

Limit on $t\bar{t}+Z$ production in the $Z \rightarrow b\bar{b}$ channel at the CMS experiment

Master Thesis

Jan van der Linden

At the Department of Physics
Institute of Experimental Particle Physics

Reviewer:	Prof. Ulrich Husemann
Second reviewer:	PD Andreas Meyer (DESY)
Advisor:	Dr. Matthias Schröder

Karlsruhe, 08. 10. 2019



This thesis has been accepted by the first reviewer of the master thesis.

PLACE, DATE

.....
(Prof. Ulrich Husemann)

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

PLACE, DATE

.....
(Jan van der Linden)

Contents

1	Theoretical Foundation	3
1.1	The Standard Model	3
1.2	Collider Physics	6
1.3	$t\bar{t}+Z$ Production	9
2	Experimental Setup	11
2.1	The Large Hadron Collider	11
2.2	Kinematic Quantities at Hadron Colliders	12
2.3	The CMS Experiment	13
3	Object Reconstruction and Event Selection	17
3.1	Object Reconstruction	17
3.2	Event Topologies	19
3.3	Event Selection	24
3.4	Background Determination	25
4	Analysis Methods	29
4.1	Multivariate Analysis Methods	29
4.2	Statistical Methods	33
5	Systematic Uncertainties	37
5.1	Treatment of Systematic Uncertainties	37
5.2	Sources of Systematic Uncertainties	38
6	Analysis Strategy	43
6.1	Related Previous Results	43
6.2	Adjustments to the $t\bar{t}+\text{jets}$ -Background Model	44
6.3	Event Reconstruction	45
6.4	Event Classification	47
6.5	Optimization	48
7	Validation of Analysis Strategy	57
7.1	Validation of Input Features	57
7.2	Validation of ANN Training	58
7.3	Fit Validation	61
7.4	Impact of $t\bar{t}+H$ Contribution	65
8	Results	71
9	Summary and Outlook	79
	Bibliography	81

Appendix	89
A χ^2 -based Event Reconstruction	90
B ANN Input Features	92
C ANN Evaluation	116
D Discriminator Distributions	120

Introduction

In a strive to understand what the universe is made of and how the basic building blocks of our world behave and interact, the Standard Model of particle physics (SM) has been developed. This model describes the elementary particles as excitation of quantized fields. It incorporates the fundamental electromagnetic, weak and strong interactions. Only the gravitational interaction has so far not been able to be integrated in this extensive model. The latest addition to the SM is the Higgs boson. It was discovered by the ATLAS and CMS experiments at the CERN LHC in 2012 [1, 2], after being predicted almost 50 years earlier [3–5]. So far, the predictions of the Standard Model have been able to be confirmed extensively. Many analyses aim at either confirming the predictions of the SM, or finding deviations from it. These deviations might be indications of physics beyond the Standard Model (BSM). For example, it is still unknown what the vast amounts of Dark Matter in the universe are made of. In recent years, BSM physics has begun to be parameterized as effective field theories (EFT). Contributions of BSM interactions can, for example, be searched for by measuring the weak coupling of the top quark to the Z boson. An analysis of the $t\bar{t}+Z$ process is a probe of this weak coupling.

In this thesis a complete analysis of a Standard Model process with data of CMS is documented. It targets the $t\bar{t}+Z$ process with a decay of the Z boson to a pair of bottom quarks. The $t\bar{t}+Z$ process has already been successfully measured in final states with equal or more than two charged leptons [6–8]. The sensitivity in this multi-lepton final state exceeds the 5σ threshold required to claim an observation of the process. In these multi-lepton analyses also first EFT interpretations are performed.

This thesis aims at initiating a measurement of the $t\bar{t}+Z$ process in an additional final state with novel analysis techniques, and thereby getting insight in the possible sensitivity to the $t\bar{t}+Z$ process in the $Z \rightarrow b\bar{b}$ final state.

From a broader perspective this thesis builds the ground work for a combined measurement of the $t\bar{t}+Z$ and $t\bar{t}+H$ processes in the $Z \rightarrow b\bar{b}$ and $H \rightarrow b\bar{b}$ final states for the full LHC Run-II. This is especially interesting, as these two processes share similar kinematic features and production probabilities, and are therefore inherently hard to separate from each other. Up until now only the $t\bar{t}+H$ process has been studied in this phase space region, without well-studied consideration of the $t\bar{t}+Z$ contribution. The analysis strategy devised in this thesis explicitly aims at distinguishing the $t\bar{t}+H$ and $t\bar{t}+Z$ contributions in the analyzed region as much as possible.

Multivariate analysis techniques are employed for a multi-classification of the sought-after signal process $t\bar{t}+Z$ and irreducible background processes like $t\bar{t}+H$ production or $t\bar{t}$ production in association with additional jets ($t\bar{t}+\text{jets}$). The Artificial Neural Networks (ANN) used for the multi-classification utilize various features of an event to predict its physics content. Furthermore, a reconstruction of the final-state objects in $t\bar{t}+Z$ and $t\bar{t}$ events is performed which enhances the classification performance of the ANNs. An upper limit on the signal-strength modifier $\mu = \sigma^{\text{observed}}/\sigma^{\text{theory}}$ is determined with

the data recorded in the year 2018 at the CMS experiment at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The amount of data corresponds to an integrated luminosity of 59.7 fb^{-1} .

