simulation.

6.1.1 Triggers

As explained in Section 3.3, pp-collision events of potential interest are captured with a suitable set of triggers. The most efficient basis for the present analysis is held by the non-prescaled lepton triggers with the lowest p_{T} threshold, in particular the `e24vhi` chain for electrons, and the `mu24i` chain for muons. The number in the name of the trigger chain indicates the minimum transverse momentum in units of GeV that the lepton candidate, as reconstructed by the trigger, must possess. The triggers also require the lepton candidate to be isolated. Since this has a somewhat detrimental effect on the efficiency for leptons with larger transverse momenta, each trigger chain is complemented by a second chain with a higher threshold and no isolation cut, namely `e60` for electrons and `mu36` for muons. Due to the choice

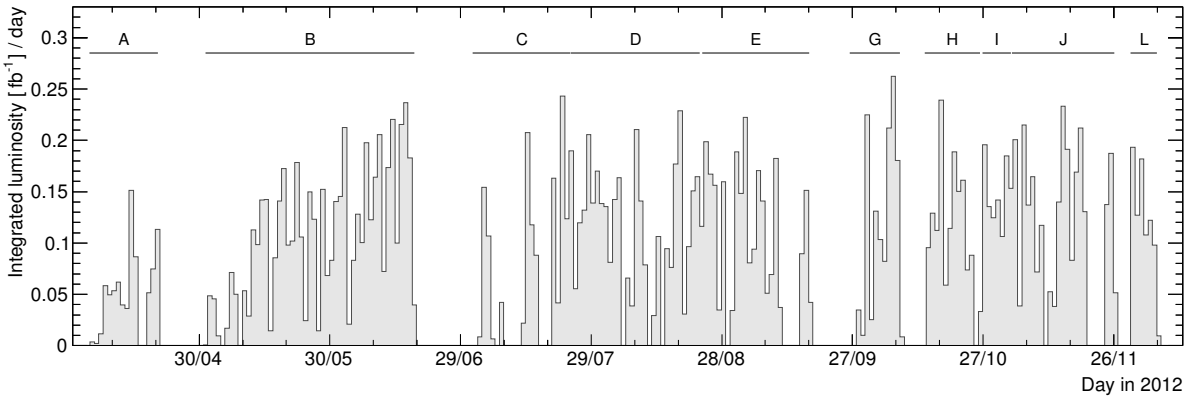

Figure 6.1: Integrated luminosity recorded on each day during the data taking in the year 2012, restricted to the luminosity blocks used for the present analysis. The extents of the data-taking periods, labelled with letters, are indicated as well. Apart from the small breaks between the stable beams phases of individual runs, there were O(1 d) interruptions for special runs like studies of cosmic-ray events and van-der-Meer scans (see Section 3.1). The major gaps in April, June and September are due to the technical stops.

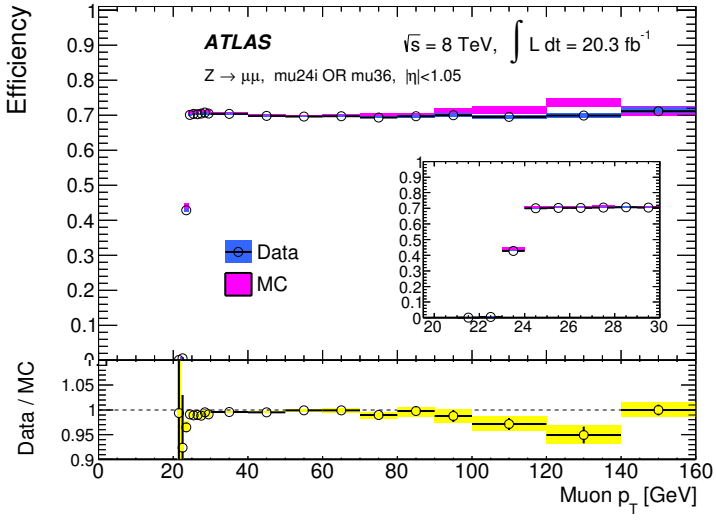

Figure 6.2: Combined efficiency of the `mu24i` and `mu36` trigger chains as function of the muon p_{T} in the barrel region in data and the simulation. The insert shows the details in the region near the turn-on threshold. The error bands include both statistical and systematic uncertainties.[109]

of triggers the EGAMMA and MUONS data streams are processed for the analysis, each of which contains about 1.4 billion good-for-physics events.

The combination of the two electron triggers has an efficiency of about 98 % for electrons in the barrel region of the LAr Calorimeter, dropping to 90 % to 95 % in the endcaps [44]. The efficiency of the muon triggers is about 70 % in the barrel region of the Muon Spectrometer, and 85 % in the endcaps [109]. The difference between the two regions arises from the limited coverage of the barrel part with Resistive Plate Chambers, which is due to the gaps at $\eta \approx 0$, the magnet ribs and the feet of the ATLAS detector. The plateau of the efficiency curve is reached between 24 and 25 GeV, as can be seen in Figure 6.2. The efficiency in data was determined with a tag-and-probe method, similar to the measurement of the reconstruction efficiency explained in Chapter 4. The systematic uncertainty on the efficiencies is 0.5 %, separately for electrons and muons.

6.2 Signal and background predictions

The procedures for obtaining Standard Model predictions for the outcomes of pp collisions are subdivided according to whether the predicted events contain a prompt lepton. Contributions from events with at least one prompt lepton are estimated using theory calculations, Monte Carlo sampling of events, and detector simulations (see Sections 2.2 and 3.5). The samples are centrally managed in the ATLAS experiment, and were used in numerous studies and analyses. There is a small but non-negligible chance that an event without a prompt lepton looks like it had one, and hence passes the event selection. This contribution is estimated using the data-driven method discussed in Section 6.2.5.

Event weights

In order to accurately reproduce the conditions under which the observed data have been recorded, the simulation is reweighted in the number of overlayed pileup interactions, and the longitudinal position of the pp interaction point. For illustration, the distribution used for the pileup reweighting is shown in Figure 6.3. The simulated samples are part of either the MC12A, which started before the data taking, or the MC12B production campaigns. The latter has been invoked at the end of data taking, and comes with an improved pileup overlay, having the distribution of the number of interactions per bunch-crossing, μ,

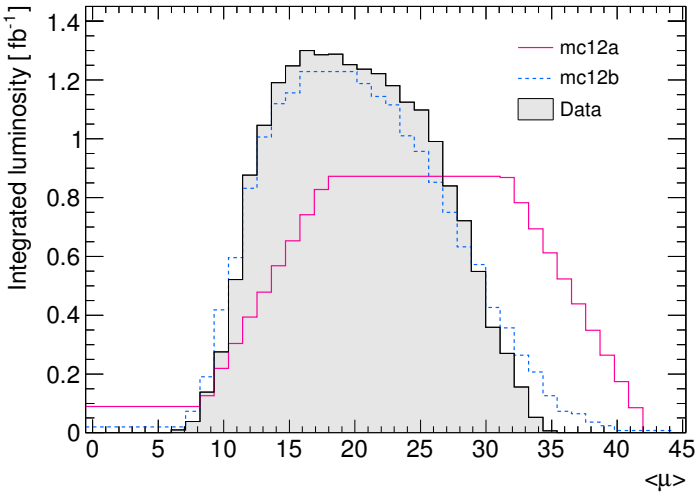

Figure 6.3: Distributions of the average number of interactions per bunch crossing, $\langle\mu\rangle$. The observed distribution is compared to the those in the MC12A and MC12B simulations before the reweighting in $\langle\mu\rangle$. The average number of interactions is measured over individual luminosity blocks, and can vary by roughly a factor of 2 over a run.

close to that observed during data taking. Each simulated event receives a weight given by the ratio of the data and MC distributions evaluated at the simulated μ. A similar reweighting is performed in the longitudinal position of the pp interaction point. Time-dependent corrections are implemented using a simulated run number, which is randomly drawn from the distribution of the integrated luminosity over the runs in observed data (essentially the distribution in Figure 6.1 with the run number instead of the date on the x axis).

Afterwards, each simulated sample is normalised to the number of events, N_{exp}, that is expected for the simulated final state:

$$N_{\text{exp}}(pp \to X) = \sigma_{\text{th}}(pp \to X) \cdot \mathcal{L}\,. \tag{6.1}$$

The cross sections are taken from theoretical calculations (see Section 5.1), and multiplied by the branching fraction of the W and Z bosons into the leptons. The uncertainties on the cross sections and the luminosity are propagated through the analysis.

Putting together the weights from the MC generator, the corrections of the simulation and the reconstruction, and the sample normalisation, the weight of the i^{th} event is given by:

$$w_i = w_i^{\text{sim}} \cdot \frac{N_{\text{exp}}}{\sum_{k=1}^{N_{\text{evt}}} w_k^{\text{sim}}} \cdot w_i^{\text{trigger eff.}} \cdot w_i^{\text{lepton eff.}} \cdot \prod_{j=1}^{N_{\text{jets}}} w_{ij}^{\text{b-tag eff.}}\,, \text{ with} \tag{6.2}$$

$$w_i^{\text{sim}} = w_i^{\text{generator}} \cdot w_i^{\text{pileup}} \cdot w_i^{\text{interaction point}}\,, \tag{6.3}$$

where N_{evt} refers to the number of entries produced by the MC generator.

While the number of observed events follows Poisson statistics, with a relative statistical error approximately given by $1/\sqrt{N_{\text{evt}}}$, the statistical error on the MC simulation depends on the weights. Some NLO generators produce events with negative weights as a result of the NLO subtraction scheme. In addition, the reweighting of the MC simulation significantly reduces the statistical power of the sample, which can be easily understood when considering the extreme case of the pileup reweighting because it assigns a weight of zero to many events and thus essentially reduces the size of the dataset. The *effective number of entries* is used to quantify the statistical power of such weighted samples:

$$N_{\text{eff}} = \left(\sum_i^{\text{entries}} w_i\right)^2 \Big/ \left(\sum_i^{\text{entries}} w_i^2\right). \tag{6.4}$$

It scales linearly with the size of the dataset. The relative statistical error of the weighted dataset can then be written $1/\sqrt{N_{\text{eff}}}$. The statistical power of the simulated events can be directly compared to an equivalent set of observed events using the ratio $N_{\text{eff}}/N_{\text{evt}}$, which is 1 for observed data. The statistical errors of the simulated samples enter the analysis as systematic uncertainties.

Generator versions, tunes, and structure functions

The NLO generators MC@NLO version 4.06 [110], and the Powheg method [111, 112] implemented in Powheg-Box-v1 [113] were used, each with the CT10 [114] set of proton structure functions. The parton shower and the hadronisation were simulated using either Herwig version 6.520 [115], or Pythia [116]. For Herwig, the AUET2 tune [117] with the CT10 structure functions[2] were used, and the underlying event was generated with Jimmy version 4.31 [118]. Pythia appears in two configuration: an older one that includes Powheg-Box-v1 revision 2129 with Pythia version 6.426, and a newer one coupling Powheg-Box-v1 revision 2330 with Pythia version 6.427[3]. There are no known differences between the two configurations for the physics processes under consideration. Unless noted otherwise, the Perugia 2011C [119] tune was used for Pythia, where the "C" stands for the variation with the CTEQ6L1 [120] set of proton structure functions. For the generator Sherpa [121–123], version 1.4.1 was used with the CT10 structure functions. The multileg generator Alpgen version 2.1.4 [124], coupled to Pythia with the Perugia 2012 tune, was used exclusively for the evaluation of systematic uncertainties. QED radiative corrections in the decays of resonances were generated with Photos [125]. The decays of (polarised) tauon were generated with Tauola [126].

The parameter for the top-quark mass was set to $m_t = 172.5\,\text{GeV}$, and the decay width of the top quark to $\Gamma_t = 1.32\,\text{GeV}$. For other physics constants, in particular the mass, decay width and branching fractions of the W and Z bosons, the electroweak mixing angle, the CKM matrix elements, and α_{em} the Particle Data Group world averages of 2010 [12] were used.

The b-tagging efficiencies in samples showered with Pythia, Herwig and Sherpa slightly differ from each other, so the appropriate set of MC-to-data corrections is used for each of the three generators.

While the dependence the cross section on the proton structure functions is part of the uncertainty on the theoretical cross section, the effect of the choice of the functions on the detector acceptance and selection efficiency is estimated approximately as an additional uncertainty of 1 % on the normalisation of the Wt signal and the $t\bar{t}$ background, following the studies in [127].

6.2.1 Wt signal

The prediction for the Wt signal was calculated with Powheg coupled to Pythia, treating the Wt–$t\bar{t}$ overlap with the diagram-removal scheme. The setup was chosen as close as possible to the setup for top-quark pair production (see below) in order to avoid spurious differences between the Wt and the $t\bar{t}$ samples. The full detector simulation is used as default, but since its statistical power is quite limited with $N_{\text{eff}}/N_{\text{evt}} \approx 1.4$, a larger sample made with the fast detector simulation is used in its place for a few selected purposes.

In the generation of this particular process, all decay modes of the prompt W bosons were included, implying that in 45.7 % of the simulated events both W bosons decay hadronically. In the present

[2] The structure functions used for the generation of the parton shower do not need to be the same as for the generation of the hard scattering process. Also, the parton showering depends only marginally on the choice of the functions because only ratios of functions enter the calculation.

[3] The choice of the older (newer) configuration coincides with the use of mc12a (mc12b), so the versions can be deduced from the tables in Appendix A.

analysis, such events are explicitly removed for the sake of consistency.[4]

Alternative samples were generated for the evaluation of systematic uncertainties. The diagram-subtraction scheme (see Section 5.2) is used to estimate the uncertainty due to the treatment of the Wt–$t\bar{t}$ overlap. The uncertainties on the parton showering and the hadronisation model are assessed by comparing the nominal prediction with a sample of Wt events that were showered with Herwig. The difference between samples produced with MC@NLO and Powheg, both coupled to Herwig, is used to estimate the uncertainty due to the choice of the NLO subtraction scheme. The uncertainties on the renormalisation scale, μ_R, and the factorisation scale, μ_F, are evaluated by varying each of the two scales independently by factors of 0.5 and 2.0. The renormalisation scale in the parton shower and the matrix element are varied simultaneously.

Both Powheg-Box and MC@NLO compute Wt events in the narrow-width approximation (see Section 2.2), but they do not apply momentum reshuffling to the top quark and the associated W boson. Consequently the two particles always appear exactly on-shell in the simulation. As shown in Appendix B, the effect is negligible in the context of this analysis. The W boson from the top-quark decay does not share this limitation.

6.2.2 Top-quark pair production

Events containing a $t\bar{t}$ pair produced via the strong interaction were generated with Powheg+Pythia. The decision for Powheg+Pythia was based on an extensive comparison between measurements and the predictions for various choices of generator and parameters, summarised in [128]. The hdamp parameter was set to m_t, deviating from the Powheg default of ∞. The full detector simulation was used to generate the sample with high statistical power of $N_{\text{eff}}/N_{\text{evt}} \approx 15$.

At least one of the two top quarks was required to decay semileptonically in the production of this sample. As a consequence the sample is normalised to the total $t\bar{t}$ cross section quoted in Section 5.1 multiplied by the branching fraction.

The same procedures as for the Wt signal (see Section 6.2.1) are employed to determine the uncertainties due to the NLO subtraction scheme and the generator for the parton shower and the hadronisation. The uncertainty due to the choice of the renormalisation scale, μ_R, is evaluated with due regard to [129]. The evaluation is based on samples produced with Alpgen, coupled to Pythia. The value for $\alpha_S(\mu_R)$ in the parton showering, usually evaluated at the scale $\mu_R = p_\perp$, is varied between $\mu_R = 0.5p_\perp$ and $\mu_R = 2p_\perp$, and simultaneously the value of μ_R in the hard scattering process is varied by a factor of 0.5 and 2.0, respectively[5]. No variation of the factorisation scale, μ_F, is performed[6].

6.2.3 Electroweak production of a W boson in association with jets

The production of a W boson in association with jets constitutes the second-most important background for the analysis. The final state for the associated production of a W boson and 3 jets looks very similar to the Wt signal, especially when considering the experimental limitations regarding the identification of b-quark jets.

Sherpa was used to generate such W+jets events, where the W boson decays leptonically. The events are subdivided according to their flavour content:

- "b" if there is a b-hadron with $|\eta| < 4$;

[4] The event selection suppresses such events by a factor of about 3 000, which is well below the sensitivity of this analysis.
[5] The variation is often referred to as "variation of the initial/final state radiation".
[6] At the time of writing, ATLAS samples generated with Powheg were still in preparation.

- "c" if there is no such b-hadron, but a c-hadron with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $|\eta| < 3$;
- "light" otherwise.

The flavour composition is allowed to vary in the analysis. The samples of the processes involving the two heavier quarks are merged together in the analysis unless indicated otherwise. The c- and b-quarks were treated as massive in the calculation. Since the total cross section of these processes is about two orders of magnitude above that of top-quark production, the fast detector simulation was used to generate samples with reasonable statistical power. The samples of the different flavours have different statistical power, ranging from about $N_{\mathrm{eff}}/N_{\mathrm{evt}} \approx 2$ times for bottom, and 0.4 for charm, to 0.1 for light flavours.

The samples were normalised to the cross section calculated at NNLO in QCD [130]:

$$\sigma(pp \to W^{\pm}X \to \ell \pm \nu_{\ell}X) = (36.3 \pm 1.8)\,\mathrm{nb}\,. \tag{6.5}$$

Each lepton flavour contributes with 7.07 nb (5.02 nb) to the total W^+ (W^-) cross sections.

6.2.4 Other prompt lepton backgrounds

Single top-quark production via s/t-channel. The samples of the backgrounds involving the production of single top quarks were made with Powheg+Pythia. Only semileptonic decays of the top quark were generated. The samples were normalised according to the cross sections and branching fractions given in Chapter 5.

Z boson production in association with jets. Sherpa was used to generate Z+jets events where the Z boson decays into a pair of charged leptons. The configuration was otherwise identical to W+jets. The samples were normalised to the cross section calculated at NNLO in QCD [130]:

$$\sigma(pp \to Z^0X \to \ell^+\ell^-X) = (3.72 \pm 0.19)\,\mathrm{nb}\,. \tag{6.6}$$

Diboson production (WW, WZ, ZZ). The production of two vector bosons was generated with Herwig at LO QCD. While Herwig simulates all possible the decay modes, a filter was applied requiring that at least one lepton with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ and $|\eta| < 2.8$ be present in the event. The remaining events were normalised to a cross section of 29.4 pb. The WW process is the only background besides $t\bar{t}$ that can produce two real W bosons. While the cross section for diboson production is comparable to the predicted signal cross section, the chance to produce a b-quark jet in addition is fairly low and acceptance is much smaller, making the diboson samples the least relevant of the simulated samples.

6.2.5 Fake-lepton background

The contribution from events containing fake leptons[7] is estimated using a data-driven method. Such events come predominantly from QCD-induced multijet production, where a jet mimics an electron, or produces a muon in a semileptonic decay. While the chance to accept such an event is small, the cross section for jet production is many orders of magnitude larger than the cross section of the electroweak processes, so their contribution cannot be neglected. The fake leptons background also includes

[7] The term refers to lepton candidates when there was actually no such lepton in the event as well as non-prompt leptons, see Chapter 4.

a tiny contribution from the variants of the previously considered electroweak processes where all of the produced real vector bosons decay hadronically.

The data-driven procedure introduced here is called *matrix method* [131]. Due to the very small acceptance for multijet events an MC simulation would be a rather inefficient approach. The idea of the matrix method is to obtain an estimate of the fake-lepton background from the observed data using the side-band of the lepton isolation: one loosens the lepton isolation cut slightly, changing the balance of the prompt and the fake leptons in favour of the fake leptons. The change in the number of accepted events can be used to extrapolate the number of fake-lepton events.

If a lepton candidate passes the relaxed isolation cut, it counts as a *loose* lepton, and if it passes the tight isolation, it also counts as a *tight* lepton. The probability that a prompt loose lepton is also identified as a tight lepton is called the *real efficiency*, ϵ_{prompt}. The *fake efficiency*, ϵ_{fake}, is similarly defined using fake leptons instead of prompt leptons. Given that the efficiencies are relatively independent of the analysis, they can be factored out:

$$\begin{aligned} N^{\text{loose}} &= N^{\text{loose}}_{\text{prompt}} + N^{\text{loose}}_{\text{fake}} \\ N^{\text{tight}} &= N^{\text{tight}}_{\text{prompt}} + N^{\text{tight}}_{\text{fake}} \\ &= \epsilon_{\text{prompt}} N^{\text{loose}}_{\text{prompt}} + \epsilon_{\text{fake}} N^{\text{loose}}_{\text{fake}} \end{aligned} \tag{6.7}$$

The N are understood to be the numbers of events with one prompt or fake lepton passing the loose or tight selection. The equation system resembles a 2×2 matrix equation, hence the name "matrix method". It can be solved for the number of fake leptons passing the analysis selection, $N^{\text{tight}}_{\text{fake}}$.

In practice the efficiencies depend slightly on the event topology, and therefore the efficiencies are parametrised as a function of variables such as the transverse momentum and polar angle of the lepton, the angular separation between the lepton and a nearby jet or the missing energy. The dependencies of the efficiencies on the variables are then combined multiplicatively. Different sets of efficiencies are used depending on the number of (b-tagged) jets, and for leptons with high momenta that are supposed to have fired the non-isolated trigger with the higher p_{T} threshold. Based on the efficiencies, a weight is assigned to each event:

$$w = \frac{\epsilon_{\text{fake}}}{\epsilon_{\text{prompt}} - \epsilon_{\text{fake}}} \left(\epsilon_{\text{prompt}} - \begin{cases} 1 & : \text{tight selection passed} \\ 0 & : \text{only loose selection passed} \end{cases} \right) \tag{6.8}$$

Events that pass only the loose selection have in general $w > 0$, whereas those passing the tight selection have $w < 0$, corresponding to an effective subtraction of the observed prompt-lepton events from the fake-lepton background.

The real and fake efficiencies were measured [131] in control regions which are representative for the signal regions in terms of kinematics. A sample of real electrons (muons) from $Z \to ee$ ($Z \to \mu\mu$) decays, obtained with a tag-and-probe method, was used to derive the real efficiencies. As it is important to choose the control region for determining the fake efficiency such that it reflects the mixture of the different sources of fake leptons in the signal region, the control region for fake electrons was defined by having $M_{\text{T}}(\ell\nu) < 20\,\text{GeV}$ and $p^{\text{miss}}_{\text{T}} + M_{\text{T}}(\ell\nu) < 60\,\text{GeV}$ instead of the "nominal" cuts that enrich leptonic W decays (see Section 6.3). This particular definition was used only as an alternative control region for fake muons. Since fake muons are often real but non-prompt muons originating from a heavy-flavour decay, the preferred control region was defined by changing the cut on the transverse impact parameter significance to $|d_0|/(\Delta d_0) > 5$. The "contamination" of the control region with prompt leptons was estimated using the MC simulation. The dominant uncertainty was due to the normalisation of the simulated processes, and is propagated into the analysis, together with additional uncertainties due to

the choice of control region and parametrisation of the efficiencies.

6.3 Event selection

This section specifies the event selection, which consists of a sequence of requirements (cuts) that are imposed on the reconstructed physics objects in each event. The selection is designed to let events with one leptonically decaying W boson and jets pass, while rejecting other events. Most of the requirements are based on the characteristics of the Wt topology, and thus favour signal events over background events that do not share these characteristics.

The selections for electron and for muon candidates are executed independently at first, and the surviving events are merged after the selection. Tauon candidates are not explicitly reconstructed, but the feed-down from tauons is included. A tauon can decay into an electron (muon) with a probability of 17.8 % (17.4 %), resulting in an observed final state that is similar to one encountered if the W boson had directly decayed into an electron (muon). The remaining 64.8 % of the time, the tauon decays hadronically, and may be identified as jets. Regardless of the decay mode, a tau neutrino is produced in the decay due to the conservation of the lepton flavour quantum numbers. In consequence of the inherent difficulty to reconstruct neutrinos, the measurement of the kinematics of the W boson is less precise when it decays into a tauon.

An overview of the sequence of cuts is presented in Figure 6.4, which will be discussed step by step in the coming paragraphs. The figure also shows the (effective) efficiencies of each cut for the expected Wt signal, the major simulated backgrounds, and observed data, defined as:

$$\epsilon(\text{cut } i) = \frac{\#\,\text{events passing cut } i}{\#\,\text{events before cut } i}\,. \tag{6.9}$$

Trigger

The event selection starts with the requirement that at least one of the relevant electron (muon) triggers (cf. Section 6.1) have fired. Events that do not contain an electron (muon) with sufficiently large transverse momentum are rejected at this stage with very high probability[8]. All simulated events used in this analysis contain at least one prompt lepton, and so have a good chance of being selected. The observed data on the other hand are dominated by backgrounds without a prompt lepton, hence only a fraction of about 2×10^{-7} of all recorded inelastic pp collisions passes the electron (muon) triggers.

Event cleaning

After the trigger selection, three cleaning cuts are applied in order to suppress non-collision events as well as events that cannot be reconstructed with the desired quality:

- Real data events with faults in the LAr/Tile calorimeter are dropped, as discussed in Section 6.1.
- At least five tracks must be associated with the hard collision primary vertex introduced in Section 4.1.
- Events with bad jets are vetoed, as explained in Section 4.2.

[8] The effective efficiency of this trigger cut is not to be confused with the trigger efficiency introduced in Section 3.3. The trigger efficiency is the conditional probability that in an event that actually contains a lepton, that lepton fires the trigger. The trigger efficiency for electrons (muons) from top-quark decays is about 98 % (76 %).

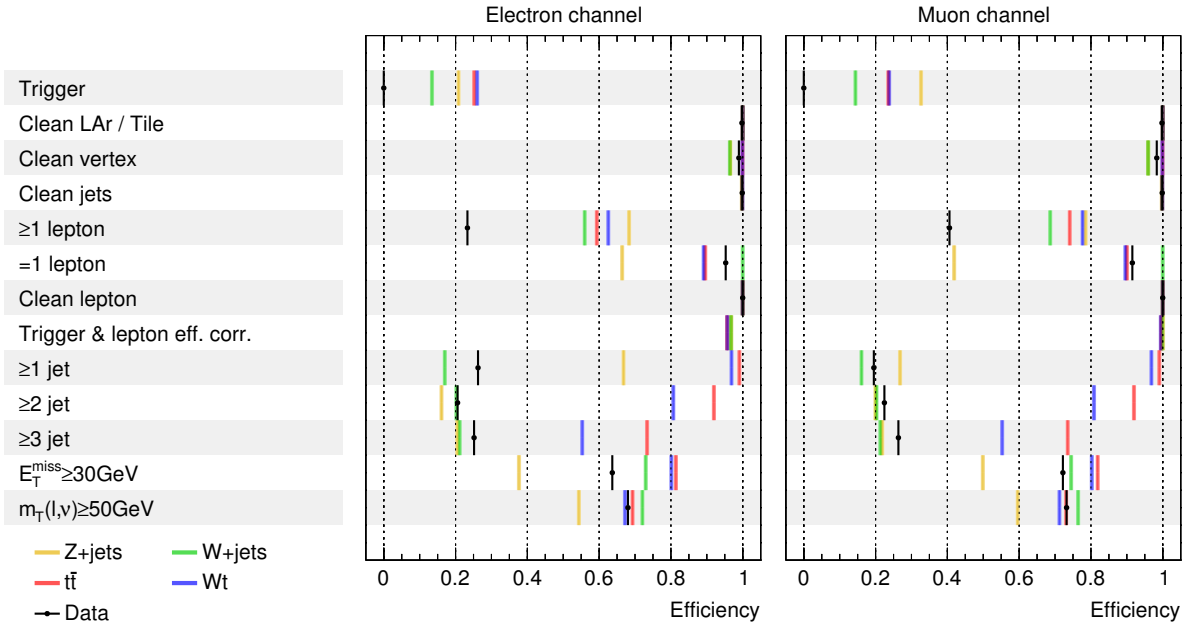

Figure 6.4: Sequential cutflows for the electron and muon selections. The efficiencies of each cut (cf. Equation 6.9) for simulated events with at least one prompt lepton are presented in different colours. Also shown are the efficiencies for observed data, which are dominated by the fake-lepton background before the lepton selection. The numbers are documented in Tables C.1 and C.2, which also include the simulated processes not shown in this figure.

Top-quark events survive these cuts with a 99.8 % chance.

Leptons

In the following, electrons and muons passing the quality criteria and the overlap removal procedure are considered. The quality criteria imply that leptons are reconstructed only in the central part of the detector, $|\eta| \lesssim 2.5$. Details can be found in Section 4.3.

At least one good electron (muon) candidate with a transverse momentum above 30 GeV is required in the electron (muon) channel. The candidate must have fired the trigger. About 70 % to 80 % of the simulated events with at least one prompt lepton pass this cut. The *Wt* signal, the top-quark pair production and the *Z*+jets background are each capable of producing two leptons with high transverse momentum, so they have a higher chance of passing this cut than the *W*+jets background. The efficiency for observed events is different from the simulation because real data events are still dominated by the fake-lepton background at this stage.

If there is another good lepton candidate with a transverse momentum above 25 GeV, the event is rejected. The threshold is slightly lower compared to the previous cut. This ensures that the analysis is orthogonal to the measurement of *Wt* production using dilepton events, meaning that there is no overlap between the sets of the events selected by each analysis. The cut also suppresses the *Z*+jets background because charged leptons are produced in pairs when the *Z* boson decays, whereas the *W*+jets does not contain a second lepton, making the cut fully efficient for *W*+jets events.

Afterwards, about 0.003 % of the events in the electron channel, where the remaining electron candidate shares an Inner Detector track with a reconstructed muon, are vetoed, referred to as *lepton cleaning*

in Figure 6.4[9]. The surviving lepton candidate is referred to simply as *the lepton* in the following.

The efficiencies of the triggering, the reconstruction and the identification of the selected lepton are corrected in the MC simulation using event weights (see Equation 6.3). As can be seen in Figure 6.4, this primarily concerns the electron channel, where the number of simulated events passing the lepton selection is corrected down by about 5 % when applying the corrections.

Jets

Jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$ are considered in this and all subsequent parts of the analysis. The jets must pass the quality criteria, including the JVF cut, and the overlap removal (see Section 4.2). The cut on η ensures that each jet is within the coverage of the Inner Detector, so that the track-based identifications of b-quark and pileup-induced jets perform efficiently.

Figure 6.5 visualises the expected composition of the selected events in terms of physics processes, separately in bins of the number of b-tagged jets, n_{tag}, and the total number of jets, n_{jet}. The best signal-to-background ratio is found for $n_{\mathrm{jet}} = 3, n_{\mathrm{tag}} = 1$, as is expected from the characteristic final state of the Wt production. When there are at least two b-tagged jets or at least three jets present, the top-quark pair production is the dominating physics process. The W+jets background on the other hand contributes the most for the events with $n_{\mathrm{tag}} = 0$. At this point of the cutflow, it is required that $n_{\mathrm{jet}} \geq 3$; further cuts on the number of (b-tagged) jets are deferred until the end of the selection.

The requirement that at least three jets be present in the event is applied incrementally, each time requiring one more jet. The Wt signal passes these cuts with a combined efficiency of about 45 %. Top-quark pair production events have an additional high-p_{T} b-quark jet that helps to survive the cuts more easily than any other physics process. The second and the third cut have about the same efficiency, which is expected from Berends-Giele scaling [132, 133]. For $Z \to e^+e^-$ events, there a considerable chance that one of the two electrons is misidentified as a jet, so the efficiency of the first cut is significantly increased in the electron channel.

Afterwards, the b-tagging scale factors associated with the selected jets are applied to the simulated events (see Equation 6.3). This modifies the ratio of b-tagged to b-untagged jets, keeping the total number of jets fixed. Since no cut on the number of b-tagged jets has been imposed to this point, the number of selected events remains constant.

Leptonic W decay

The selection concludes with two cuts that further enrich events containing a leptonically decaying W boson. As a neutrino is produced in the decay, one expects some missing transverse momentum, $p_{\mathrm{T}}^{\mathrm{miss}}$, for such events. The missing transverse momentum is required to be above 30 GeV. QCD production of multijets, the dominant physics process behind the selected fake-lepton events, is expected to have vanishing $p_{\mathrm{T}}^{\mathrm{miss}}$.[10] The fake-lepton and Z+jets backgrounds are suppressed by more than a factor of 2. The resolution of $p_{\mathrm{T}}^{\mathrm{miss}}$ strongly limits the suppression of the multijets background.

In samples with one leptonically decaying W boson, one expects the distribution of the invariant mass of the lepton and the neutrino to peak around the mass of the W boson. Experimentally only the transverse components of the missing momentum are measured, so invariant mass cannot be reconstructed,

[9] The muon definition used here is slightly looser than the one used in the analysis, otherwise there could be no reconstructed muon left after the vetoing events with two leptons.

[10] In practice, $p_{\mathrm{T}}^{\mathrm{miss}}$ is severely smeared out due to resolution effects (see Section 4.4), and a little missing momentum can arise from neutrinos that emerge from decays in b-/c-quark jets.

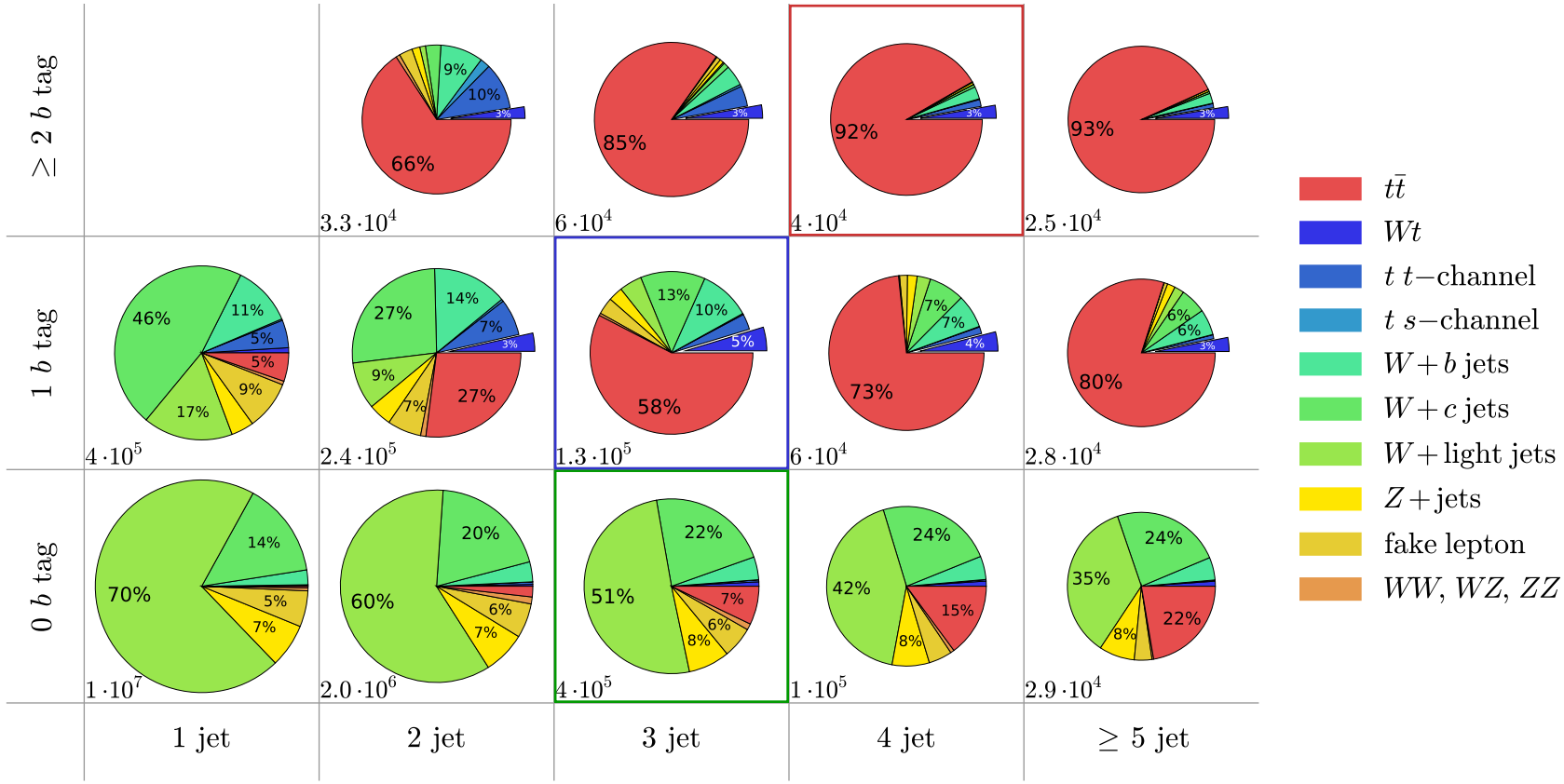

Figure 6.5: Pie diagrams presenting the expected composition of the selected events in terms of physics processes, separately in bins of the number of b-tagged jets and the total number of jets. At least one jet is required in order to have a reliable estimate of the fake-lepton background. All the other selection criteria discussed in the text have been applied. The expected number of events in each bin is shown as well. The regions relevant for this thesis are highlighted.

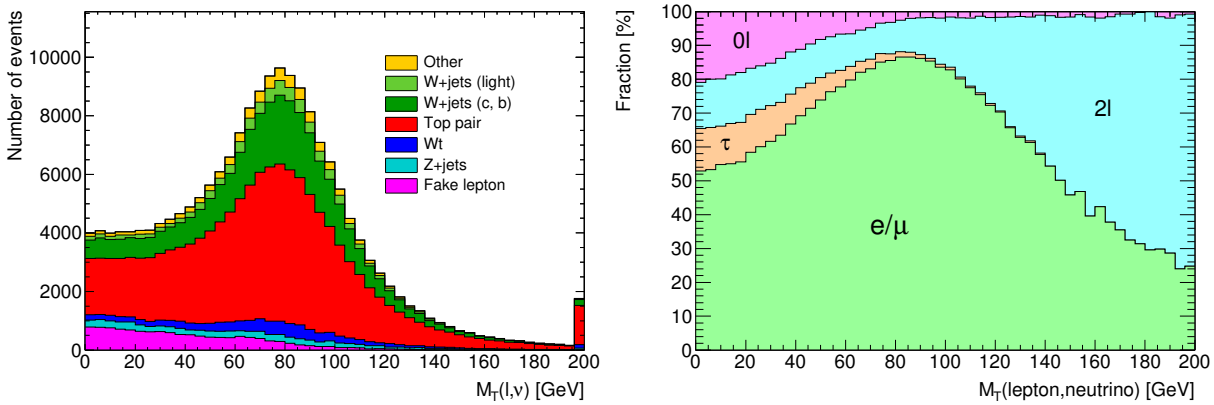

Figure 6.6: The left plot shows the expected distribution of $M_{\mathrm{T}}(\ell\nu)$ after the event selection. Colours are used to distinguish the contributions from the individual physics processes. The right plot distinguishes by the number of prompt lepton in the final state. It focuses solely on the fractions (by normalising each bin to unity). The cut on $M_{\mathrm{T}}(\ell\nu)$ has been omitted in the plots.

and the *transverse mass*, $M_{\mathrm{T}}(\ell\nu)$, is used instead:

$$M_{\mathrm{T}}(\ell\nu) = \sqrt{2p_{\mathrm{T}}(\ell)p_{\mathrm{T}}(\nu)\,(1 - \cos\,\Delta\phi(\ell,\nu))}\,, \tag{6.10}$$

where the masses of both particles have been neglected. It captures only a fraction of the W-boson mass (depending on $|\eta(\nu)|$), and is smeared out due to the experimental resolution, but still one can clearly identify the peak around 80 GeV in Figure 6.6. The peak is obviously not present for the fake-lepton background, which populates the region with smaller values of $M_{\mathrm{T}}(\ell\nu)$[11]. The observable can therefore be used to suppress the remaining fake-lepton background. The transverse mass of the leptonically decaying W boson, $M_{\mathrm{T}}(\ell\nu)$, is required to be above 50 GeV. As mentioned above, tau events are less precisely reconstructed, and $M_{\mathrm{T}}(\ell\nu)$ is clearly pulled away from the W peak towards to 0. So the cut also helps to reject these less precisely reconstructed events. The right plot might suggest that an upper cut on the $M_{\mathrm{T}}(\ell\nu)$ could help to suppress misreconstructed two-lepton events; however the systematic uncertainties on such a cut outweigh the benefit of the additional rejection against two-lepton events in the analysis, especially because events without a hadronically decaying W boson will be suppressed using a different technique later.

Merging of electron and muon channels

Due to lepton universality and the fact that the phase space covered by the electron and muon selections is very similar, there is no motivation for measuring the Wt cross section separately for electron and muon events. Therefore, the events yielded by the electron and muon selections, which are orthogonal to each other in consequence of the two-lepton veto, are merged together. The studies in Section 6.4 confirm that this simplification is indeed justified given the observed data.

[11] When a jet is reconstructed as lepton, the wrong energy calibration is applied, which results in spurious missing momentum along the lepton axis, often in the direction of the lepton. Consequently the azimuthal opening angle between the lepton and the missing momentum is underestimated, leading to a smaller $M_{\mathrm{T}}(\ell\nu)$ – or in the less likely case that the spurious missing momentum points in the opposite direction, a peak near the sum of the values used in the p_{T} cuts on the lepton and the missing momentum.

Cut optimisation

The performance of the analysis is largely independent of the choices made for the p_{T} thresholds in the event selection. The performance is quantified by the expected total uncertainty on the cross-section measurement (see Section 8.3). The expected uncertainty on the measurement changes from 38 % by less than 1 percentage point when varying the p_{T} cuts imposed on the lepton and/or the jets within ±5 GeV about the nominal cut value of 30 GeV. When using thresholds well above 35 GeV, the acceptance for signal events drops considerably, and with it the signal-to-$t\bar{t}$ ratio. Relaxed cuts, while improving the signal-to-$t\bar{t}$ ratio, would lead to larger the systematic uncertainties due to the low-p_{T} jets. They would also increase the W+jets and fake-lepton backgrounds, which would require a more precise treatment of the uncertainties on their modelling. The performance of the analysis also does not depend on whether events with more than two light-flavour jets are allowed to pass the selection.

Signal and validation regions

Following the discussion of Figure 6.5, the *signal region* is defined by the requirement of exactly one b-tagged jet and exactly two b-untagged jets. The selection for the signal region has an efficiency of 2.7 % for Wt and $t\bar{t}$ events. Background events with at least one prompt lepton but no top quark survive the selection with a chance of about 1 : 20 000. The regions that are close to the signal region, but have a very small expected signal-to-background ratio, are useful *validation regions*, where the modelling of the major backgrounds can be checked in events that have similar kinematic distributions as the signal region. The region with $n_{\mathrm{jet}} = 4$ and $n_{\mathrm{tag}} = 2$ contains a very pure sample of top-quark pair production events, while the region with $n_{\mathrm{jet}} = 3$ and $n_{\mathrm{tag}} = 0$ provides a good measure to check the modelling of the W+jets background.

Process	Signal region	$t\bar{t}$ region	W+jets region
Wt, $\sigma = 22.4$ pb	6 300 ± 600	1 110 ± 330	3 430 ± 330
$t\bar{t}$	77 000 ± 6 000	37 000 ± 6 000	30 000 ± 3 500
t, t-channel	4 190 ± 310	610 ± 60	1 530 ± 150
t, s-channel	308 ± 21	66 ± 7	94 ± 10
W+jets, heavy flavour	31 000 ± 15 000	1 300 ± 700	110 000 ± 50 000
W+jets, light	6 000 ± 3 000	31 ± 24	210 000 ± 100 000
Z+jets	3 900 ± 1 700	190 ± 90	32 000 ± 14 000
$WW/WZ/ZZ$	650 ± 280	25 ± 13	4 800 ± 2 100
Fake leptons	4 300 ± 1 900	–	23 000 ± 7 000
Total background	128 000 ± 18 000	39 000 ± 6 000	420 000 ± 140 000
Total model	134 000 ± 18 000	40 000 ± 6 000	420 000 ± 140 000
Observed	134 216	41 480	422 185

Table 6.1: Expected and observed numbers of events in the signal and validation regions, before fitting. The quoted errors are the standard errors due to all systematic uncertainties (cf. Section 8.2). The cross section for Wt production has been fixed to the theory prediction. The rounding of the numbers follows the recommendation of the Particle Data Group [7].

6.4 Control plots

Table 6.1 shows the expected and the observed number of events in the signal region and the validation regions after the event selection. The small backgrounds, in particular Z+jets, diboson, fake leptons and single top-quark production via the s-/t-channel exchange of a W boson, will often be subsumed under *other backgrounds* in the following. Since the cross section for the Wt production is yet to be measured, the number of Wt events is normalised using the theoretical prediction. The predicted total number of events, especially in the signal region, depends on this choice. The uncertainties on the prediction have been calculated by random sampling of all systematic uncertainties. Their construction and interpretation are described in detail in Section 8.4. The systematic uncertainties on the numbers of the individual physics processes do not add up in quadrature to the uncertainty on the total number of expected events because uncertainties due to the reconstruction as well as the luminosity measurement are correlated among the processes. The totals for expected and observed events coincide within just a few percent, which is notably accurate given the uncertainties. For the signal and the W+jets validation region, this is mainly because the uncertainties on the W+jets normalisation are rather conservative, given that a lot of progress has been made in the field of W+jets event generators since the uncertainties were determined.

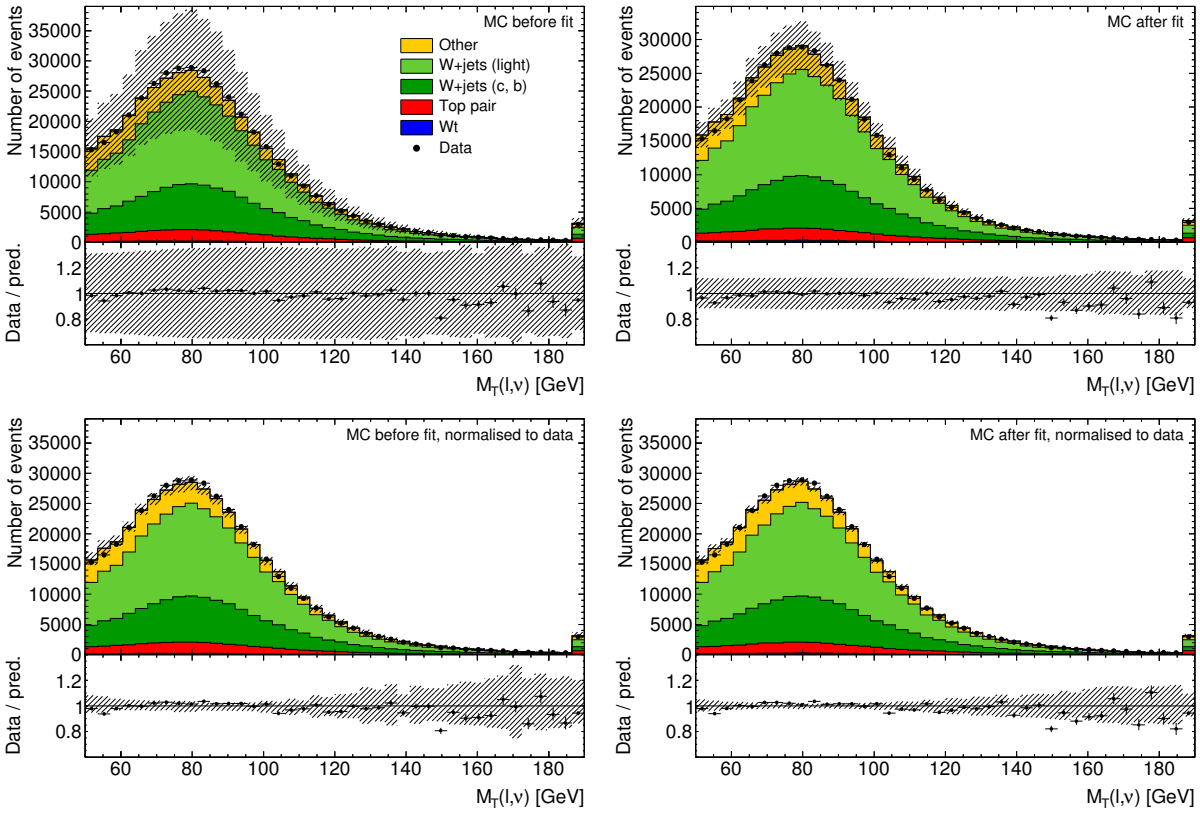

Figure 6.7: Distribution of the transverse mass in the leptonic W decay in the W+jets validation region. The left (right) plots compare the observed data with the model before (after) the fit. The Wt signal in the left plots is normalised to the theoretical cross section. The size of the uncertainty band is still ~20 % after the fit because of the uncertainty on the extrapolation of the W+jets background from the signal to the W+jets validation region. For the bottom plots the expected number of events has been normalised to the observed number.

It is crucial to check that the model is able to reproduce the observed data within the uncertainties. These distributions will be visualised as histograms. Each observed event in the considered signal or validation region contributes with a weight of 1, and consequently the statistical error on the observed number of events is approximated by its square root. The first (last) bin also includes the underflow (overflow) that occurs when the quantity on the x axis is outside of the range spanned by the axis. The expected distributions of the Wt signal and the major backgrounds are presented using a stacked bar chart, with a hashed band to represent the uncertainty on the total expectation in each bin. Beneath each such *stack plot*, the ratio of the observed over the expected number of events in each bin is shown as a visual aid for the comparison of the two. In order to ease the comparison of the shapes of the expected and the observed distributions, the distributions are normalised to the number of observed events, separately in each region. If the major physics processes yield similarly distributed events, the typically large normalisation uncertainties on the individual processes cancel out to a large degree, allowing a more critical review of the modelling.

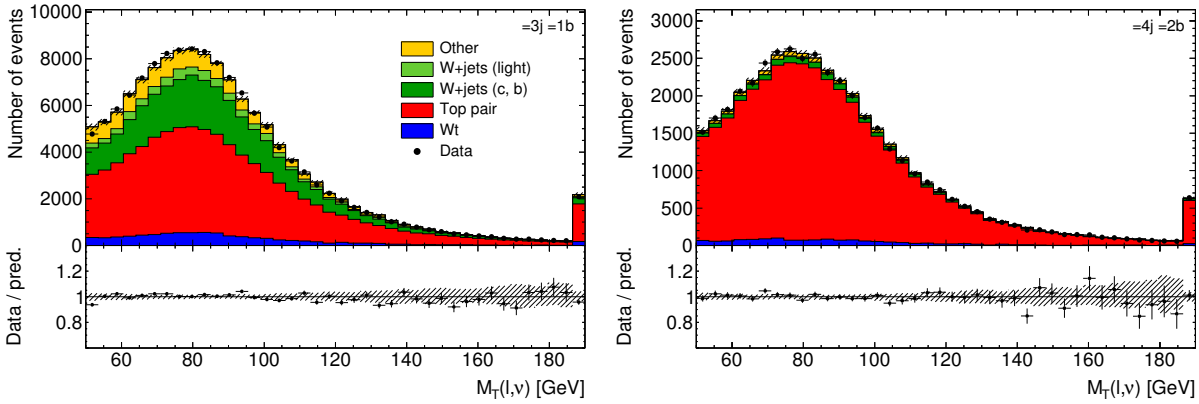

Figure 6.8: Distributions of the transverse mass in the leptonic W decay in the signal region (left) and the $t\bar{t}$ validation region (right). The shapes of the expected distributions have been explained in Section 6.3.

Since the Wt cross section is to be measured, it must be treated as an essentially unknown parameter until the statistical analysis is performed. The definite value and the uncertainty that are needed in order to create the stack plots is therefore anticipated from the result given in Equation 8.13. The fit also yields improvements for other model parameters, such as the normalisation of the backgrounds, and some of the uncertainties, most prominently the conservative uncertainty on the estimate of the W+jets background. The fitted Wt normalisation as well as the corrections are used in the preparation of stack plots hereafter unless noted otherwise. A detailed discussion of the involved statistical methods will follow in Chapter 8. For illustration, Figure 6.7 shows the distribution of $M_{\mathrm{T}}(\ell\nu)$ in the W+jets validation region before and after the aforementioned improvements.

This powerful visualisation tool is used to verify the modelling first in the validation regions, where the signal contribution is small. The $t\bar{t}$ validation region is especially useful due to its smaller uncertainties. The fact that the Wt signal and the $t\bar{t}$ background are rather similar also enables an indirect test of the acceptance modelling of the signal in the $t\bar{t}$-enriched region. The following discussion will therefore focus more on the $t\bar{t}$ validation region than the W+jets validation regions, though the latter has been checked with the same diligence. In Figure 6.8 it can be seen that the transverse mass of the leptonically decaying W boson is well-modelled in the signal and the $t\bar{t}$ validation regions.