In Chapter 1, the theoretical foundation of the Standard Model, physics at hadron colliders and the $t\bar{t}+Z$ process are introduced. The experimental setup of the LHC and the CMS experiment are introduced thereafter in Chapter 2. In Chapter 3 the reconstruction of physics objects from CMS detector measurements is described. Also included in this chapter is the selection of events for the analysis of this thesis and a description of the event simulation. The multivariate analysis methods and the statistical methods used for the multi-classification and the evaluation of the ML fit, respectively, are introduced in Chapter 4. Thereafter, in Chapter 5, the systematic uncertainties applied for the final result are summarized. The analysis strategy is summarized and motivated in Chapter 6. This strategy is validated in Chapter 7. In this chapter also the fit to a pseudo data set is discussed and evaluated. Finally, in Chapter 8, the results of the analysis performed with real data events and the corresponding limit on the signal-strength modifier of $t\bar{t}+Z$ production is presented.

1 Theoretical Foundation

In this chapter, the theoretical background for the physics analysis presented in this thesis is reviewed. The Standard Model of particle physics (SM) describes the known elementary particles and their interactions and will be introduced in section 1.1. This analysis is based on events observed at the CMS experiment at the LHC, both introduced in chapter 2. Particle colliders can be used to probe the predictions of the SM, especially hadron colliders are well-suited for probing the SM at high energies. An overview of the physics concepts needed for the prediction of results in this hadron collider context will be given in section 1.2. Finally, the analyzed signal process will be briefly discussed in section 1.3 and later in section 3.2.

1.1 The Standard Model

The Standard Model of particle physics is the most advanced description of fundamental interactions and particles in nature [3–5, 9–21] and has been tested and verified in many ways [1, 2, 22–32]. All particles described by the SM interact via fundamental forces which are described by the theory of gauge interactions; these will be introduced in the following. The elementary particles are introduced thereafter.

The following explanations were adapted from [33, 34].

1.1.1 Gauge Theories and Interactions

The particles of the SM interact via electromagnetic, weak and strong forces. These interactions are mediated by gauge bosons. At a mathematical level, the gauge theories are quantum field theories (QFT), where particles are interpreted as excitations of the fields. A field is described by operators $\phi(x)$, which depend on the space-time coordinates x . The state of a field can be described via Lagrangian densities $\mathcal{L}(\phi(x), \partial_\mu \phi(x))$. Equations of motion for fields can be obtained by solving the Euler-Lagrange equation for the corresponding Lagrangian densities. By requiring the Lagrangian density, i.e. the state of a field, to be invariant under local unitary transformations, as postulated for physics systems, a description of interactions of gauge bosons and elementary fermions is obtained.

For example, the Lagrangian density for a free fermion (which is described by a four-component Dirac spinor $\psi(x)$) is given by

$$\mathcal{L}_{\text{fermion}} = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x), \quad (1.1)$$

where γ_μ are the Dirac matrices and m is the mass of the fermion. The fundamental physics theories are based on symmetry principles. A local unitary transformation $U(1)$ can be introduced as

$$\psi(x) \rightarrow \psi'(x) = e^{iq\alpha(x)}\psi(x), \quad (1.2)$$

where $\alpha(x)$ is a space-time dependent phase. To achieve symmetry under this transformation, i.e. invariance of the Lagrangian density, the derivative ∂_μ in eq. 1.1 has to be replaced by a covariant derivative $D_\mu = \partial_\mu + iqA_\mu(x)$. The field $A_\mu(x)$ is introduced as a bosonic vector field transforming as $A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu\alpha(x)$. The associated boson carries no mass, as this would break the proposed symmetry. According to Noether's theorem the introduction of a continuous symmetry leads to the conservation of a quantity q , which can in this case be identified as the electric charge. The field A_μ corresponds to the photon field, whose dynamics are described via the electromagnetic tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ as $\mathcal{L}_{\text{kin}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$.

Together, this yields the Lagrangian density of quantum electrodynamics (QED) [10–12] as

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(x)(i\gamma^\mu\partial_\mu - m)\psi(x) - q\bar{\psi}(x)\gamma^\mu A_\mu(x)\psi(x) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (1.3)$$

Each term can be associated with a basic interaction of QED. The first term ($\bar{\psi}\gamma^\mu\partial_\mu\psi$) and the second term ($m\bar{\psi}\psi$) describe the kinematics and the propagation of a fermion with mass m . The third term ($q\bar{\psi}A_\mu\psi$) describes the interaction of a fermion with the photon field A_μ . The strength of this interaction is proportional to the electric charge q .

Similarly, quantum chromodynamics (QCD) can be introduced by requiring invariance of the Lagrangian density under $SU(3)$ transformations. This yields eight physical gluon fields and couplings of fermions to gluons. In QCD there are three conserved Noether charges, which are referred to as color charges [13]. In addition, the Lagrangian density of QCD also incorporates self-interaction terms of the gluon fields as a consequence of the non-zero structure constant of $SU(3)$, enabling triple and quartic gluon gauge couplings.