When assessing whether the prediction describes the detector acceptance and the efficiencies correctly, the distributions of the kinematic properties of the reconstructed objects, in particular the lepton,

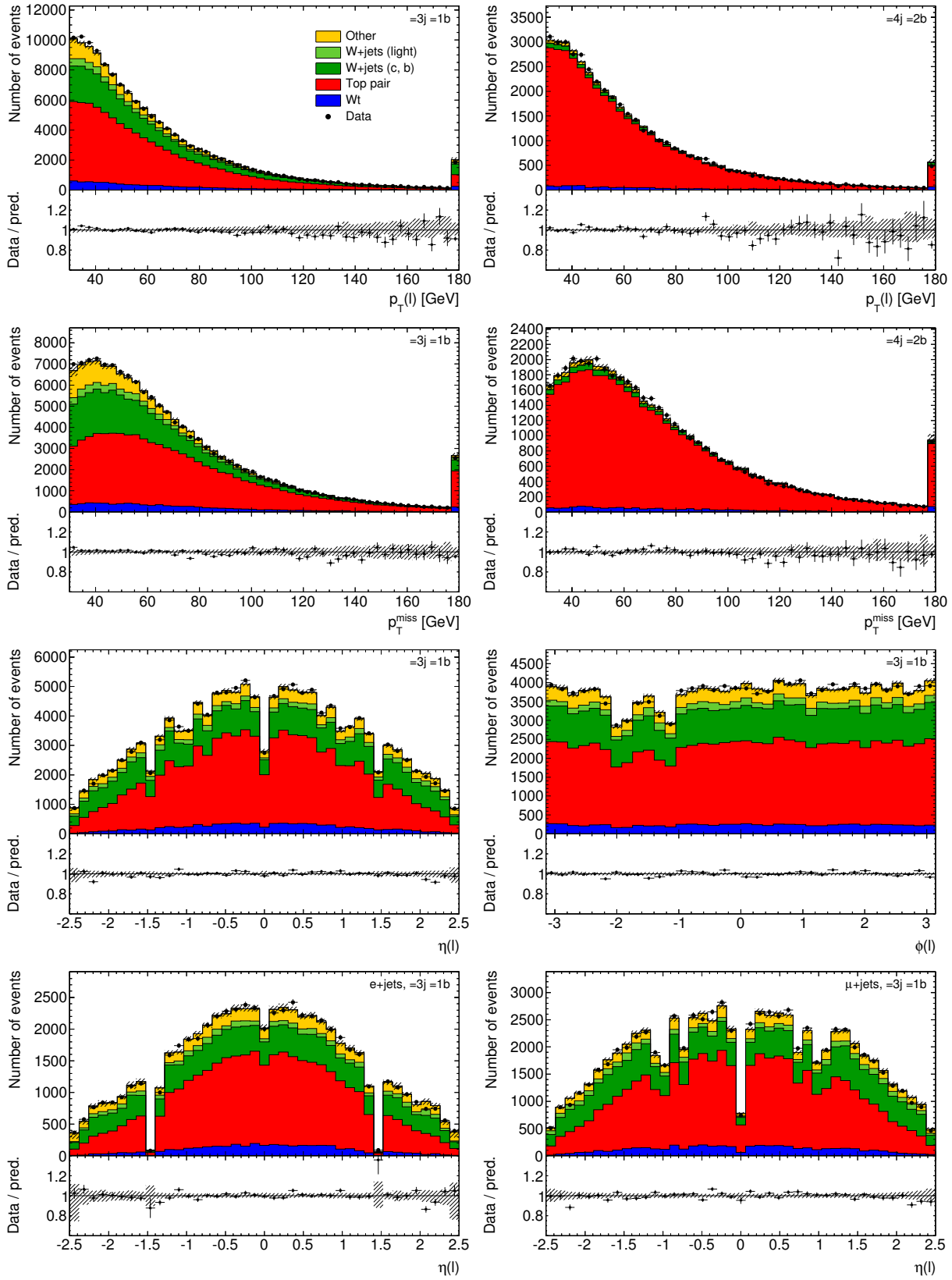

Figure 6.9: Distributions of the transverse momentum of the lepton, and the missing energy in the signal region (left) and the $t\bar{t}$ validation region (right), and the angular distributions of the lepton in the signal region. The bottom left (right) plot shows the pseudorapidity of the electron (muon) only.

the missing energy, and the jets, are particularly useful. While modelling issues in the physics simulation can also show up here, they are often seen more distinctively in the kinematics of "higher level" objects such as the W bosons or the top quark. Figure 6.9 illustrates some of the distributions of the lepton and the missing energy. The observed distributions of the missing energy and the transverse momentum of the lepton are described reasonably well by the model. When a heavy particle like the W boson decays into two particles with negligible mass, each receives an additional momentum of half of the heavy particle's mass on top of the heavy particle's original momentum. In consequence small values of p_{T} are disfavoured. Acceptance and efficiency also play an important role especially at small values of the momentum. The reconstruction and identification efficiencies are highly p_{T}-dependent (see Figure 4.5). The distribution of $p_{\mathrm{T}}^{\mathrm{miss}}$ on the other hand is strongly affected by the relatively poor $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution of about 30 GeV. Also as the lepton and the neutrino emerge from the same W boson, the two are correlated such that the cut of the transverse momentum of the one affects the p_{T} acceptance of the other. The tails of the distributions strongly depend on the energy available to the parton collision, so the event rate falls with a power law.

Since the physics behind the pp collisions are rotationally invariant about the beam axis, the azimuthal distributions of the objects would be flat if it were not for acceptance effects. A similar argument holds for the sign of the pseudorapidity. One expects top quarks to be produced relatively central because the relevant structure functions of the proton fall off rapidly towards larger longitudinal momentum transfers. The peculiar shape of the pseudorapidity distribution of the lepton is the result of the geometry of the detector components that are involved in the electron and muon reconstructions. To understand this a bit better, the distribution is also shown separately for electron and muon events. The efficiency for electrons drops to 0 near $|\eta| \approx 1.5$ due to the quality cuts that exclude the transition region between the barrel and the end-caps of the calorimeter. The dip at $\eta \approx 0$ is the result of the gap in between the $\eta < 0$ and $\eta > 0$ parts of the Transition Radiation Tracker. This feature is also seen for muons because the Muon Spectrometer is only partially equipped with muon chambers at $\eta \approx 0$. In addition, the trigger efficiency drops strongly at 0.4, 0.7 and 1.0, where the magnet ribs and the feet of the ATLAS detector are located. The feet are also responsible for the pronounced dip at $\phi = -\pi/2$ in the distribution of the azimuthal angle of the lepton. Both angular distributions are modelled well.

While separating the electron and the muon channel is instructive for the understanding of the effects of the detector geometry, differences between the channels can be checked much better using ratios. In the ratio of the number of electron over the number of muon events, many systematic uncertainties, except those on (fake) leptons, cancel out to a large degree, allowing a critical comparison between the two. An expression somewhat more convenient than the ratio is the fraction of selected events that is contributed by the muon channel:

$$\frac{N(\mu)}{N(e) + N(\mu)} = 1 - \frac{N(e)}{N(e) + N(\mu)}\,. \tag{6.11}$$

It is found to be 0.563 ± 0.001 in observed data, consistent with the expectation of 0.559 ± 0.011. It is above 0.5 because the muons are reconstructed with a higher efficiency than electrons. The ratio as a function of the transverse momentum of the lepton, displayed in Figure 6.10, clearly reflects the p_{T}-dependence the electron efficiency below ~100 GeV (see Section 4.3.1). The figure also shows the ratios as functions of $p_{\mathrm{T}}^{\mathrm{miss}}$ and $M_{\mathrm{T}}(\ell\nu)$, which are very useful to validate the estimate of the fake-electron and the fake-muon backgrounds against each other. A significant disagreement at low values, populated by the fake-lepton background, would indicate a problem, but none is seen. The transverse momentum of the b-tagged jet exemplifies the behaviour of most variables: neither the electron–muon ratio and nor its modelling are notably correlated with it.

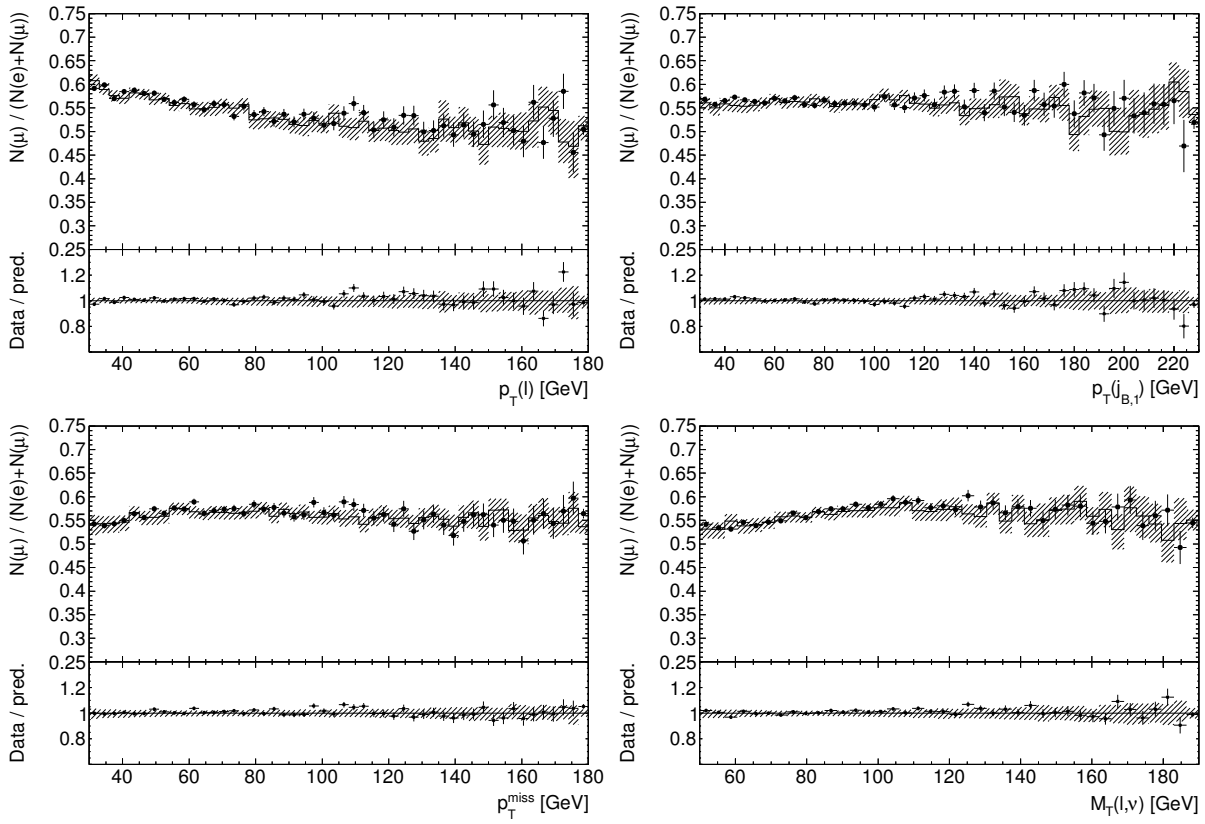

Figure 6.10: Fraction of selected events that is contributed by the muon channel in the signal region, separately as function of the transverse momentum of the lepton, the b-tagged jet, the missing energy and the transverse mass in the leptonic W decay.

Figure 6.11 shows several kinematic properties of the b-tagged jet and the two b-untagged jets, labelled j_{B}, $j_{\mathrm{L},1}$, and $j_{\mathrm{L},2}$ in the plots. In the $t\bar{t}$ validation region the b-tagged jet with the larger transverse momentum has been picked. A qualitative discussion of the general shapes expected for p_{T} spectra has already been presented in the context of the lepton and the missing energy. For example, the turn-on in the p_{T} spectrum of the leading b-untagged jet is a consequence of the p_{T} cut on the sub-leading b-untagged jet. For the leading b-tagged jet in the signal region, only the distributions for top-quark processes show such a turn-on because the kinematics of the b-tagged jet are strongly correlated with the other decay products of the top quark, in contrast to processes such as W+jets. Since the detector acceptance for jets is relatively flat in the polar and azimuthal angles, the angular distributions do not exhibit any strong features. The model seems to slightly underestimate the p_{T} spectra of the b-tagged jet and the leading b-untagged jet around 60 GeV, and slightly overestimate the p_{T} spectra above ~ 90 GeV, but this is acceptable given the uncertainties. The other distributions are described well by the model.

Having gained confidence that the detector, its simulation, and the data-driven estimates work well for the basic reconstructed objects, it is time to investigate some more complex variables that have a stronger physics interpretation, starting with the mass of the top quark. By adding the four-momentum vectors of the charged lepton and the neutrino, the candidate for the leptonically decaying W boson, W_{L}, is reconstructed. Similarly the hadronically decaying W boson candidate, W_{H}, is constructed from the two b-untagged jets. Since it is not known with which of the two W bosons the b-tagged jet emerged from a top quark, two top-quark candidates are reconstructed by combining the leading b-tagged jet and

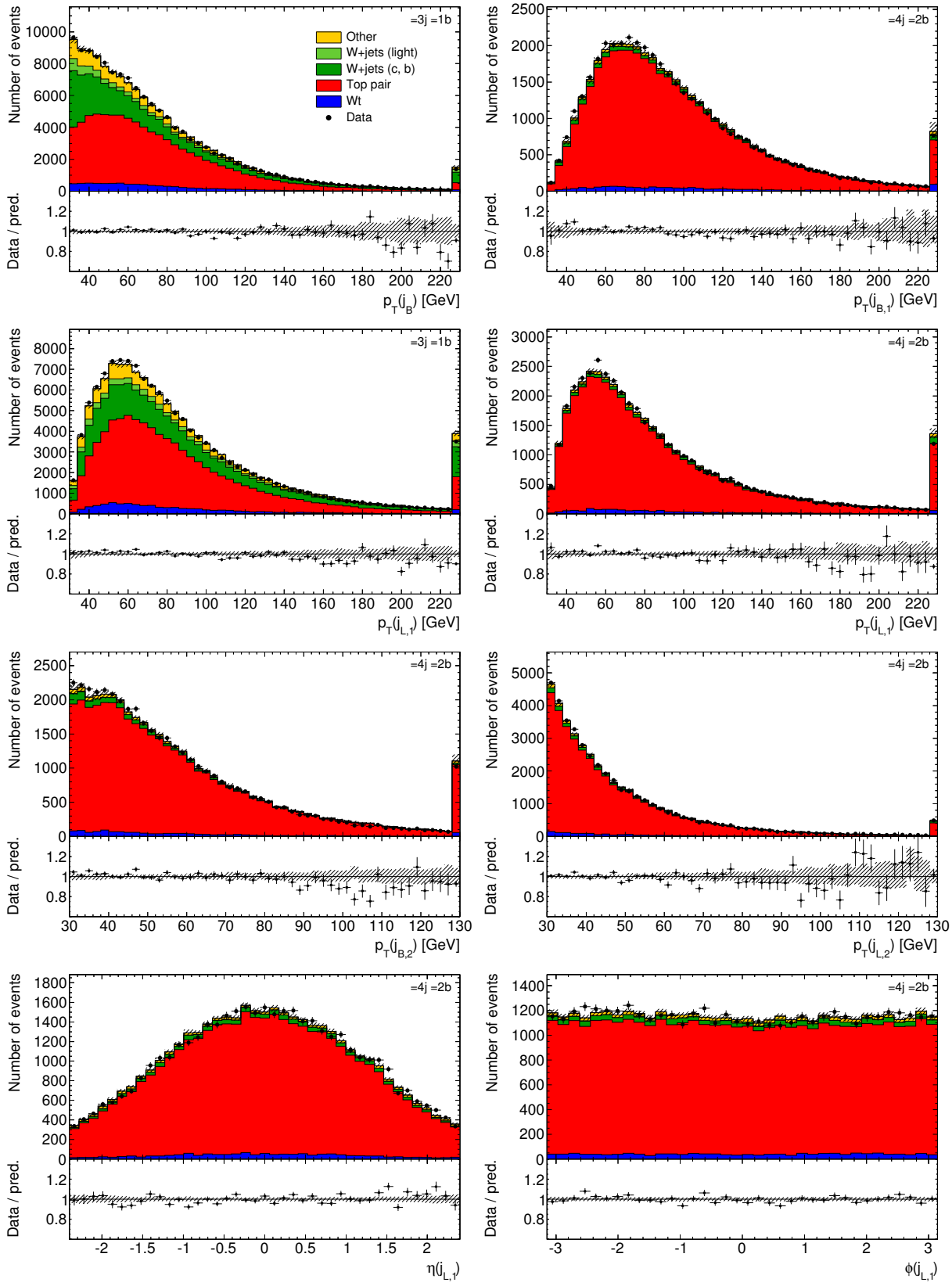

Figure 6.11: Distributions of kinematic properties of the selected jets in the signal region and the $t\bar{t}$ validation region.

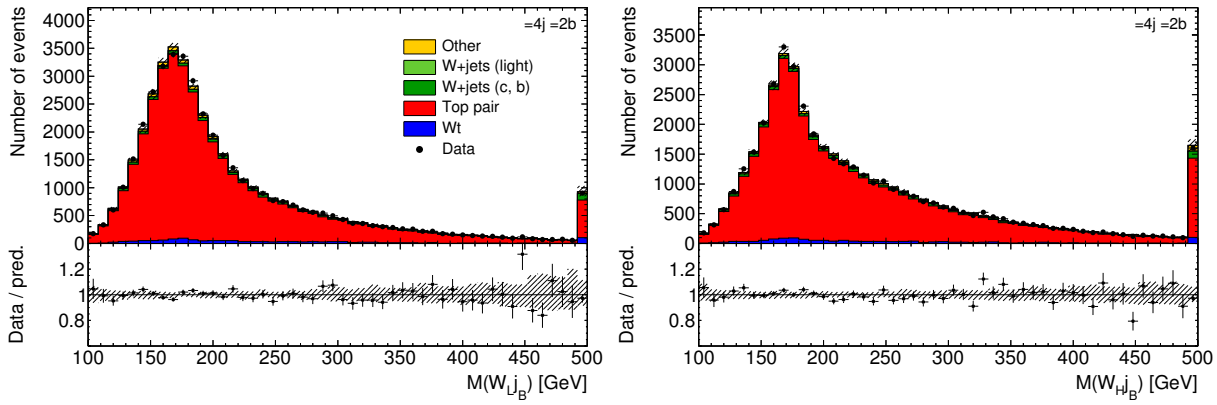

Figure 6.12: Distributions of the top-quark mass reconstructed from the leading b-tagged jet and the leptonically (hadronically) decaying W boson.

either W_{L} or W_{H}, again using four-momentum addition. To summarise, the four-momenta are combined as follows:

$$\begin{aligned} p(W_{\mathrm{L}}) &= p(\ell) + p(\nu) \\ p(W_{\mathrm{H}}) &= p(j_{\mathrm{L},1}) + p(j_{\mathrm{L},2}) \\ p(t_{\mathrm{L}}) &= p(j_{\mathrm{B}}) + p(W_{\mathrm{L}}) \\ p(t_{\mathrm{H}}) &= p(j_{\mathrm{B}}) + p(W_{\mathrm{H}})\,, \end{aligned} \tag{6.12}$$

(6.13)

where b-untagged jets are assumed to be massless, and a mass of 5 GeV is assigned to the b-tagged jets.

Figure 6.12 shows the spectrum of the invariant mass of the top-quark candidates. The peaks around 172.5 GeV[12] are clearly visible, sitting on top of combinatorial background that results from combinations of the b-tagged jet with the wrong top quark, or misreconstruction of the W boson candidate. Again the good modelling of the observed data is evident.

6.5 In a nutshell

The major criteria used for the event selection have been presented:

- exactly one central lepton with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$;
- missing energy, $p_{\mathrm{T}}^{\mathrm{miss}} > 30\,\mathrm{GeV}$;
- at least three jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$;
- consistent with leptonically decaying W boson, $M_{\mathrm{T}}(\ell\nu) > 50\,\mathrm{GeV}$.

Signal and validation regions have been defined depending on the number of (b-tagged) jets:

[12] It is not expected that the simulated distribution peaks at 172.5 GeV, the value used for the mass of the top quark in the configuration of the MC simulation, for various reasons such as shifts due to the jet energy scale, radiation that is not captured by the jets, energy contributions from the underlying event that fall into the jet cones, and apparent shifts resulting from a combinatorial background that is not flat in the peak region.

- signal region: $n_{\text{jet}} = 3$, $n_{\text{tag}} = 1$;
- W+jets validation region: $n_{\text{jet}} = 3$, $n_{\text{tag}} = 0$;
- $t\bar{t}$ validation region: $n_{\text{jet}} = 4$, $n_{\text{tag}} = 2$.

Wt events constitute about 5 % of the signal region. The major backgrounds are the top-quark pair production with 55 %, and the W+jets background with about 30 %. An MC simulation has been used to model the signal and backgrounds for events with at least one prompt lepton, while the contribution from the fake-lepton background has been estimated using a data-driven method. This combination has been confronted with the observed data using a variety of control plots, and has been found to describe the data very well.

CHAPTER 7

Multivariate Analysis

While the bulk of the background is efficiently suppressed by the event selection, the large uncertainty on the remaining background events exclude the possibility to measure the *Wt* cross section by simple event counting. This chapter focuses on improving the separation of the signal from the remaining background in order to make the measurement viable. Section 7.1 analyses what the signal and the backgrounds "look like", aiming at the kinematic properties that can be used to distinguish the signal from the background. These properties can be combined in an optimised way using NeuroBayes, a multivariate technique based on artificial neutral networks, which is portrayed in Section 7.2. After identifying a specific shortcoming of this multivariate technique for the problem in hand (Section 7.3), a modified approach is presented that can cope with the large systematic uncertainties (Section 7.4). At the end of the chapter, a multivariate function is obtained that projects the measured properties of an event into a binned discriminant that is suitable for the statistical analysis.

7.1 Discriminating variables

In an event-counting analysis, the number of signal events is obtained by subtracting the expected number of background events from the observed number of events. In view of the uncertainty on the backgrounds (cf. Table 6.1), it is clear that a more sophisticated treatment is required for the measurement of the *Wt* cross section. It relies on the reduction of the uncertainties on the background as well as fingerprinting based on the expected differences between the probability density functions of the signal and the backgrounds. Such differences are already utilised for the event selection and the definition of the signal region. The number of (*b*-tagged) jets is a good example of a quantity that sets the *Wt* signal apart, to a certain extent, from the top-quark pair production. This section explores other variables that distinguish the signal from the backgrounds, with a strong focus on the two major backgrounds: top-quark pair production and the production of a *W* boson in association with heavy-flavour jets.

Figure 7.1 presents a different view on a few of the distributions known from Section 6.4. Instead of stacking the contributions of the processes upon each other, the distributions are normalised to unity, and compared directly to each other. The uncertainty bands are constructed as described in Section 8.4. The fast detector simulation is used for this mode of presentation in order to avoid large statistical fluctuations of the *Wt* signal due to the limited sample size for the full detector simulation. The remaining fluctuations in the error bands are due to the limited size of the samples used to evaluate the uncertainties on the choice of generator as well as renormalisation and factorisation scales for the *Wt* process.

The objects produced in *W*+jets events pass the detector less centrally, as illustrated taking the example of the pseudorapidity of the lepton. The *W*+jets events also populate the phase space regions with a soft *b*-tagged jet more strongly than the events where the *b*-quark jet can receive a large momentum transfer due to the decay of a massive top quark. While significant differences show up in all of these distributions for the *W*+jets background, the $t\bar{t}$ background looks very similar to the signal. A notable

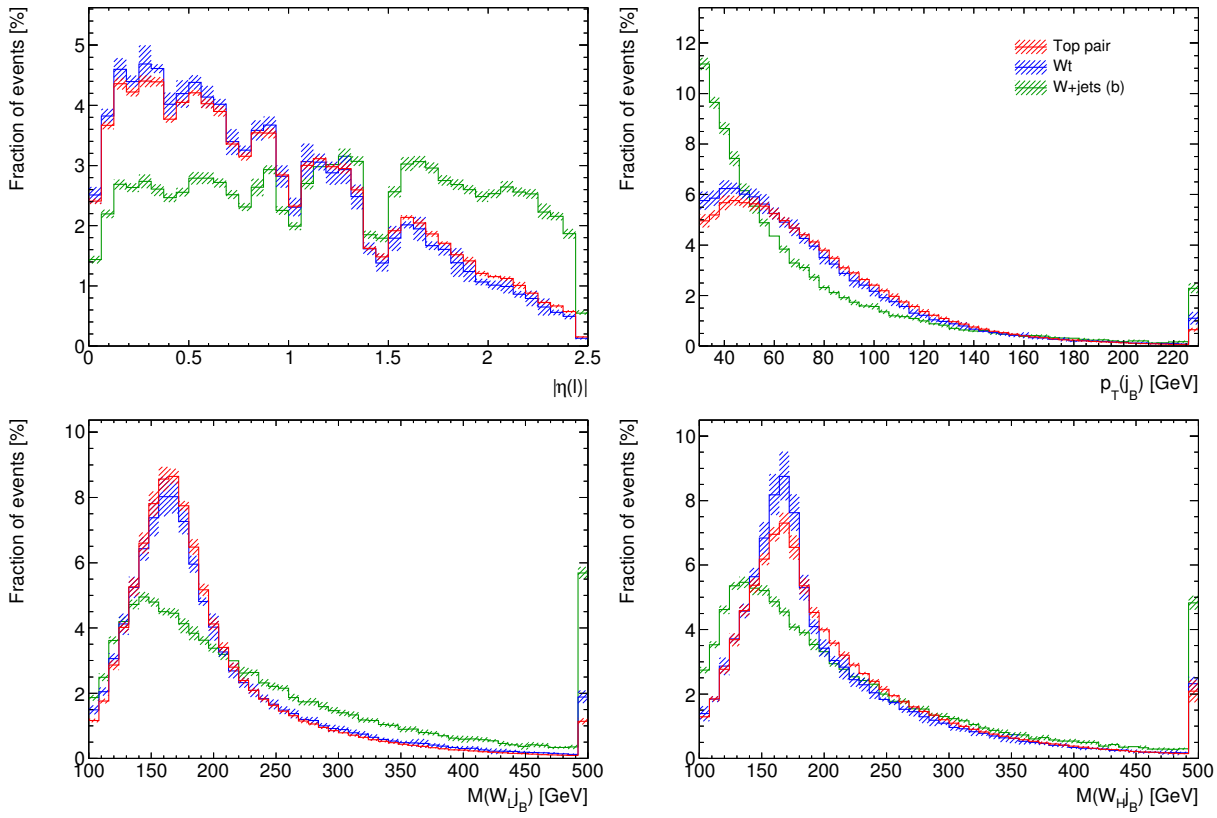

Figure 7.1: Spectra of the pseudorapidity of the lepton, the transverse momentum of the b-tagged jet, and the reconstructed masses of the top-quark candidates (calculated using Equation 6.13) in the signal region, compared to the Wt signal and the two major backgrounds. The hatched bands indicate the effects of the systematic uncertainties on the shapes. Since all distributions are normalised to unity, the uncertainties on the overall normalisation of every process cancel out.

difference is found only in the spectrum of the invariant mass of the three jets. The peak resulting from the hadronic decays of the top quarks is more pronounced for the Wt signal (for a reason that will become clear in the discussion of the reconstructed mass spectrum of the hadronically decaying W boson later on). The invariant mass of the three jets is sensitive to a variety of effects, predominantly the details of the parton shower and hadronisation model, so it cannot be fully relied upon for the separation of the Wt signal and the $t\bar{t}$ background.

Figure 7.2 shows an example of a variable that makes direct use of the very well-measured lepton, largely avoiding[1] the uncertainties on the reconstruction of the jet energies and the missing transverse momentum: the angular separation, $\Delta R(\ell, j_\mathrm{B})$, between the lepton and the b-tagged jet, expressed in the η–ϕ plane. In Wt and $t\bar{t}$ events, the lepton and the b-tagged jet tend to be back-to-back when the b-tagged jet is produced in the semileptonic decay of a top quark, resulting in the peak around π. Otherwise, the distribution of the azimuthal difference, $\Delta\phi(\ell, j_\mathrm{B})$, peaks near 1.0, which is reflected in the distribution of ΔR. The variable exhibits good separation power between the signal and W+jets background. When a $b\bar{b}$ ($c\bar{c}$) pair recoils against the leptonically decaying W boson in a W+jets event, a back-to-back configuration is preferred again. This time however $|\Delta\eta(\ell, j_\mathrm{B})|$ tends to be large when compared to Wt and $t\bar{t}$ events, resulting in distribution that falls more slowly for $\Delta R(\ell, j_\mathrm{B}) > \pi$.

[1] Modulo any biases due to the event selection, which incorporates cuts on $p_\mathrm{T}^\mathrm{miss}$ and jet energies.

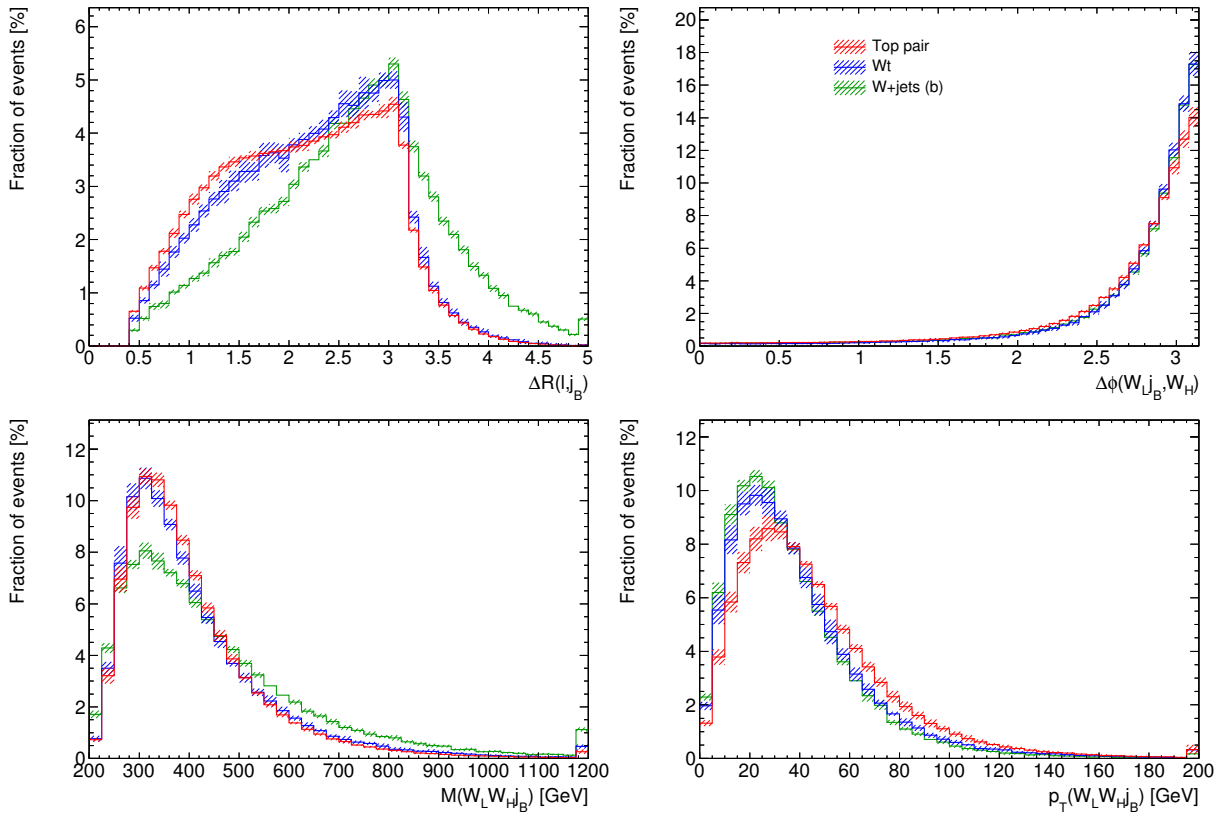

Figure 7.2: Spectra of the angular difference between the lepton and the b-tagged jet in the η–ϕ plane, the azimuthal angle between the reconstructed semileptonically decaying top-quark candidate and the hadronically decaying W boson candidate, and the invariant mass as well as the transverse momentum of all reconstructed objects, compared to the Wt signal and the two major backgrounds. The hatched bands indicate the effects of the systematic uncertainties on the shapes. Since all distributions are normalised to unity, the uncertainties on the overall normalisation of every process cancel out.

In order to find observables that distinguish between Wt and $t\bar{t}$ events, it is useful to reconstruct the four-momentum of the entire Wt system candidate:

$$p(W_\mathrm{L} W_\mathrm{H} j_\mathrm{B}) = p(W_\mathrm{L}) + p(W_\mathrm{H}) + p(j_\mathrm{B})\,, \tag{7.1}$$

where the four-momenta of the W-boson and top-quark candidates are again reconstructed according to Equation 6.13. According to the findings presented in Section 5.2, the invariant mass of the Wt system, $M(W_\mathrm{L} W_\mathrm{H} j_\mathrm{B})$, and its transverse momentum, $p_\mathrm{T}(W_\mathrm{L} W_\mathrm{H} j_\mathrm{B})$, are candidates with promising separation between the Wt signal and the $t\bar{t}$ background. The Wt signal generally populates the lower-energy regime compared to the $t\bar{t}$ production, which is reflected in the mass spectrum shown in Figure 7.2, albeit not very pronounced. The W+jets background generally tends to have lower energies but after the event selection, a significant fraction of events with large $M(W_\mathrm{L} W_\mathrm{H} j_\mathrm{B})$ remains.

Considering the kinematics of the diagrams for Wt and $t\bar{t}$ at leading order in α_S, it is natural to identify Wt events with the back-to-back production of the top quark and the associated W boson, or similarly, $p_\mathrm{T}(WWb) = 0$. Because the reconstruction of the top-quark candidate is ambiguous, two variables are constructed: $\Delta\phi(t_\mathrm{L}, W_\mathrm{H})$ and $\Delta\phi(t_\mathrm{H}, W_\mathrm{L})$. The former is shown in Figure 7.2, which reveals that the reconstructed candidates for the top quark and the associated W boson are indeed back-to-back more

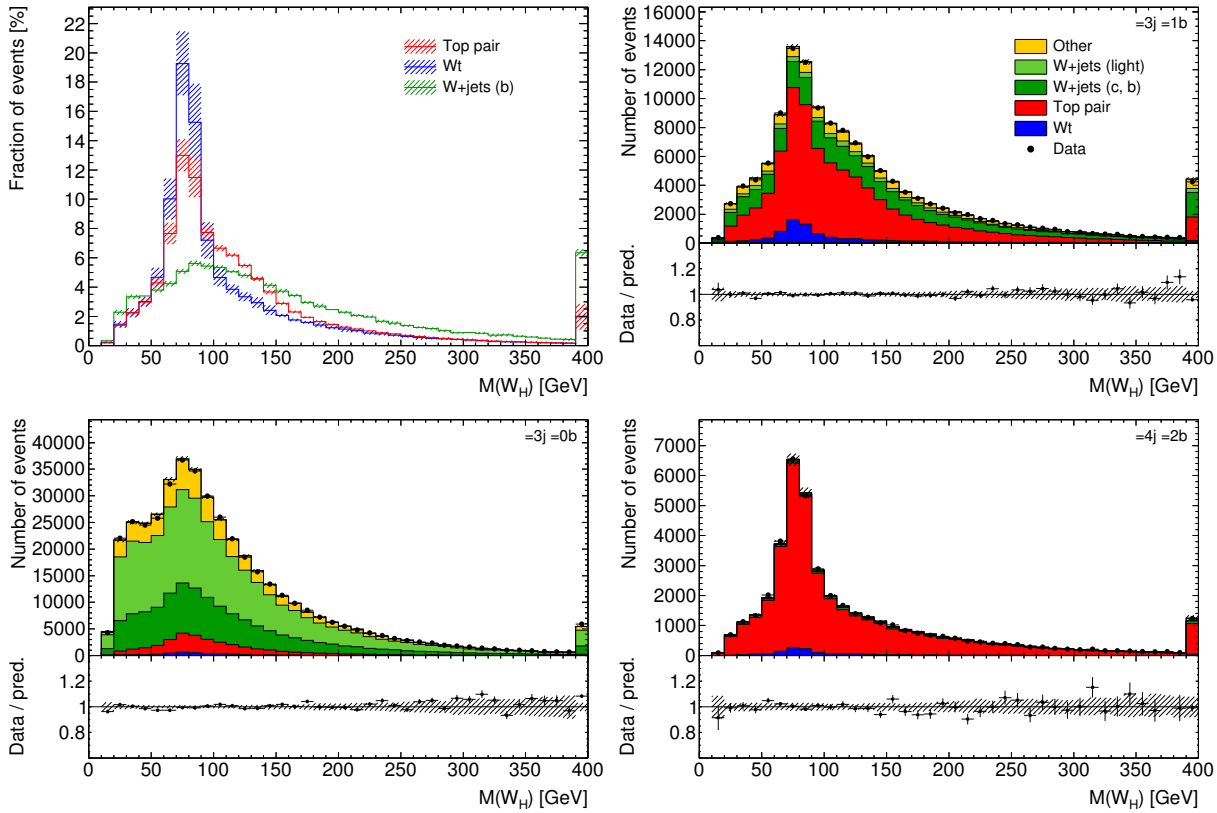

Figure 7.3: Spectrum of the reconstructed mass of the hadronic W boson candidate compared to the Wt signal and the two major backgrounds (top left). The hatched bands indicate the effects of the systematic uncertainties on the shapes. Since all distributions are normalised to unity, the uncertainties on the overall normalisation of every process cancel out. The distributions of the reconstructed mass are compared to the observed data in the signal (top right) and the validation regions (bottom row).

often for the Wt signal when compared to the $t\bar{t}$ background. A more attractive variable is $p_{\mathrm{T}}(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$, which is unaffected by the aforementioned ambiguity in the top-quark reconstruction. The distribution of the reconstructed $p_{\mathrm{T}}(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$ is considerably softer for the Wt signal than for the $t\bar{t}$ background. The region below 30 GeV is somewhat disturbed by uncertainties due the reconstruction of the jet energies as well as the theory modelling. Also, the W+jets background is large at small values of $p_{\mathrm{T}}(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$ even more than the signal, making the variable a little less appealing. It still constitutes a promising discriminant between the signal and the backgrounds, and will be used in a modified way later in the analysis. The rapid fall-off setting in at 30 GeV is due to the event selection: a large $p_{\mathrm{T}}(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$ is strongly correlated with the presence of another jet with large transverse momentum. Such events are rejected unless the jet is sufficiently non-central ($|\eta| > 2.4$).

The reconstructed mass of W_{L} is fixed to m_W by construction (see Section 4.4).[2] The reconstructed mass spectrum of W_{H} is shown in Figure 7.3. The observed distribution of the variable is modelled well in the signal and the control regions.[3] On the face of it, $M(W_{\mathrm{H}})$ displays exceptional separation

[2] Remember that the presence of a leptonically decaying W boson is detected using $M_{\mathrm{T}}(\ell\nu)$, which is built from the raw $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ rather than $\vec{p}(\nu)$.

[3] When constructing equivalent variables in the W+jets control region, the jet with largest transverse momentum is substituted for the b-tagged jet.

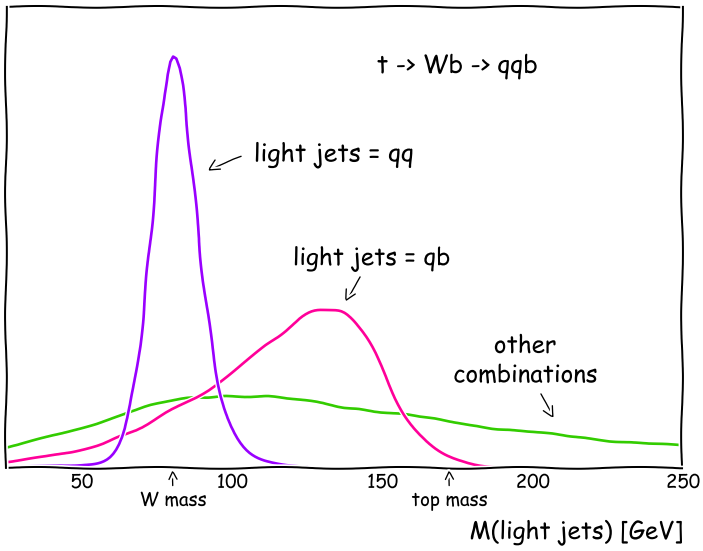

Figure 7.4: Sketch of the different contributions to the $M(W_{\mathrm{H}})$ distribution for the top-quark pair production. Depending on which partons the b-untagged jets actually correspond to, different spectra for $M(W_{\mathrm{H}})$ are reconstructed.

power between the Wt signal and all major backgrounds. For the $t\bar{t}$ background and even more so the Wt signal, the distribution peaks at m_W. The peak does not show up for the W+jets and the other major backgrounds, which do not produce a second W boson (in addition to the leptonically decaying W boson required by the event selection). Of course the physical mass of the W boson is the same for the Wt and the $t\bar{t}$ processes but the peak sits on top of a non-trivial combinatorial background that depends on the physics process. In order to understand the differences between the Wt signal and the $t\bar{t}$ background better, the samples can be broken down into three contributions, whose probability density functions are sketched[4] in Figure 7.4:

1. The two b-untagged jets actually coincide with the decay products of a W boson, giving rise to the peak at m_W. The width of the peak is primarily due to the energy resolution of the jet reconstruction. Radiative effects, which would give long tails of the distribution, are neglected in the figure.

2. One of the b-untagged jets coincides with a decay product of the W boson, whereas the other is actually a misidentified b-quark jet, and both originate from the same top quark. This contribution shows up as a peak-like structure to the right of m_W, and has an end-point (slightly smeared out) at the top-quark mass.

3. The b-untagged jets do not originate from the same top quark.

The relative abundances of the combinations differ between the Wt and the $t\bar{t}$ processes: the presence of a second b-quark jet with a large transverse momentum in $t\bar{t}$ events strongly enhances the second over the first contribution with respect to Wt events. The different admixtures of these contributions explain the distinct shapes of the $M(W_{\mathrm{H}})$ distribution for the Wt signal and the $t\bar{t}$ background. Unfortunately, the admixtures as well as the shapes of the individual contributions are affected by significant uncertainties,

[4] The sketch is based on the partonic final states in $t\bar{t}$ events generated with POWHEG, folded with transfer functions and efficiency maps for an approximate description of the detector response. Since no minimum separation is required for the two partons emerging from the decay of the W boson, $M(W_{\mathrm{H}})$ can go down all the way to 0. In a more realistic setting, the "cone size" used for jet finding and the cut applied on the transverse momentum of jets together lead to a non-zero end-point at the left side of the distribution.

especially due to the jet energy scale and resolution, and the theory modelling, which results in the large error bands visible in Figure 7.3. Nevertheless, the variable will play an important role in the later parts of the analysis, where it not only serves to suppress events without a hadronic decay of a W boson but also its sensitivity to the jet energy scale is exploited through the position of the peak.

While there are many more variables that separate the signal from the backgrounds, they are often correlated, in a similar way as $p_{\mathrm{T}}(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$ and $\Delta\phi(t, W)$, $M(W_{\mathrm{H}})$ and $M(t_{\mathrm{H}})$, or $\Delta R(\ell, j_{\mathrm{B}})$ and $M(t_{\mathrm{L}})$ are. The next section discusses how these correlations can be systematically taken into account.

7.2 Classification with NeuroBayes

For every event in the signal region, there are 14 independent measurements[5] of the kinematics of the reconstructed objects. In order to take the correlations among the measured values for the signal and the background into account, their joint probability densities needs to be analysed. The Neyman-Pearson lemma [134] suggests that a likelihood ratio[6] based on the joint probability density functions for the signal and the background achieves optimum separation. Unfortunately, the required probability density functions are not available as explicit formulae but can only be obtained by means of MC sampling. Using MC sampling, it is not even remotely possible to populate a sufficiently fine grid in the interesting regions of the 14-dimensional space. Therefore, the dimensionality needs to be reduced to a manageable level before the statistical analysis. This is where *multivariate analysis* (MVA) techniques often come into play in contemporary high-energy physics analyses.

Neural networks

The present analysis employs an MVA technique known as *artificial neural networks*. When presented with a representative set of events whose nature (signal or background) is known, the network learns to distinguish the signal and the background. In particular, NeuroBayes [135, 136] is utilised, which implements a *feed-forward multilayer perceptron*: an artificial neural network where the neurons are arranged into an input layer, an internal layer, and an output layer. It is depicted as a directed graph in Figure 7.5. The neurons in the input layer receive the values of the input variables, and are connected to neurons in the internal layer, which in turn propagate information to the output layer. The neuron in the output layer emits the response of the network, a value in the range from -1 to $+1$, with greater values indicating more signal-like events. The response has the interpretation as the probability of an event being a signal event, assuming that the examined sample and the training sample are identically distributed. This implies that the signal purity is expected to increase gradually as the network response increases, making it a very suitable discriminant for the statistical analysis (see Section 8.1).

Each of the connections has its individual *weight*, a multiplicative factor that is applied to every value transmitted over the connection. Every node in the internal and the output layer applies an *activation function*, $S(x)$, to the sum of the values received at its incoming connections. Hence, the network response can be expressed as:

$$\hat{y} = S\Big(\sum w_{j1}^{2\to 3} S\Big(\sum w_{ij}^{1\to 2} x_i\Big)\Big)\,, \tag{7.2}$$

where x_i denotes the value of the i^{th} input variable, and the w stand for the weights of the connections

[5] There are 3 degrees of freedom for the lepton, 2 for the missing transverse momentum, and 3 for each jet. The discrete degrees of freedom are ignored in the text: 1 for the charge of the lepton, 1 for its flavour. One continuous and one discrete degree of freedom could be subtracted because the physics is invariant under rotations and reflections; however, the region around the interaction point and the detector do not perfectly share this feature.

[6] The concept of the likelihood function is introduced in the description of the statistical analysis, see Chapter 8.

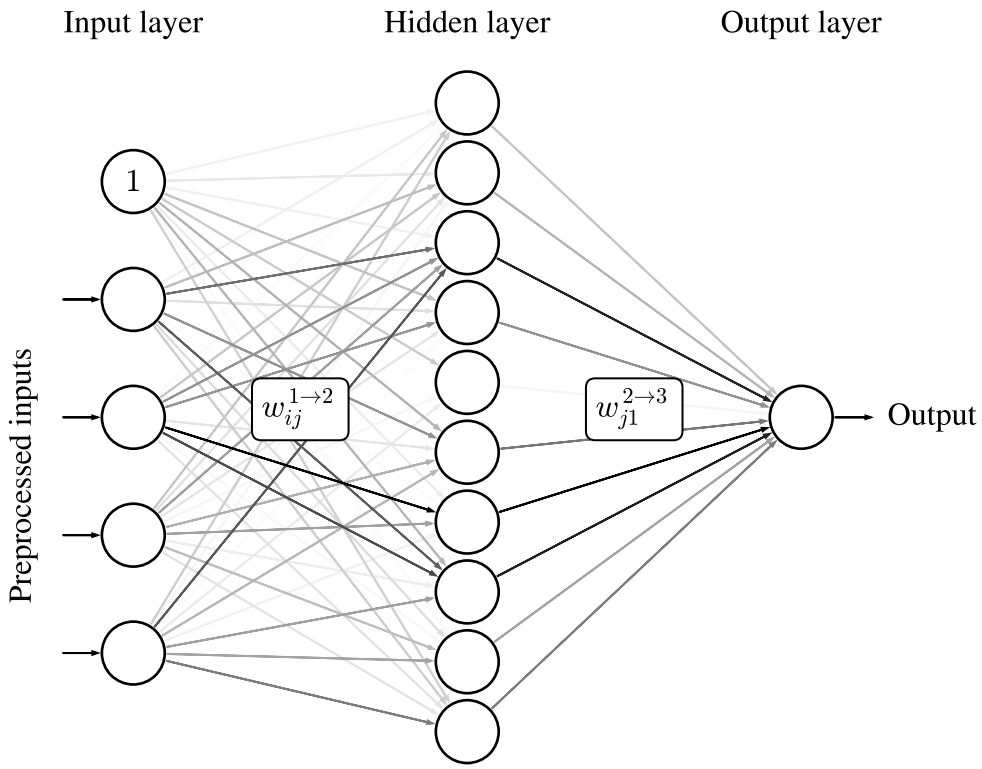

Figure 7.5: Illustration of a feed-forward multilayer perceptron. Every circle represents a neuron. The neurons at the left side (with the exception of the first node) receive the input values for each event. The output value is extracted at the right side. The significance of the weight of each connection is visualised as the intensity of the connecting line. The number of nodes and the weights have been taken from the network that will be used to separate the *Wt* signal from the $t\bar{t}$ background for the measurement of the *Wt* production cross section.

(cf. Figure 7.5). A sigmoidal function that maps the real axis non-linearly to the range of -1 to 1 is used as the activation function:

$$S(x) = \tanh 2x\,. \tag{7.3}$$

This function has the properties of being almost linear near $x = 0$, and approaching a value of ±1 asymptotically (*saturating*) for large $\pm x$. There is one node in the input layer sends a constant value of 1 to the internal layer, which can shift the argument of each activation function in the internal layer by the weight of the connection. This tunes the value at which the internal node is most sensitive to changes of its inputs. A neural network with the given activation function and a single internal layer can be trained to approximate any continuous function in a compact subspace of $\mathbb{R}^n$ [137].

Training sample

The training uses a different set of events from the rest of the analysis. For the signal, a sample based on the fast detector simulation was chosen, which adds the benefit of having a larger sample available for the training. Accordingly, the fast simulation is used for the background as well; otherwise, the neural network could erroneously attribute differences due to the kind of detector simulation as distinctive

features of the signal and the background. Only a fraction of the available background events are used in order to equalise the sample sizes for the signal and the background. The use of orthogonal samples for the training and the rest of analysis constitutes a perfect safeguard against any bias occurring when statistical fluctuations in the training samples are learned by the neural network (*overtraining*), which is already unlikely to happen due to the inherent protection offered by NeuroBayes.

Training algorithm

In order to improve the learning conditions [138], NeuroBayes draws the initial weights randomly from a normal distribution with a mean of 0 and a variance equal to the number of input connections. The training consists of several iterations over the events in the training sample. At the beginning of each iteration, all events are shuffled. Then a *backpropagation* [139] algorithm is used to adjust the weights such that the *entropy loss function* is minimised:

$$E = \sum_{i}^{\text{events}} \log \frac{1 + y_i' \hat{y}_i}{2}, \tag{7.4}$$

where y_i' denotes the actual class (*target*) of the i^{th} event: $+1$ for a signal event, -1 for a background event. The function quantifies the rate of misclassification that would occur in the events given the current weights of the network. After every 200 events, the weights are updated based on how much E changes when the weight of a connection to a neuron is varied. The training ends when convergence is reached. Overtraining (i.e. learning to "distinguish" signal and background on the basis of statistical fluctuations in the training sample) is prevented by pruning connections with insignificant weights, and penalising large weights. The procedure is explained in more detail in Ref. [135].

Preprocessing

NeuroBayes transforms the inputs before they reach the input layer of the neural network. The transformation is derived from the training samples before the actual training phase starts. The strategy is outlined in [140], and illustrated in Figure 7.6, exemplified by $M(W_{\text{H}})$. First, the distribution of $M(W_{\text{H}})$ (cf. Figure 7.3) is binned such every bin contains the approximately same number of events. A spline fit [141] is performed in order to obtain a smooth function estimating the signal purity in each bin. The function is then applied to the original variable, followed by an affine transformation that shifts the expectation value of the transformed variable to 0, and scales its variance to 1, ensuring optimal learning conditions for the neural network [138]. The transformation also guards against outliers that could otherwise cause saturation of neurons. Finally, the transformed variables are decorrelated. The correlation factors of the pairs of all transformed variables are shown in Figure 7.7.

There is a little caveat: unless the spline is strictly monotonic, information is lost through the transformation. An example of this can be seen in Figure 7.6: since the signal purity is 55 % for $M(W_{\text{H}}) \approx 58\,\text{GeV}$ as well as for $M(W_{\text{H}}) \approx 88\,\text{GeV}$, the neural network cannot know to which of the two original values the transformed value of about 0.5 corresponds to. Hence, it can sometimes be helpful to feed more variables into the neural network than the number of independent observables would suggest. Another consequence is that the network will simply ignore the sign of an input variable that is distributed symmetrically about 0. For such variables, it is actually better to remove the sign before presenting them to the preprocessor, making the spline approximation easier. Even better is the use of rapidity differences, which could otherwise not be unambiguously calculated from the transformed rapidity variables. A similar argument applies to azimuthal angles.

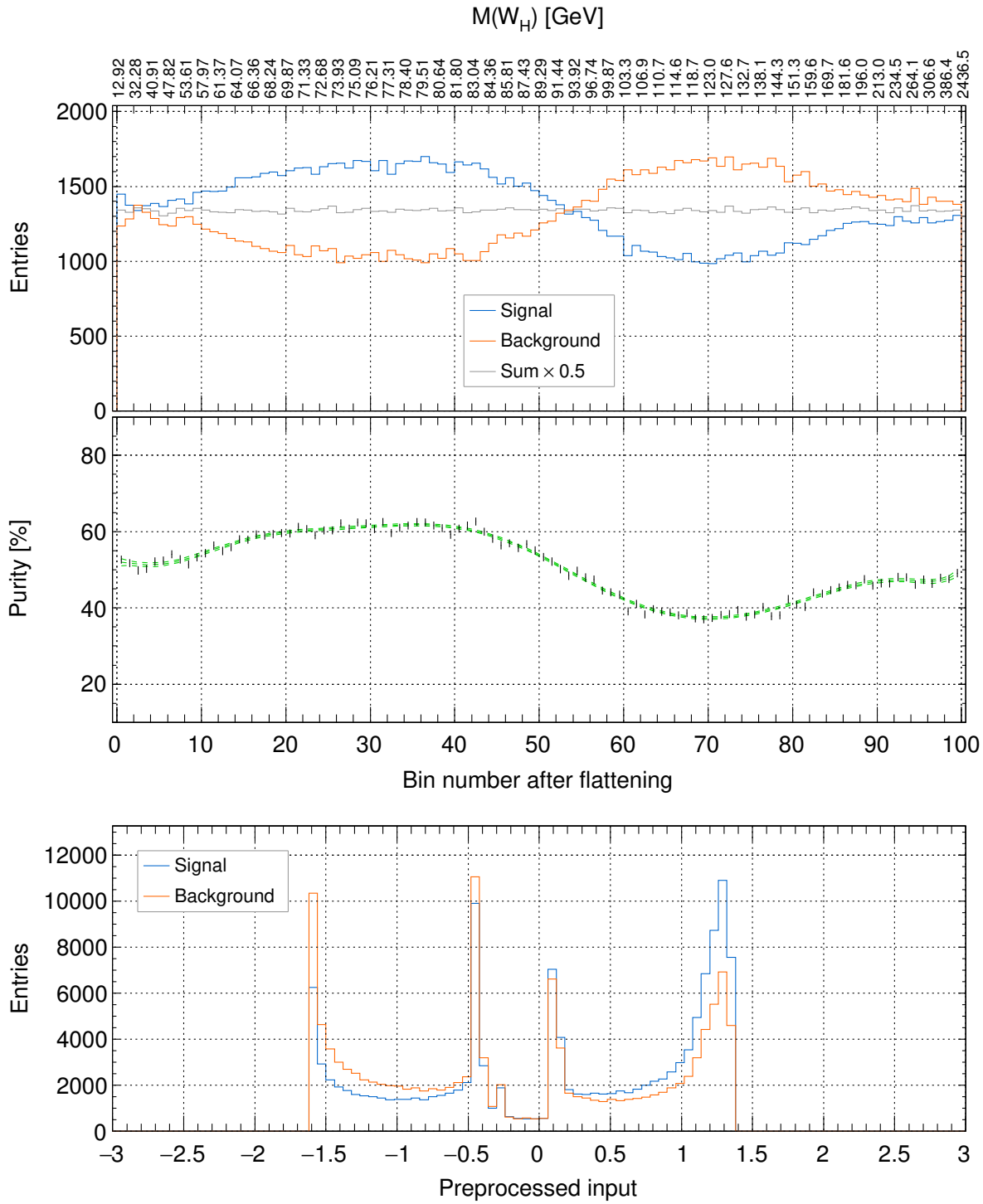

Figure 7.6: Illustration of the preprocessing of an input with NeuroBayes, exemplified by $M(W_{\mathrm{H}})$. The upper histogram shows the distribution of the variable, after binning it such that every bin contains the approximately same number of events. In the middle, the signal purity in each bin is shown, and fitted with a spline. The histogram at the bottom shows the distribution of the input variable after the applying the transformation.

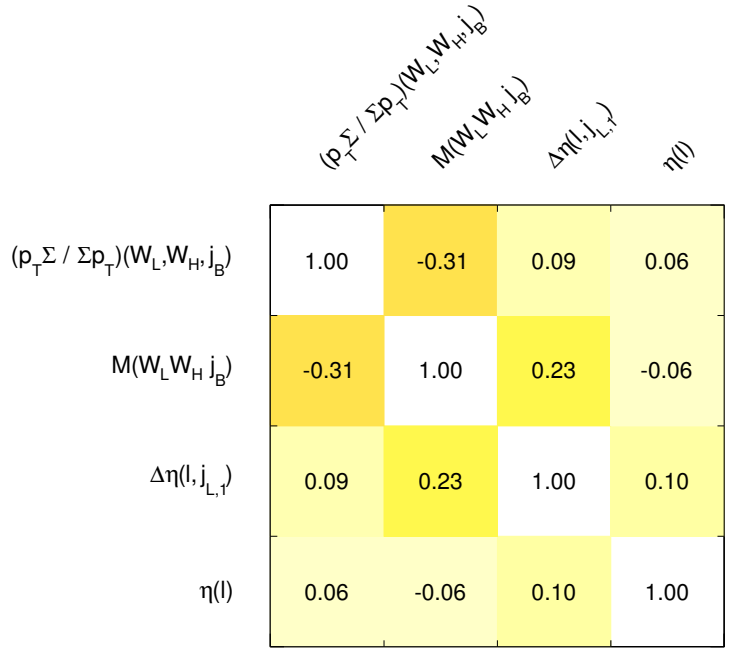

Figure 7.7: Linear correlations of the transformed input variables in the training sample. The variables and the construction of the training sample will be explained in Section 7.4.

Relation to cut-based event selection

Note that the cut-based event selection portrayed in Section 6.3 can be seen as a specific case of a manually tuned multivariate classification. Here, $\hat{y}$ can be expressed as a product of several Heaviside functions, one for each cut. If $\hat{y} = 1$, the event is selected, else if $\hat{y} = 0$, the event is discarded. The correlations between variables used in the cuts are not systematically taken into account. As a consequence, an event barely failing a single cut will be discarded even if it looks perfectly signal-like in all other variables. A better performance would be achieved by placing a cut on the response of an MVA, i.e. $\hat{y} > y_{\text{cut}}$. For a given value of y_{cut}, the multivariate estimator, $\hat{y}$, defines a hyperplane that separates a signal-enriched from a background-enriched region. The event selection as is does, however, have the distinct advantage that it allows the removal of regions where events cannot be reconstructed well, or the MC simulation fails to model the data.

7.3 Previous approaches

Usually, the neural network would be trained to separate the signal from a mixture of all background processes. However, it was realised in previous searches for the *Wt* production that this strategy produces neural network that are very sensitive to the effects of systematic uncertainties. This applied not only to neural networks, but also to strategies involving a kinematic fit [142] or boosted decision trees [143]. Hence, modified strategies were explored. In Ref. [144], which presents the search for *Wt* production at $\sqrt{s} = 7\,\text{TeV}$ with the ATLAS detector, a combination of two neural networks was used. One was dedicated to the separation from the $t\bar{t}$ background, while the second network was trained against the remaining backgrounds. The approach was motivated by the fact that the *Wt* signal is much harder to separate from the $t\bar{t}$ background than from any other background, and together with an automated

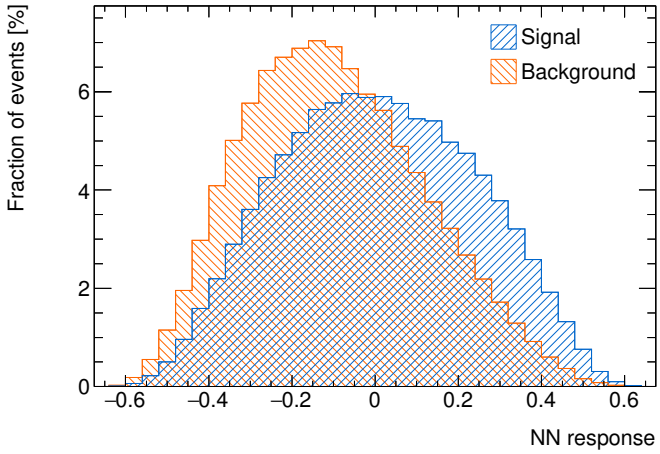

Figure 7.8: Spectra of the response of the neural network adapted from [144]. The signal (Wt) and the background ($t\bar{t}$) are distinguished by colours.

optimisation of the set of input variables, led to an improvement of the expected discovery significance[7] from 0.9 to 1.5 standard deviations. In Ref. [145] a similar approach was pursued, but the second network was trained to distinguish events with a well-reconstructed hadronic W-boson decay, based on a matching between the reconstructed jets and the partons in the MC simulation [146]. Here, an expected significance of 1.4 was achieved.

In every case the sensitivity was strongly limited by the uncertainties on the $t\bar{t}$ background. The upgrade of the collision energy from $\sqrt{s} = 7\,\text{TeV}$ to 8 TeV has a negligible influence on the ratio of the Wt and the $t\bar{t}$ cross sections. Only very little sensitivity is gained by just increasing the amount of data from the $4.7\,\text{fb}^{-1}$ used by the previous analyses to the $20.3\,\text{fb}^{-1}$ used in this analysis.

It is instructive to reproduce one of the neural networks used in the previous analyses, and explore possible improvements from there, focusing on the separation of the Wt signal and the $t\bar{t}$ background. A neural network is trained using the optimised set of 19 input variables from Ref. [144] and 10 nodes in the internal layer.[8] The two by far most relevant input variables are $M(W_\text{H})$ and $p_\text{T}(WWb)$, as expected from the studies in Section 7.1. The full list of variables can be found in Table E.2. Figure 7.8 shows the distributions of the neural network response to the signal and the background. Their shapes are clearly different but there is a strong overlap between the two distributions. The region with a neural network response close to 1, where a strongly signal-enriched region is typically found, is not populated. It is evident that the $t\bar{t}$ background cannot be efficiently suppressed by a simple cut on the response, and that a full analysis of the shape of the distributions will be necessary in order to extract the Wt signal.

Figure 7.9 presents the neural network response in bins of the most important input variable, $M(W_\text{H})$. The average response as a function of $M(W_\text{H})$ follows approximately the signal-over-background ratio in the training sample. Since the peak near m_W (see Section 7.1) is more pronounced in the Wt signal than in the $t\bar{t}$ background, the response is very high when $M(W_\text{H})$ is close to m_W. The relatively small difference between the average response of the signal and the background reveals that the other input variables do not contribute as much to the separation of signal and background as m_W alone does. The figure helps to understand why the uncertainties on the jet energy scale and resolution as well as the modelling of final-state radiation are problematic for an analysis that uses such a neural network: small variations of the jet energy near the sharply falling edge at 90 GeV translate into large variations of the response, making it very hard to distinguish a change in the fraction of signal events from a change of the

[7] See Section 8.5 for an explanation of the term *discovery significance*.

[8] The original training used 20 input variables, but the difference of the masses of the b-tagged jet and the leading b-untagged jet is dropped because the systematic uncertainties on the reconstruction of the jet mass are not well known.

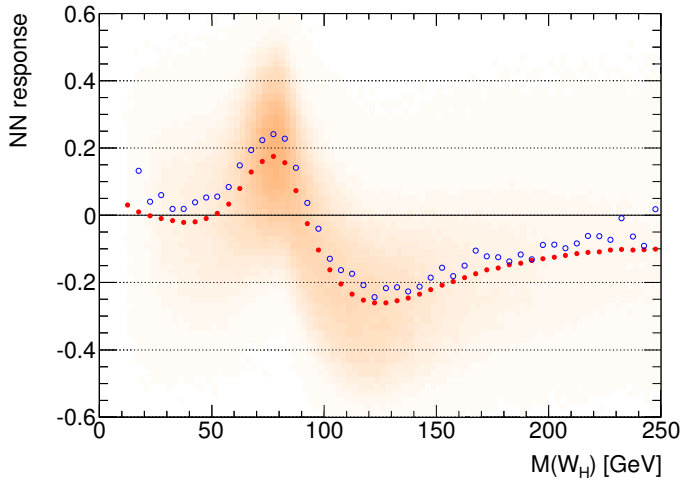

Figure 7.9: Distribution of events in bins of the neural network response and $M(W_{\mathrm{H}})$ for the $t\bar{t}$ background. Darker colours correspond to more events. The two curves show the average neural network response as function of $M(W_{\mathrm{H}})$, separately for the signal and the background. The fluctuations in the curve for the Wt signal are due to the limited statistical power of the sample used to probe the neural network.

jet energy scale. As the peak near m_W is essentially frozen into the neural network during the training phase, there is no freedom to adopt to a change of the jet energy scale. Clearly, a signal extraction strategy must be developed that can take such a shifting peak into account.

7.4 A new discriminant

As outlined in the previous section, the use of $M(W_{\mathrm{H}})$ as an input to a neural network does not lead to a satisfactory discriminant between the Wt signal and the $t\bar{t}$ background. However, the information carried by $M(W_{\mathrm{H}})$ is too valuable to be ignored. From the investigation of its shape (see Section 7.1), it follows that the Wt-to-$t\bar{t}$ ratio could be doubled, and most non-$t\bar{t}$ backgrounds removed if there were a way to select only events with a well-reconstructed hadronic W-boson decay. This could be accomplished to some extent by adding a sufficiently tight cut on $M(W_{\mathrm{H}})$ to the event selection.[9] Unfortunately, the performance of such a cut would be strongly limited by the very systematic uncertainties that already degrade the performance of the neural network considered in the previous section. To make things worse, the typical momenta of the W bosons are not much larger than their mass, so their decays are rather isotropic. Therefore, alternative variables like the angular separation of the two b-untagged jets cannot be efficiently used to identify the hadronic decay products of a W boson.

In order to solve the problem, an uncommon approach is adopted. As explained in the beginning of Section 7.2, the purpose of the neural network is to reduce the dimensionality of the variables representing the measurement of the final state in an event. Usually, all information would be mapped onto a single real axis, but for the present analysis, it is better to keep $M(W_{\mathrm{H}})$ out of this process. The two-dimensional distribution of $M(W_{\mathrm{H}})$ and the neural network response can then serve as a discriminant for the statistical analysis.

The neural network is trained using only events with a well-reconstructed hadronic W-boson decay. The training sample contains 50 000 signal events, and an equal amount of background events. After neglecting a tiny contribution from the diboson production, the training sample consists only of Wt and $t\bar{t}$ events. W+jets and other background events cannot have a well-reconstructed hadronic W-boson decay because the W boson must decay leptonically for event to pass the selection. Four input variables are used:

[9] See Ref. [147] for a study of a measurement of the Wt production cross section that involves such a cut. Note that the estimates of the systematic uncertainties therein were somewhat optimistic.

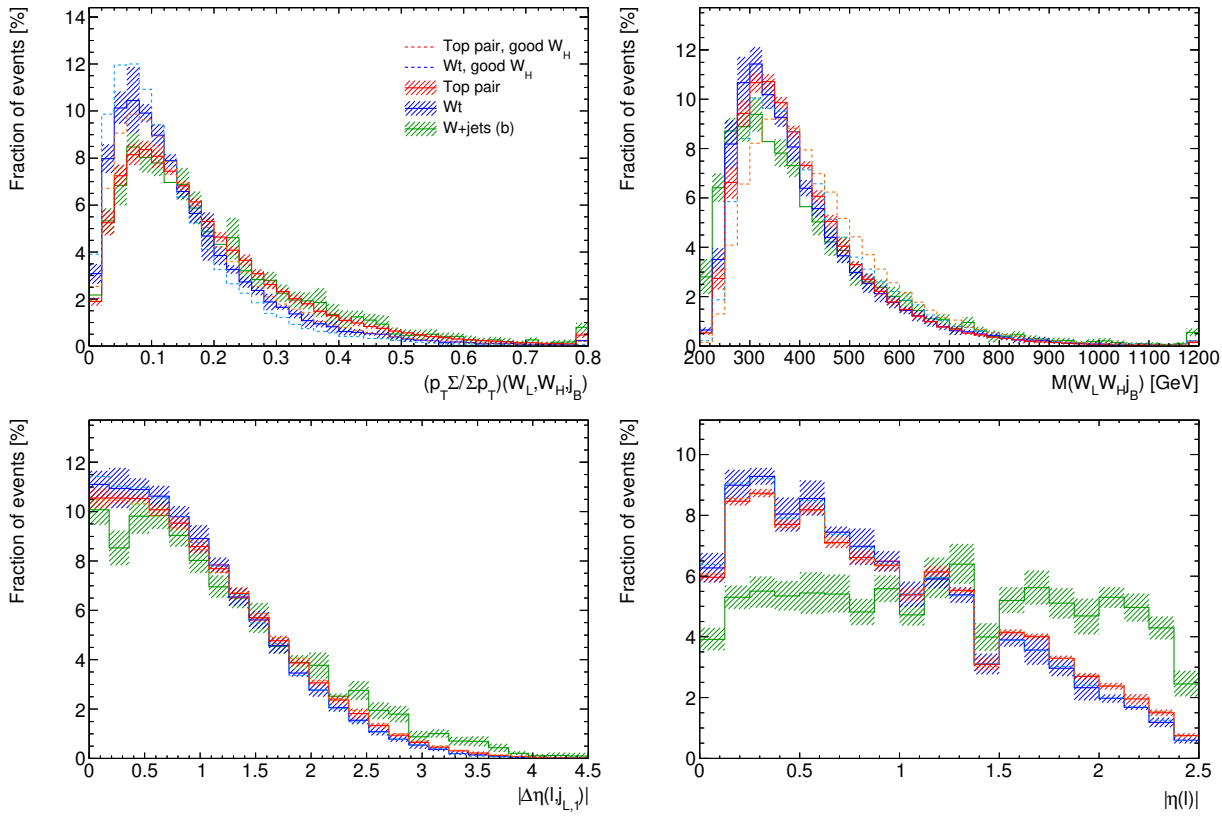

Figure 7.10: Spectra of the input variables of the neural network in the signal region for events satisfying 69 GeV $< M(W_\mathrm{H}) <$ 88 GeV. Dashed lines display the spectra with the additional requirement of a well-reconstructed hadronic W decay. The hatched bands indicate the effects of the systematic uncertainties. Since all distributions are normalised to unity, the uncertainties on the overall normalisation of every process cancel out.

1. the transverse momentum balance of the two reconstructed W bosons and the b-tagged jet:

$$(p_\mathrm{T}\Sigma/\Sigma p_\mathrm{T})(W_\mathrm{L}, W_\mathrm{H}, j_\mathrm{B}) = \frac{p_\mathrm{T}(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B})}{p_\mathrm{T}(W_\mathrm{L}) + p_\mathrm{T}(W_\mathrm{H}) + p_\mathrm{T}(j_\mathrm{B})};$$

2. the invariant mass, $M(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B})$, of the reconstructed W bosons and the b-tagged jet;

3. the difference, $\Delta\eta(\ell, j_{\mathrm{L},1})$, of the pseudorapidities of the lepton and the harder b-untagged jet;

4. the pseudorapidity, $\eta(\ell)$, of the lepton.

If $p_\mathrm{T}(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B})$ were used instead of the transverse momentum balance, the response of the neural network to W+jets events would tend to more positive values. Since this increase of the background in the signal-like region of the neural network response would result in a loss of sensitivity, the momentum balance is preferred.

Figure 7.10 shows the distributions of the input variables. Neglecting systematic uncertainties, the first two variables exhibit strong separation power on their own, with a considerable gain achieved by combining them (as explained in Section 5.2). Variables that are not symmetric under the exchange of W_L and W_H, such as the individual masses of the two reconstructed top-quark candidates, are consider-

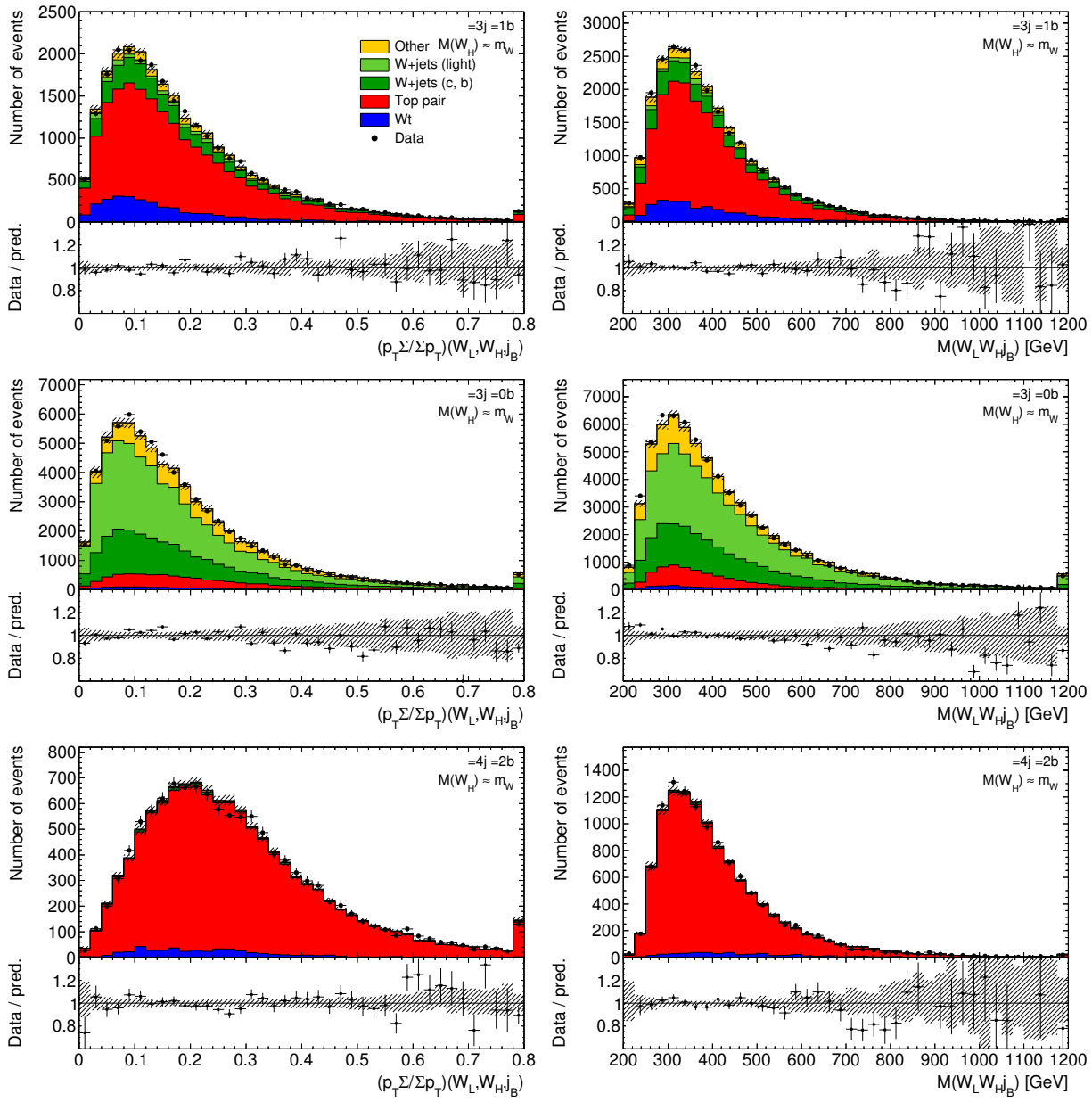

Figure 7.11: Distribution of the first two variables used as MVA inputs, $(p_{\mathrm{T}}\Sigma/\Sigma p_{\mathrm{T}})(W_{\mathrm{H}}, W_{\mathrm{L}}, j_{\mathrm{B}})$ (left) and $M(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$ (right), in the signal region (top row), the W+jets validation region (middle row), and the $t\bar{t}$ validation region (bottom row).

ably less powerful.[10] The pseudorapidity of the lepton is still a good pick because it is less sensitive to systematic uncertainties. Interestingly, $\Delta\eta(\ell, j_{\mathrm{L},1})$ exhibits no separation power on its own but becomes a lot more powerful when combined with the invariant mass. For a more quantitative assessment of the relevance of the input variables, see Appendix E. Because only events with a well-reconstructed hadronic W-boson decay are used in the training phase, variables such as $M(W_{\mathrm{H}})$ do not provide a significant gain.

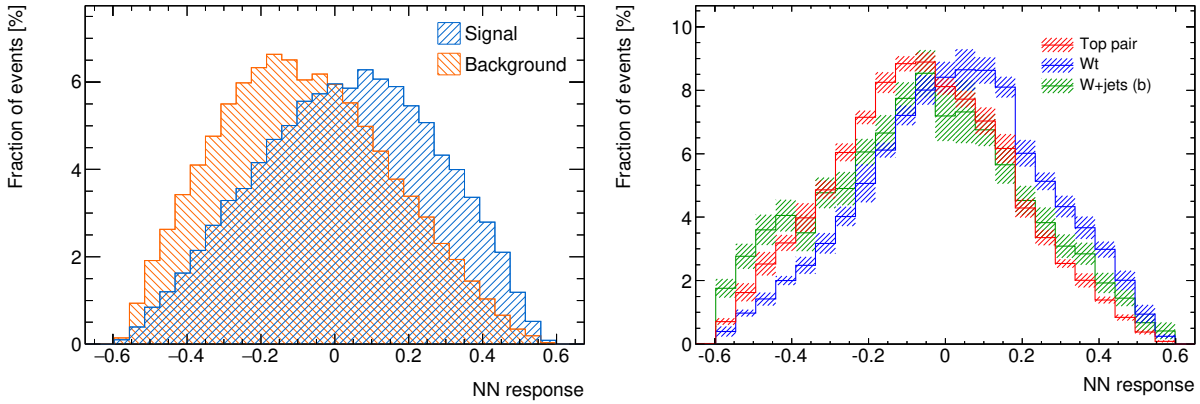

Figure 7.12: Spectra of the response of the neural network that separates the Wt and the $t\bar{t}$ production. The histograms on the left are based on the training samples, i.e. they include only events with a well-reconstructed hadronic W-boson decay, whereas the right plot shows the spectra for the events in the signal region with $69\,\mathrm{GeV} < M(W_{\mathrm{H}}) < 88\,\mathrm{GeV}$. All histograms are normalised to unity. The hatched bands indicate the effects of the systematic uncertainties.

Distributions for the first two input variables are shown in Figure 7.11 after cutting on $M(W_{\mathrm{H}})$ to remove the events for which the neural network is not used. The observed distribution of $M(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$ in the W+jets validation region is slightly softer than expected, but the difference is still within the uncertainties. The distributions for the two other variables are displayed in Figure D.2. In total, very good modelling of the four selected input variables is observed in the signal and the validation regions.

The tail of the $M(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$ distribution appears to be sensitive to the treatment of the interference between Wt and $t\bar{t}$. With the diagram removal scheme, a rising signal-to-background ratio is predicted in the tail, an effect not seen with the diagram subtraction scheme. As a precaution, only events with $M(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}}) < 500\,\mathrm{GeV}$ are used in the training process.[11] As a consequence, the response to events with $M(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}}) > 500\,\mathrm{GeV}$ is extrapolated by means of the transformation of the input variables (Section 7.2).

Figure 7.12 shows distributions of the neural network response. Incidentally, the separation power is not spoilt when the requirement of a well-reconstructed hadronic W-boson decay, which can only be applied in the simulation, is replaced by a simple cut on $M(W_{\mathrm{H}})$. The separation between the signal and the background is comparable to the approach presented in Section 7.3, although the network did not learn to use $M(W_{\mathrm{H}})$ or related variables. For further analysis, the distribution of the neural network

[10] It is possible to symmetrise by identifying which of the two W bosons originate from the top-quark decay (such as picking the candidate with a reconstructed mass close to m_t or using a neural network to distinguish the two cases). Unfortunately, attempts to do so did not boost the sensitivity. This is again due to the rather isotropic decays of the mother particle, here the top quark, combined with the limited reconstruction of the jet energy.

[11] An interesting alternative, training on a mixture of the two schemes, is not viable due to the very limited sample size for diagram subtraction scheme.

response is subdivided into eight bins, with the edges placed at:

$$-1, -0.30407, -0.17779, -0.09170, -0.01704, 0.05733, 0.14100, 0.26829, 1 \tag{7.5}$$

The binning is chosen such that the number of events, averaged over the Wt and the $t\bar{t}$ samples, is approximately constant over the bins. Since the neural network response is only meaningful for events with a well-reconstructed hadronic W-boson decay, it is used only when the value of $M(W_{\mathrm{H}})$ is between 65 GeV and 92.5 GeV.

The final discriminant is the two-dimensional distribution of the neural network response and $M(W_{\mathrm{H}})$, as shown in Figure 7.13. The events with $M(W_{\mathrm{H}}) > 225$ GeV are included in the last bin.[12] In order to visualise the histograms (and their uncertainties) in the usual manner, the bins are rearranged onto a one-dimensional axis by scanning them from left to right, bottom to top. The resulting one-dimensional distribution is presented in Figure 7.14, together with a comparison of the shapes. A review of the modelling of the discriminant is deferred until Section 8.4.

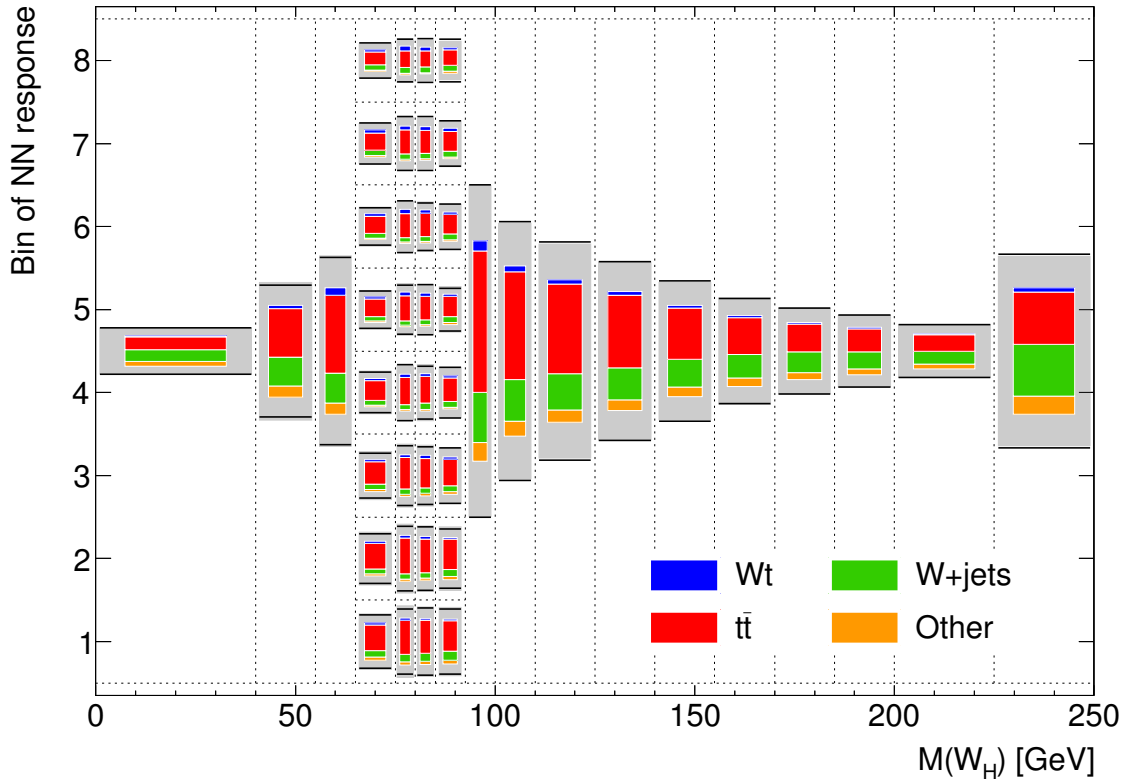

Figure 7.13: Predicted distribution of the discriminant in the signal region. The area contoured by each of the grey boxes is proportional to the number of expected entries in a particular bin. The proportions of the coloured areas reflect the expected composition in terms of Wt, $t\bar{t}$, W+jets and other processes. Also shown is the observed number of events in each bin, which is proportional to the area spanned by the pair of markers.

[12] Whether this overflow is included in the last bin has a negligible effect on the expected sensitivity.

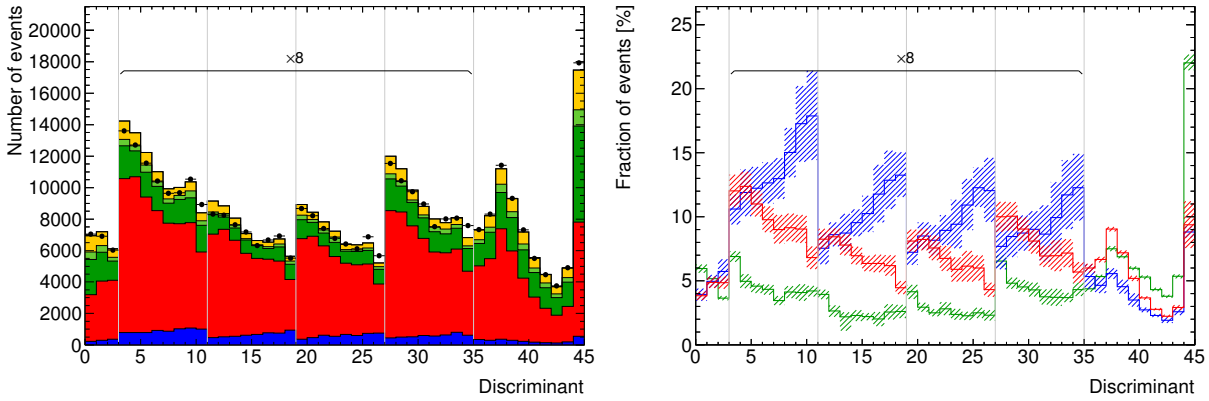

Figure 7.14: Predicted distribution of the discriminant in the signal region, rearranged onto a 1-dimensional axis. In the right plot, the distributions of the signal and the major backgrounds are normalised to 1 and not stacked upon each other, allowing a direct comparison of their shapes. In the bins where the neural network is used, the numbers of events are scaled by a factor of 8 for better visibility.

CHAPTER 8

Statistical Analysis

Using the distribution of the discriminant introduced in the previous chapter, the number of *Wt* events in the observed data can be determined, and translated into the measured cross section for *Wt* production. In this analysis, a binned profile likelihood fit is used to extract the cross section from the observed and expected distributions of the discriminant. In addition to the cross section for *Wt* production, the fit has additional degrees of freedom (nuisance parameters), which enable it to adopt to the effects of the systematic uncertainties on the discriminant. The fit also yields improved estimates for the nuisance parameters based on the observed data, leading to a reduction of initially overestimated uncertainties. The expected numbers and distributions for the signal and background processes are updated accordingly; these new expectations can be used to validate the fit result. Note that the figures in the previous chapters, with only a few exceptions, have already shown the updated (*post-fit*) versions of the distributions.

The chapter starts off with an introduction into the methodology of profile maximum-likelihood fits (Section 8.1). Afterwards, the sources of systematic uncertainty affecting the discriminant are recapitulated (Section 8.2). The presentation of the measured cross section is followed by a detailed discussion of the reduction of systematic uncertainties (Section 8.3). The statistical methods used to verify the modelling of the observed data after the reduction of the uncertainties are portrayed (Section 8.4). The chapter closes with the validation of the discriminant.

8.1 Likelihood Function

The signal cross section is extracted from the observed distribution of the discriminant using the principle of *maximum likelihood*. Given a parametrised model for the expected number events, one determines the parameter values that *fit* the observed data best. Let ν be the true event rate in the signal region, summed over the signal and all background processes:

$$\nu(\mu) = \mathcal{L}\epsilon_0\sigma_0\mu + \mathcal{L}\sum_{j}^{\text{bkg.}} \epsilon_j\sigma_j\,, \tag{8.1}$$

where $\mathcal{L}$ denotes the integrated luminosity, ϵ_j the efficiency for the reconstruction and selection of the events, and σ_j the cross section for each physics process, j. Because the signal cross section is to be measured, it is conventionally substituted with the theoretical cross section multiplied by a free parameter, μ. Neglecting systematic uncertainties, the probability of observing n events follows a Poison distribution:

$$\text{Pois}(n;\nu) = \frac{\nu^n}{n!}\mathrm{e}^{-\nu}\,. \tag{8.2}$$

The discriminant introduced in Section 7.4 comprises 45 bins. Since each bin constitutes a statistically independent measurement, the probabilities for the individual bins can be multiplied together, yielding

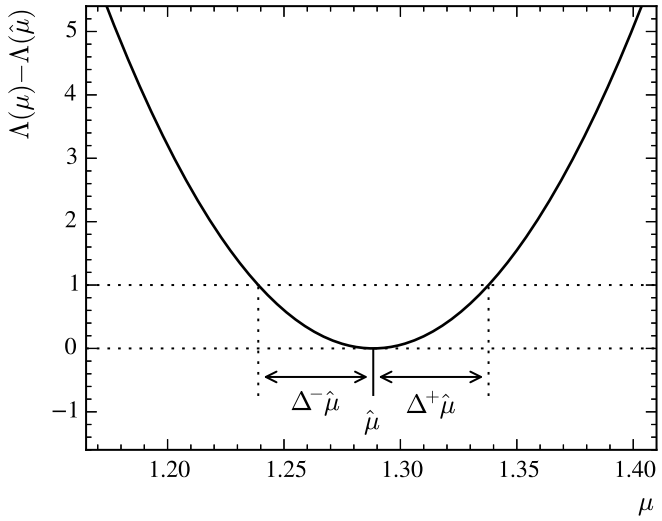

Figure 8.1: The log-likelihood function $\Lambda(\mu)$ plotted against μ after subtraction of an offset, $\Lambda(\hat{\mu})$. The construction of the error estimate, $\Delta\hat{\mu}$, using a scan of the likelihood function around the minimum at $\hat{\mu}$ is illustrated.

the *likelihood function*:

$$L(\mu; \vec{n}) = \prod_{i}^{\text{bins}} \text{Pois}(n_i; \nu_i(\mu)) \,, \tag{8.3}$$

where n and ν have been promoted to vectors, with an index i identifying the bin of the discriminant.

In practice it is more convenient to use the logarithm of the likelihood function, turning the multiplication over all bins into a sum of logarithms:

$$\Lambda(\mu; \vec{n}) = -2 \log L(\mu; \vec{n}) \tag{8.4}$$

The factor of -2 is a common convention. Given the observed distribution of the discriminant, the most likely value, $\hat{\mu}$, of the signal strength is found at the value of μ that minimises the log-likelihood function Λ, or equivalently maximises the likelihood function L. The standard error, $\Delta\mu$, on $\hat{\mu}$ is constructed by scanning the neighbourhood of $\hat{\mu}$ until Λ increases by one unit [7]. This is illustrated in Figure 8.1, which also demonstrates the typical shape of $\Lambda(\mu)$.

The likelihood function is composed and evaluated with the help of the HistFactory programme [148], part of the RooStats framework [149]. The minimisation is performed with the Minuit package [150].

In reality neither the cross sections of the background processes nor the efficiencies are known exactly. In order to take this into account, the likelihood function is extended by *nuisance parameters* that describe how the uncertainties change the event rates. The concept will be demonstrated using the simple example of a (fictitious) uncertainty of $\delta = 10\,\%$ on the overall detector acceptance. A nuisance parameter, θ, is introduced into Formulae 8.1 and 8.3:

$$\nu_i(\mu, \theta) = (1 + \delta \cdot \theta)\, \nu_i(\mu, 0) \tag{8.5}$$

$$L(\mu, \theta; \vec{n}) = \prod_{i}^{\text{bins}} \text{Pois}(n_i; \nu_i(\mu, \theta))\, f_{\mathcal{N}}(\theta; 0, 1) \,. \tag{8.6}$$

A value of $\theta = +1$ ($\theta = -1$) is identified with the increase (decrease) of the acceptance, and thus the event rate, by δ. The term $f_{\mathcal{N}}(\theta; 0, 1)$ expresses that the p.d.f. of θ follows a normal distribution with a mean of 0 and a width of 1. The formulae for ν_i can be generalised to variations that modify the rates of the individual processes by different factors, or change the distribution of the events over the individual

bins. The recipe is applied separately to the overall normalisation of each process, and to each bin of the shape histogram of the discriminant for each process. For the sake of conciseness, the index i will be omitted from here on.

In total 74 sources[1] of systematic uncertainties are considered for the present analysis. The effects of the individual nuisance parameters are combined additively:

$$\nu(\mu, \vec{\theta}) = \nu(\mu, \vec{0}) \left(1 + \sum_k \delta_k \theta_k \right), \tag{8.7}$$

where k identifies the nuisance parameter. There is no universal expression for the combination; any function with a first-order Taylor series expansion given by Equation 8.7 is a valid choice. Therefore, HistFactory offers several alternatives to this "linear interpolation" scheme [148]. A commonly adopted strategy is to use the linear interpolation only for the shapes, and an exponential interpolation for the overall normalisation of each physics process. The exponential scheme is similar to the linear scheme but operates on $\log \nu$ instead of ν, motivated by the fact the number of selected events is a product of efficiencies and the cross section. As an example, if two nuisance parameters increase the event rate by 10 % each, their combined effect is a 21 % (because $1.10 \cdot 1.10 = 1.21$) increase instead of the 20 % increase obtained with the additive scheme. However, the differences between the schemes are small unless the variations are large and the nuisance parameters have values far away from 0. For the present analysis, the choice of interpolation scheme has a negligible effect on the result, and therefore the simpler additive scheme is preferred.

The presence of nuisance parameters slightly complicates the construction of the uncertainty on μ. The picture in Figure 8.1 still applies, however, for each scanned value of μ, $\Lambda(\mu, \vec{\theta})$ has to be minimised with respect to $\vec{\theta}$. For every $\mu \neq \hat{\mu}$, there are in general values of $\vec{\theta}$ that lead to a smaller Λ than $\Lambda(\hat{\mu}, \hat{\vec{\theta}})$, so the parabolic shape broadens, corresponding to an increase in the uncertainty on $\hat{\mu}$. The calculation of the error interval of a nuisance parameter and of the error contour of a pair of parameters works analogously. Alternatively, the error intervals and contours can be approximated using the covariance matrix, V_{kl}, of the parameter estimates, which is calculated numerically from the second-order derivatives of the likelihood function at its minimum:

$$V_{kl}^{-1} = -\frac{\partial^2 L}{\partial \hat{\theta}_k \, \partial \hat{\theta}_l}, \tag{8.8}$$

letting $\hat{\theta}_0 \equiv \hat{\mu}$ in this particular formula in order to keep the expression simple.

In order to judge the relevance of systematic uncertainties for the measurement, the *impact* of each systematic uncertainty on the measured cross section needs to be known. It is defined as the shift of the maximum-likelihood estimate for μ when changing the value of the nuisance parameter from $\hat{\theta}_k$ to $\hat{\theta}_k \pm \Delta\theta_k$. Again, it can be approximated using the covariance matrix:

$$\frac{\operatorname{cov}\left[\hat{\mu}, \hat{\theta}_k\right]}{\sqrt{\operatorname{cov}\left[\hat{\theta}_k, \hat{\theta}_k\right]}}. \tag{8.9}$$

The results obtained from the covariance and from the construction using the conditional[2] maximum-likelihood fit are very similar (cf. Table G.1).

Assuming that all systematic effects are correctly taken into account, each error interval defines a

[1] Not counting the 45 independent uncertainties due to the limited sample size of the simulation.

[2] The likelihood function is called *conditional* if one or more of its parameters (here $\hat{\theta}_k$) are fixed to a constant value.

68.3 %-level *confidence interval* on the respective parameter. The interpretation of the confidence interval is: if the full experiment (including any auxiliary measurements such as the determination of the jet energy scale) were repeated a large number of times, the true value of the parameter would be inside the interval for 68.3 % of the experiments.

The nuisance parameters not only fulfil the purpose of error propagation: with a suitable discriminant, they will actually be measured from, or *constrained* by, the observed data. Often, the initial uncertainties put into the model, such as the uncertainty on the W+jets normalisation, are deliberately conservative, and can be more reliably estimated from the observed data. As an example, an experiment with a initial uncertainty of 10 % on the background normalisation is considered. If the discriminant comprised just a single bin, the uncertainty would propagate directly into the measured signal strength, yielding a systematic uncertainty of approximately 0.1 divided by the signal-to-background ratio. However, when adding a second bin with a different signal-to-background ratio,[3] the signal and the background become distinguishable, and the background normalisation can be determined from the observed data. If the second bin is free of signal events, the number of background events is constrained to the statistical error of the number of events in that bin. Hence with a sufficient amount of data, large uncertainties can be effectively reduced. The technique works not only for the background normalisation but for any parameter to which the discriminant is sensitive. A good example is the jet energy scale, which is estimated using the peak in the distribution of the invariant mass of the two b-untagged jets.

This sort of "calibration" relies on the correctness and the completeness of the model.[4] In the above fictitious fit of two bins, there will in general be other sources of uncertainties that add an error on the extrapolation of the background from the second to the first bin. Unless negligible, these uncertainties must be taken into account, as otherwise the background normalisation in the first bin will be *overconstrained*. Whether the reduction of an uncertainty is reasonable, has to be judged on a case-by-case basis. Two sections, Section 8.3 and Section 8.4, have a strong focus on checks of this delicate procedure.

Sometimes knowledge of the expected sensitivity of the measurement and the constraints on the nuisance parameters is needed without access to the observed data, for example in order to optimise the analysis. This is achieved by replacing the observed data with an *Asimov dataset* [152] that is obtained by substituting $\nu_i(1, \vec{0})$ for n_i in the likelihood function.

8.2 Overview of systematic uncertainties

The value of every δ_k is, in general, computed by modifying a specific aspect of the simulation or the reconstruction, and propagating the effect into the discriminant. As an example, the uncertainty of the jet energy scale is known from the auxiliary measurements described in Section 4.2. When the jet energy scale in the simulation is increased (decreased) by this uncertainty, the number of selected events changes from ν to ν_k^+ (ν_k^-). δ_k is calculated from the difference:

$$\delta_k = \frac{\nu_k^+ - \nu_k^-}{2\nu} \,. \tag{8.10}$$

The $\nu^{(\pm)}$ may refer to the total number of events for a given process, or to the number inside a single bin of the discriminant. For those uncertainties that do not have separate variations in the upwards ($\theta_k = +1$) and the downwards ($\theta_k = -1$) direction, δ_k is taken to be the relative deviation of the varied from the nominal rate.

[3] Recall from Section 7.2 that NeuroBayes (as well as other multivariate analysis techniques) produces a discriminant with a monotonically increasing signal-to-background ratio.

[4] Some helpful advices regarding the best practices in dealing with nuisance parameters are summarised in Ref. [151].

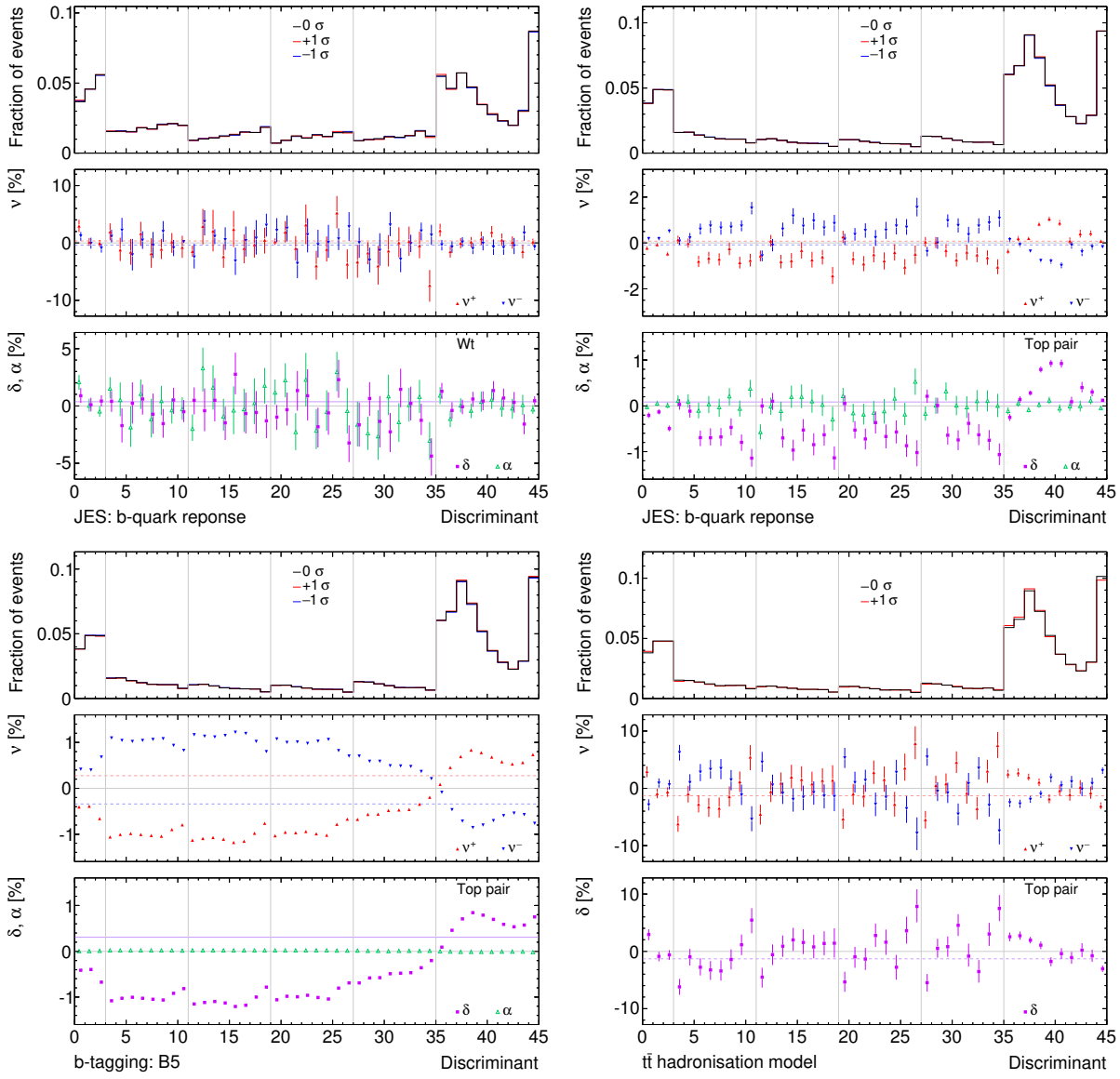

Figure 8.2: Relative deviations in normalised distribution of the discriminant due to systematic effects. Each plot is subdivided into three parts: the nominal and the varied distributions (upper part), the relative deviation of the upwards/downwards variation with respect to the nominal (middle part), and the corresponding δ and α (lower part). The middle and the lower part may have different scales on the y-axis. The deviations due to variations of the detector response for b-quark jets are shown in the top row, separately for the Wt signal (top left) and the $t\bar{t}$ background (top right). In the bottom row, the deviations in the distribution of the $t\bar{t}$ background due to the most relevant eigenvariation of the b-tagging efficiencies (bottom left) and the variation of the hadronisation model (bottom right) is presented.

Uncertainty	Wt		$t\bar{t}$		W+jets, beauty	
	δ_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	δ_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	δ_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$
b-tagging: B0	0.0 ± 0.0	1.4	0.0 ± 0.0	78	0.0 ± 0.0	1.2
b-tagging: B1	-0.1 ± 0.0	2.5	0.0 ± 0.0	580	-0.2 ± 0.0	10
b-tagging: B2	-0.3 ± 0.0	14	0.1 ± 0.0	1 800	-0.3 ± 0.0	3.7
b-tagging: B3	-0.4 ± 0.0	16	0.1 ± 0.0	2 500	-0.4 ± 0.0	5.7
b-tagging: B4	-0.1 ± 0.0	4.4	0.0 ± 0.0	440	-0.1 ± 0.0	4.4
b-tagging: B5	-0.9 ± 0.0	14	0.3 ± 0.0	2 200	-1.0 ± 0.0	4.6
b-tagging: C	-0.2 ± 0.0	7.6	-0.3 ± 0.0	220	0.2 ± 0.0	3.1
b-tagging: L	-0.1 ± 0.0	20	-0.1 ± 0.0	19	0.3 ± 0.0	3.0
JES: b-quark reponse	0.4 ± 0.1	1.0	0.1 ± 0.0	20	1.0 ± 0.0	1.0
JES: detector 1	0.2 ± 0.1	2.0	0.0 ± 0.0	38	0.3 ± 0.0	1.2
JES: detector 2	0.2 ± 0.1	1.4	0.0 ± 0.0	3.8	0.3 ± 0.0	1.4
JES: detector 3	0.1 ± 0.0	1.4	0.0 ± 0.0	5.1	0.0 ± 0.0	1.2
JES: mixed 1	0.2 ± 0.1	1.7	0.0 ± 0.0	27	0.4 ± 0.0	0.7
JES: mixed 2	-0.2 ± 0.1	1.5	0.0 ± 0.0	6.0	-0.5 ± 0.0	1.4
JES: mixed 3	0.1 ± 0.0	1.2	0.0 ± 0.0	18	0.1 ± 0.0	2.1
JES: mixed 4	0.0 ± 0.0	0.8	0.0 ± 0.0	3.1	0.0 ± 0.0	1.2
JES: modelling 1	2.0 ± 0.2	3.0	-0.1 ± 0.0	62	3.0 ± 0.1	2.9
JES: modelling 2	-0.1 ± 0.0	1.4	0.0 ± 0.0	3.0	-0.1 ± 0.0	1.2
JES: modelling 3	0.0 ± 0.0	1.2	0.0 ± 0.0	32	0.0 ± 0.0	2.4
JES: modelling 4	0.0 ± 0.0	0.9	0.0 ± 0.0	7.4	0.1 ± 0.0	1.1
JES: statistical 1	0.6 ± 0.1	1.3	0.0 ± 0.0	13	1.1 ± 0.0	2.4
JES: statistical 2	-0.1 ± 0.0	1.2	0.0 ± 0.0	3.6	-0.1 ± 0.0	1.4
JES: statistical 3	-0.2 ± 0.1	1.5	0.0 ± 0.0	21	-0.3 ± 0.0	0.7
JES: statistical 4	0.2 ± 0.0	1.1	0.0 ± 0.0	9.0	0.2 ± 0.0	1.2
JES: η intercalibration: modelling	0.4 ± 0.1	1.6	0.0 ± 0.0	21	0.7 ± 0.0	0.8
JES: η intercalibration: statistical	0.3 ± 0.1	1.3	0.0 ± 0.0	12	0.6 ± 0.0	0.8
JES: quark/gluon composition	1.7 ± 0.2	3.1	-0.3 ± 0.0	66	1.8 ± 0.1	3.2
JES: quark/gluon response	-0.9 ± 0.2	2.5	0.2 ± 0.0	57	-1.2 ± 0.1	1.9
JES: pile-up: OffsetMu	-0.1 ± 0.1	0.9	0.0 ± 0.0	3.5	-0.2 ± 0.0	1.1
JES: pile-up: OffsetNPV	-0.2 ± 0.2	1.1	0.1 ± 0.0	6.5	-0.2 ± 0.1	1.3
JES: pile-up: Pt	0.0 ± 0.1	1.2	0.0 ± 0.0	1.9	0.1 ± 0.0	1.1
JES: pile-up: RhoTopology	1.1 ± 0.2	2.0	-0.1 ± 0.0	37	2.1 ± 0.1	3.0
JES: punch through	0.0 ± 0.0	0.6	0.0 ± 0.0	2.2	0.0 ± 0.0	0.1
JES: high-p_{T} single particle	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0
Jet reconstruction eff.	0.0 ± 0.0	0.6	0.0 ± 0.0	1.7	0.0 ± 0.0	1.8
Jet energy resolution	0.6 ± 0.8	1.8	-0.3 ± 0.1	17	2.3 ± 0.3	1.4
Jet vertex fraction	-0.3 ± 0.1	2.5	0.0 ± 0.0	21	-0.5 ± 0.0	4.8
$e^{\pm}$ energy resolution	0.1 ± 0.1	1.1	0.0 ± 0.0	0.9	0.0 ± 0.0	1.0
$e^{\pm}$ energy scale	0.8 ± 0.1	1.0	0.7 ± 0.0	4.7	0.9 ± 0.0	0.9
$e^{\pm}$ identification eff.	1.0 ± 0.0	1.1	1.0 ± 0.0	4.3	1.0 ± 0.0	1.0
$e^{\pm}$ reconstruction eff.	0.1 ± 0.0	2.0	0.1 ± 0.0	48	0.1 ± 0.0	3.5
$e^{\pm}$ trigger eff.	0.2 ± 0.0	1.5	0.2 ± 0.0	29	0.3 ± 0.0	4.3
$\mu^{\pm}$ energy resolution: inner detector	0.1 ± 0.1	1.4	0.0 ± 0.0	0.8	0.0 ± 0.0	1.3
$\mu^{\pm}$ energy resolution: spectrometer	-0.1 ± 0.1	1.2	0.0 ± 0.0	1.0	0.0 ± 0.0	0.7
$\mu^{\pm}$ energy scale	0.0 ± 0.0	4.2	0.0 ± 0.0	1.2	0.1 ± 0.0	2.6
$\mu^{\pm}$ identification eff.	0.3 ± 0.0	1.0	0.3 ± 0.0	2.0	0.3 ± 0.0	1.1
$\mu^{\pm}$ reconstruction eff.	0.1 ± 0.0	1.1	0.1 ± 0.0	4.0	0.1 ± 0.0	2.1
$\mu^{\pm}$ trigger eff.	1.0 ± 0.0	1.1	1.0 ± 0.0	1.4	1.0 ± 0.0	1.0
$p_{\text{T}}^{\text{miss}}$ soft term resolution	0.1 ± 0.1	1.1	0.0 ± 0.0	1.2	0.1 ± 0.0	1.4
$p_{\text{T}}^{\text{miss}}$ soft term scale	0.1 ± 0.1	1.5	0.1 ± 0.0	26	0.2 ± 0.0	1.4
NLO subtraction scheme	-0.9 ± 0.6	2.8	5.3 ± 0.3	12	–	–
Hadronisation model	0.6 ± 0.4	1.5	-1.3 ± 0.2	4.4	–	–
Renormalisation scale	-6.9 ± 0.6	3.3	-2.2 ± 0.1	54	–	–
Wt-$t\bar{t}$ interference	2.4 ± 1.6	1.6	–	–	–	–

Table 8.1: Effect of the systematic uncertainties (excluding uncertainties on the overall normalisation, which are given in the text) on the number of selected events for the Wt signal, the top-quark pair production, and W-boson production in association with b-quark jets. The statistical error on the number is computed using the bootstrap method. Also shown is the result of the χ^2 test comparing the nominal and the varied shape of the discriminant. See Appendix F for an explanation of the bootstrap method, and more tables that cover the other processes and demonstrate that the asymmetries of the uncertainties can be neglected.