The formulation of a theory to describe the weak interaction [20] had a number of problems. It was observed that only left-handed particles and right-handed antiparticles interact weakly [22, 23]. Hence, in the description of weak interaction the distinction between chiralities of particles has to be considered. Furthermore, the gauge bosons of the weak interaction were observed to be massive [27–29]. However, the introduction of gauge fields to conserve the proposed symmetry of the Lagrangian density does not allow for massive gauge bosons. These problems can be solved by formulating the theory as before with massless gauge bosons and introducing the masses of particles via the Higgs mechanism [3–5].

The left-handed and right-handed components of the fermions are defined as the projections $\psi_{\text{L/R}} = \frac{1}{2}(1 \mp \gamma^5)\psi$, where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. Without mass terms, the Lagrangian of one generation of leptons then is

$$\mathcal{L} = \bar{\psi}_{\text{L}}\gamma^\mu D_\mu\psi_{\text{L}} + \bar{\psi}_{\text{R}}\gamma^\mu D_\mu\psi_{\text{R}}. \quad (1.4)$$

A combined electroweak interaction follows from symmetry under the $SU(2)_{\text{L}} \times U(1)_{\text{Y}}$ gauge group [14, 18, 19]. This incorporates three massless gauge fields W_μ^a with $a = 1, 2, 3$ for $SU(2)_{\text{L}}$, acting only on the left-handed particles (and right-handed antiparticles), and one massless gauge field B_μ for $U(1)_{\text{Y}}$. The resulting bosons mix to form the observed physical bosons $W^\pm$ from W^1 and W^2 , and Z and the photon A from a superposition of W^3 and B . As a consequence, the Z boson and the photon both couple to both chirality states, while the W bosons couple only to the left-chiral parts.

To obtain the particle masses, a doublet of a complex scalar field ϕ can be introduced, which has to transform under $SU(2)_{\text{L}} \times U(1)_{\text{Y}}$ the same way as the left-chiral fermion

Table 1.1: **Gauge and Higgs bosons of the Standard Model.** All gauge bosons carry spin 1 in units of $\hbar$, the Higgs boson carries no spin. The values are taken from [36].

	interaction	mass	charges
photon (γ)	electromagnetic	—	—
gluon (g)	strong	—	color charge
$W^\pm$ bosons	electroweak	$(80.38 \pm 0.01) \text{ GeV}$	electric and weak charge
Z boson	electroweak	$(91.188 \pm 0.002) \text{ GeV}$	weak charge
Higgs boson		$(125.1 \pm 0.1) \text{ GeV}$	—

fields. The Lagrangian density for this field is

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \quad (1.5)$$

The last two terms represent a potential for the field that with $\lambda > 0$ and $\mu^2 < 0$ results in infinitely many ground states where the field is non-zero. By choosing a specific ground state of the potential $\phi_0 \sim (0, v)^T$ with the Higgs vacuum expectation value $v = \sqrt{-\mu^2/(2\lambda)}$, the $SU(2)_L \times U(1)_Y$ symmetry is spontaneously broken. From this ground state the field ϕ can be expanded via a perturbation approach

$$\phi = \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \quad (1.6)$$

leading to a scalar boson H with mass $m_H = \sqrt{2\lambda}v$, the Higgs boson. Furthermore, three additional massless fields are obtained [16], which can be absorbed by the fundamentally massless gauge bosons $W^\pm$ and Z via an $U(1)$ transformation as additional degrees of freedom, manifesting themselves as the gauge boson masses. The gauge boson masses are related to the vacuum expectation value v as $m_W = \frac{1}{2}gv$ and $m_Z = \frac{1}{2}\sqrt{g^2 + g'^2}v$, where g and g' are gauge group specific coupling constants. The expansion also introduces terms related to the coupling of the Higgs boson to the massive gauge bosons and triple and quartic Higgs boson couplings.

Masses of fermions are obtained from introducing Yukawa interaction terms [35]

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi}_L(x) \phi(x) \psi_R(x) \quad (1.7)$$

to the Lagrangian density, which can be interpreted as mass terms for fermions ($m\bar{\psi}\psi$) and the coupling of fermions to the Higgs boson ($\bar{\psi}H\psi$) when expanding the Higgs doublet around its ground state (see eq. 1.6). The fermion masses are related to the vacuum expectation value as $m_f = \frac{1}{\sqrt{2}}y_f v$, where y_f is the fermion-specific Yukawa coupling constant.

In summary, three massive bosons ($W^\pm$, Z) and one massless boson (γ) are introduced from electroweak unification, and eight massless gluons (g) are introduced from QCD. A massive boson (H) is obtained from the introduction of the Higgs field and its potential to the Lagrangian density, providing a mechanism to generate the masses of gauge bosons, and fermion masses from additional Yukawa interaction terms. Some properties of the bosons are summarized in Table 1.1. The remaining known fundamental interaction, namely gravity, has so far not been able to be incorporated consistently into the Standard Model. The elementary particles will be discussed in the following.