Figure 8.2 exemplifies the effects that a few of the systematic uncertainties have on the shape of the discriminants. Since the δ are computed using finite samples, they may suffer from statistical fluctuations that are larger than the actual systematic effects. It is important to include only statistically significant effects in the fit, in particular for the shapes of the discriminant, as otherwise the fluctuations may lead to a double-counting of the statistical error or spurious constraints on the nuisance parameters. For most sources of uncertainty, the $\nu^{(\pm)}$ are determined from the same underlying set of events, making the use of the bootstrap method (see Appendix F) necessary in order to calculate the statistical error of δ. For every systematic uncertainty, the significance of the deviations in shape is assessed using a χ^2 for compatibility with zero on the δ computed in each bin of the discriminant (i.e. on the graphs labelled with "δ" in Figure 8.2). A small value of χ^2 ($\lesssim 1.5$) suggests that the estimate of the variation on the shape is dominated by statistical fluctuations, and is used as an indicator to neglect[5] the particular variation. Note that the effects of some sources of systematic uncertainty, e.g. the *b*-tagging efficiencies, are evaluated without changing the set of the selected events – only the event weights are modified a little. This results in a much smaller statistical error on the δ, and consequently to a large χ^2. Hence for these kinds of systematic uncertainties, the deviations in shape are always taken into account (which does *not* necessarily mean that the discriminant is very sensitive to the uncertainty!). Table 8.1 summarises the deviations of the rate and the result of the χ^2 test for every source of systematic uncertainty for the *Wt* signal and the two major backgrounds. The individual systematics are discussed in the following subsections.

For every uncertainty, the asymmetry,

$$\alpha = \frac{\nu^+ + \nu^-}{2\nu}\,, \tag{8.11}$$

is found to be consistent with 0 or much smaller than the corresponding δ. Hence, it is safely neglected for the statistical analysis, avoiding the complications of the treatment of asymmetric uncertainties as well as reducing the statistical error of the δ by about 30 % (as can be seen by comparing the errors of $\nu^\pm$ and of δ in Figure 8.2).

8.2.1 Reconstruction uncertainties

This section summarises the uncertainties introduced in Chapter 4, and portrays their effects on the discriminant. Variations of efficiencies are performed by either changing the scale factors that correct the simulated efficiencies to the observed ones, or by dropping a random fraction of the reconstructed objects in the simulation. The uncertainties on the reconstructed four-momenta are evaluated by adjusting the corresponding energy corrections in the simulation.

Jet energy scale

The uncertainty due to the jet energy scale (JES) is estimated by varying the reconstructed energies of the jets in the simulated samples. The JES is described using multiple nuisance parameters, taking into account that the uncertainties on jets with different energies or pseudorapidities are not fully correlated.

The component "modelling 1" is dominant among the components. It describes a variation of the JES that increases as the transverse momentum of the jet decreases, and is independent of the direction of the jet. The magnitude of the effect is about 2 % to 1 % for jets with transverse momenta between 20 and 60 GeV. Like the other "modelling" components, it originates from discrepancies that were seen when

[5] Without this pruning mechanism, the total uncertainty on the measured cross section would increase by only 1 %, however, the estimates of the nuisance parameters would fluctuate much more strongly (compare Figure 8.3 with G.4).

switching the MC generator in the jet calibration procedure, especially for smaller transverse momenta of the jets. Other components describe uncertainties due to the detector description ("detector"), the statistical precision and fitting method ("statistical"), or an interplay of the physics modelling and the detector description ("mixed").

Additional uncertainties are related to the subtraction of pileup, and differences in the detector response for jets initiated by a gluon, a light quark, or a b-quark. The additional uncertainty on the response for b-quark jets is quantitatively similar to the "modelling 1" component. Pileup is relevant primarily for jets with transverse momenta below 30 GeV, which affect the analysis indirectly through the calculation of the missing transverse momentum.

Two more uncertainties have been evaluated but not included in the statistical analysis because their effect is essentially zero in this analysis. One is due to the extrapolation of the JES to jets with $p_{\mathrm{T}} > 1.5$ TeV, which do not appear in this analysis. The other is related to "punch-through", where the energy of a highly energetic jet leaks out beyond the calorimeter.

The discriminant is designed to be sensitive to the JES of the light-quark jets (see Section 7.4), allowing the JES to be constrained by the observed data.

Jet energy resolution

In order to estimate the effect of the jet energy resolution (JER) uncertainty, the energy of each jet is smeared by a conservative estimate of the uncertainty of its resolution. As it is nearly impossible to reduce the resolution after running the detector simulation, the -1σ variation is not explicitly constructed but obtained via interpolation (see the comment below Equation 8.10). Variations of the JER widen or narrow the mass peak of the W boson in the distribution of $M(W_{\mathrm{H}})$. Similarly to the JES uncertainties, it is expected that the uncertainty due to the JER can be constrained by the data.

Jet reconstruction efficiency

Apart from the uncertainties on the reconstructed energy of the jets, there are two more uncertainties related to jet reconstruction:

- the very small uncertainty on whether a jet is reconstructed at all, estimated by randomly dropping simulated jets;
- the uncertainty on the JVF quantity, which is used to suppress jets from pileup.

Identification of b-quark jets

The uncertainty due to the identification of b-quark jets is estimated by varying the efficiency correction in the MC simulation. The efficiencies for b-quark, c-quark and light-flavour jets are varied separately, as they are the results of independent measurements (see Section 4.2.1). The uncertainty on fraction of b-quark jets that are erroneously identified as light-flavour jets has a large effect on the discriminant (see discussion of $M(W_{\mathrm{H}})$ in Section 7.1). It is expected that the b-tagging efficiency is constrained by the observed data.

As for the JES, the description of the b-tagging may depend on the kinematics of the jet, so the variation of the efficiency for true b-quark jets is subdivided into 6 components. The last component is the dominant one, and corresponds roughly (given that most b-quark jets in this analysis have a

transverse momentum between 30 GeV and 100 GeV) to a variation of the overall tagging efficiency for b-quark jets by 1.5 %.[6]

Leptons

The uncertainties due to the reconstruction of leptons are subdivided into:

- the energy scale;
- the energy resolution, which for muon is treated independently in the Inner Detector and in the Muon Spectrometer;
- the tight identification cuts;
- the efficiency whether the lepton is reconstructed at all;
- the efficiency of the trigger chain.

The uncertainties for electrons and for muons are treated as uncorrelated between each other. In total they add up to an uncertainty of about 1 % on the event rates, with very little effect on the shape of the discriminant.

Fake-lepton estimate

There are various uncertainties on the modelling of the fake-lepton background (see Section 6.2.5):

- choice of control region for the determination of the "fake efficiency";
- choice of control region for the determination of the "real efficiency";
- choice of parametrisation of the efficiencies;
- the normalisation of the prompt-lepton backgrounds in the determination of the efficiencies.

These uncertainties affect primarily the normalisation of the fake-lepton background. In the electron channel, the choice of control region for the "fake efficiency" also has an influence on how hard the expected spectra of $M(W_{\mathrm{H}})$ and the first two input variables of the neural network are; this translates into larger variations of the shape of the discriminant (cf. Figure F.13).

Missing transverse momentum

All uncertainties due to the reconstruction of jets and leptons are propagated coherently into the missing transverse momentum. In addition, uncertainties on the soft terms of the missing transverse momentum are taken into account. The distribution of the neural network response is sensitive to the scale of the soft terms, whereas the resolution has little effect.

[6] The $+1\sigma$ variation corresponds to a reduction (not an increase) of the efficiency by 1.5 %.

8.2.2 Generator uncertainties

The uncertainties due the MC modelling of the physics processes are evaluated by means of alternative MC samples, as described in Section 6.2. Because similar shapes are predicted for the Wt signal and the $t\bar{t}$ background, and the latter contributes more than 50 % of the events in the signal region, it is particularly important to assess the uncertainties on their modelling, as even small deviations may translate into a large effect on the measured cross section. On the other hand, the large statistical error of the W+jets sample and its large normalisation uncertainties dominate over the uncertainties due to the choice of generator and renormalisation scale for the W+jets background, so no dedicated uncertainties need to be used. Only the effect on the acceptance, not the theory cross section, is considered in this subsection.

Hadronisation model

In order to estimate the uncertainty due to the modelling of the parton shower and the hadronisation, predictions by (Powheg+)Pythia and (Powheg+)Herwig are compared, independently for the Wt and $t\bar{t}$ processes. Herwig predicts more reconstructed jets, and attributes a softer transverse momentum spectrum to the leading jet for events in the signal region[7]. The effect on the discriminant resembles that of the uncertainties due to the identification or the energy scale of b-quark jets.

NLO subtraction scheme

The uncertainty due to the NLO subtraction scheme is taken to be the difference of MC@NLO(+Herwig) with respect to Powheg(+Herwig). Apart from a difference in the normalisation, the spectrum of $M(W_{\mathrm{H}})$ falls much less rapidly for MC@NLO. The effect is most pronounced in the last bin of the discriminant, where MC@NLO predicts about 15 % more events than Powheg. Also, the predicted spectrum of $M(W_{\mathrm{L}}W_{\mathrm{H}}j_{\mathrm{B}})$ is harder in MC@NLO, which affects primarily the first bin of the neural network response.

QCD renormalisation and factorisation scale

The choice of the renormalisation and the factorisation scale is varied by a factor of two, as described in Section 6.2. A higher scale corresponds to a smaller value of α_{S}, and therefore less QCD radiation. While the effect on the input variables of the neural network is almost negligible for events with a well-reconstructed hadronic W-boson decay, the distribution of $M(W_{\mathrm{H}})$ itself is highly sensitive to the setting of the scale. The shift of the mass peak of the W boson in the $M(W_{\mathrm{H}})$ distribution due to the choice of the renormalisation scale is comparable to the shift due to all JES uncertainties combined, limiting the in-situ determination of the JES. In addition, the fraction of events with a well-reconstructed hadronic W-boson decay increases as the QCD radiation decreases.

8.2.3 Theory cross sections

The uncertainties on the inclusive cross sections are about 4 % to 6 % depending on the process (see Section 2.2). Also, the uncertainty on the luminosity directly translates into an uncertainty of 2.8 % on the overall normalisation of all simulated processes (see Equation 6.1).

[7] Without the veto against a 4$^{\mathrm{th}}$ jet, Pythia and Herwig make compatible predictions for the transverse momentum of the leading jet.

In addition, there are large uncertainties on the production of jets associated with the intermediate vector bosons. For every jet, an additional uncertainty of 24 % is assumed (see Section 2.2). Consequently, the uncertainty on the normalisation of the production of a W/Z boson in association with three jets is 42 %. On top of that, the ratio of W+jets events with heavy-flavour over light-flavour content is allowed to vary by 20 %. The size of the uncertainty is permissive enough to allow the uncertainty to be constrained by the observed data.

8.2.4 Simulation statistics

The uncertainty due to the limited size of the simulated samples is estimated by applying the *Barlow-Beeston light* method [153, 154]: for every bin of the discriminant, an independent nuisance parameter is assigned which describes the variation of the predicted event rate by its statistical error.

8.3 Fit result

Using the procedure outlined in Section 8.1, the value of the signal strength parameter is measured to be:

$$\hat{\mu} = 1.29^{+0.37}_{-0.34}\,, \tag{8.12}$$

consistent with the Standard Model expectation ($\mu = 1$). The error includes the statistical error of the observed data as well as all systematic uncertainties. The measured signal strength translates into an observed cross section of:

$$\sigma_{Wt}^{\text{obs}} = (29 \pm 8)\,\text{pb}\,. \tag{8.13}$$

Figure 8.3 shows the confidence intervals on the nuisance parameters extracted from the observed data, together with the impact of each source of systematic uncertainty. The impact due to the limited size of the signal and background samples, whose 45 components are not shown in the figure, is 17 %. Besides that, the dominant uncertainties are due to the choice of renormalisation scale and the jet energy scale, whose impact is 12 % summed over all components. At the other end of the spectrum, the reconstruction of leptons and the estimate of the fake-lepton background are found to have little effect on the result, as expected.

The intervals as well as the impacts shown in Figure 8.3 are computed using scans of the likelihood function. Figure 8.5 (left) illustrates the similarity[8] of the error contours produced by each of the two methods in the plane of the signal strength parameter and one of the nuisance parameters. The impact can be visualised as the distance in $\hat{\mu}$ between the leftmost (rightmost) point and the centre of the contour. The expected uncertainty, obtained from the fit of the model to the Asimov dataset, is 38 %; the corresponding plot of the nuisance parameters can be found in Figure G.2.

Many nuisance parameters are not significantly constrained by the fit, i.e. their error estimate is almost as large as the initial uncertainty ($\pm 1\sigma$). This is the expected behaviour for parameters with a small effect on the discriminant. The best-fit values of such parameters are expected to be close to 0 because (almost) no information can be extracted from the data that would make the value shift away from 0. For a few parameters, a significant reduction of the uncertainty is found. In principle, the value of such a parameter is expected to fluctuate by $(1 - (\Delta\hat{\theta})^2)^{1/2}$. In practice, the values fluctuate less, primarily because many uncertainties are modelled in a conservative way. The significantly constrained parameters will be discussed in the following.

[8] Note however that this similarity does not always extend to a higher number of standard deviations.

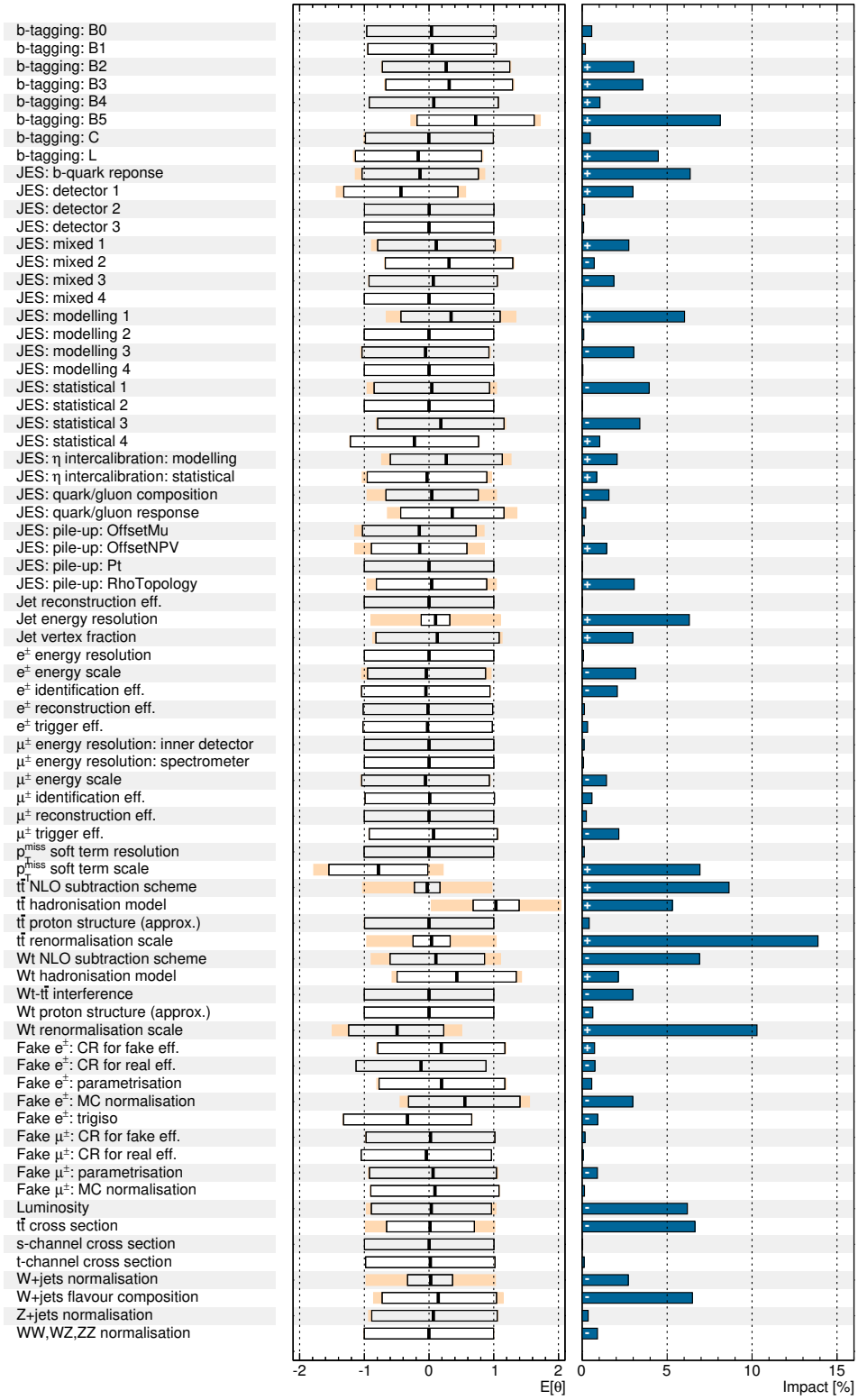

Figure 8.3: The measured values of the nuisance parameters after fitting the model to the observed data. Their errors are displayed as boxes. As a visual aid to find constrained parameters, each box is placed on a light red background, which indicates the $\pm 1\sigma$ range about the measured value. The right pane shows the impact, that is the absolute contribution of each source of systematic uncertainty to the total uncertainty on the measured signal strength. The sign indicates whether the measured signal strength has a positive or a negative correlation to the nuisance parameter. See Table G.1 for the numbers.

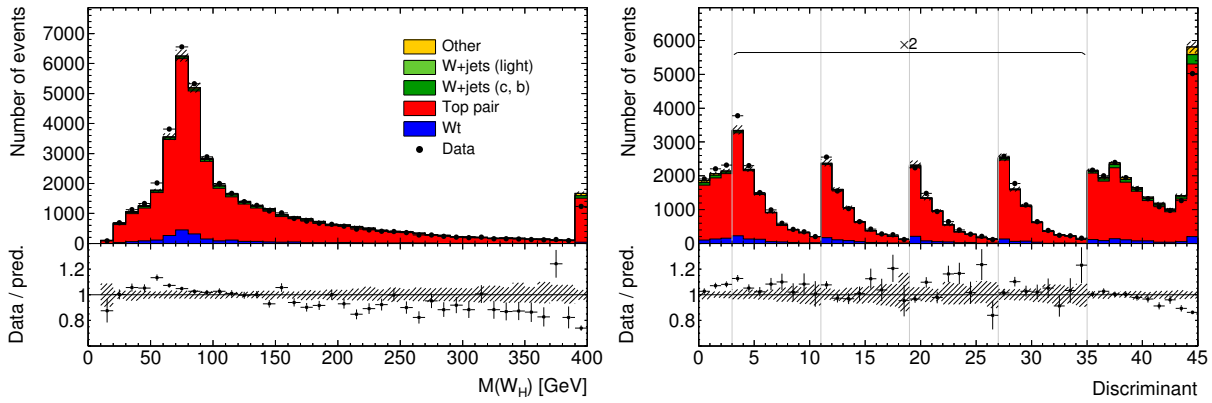

Figure 8.4: Distributions of $M(W_\mathrm{H})$ and the discriminant in the validation region, after a conditional likelihood fit of the MC@NLO-based model to the observed data, where the nuisance parameter for the NLO subtraction scheme is kept fixed.

One parameter in Figure 8.3 is particularly striking because the nominal value, 0, is well outside of the confidence interval, 1.03 ± 0.36: the choice of the hadronisation model for top-quark pair production. Apparently, Herwig is favoured significantly over the Pythia. In order to verify that this does not cause a significant bias in the measured signal strength, the measurement is repeated using[9] the Pythia hadronisation model. With the latter, a 9 % smaller signal strength is obtained. This is small compared to the total uncertainty, so the potential bias can be safely tolerated. Care should be taken not to interpret this as direct support for the hadronisation model used in Herwig: the Pythia and Herwig samples differ in other points as well, for example in the decays of heavy-flavour particles [155]. Incidentally, recent measurements of the top-quark pair production by CMS [156] and ATLAS [157] found that Powheg+Herwig provides a better description of the transverse momentum spectrum of the top quark than Powheg+Pythia.

The jet energy resolution is one of the parameters with the strongest constraint from the data, the uncertainty being reduced to 1/4 of the initial uncertainty. This comes as no surprise considering the sensitivity of the discriminant to the W-boson mass peak, and the conservativeness of the initial uncertainty. The narrow confidence interval on the nuisance parameter covers 0, as expected for a conservative uncertainty.

The situation is similar for the NLO subtraction scheme. Compared to Powheg, MC@NLO predicts a very different slope for the combinatorial background in the spectrum of $M(W_\mathrm{H})$, which is strongly disfavoured by the data. In order to investigate this more closely, an MC@NLO-based model was created by fixing the nuisance parameter for the NLO subtraction scheme to 1, and fitting to the data. Figure 8.4 clearly demonstrates that this MC@NLO-based model is unable to describe the observed spectrum of $M(W_\mathrm{H})$, and that it extrapolates badly into the validation region. The pivotal role of the slope of the combinatorial background for the constraint on the nuisance parameter is checked using another likelihood fit. The nuisance parameter is not kept fixed anymore, but the last bin of the discriminant, where the relative deviation between Powheg and MC@NLO is largest, is removed. This relaxes the constraint to 30 % with negligible changes of the impact, the signal strength and its total uncertainty change. As with the comparison between Pythia and Herwig, the NLO subtraction scheme is of course not the only difference between the MC@NLO and Powheg.

[9] Technically the conditional likelihood function with the corresponding nuisance parameter fixed at 0 is fitted to the observed data.

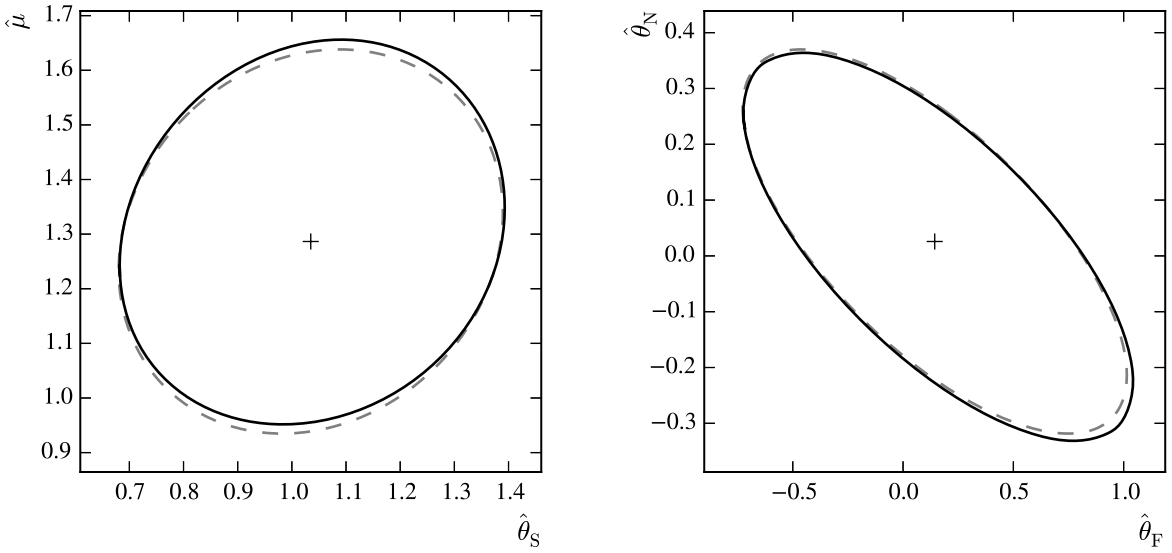

Figure 8.5: Error contours of the signal strength and the hadronisation model for $t\bar{t}$ events (left), and the parameter estimates for the W+jets normalisation and flavour composition (right), corresponding to one standard deviation. The contours obtained from a scan of the likelihood function (solid lines) are compared to the ellipses calculated from the covariance matrix of the maximum-likelihood fit (dashed lines).

It is important to also bear in mind the correlations between the measurements of the nuisance parameters. These correlations appear when multiple parameters are associated with similar effects on the discriminant. In the presence of strong correlations, constraints may exist that are not obvious from the uncertainties of the individual nuisance parameters [151]. This is exactly what happens when trying to determine the W+jets normalisation and the flavour composition. The uncertainty on the number of W+jets events is reduced from almost 50 % to about 10 % (Table 8.2), but this is not fully reflected in the constraints on the individual uncertainties (Figure 8.3). The number can be understood using error propagation, taking the constraints *and* the correlation of the measurements into account:

$$\sqrt{\left(\delta_{\mathrm{N}}\,\Delta\hat{\theta}_{\mathrm{N}}\right)^2 + \left(\delta_{\mathrm{F}}\,\Delta\hat{\theta}_{\mathrm{F}}\right)^2 + 2\,\delta_{\mathrm{N}}\delta_{\mathrm{F}}\,\Delta\hat{\theta}_{\mathrm{N}}\Delta\hat{\theta}_{\mathrm{F}}\,\hat{\rho}_{\mathrm{N,F}}}\,, \tag{8.14}$$

where the labels N and F identify the normalisation and the flavour composition components. δ_{N} is the initial uncertainty on the number of selected W+jets events due to the W+jets normalisation (see Section 8.2). δ_{F} is the change in the number of events when increasing the number of W+jets events with heavy-flavour content by 20 % while reducing the number of W+jets events with light-flavour content by 20 % (cf. Section 8.2 and Table 6.1). The $\Delta\hat{\theta}$ can be taken directly from Table G.1, and the correlation coefficient, denoted by $\hat{\rho}_{\mathrm{N,F}}$, from Figure G.1. To summarise:

$$\delta_{\mathrm{N}} = 0.42 \qquad \Delta\hat{\theta}_{\mathrm{N}} = 0.35$$
$$\delta_{\mathrm{F}} = 0.14 \qquad \Delta\hat{\theta}_{\mathrm{F}} = 0.87$$

$$\hat{\rho}_{\mathrm{N,F}} = 0.70\,.$$

The hidden constraint also reveals itself as a reduction of the length of the minor axis of the error contour of the two nuisance parameters, visualised in Figure 8.5 (right). Given that the initial uncertainty on the

W+jets normalisation is highly conservative, and that W+jets events can be separated well from Wt and $t\bar{t}$ events, the constraint is very reasonable.

A similar argument applies for the JES parameters. Several of these parameters have a slightly reduced uncertainty, and their estimates are correlated among each other, indicating that the overall JES is constrained by the data. This is expected because the discriminant is sensitive to the position of the W-boson mass peak in the spectrum of $M(W_{\mathrm{H}})$. Since variations of the JES produce a similar shift in the peak as variations of the renormalisation scale in the parton shower, the estimates of the JES and the renormalisation scale are correlated. Varying the latter by more than 35 % of the initial uncertainty is disfavoured by the data.

Naturally, the estimate of the $t\bar{t}$ cross section tends to be correlated with other parameters that have a strong effect on the number of selected $t\bar{t}$ events. Other correlations than the ones already mentioned are small enough to be ignored in the discussion of the fit result. As shown in Figure G.1, most coefficients are close to 0, only a few are as large as 25 %. Note that correlations are taken into account in the definition of the impact: given that the asymptotic approximation holds, the impact of two or more uncertainties combined is equal to the quadratic sum of the impacts of the individual uncertainties.

In summary, the nuisance parameters behave reasonably in the fit to the observed data, and the few stronger constraints can be motivated well. A noticeable exception is that the predictions using Herwig for the parton shower and hadronisation are significantly preferred over those using Pythia. This does not pose a problem, because the resulting potential bias is small compared to the uncertainty of the measurement.

However, the checks presented in this section alone are not sufficient to judge whether the model still describes the observed data within the reduced uncertainties – this will be addressed in the following section.

8.4 Advanced modelling checks

In order to confirm the validity of the post-fit model, including its adjustments on the parameters and the reduction of the uncertainties, its predictions are confronted again with the observed data. The adjustments lead to an updated *post-fit* model expectation for the event rate, given by $\nu(\hat{\mu}, \hat{\vec{\theta}})$. The uncertainty

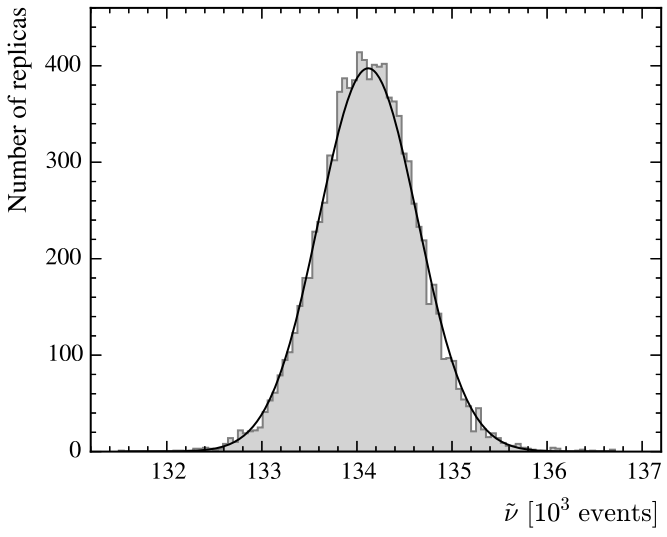

Figure 8.6: Distribution of the expected number of events, $\tilde{\nu}$, in the 10 000 replicas of the post-fit model. It approximately follows a normal distribution, which is also shown in the figure. The width of normal distribution quantifies the uncertainty, $\Delta\nu$, on the expected number of events after the fit. The interval $\nu \pm \Delta\nu$ is highlighted in the figure.

Process	Signal region	$t\bar{t}$ region	W+jets region
Wt	8 400 ± 2 200	1 400 ± 600	4 400 ± 1 300
$t\bar{t}$	75 300 ± 2 300	37 700 ± 2 400	29 500 ± 1 700
t, t-channel	4 230 ± 220	600 ± 40	1 610 ± 100
t, s-channel	313 ± 16	64 ± 4	99 ± 7
W+jets, heavy flavour	32 300 ± 3 400	1 300 ± 500	114 000 ± 28 000
W+jets, light	5 800 ± 1 900	29 ± 18	220 000 ± 50 000
Z+jets	3 900 ± 1 600	190 ± 90	33 000 ± 13 000
$WW/WZ/ZZ$	650 ± 270	25 ± 12	4 800 ± 2 000
Fake leptons	3 300 ± 1 600	–	23 000 ± 7 000
Total background	125 800 ± 2 200	39 900 ± 2 500	420 000 ± 50 000
Total model	134 200 ± 500	41 300 ± 2 400	430 000 ± 50 000
Observed	134 216	41 480	422 185

Table 8.2: Expected and observed numbers of events in the signal and the validation regions, after fitting to the observed data. The quoted errors are the standard errors due to all systematic uncertainties (cf. Section 8.2). The rounding of the numbers follows the recommendation of the Particle Data Group [7].

of the latter can be estimated using a large number (here 10 000) of replicas of the model. Each replica is generated by drawing random numbers for the signal strength and the nuisance parameters from the multivariate distribution defined by the covariance matrix of the fit result, ensuring that all linear correlations among the parameter estimates are taken into account. Some uncertainties do not apply to the discriminant in the signal region, as for example the uncertainty due to the extrapolation from the production of a W boson in association with 3 jets to 4 jets. The corresponding parameters are randomised according to their initial p.d.f.s, so they follow a normal distribution with width 1. For every replica, the expected event rate, $\tilde{\nu}$ is calculated using the randomised parameters. Finally, the uncertainty on ν is extracted from the distribution of the $\tilde{\nu}$, which is shown in Figure 8.6.

Occasionally, it can be useful to know the event rate expected from the model when using the initial parameter values and uncertainties. Here, the rate of the model is simply given by the nominal expectation, $\nu(1, \vec{0})$, i.e. the Wt signal is normalised to the theoretically predicted cross section for the Standard Model, and all nuisance parameters are set to zero. When generating the replica, all nuisance parameters are randomised according to their initial uncertainties. As setting $\mu = 1$ can easily bias conclusions drawn from comparisons of the pre-fit model and the observed data, in particular outside of signal-depleted validation regions, such comparisons has been generally avoided in this analysis. It is used only in Table 6.1 and Figure 6.7, as well as in the spectra shown in Chapter 7, which do not depend on the normalisation.

Table 8.2 presents the expectations for the event rates in the signal and validation regions, using the model parameters estimated from the observed data in the signal region, and compares them to the observed rates. The total number of events in the signal region is reproduced very well by the post-fit model. As a fit in the signal region has been used to improve the model, the expected rates are statistically independent of the observed rates only in the validation regions. In the latter, the expected and the observed numbers are in excellent agreement as well, indicating that the normalisations of the $t\bar{t}$ and the W+jets backgrounds are correctly determined by the fit in the signal region. Similar to the pre-fit expectation (see Table 6.1), the uncertainties for the individual processes need not add up to the uncertainty on the total expected rate due to correlations between the uncertainties. As an example,

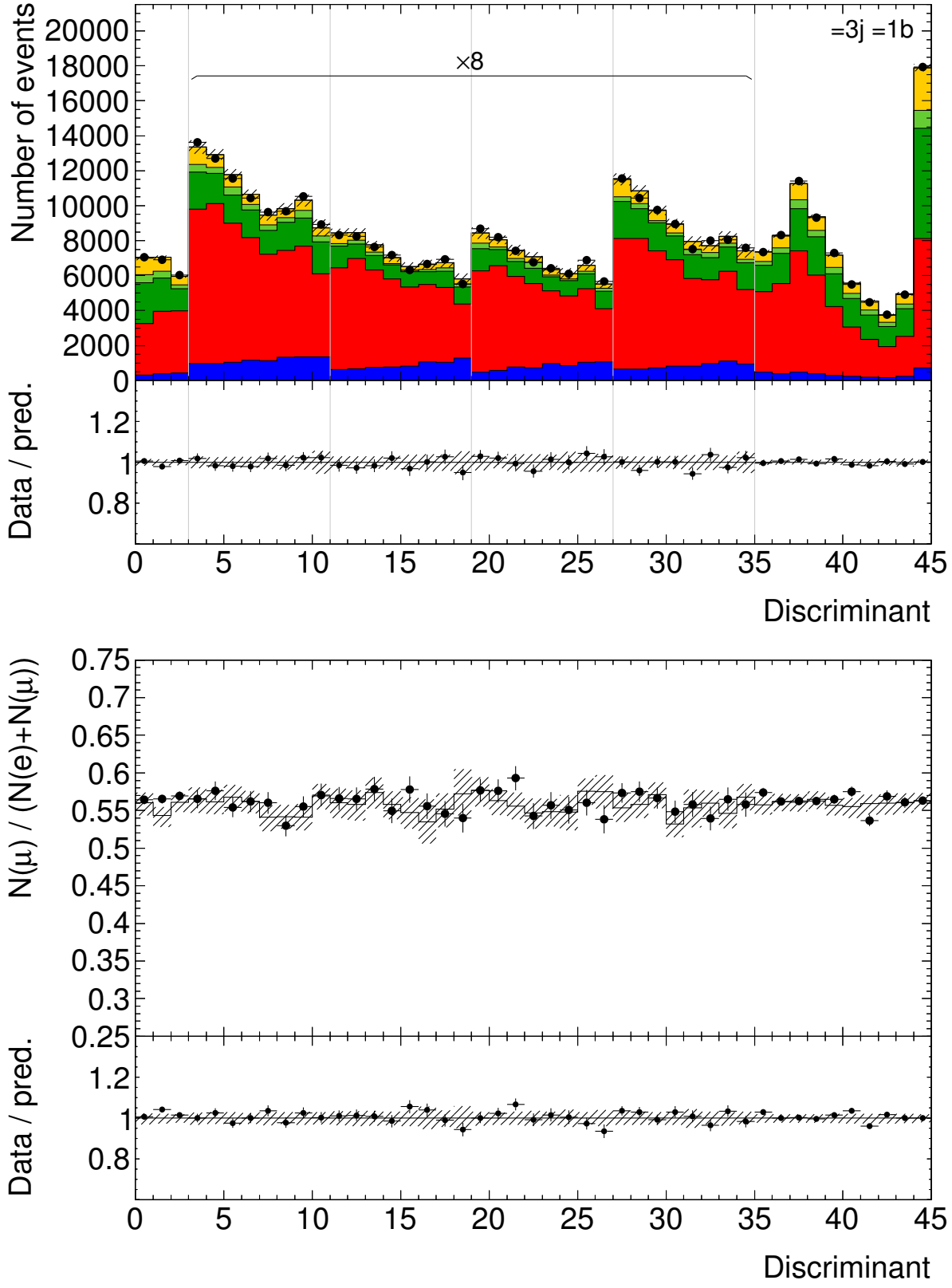

Figure 8.7: Post-fit distribution of the discriminant in the signal region. In the bins where the neural network is used, the numbers of events are scaled by a factor of 8 for better visibility. Also shown are the fraction of events with a muon as opposed to an electron in each bin of the discriminant, presented in the manner explained in Section 6.4. The expectations and their uncertainties are based on the improved model parameters, and compared to the observed data.

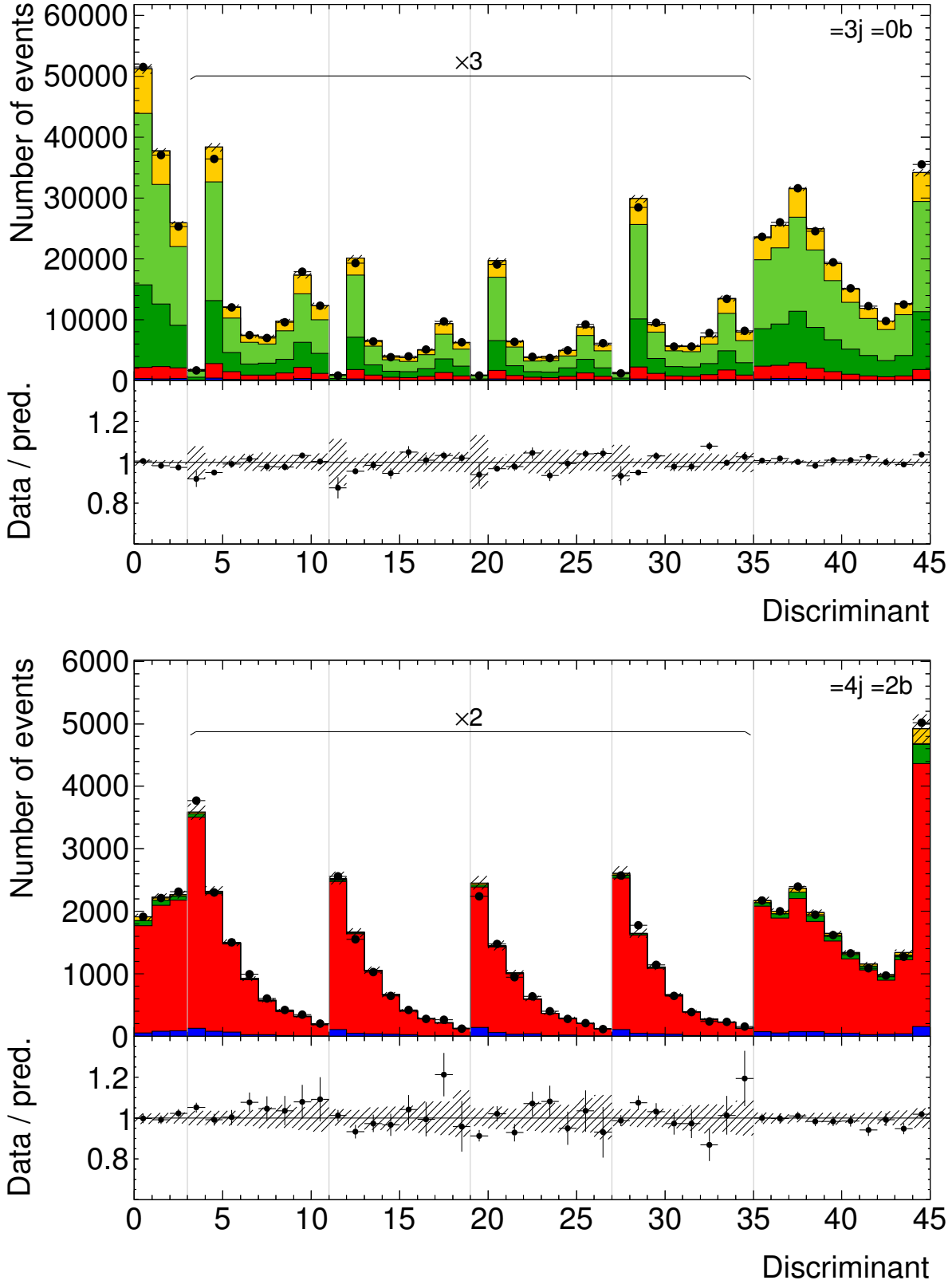

Figure 8.8: Post-fit distributions of the discriminant in the validation regions. The expectations and their uncertainties are based on the improved model parameters, and compared to the observed data. In the bins where the neural network is used, the numbers of events are scaled by a factor for better visibility.

the estimates of the rates for the Wt and the $t\bar{t}$ events are strongly anti-correlated after the fit because their sum can be extracted very well from the observed data, while their ratio is difficult to determine. Compared to the pre-fit expectation, the background rates change only little, while their uncertainties are greatly reduced.

The expectations for distributions of variables, as presented in Section 6.4 and Chapter 7, are calculated by applying the aforementioned procedure bin by bin. In order to prevent the large uncertainties on the overall normalisation from dominating the uncertainty bands, the distributions are normalised to the observed number of events in the particular signal or validation region.[10] The thinner uncertainty bands allow much stricter checks, while the normalisation can still be compared separately in Tables 6.1 and 8.2. It matters chiefly in the validation regions, whose post-fit normalisations are not constrained to the number of observed events. The construction of the uncertainties aims for a coverage probability[11] of 68 %. In the W+jets validation region, the coverage probability may be slightly higher because the conservative uncertainty on the flavour fractions in W+jets events is not constrained strongly by the fit. In contrast, the uncertainties bands for distributions of azimuthal angles of physics objects are permissive because uncertainties on the angles are not included in the model, as they are not relevant for the fit.

While ATLAS top-quark physics analyses are moving towards this or similar strategies for the construction of uncertainty bands in control plots, the coverage probability of the bands varies strongly when comparing existing measurements of single-top quark production.[12]

Post-fit distributions of many variables are presented throughout this thesis (Sections 6.4, 7.1 and 7.4 as well as Appendix D), confirming the excellent performance of the post-fit model. Figure 8.7 and Figure 8.8 conclude this series of control plots by showing the distributions of the discriminant in the signal and the validation regions. The agreement in all regions is remarkable. The consistency between the electron and the muon channel not only proves that the two channels can be treated on the same footing but also supports the reliability of the fake-lepton estimate. While it is also reassuring to see that there are no unexpected features in the signal region that cannot be described by the model, the strongest support for the validity of the fit result comes from the comparisons of the expected and the observed distributions in the validation regions. The $t\bar{t}$ validation region in particular, where the uncertainty due to the extrapolation from the signal region is small, provides a stringent test that the main background is understood extremely well.

8.5 Hypothesis testing

Given that no experimental evidence for the production of a single top quark in association with a W boson in single-lepton events has been published so far, it is interesting to test the statistical significance of the result presented in this thesis. This is done by comparing the two hypotheses:

[10] Mathematical operations involving the event rates, such as normalising a distribution to the observed number of events, or taking the ratio of the rates in the electron and the muon channel, are performed coherently in all replicas before the error is extracted.

[11] In a large number of hypothetical repetitions of the experiment, the *coverage probability* is the fraction of times that the true value is contained in the uncertainty band.

[12] Some analyses, where one source of systematic uncertainty dominates over the others, included only one important source of uncertainty in the band, which can lead to permissive (e.g. in Ref. [158]) or conservative (e.g. in Ref. [144, 159]) uncertainty bands, depending on the fit model and the nature of the uncertainty. Recently, analyses of single-top-quark production moved towards showing the post-fit uncertainties, either for all distributions (e.g. Ref. [96]) or at least for the discriminant (e.g. Ref. [95]). Conservative uncertainty bands in control distributions can suggest that the data are modelled much better than they actually are.

1. the null or background-only hypothesis, H_0, which states that the signal does not exist, and the observed data can be explained using only the background processes;

2. the alternative or signal-plus-background hypothesis, H_1, stating that the signal does exist.

The *p-value* is the probability that assuming H_0 is true, the data fluctuate away from the expectation for H_0 at least as much as the actually observed data. Extreme p-values suggest that H_0 should be rejected because it cannot explain the observed data. The p-value is commonly translated into a significance, Z, expressed in units of standard deviations:

$$Z = \Phi^{-1}(1 - p)\,, \tag{8.15}$$

where Φ^{-1} denotes the quantile (the inverse of the cumulative distribution function) of the normal distribution with a mean of 0 and a width of 1. The generally accepted threshold for claiming evidence for a hypothesised signal is 3 standard deviations, corresponding to a p-value of about 1 : 750. For a discovery claim, $Z \geq 5$ is required, equivalent to a p-value of less than 1 : 3 500 000. Although the p-value does not directly indicate whether H_1 should be considered more likely than H_0, the extreme p-values associated with discovery claims usually leads to the belief that H_1 is true unless there are severe reasons not to believe[13] H_1.

The two hypotheses are compared using a test statistic based on the profile likelihood ratio:

$$t = -2 \log \frac{L(0)}{L(\hat{\mu})} = \Lambda(0) - \Lambda(\hat{\mu})\,, \tag{8.16}$$

where the nuisance parameters are implied to minimise $\Lambda(\mu)$ for the given μ. Setting $\mu = 0$ in the numerator effectively removes the signal from the likelihood function, converting into a likelihood function for H_0. The alternative hypothesis is based on the observed signal strength.[14] The significance Z can be computed directly from t using the asymptotic formula [152]:

$$Z = \sqrt{t}\,. \tag{8.17}$$

This equation is based on the same principle as the estimate of the uncertainty on the measured cross section (depicted in Figure 8.1). However, instead of calculating by how much μ needs to moved in order to change Λ by one unit, one calculates by how much Λ changes when moving μ from $\hat{\mu}$ to 0. The observed significance is 4.2 standard deviations, equivalent to a p-value of 1 : 75 000, providing strong support for the existence of Wt production. The expected significance, computed analogously to the observed significance except that μ is assumed to be 1.0, is 3.5 standard deviations.

[13] A quantitative comparison of the probabilities of H_0 and H_1, taking their a-priori probabilities into account, can be achieved in the framework of Bayesian interference, as opposed to the frequentist interpretation that is commonly used in contemporary high-energy physics analyses. Other factors related to decision making, like the cost of making a wrong discovery claim, or the result of other measurements of single-top quark production, are not explicitly taken into account here.

[14] This is common for the experiments at the LHC, as opposed to LEP or the Tevatron, where the signal strength for the alternative was usually formulated with a fixed $\mu = 1$.

CHAPTER 9

Interpretation of the Result

The result of the measurement presented in this thesis is summarised together with other recent measurements of the production of single top quarks at the LHC in Figure 9.1. The cross section measured in this analysis is consistent with the theory predictions and the other measurements of the production of a single top quark in association with a W boson at $\sqrt{s} = 8\,\text{TeV}$. The latter were performed using events with two leptons in the final state, and are nicely complemented by the analysis presented in this thesis.

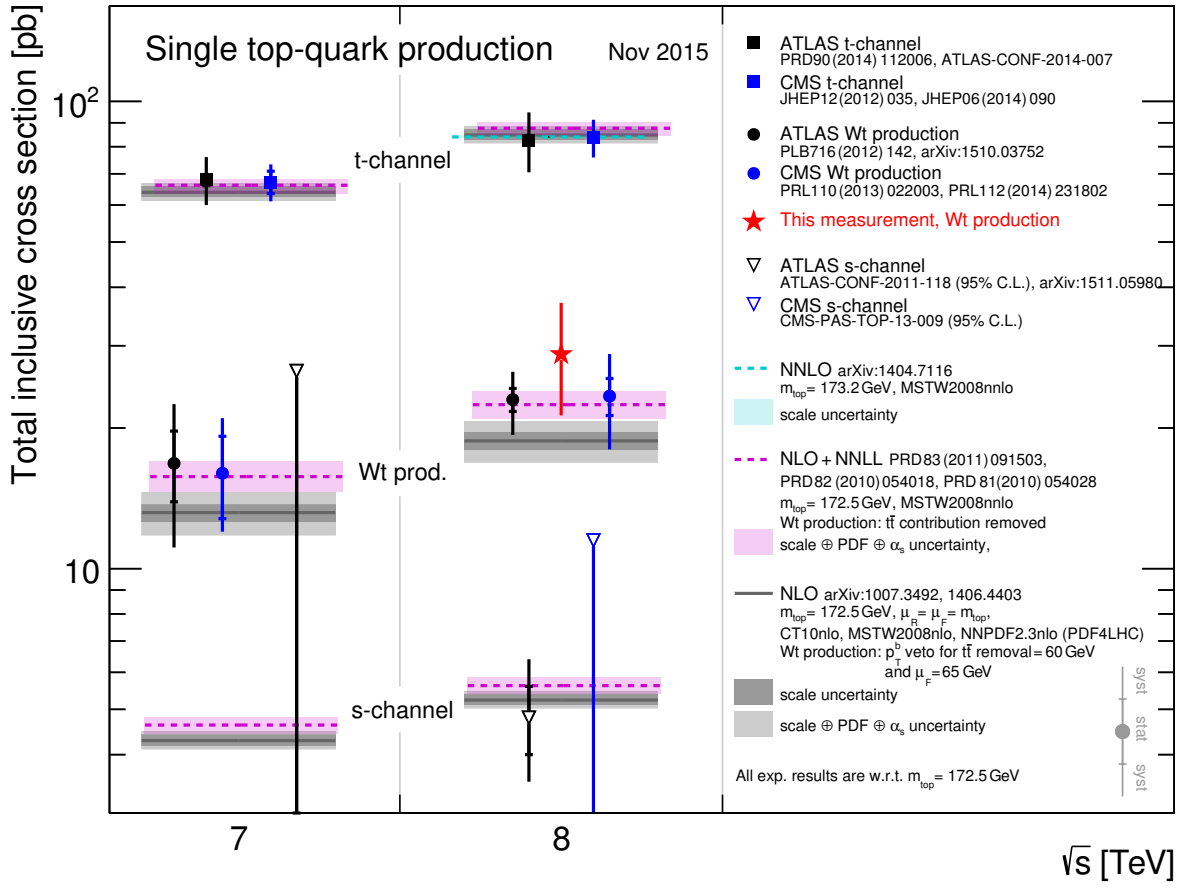

Figure 9.1: The results of this and other recent measurements of the production of single top quarks at centre-of-mass energies of 7 and 8 TeV, compared to theory predictions. The figure shows a modified version of a summary plot published by the LHC*top*WG [160, 161].

The measured cross section can be interpreted in terms of the CKM matrix element V_{tb}. Some moderate assumptions are required:

- The magnitudes of the CKM matrix elements V_{ts} and V_{td} are negligible compared to V_{tb}.
- The structure of the W–t–b coupling is not modified by physics beyond the Standard Model, i.e. the W–t–b vertex is described by a left-handed vector coupling without an additional form factor.
- Decays of the top quark into particles not described by the Standard Model can be neglected.

No assumptions on the unitarity of the CKM matrix, or the number of quark generations is required. In addition to the uncertainties on the measurements, the theoretical uncertainty on the relation between the Wt cross section and the coupling needs to be taken into account, which amounts to 1.57 pb due to the renormalisation and factorisation scales, the proton structure functions, and the mass of the top quark (± 1 GeV). The cross section measured in the present analysis then translates into:

$$|V_{tb}| = \sqrt{\frac{\sigma_{Wt}^{\text{obs}}}{\sigma_{Wt}^{\text{theo}}}} = 1.13 \pm 0.16\,. \tag{9.1}$$

The individual measurements are an important step towards a combined determination of the production cross sections in the s- and t-channel as well as the Wt mode. Their interpretation in terms of effective couplings (see Section 5.1) will allow to systematically look for new physics effects at the TeV scale even beyond the reach of direct searches [99].

CHAPTER 10

Summary

Evidence for the associated production of a single top quark and a W boson (Wt) in final states with a single charged lepton was presented, using proton–proton collisions at a centre-of-mass energy of 8 TeV recorded with the ATLAS detector at the LHC in the year 2012. An observed statistical significance of 4.2 standard deviations was found. The inclusive cross section for Wt production was measured:

$$\sigma_{Wt}^{\text{obs}} = (29 \pm 8)\,\text{pb}\,, \tag{10.1}$$

consistent with the theoretical prediction as well as the measurement of singly produced top quarks in final states with two charged leptons. The dominating systematic uncertainties are due to the Monte Carlo modelling of the top-quark pair production background.

The cross section was also interpreted in terms of the Cabibbo-Kobayashi-Maskawa matrix element V_{tb}, assuming that V_{tu} and V_{ts} are small, that the W–t–b coupling is Standard Model-like, and that decays of the top quark into any unknown particles are negligible:

$$|V_{tb}| = 1.13 \pm 0.16\,. \tag{10.2}$$

The value is in agreement with other direct and indirect determinations of V_{tb}.

As the cross sections for the production of single top quarks are sensitive to potential new physics phenomena beyond the Standard Model, their study is of great interest. A phenomenological analysis of the Wt process was performed with a strong focus on separating the production of Wt events from that of top-quark pairs, which constitute the dominant source of background. It was pointed out that the combination of two observables is particularly suited for the separation: the transverse momentum and the invariant mass of the system of the two W bosons and the b-quark jet produced in Wt events. In the experiment, only the decay products of the W bosons can be measured, so the signature for the single-lepton final state is characterised by one charged lepton, two light-flavour jets, one b-quark jet, and missing transverse momentum. It was explored how their measurable properties can be used to filter out a large fraction of background events, and to identify Wt events in remaining events. The hadronic decay of the W boson was found to be very difficult to identify, which explains why studies of Wt production using single-lepton final states have proven to be so much more challenging compared to the dilepton final states. The behaviour of the artificial neural network used in a previous measurement of Wt production at a centre-of-mass energy of 7 TeV was investigated closely. It was demonstrated that the network learned to identify hadronically decaying W bosons based on the invariant mass of the two light-flavour jets but was not flexible enough to take into account the large systematic uncertainties affecting the distribution of the invariant mass. This conclusion is not limited to neural networks but applies to other multivariate analysis techniques as well. Recalling that these techniques are just tools necessary to reduce the dimensionality of the measured final state down to a level where a binned likelihood fit can be applied, an unorthodox solution was devised: the invariant mass of the two light-flavour jets was

excluded from the dimensional reduction, and instead combined with the neural network response. This shifted the problem of identifying the W boson from the neural network to the statistical analysis. For the statistical analysis, a profile-likelihood fit was used, which had the necessary degrees of freedom to adopt to the effects of the systematic uncertainties on the distribution of the invariant mass. The profile-likelihood fit not only served to extract the signal cross section but also allowed the reduction of initially overestimated uncertainties such as the highly conservative uncertainty on the number of W+jets background events. State-of-the-art statistical methods were implemented into the analysis in order to validate the modelling of the data after the application of the newly developed strategy and the reduction of the uncertainties. Excellent modelling of the data was demonstrated not only in the signal region but also in two dedicated validation regions.