Table 1.2: **Fundamental particles of the Standard Model.** The first group are the six quarks (and antiquarks), the second group are the six leptons (and antileptons). All quarks carry color charge and thus interact via the strong force. All left-handed particles (and right-handed antiparticles) carry weak charge and thus interact via the weak force. All but the neutrinos (and antineutrinos) carry electric charge and thus interact via the electromagnetic force. If not stated otherwise the values are taken from [36].

	electric charge	mass
up quark (u)	+2/3	$2.2^{+0.5}_{-0.3}$ MeV
down quark (d)	-1/3	$4.7^{+0.5}_{-0.2}$ MeV
charm quark (c)	+2/3	(1.27 ± 0.02) GeV
strange quark (s)	-1/3	93^{+11}_{-5} MeV
top quark (t)	+2/3	(172.9 ± 0.4) GeV
bottom quark (b)	-1/3	$4.18^{+0.03}_{-0.02}$ GeV
electron neutrino (ν_e)	0	< 1.10 eV [37]
electron (e)	-1	0.51 MeV
muon neutrino (ν_μ)	0	< 1.10 eV [37]
muon (μ)	-1	105.60 MeV
tau neutrino (ν_τ)	0	< 1.10 eV [37]
tauron (τ)	-1	1776.90 MeV

1.1.2 Elementary Particles

In Table 1.2, the fundamental fermions described by the SM are summarized. They exist in three generations, each consisting of two quarks, a charged lepton and its associated neutrino. The fermions of the different generations differ only in mass. For each of these particles an antiparticle exists that carries opposite charge. Stable matter is composed of the fermions of the first generation, i.e. up and down quarks and the electron. The charged leptons and quarks of the second and third generation are not stable and are observed to eventually decay into the fermions of the first generation.

Quarks carry color charge and are subject to the confinement of color charge, which does not allow particles with a net color charge to exist as free particles. Therefore, quarks are observed to exist in groups of two or three, called mesons and baryons, respectively [15, 17]. All fermions are subject to the weak interaction, all but the neutrinos participate in the electromagnetic interaction and only quarks are subject to the strong interaction.

1.2 Collider Physics

The results presented in this thesis are obtained in a hadron collider context, probing the SM at high energies. The description of the SM in the previous section is of mathematical nature, which needs to be translated into physics observables at hadron colliders. In this section, the concept of cross sections and decay rates will be introduced first. This lays out the basis for the interpretation of results observed at colliders, by which the predictions made by the SM or deviations from these expectations can be confirmed. Thereafter, the physics of partons will be introduced, which is an important concept at hadron colliders, translating the clean initial states expected from theory predictions of the Standard Model into the observables.

1.2.1 Cross Sections and Decay Rates

The observables commonly measured at collider experiments are the rates of specific processes predicted by the SM or theories covering physics beyond the Standard Model (BSM). Generally, two different rates are distinguished:

Interaction Rate

On the one hand, the interaction rate can be an observable of interest. It expresses the probability of a certain process to take place from the collision of protons and is commonly expressed in terms of the production cross section σ of a process. The interaction rate for this is given by

$$\frac{dN}{dt} = \sigma \cdot L, \quad (1.8)$$

where L is the instantaneous luminosity, quantifying the flux of colliding particles at the accelerator normalized to unit area and time.

Decay Rate

On the other hand, almost all particles have finite mean lifetimes τ and decay into other particles, unless this is prevented by some conservation law. The probability for a particle to decay follows an exponential density distribution $e^{-t/\tau}$. The decay rate is the inverse of the mean lifetime $\Gamma = 1/\tau$. Most particles are able to decay into other (lighter) particles, each being associated with a separate decay rate Γ_i . The total decay rate Γ_{tot} is then given as the sum of all partial decay rates. Commonly, the fraction of one particle decaying via one decay mode is expressed in terms of the branching fraction $\mathcal{B}_i = \Gamma_i/\Gamma_{\text{tot}}$.

These observable quantities are related to the quantum mechanical transition amplitudes $|\mathcal{M}|^2 = |\langle\psi_f|V|\psi_i\rangle|^2$, describing the probability of a quantum mechanical initial state $|\psi_i\rangle$ to transition into a final state $|\psi_f\rangle$ via an interaction potential V . Fermi's Golden Rule allows the derivation of scattering and decay probabilities from this matrix element $|\mathcal{M}|^2$.

The matrix elements themselves can be calculated from the Lagrangian densities of the SM via a procedure initially developed by Richard Feynman [38] using a perturbation approach. The terms of this perturbation expansion can be visualized via Feynman diagrams which can directly be translated into mathematical expression for the calculation of the matrix elements. The first term of the expansion is referred to as leading order term, which encompasses the most direct way, i.e. the way with the smallest number of interaction vertices, to reach the final state from the initial state. Further orders of perturbation theory are ordered by the additional number of interactions (i.e. vertices). Each vertex is associated with the introduction of another factor of the squared coupling constant to the matrix element. These coupling constants are in general small, thus, an expansion in orders of the coupling constant, i.e. number of vertices can be performed, which is expected to converge for higher orders. Exemplary, the leading order diagram for $e^+e^- \rightarrow e^+e^-$ scattering is shown in Figure 1.1 on the left and possible higher order contribution on the right. Higher order corrections can to a certain degree be neglected and the leading order calculation already gives a good estimate of the matrix element¹.

The observed value of the coupling constant α_S of the strong interaction changes considerably at different energies (referred to as running coupling). At high energies, the strong coupling constant is observed to be small, at lower energies it is large. Hence, this perturbation approach is no longer valid at low energies for QCD interactions, as the perturbation series (expanded in the orders of the coupling constant) no longer converges. This consequently

¹With the advance in precision measurements this is not necessarily the case anymore, as the measured cross sections are by now well in the region of being sensitive to additional contributions to the matrix elements from higher order contributions.

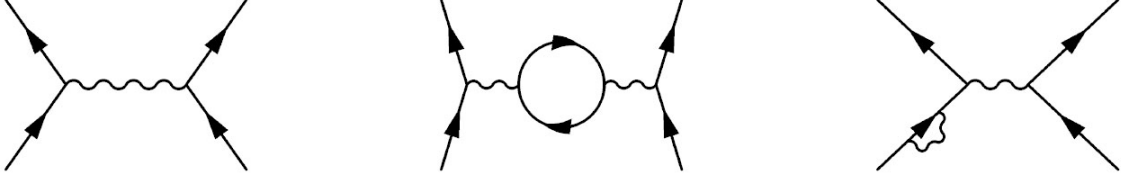


Figure 1.1: **Exemplary Feynman diagrams for $e^+e^- \rightarrow e^+e^-$ scattering.** Incoming fermions correspond to lines on the left, outgoing fermions to lines on the right of a diagram. Intermediate interactions happen at vertices in between. Wiggly lines represent photons. The left diagram is at leading order, containing two vertices corresponding to the coupling of a fermion to a photon. The diagrams in the middle and on the right are diagrams of higher order, containing four interaction vertices each, resulting in a suppression with higher powers of the coupling constant. Taken from [39].

requires a different description of low energy interactions, covered by phenomenological parton showering and hadronization models.

1.2.2 Physics of Partons

Protons are composite objects, consisting of three valence quarks: two up quarks and one down quark. These interact with each other, emitting or absorbing gluons which in turn can produce a quark and antiquark pair of same quark flavor. The content of the protons at different energies can be determined via deep inelastic scattering measurements [41], where elementary particles, e.g. electrons, are collided with protons, probing the substructure of the proton. From these measurements and theory predictions parton distribution functions (PDF) of the protons can be determined, expressing the probability of a parton (i.e. quarks or gluons) to be found with a certain momentum fraction inside the proton. As an example, one PDF set is shown in Figure 1.2. The derived PDFs can be used to sample the initial states expected in proton-proton (or parton-parton) interactions to calculate the matrix elements.

For the interaction of partons inside the proton the four-momentum transfers are low, i.e the coupling constant of the strong interaction is large. At the same time, the interactions of interest at hadron colliders happen at higher energies, where the strong coupling constant is very small. This allows for a factorization of these two stages, i.e. the soft (low-energy) interactions inside the proton and the hard (high-energy) interactions of the colliding partons. The factorization of these processes leads to the introduction of a factorization energy scale μ_F defining the threshold for soft and hard interactions. The cross section σ for the process of proton-proton interaction to some final state X ($pp \rightarrow X$) can thereby be expressed as

$$\sigma_{pp \rightarrow X}(p_1, p_2) = \sum_{i,j}^{\text{partons}} \int dx_i \int dx_j \cdot f_i(x_i, \mu_F^2) f_j(x_j, \mu_F^2) \cdot \hat{\sigma}_{\hat{p}_i, \hat{p}_j \rightarrow X}(x_i p_1, x_j p_2). \quad (1.9)$$

Here, $f_{i,j}$ are the PDFs for partons $\hat{p}_{i,j}$ to be found at a momentum fraction x_i inside the proton, $p_{1,2}$ are the momenta of the colliding protons. The partonic cross section $\hat{\sigma}_{\hat{p}_i, \hat{p}_j \rightarrow X}$ encompasses the hard scattering that can be calculated via perturbation theory.

After calculating the hard matrix element, the final state particles are expected (in the case of quarks and gluons) to collimate in color neutral groups and subsequently decay into lower energy states. These processes, again, happen at lower energies, introducing another