This analysis complements a recently published measurement of Wt production in the dilepton final state. Although it cannot quite compete in terms of the systematic uncertainty, it is reassuring to see that a similar result is obtained in a different final state using different analysis methods. The thesis also provides insights and techniques that are valuable for future analyses employing neural networks or similar multivariate analysis techniques in final states with a hadronic decay of a W or a Z boson. There are efforts ongoing in the ATLAS Collaboration aiming to publish the results presented here in a scientific journal.

APPENDIX A

Technical details

The analysis is based on observed data on the standard good-runs list `data12_8TeV.AllYear` using the defects `PHYS_StandardGRL_All_Good` in version `DetStatus-v61-pro14-02`. Table A.1 presents the analysed dataset containers. The numbers do not include an additional count of 5 427 486 events from the EGAMMA stream and 33 945 933 events from the MUONS stream that are part of the containers but not of a run on the good-runs list. One run, number 209024, is a very short run that does not have data in the EGAMMA stream. The number of events processed in the analysis agrees exactly with the number of events provided by the ATLAS bookkeeping system AMI.

Dataset container	Good runs	Events
EGAMMA stream		
data12_8TeV.periodA.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1517_p1562	23	43 024 225
data12_8TeV.periodB.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	70	177 088 965
data12_8TeV.periodC.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	17	48 944 245
data12_8TeV.periodD.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	45	112 911 766
data12_8TeV.periodE.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	27	85 964 025
data12_8TeV.periodG.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	15	43 683 021
data12_8TeV.periodH.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	19	52 519 263
data12_8TeV.periodI.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1562	13	36 732 329
data12_8TeV.periodJ.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1562	26	95 304 168
data12_8TeV.periodL.physics_Egamma.PhysCont.NTUP_COMMON.grp14_v01_p1562	9	30 644 241
MUONS stream		
data12_8TeV.periodA.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1517_p1562	23	43 593 454
data12_8TeV.periodB.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	70	163 080 657
data12_8TeV.periodC.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	17	49 403 586
data12_8TeV.periodD.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	46	112 478 037
data12_8TeV.periodE.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	27	80 992 129
data12_8TeV.periodG.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	15	41 026 367
data12_8TeV.periodH.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1278_p1562	19	48 875 862
data12_8TeV.periodI.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1562	13	34 530 597
data12_8TeV.periodJ.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1562	26	88 410 196
data12_8TeV.periodL.physics_Muons.PhysCont.NTUP_COMMON.grp14_v01_p1562	9	29 010 162

Table A.1: List of the dataset containers processed for observed data. The numbers include only good runs.

Tables A.2 and A.3 show the details about the individual datasets that make up the predicted signal and background samples.

Typical ATLAS analyses do not access the files containing the byte streams from the detector (or the generated particles from the simulation) directly but instead rely on the derived datasets (as listed in the tables), which are generated centrally using the Athena framework [162] (release 17). These datasets already contain the reconstructed and partially calibrated physics objects. The total size of the datasets is about 300 TB for observed data, and another 300 TB for the simulated samples used in the analysis. The datasets are essentially large tables, using the data format understood by the data-analysis library Root [163], which supports efficient [164] reading of some or all columns. Each cell stores a number or an array of numbers, where floating-point numbers are generally stored in 32-bit format (i.e. with about 6 significant digits). Since the datasets are shared by many analyses and studies, they contain many columns that are not needed for the specific analysis. Only $\mathcal{O}(1\,\%)$ of the information stored in the files is actually needed for the present analysis.

AnalysisTop-1.9 [44] is then used extract the fully corrected four-momenta of the physics objects for the analysis. The CPU-intensive process typically takes a few days when using the grid computing infrastructure of the ATLAS experiment. Events that pass a loose version of the selection criteria discussed in Section 6.3 are stored on disk, amounting to about 100 GB including all systematic variations. Later stages can process this in a matter of minutes, and together with a high degree of automation, fast turnaround times are achieved. This is particularly useful when varying the event selection as part of the cut optimisation, or when new variables are investigated. The bodies and surroundings of the loops over events are written in C++, with some parts (such as the allocation and filling of histograms) automatically generated by Python code. Python is also used in many places outside of event loops, such as plotting or configuration. Observed data, simulated data, and each of its systematic variations are processed independently of each other, and brought together in the statistical analysis. The plots shown in this thesis are made with Root (sans-serif labels) and matplotlib (serif labels), respectively.

Process	ID	Tags	Entries	$\sigma \times \mathcal{B}$	$N_{\text{eff}}/N_{\text{evt}}$
Wt, POWHEG+PYTHIA, MC12A GEANT					
any	110140	e1743 s1581 s1586 r3925 r3549 p1575	999 692	22.4	1.36
Wt, POWHEG+PYTHIA, MC12A ATLFAST2					
any	110140	e1743 a188 a171 r3549 p1575	17 448 984	22.4	23.6
single *t*, *s*-channel, POWHEG+PYTHIA, MC12A GEANT					
ℓ	110119	e1720 s1581 s1586 r3658 r3549 p1575	5 999 781	1.8	100
single *t*, *t*-channel, POWHEG+PYTHIA, MC12B GEANT					
ℓ, t	110090	e2575 s1773 s1776 r4485 r4540 p1575	4 994 481	18.4	10.3
ℓ, $\bar{t}$	110091	e2575 s1773 s1776 r4485 r4540 p1575	4 999 879	10.0	19.0
$t\bar{t}$, POWHEG+PYTHIA, MC12B GEANT					
ℓ	110404	e3151 s1773 s1776 r4485 r4540 p1575	44 953 451	137.3	13.9
$t\bar{t}$, POWHEG+PYTHIA, MC12B ATLFAST2					
ℓ	110404	e3151 a220 a205 r4540 p1575	49 980 940	137.3	15.4
W+jets, SHERPA, MC12A ATLFAST2					
$e\nu_e$, beauty	167740	e1585 a159 a171 r3549 p1562	14 997 980	154.4	2.14
$e\nu_e$, charm	167741	e1585 a159 a171 r3549 p1562	9 998 989	591.6	0.40
$e\nu_e$, light	167742	e1585 a159 a171 r3549 p1562	48 250 968	11 324.5	0.10
$\mu\nu_\mu$, beauty	167743	e1585 a159 a171 r3549 p1562	14 989 485	154.4	2.02
$\mu\nu_\mu$, charm	167744	e1585 a159 a171 r3549 p1562	9 992 484	513.1	0.46
$\mu\nu_\mu$, light	167745	e1585 a159 a171 r3549 p1562	49 781 965	11 404.8	0.10
$\tau\nu_\tau$, beauty	167746	e1585 a159 a171 r3549 p1562	14 955 982	154.4	2.66
$\tau\nu_\tau$, charm	167747	e1585 a159 a171 r3549 p1562	9 993 984	557.1	0.49
$\tau\nu_\tau$, light	167748	e1585 a159 a171 r3549 p1562	49 880 968	11 359.7	0.13
Z+jets, SHERPA, MC12A ATLFAST2					
ee, beauty	167749	e1585 a159 a171 r3549 p1575	3 829 000	34.8	2.63
ee, charm	167750	e1585 a159 a171 r3549 p1575	2 999 995	352.0	0.20
ee, light	167751	e1585 a159 a171 r3549 p1575	4 978 999	856.1	0.15
$\mu\mu$, beauty	167752	e1585 a159 a171 r3549 p1575	3 997 997	34.8	2.72
$\mu\mu$, charm	167753	e1585 a159 a171 r3549 p1575	2 997 995	352.6	0.20
$\mu\mu$, light	167754	e1585 a159 a171 r3549 p1575	4 993 999	856.2	0.15
$\tau\tau$, beauty	167755	e1585 a159 a171 r3549 p1575	3 997 994	34.7	2.95
$\tau\tau$, charm	167756	e1587 a159 a171 r3549 p1575	2 978 998	352.2	0.21
$\tau\tau$, light	167757	e1587 a159 a171 r3549 p1575	4 814 999	856.3	0.15
WW, *ZZ*, *WZ*, HERWIG, MC12A GEANT					
ℓ, *WW*	105985	e1576 s1499 s1504 r3658 r3549 p1575	2 499 890	20.9	3.6
ℓ, *ZZ*	105986	e1576 s1499 s1504 r3658 r3549 p1575	245 000	1.5	4.9
ℓ, *WZ*	105987	e1576 s1499 s1504 r3658 r3549 p1575	999 998	7.0	4.4

Table A.2: Overview of the simulated datasets, excluding those that are used exclusively for systematics.

Process	ID	Tags	Entries	$\sigma \times \mathcal{B}$	$N_{\text{eff}}/N_{\text{evt}}$
Wt diagram subtraction, Powheg+Pythia, mc12a Geant					
any	110142	e1743 s1581 s1586 r3925 r3549 p1575	994 894	22.4	1.32
Wt, Herwig, mc12b AtlFast2					
any	110144	e1743 a220 a205 r4540 p1575	9 994 488	22.4	18.9
Wt, MC@NLO+Herwig, mc12a AtlFast2					
any	108346	e1525 a159 a171 r3549 p1575	4 996 492	22.4	5.3
Wt, $\mu = 2.0\mu_0$, Powheg+Pythia, mc12b AtlFast2					
any	110046	e2727 a220 a205 r4540 p1575	999 999	22.4	1.89
Wt, $\mu = 0.5\mu_0$, Powheg+Pythia, mc12b AtlFast2					
any	110051	e2727 a220 a205 r4540 p1575	997 999	22.4	1.89
$t\bar{t}$, hdamp=∞, Powheg+Pythia, mc12a AtlFast2					
ℓ	117050	e1727 a188 a171 r3549 p1575	49 955 949	137.3	11.0
$t\bar{t}$, Herwig, mc12b AtlFast2					
ℓ	105860	e1576 a159 a222 r4540 p1575	27 452 590	137.3	6.0
$t\bar{t}$, MC@NLO+Herwig, mc12a AtlFast2					
ℓ	105200	e1513 a159 a171 r3549 p1575	27 796 962	137.3	3.7
$t\bar{t}$, $\mu = 2.0\mu_0$, Alpgen+Pythia, mc12b AtlFast2					
ℓ	201030	e2356 a188 a222 r4540 p1575	2 643 997	7.2	11.1
ℓ	201031	e2356 a188 a222 r4540 p1575	3 028 498	8.2	11.1
ℓ	201032	e2356 a188 a222 r4540 p1575	2 176 000	5.9	11.1
ℓ	201033	e2356 a188 a222 r4540 p1575	991 997	3.1	9.8
ℓ	201034	e2356 a188 a222 r4540 p1575	626 998	2.1	8.9
ℓ	201230	e2356 a188 a222 r4540 p1575	5 336 998	30.0	5.4
ℓ	201231	e2356 a188 a222 r4540 p1575	6 183 996	34.3	5.4
ℓ	201232	e2356 a188 a222 r4540 p1575	4 452 994	24.8	5.44
ℓ	201233	e2356 a188 a222 r4540 p1575	1 967 495	12.8	4.6
ℓ	201234	e2499 a188 a222 r4540 p1575	1 530 499	8.8	5.3
$t\bar{t}$, $\mu = 0.5\mu_0$, Alpgen+Pythia, mc12b AtlFast2					
ℓ	201040	e2356 a188 a222 r4540 p1575	3 489 998	9.5	11.1
ℓ	201041	e2356 a188 a222 r4540 p1575	3 296 999	8.8	11.3
ℓ	201042	e2356 a188 a222 r4540 p1575	1 890 999	5.1	11.3
ℓ	201043	e2356 a188 a222 r4540 p1575	842 999	2.1	12.0
ℓ	201044	e2356 a188 a222 r4540 p1575	495 999	1.1	14.2
ℓ	201240	e2356 a188 a222 r4540 p1575	7 097 992	39.6	5.4
ℓ	201241	e2356 a188 a222 r4540 p1575	6 672 995	36.7	5.5
ℓ	201242	e2356 a188 a222 r4540 p1575	3 887 998	21.2	5.6
ℓ	201243	e2356 a188 a222 r4540 p1575	1 678 497	8.9	5.8
ℓ	201244	e2356 a188 a222 r4540 p1575	894 999	4.4	6.1

Table A.3: Overview of the simulated datasets used exclusively for systematics.

APPENDIX B

Finite width effects in Wt

As introduced in Section 6.2, the top quark and the associated W boson are generated on-shell, completely neglecting their natural width. This leads to spurious differences between Wt and $t\bar{t}$ events in observables constructed from W bosons and jets at the parton level. While the detector resolution washes out any differences between a delta distribution and a Breit-Wigner distribution near their cores, the differences in their tails (i.e. the fact that the Breit-Wigner has one while the delta distribution does not) may have unforeseen consequences for the measurement.

The impact of this imperfection can actually be estimated from the $t\bar{t}$ simulation, exploiting the similarity of their final states. The idea is to artificially narrow the width of particles in the $t\bar{t}$ simulation, and check the effect. This can be achieved by a reweighting in the four-momentum magnitude of each particle from a Breit-Wigner distribution with width Γ to a Breit-Wigner distribution with a smaller width Γ'. Here, a slightly simpler recipe is pursued: calculate the virtuality, λ, of the particle, which is defined as $\lambda = p^2 - m^2$, where p is the four-momentum vector of the particle, and m its on-shell mass. If $|\lambda|$ is above a cut-off, a weight of 0 is assigned, otherwise a weight of 1. A cut-off of 0.1 GeV^2 is chosen. The result does not change qualitatively for any reasonable (not much larger than Γ) value of the cut-off.

Because the $t\bar{t}$ process has two top quarks, there is an ambiguity as to which of the two top quarks and to which the two W bosons the cut should be applied. In practice, for a $t\bar{t}$ event to mimic a Wt event, one of the two b quarks from the decay of the top quark needs to have small transverse momentum. This b quark is ignored, and its "sibling", the W boson, assumes the role of the prompt W boson. The top quark is then defined unambiguously, too.

Figure B.1 presents the effect on the variable that is found to be the most sensitive to the narrow-width approximation. The difference in normalisation, not shown in the plot, is less than about 0.2 %. Effects on other observables, in particular those relevant for the separation of Wt and $t\bar{t}$, but also observables like $p_{\mathrm{T}}(\ell)$, are found to be well below 1 %. It is therefore concluded that the effect is negligible given the precision of the analysis.

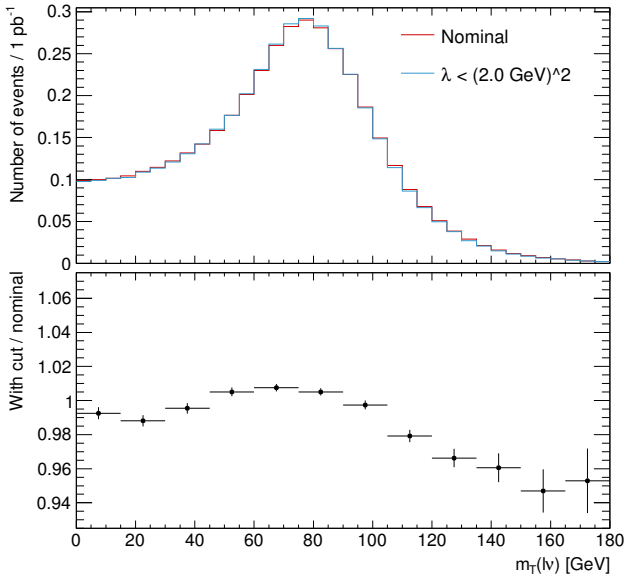

Figure B.1: Distributions of the reconstructed transverse mass of the leptonically decaying W boson, $M_{\mathrm{T}}(\ell\nu)$, showing the estimated impact of the missing momentum reshuffling in the Wt events. A cut is applied on the virtualities, λ, of the W boson and the top quark. The distributions are obtained from the $t\bar{t}$ sample with exactly one lepton in the final state, each normalised to the $t\bar{t} \to \ell\nu bqqb$ cross section. The lower panel shows the ratio of the two distributions. All selection cuts for the signal region except for the $M_{\mathrm{T}}(\ell\nu)$ cut have been applied.

APPENDIX C

Selection efficiencies

	Wt	$t\bar{t}$	t, t-channel	t, s-channel	W+jets, heavy flavour	W+jets, light flavour	Z+jets	$WW/WZ/ZZ$	Data
Trigger	0.260	0.252	0.194	0.187	0.176	0.132	0.208	0.300	$\sim 10^{-7}$
Clean LAr / Tile	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
Clean vertex	0.999	1.000	0.997	0.998	0.993	0.962	0.966	0.987	0.989
Clean jets	0.999	0.999	0.998	0.999	0.997	0.998	0.998	0.998	0.999
≥ 1 lepton	0.625	0.594	0.605	0.586	0.592	0.557	0.683	0.628	0.233
$= 1$ lepton	0.891	0.896	1.000	1.000	1.000	1.000	0.664	0.871	0.953
Clean lepton	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Trigger & lepton eff.	0.957	0.957	0.954	0.955	0.968	0.967	0.968	0.955	–
≥ 1 jet	0.969	0.991	0.892	0.912	0.505	0.139	0.668	0.733	0.263
≥ 2 jet	0.807	0.920	0.518	0.634	0.281	0.176	0.161	0.430	0.206
≥ 3 jet	0.553	0.734	0.276	0.301	0.258	0.187	0.205	0.229	0.252
$p_{\mathrm{T}}^{\mathrm{miss}} \geq 30\,\mathrm{GeV}$	0.801	0.814	0.780	0.787	0.729	0.730	0.377	0.655	0.637
$M_{\mathrm{T}}(\ell\nu) \geq 50\,\mathrm{GeV}$	0.672	0.693	0.717	0.705	0.724	0.718	0.544	0.675	0.680

Table C.1: Sequential cutflow for the electron selection. The efficiencies of each cut (cf. Equation 6.9) for simulated events with at least one prompt lepton are presented in different columns. The last column shows the observed efficiencies.

	Wt	$t\bar{t}$	t, t-channel	t, s-channel	W+jets, heavy flavour	W+jets, light flavour	Z+jets	$WW/WZ/ZZ$	Data
Trigger	0.238	0.235	0.185	0.188	0.159	0.143	0.327	0.290	$\sim 10^{-7}$
Clean LAr / Tile	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
Clean vertex	0.999	1.000	0.997	0.998	0.993	0.957	0.959	0.986	0.983
Clean jets	0.999	0.999	0.998	0.999	0.997	0.998	0.997	0.998	0.999
≥ 1 lepton	0.777	0.741	0.740	0.717	0.752	0.682	0.785	0.776	0.406
$= 1$ lepton	0.896	0.900	1.000	1.000	1.000	1.000	0.419	0.850	0.915
Clean lepton	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Trigger & lepton eff.	0.995	0.996	0.996	0.998	0.998	0.999	1.001	0.998	–
≥ 1 jet	0.968	0.990	0.885	0.903	0.522	0.132	0.268	0.710	0.195
≥ 2 jet	0.808	0.920	0.514	0.623	0.287	0.176	0.200	0.408	0.225
≥ 3 jet	0.553	0.735	0.279	0.301	0.262	0.188	0.219	0.210	0.264
$p_{\mathrm{T}}^{\mathrm{miss}} \geq 30\,\mathrm{GeV}$	0.802	0.819	0.793	0.795	0.744	0.745	0.499	0.736	0.722
$M_{\mathrm{T}}(\ell\nu) \geq 50\,\mathrm{GeV}$	0.712	0.731	0.761	0.749	0.765	0.764	0.596	0.730	0.732

Table C.2: Sequential cutflow for the muon selection. The efficiencies of each cut (cf. Equation 6.9) for simulated events with at least one prompt lepton are presented in different columns. The last column shows the observed efficiencies.

APPENDIX D

Supplemental control plots

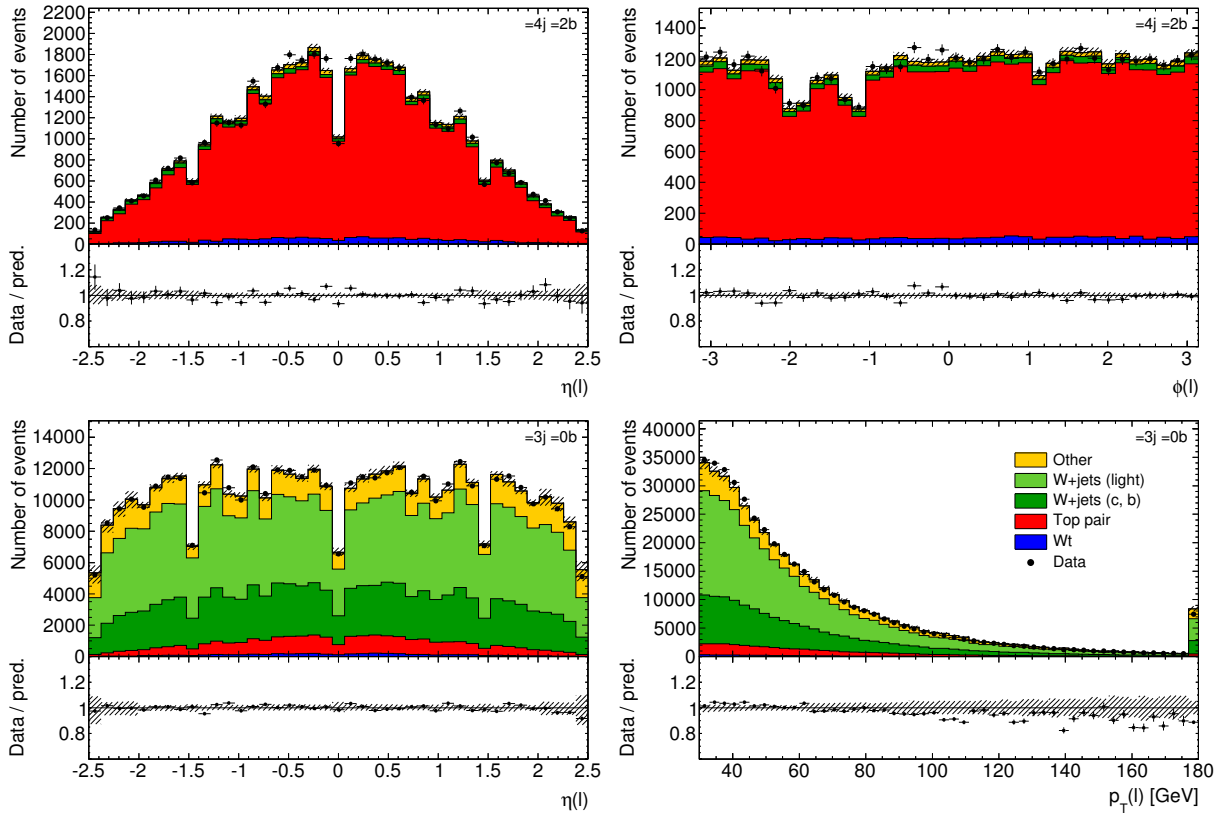

Figure D.1: Kinematic distributions of the lepton in the validation regions.

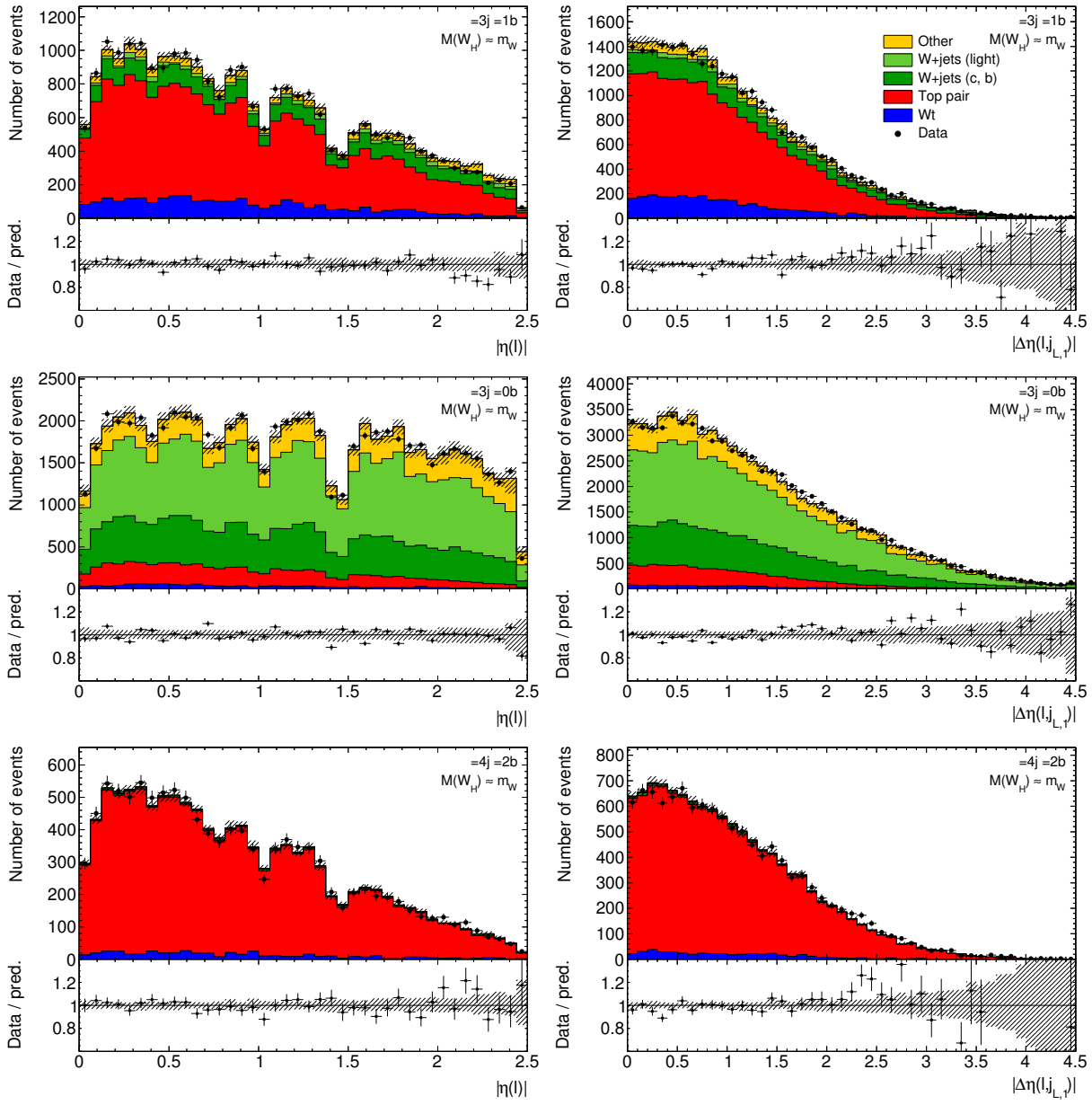

Figure D.2: Distribution of variables used as MVA inputs in the signal region (top row), the W+jets validation region (middle row), and the $t\bar{t}$ validation region (bottom row): $|\Delta\eta\ell, j_{\mathrm{L},1}|$ (left), and $|\eta(\ell)|$ (right).

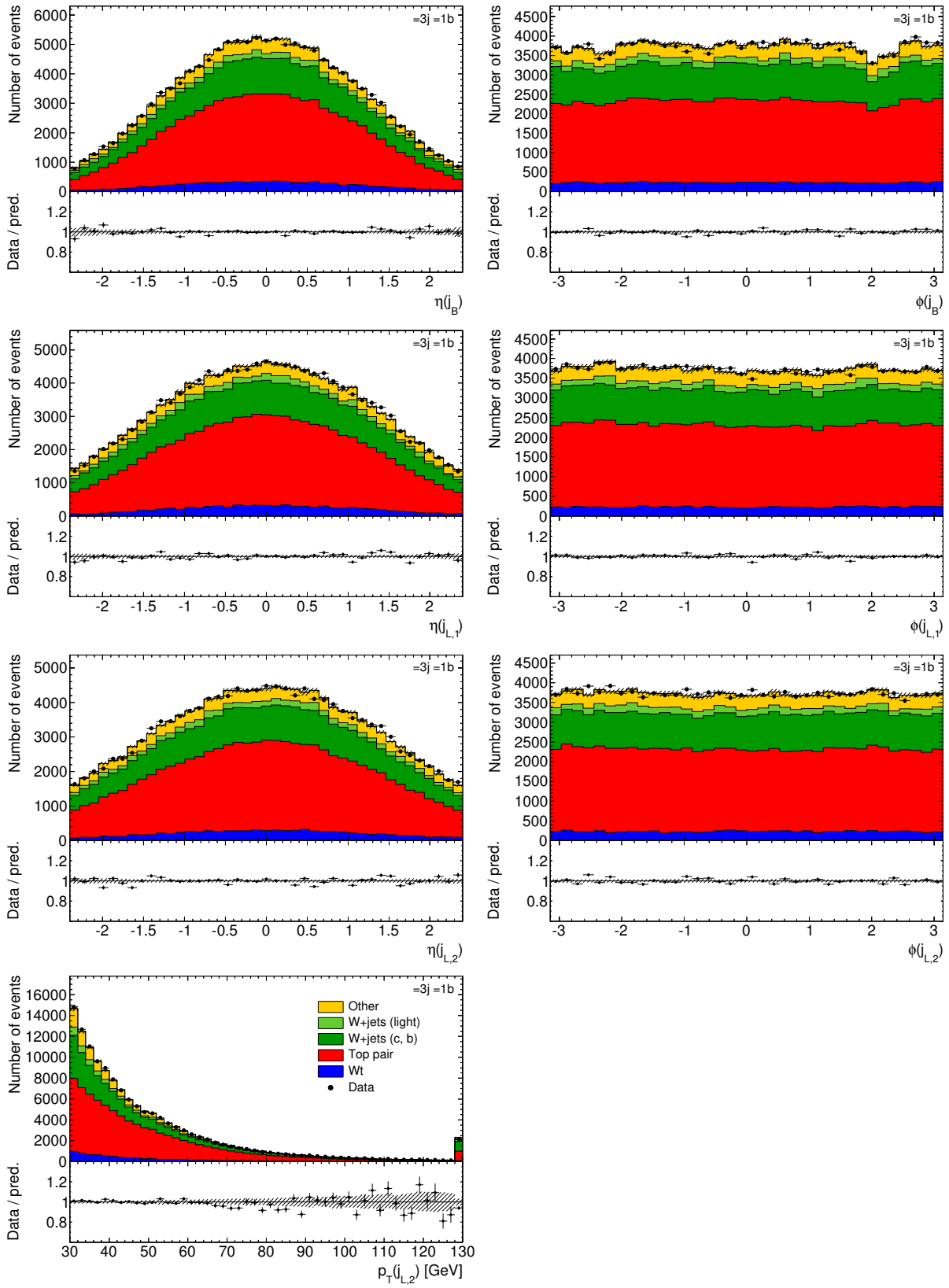

Figure D.3: Kinematic distributions of the selected jets in the signal region.

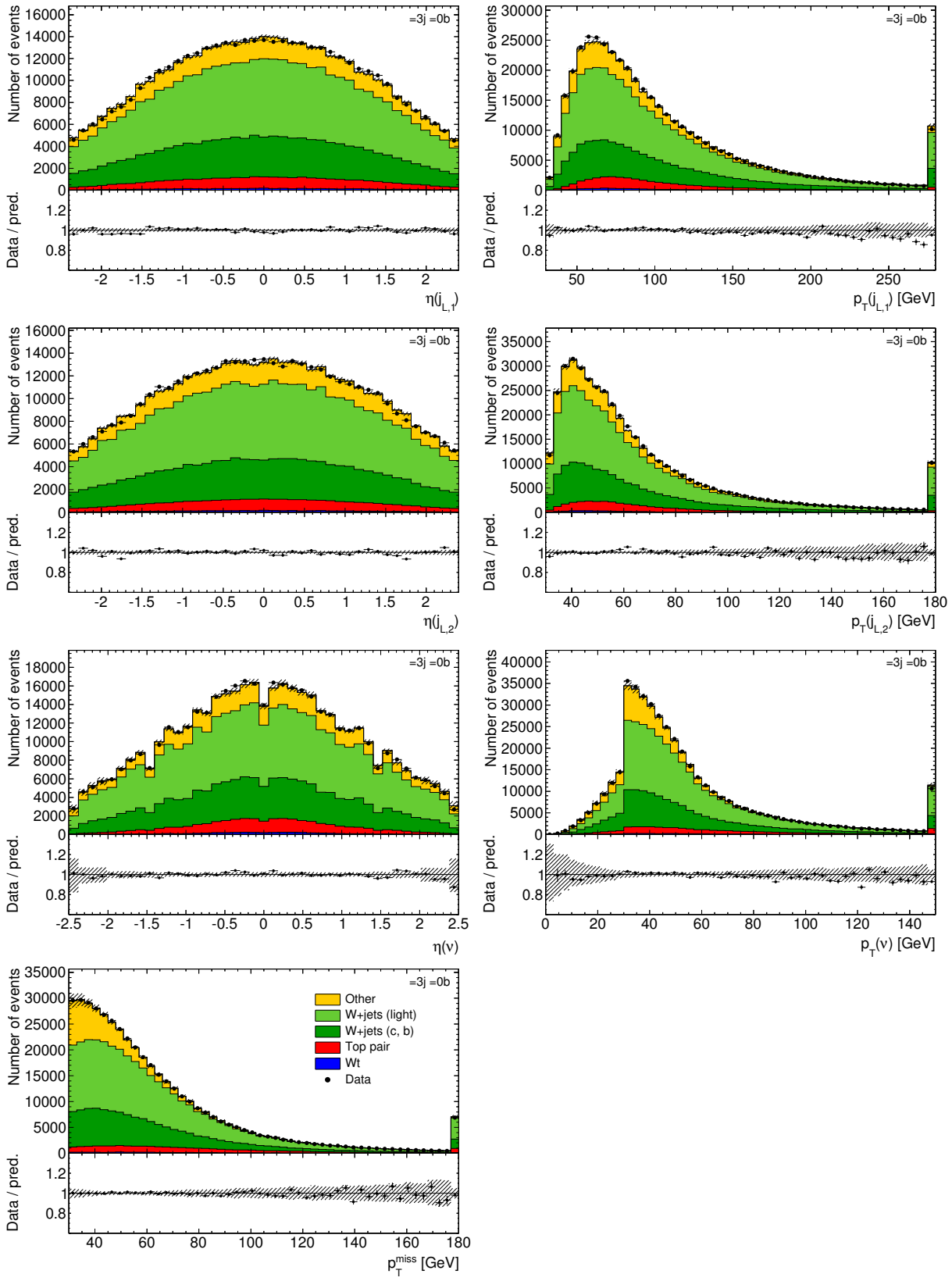

Figure D.4: Kinematic distributions of the selected jets, the neutrino, and the missing transverse momentum in the *W*+jets validation region.

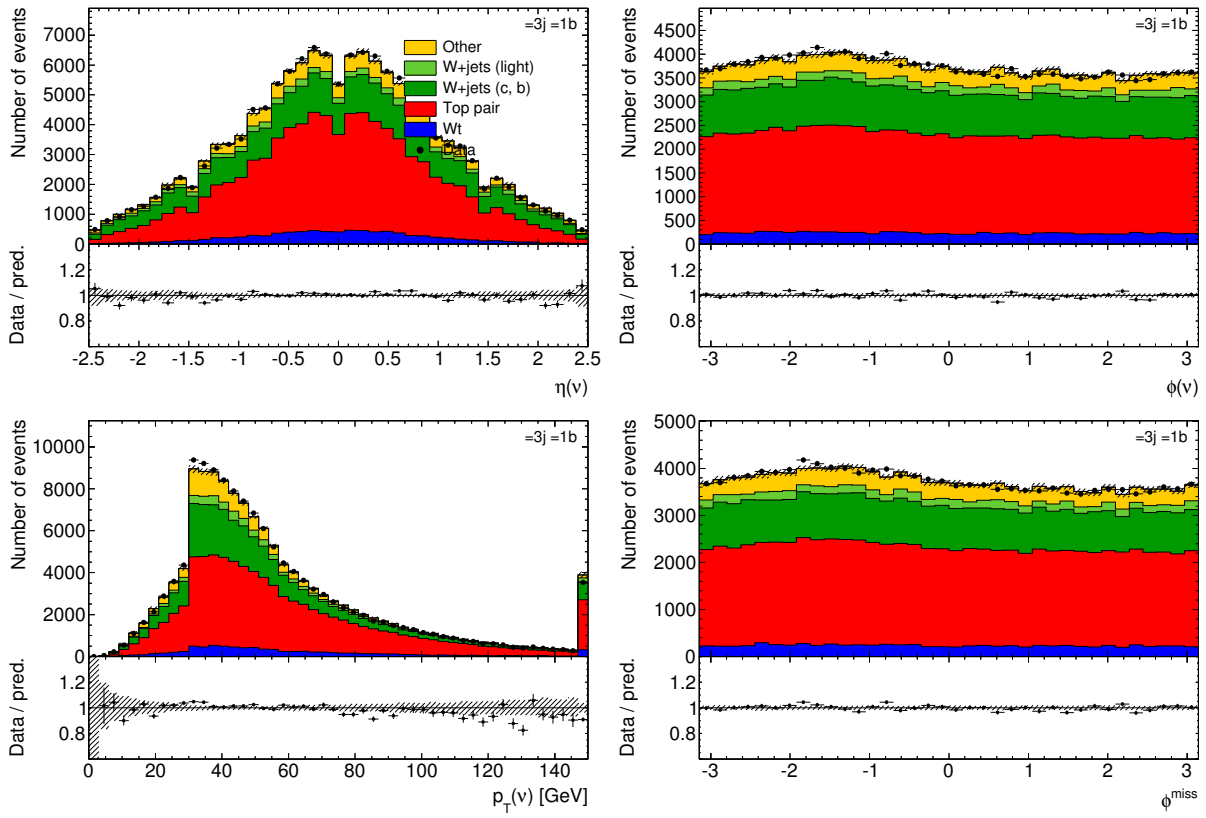

Figure D.5: Distributions of the momentum of the reconstructed neutrino as well as the azimutal angle of the missing transverse momentum in the signal region.

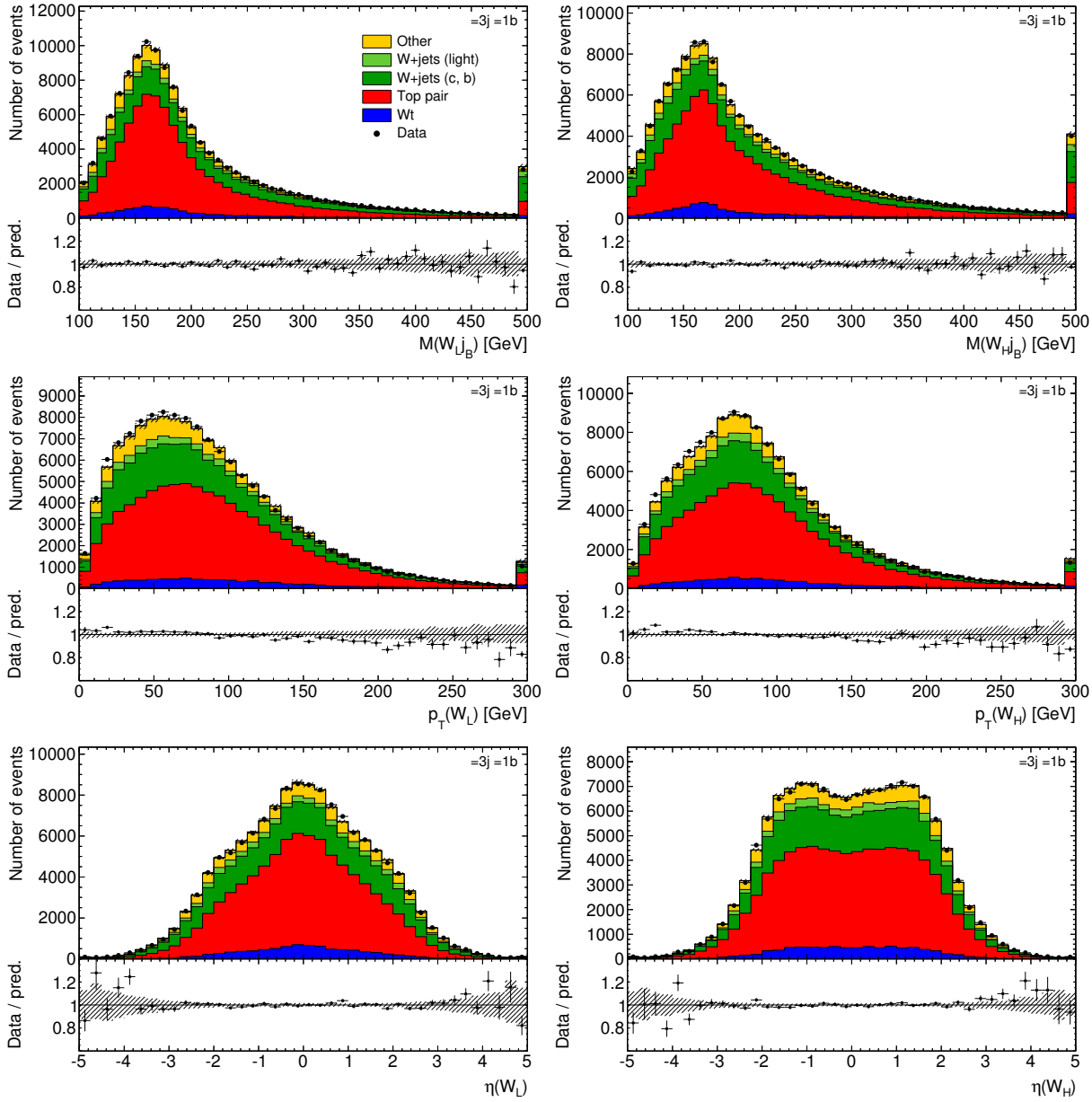

Figure D.6: Distributions of the reconstructed top-quark and the *W* boson candidates in the signal region.

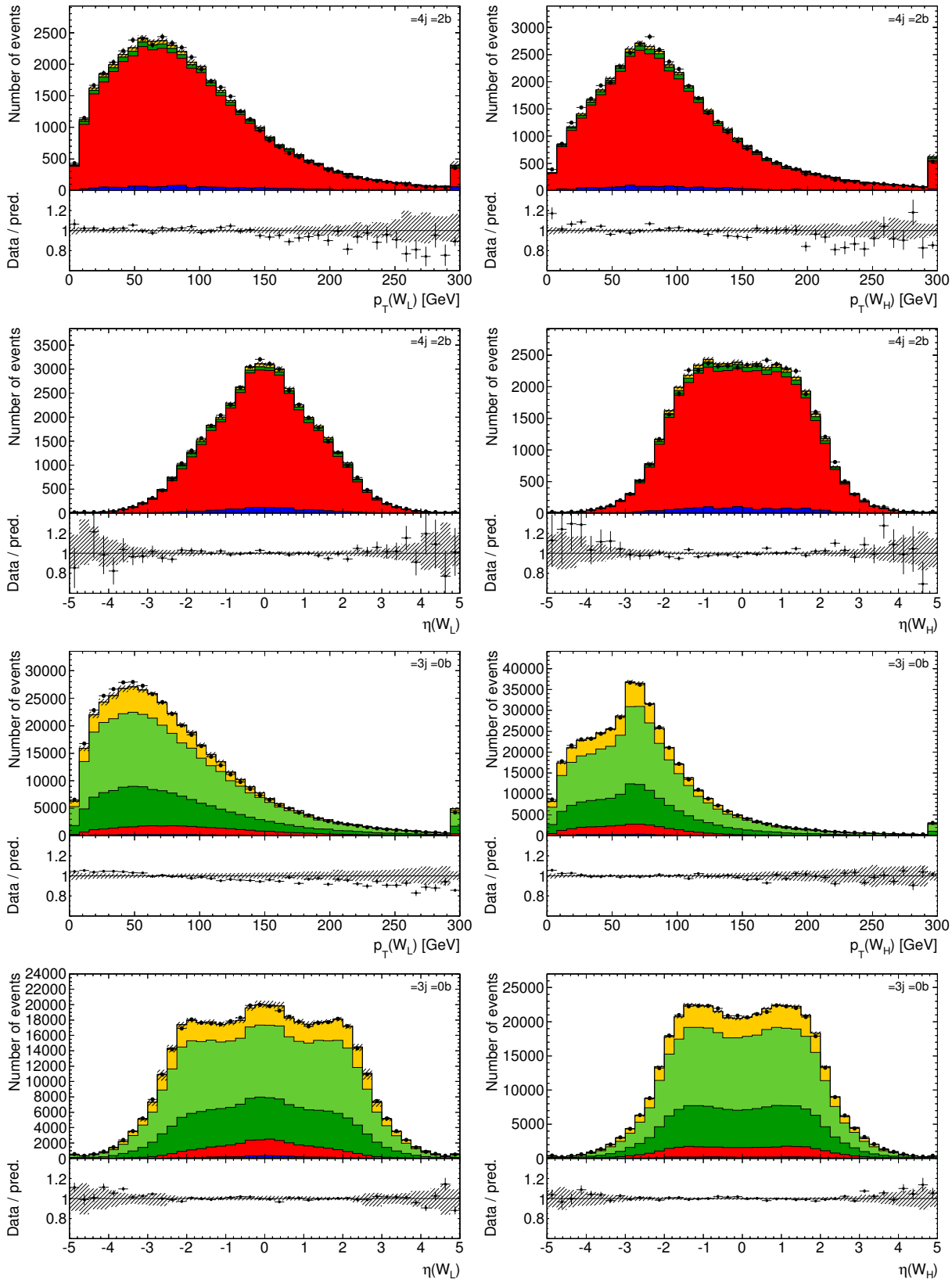

Figure D.7: Distributions of the reconstructed W boson candidates in the $t\bar{t}$ and the W+jets validation regions.

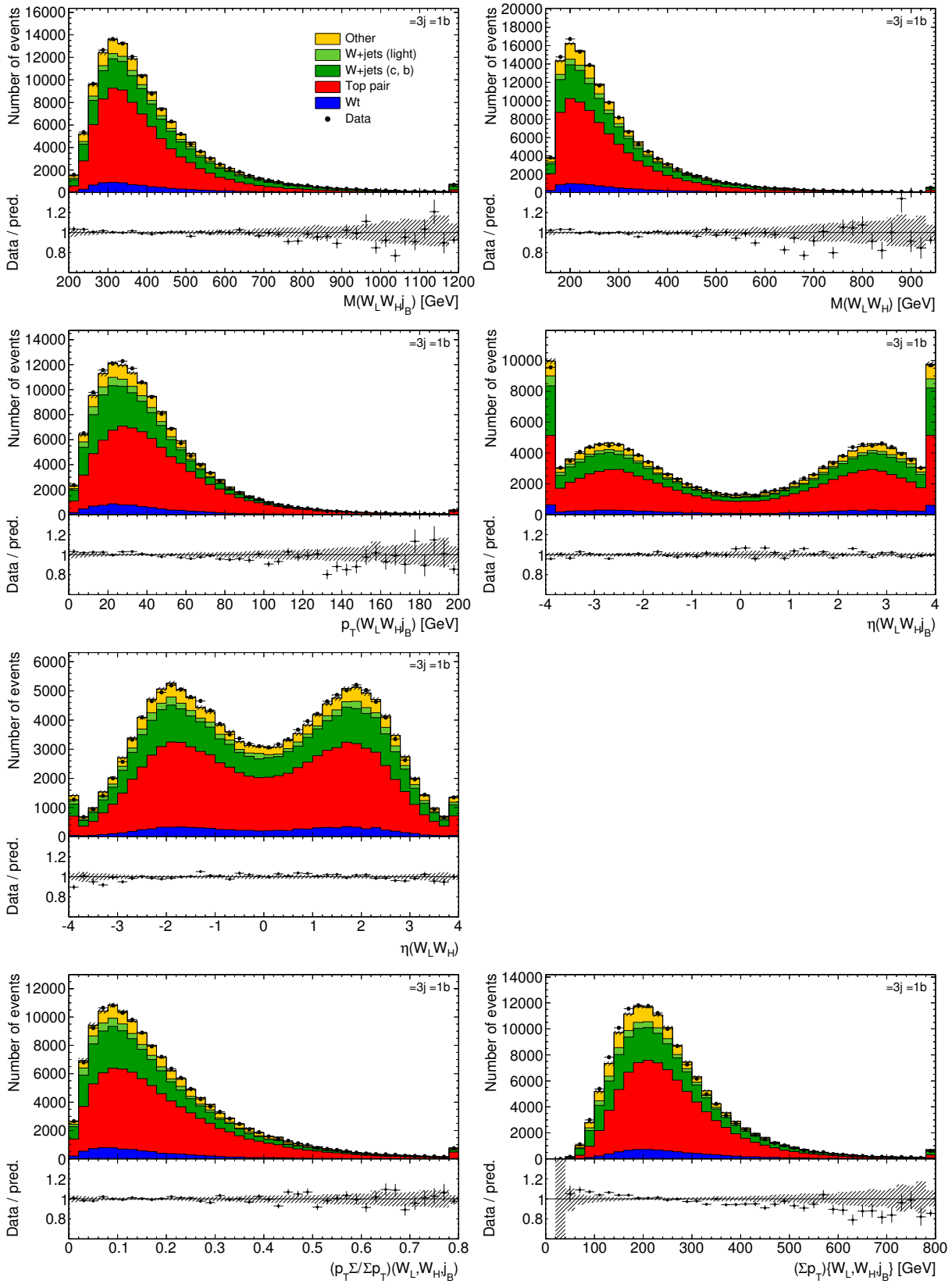

Figure D.8: Distributions in the signal region before applying the cut on $M(W_{\mathrm{H}})$.

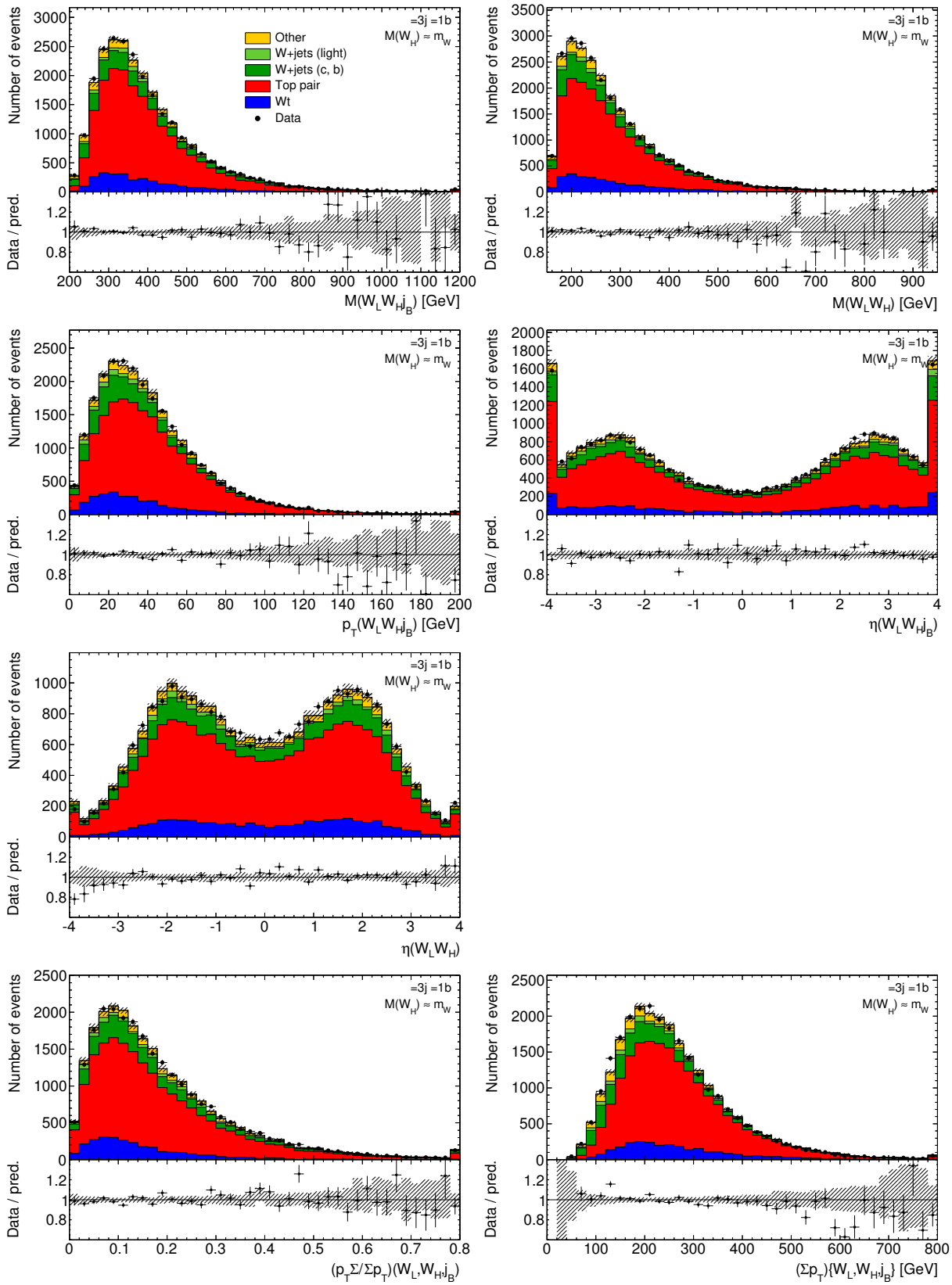

Figure D.9: Distributions in the signal region after applying the cut on $M(W_{\mathrm{H}})$.

APPENDIX E

Neural network training

Variable	Additional significance	Single significance	Significance loss	Global correlation [%]
$(p_{\mathrm{T}}\Sigma/\Sigma p_{\mathrm{T}})(W_{\mathrm{L}}, W_{\mathrm{H}}, j_{\mathrm{B}})$	38.7	38.7	48.1	35.1
$M(W_{\mathrm{L}} W_{\mathrm{H}} j_{\mathrm{B}})$	33.3	19.6	35.5	40.8
$\Delta\eta\ell$, $j_{\mathrm{L},1}$	10.7	2.0	11.6	30.2
$\eta(\ell)$	7.9	7.4	7.9	13.8

Table E.1: Input variables of the neural network used to enrich *Wt* events for the measurement presented in this thesis, ranked by their relevance. The *additional significance* quantifies by how much the linear correlation to the target is increased by each variable when reading the table top-down. The *single significance* gives the remaining correlation when removing all other variables, and the *significance loss* the decrease of correlation when removing only that particular variable. The correlation factors are commonly expressed in terms of a statistical significance, obtained by multiplying the correlation factor with statistical error on the total number of events in the training sample. The last column shows the correlation of each variable to all other variables. See Ref. [140] for a detailed description of the ranking algorithm.

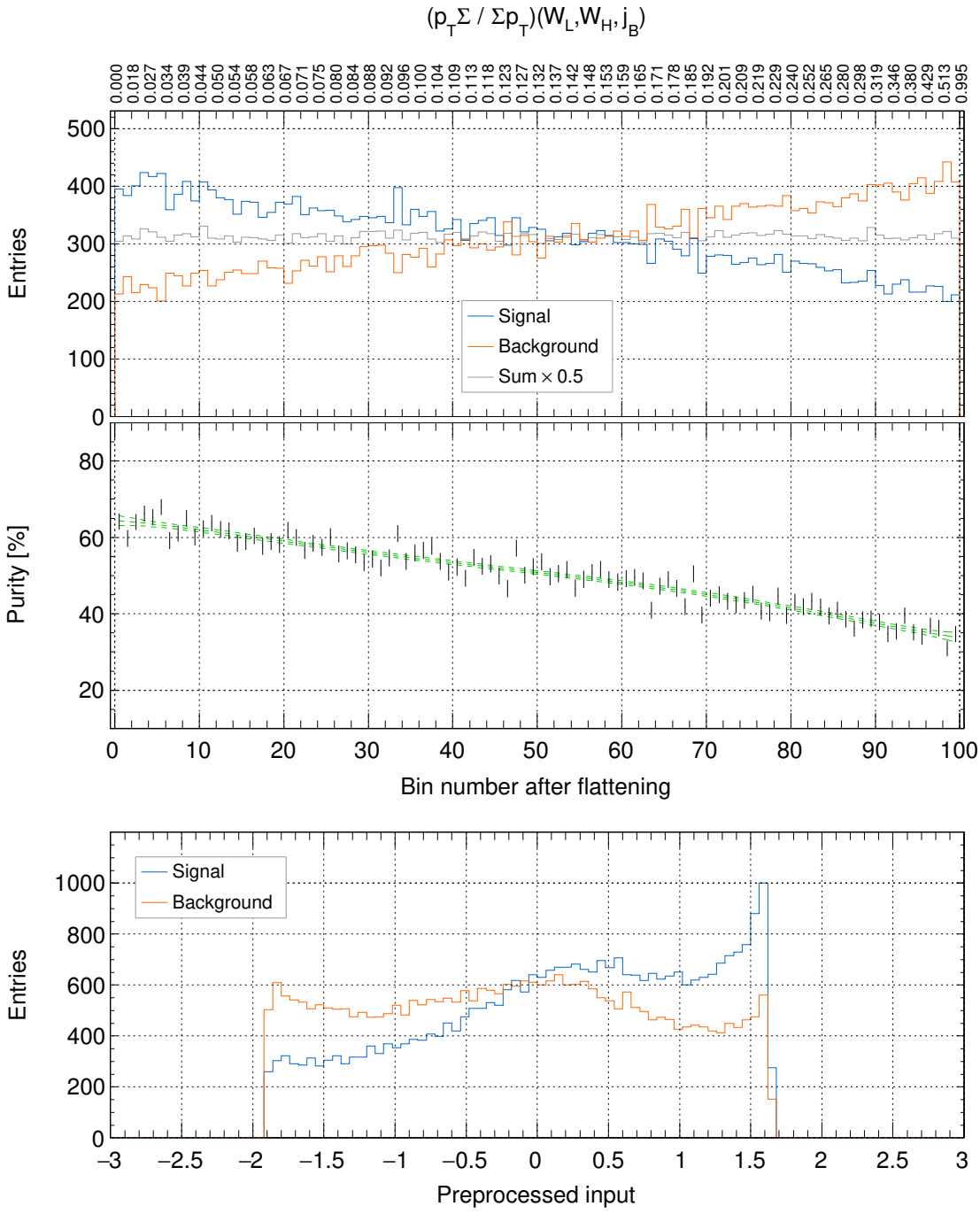

Figure E.1: Illustration of the preprocessing of $(p_T\Sigma/\Sigma p_T)(W_L, W_H, j_B)$ in the training sample.

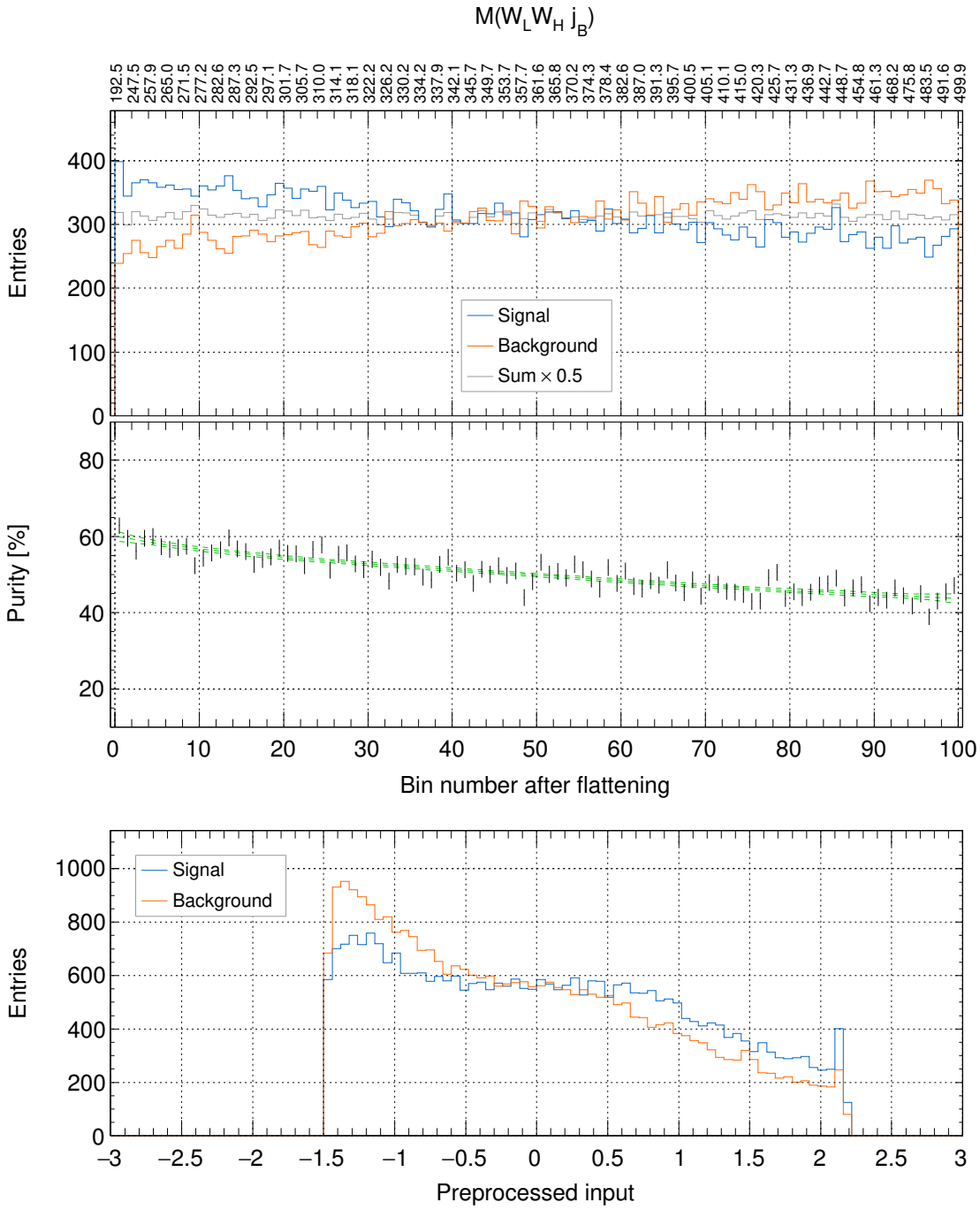

Figure E.2: Illustration of the preprocessing of $M(W_{\mathrm{L}} W_{\mathrm{H}} j_{\mathrm{B}})$ in the training sample.

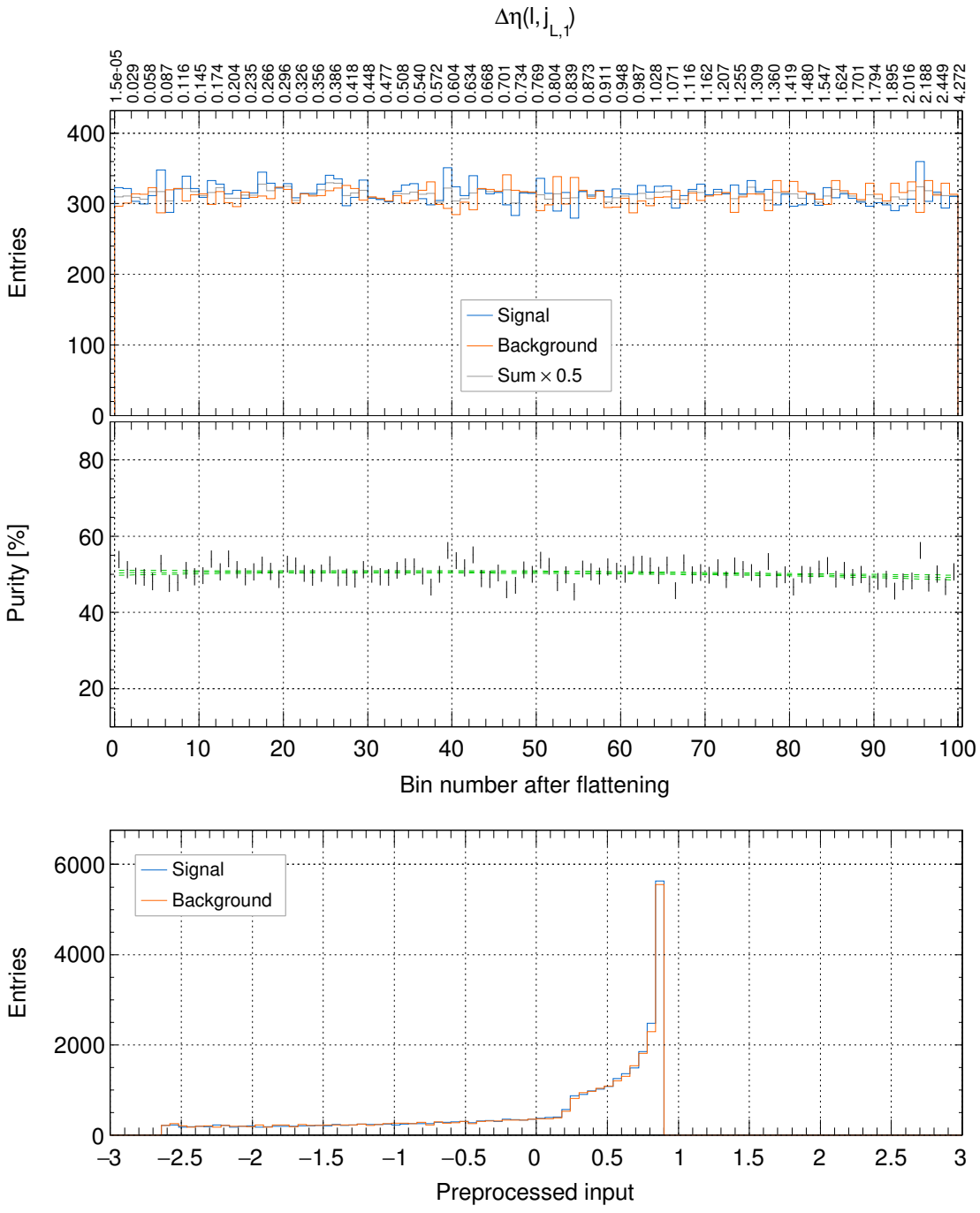

Figure E.3: Illustration of the preprocessing of $\eta(\ell)$ in the training sample.

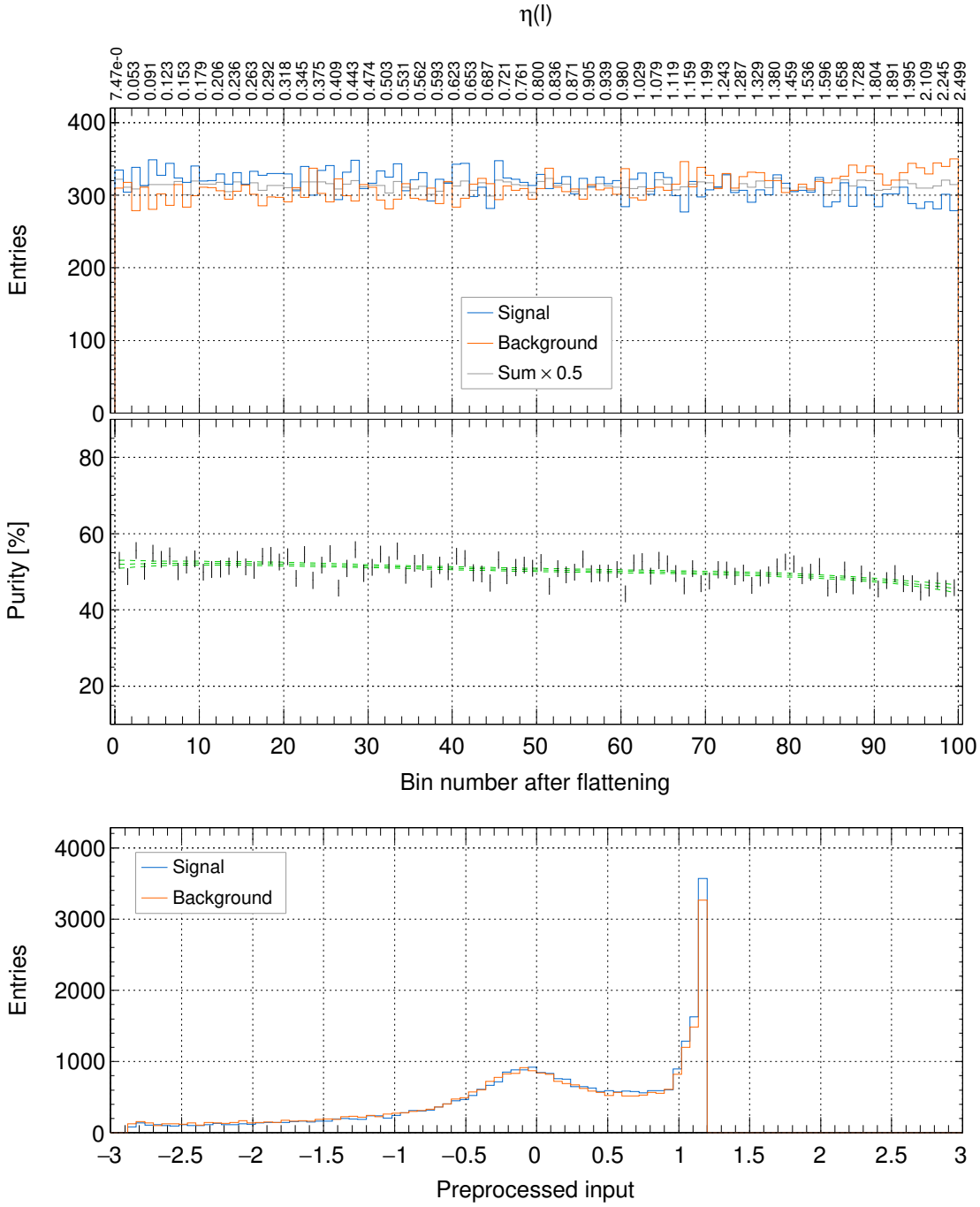

Figure E.4: Illustration of the preprocessing of $\Delta\eta\ell$, $j_{\mathrm{L},1}$ in the training sample.

Variable	Additional significance	Single significance	Significance loss	Global correlation [%]
$M(W_\mathrm{H})$	32.2	32.2	23.0	57.8
$p_\mathrm{T}(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B})$	23.0	24.9	18.9	42.1
$(\Sigma p_\mathrm{T})(\ell, \nu, j_{\mathrm{L},1}, j_{\mathrm{L},2}, j_\mathrm{B})$	14.7	20.4	10.4	58.0
$\Delta\phi(\nu, j_\mathrm{B})$	6.8	5.3	5.8	51.4
$\Delta\phi(W_\mathrm{L}j_\mathrm{B}, j_\mathrm{B})$	6.4	4.2	4.7	67.6
$\eta(W_\mathrm{H})$	5.6	9.4	5.0	47.7
$\eta(\ell)$	5.2	6.0	5.9	35.7
$p_\mathrm{T}^\mathrm{miss}$	5.0	8.2	5.3	37.2
$\Delta\phi(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B}, j_\mathrm{B})$	4.6	6.9	4.9	72.5
$\Delta\eta(W_\mathrm{L}j_\mathrm{B}, j_\mathrm{B})$	4.2	2.0	4.0	38.1
$\Delta\phi(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B}, j_{\mathrm{L},1})$	4.1	4.1	4.4	76.5
$\Delta\phi(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B}, j_{\mathrm{L},2})$	4.1	2.2	3.7	55.0
$\Delta R(W_\mathrm{H}, j_{\mathrm{L},2})$	3.0	13.5	3.9	86.6
$M(W_\mathrm{H}j_\mathrm{B})$	2.8	21.2	3.3	62.5
$p_\mathrm{T}(W_\mathrm{L})$	2.7	9.0	2.7	47.9
$\Delta\phi(j_{\mathrm{L},1}, j_{\mathrm{L},2})$	2.6	11.2	2.5	86.1
$\Delta\eta(W_\mathrm{L}, \ell)$	2.3	4.0	2.3	34.4
$\Delta\phi(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B}, \ell)$	1.9	1.4	1.6	68.9
$\Delta\phi(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B}, W_\mathrm{L})$	0.5	1.0	0.5	65.4
$\Delta\phi(W_\mathrm{L}W_\mathrm{H}j_\mathrm{B}, \nu)$	0.1	2.2	0.1	70.6

Table E.2: Input variables of the neural network described in Section 7.3, ranked by their relevance. See Table E.1 for a description of the meaning of the numbers. For better comparison with Table 9.3 of Ref. [144], the significances are calculated for a sample size of 37 667.

APPENDIX F

Systematic variations

Bootstrap method

The statistical error can be approximated by means of *bootstrapping*, a technique that is similar to the sampling of the systematic uncertainties (see Section 8.4) but performs resampling of the dataset instead of the random sampling of a nuisance parameter. 1 000 replicas of the sample are created by random sampling with replacement, i.e. each replica is almost identical to the original sample but the weight of every event is multiplied by a random number drawn from a Poisson distribution with mean 1. Each replica, identified by i, is processed in the same way as the original sample, yielding an event rate for the $+1\sigma$ and the -1σ variations, from which a $\tilde{\delta}_i$ is computed. The $\tilde{\delta}_i$ differs from the original δ only by random fluctuations, and its distribution reflects the p.d.f. of δ. It follows approximately a normal distribution with a mean of δ, and its width corresponds to the sought-after statistical error.

The bootstrap method was previously applied for jet physics with the ATLAS detector in order to control the effect of statistical fluctuations in the fit. The implementation in the BootstrapGenerator in the ATLAS software repository is used.

Systematic variations

	Fake leptons	
Uncertainty	δ_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$
Fake $e^{\pm}$: CR for fake eff.	-2.8 ± 0.3	8.9
Fake $e^{\pm}$: CR for real eff.	-4.5 ± 0.3	2.3
Fake $e^{\pm}$: parametrisation	-19.1 ± 0.4	2.0
Fake $e^{\pm}$: MC normalisation	-36.3 ± 0.9	1.7
Fake $e^{\pm}$: trigiso	-1.1 ± 0.5	1.9
Fake $\mu^{\pm}$: CR for fake eff.	-2.4 ± 0.2	1.4
Fake $\mu^{\pm}$: CR for real eff.	-7.0 ± 0.2	4.1
Fake $\mu^{\pm}$: parametrisation	-9.1 ± 0.3	1.2
Fake $\mu^{\pm}$: MC normalisation	-5.2 ± 0.2	1.3

Table F.1: Effects of the systematic uncertainties on the number of selected events in the signal region. The statistical errors on the numbers are computed using the bootstrap method. Also shown are the results of the χ^2 test for compatibility between the nominal and the varied shape of the discriminant.

Uncertainty	W+jets, heavy flavour δ_{overall} [%]	W+jets, heavy flavour $\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	W+jets, light δ_{overall} [%]	W+jets, light $\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	Z+jets δ_{overall} [%]	Z+jets $\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$
b-tagging: B0	0.0 ± 0.0	1.3	0.0 ± 0.0	0.0	0.0 ± 0.0	0.8
b-tagging: B1	-0.1 ± 0.0	14	0.0 ± 0.0	0.0	-0.1 ± 0.0	1.7
b-tagging: B2	-0.1 ± 0.0	6.5	0.0 ± 0.0	0.0	-0.2 ± 0.0	1.7
b-tagging: B3	-0.2 ± 0.0	2.8	0.0 ± 0.0	0.0	-0.2 ± 0.0	2.0
b-tagging: B4	0.0 ± 0.0	3.8	0.0 ± 0.0	0.0	0.0 ± 0.0	2.3
b-tagging: B5	-0.4 ± 0.0	3.3	0.0 ± 0.0	0.0	-0.5 ± 0.0	1.6
b-tagging: C	3.9 ± 0.0	6.0	0.0 ± 0.0	0.7	2.2 ± 0.1	1.4
b-tagging: L	1.0 ± 0.0	1.2	16.5 ± 0.1	0.9	3.0 ± 0.2	5.1
JES: b-quark reponse	0.4 ± 0.0	0.9	0.0 ± 0.0	0.0	0.0 ± 0.1	1.2
JES: detector 1	0.3 ± 0.0	1.5	0.7 ± 0.2	1.4	0.0 ± 0.1	1.2
JES: detector 2	0.3 ± 0.0	1.1	0.2 ± 0.1	1.2	0.4 ± 0.1	1.3
JES: detector 3	0.0 ± 0.0	1.4	0.2 ± 0.1	1.8	0.0 ± 0.0	0.7
JES: mixed 1	0.5 ± 0.0	1.0	0.8 ± 0.2	1.3	0.4 ± 0.2	1.5
JES: mixed 2	-0.6 ± 0.0	1.4	-0.5 ± 0.2	1.5	-0.8 ± 0.2	1.3
JES: mixed 3	0.1 ± 0.0	1.8	0.1 ± 0.1	0.7	0.3 ± 0.1	1.0
JES: mixed 4	0.0 ± 0.0	0.9	0.0 ± 0.0	0.9	0.0 ± 0.0	1.5
JES: modelling 1	3.4 ± 0.1	2.7	4.2 ± 0.5	1.2	3.6 ± 0.4	1.4
JES: modelling 2	-0.1 ± 0.0	0.8	0.0 ± 0.1	0.5	-0.2 ± 0.1	1.7
JES: modelling 3	0.0 ± 0.0	1.8	-0.1 ± 0.1	0.7	0.2 ± 0.1	1.0
JES: modelling 4	0.1 ± 0.0	1.2	0.1 ± 0.1	1.6	-0.1 ± 0.1	0.9
JES: statistical 1	1.2 ± 0.1	2.4	1.3 ± 0.3	1.1	1.6 ± 0.3	1.3
JES: statistical 2	-0.1 ± 0.0	1.1	-0.2 ± 0.1	2.9	-0.1 ± 0.0	2.8
JES: statistical 3	-0.3 ± 0.0	0.9	-0.3 ± 0.2	1.0	-0.1 ± 0.1	1.2
JES: statistical 4	0.2 ± 0.0	1.2	0.3 ± 0.2	1.4	0.0 ± 0.1	0.7
JES: η intercalibration: modelling	0.7 ± 0.1	0.9	0.7 ± 0.3	1.0	0.9 ± 0.2	1.3
JES: η intercalibration: statistical	0.7 ± 0.0	1.1	1.0 ± 0.3	1.1	1.0 ± 0.3	1.3
JES: quark/gluon composition	3.2 ± 0.1	3.0	5.1 ± 0.6	1.2	3.9 ± 0.5	1.5
JES: quark/gluon response	-2.1 ± 0.1	1.7	-3.1 ± 0.5	1.0	-2.7 ± 0.4	1.7
JES: pile-up: OffsetMu	-0.2 ± 0.0	1.0	-0.1 ± 0.2	1.0	-0.3 ± 0.2	1.0
JES: pile-up: OffsetNPV	-0.2 ± 0.1	1.1	-0.2 ± 0.4	1.0	-0.2 ± 0.4	0.9
JES: pile-up: Pt	0.1 ± 0.0	0.8	0.0 ± 0.2	0.8	0.0 ± 0.1	1.0
JES: pile-up: RhoTopology	2.2 ± 0.1	2.4	2.5 ± 0.4	0.9	2.8 ± 0.4	1.6
JES: punch through	0.0 ± 0.0	0.1	0.0 ± 0.0	0.9	0.0 ± 0.0	1.9
JES: high-p_{T} single particle	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0
Jet reconstruction eff.	0.0 ± 0.0	0.9	0.0 ± 0.1	0.4	0.1 ± 0.0	1.8
Jet energy resolution	2.7 ± 0.4	1.3	5.7 ± 2.0	0.8	9.7 ± 1.5	1.2
Jet vertex fraction	-0.6 ± 0.0	4.0	-0.2 ± 0.2	1.0	-0.9 ± 0.1	2.6
$e^{\pm}$ energy resolution	0.0 ± 0.0	0.8	-0.1 ± 0.2	0.6	0.2 ± 0.2	0.9
$e^{\pm}$ energy scale	1.0 ± 0.0	1.7	0.9 ± 0.2	3.9	1.8 ± 0.2	1.7
$e^{\pm}$ identification eff.	1.0 ± 0.0	0.6	1.0 ± 0.0	1.1	1.5 ± 0.0	1.4
$e^{\pm}$ reconstruction eff.	0.1 ± 0.0	2.5	0.1 ± 0.0	1.6	0.2 ± 0.0	1.2
$e^{\pm}$ trigger eff.	0.3 ± 0.0	2.9	0.3 ± 0.0	1.4	0.5 ± 0.0	1.3
$\mu^{\pm}$ energy resolution: inner detector	0.0 ± 0.0	1.2	-0.1 ± 0.1	0.5	-0.1 ± 0.1	0.8
$\mu^{\pm}$ energy resolution: spectrometer	0.0 ± 0.0	1.0	-0.2 ± 0.2	0.9	-0.1 ± 0.1	0.9
$\mu^{\pm}$ energy scale	0.1 ± 0.0	2.6	0.2 ± 0.1	2.3	0.1 ± 0.0	4.5
$\mu^{\pm}$ identification eff.	0.3 ± 0.0	0.5	0.3 ± 0.0	1.0	0.2 ± 0.0	1.5
$\mu^{\pm}$ reconstruction eff.	0.1 ± 0.0	1.4	0.1 ± 0.0	0.8	0.1 ± 0.0	1.7
$\mu^{\pm}$ trigger eff.	1.0 ± 0.0	0.7	1.0 ± 0.0	1.1	0.7 ± 0.0	1.4
$p_{\text{T}}^{\text{miss}}$ soft term resolution	0.0 ± 0.1	0.9	0.3 ± 0.3	1.2	-0.5 ± 0.3	1.3
$p_{\text{T}}^{\text{miss}}$ soft term scale	0.2 ± 0.0	1.5	0.5 ± 0.3	1.4	1.7 ± 0.3	1.4

Table F.2: Effects of the systematic uncertainties on the numbers of selected events in the signal region. The statistical errors on the numbers are computed using the bootstrap method. Also shown are the results of the χ^2 test for compatibility between the nominal and the varied shape of the discriminant.

Uncertainty	t, t-channel		t, s-channel		$WW/WZ/ZZ$	
	δ_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	δ_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	δ_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$
b-tagging: B0	0.0 ± 0.0	2.4	0.0 ± 0.0	3.4	0.0 ± 0.0	1.0
b-tagging: B1	0.1 ± 0.0	24	0.1 ± 0.0	49	0.0 ± 0.0	1.4
b-tagging: B2	0.1 ± 0.0	22	0.3 ± 0.0	18	0.0 ± 0.0	1.4
b-tagging: B3	0.1 ± 0.0	61	0.5 ± 0.0	23	0.0 ± 0.0	1.3
b-tagging: B4	0.0 ± 0.0	6.1	0.1 ± 0.0	9.7	0.0 ± 0.0	1.1
b-tagging: B5	0.2 ± 0.0	55	1.2 ± 0.0	35	0.0 ± 0.0	1.5
b-tagging: C	-0.1 ± 0.0	2.0	-0.1 ± 0.0	1.1	4.5 ± 0.1	2.0
b-tagging: L	-0.1 ± 0.0	13	-0.1 ± 0.0	9.3	2.3 ± 0.1	1.6
JES: b-quark reponse	0.8 ± 0.0	1.4	0.6 ± 0.1	0.8	0.3 ± 0.1	1.1
JES: detector 1	0.4 ± 0.0	2.3	0.2 ± 0.0	1.2	0.3 ± 0.1	1.0
JES: detector 2	0.3 ± 0.0	1.2	0.2 ± 0.0	0.9	0.3 ± 0.1	1.5
JES: detector 3	0.1 ± 0.0	0.9	0.0 ± 0.0	1.6	0.0 ± 0.1	0.4
JES: mixed 1	0.5 ± 0.0	1.4	0.3 ± 0.0	0.9	0.5 ± 0.2	1.1
JES: mixed 2	-0.5 ± 0.0	1.5	-0.3 ± 0.0	0.7	-0.5 ± 0.1	1.5
JES: mixed 3	0.1 ± 0.0	1.1	0.1 ± 0.0	1.7	0.0 ± 0.1	1.0
JES: mixed 4	0.0 ± 0.0	1.4	0.0 ± 0.0	1.0	0.0 ± 0.0	0.2
JES: modelling 1	3.0 ± 0.1	1.4	2.4 ± 0.1	1.9	4.6 ± 0.4	1.4
JES: modelling 2	-0.1 ± 0.0	1.1	-0.1 ± 0.0	1.1	0.0 ± 0.1	0.6
JES: modelling 3	0.0 ± 0.0	1.8	0.0 ± 0.0	1.1	0.0 ± 0.1	0.9
JES: modelling 4	0.1 ± 0.0	1.0	0.0 ± 0.0	1.4	0.0 ± 0.1	0.7
JES: statistical 1	1.0 ± 0.0	1.9	0.8 ± 0.1	1.0	1.1 ± 0.2	0.9
JES: statistical 2	-0.1 ± 0.0	1.0	-0.1 ± 0.0	1.4	0.0 ± 0.1	0.6
JES: statistical 3	-0.2 ± 0.0	1.5	-0.2 ± 0.0	1.0	-0.2 ± 0.1	1.3
JES: statistical 4	0.2 ± 0.0	1.2	0.1 ± 0.0	0.8	0.2 ± 0.1	1.4
JES: η intercalibration: modelling	0.8 ± 0.0	1.5	0.5 ± 0.1	1.1	0.5 ± 0.2	1.2
JES: η intercalibration: statistical	0.7 ± 0.0	1.2	0.5 ± 0.0	0.5	0.7 ± 0.2	1.1
JES: quark/gluon composition	2.2 ± 0.1	2.6	1.7 ± 0.1	2.2	4.8 ± 0.4	1.7
JES: quark/gluon response	-1.5 ± 0.1	1.7	-1.1 ± 0.1	1.8	-3.0 ± 0.3	1.6
JES: pile-up: OffsetMu	-0.2 ± 0.0	0.9	-0.1 ± 0.0	1.0	-0.3 ± 0.2	1.1
JES: pile-up: OffsetNPV	-0.5 ± 0.1	1.0	-0.3 ± 0.1	1.0	-0.2 ± 0.3	0.9
JES: pile-up: Pt	0.0 ± 0.0	0.9	0.1 ± 0.0	1.6	0.1 ± 0.1	0.9
JES: pile-up: RhoTopology	2.1 ± 0.1	1.5	1.6 ± 0.1	1.6	2.7 ± 0.3	1.3
JES: punch through	0.0 ± 0.0	0.7	0.0 ± 0.0	0.6	0.0 ± 0.0	0.3
JES: high-p_{T} single particle	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0
Jet reconstruction eff.	0.0 ± 0.0	0.9	0.0 ± 0.0	0.8	0.1 ± 0.1	1.2
Jet energy resolution	2.1 ± 0.3	0.9	1.6 ± 0.4	0.7	6.6 ± 1.6	0.8
Jet vertex fraction	-0.5 ± 0.0	2.1	-0.5 ± 0.0	2.9	-0.7 ± 0.2	2.4
$e^{\pm}$ energy resolution	0.0 ± 0.0	1.0	0.0 ± 0.0	1.4	-0.2 ± 0.1	1.1
$e^{\pm}$ energy scale	0.9 ± 0.0	1.4	0.9 ± 0.0	1.2	0.7 ± 0.2	1.7
$e^{\pm}$ identification eff.	1.0 ± 0.0	1.4	1.0 ± 0.0	1.5	1.0 ± 0.0	1.1
$e^{\pm}$ reconstruction eff.	0.1 ± 0.0	2.6	0.1 ± 0.0	2.4	0.1 ± 0.0	1.7
$e^{\pm}$ trigger eff.	0.2 ± 0.0	1.3	0.2 ± 0.0	1.9	0.3 ± 0.0	2.0
$\mu^{\pm}$ energy resolution: inner detector	0.0 ± 0.0	0.9	0.0 ± 0.0	1.3	0.1 ± 0.1	2.0
$\mu^{\pm}$ energy resolution: spectrometer	0.0 ± 0.0	0.7	0.0 ± 0.0	0.9	-0.1 ± 0.2	0.5
$\mu^{\pm}$ energy scale	0.1 ± 0.0	10	0.1 ± 0.0	12	0.1 ± 0.0	1.9
$\mu^{\pm}$ identification eff.	0.3 ± 0.0	1.2	0.3 ± 0.0	1.5	0.3 ± 0.0	1.1
$\mu^{\pm}$ reconstruction eff.	0.1 ± 0.0	0.9	0.1 ± 0.0	1.3	0.1 ± 0.0	1.5
$\mu^{\pm}$ trigger eff.	1.0 ± 0.0	1.2	1.0 ± 0.0	1.3	1.0 ± 0.0	1.0
$p_{\text{T}}^{\text{miss}}$ soft term resolution	-0.1 ± 0.0	1.1	0.0 ± 0.1	1.1	-0.2 ± 0.2	0.9
$p_{\text{T}}^{\text{miss}}$ soft term scale	0.0 ± 0.0	2.0	0.1 ± 0.0	1.7	0.3 ± 0.2	1.1

Table F.3: Effects of the systematic uncertainties on the numbers of selected events in the signal region. The statistical errors on the numbers are computed using the bootstrap method. Also shown are the results of the χ^2 test for compatibility between the nominal and the varied shape of the discriminant.

Uncertainty	Wt		$t\bar{t}$		W+jets, beauty	
	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$
b-tagging: B0	0.0 ± 0.0	2.0	0.0 ± 0.0	110	0.0 ± 0.0	1.1
b-tagging: B1	0.0 ± 0.0	2.0	0.0 ± 0.0	120	0.0 ± 0.0	5.7
b-tagging: B2	0.0 ± 0.0	2.0	0.0 ± 0.0	120	0.0 ± 0.0	6.0
b-tagging: B3	0.0 ± 0.0	2.2	0.0 ± 0.0	67	0.0 ± 0.0	3.4
b-tagging: B4	0.0 ± 0.0	1.9	0.0 ± 0.0	83	0.0 ± 0.0	1.4
b-tagging: B5	0.0 ± 0.0	4.4	0.0 ± 0.0	540	0.0 ± 0.0	14
b-tagging: C	0.0 ± 0.0	3.8	0.0 ± 0.0	84	0.0 ± 0.0	5.1
b-tagging: L	0.0 ± 0.0	5.0	0.0 ± 0.0	3.2	0.0 ± 0.0	3.0
JES: b-quark reponse	0.0 ± 0.1	1.2	0.0 ± 0.0	1.1	0.1 ± 0.0	0.8
JES: detector 1	0.0 ± 0.1	0.8	0.0 ± 0.0	0.8	0.0 ± 0.0	0.9
JES: detector 2	0.0 ± 0.1	0.9	0.0 ± 0.0	1.1	0.0 ± 0.0	0.9
JES: detector 3	0.0 ± 0.0	1.1	0.0 ± 0.0	0.8	0.0 ± 0.0	0.8
JES: mixed 1	0.0 ± 0.1	0.9	0.0 ± 0.0	1.2	0.0 ± 0.0	1.1
JES: mixed 2	0.0 ± 0.1	1.2	0.0 ± 0.0	1.2	0.0 ± 0.0	1.4
JES: mixed 3	0.0 ± 0.0	1.1	0.0 ± 0.0	0.9	0.0 ± 0.0	0.8
JES: mixed 4	0.0 ± 0.0	0.8	0.0 ± 0.0	0.9	0.0 ± 0.0	1.1
JES: modelling 1	-0.2 ± 0.2	0.9	0.0 ± 0.0	1.2	0.0 ± 0.1	1.1
JES: modelling 2	0.0 ± 0.0	1.1	0.0 ± 0.0	0.7	0.0 ± 0.0	0.7
JES: modelling 3	0.0 ± 0.0	0.8	0.0 ± 0.0	0.9	0.0 ± 0.0	1.0
JES: modelling 4	0.0 ± 0.0	1.0	0.0 ± 0.0	1.0	0.0 ± 0.0	0.9
JES: statistical 1	0.0 ± 0.1	0.9	0.0 ± 0.0	1.1	0.0 ± 0.1	1.6
JES: statistical 2	0.0 ± 0.0	1.2	0.0 ± 0.0	0.7	0.0 ± 0.0	0.9
JES: statistical 3	0.0 ± 0.1	0.9	0.0 ± 0.0	0.8	0.0 ± 0.0	1.2
JES: statistical 4	0.0 ± 0.0	1.1	0.0 ± 0.0	0.9	0.0 ± 0.0	1.0
JES: η intercalibration: modelling	0.0 ± 0.1	0.9	0.0 ± 0.0	0.5	0.0 ± 0.0	1.1
JES: η intercalibration: statistical	0.0 ± 0.1	0.9	0.0 ± 0.0	1.0	0.0 ± 0.0	1.4
JES: quark/gluon composition	0.1 ± 0.2	0.9	0.0 ± 0.0	1.1	0.0 ± 0.1	1.4
JES: quark/gluon response	0.0 ± 0.2	1.0	0.0 ± 0.0	1.1	0.0 ± 0.1	1.2
JES: pile-up: OffsetMu	0.0 ± 0.1	1.0	0.0 ± 0.0	0.9	0.0 ± 0.0	1.4
JES: pile-up: OffsetNPV	-0.1 ± 0.2	1.0	0.0 ± 0.0	1.0	0.1 ± 0.1	1.3
JES: pile-up: Pt	-0.1 ± 0.1	1.3	0.0 ± 0.0	0.7	0.0 ± 0.0	1.3
JES: pile-up: RhoTopology	-0.1 ± 0.2	1.2	0.0 ± 0.0	1.1	0.1 ± 0.1	1.1
JES: punch through	0.0 ± 0.0	1.2	0.0 ± 0.0	0.9	0.0 ± 0.0	0.3
JES: high-p_{T} single particle	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0
Jet reconstruction eff.	–	–	–	–	–	–
Jet energy resolution	–	–	–	–	–	–
Jet vertex fraction	0.0 ± 0.1	1.1	0.0 ± 0.0	1.1	-0.1 ± 0.0	1.5
$e^{\pm}$ energy resolution	0.0 ± 0.1	1.0	0.0 ± 0.0	0.8	0.1 ± 0.0	1.0
$e^{\pm}$ energy scale	-0.1 ± 0.1	1.3	0.0 ± 0.0	0.6	0.0 ± 0.0	1.1
$e^{\pm}$ identification eff.	0.0 ± 0.0	1.1	0.0 ± 0.0	4.3	0.0 ± 0.0	1.0
$e^{\pm}$ reconstruction eff.	0.0 ± 0.0	2.0	0.0 ± 0.0	47	0.0 ± 0.0	3.5
$e^{\pm}$ trigger eff.	0.0 ± 0.0	1.5	0.0 ± 0.0	29	0.0 ± 0.0	4.2
$\mu^{\pm}$ energy resolution: inner detector	–	–	–	–	–	–
$\mu^{\pm}$ energy resolution: spectrometer	–	–	–	–	–	–
$\mu^{\pm}$ energy scale	0.0 ± 0.0	0.7	0.0 ± 0.0	0.9	0.0 ± 0.0	1.2
$\mu^{\pm}$ identification eff.	0.0 ± 0.0	1.0	0.0 ± 0.0	2.0	0.0 ± 0.0	1.1
$\mu^{\pm}$ reconstruction eff.	0.0 ± 0.0	1.1	0.0 ± 0.0	3.9	0.0 ± 0.0	2.1
$\mu^{\pm}$ trigger eff.	0.0 ± 0.0	1.1	0.0 ± 0.0	1.4	0.0 ± 0.0	1.0
$p_{\text{T}}^{\text{miss}}$ soft term resolution	0.1 ± 0.2	1.1	0.0 ± 0.0	0.8	0.1 ± 0.1	0.7
$p_{\text{T}}^{\text{miss}}$ soft term scale	0.0 ± 0.1	1.2	0.0 ± 0.0	0.6	0.0 ± 0.0	1.0
NLO subtraction scheme	–	–	–	–	–	–
Hadronisation model	–	–	–	–	–	–
Renormalisation scale	–	–	–	–	–	–
Wt-$t\bar{t}$ interference	–	–	–	–	–	–

Table F.4: Asymmetries of the systematic uncertainties on the numbers of selected events in the signal region. The statistical errors on the numbers are computed using the bootstrap method. Also shown are the results of the χ^2 test for the hypothesis that the shape variations are symmetric.

Uncertainty	W+jets, heavy flavour		W+jets, light		Z+jets	
	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$
b-tagging: B0	0.0 ± 0.0	1.1	0.0 ± 0.0	0.0	0.0 ± 0.0	2.2
b-tagging: B1	0.0 ± 0.0	5.9	0.0 ± 0.0	0.0	0.0 ± 0.0	1.8
b-tagging: B2	0.0 ± 0.0	5.5	0.0 ± 0.0	0.0	0.0 ± 0.0	1.7
b-tagging: B3	0.0 ± 0.0	8.9	0.0 ± 0.0	0.0	0.0 ± 0.0	1.7
b-tagging: B4	0.0 ± 0.0	1.6	0.0 ± 0.0	0.0	0.0 ± 0.0	2.2
b-tagging: B5	0.0 ± 0.0	15	0.0 ± 0.0	0.0	0.0 ± 0.0	1.6
b-tagging: C	0.0 ± 0.0	5.2	0.0 ± 0.0	0.1	0.0 ± 0.0	1.3
b-tagging: L	0.0 ± 0.0	1.2	-0.1 ± 0.0	0.9	0.0 ± 0.0	4.5
JES: b-quark reponse	0.0 ± 0.0	0.8	0.0 ± 0.0	0.0	0.0 ± 0.1	1.3
JES: detector 1	0.0 ± 0.0	0.8	0.0 ± 0.2	0.9	0.1 ± 0.2	1.0
JES: detector 2	0.0 ± 0.0	1.0	0.1 ± 0.1	0.8	0.1 ± 0.1	1.0
JES: detector 3	0.0 ± 0.0	0.7	0.0 ± 0.1	0.5	-0.1 ± 0.0	1.1
JES: mixed 1	0.0 ± 0.0	0.5	-0.2 ± 0.2	1.1	0.2 ± 0.2	1.0
JES: mixed 2	0.0 ± 0.0	1.1	0.0 ± 0.2	0.9	0.3 ± 0.2	1.2
JES: mixed 3	0.0 ± 0.0	0.9	0.0 ± 0.1	0.4	0.0 ± 0.1	1.1
JES: mixed 4	0.0 ± 0.0	1.1	0.0 ± 0.0	0.8	0.0 ± 0.0	0.9
JES: modelling 1	0.1 ± 0.1	0.8	-0.5 ± 0.6	1.1	-0.1 ± 0.4	1.0
JES: modelling 2	0.0 ± 0.0	0.7	0.1 ± 0.1	0.8	0.0 ± 0.1	0.9
JES: modelling 3	0.0 ± 0.0	0.9	-0.1 ± 0.1	0.9	0.0 ± 0.1	0.7
JES: modelling 4	0.0 ± 0.0	0.9	0.0 ± 0.1	0.5	0.0 ± 0.1	0.9
JES: statistical 1	0.0 ± 0.1	0.9	-0.2 ± 0.3	1.2	0.1 ± 0.3	1.2
JES: statistical 2	0.0 ± 0.0	1.1	0.1 ± 0.1	0.6	0.0 ± 0.0	1.1
JES: statistical 3	0.0 ± 0.0	0.7	-0.1 ± 0.2	0.8	0.0 ± 0.1	0.8
JES: statistical 4	0.0 ± 0.0	0.9	-0.1 ± 0.1	0.7	-0.1 ± 0.1	0.9
JES: η intercalibration: modelling	0.0 ± 0.1	0.9	0.1 ± 0.3	1.1	-0.1 ± 0.2	1.2
JES: η intercalibration: statistical	0.0 ± 0.0	0.8	-0.2 ± 0.3	1.1	0.1 ± 0.3	1.2
JES: quark/gluon composition	0.0 ± 0.1	0.8	-0.8 ± 0.6	1.2	0.1 ± 0.5	1.2
JES: quark/gluon response	0.0 ± 0.1	0.7	-0.4 ± 0.5	1.4	0.2 ± 0.4	1.0
JES: pile-up: OffsetMu	0.0 ± 0.0	1.2	0.1 ± 0.2	1.0	0.1 ± 0.2	1.1
JES: pile-up: OffsetNPV	0.1 ± 0.1	0.8	-0.4 ± 0.5	1.1	0.4 ± 0.4	1.4
JES: pile-up: Pt	0.0 ± 0.0	1.0	-0.2 ± 0.2	0.8	0.1 ± 0.1	0.9
JES: pile-up: RhoTopology	0.0 ± 0.1	0.7	-0.1 ± 0.4	1.5	0.1 ± 0.4	1.0
JES: punch through	0.0 ± 0.0	0.3	0.0 ± 0.0	0.9	0.0 ± 0.0	0.1
JES: high-p_{T} single particle	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0
Jet reconstruction eff.	–	–	–	–	–	–
Jet energy resolution	–	–	–	–	–	–
Jet vertex fraction	-0.1 ± 0.0	1.3	0.0 ± 0.2	0.4	-0.2 ± 0.1	0.9
$e^{\pm}$ energy resolution	-0.1 ± 0.1	1.1	0.0 ± 0.2	0.5	-0.2 ± 0.4	0.8
$e^{\pm}$ energy scale	-0.1 ± 0.0	0.7	0.2 ± 0.2	0.8	0.1 ± 0.2	0.9
$e^{\pm}$ identification eff.	0.0 ± 0.0	0.6	0.0 ± 0.0	1.1	0.0 ± 0.0	1.4
$e^{\pm}$ reconstruction eff.	0.0 ± 0.0	2.5	0.0 ± 0.0	1.6	0.0 ± 0.0	1.2
$e^{\pm}$ trigger eff.	0.0 ± 0.0	2.9	0.0 ± 0.0	1.4	0.0 ± 0.0	1.3
$\mu^{\pm}$ energy resolution: inner detector	–	–	–	–	–	–
$\mu^{\pm}$ energy resolution: spectrometer	–	–	–	–	–	–
$\mu^{\pm}$ energy scale	0.0 ± 0.0	1.1	-0.1 ± 0.1	0.8	0.0 ± 0.0	0.3
$\mu^{\pm}$ identification eff.	0.0 ± 0.0	0.5	0.0 ± 0.0	1.0	0.0 ± 0.0	1.5
$\mu^{\pm}$ reconstruction eff.	0.0 ± 0.0	1.4	0.0 ± 0.0	0.8	0.0 ± 0.0	1.7
$\mu^{\pm}$ trigger eff.	0.0 ± 0.0	0.7	0.0 ± 0.0	1.1	0.0 ± 0.0	1.4
$p_{\text{T}}^{\text{miss}}$ soft term resolution	0.0 ± 0.1	1.1	0.2 ± 0.4	1.2	0.6 ± 0.5	1.2
$p_{\text{T}}^{\text{miss}}$ soft term scale	0.0 ± 0.0	1.0	0.2 ± 0.3	1.0	0.1 ± 0.3	1.2

Table F.5: Asymmetries of the systematic uncertainties on the numbers of selected events in the signal region. The statistical errors on the numbers are computed using the bootstrap method. Also shown are the results of the χ^2 test for the hypothesis that the shape variations are symmetric.

Uncertainty	t, t-channel		t, s-channel		$WW/WZ/ZZ$	
	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$	α_{overall} [%]	$\frac{\chi^2_{\text{shape}}}{\text{n.d.f.}}$
b-tagging: B0	0.0 ± 0.0	1.2	0.0 ± 0.0	1.5	0.0 ± 0.0	1.1
b-tagging: B1	0.0 ± 0.0	10	0.0 ± 0.0	9.4	0.0 ± 0.0	0.9
b-tagging: B2	0.0 ± 0.0	16	0.0 ± 0.0	23	0.0 ± 0.0	0.8
b-tagging: B3	0.0 ± 0.0	38	0.0 ± 0.0	16	0.0 ± 0.0	0.7
b-tagging: B4	0.0 ± 0.0	1.7	0.0 ± 0.0	8.7	0.0 ± 0.0	0.4
b-tagging: B5	0.0 ± 0.0	55	-0.1 ± 0.0	53	0.0 ± 0.0	2.8
b-tagging: C	0.0 ± 0.0	2.0	0.0 ± 0.0	2.4	0.0 ± 0.0	1.9
b-tagging: L	0.0 ± 0.0	4.6	0.0 ± 0.0	1.3	0.0 ± 0.0	1.5
JES: b-quark reponse	0.0 ± 0.0	0.7	-0.1 ± 0.1	0.6	0.4 ± 0.1	2.1
JES: detector 1	0.0 ± 0.0	0.7	0.0 ± 0.0	1.1	0.1 ± 0.1	1.2
JES: detector 2	0.0 ± 0.0	0.7	0.0 ± 0.0	1.0	0.0 ± 0.1	0.9
JES: detector 3	0.0 ± 0.0	0.9	0.0 ± 0.0	1.1	-0.1 ± 0.1	0.8
JES: mixed 1	0.0 ± 0.0	0.7	0.0 ± 0.0	0.9	0.0 ± 0.1	1.0
JES: mixed 2	0.0 ± 0.0	0.8	0.0 ± 0.0	1.2	0.0 ± 0.1	1.1
JES: mixed 3	0.0 ± 0.0	0.9	0.0 ± 0.0	1.1	-0.1 ± 0.1	1.1
JES: mixed 4	0.0 ± 0.0	0.7	0.0 ± 0.0	0.8	0.0 ± 0.0	0.2
JES: modelling 1	0.0 ± 0.1	0.9	-0.1 ± 0.1	0.7	0.7 ± 0.4	0.7
JES: modelling 2	0.0 ± 0.0	1.0	0.0 ± 0.0	0.8	-0.1 ± 0.1	0.7
JES: modelling 3	0.0 ± 0.0	1.1	0.0 ± 0.0	1.2	0.0 ± 0.1	0.8
JES: modelling 4	0.0 ± 0.0	1.2	0.0 ± 0.0	1.2	-0.1 ± 0.1	1.0
JES: statistical 1	0.0 ± 0.0	0.9	0.1 ± 0.1	0.9	0.2 ± 0.2	1.0
JES: statistical 2	0.0 ± 0.0	0.9	0.0 ± 0.0	0.9	-0.1 ± 0.1	1.6
JES: statistical 3	0.0 ± 0.0	0.9	0.0 ± 0.0	1.2	0.0 ± 0.1	1.1
JES: statistical 4	0.0 ± 0.0	1.1	0.0 ± 0.0	1.1	0.0 ± 0.1	1.1
JES: η intercalibration: modelling	0.0 ± 0.0	0.7	0.0 ± 0.1	1.0	0.2 ± 0.2	0.9
JES: η intercalibration: statistical	0.0 ± 0.0	1.0	0.0 ± 0.0	1.1	0.0 ± 0.2	1.2
JES: quark/gluon composition	0.1 ± 0.1	0.8	0.0 ± 0.1	0.9	0.3 ± 0.4	0.9
JES: quark/gluon response	0.0 ± 0.1	0.8	0.0 ± 0.1	0.8	0.4 ± 0.3	0.8
JES: pile-up: OffsetMu	0.0 ± 0.0	0.9	0.0 ± 0.0	1.4	0.1 ± 0.2	1.2
JES: pile-up: OffsetNPV	-0.1 ± 0.1	0.6	0.0 ± 0.1	1.1	0.7 ± 0.3	0.9
JES: pile-up: Pt	0.0 ± 0.0	0.6	0.0 ± 0.0	1.0	0.0 ± 0.1	1.0
JES: pile-up: RhoTopology	-0.1 ± 0.1	0.9	0.0 ± 0.1	0.7	0.4 ± 0.3	0.7
JES: punch through	0.0 ± 0.0	0.6	0.0 ± 0.0	0.8	0.0 ± 0.0	0.6
JES: high-p_{T} single particle	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0	0.0 ± 0.0	0.0
Jet reconstruction eff.	–	–	–	–	–	–
Jet energy resolution	–	–	–	–	–	–
Jet vertex fraction	-0.1 ± 0.0	1.5	0.0 ± 0.0	1.2	0.0 ± 0.2	0.7
$e^{\pm}$ energy resolution	0.0 ± 0.0	1.1	-0.1 ± 0.1	1.1	0.1 ± 0.2	0.9
$e^{\pm}$ energy scale	0.0 ± 0.0	0.6	-0.1 ± 0.0	1.1	0.1 ± 0.1	0.9
$e^{\pm}$ identification eff.	0.0 ± 0.0	1.4	0.0 ± 0.0	1.5	0.0 ± 0.0	1.1
$e^{\pm}$ reconstruction eff.	0.0 ± 0.0	2.6	0.0 ± 0.0	2.3	0.0 ± 0.0	1.7
$e^{\pm}$ trigger eff.	0.0 ± 0.0	1.3	0.0 ± 0.0	1.8	0.0 ± 0.0	1.8
$\mu^{\pm}$ energy resolution: inner detector	–	–	–	–	–	–
$\mu^{\pm}$ energy resolution: spectrometer	–	–	–	–	–	–
$\mu^{\pm}$ energy scale	0.0 ± 0.0	1.5	0.0 ± 0.0	0.6	-0.1 ± 0.0	1.9
$\mu^{\pm}$ identification eff.	0.0 ± 0.0	1.2	0.0 ± 0.0	1.5	0.0 ± 0.0	1.1
$\mu^{\pm}$ reconstruction eff.	0.0 ± 0.0	0.9	0.0 ± 0.0	1.3	0.0 ± 0.0	1.5
$\mu^{\pm}$ trigger eff.	0.0 ± 0.0	1.2	0.0 ± 0.0	1.3	0.0 ± 0.0	1.0
$p_{\text{T}}^{\text{miss}}$ soft term resolution	0.0 ± 0.1	0.9	0.0 ± 0.1	1.1	0.5 ± 0.3	1.2
$p_{\text{T}}^{\text{miss}}$ soft term scale	-0.1 ± 0.0	0.6	0.0 ± 0.0	1.1	0.4 ± 0.2	0.9

Table F.6: Asymmetries of the systematic uncertainties on the numbers of selected events in the signal region. The statistical errors on the numbers are computed using the bootstrap method. Also shown are the results of the χ^2 test for the hypothesis that the shape variations are symmetric.