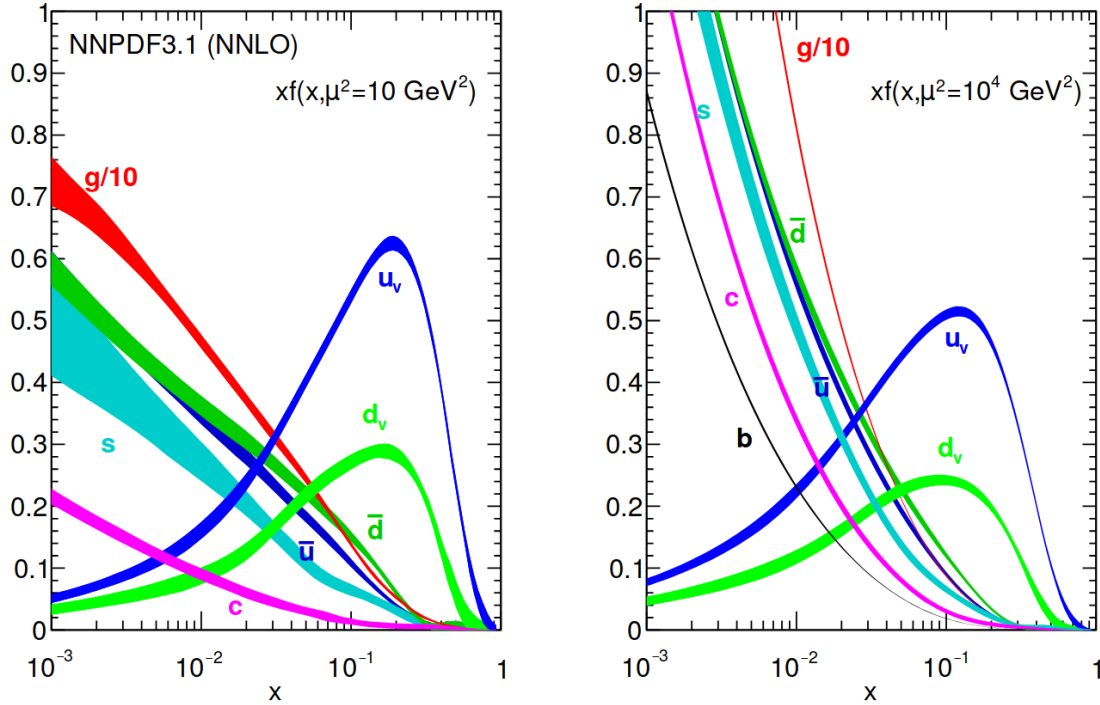


Figure 1.2: **Exemplary proton PDF sets at different four-momentum as a function of the momentum fraction x of the parton.** At higher x the proton content is dominated by the valence quarks up and down. Especially at low momentum fractions, gluons are dominant by multiple orders of magnitude. Taken from [40].

factorization scale where this hadronization and parton showering process takes over from the matrix element final states. This ultimately leads to the final set of particles that can be observed in the detector.

1.3 $t\bar{t}+Z$ Production

Top quarks are the heaviest known elementary particles, with a mass of approximately 172.9 GeV [36]. The transitions of quarks to quarks of other flavors via weak interaction is described by the CKM formalism [21]. From this it is expected and also observed that the top quark almost exclusively decays to bottom quarks in association with a W boson. Due to the high mass difference relative to other quarks and the CKM-favored decay, the life time of the top quark is very short, of the order of 10^{-25} seconds, which is shorter than the expected time to hadronize (i.e. form hadrons). Thus, the top quark decays before participating in the hadronization process, distinguishing it from other quarks. The decay modes of a pair of top quarks ($t\bar{t}$) are classified via the decay modes of the W bosons. One W boson decaying leptonically via $W \rightarrow \ell\nu$ with $\ell = e, \mu, \tau$ (branching fraction 33%) and one W boson decaying hadronically via $W \rightarrow q\bar{q}'$ (branching fraction 67%) is referred to as the semileptonic $t\bar{t}$ decay mode. It has a branching fraction of 44%.

Z bosons are the neutral gauge bosons of the electroweak force with a mass of 91.2 GeV [36]. With a mean lifetime of approximately 10^{-25} seconds, the Z boson cannot be detected in the detectors directly. It decays into pairs of charged or neutral leptons with a combined branching fraction of 30%, or hadronically into $q\bar{q}$ pairs with a branching fraction of 70%. Due to lepton universality, the branching fraction is equal for all generation of leptons.

The Z boson cannot decay into pairs of top quarks as the mass of these are too high. The decay into a pair of bottom quarks has a branching fraction of 15.6%.

$t\bar{t}+Z$ production is a direct probe of the weak coupling of the Z boson to the top quark, as the Z boson couples directly to one of the top quarks in the dominant production channel (gluon-gluon initial state), as visualized in Figure 3.1.

Contributions of BSM interactions, changing the coupling of the Z boson to the top quarks, might be identified by studying this process [42]. This can be advanced further by measuring differential cross sections of $t\bar{t}+Z$ production, e.g. the cross-section of $t\bar{t}+Z$ production as a function of the transverse momentum of the Z boson. Contributions from BSM physics can be detected (or limits can be set) in these differential measurements. BSM effects can be parameterized in the scope of effective field theories (EFT) explained elsewhere [43]. In the scope of this thesis, the sensitivities for such an interpretation cannot be reached, however, first EFT interpretations in the context of $t\bar{t}+Z$ are performed by the CMS Collaboration in multi-lepton final states [8].

The latest cross-section calculation of the $t\bar{t}+Z$ production process [44] is performed at next-to-leading-order (NLO) accuracy, including electroweak and next-to-next-to-leading-logarithmic (NNLL) corrections. From these calculations the cross-section of $t\bar{t}+Z$ production is predicted to be

$$\sigma(t\bar{t}+Z)^{\text{theory}} = 0.86^{+0.07}_{-0.08} (\text{scale}) \pm 0.03 (\text{PDF} + \alpha_S) \text{ pb}. \quad (1.10)$$

2 Experimental Setup

In this chapter the experimental setup is described, including a brief description of the Large Hadron Collider (LHC), the Compact Muon Solenoid (CMS) experiment and an explanation of some quantities used for the description and interpretation of physics results in high energy physics at hadron colliders.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is designed to accelerate and store two beams of either protons or lead-ions that are brought to collision at four different points where the experiments ALICE [45], ATLAS [46], CMS [47] and LHCb [48] are placed. The ATLAS and CMS detectors are multi-purpose detectors, designed to probe the Standard Model in as many ways as possible. The LHCb experiment focuses its research on physics involving B hadrons, while the ALICE experiments focuses on heavy-ion research. The protons and lead-ions are accelerated to energies of 6.5 TeV or 2.8 TeV per nucleon, respectively. A more detailed description of the LHC and its accelerator complex can be found at [49, 50], only a brief introduction will be given in the following.

For the LHC to reach a center-of-mass energy of 13 TeV the protons have to pass through a chain of accelerators, since the LHC synchrotron itself is not designed to accelerate particles at low energies. The whole accelerator complex is shown in Figure 2.1. At first, hydrogen gas is ionized to obtain the protons. These are injected into the LINAC2, a linear accelerator, where the continuous beam is separated into bunches that are brought to energies of 50 MeV. Subsequently, the protons are lead through the Proton Synchrotron Booster (BOOSTER), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), reaching energies of 450 GeV. The instantaneous luminosity, which is a measure of the number of expected proton-proton collisions per time interval, is defined as

$$L = f \frac{n_b N_b^2}{4\pi\sigma_x\sigma_y}. \quad (2.1)$$

Here, f denotes the revolution frequency of the proton bunches, N_b the number of protons per bunch and n_b the number of colliding pairs of bunches in the accelerator. The geometrical spread of a bunch is quantified with σ_x and σ_y perpendicular to its flight direction. With bunches of about 10^{11} protons with a spacing of 25 ns, the LHC is designed

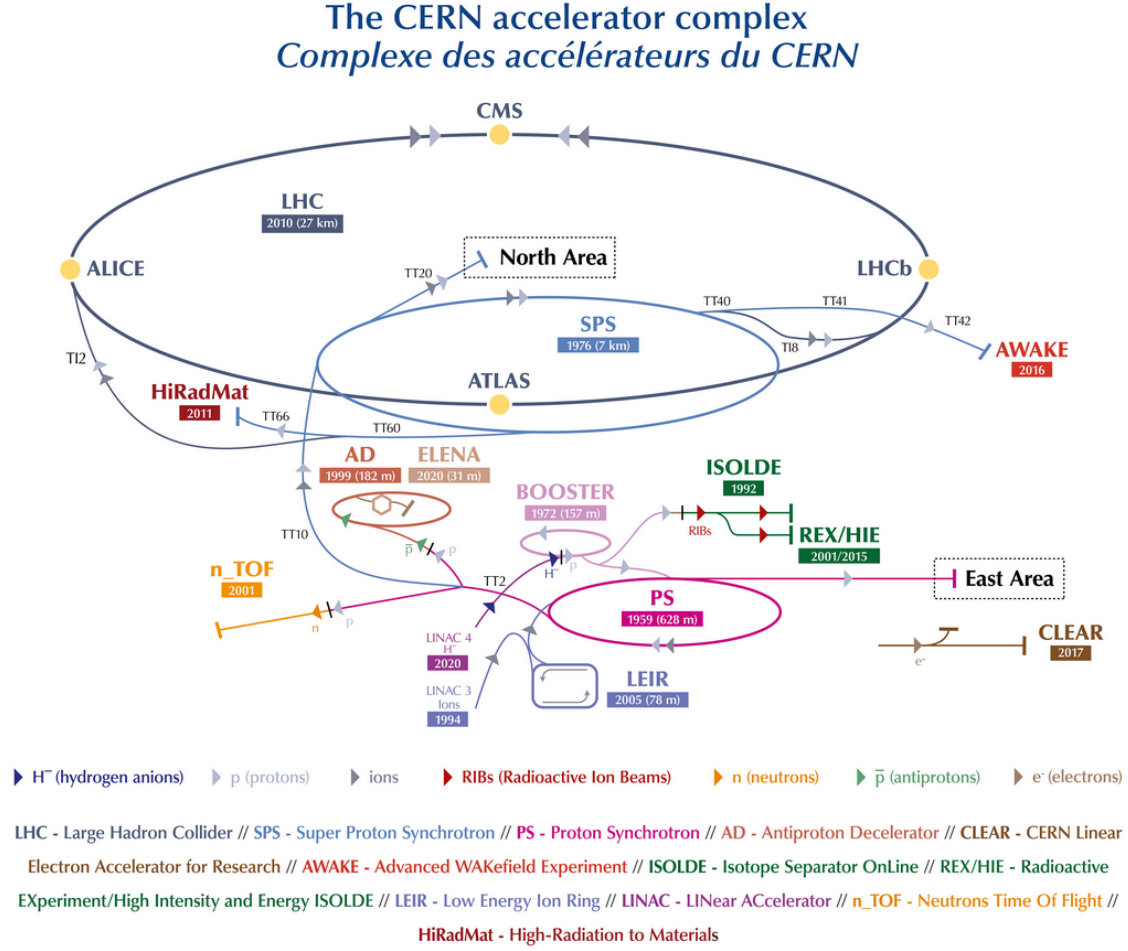


Figure 2.1: **LHC accelerator complex.** This graphic was taken from [52].

to reach an instantaneous luminosity of $L = 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. To quantify the total amount of data, the integrated luminosity

$$\mathcal{L} = \int dt L \quad (2.2)$$

is used. During the year 2018 an integrated luminosity of 59.7 fb^{-1} was reached [51]. With this amount of data, a hypothetical process with a production cross-section of 1 fb is expected to be produced about 60 times in this period.

2.2 Kinematic Quantities at Hadron Colliders

Some common quantities in collider physics that are used in this thesis are introduced in the following.

The coordinate system of the CMS experiment is centered around the interaction point. The z -axis is pointed in the direction of one proton beam. Due to the rotational invariance of physics processes and the detector, cylindrical coordinates are used. The distance r and the azimuthal angle ϕ are defined in the (x, y) -plane with $\phi = 0$ pointing to the center of the LHC. The polar angle θ is defined with respect to the z -axis.