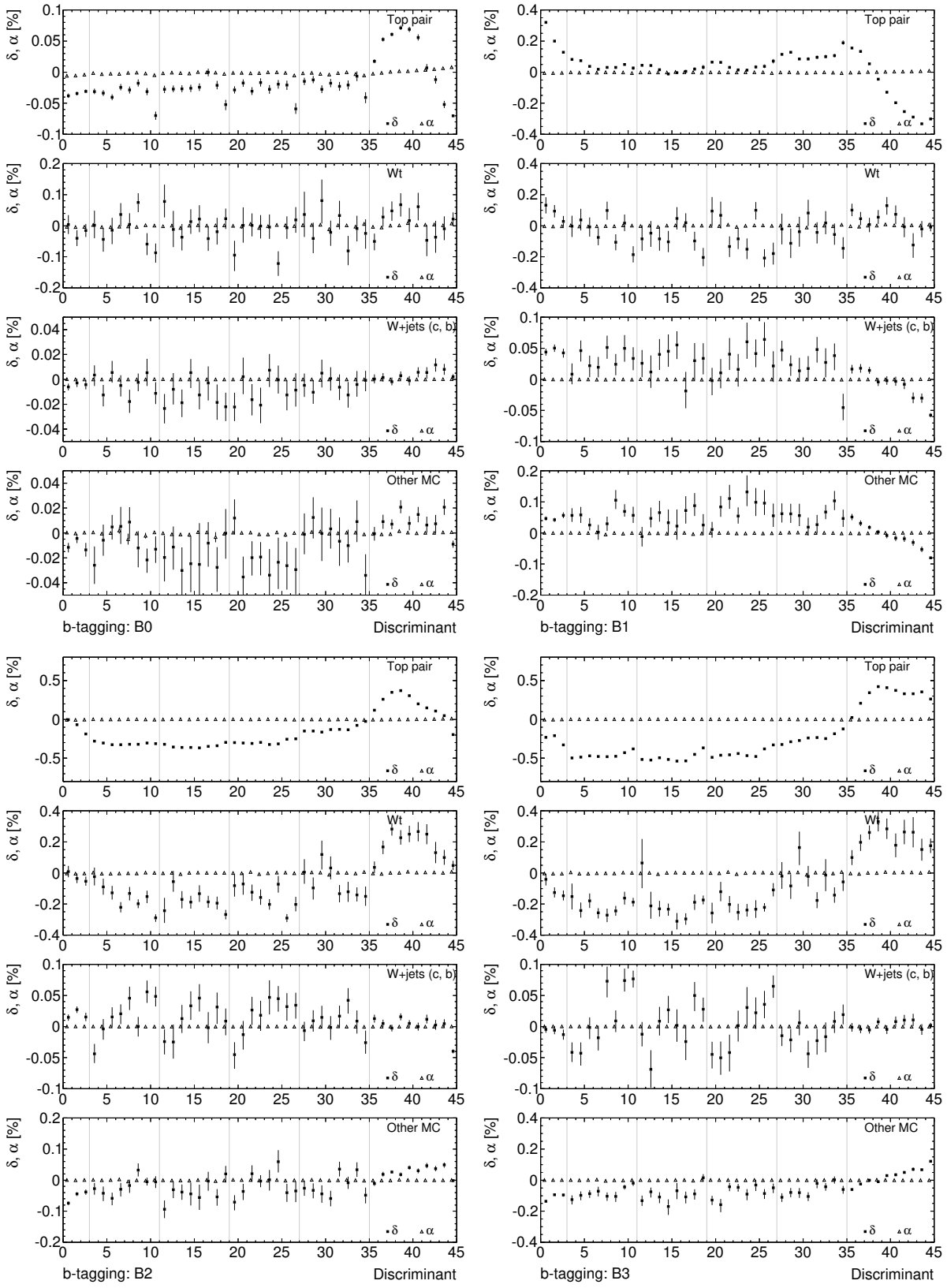

Figure F.1: Effects of the systematic uncertainties on the normalised distributions in the signal region.

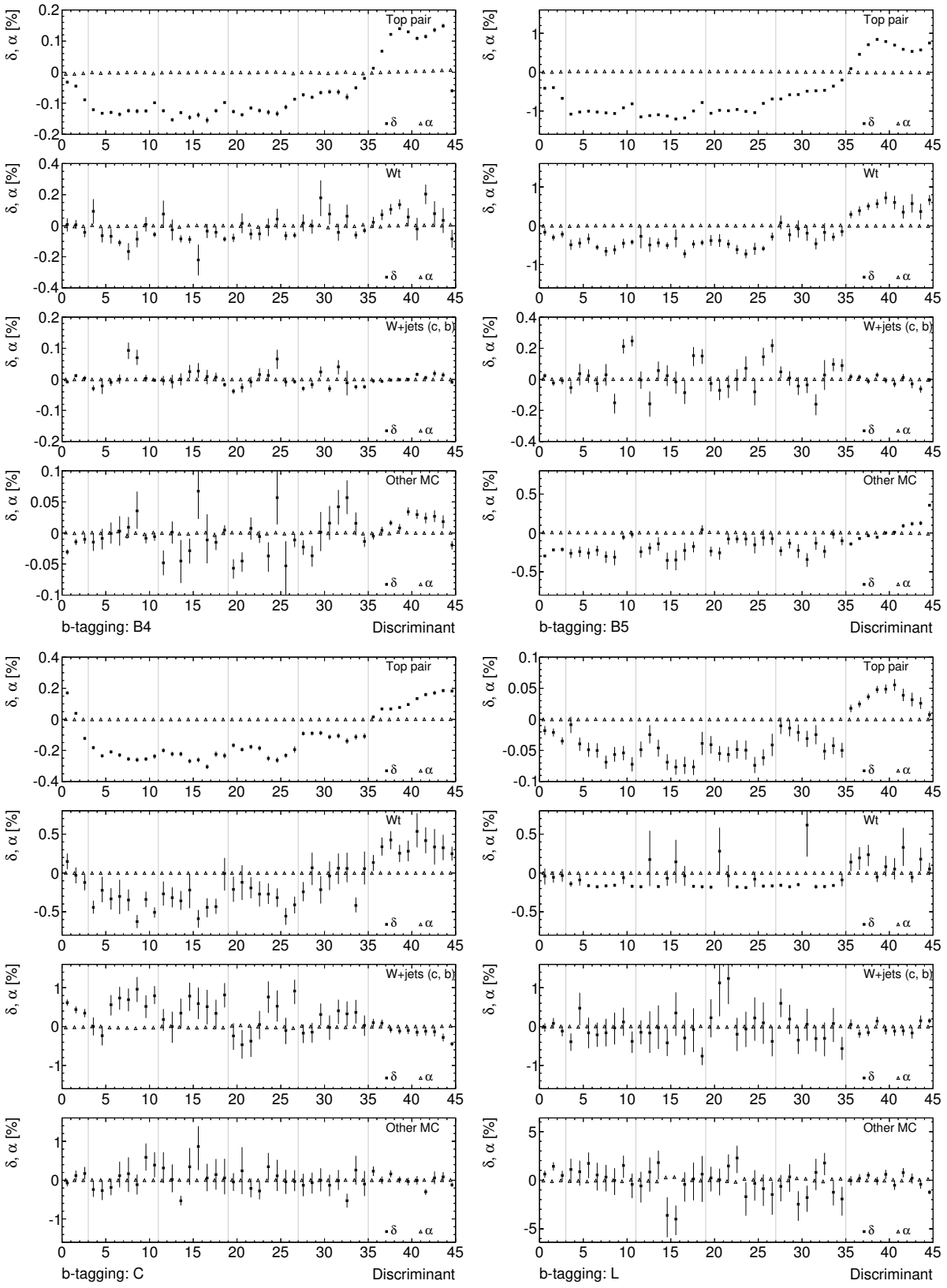

Figure F.2: Effects of the systematic uncertainties on the normalised distributions in the signal region.

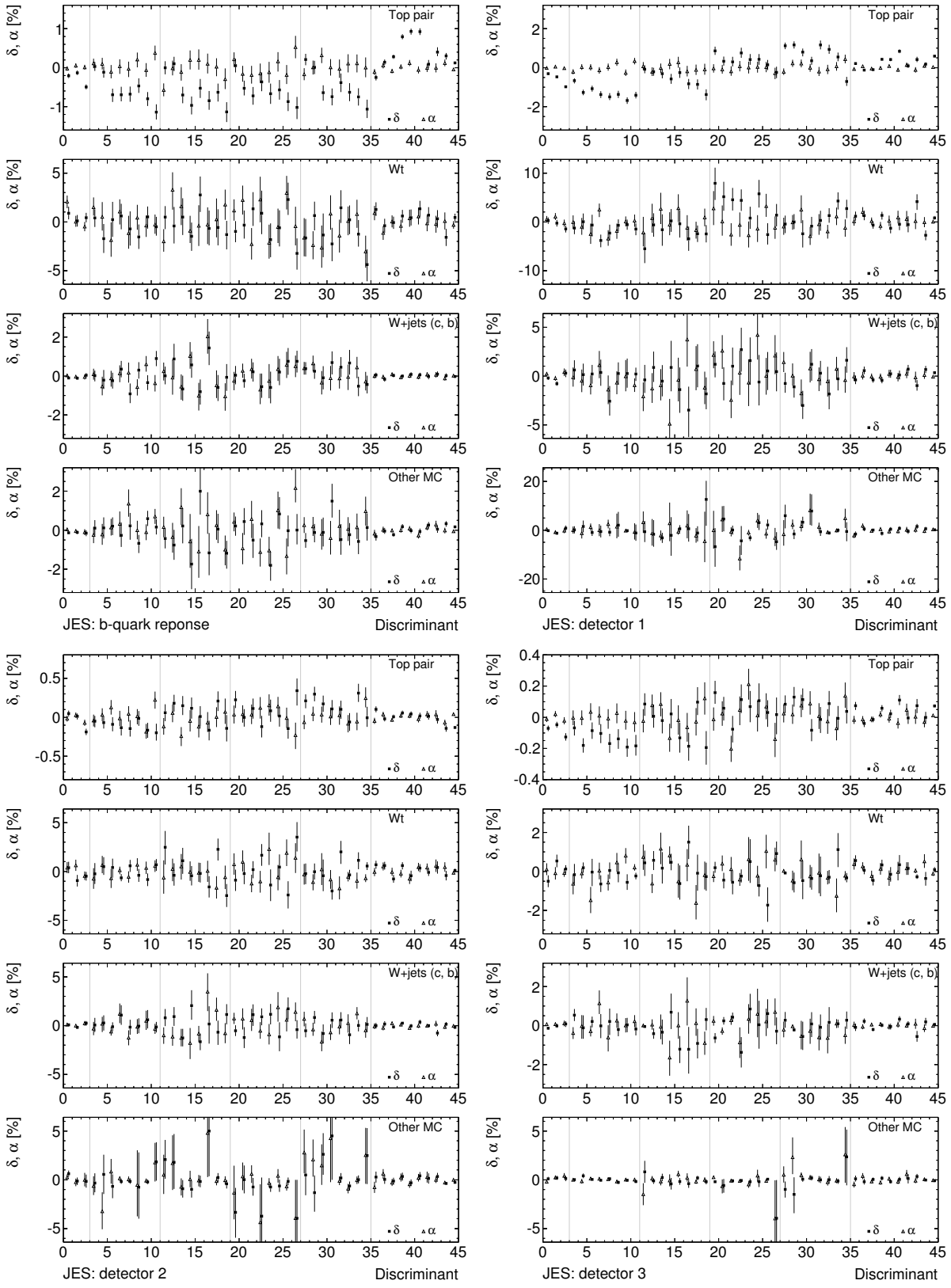

Figure F.3: Effects of the systematic uncertainties on the normalised distributions in the signal region.

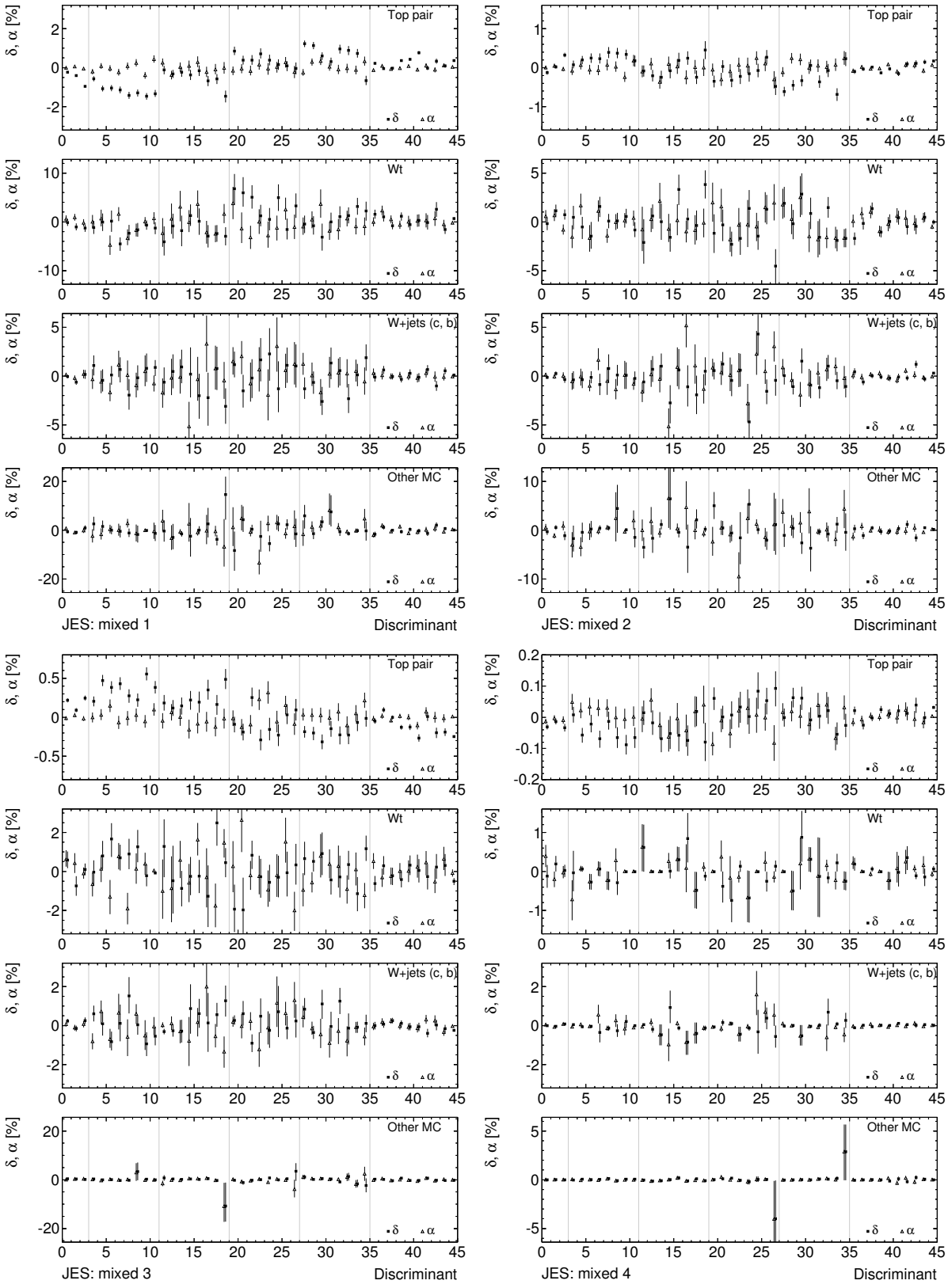

Figure F.4: Effects of the systematic uncertainties on the normalised distributions in the signal region.

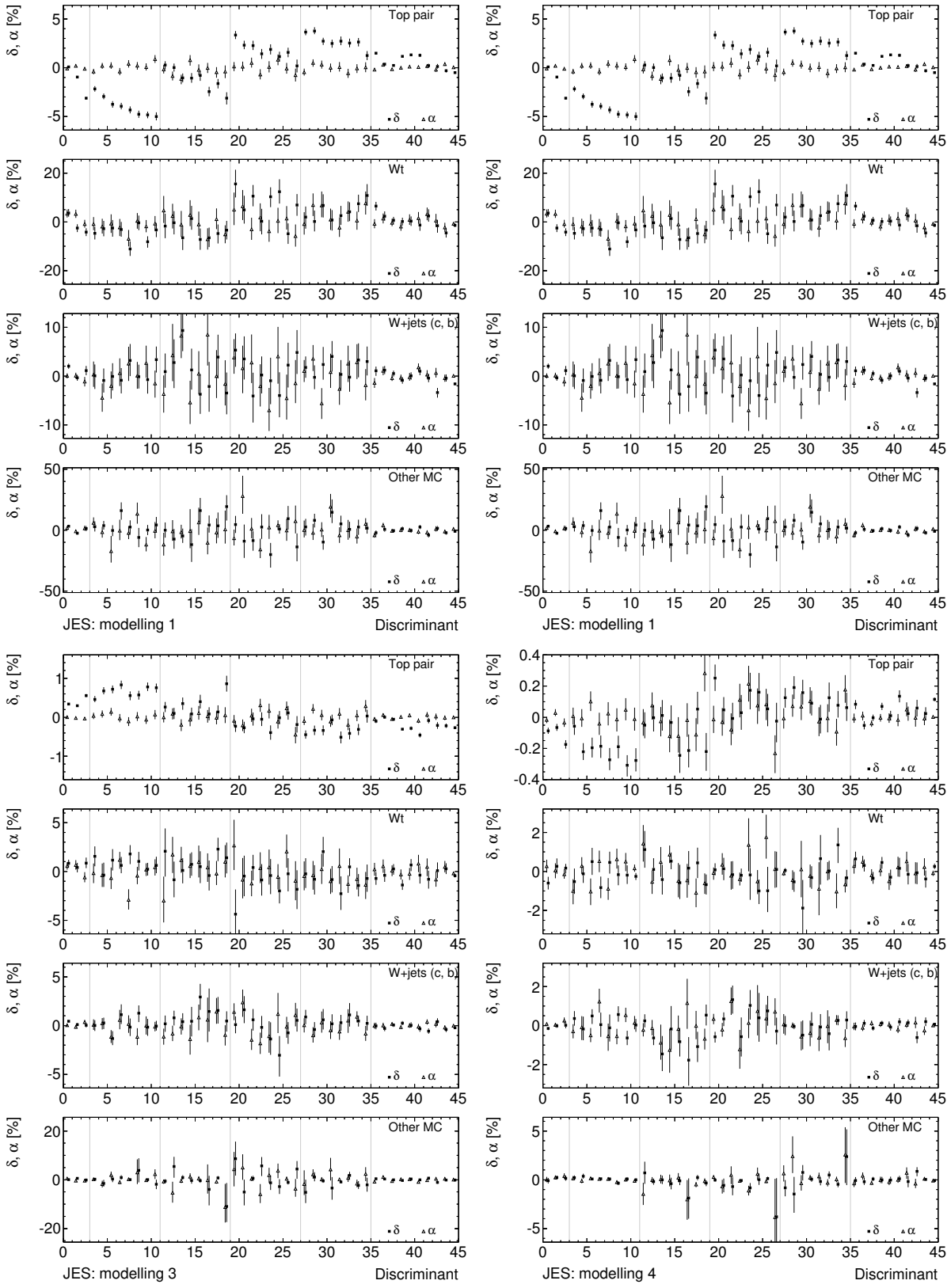

Figure F.5: Effects of the systematic uncertainties on the normalised distributions in the signal region.

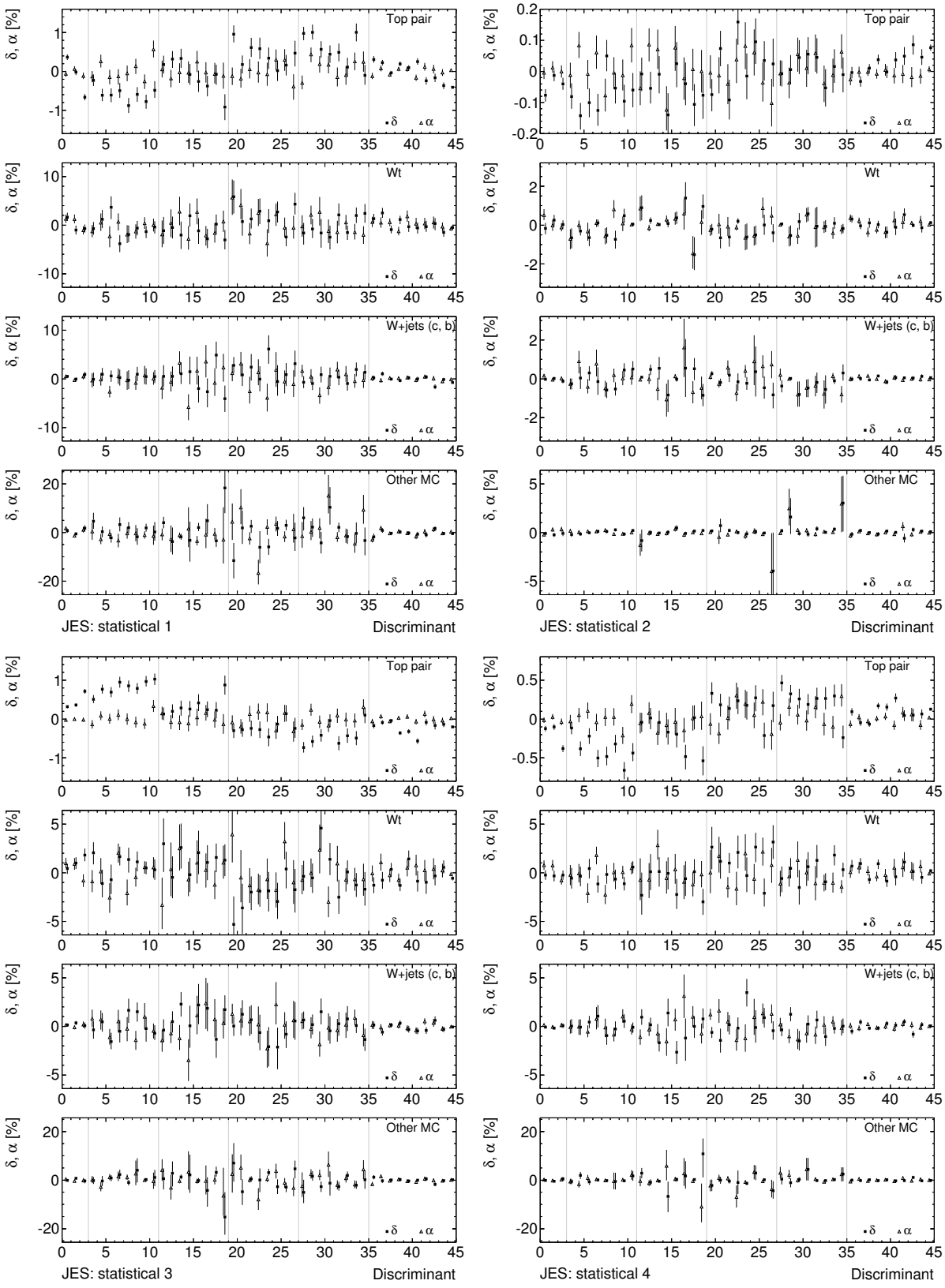

Figure F.6: Effects of the systematic uncertainties on the normalised distributions in the signal region.

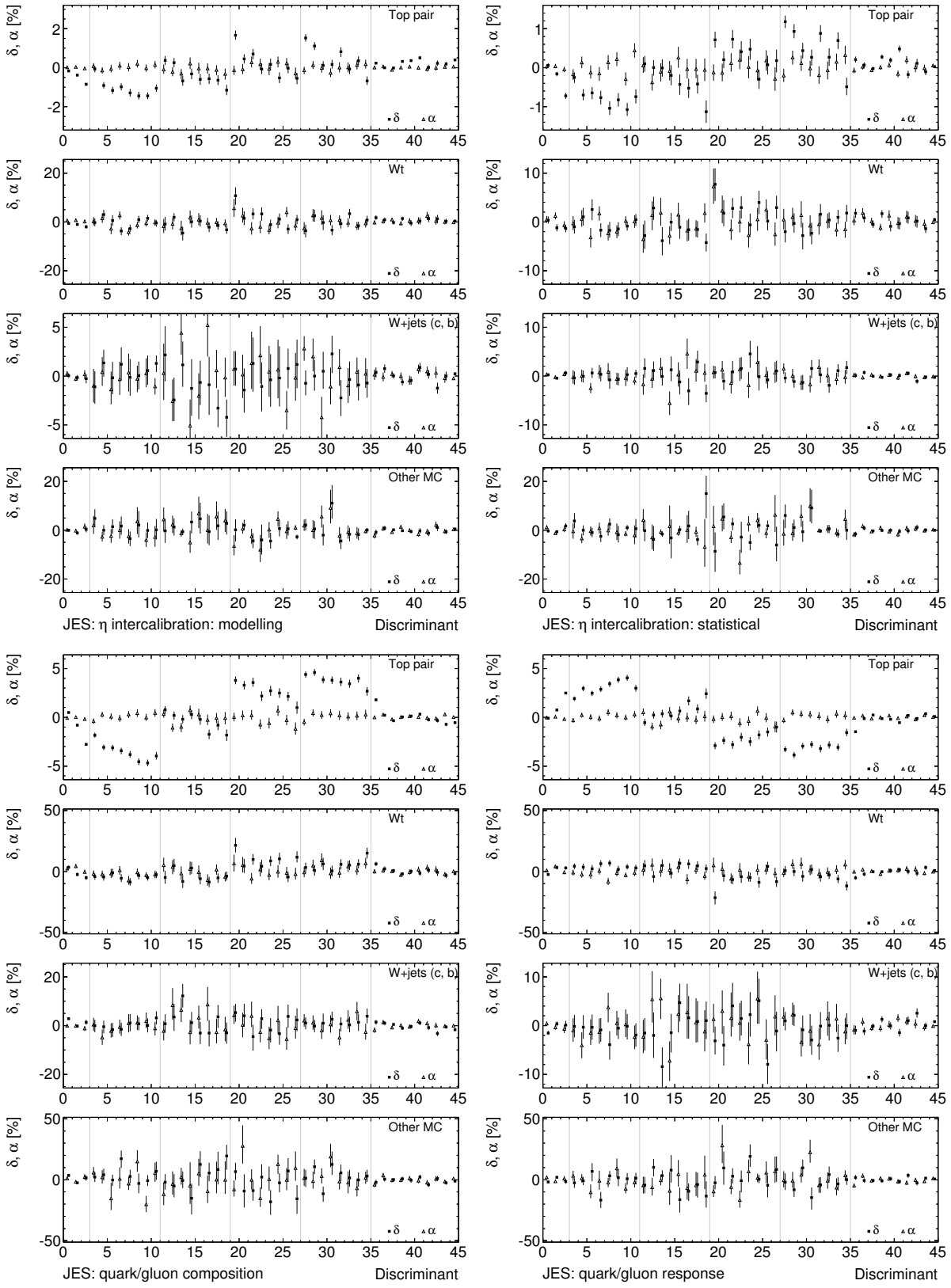

Figure F.7: Effects of the systematic uncertainties on the normalised distributions in the signal region.

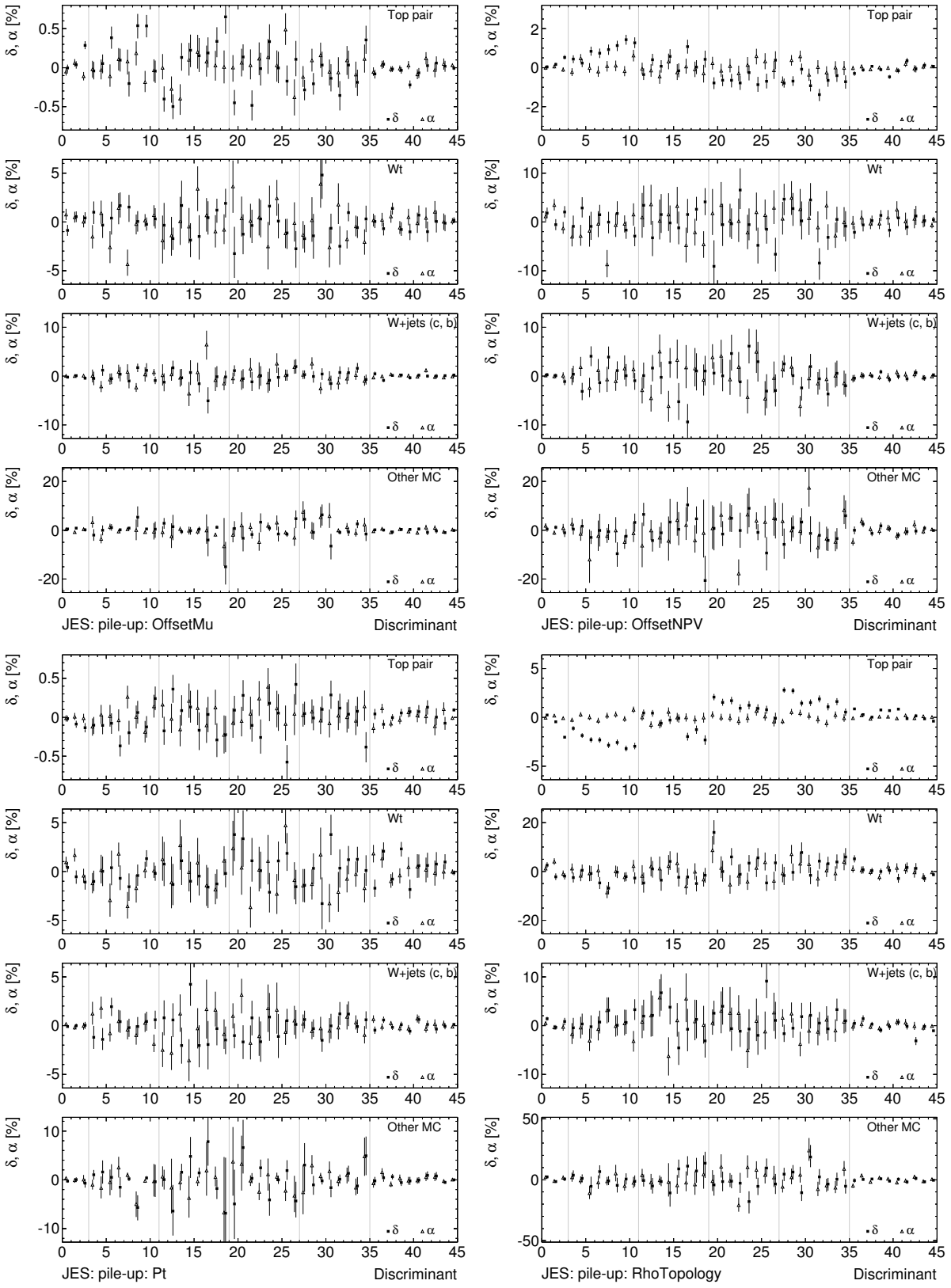

Figure F.8: Effects of the systematic uncertainties on the normalised distributions in the signal region.

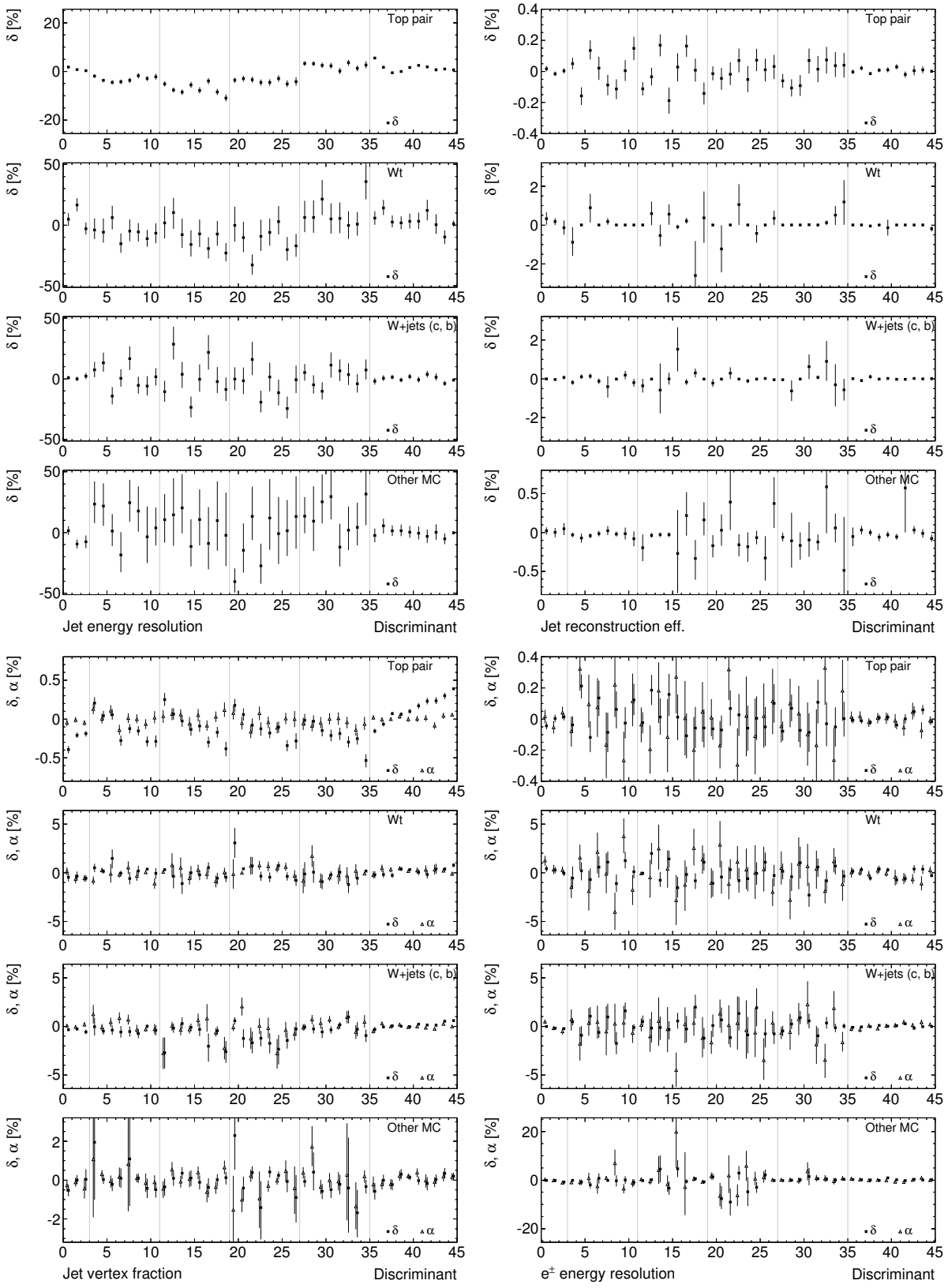

Figure F.9: Effects of the systematic uncertainties on the normalised distributions in the signal region.

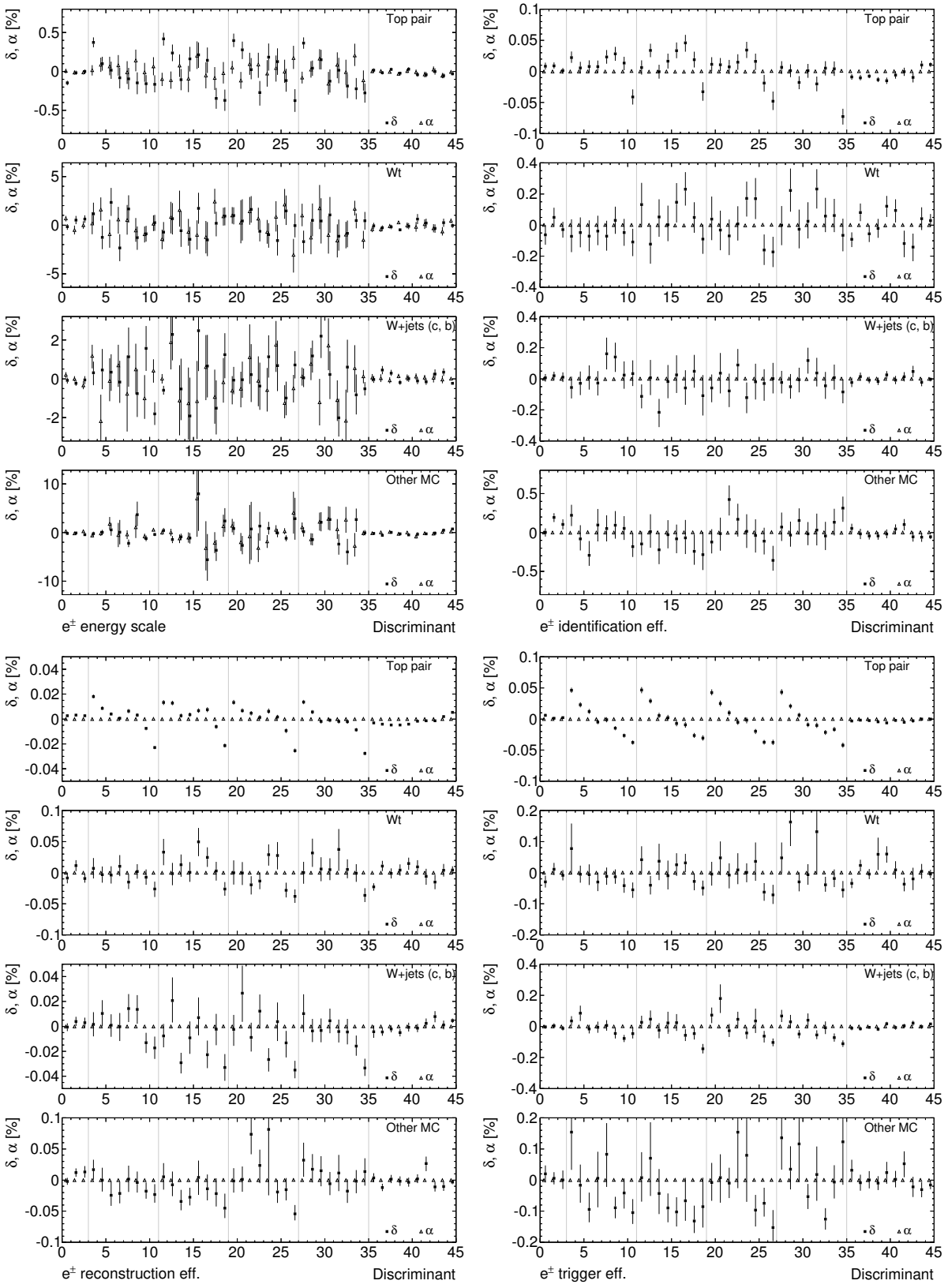

Figure F.10: Effects of the systematic uncertainties on the normalised distributions in the signal region.

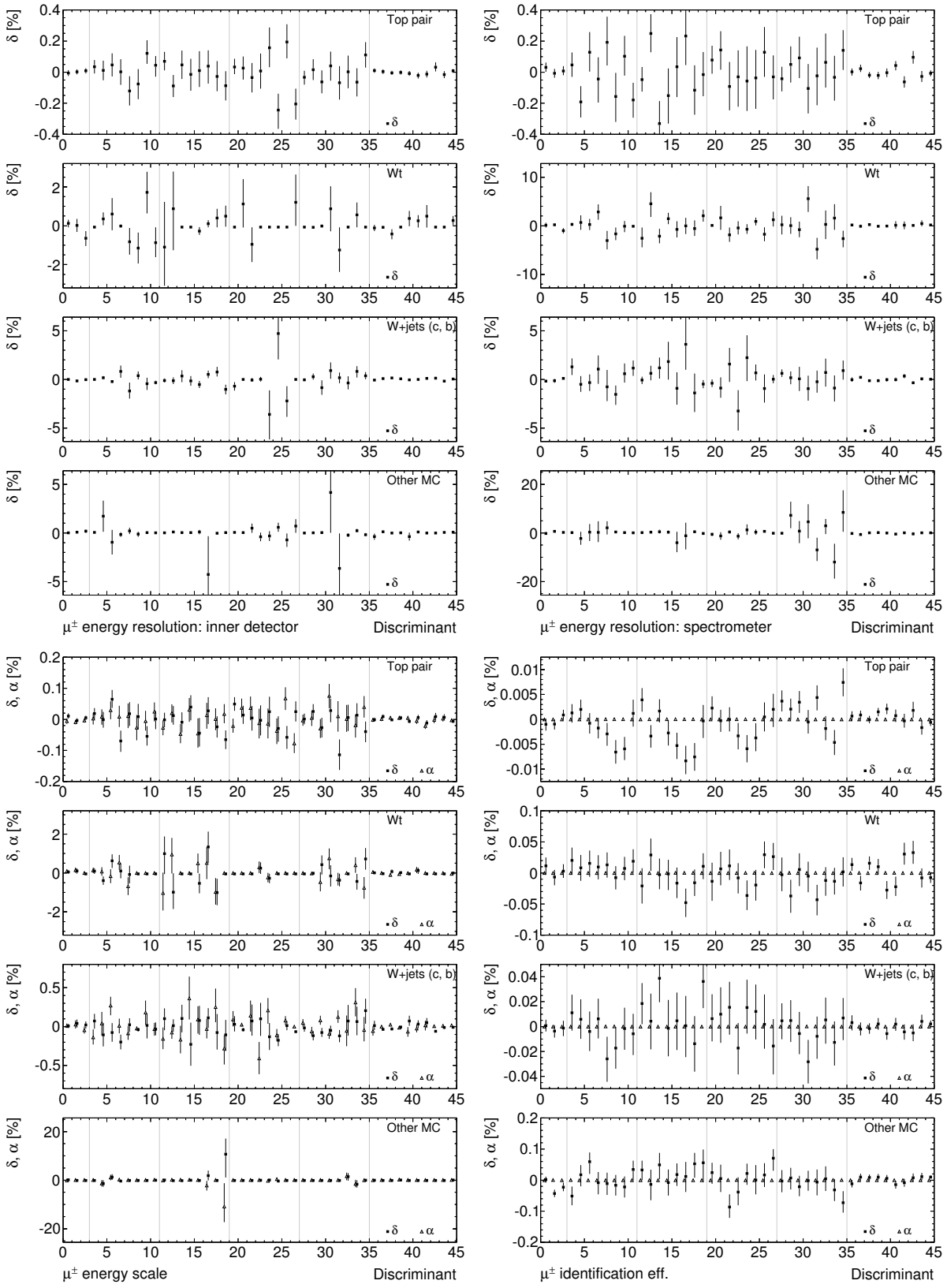

Figure F.11: Effects of the systematic uncertainties on the normalised distributions in the signal region.

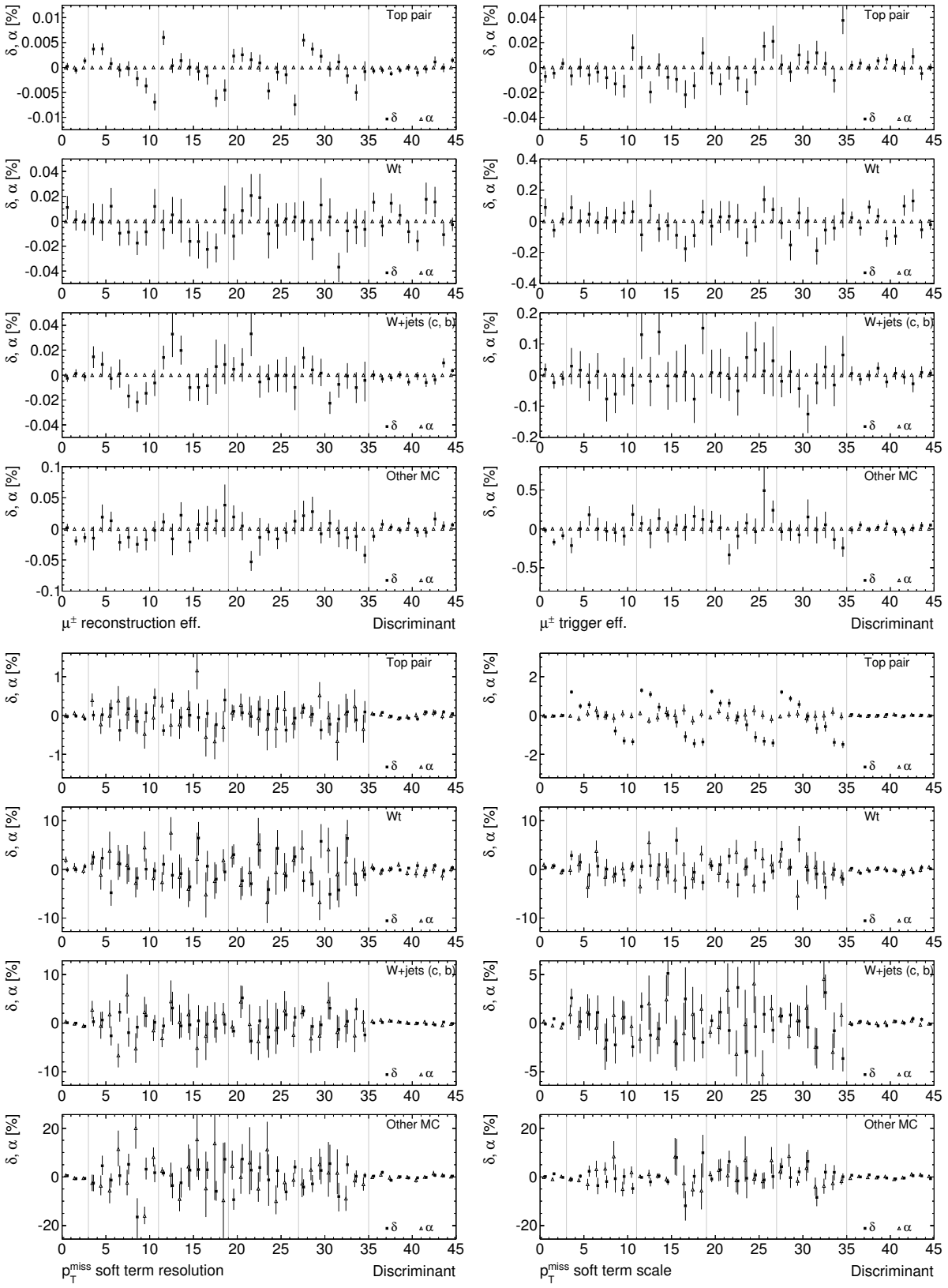

Figure F.12: Effects of the systematic uncertainties on the normalised distributions in the signal region.

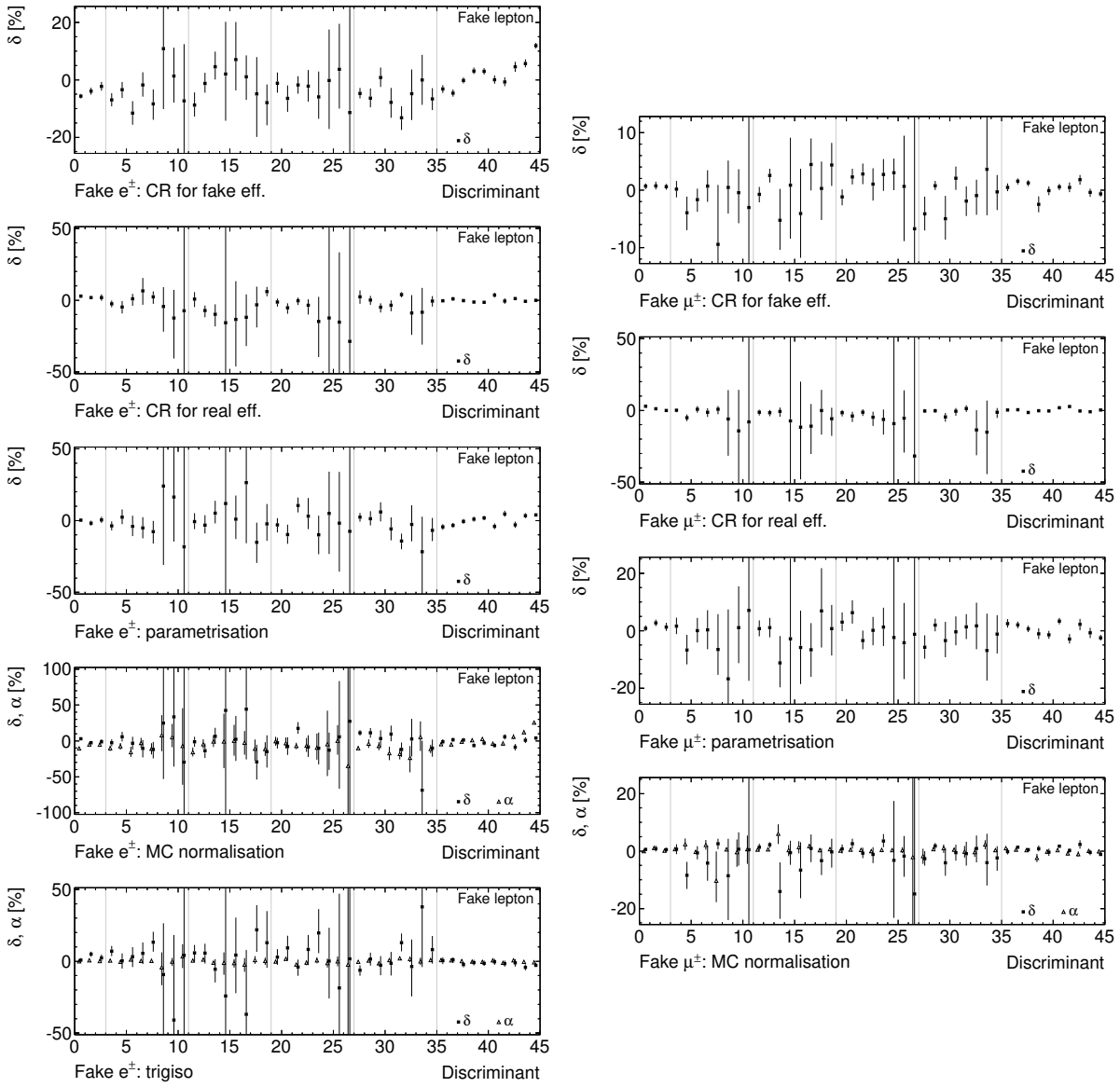

Figure F.13: Effects of the systematic uncertainties on the normalised distributions in the signal region.

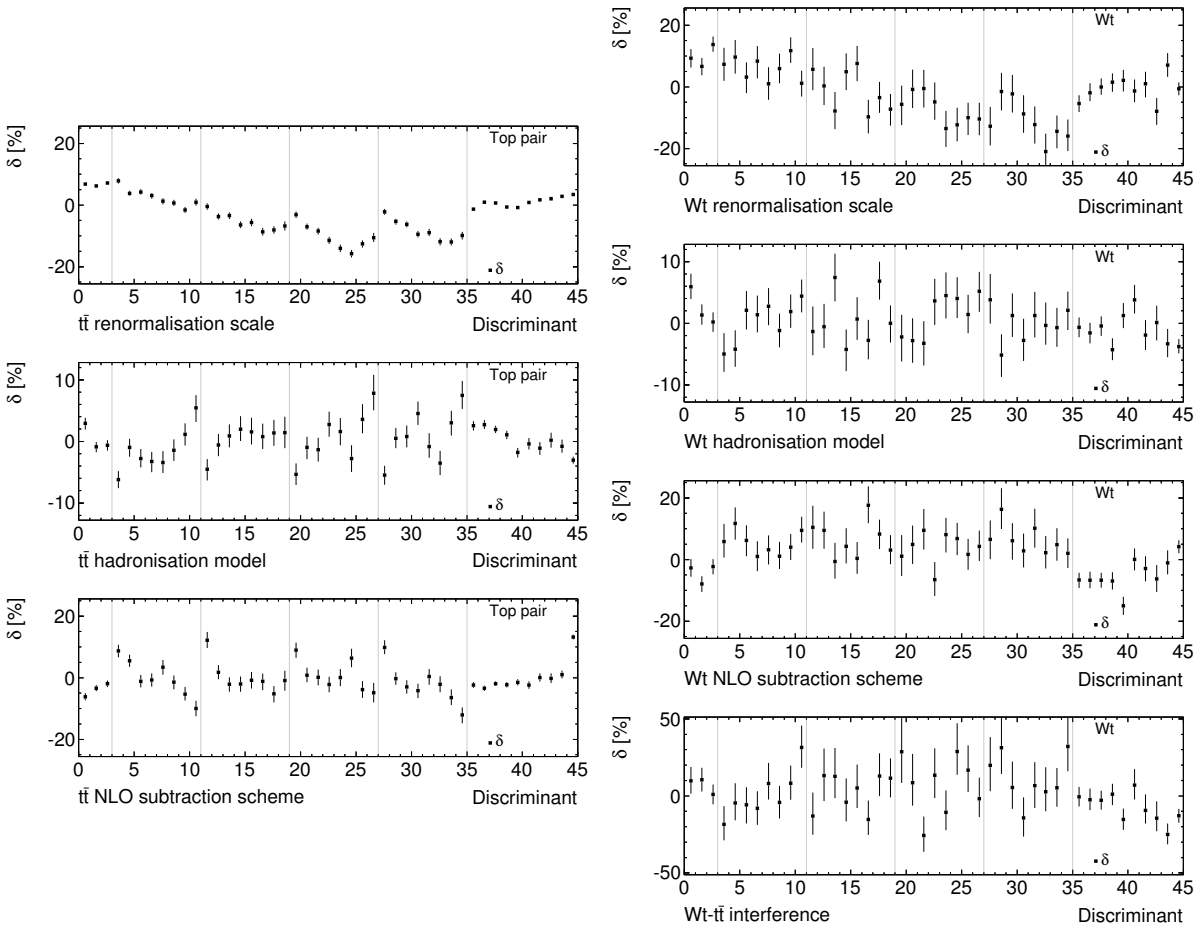

Figure F.14: Effects of the systematic uncertainties on the normalised distributions in the signal region.

APPENDIX G

Supplemental fit results

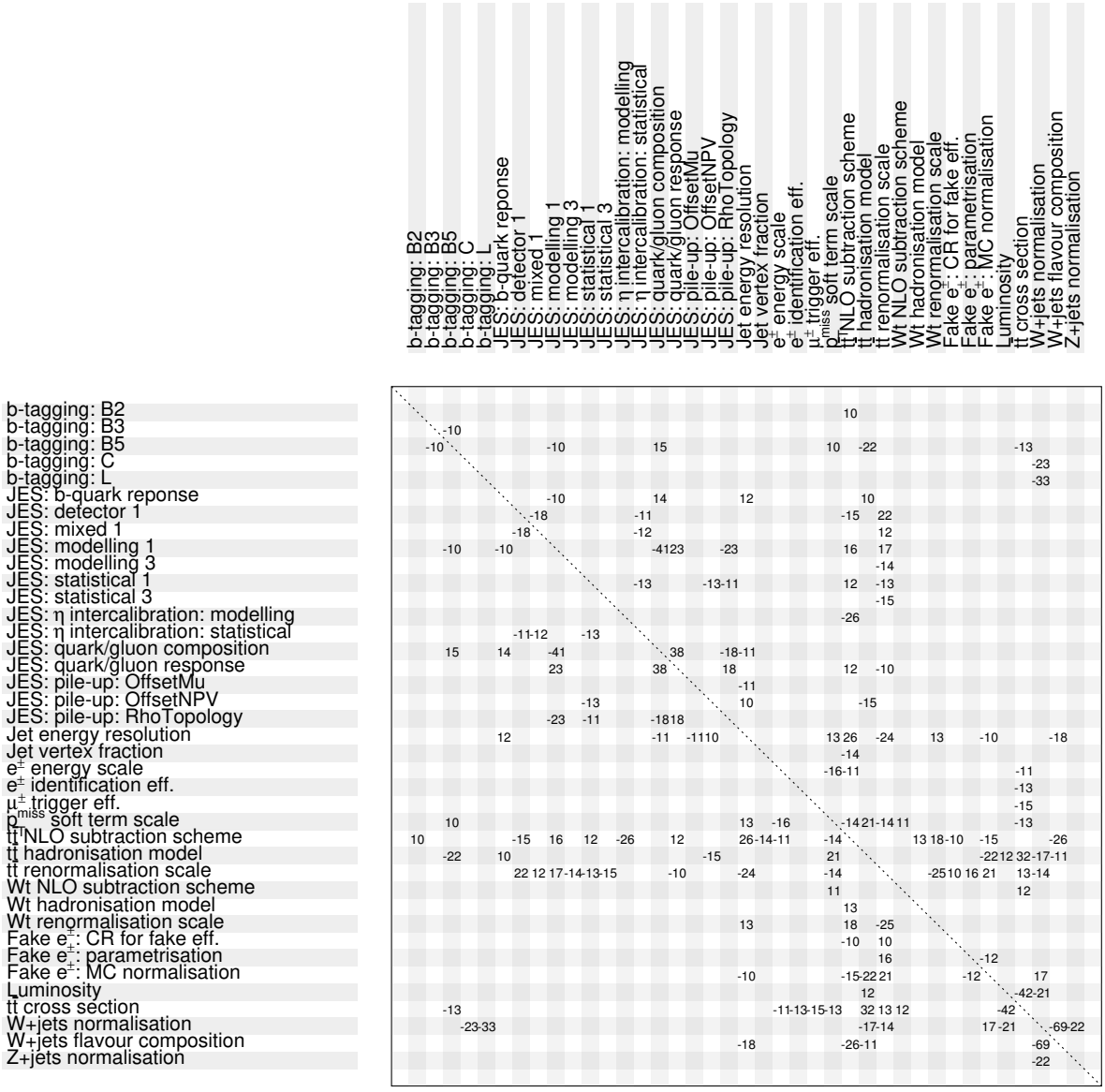

Figure G.1: Covariance matrix of the nuisance parameters after fitting the model to observed data. Only entries above 10 % are shown.

	$\hat{\theta}$	Error estimate			Impact on $\hat{\mu}$ [%]		
		Cov. mat.	Scan of Λ		Cov. mat.	Scan of Λ	
b-tagging: B0	0.04	0.99	−1.00	1.00	0.4	0.5	−0.3
b-tagging: B1	0.05	0.99	−0.99	0.99	−0.2	−0.1	0.2
b-tagging: B2	0.27	0.98	−0.99	0.98	2.6	2.3	−2.5
b-tagging: B3	0.31	0.97	−0.98	0.98	3.0	2.6	−2.9
b-tagging: B4	0.07	0.99	−1.00	1.00	1.1	0.7	−1.0
b-tagging: B5	0.72	0.90	−0.91	0.91	6.3	6.6	−6.1
b-tagging: C	0.00	0.98	−0.98	1.00	0.4	0.5	−0.3
b-tagging: L	−0.17	0.97	−0.97	0.98	3.3	3.8	−3.1
JES: b-quark reponse	−0.14	0.89	−0.90	0.90	4.9	5.0	−4.9
JES: detector 1	−0.43	0.88	−0.88	0.88	2.3	2.1	−2.5
JES: detector 2	0.00	0.99	−1.00	1.00	−0.2	−0.1	0.1
JES: detector 3	0.00	0.99	−1.00	1.00	−0.1	0.0	0.2
JES: mixed 1	0.11	0.90	−0.91	0.91	2.1	2.0	−2.3
JES: mixed 2	0.31	0.98	−0.99	0.98	−0.6	−0.4	0.7
JES: mixed 3	0.07	0.98	−0.99	0.99	−1.4	−1.3	1.6
JES: mixed 4	0.00	0.99	−1.00	1.00	0.0	0.1	0.1
JES: modelling 1	0.34	0.76	−0.78	0.76	4.7	4.6	−4.8
JES: modelling 2	0.00	0.99	−1.00	1.00	0.1	0.2	0.0
JES: modelling 3	−0.05	0.97	−0.98	0.98	−2.4	−2.2	2.5
JES: modelling 4	0.00	0.99	−1.00	1.00	0.0	0.1	0.1
JES: statistical 1	0.04	0.89	−0.89	0.89	−3.1	−3.1	3.0
JES: statistical 2	0.00	0.99	−1.00	1.00	0.0	0.0	0.1
JES: statistical 3	0.18	0.97	−0.98	0.98	−2.6	−2.5	2.8
JES: statistical 4	−0.23	0.98	−0.99	0.99	1.0	1.1	−0.5
JES: η intercalibration: modelling	0.27	0.86	−0.87	0.87	1.6	0.8	−2.4
JES: η intercalibration: statistical	−0.03	0.92	−0.92	0.93	1.1	0.4	−0.9
JES: quark/gluon composition	0.04	0.71	−0.71	0.72	−1.6	−0.7	1.7
JES: quark/gluon response	0.36	0.79	−0.80	0.80	0.5	−0.2	0.2
JES: pile-up: OffsetMu	−0.15	0.87	−0.88	0.88	−0.1	−0.1	0.1
JES: pile-up: OffsetNPV	−0.15	0.73	−0.74	0.73	1.1	0.8	−1.5
JES: pile-up: Pt	0.00	0.99	−1.00	1.00	0.0	0.1	0.1
JES: pile-up: RhoTopology	0.04	0.85	−0.85	0.85	2.7	2.2	−2.5
Jet reconstruction eff.	0.00	0.99	−1.00	1.00	0.0	0.1	0.1
Jet energy resolution	0.10	0.22	−0.22	0.22	5.2	4.0	−5.8
Jet vertex fraction	0.13	0.95	−0.95	0.96	2.8	2.3	−2.3
$e^{\pm}$ energy resolution	0.00	0.99	−1.00	1.00	−0.1	0.0	0.2
$e^{\pm}$ energy scale	−0.04	0.91	−0.91	0.91	−2.5	−2.4	2.5
$e^{\pm}$ identification eff.	−0.05	0.98	−0.99	0.99	−1.6	−1.5	1.7
$e^{\pm}$ reconstruction eff.	−0.02	0.99	−1.00	1.00	−0.1	0.0	0.2
$e^{\pm}$ trigger eff.	−0.02	0.99	−1.00	1.00	−0.2	−0.1	0.4
$\mu^{\pm}$ energy resolution: inner detector	0.00	0.99	−1.00	1.00	−0.1	0.0	0.2
$\mu^{\pm}$ energy resolution: spectrometer	0.00	0.99	−1.00	1.00	0.1	0.2	0.0
$\mu^{\pm}$ energy scale	−0.06	0.98	−0.98	0.99	−1.1	−1.1	1.2
$\mu^{\pm}$ identification eff.	0.01	0.99	−1.00	1.00	−0.5	−0.3	0.5
$\mu^{\pm}$ reconstruction eff.	0.00	0.99	−1.00	1.00	−0.2	−0.1	0.3
$\mu^{\pm}$ trigger eff.	0.07	0.98	−0.99	0.99	−1.7	−1.6	1.8
$p_{\mathrm{T}}^{\mathrm{miss}}$ soft term resolution	0.00	0.99	−1.00	1.00	−0.1	0.0	0.2
$p_{\mathrm{T}}^{\mathrm{miss}}$ soft term scale	−0.78	0.76	−0.77	0.76	5.5	5.3	−5.5
$t\bar{t}$ NLO subtraction scheme	−0.03	0.20	−0.20	0.20	6.7	7.8	−5.7
$t\bar{t}$ hadronisation model	1.03	0.35	−0.35	0.36	4.1	4.8	−3.4
$t\bar{t}$ proton structure (approx.)	0.00	0.99	−1.00	1.00	−0.3	−0.2	0.4
$t\bar{t}$ renormalisation scale	0.04	0.29	−0.28	0.29	10.8	11.2	−10.4
Wt NLO subtraction scheme	0.11	0.72	−0.71	0.75	−5.9	−8.0	2.8
Wt hadronisation model	0.43	0.91	−0.92	0.92	1.8	0.9	−2.4
Wt-$t\bar{t}$ interference	0.00	0.99	−1.00	1.00	−2.3	−2.2	2.5
Wt proton structure (approx.)	0.00	0.99	−1.00	1.00	−0.5	−0.4	0.6
Wt renormalisation scale	−0.49	0.73	−0.75	0.72	8.3	5.7	−10.3
Fake $e^{\pm}$: CR for fake eff.	0.19	0.98	−0.98	0.98	0.6	0.6	−0.5
Fake $e^{\pm}$: CR for real eff.	−0.12	1.00	−1.01	1.00	−0.6	−0.5	0.7
Fake $e^{\pm}$: parametrisation	0.19	0.97	−0.97	0.98	−0.4	−0.5	0.4
Fake $e^{\pm}$: MC normalisation	0.55	0.86	−0.87	0.85	−2.3	−2.0	2.6
Fake $e^{\pm}$: trigiso	−0.33	0.98	−0.99	0.99	−0.7	−0.6	0.8
Fake $\mu^{\pm}$: CR for fake eff.	0.02	0.99	−1.00	1.00	−0.1	0.0	0.2
Fake $\mu^{\pm}$: CR for real eff.	−0.04	1.00	−1.01	1.01	−0.1	0.2	0.1
Fake $\mu^{\pm}$: parametrisation	0.06	0.97	−0.98	0.98	−0.7	−0.6	0.8
Fake $\mu^{\pm}$: MC normalisation	0.09	0.98	−0.99	0.99	−0.1	0.0	0.2
Luminosity	0.03	0.03	−0.93	0.93	−5.0	−4.3	5.3
$t\bar{t}$ cross section	0.02	0.67	−0.67	0.68	−5.1	−4.8	5.5
s-channel cross section	0.00	0.99	−1.00	1.00	0.0	0.1	0.1
t-channel cross section	0.02	0.99	−1.00	1.00	0.1	0.2	0.0
W+jets normalisation	0.03	0.34	−0.36	0.34	−1.9	−2.1	2.1
W+jets flavour composition	0.14	0.87	−0.87	0.90	−5.2	−4.6	5.5
Z+jets normalisation	0.07	0.96	−0.95	0.99	0.3	0.4	−0.2
$WW/WZ/ZZ$ normalisation	0.00	0.99	−1.00	1.00	−0.9	−0.8	0.6

Table G.1: The estimated values of the nuisance parameters, their errors and their impact on the estimate of the signal strength after the fitting the model to observed data. The table shows the (symmetric) errors obtained with the covariance matrix as well as the (asymmetric) errors from the scan of the likelihood function.

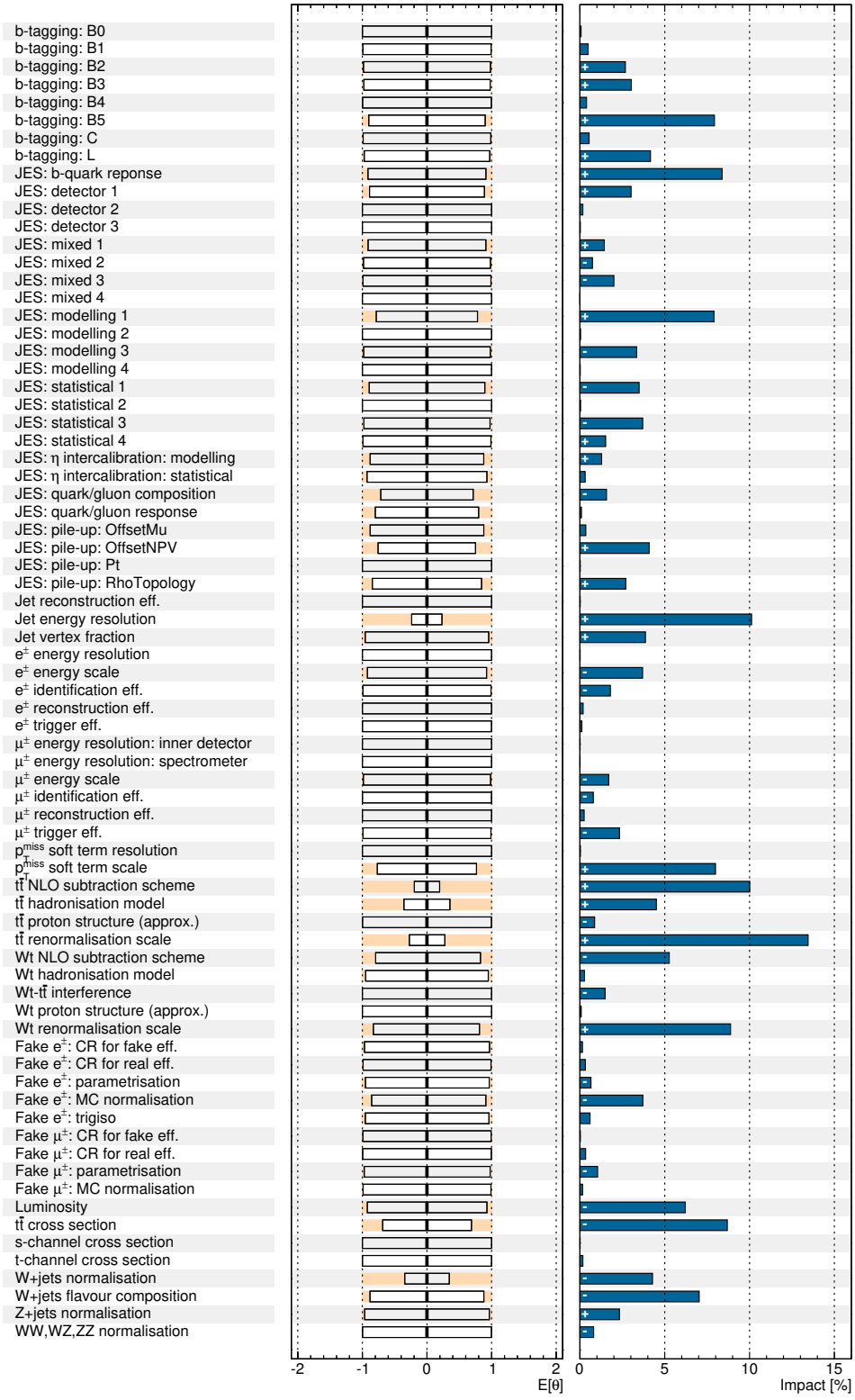

Figure G.2: The estimated values of the nuisance parameters and their errors after the fitting the model to Asimov data. The right pane shows the impact of each systematic uncertainty. The sign indicates whether the measured signal strength has a positive or a negative correlation to the nuisance parameter.

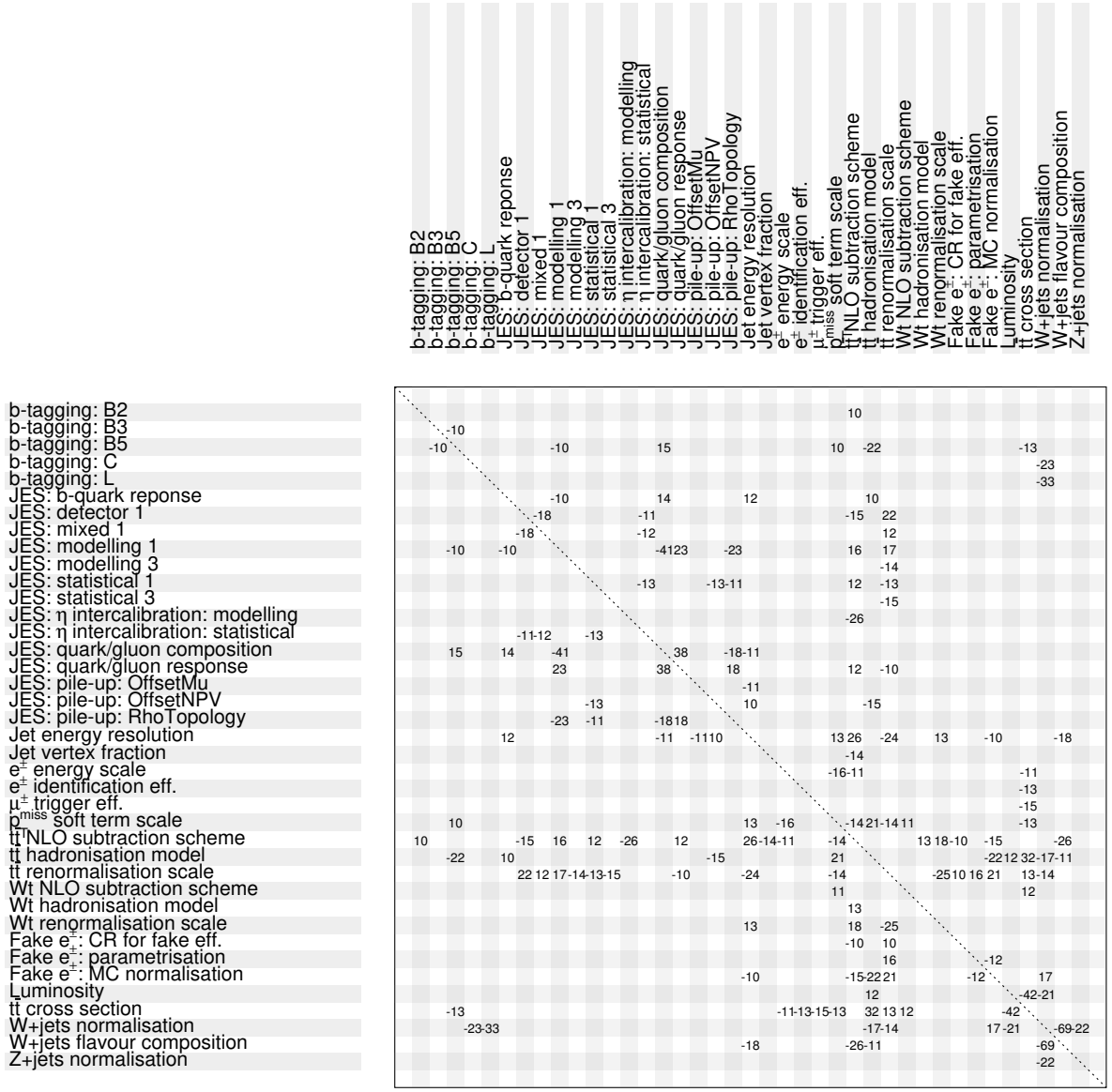

Figure G.3: Covariance matrix of the nuisance parameters after fitting the model to Asimov data. Only entries above 10 % are shown.

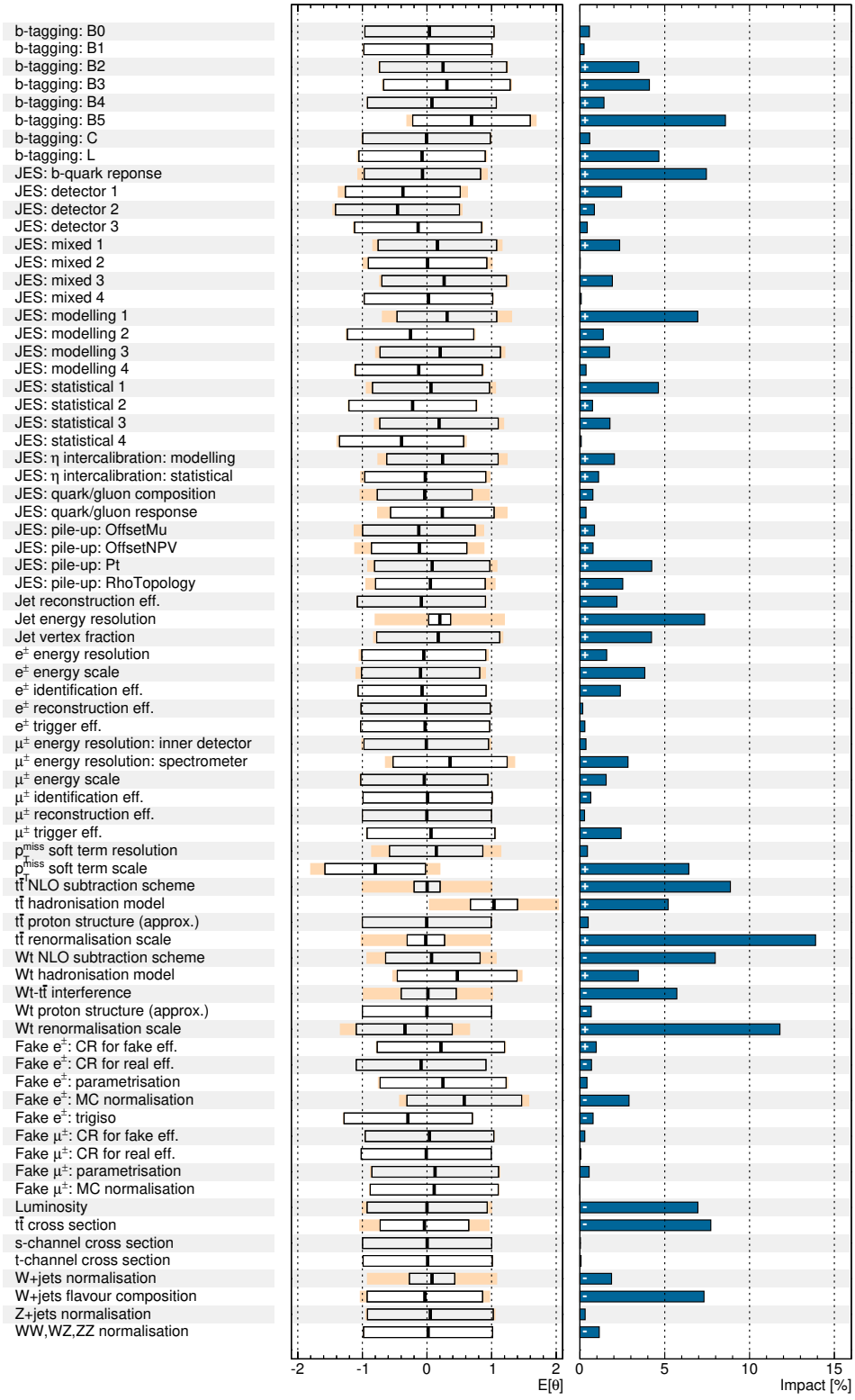

Figure G.4: The estimated values of the nuisance parameters and their errors after the fitting the model to observed data without pruning insignificant shape variations. The right pane shows the impact of each systematic uncertainty. The sign indicates whether the measured signal strength has a positive or a negative correlation to the nuisance parameter.

Bibliography

[1] F. Halzen and A. D. Martin,
Quarks and Leptons: An Introductory Course in Modern Particle Physics, 1984,
ISBN: 9780471887416 (cit. on pp. 3, 4).

[2] J. J. Sakurai, *Modern quantum mechanics*, Addison-Wesley, 1994 (cit. on p. 3).

[3] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, 1995,
ISBN: 0-201-50934-2 (cit. on p. 3).

[4] A. Zee, *Quantum Field Theory in a Nutshell*, Princeton Univ. Press, 2003 (cit. on p. 3).

[5] D. Griffiths, *Introduction to Elementary Particles*, Wiley-VCH, 2008 (cit. on p. 3).

[6] I. C. Brock, T. Schorner-Sadenius et al., *Physics at the Terascale*, Wiley-VCH, 2011
(cit. on p. 3).

[7] Particle Data Group, K. A. Olive et al., *Review of Particle Physics*,
Chin. Phys. C **38** (2014) 090001 (cit. on pp. 4, 56, 84, 98).

[8] Super-Kamiokande Collaboration, *Evidence for oscillation of atmospheric neutrinos*,
Phys. Rev. Lett. **81** (1998) 1562, arXiv:hep-ex/9807003 [hep-ex] (cit. on p. 5).

[9] SNO Collaboration, *Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^{8}B solar neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (2001) 071301,
arXiv:nucl-ex/0106015 [nucl-ex] (cit. on p. 5).

[10] Daya Bay Collaboration, *Observation of electron-antineutrino disappearance at Daya Bay*,
Phys. Rev. Lett. **108** (2012) 171803, arXiv:1203.1669 [hep-ex] (cit. on p. 5).

[11] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*,
Phys. Rev. **76** (6 Sept. 1949) 769 (cit. on p. 5).

[12] Particle Data Group, K. Nakamura et al., *Review of Particle Physics*,
J. Phys. G **37** (2010) 075021, URL: http://pdg.lbl.gov (cit. on pp. 6, 47).

[13] 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. B **716** (2012) 1,
arXiv:1207.7214 [hep-ex] (cit. on p. 6).

[14] CMS Collaboration,
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,
Phys. Lett B. **716** (2012) 30, arXiv:1207.7235 [hep-ex] (cit. on p. 6).

[15] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
Phys. Rev. Lett. **13** (1964) 321 (cit. on p. 6).

[16] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,
Phys. Rev. Lett. **13** (1964) 508 (cit. on p. 6).

[17] CMS Collaboration,
Measurement of the inclusive 3-jet production differential cross section in proton–proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range,
Eur. Phys. J. C **75** (2015) 186, arXiv:1412.1633 [hep-ex] (cit. on p. 7).

[18] H. Abramowicz et al., *Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data*, Eur. Phys. J. C **75** (2015) 580, arXiv:1506.06042 [hep-ex] (cit. on p. 8).

[19] P. Duinker, *Review of e^+e^- physics at PETRA*, Rev. Mod. Phys. **54** (2 Apr. 1982) 325, URL: http://link.aps.org/doi/10.1103/RevModPhys.54.325 (cit. on p. 10).

[20] G. P. Salam, *Towards Jetography*, Eur. Phys. J. C **67** (2010) 637, arXiv:0906.1833 [hep-ph] (cit. on p. 10).

[21] S. Catani et al., *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, Nucl. Phys. B **406** (1993) 187 (cit. on p. 11).

[22] 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] (cit. on p. 11).

[23] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph] (cit. on p. 11).

[24] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. B **641** (2006) 57, arXiv:hep-ph/0512210 (cit. on p. 11).

[25] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001, ed. by L. Evans (cit. on p. 13).

[26] M. Lamont, *The LHC's first long run*, (2013),
URL: http://cerncourier.com/cws/article/cern/54381 (cit. on p. 13).

[27] CERN, "The CERN accelerator complex", Dec. 2008,
URL: http://cds.cern.ch/record/1260465 (cit. on p. 14).

[28] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 13, 15).

[29] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 13).

[30] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 13).

[31] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002 (cit. on p. 14).

[32] TOTEM Collaboration, *The TOTEM experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007 (cit. on p. 14).

[33] ATLAS Collaboration, URL: http://atlas.ch/photos/index.html (cit. on pp. 15–18).

[34] A. Krasznahorkay,
The evolution of the Trigger and Data Acquisition System in the ATLAS experiment,
J. Phys. Conf. Ser. **523** (2014) 012019 (cit. on p. 20).

[35] TOTEM Collaboration, *Luminosity-Independent Measurement of the Proton-Proton Total Cross Section at $\sqrt{s} = 8$ TeV*, Phys. Rev. Lett. **111** (2013) 012001 (cit. on p. 20).

[36] ATLAS Collaboration, *Measurement of inclusive two-particle angular correlations in pp collisions with the ATLAS detector at the LHC*, JHEP **1205** (2012) 157, arXiv:1203.3549 [hep-ex] (cit. on p. 20).

[37] TOTEM Collaboration, *Luminosity-independent measurements of total, elastic and inelastic cross-sections at* $\sqrt{s} = 7$ TeV, Europhys. Lett. **101** (2013) 21004 (cit. on p. 20).

[38] 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] (cit. on pp. 20, 21).

[39] S. van der Meer, "Calibration of the effective beam height in the ISR", tech. rep. CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, 1968, URL: https://cds.cern.ch/record/296752 (cit. on p. 21).

[40] C. Rubbia, "Measurement of the luminosity of p−*overlinep* collider with a (generalized) Van der Meer Method", tech. rep. CERN-p$\overline{p}$-Note-38, CERN, Nov. 1977, URL: https://cds.cern.ch/record/1025746 (cit. on p. 21).

[41] S. Agostinelli et al., *GEANT4: A simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on p. 21).

[42] T. Yamanaka, *The ATLAS calorimeter simulation FastCaloSim*, J. Phys. Conf. Ser. **331** (2011) 032053 (cit. on p. 21).

[43] ATLAS collaboration, *Expected performance of the ATLAS experiment: detector, trigger and physics*, Geneva: CERN, 2009, URL: https://cds.cern.ch/record/1125884 (cit. on p. 23).

[44] ATLAS Collaboration, "Object selection and calibration, background estimations and MC samples for the winter 2013 top quark analyses with 2012 data", ATL-COM-PHYS-2013-088, Draft 0.10: CERN, July 2013, URL: https://cds.cern.ch/record/1509562 (cit. on pp. 23, 31, 45, 108).

[45] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector and its Performance in 2012*, ATLAS-CONF-2014-047, 2014, URL: http://cdsweb.cern.ch/record/1741021 (cit. on p. 23).

[46] ATLAS Collaboration, *Study of alignment-related systematic effects on the ATLAS Inner Detector track reconstruction*, ATLAS-CONF-2012-141, 2012, URL: http://cdsweb.cern.ch/record/1483518 (cit. on p. 23).

[47] ATLAS Collaboration, *Muon reconstruction efficiency and momentum resolution of the ATLAS experiment in proton–proton collisions at* $\sqrt{s} = 7$ *TeV in 2010*, Eur. Phys. J. C **74** (2014) 3034, arXiv:1404.4562 [hep-ex] (cit. on p. 24).

[48] ATLAS Collaboration, *Calibration of the ATLAS Transition Radiation Tracker*, ATLAS-CONF-2011-006, 2011, URL: http://cdsweb.cern.ch/record/1330712 (cit. on p. 23).

[49] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Journal of Fluids Engineering **82** (1960) 35 (cit. on p. 24).

[50] T. Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*, J. Phys. Conf. Ser. **119** (2008) 032014 (cit. on p. 25).

[51] T. Cornelissen et al.,
Single track performance of the inner detector new track reconstruction (NEWT), (2008) (cit. on p. 25).

[52] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in High Pile-Up LHC Environment*, ATLAS-CONF-2012-042, 2012, URL: http://cdsweb.cern.ch/record/1435196 (cit. on p. 25).

[53] G. Piacquadio, K. Prokofiev and A. Wildauer,
Primary vertex reconstruction in the ATLAS experiment at LHC,
J. Phys. Conf. Ser. **119** (2008) 032033 (cit. on p. 25).

[54] 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] (cit. on pp. 25, 26, 29).

[55] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, URL: http://cds.cern.ch/record/1099735 (cit. on p. 26).

[56] 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] (cit. on p. 26).

[57] 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, URL: http://cdsweb.cern.ch/record/2044941 (cit. on p. 26).

[58] ATLAS Collaboration, *Jet global sequential corrections with the ATLAS detector in proton–proton collisions at* $\sqrt{s} = 8$ *TeV*, ATLAS-CONF-2015-002, 2015, URL: http://cdsweb.cern.ch/record/2001682 (cit. on p. 26).

[59] ATLAS Collaboration, *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at* $\sqrt{s} = 8$ *TeV*, ATLAS-CONF-2015-017, 2015, URL: http://cdsweb.cern.ch/record/2008678 (cit. on p. 27).

[60] ATLAS Collaboration, *Characterisation and mitigation of beam-induced backgrounds observed in the ATLAS detector during the 2011 proton–proton run*, JINST **8** (2013) P07004, arXiv:1303.0223 [hep-ex] (cit. on p. 27).

[61] ATLAS Collaboration, *Selection of jets produced in proton–proton collisions with the ATLAS detector using 2011 data*, ATLAS-CONF-2012-020, 2012, URL: http://cdsweb.cern.ch/record/1430034 (cit. on p. 27).

[62] ATLAS Collaboration, *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083, 2013, URL: http://cdsweb.cern.ch/record/1570994 (cit. on p. 28).

[63] ATLAS Collaboration, *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046, 2014, URL: http://cdsweb.cern.ch/record/1741020 (cit. on pp. 28, 29).

[64] ATLAS Collaboration, *Commissioning of the ATLAS high performance b-tagging algorithms in the* 7 *TeV collision data*, ATLAS-CONF-2011-102, 2011, URL: http://cdsweb.cern.ch/record/1369219 (cit. on p. 28).

[65] G. Piacquadio and C. Weiser,
"A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS",
Journal of Physics: Conference Series, vol. 119, 3, IOP Publishing, 2008 032032
(cit. on p. 28).

[66] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004, 2014, URL: http://cdsweb.cern.ch/record/1664335 (cit. on p. 29).

[67] P. Seema,
"Search for single top-quark production in the *Wt*-channel with 1 lepton and 2 jets at ATLAS",
BONN-IB-2013-01, Rheinische Friedrich-Wilhelms-Universität Bonn, 2013, URL: https://www.brock.physik.uni-bonn.de/research/atlas-experiment/publications-1
(cit. on p. 29).

[68] ATLAS Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton–proton collision data*, Eur. Phys. J. C **72** (2012) 1909,
arXiv:1110.3174 [hep-ex] (cit. on p. 30).

[69] ATLAS Collaboration, *Expected electron performance in the ATLAS experiment*,
ATL-PHYS-PUB-2011-006, 2011, URL: http://cdsweb.cern.ch/record/1345327
(cit. on p. 30).

[70] ATLAS Collaboration,
Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data,
Eur. Phys. J. C **74** (2014) 3071, arXiv:1407.5063 [hep-ex] (cit. on p. 31).

[71] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton–proton collision data*, ATLAS-CONF-2014-032, 2014,
URL: http://cdsweb.cern.ch/record/1706245 (cit. on p. 31).

[72] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*,
Eur. Phys. J. C **74** (2014) 3130, arXiv:1407.3935 [hep-ex] (cit. on pp. 31, 32).

[73] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction in proton–proton collisions at* $\sqrt{s} = 7$ *TeV with ATLAS*, Eur. Phys. J. C **72** (2012) 1844,
arXiv:1108.5602 [hep-ex] (cit. on p. 33).

[74] T. Chwalek, "Measurement of the W boson helicity in top-antitop quark events with the CDF II experiment", 2006,
URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?masters-2006-04 (cit. on p. 34).

[75] P. Sturm, "Measurement of the *t*-Channel Single Top-Quark Production Cross-Section with the ATLAS Detector at $\sqrt{s} = 7$ TeV", 2012, URL: https://cds.cern.ch/record/1501083
(cit. on p. 34).

[76] ATLAS Collaboration, *Comprehensive measurements of t-channel single top-quark production cross sections at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, Phys. Rev. D **90** (2014) 112006,
arXiv:1406.7844 [hep-ex] (cit. on p. 34).

[77] ATLAS Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton–Proton Collisions recorded in 2012 at* $\sqrt{s} = 8$ *TeV*,
ATLAS-CONF-2013-082, 2013, URL: http://cdsweb.cern.ch/record/1570993
(cit. on p. 34).

[78] S. W. Herb et al.,
Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions,
Phys. Rev. Lett. **39** (1977) 252 (cit. on p. 35).

[79] F. Abe et al., *Observation of top quark production in $\bar{p}p$ collisions*,
Phys. Rev. Lett. **74** (1995) 2626, arXiv:hep-ex/9503002 [hep-ex] (cit. on p. 35).

[80] D0 Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632,
arXiv:hep-ex/9503003 [hep-ex] (cit. on p. 35).

[81] ATLAS Collaboration,
Summary of Standard Model total production cross section measurements, Nov. 2015,
URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/ATLAS_c_SMSummary_TotalXsect_rotated/history.html (visited on 12/03/2016)
(cit. on p. 36).

[82] V. del Duca and E. Laenen, *Top physics at the LHC*, Int. J. Mod. Phys. A **30** (2015),
arXiv:1510.06690 [hep-ph] (cit. on p. 35).

[83] ATLAS, CDF, CMS and D0 Collaborations,
First combination of Tevatron and LHC measurements of the top-quark mass, (2014),
arXiv:1403.4427 [hep-ex] (cit. on p. 35).

[84] M. Czakon, P. Fiedler and A. Mitov,
Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$,
Phys. Rev. Lett. **110** (2013) 252004, arXiv:1303.6254 [hep-ph] (cit. on p. 37).

[85] M. Botje et al., *The PDF4LHC Working Group Interim Recommendations*, (2011),
arXiv:1101.0538 [hep-ph] (cit. on p. 37).

[86] ATLAS and CMS Collaborations, *Combination of ATLAS and CMS top quark pair cross section measurements in the $e\mu$ final state using proton–proton collisions at $\sqrt{s} = 8$ TeV*,
ATLAS-CONF-2014-054, 2014, URL: http://cdsweb.cern.ch/record/1951322
(cit. on p. 37).

[87] N. Kidonakis, *NNLL resummation for s-channel single top quark production*,
Phys. Rev. D **81** (2010) 054028, arXiv:1001.5034 [hep-ph] (cit. on p. 37).

[88] N. Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. D **83** (2011) 091503, arXiv:1103.2792 [hep-ph]
(cit. on p. 37).

[89] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-*, Phys. Rev. D **82** (2010) 054018, arXiv:1005.4451 [hep-ph] (cit. on p. 37).

[90] CDF Collaboration, *First Observation of Electroweak Single Top Quark Production*,
Phys. Rev. Lett. **103** (2009) 092002, arXiv:0903.0885 [hep-ex] (cit. on p. 37).

[91] D0 Collaboration, *Observation of Single Top Quark Production*,
Phys. Rev. Lett. **103** (2009) 092001, arXiv:0903.0850 [hep-ex] (cit. on p. 37).

[92] CMS Collaboration, *Evidence for associated production of a single top quark and W boson in pp collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. Lett. **110** (2013) 022003,
arXiv:1209.3489 [hep-ex] (cit. on p. 37).

[93] ATLAS Collaboration,
Evidence for the associated production of a W boson and a top quark in ATLAS at $\sqrt{s} = 7$ TeV,
Phys. Lett. B **716** (2012) 142, arXiv:1205.5764 [hep-ex] (cit. on p. 37).

[94] CMS Collaboration, *Observation of the associated production of a single top quark and a W boson in pp collisions at* $\sqrt{s} = 8$ *TeV*, Phys. Rev. Lett. **112** (2014) 231802, arXiv:1401.2942 [hep-ex] (cit. on p. 37).

[95] ATLAS Collaboration, *Measurement of the production cross-section of a single top quark in association with a W boson at* 8 *TeV with the ATLAS experiment*, JHEP **1601** (2016) 064, arXiv:1510.03752 [hep-ex] (cit. on pp. 37, 101).

[96] ATLAS Collaboration, *Search for the production of single vector-like and excited quarks in the Wt final state in pp collisions at* $\sqrt{s} = 8$ *TeV with the ATLAS detector*, (2015), arXiv:1510.02664 [hep-ex] (cit. on pp. 38, 101).

[97] ATLAS Collaboration, *Search for top squarks in final states with one isolated lepton, jets, and missing transverse momentum in* $\sqrt{s} = 13$ *TeV pp collisions of ATLAS data*, ATLAS-CONF-2016-007, 2016, URL: https://cds.cern.ch/record/2139641 (cit. on p. 38).

[98] T. M. P. Tait and C. Yuan, *Single top quark production as a window to physics beyond the standard model*, Phys. Rev. D **63** (2000) 014018, arXiv:hep-ph/0007298 [hep-ph] (cit. on p. 38).

[99] Q.-H. Cao, J. Wudka and C. Yuan, *Search for new physics via single top production at the LHC*, Phys. Lett. B **658** (2007) 50, arXiv:0704.2809 [hep-ph] (cit. on pp. 38, 104).

[100] T. M. P. Tait, *The* tW^- *mode of single top production*, Phys. Rev. D **61** (1999) 034001, arXiv:hep-ph/9909352 [hep-ph] (cit. on p. 39).

[101] J. M. Campbell and F. Tramontano, *Next-to-leading order corrections to Wt production and decay*, Nucl. Phys. B **726** (2005) 109, arXiv:hep-ph/0506289 [hep-ph] (cit. on p. 39).

[102] S. Frixione et al., *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, arXiv:0805.3067 [hep-ph] (cit. on p. 39).

[103] C. D. White et al., *Isolating Wt production at the LHC*, JHEP **0911** (2009) 074, arXiv:0908.0631 [hep-ph] (cit. on p. 39).

[104] R. Frederix, *Top Quark Induced Backgrounds to Higgs Production in the* $WW^{(*)} \to llvv$ *Decay Channel at Next-to-Leading-Order in QCD*, Phys. Rev. Lett. **112** (2014) 082002, arXiv:1311.4893 [hep-ph] (cit. on p. 39).

[105] S. Gangal, M. Stahlhofen and F. J. Tackmann, *Rapidity-Dependent Jet Vetoes*, Phys. Rev. D **91** (2015) 054023, arXiv:1412.4792 [hep-ph] (cit. on p. 40).

[106] A. Buckley and C. Pollard, *QCD-aware partonic jet clustering for truth-jet flavour labelling*, Eur. Phys. J. C **76** (2016) 71, arXiv:1507.00508 [hep-ph] (cit. on p. 40).

[107] T. Golling et al., *The ATLAS Data Quality Defect Database System*, Eur. Phys. J. C **72** (2012) 1960, arXiv:1110.6119 [physics.ins-det] (cit. on p. 44).

[108] ATLAS Collaboration, *Performance of the ATLAS Liquid Argon Calorimeter after three years of LHC operation and plans for a future upgrade*, JINST **9** (2014) C09007 (cit. on p. 44).

[109] ATLAS Collaboration, *Performance of the ATLAS muon trigger in pp collisions at* $\sqrt{s} = 8$ *TeV*, Eur. Phys. J. C **75** (2015) 120, arXiv:1408.3179 [hep-ex] (cit. on p. 45).

[110] S. Frixione and B. R. Webber,
Matching NLO QCD computations and parton shower simulations, JHEP **06** (2002) 029, arXiv:hep-ph/0204244 (cit. on p. 47).

[111] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **0411** (2004) 040, arXiv:hep-ph/0409146 [hep-ph] (cit. on p. 47).

[112] S. Frixione, P. Nason and C. Oleari,
Matching NLO QCD computations with Parton Shower simulations: the POWHEG method, JHEP **0711** (2007) 070, arXiv:0709.2092 [hep-ph] (cit. on p. 47).

[113] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **1006** (2010) 043, arXiv:1002.2581 [hep-ph] (cit. on p. 47).

[114] H.-L. Lai et al., *New parton distributions for collider physics*, Phys. Rev. D **82** (2010) 074024, arXiv:1007.2241 [hep-ph] (cit. on p. 47).

[115] G. Corcella et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **0101** (2001) 010, arXiv:hep-ph/0011363 [hep-ph] (cit. on p. 47).

[116] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605** (2006) 026, arXiv:hep-ph/0603175 (cit. on p. 47).

[117] ATLAS Collaboration, *New ATLAS event generator tunes to 2010 data*, ATL-PHYS-PUB-2011-008, 2011, URL: http://cds.cern.ch/record/1345343 (cit. on p. 47).

[118] J. Butterworth, J. R. Forshaw and M. Seymour,
Multiparton interactions in photoproduction at HERA, Z. Phys. C **72** (1996) 637, arXiv:hep-ph/9601371 [hep-ph] (cit. on p. 47).

[119] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. D **82** (2010) 074018, arXiv:1005.3457 [hep-ph] (cit. on p. 47).

[120] P. M. Nadolsky et al., *Implications of CTEQ global analysis for collider observables*, Phys. Rev. D **78** (2008) 013004, arXiv:0802.0007 [hep-ph] (cit. on p. 47).

[121] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph] (cit. on p. 47).

[122] S. Höche et al., *QCD matrix elements and truncated showers*, JHEP **0905** (2009) 053, arXiv:0903.1219 [hep-ph] (cit. on p. 47).

[123] S. Schumann and F. Krauss,
A Parton shower algorithm based on Catani-Seymour dipole factorisation, JHEP **0803** (2008) 038, arXiv:0709.1027 [hep-ph] (cit. on p. 47).

[124] M. L. Mangano et al.,
ALPGEN, a generator for hard multiparton processes in hadronic collisions, JHEP **0307** (2003) 001, arXiv:hep-ph/0206293 [hep-ph] (cit. on p. 47).

[125] P. Golonka and Z. Was,
PHOTOS *Monte Carlo: A Precision tool for QED corrections in Z and W decays*, Eur. Phys. J. C **45** (2006) 97, arXiv:hep-ph/0506026 [hep-ph] (cit. on p. 47).

[126] S. Jadach, J. H. Kühn and Z. Was,
TAUOLA - *a library of Monte Carlo programs to simulate decays of polarized τ leptons*,
Computer Physics Communications **64** (1991) 275 (cit. on p. 47).

[127] ATLAS Collaboration,
Study of correlation of PDF uncertainty in single top and top pair production at the LHC,
ATL-PHYS-PUB-2015-010, 2015, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-010
(cit. on p. 47).

[128] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions to ATLAS measurements of top pair production at 7 TeV*, ATL-PHYS-PUB-2015-002, 2015,
URL: http://cdsweb.cern.ch/record/1981319 (cit. on p. 48).

[129] B. Cooper et al.,
Importance of a consistent choice of α_S in the matching of AlpGen and Pythia,
Eur. Phys. J. C **72** (2012) 2078, arXiv:1109.5295 [hep-ph] (cit. on p. 48).

[130] ATLAS Collaboration,
"Single Boson and Diboson Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV",
ATL-COM-PHYS-2010-695, CERN, Aug. 2010,
URL: https://cds.cern.ch/record/1287902 (cit. on p. 49).

[131] ATLAS Collaboration, *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector*,
ATLAS-CONF-2014-058, 2014, URL: http://cdsweb.cern.ch/record/1951336
(cit. on p. 50).

[132] S. D. Ellis, R. Kleiss and W. J. Stirling, *W's, Z's and Jets*, Phys. Lett. B **154** (1985) 435
(cit. on p. 53).

[133] F. A. Berends et al., *Multi - Jet Production in W, Z Events at $p\bar{p}$ Colliders*,
Phys. Lett. B **224** (1989) 237 (cit. on p. 53).

[134] J. Neyman and E. S. Pearson,
On the problem of the most efficient tests of statistical hypotheses,
Philosophical Transactions of the Royal Society A **231** (1933) 289 (cit. on p. 70).

[135] M. Feindt, *A Neural Bayesian Estimator for Conditional Probability Densities*, (2004),
arXiv:physics/0402093 [physics.data-an] (cit. on pp. 70, 72).

[136] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*,
Nucl. Instrum. Meth. A **559** (2006) 190 (cit. on p. 70).

[137] H. Mhaskar, *Neural networks for optimal approximation of smooth and analytic functions*,
Neural Computation **8** (1996) 164 (cit. on p. 71).

[138] Y. LeCun et al., "Efficient BackProp", *Neural Networks: Tricks of the trade*, Springer, 1998 9
(cit. on p. 72).

[139] D. E. Rumelhart, G. E. Hinton and R. J. Williams,
Learning representations by back-propagating errors, Nature **323** (1986) 533 (cit. on p. 72).

[140] Phi-T Physics Information Technologies GmbH, *The NeuroBayes® User's Guide*, (Apr. 2010),
URL: http://neurobayes.phi-t.de/nb_doc/NeuroBayes-HowTo.pdf
(cit. on pp. 72, 125).

[141] C. de Boor, *A practical guide to splines*, Applied Mathematical Sciences **27** (1978) (cit. on p. 72).

[142] G. Michelangelo, "Measurement of the production cross-section of single top quarks in association with *W* bosons at ATLAS", dissertation in preparation (cit. on p. 74).

[143] ATLAS Collaboration,
"Measurement of the *Wt* production in the lepton+jets channel using a likelihood method", ATL-COM-PHYS-2011-1704, CERN, Dec. 2011,
URL: https://cds.cern.ch/record/1407934 (cit. on p. 74).

[144] J. A. Stillings, "Search for the associated production of a *W* boson and a top quark with the ATLAS detector at 7 TeV", BONN-IR-2015-04,
Rheinische Friedrich-Wilhelms-Universität Bonn, 2015, URL: https://www.brock.physik.uni-bonn.de/research/atlas-experiment/publications-1 (cit. on pp. 74, 75, 101, 130).

[145] I. Cioară, "Optimization of the single top-quark production analysis in the *Wt* channel at ATLAS using kinematic fitting and neural networks", BONN-IB-2014-08,
Rheinische Friedrich-Wilhelms-Universität Bonn, 2014, URL: https://www.brock.physik.uni-bonn.de/research/atlas-experiment/publications-1 (cit. on p. 75).

[146] T. Loddenkötter,
"Implementation of a kinematic fit of single top-quark production in association with a *W* boson and its implementation in a neural-network-based analysis in ATLAS",
Rheinische Friedrich-Wilhelms-Universität Bonn, 2012, URL: https://www.brock.physik.uni-bonn.de/research/atlas-experiment/publications-1 (cit. on p. 75).

[147] ATLAS Collaboration,
"Expected Performance of the ATLAS Experiment – Detector, Trigger and Physics", 2009, arXiv:0901.0512 (cit. on p. 76).

[148] K. Cranmer et al.,
HistFactory*: A tool for creating statistical models for use with* RooFit *and* RooStats, (2012) (cit. on pp. 84, 85).

[149] L. Moneta et al., *The* RooStats *Project*, PoS **ACAT2010** (2010) 057,
arXiv:1009.1003 [physics.data-an] (cit. on p. 84).

[150] F. James and M. Roos, Minuit*: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Comput. Phys. Commun. **10** (1975) 343 (cit. on p. 84).

[151] W. Verkerke, "Guide to parameterized likelihood analyses", Draft 1.07, 2012,
URL: https://svnweb.cern.ch/cern/wsvn/atlasgrp/Physics/StatForum/FrequentistRecommendation/doc/profiling_recommendations_draft.pdf (cit. on pp. 86, 96).

[152] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur. Phys. J. C 73, 2501 (2013)],
arXiv:1007.1727 [physics.data-an] (cit. on pp. 86, 102).

[153] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Comput. Phys. Commun. **77** (1993) 219 (cit. on p. 93).

[154] J. Conway, *Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra*, (2011), arXiv:1103.0354 [physics.data-an] (cit. on p. 93).

[155] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions for bottom and charm hadrons in the decays of top quarks and the fragmentation of high p_{T} jets*, ATL-PHYS-PUB-2014-008, 2014, URL: http://cdsweb.cern.ch/record/1709132 (cit. on p. 95).

[156] CMS Collaboration, *Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. C **75** (2015) 542, arXiv:1505.04480 [hep-ex] (cit. on p. 95).

[157] ATLAS Collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, (2015), arXiv:1511.04716 [hep-ex] (cit. on p. 95).

[158] ATLAS Collaboration, *Measurement of the Inclusive and Fiducial Cross-Section of Single Top-Quark t-Channel Events in pp Collisions at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2014-007, 2014, URL: http://cdsweb.cern.ch/record/1668960 (cit. on p. 101).

[159] ATLAS Collaboration, *Measurement of the top quark mass in topologies enhanced with single top quarks produced in the t-channel at $\sqrt{s} = 8$ TeV using the ATLAS experiment*, ATLAS-CONF-2014-055, 2014, URL: http://cdsweb.cern.ch/record/1951323 (cit. on p. 101).

[160] L. top working-group, *Summary plot of single top-quark production*, Jan. 2015, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SingleTopAllChannelsHistory (visited on 18/02/2015) (cit. on p. 103).

[161] O. M. Kind, "Computer code used to generate a summary plot of single top-quark production", private communication, 10th Mar. 2016 (cit. on p. 103).

[162] ATLAS Collaboration, *ATLAS Computing: technical design report*, Technical Design Report ATLAS, Geneva: CERN, 2005, URL: http://cds.cern.ch/record/837738 (cit. on p. 108).

[163] I. Antcheva et al., *ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization*, Comput. Phys. Commun. **180** (2009) 2499 (cit. on p. 108).

[164] W. Bhimji et al., "The ATLAS Root-based data formats: recent improvements and performance measurements", CERN, Geneva, 2012, URL: https://cds.cern.ch/record/1448601 (cit. on p. 108).

[165] I. C. Brock, *Users Guide to Writing a Thesis in a Physics/Astronomy Institute of the Universität Bonn*, URL: http://pi.physik.uni-bonn.de/pi_only/thesis.php (cit. on p. 169).

Acknowledgements

I would like to express my gratitude to my supervisor, Prof. Ian Brock, for his continuous support for my Ph.D. studies. He gave me the freedom to explore on my own, had his door open at all times, and also made sure that I had a safe job. I also thank him and the other committee members, Prof. Klaus Desch, Dr. Bernard Metsch and Prof. Dorothea Bartels, for their interest in my dissertation as well as the discussions.

I want to thank all (current and former) members of the Brock group. I very much enjoyed working in the "7 TeV *Wt* lepton+jets analysis team" (actually more than just an analysis team): Jan Stillings, Thomas Loddenkötter, Pienpen Seema and Irina Cioară – thank all of you. Kudos to Jan who kept the computing infrastructure in Bonn running day and night! Ian's thesis guide [165] and his English tips, together with the proofreading by Pienpen and Regina Moles-Valls, were very valuable for the form of this manuscript. Without Ozan Arslan, there would have been considerably less fun at and after work as well as in my doktorwagen. Thanks go to Anjishnu Bandyopadhyay and Rui Zhang for taking over my vector-like quark duties when I needed to focus on thesis writing. Also, the coffee chats with Anjishnu, Muhammad Alhroob and Thomas will not be forgotten. (Jan would certainly have made it on the list if only he drank coffee.) I'm grateful for the numerous suggestions, discussions and critical questions that I received from members of the Desch group (especially Philip Bechtle at the biannual Jamborees) as well as my ATLAS colleagues (especially Dominic Hirschbühl, Oliver Kind, Reinhard Schwienhorst and Wolfgang Wagner). I'm also indebted to Markus Jüngst and Ewald Paul, who were excellent advisors during my diplom and Ph.D. times.

Finally, very special thanks go to my parents, not only for my own existence and their love but also for answering like a million "why?" questions when I was a little child, and their support for my academic education later on; and, of course, to my loving, patient and understanding girlfriend, Nokiang, who had to put up with a lot during my thesis writing but still cheered me up, and supported me all the time.