The rapidity y and **pseudo rapidity** η are used to describe the direction of final-state particles in the θ -direction, as differences in the rapidity

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad (2.3)$$

are invariant under Lorentz transformation. This reduces to the pseudo rapidity

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right], \quad (2.4)$$

in the limit of $m \ll p$, which is easier to measure than the rapidity, hence it is used as a proxy. A pseudo rapidity of $\eta = 0$ corresponds to the direction transverse, and $\eta \rightarrow \pm\infty$ parallel to the beam axis.

Spatial differences are commonly expressed in the (η, ϕ) -plane as

$$\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}. \quad (2.5)$$

The center-of-mass energy $\sqrt{s}$ denotes the energy of the proton-proton collision. At the high energies of the LHC, the collision no longer take place between the protons themselves, but rather between partons of the protons (see section 1.2). These only carry fractions x_1 and x_2 of the proton's momenta, resulting in the partonic center-of-mass energy of $\sqrt{\hat{s}} = \sqrt{x_1 x_2 s}$. The partons do not necessarily carry the same momentum fractions, which is why the center-of-mass frame of the actual parton-parton collision is boosted along the z -axis.

The transverse momentum p_T , the component of the momentum in the (x, y) -plane is commonly used instead of the momentum due to the unknown center-of-mass frame of the parton-parton collision.

Missing transverse energy $\cancel{E}_T$ is used to describe the energy imbalance in the vectorial sum of all transverse momenta:

$$\cancel{E}_T = \left| -\sum_i \vec{p}_{T,i} \right|. \quad (2.6)$$

This imbalance hints at particles which cannot be detected with the detector, as the total vectorial sum of transverse momenta should amount to zero, because the colliding partons do not carry considerable amounts of transverse momentum.

2.3 The CMS Experiment

The Compact Muon Solenoid (CMS) detector is designed as a multi-purpose detector. It is supposed to detect and identify as many final-state particles as possible. It is constructed in various layers, each designed for different purposes. The innermost layer consists of silicon pixel and strip detectors (tracker), these are surrounded by a calorimeter system. Each of these systems is composed of a barrel section for particles at lower pseudo rapidities, and two endcap sections for particles produced at higher pseudo rapidities. Outside the calorimeter and tracking systems is a superconducting solenoid with an internal diameter of 6 m, providing a magnetic field of 3.8 T. The return yoke of the solenoid magnet is intertwined within the muon chambers that are placed outside the solenoid. An illustration of the CMS detector and its components is shown in Figure 2.2. In the following, the detector components are discussed briefly. The information is based on [47] where a more detailed description can be found.

The Tracker System is designed to provide a precise measurement of the trajectories of charged particles passing through the detector within the pseudo-rapidity range of $|\eta| < 2.5$. Based on this, the positions of the original proton-proton interaction and secondary decay vertices are reconstructed. The tracker consists of four layers of silicon pixel detectors,

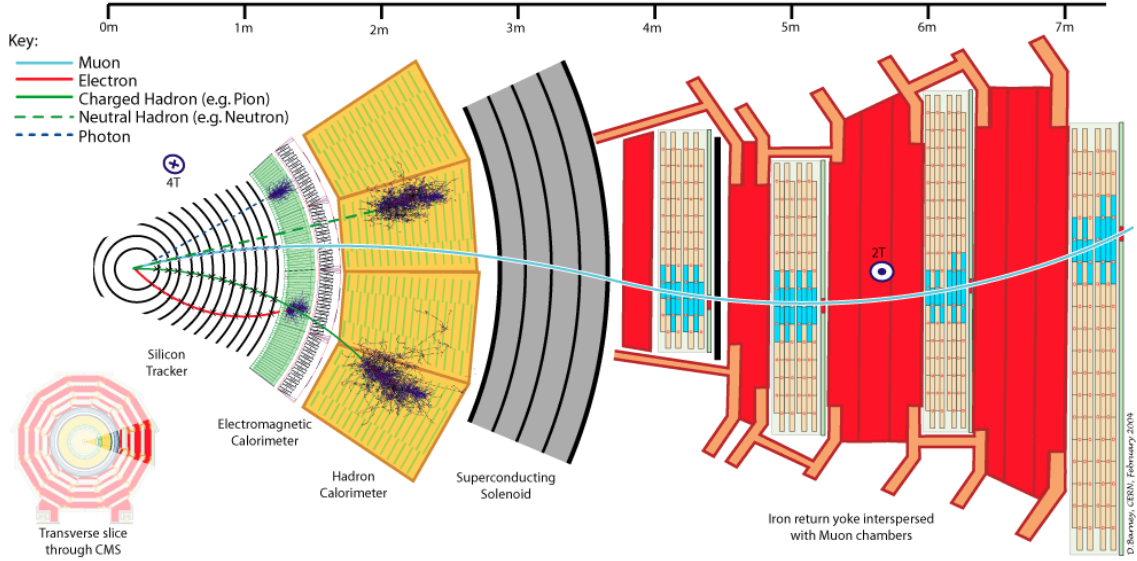


Figure 2.2: **Transverse slice through the CMS experiment.** The innermost point of the detector system, where the particle interactions take place, is on the left. From left to right the sub components of the CMS experiment are the tracker, the calorimeter system, the solenoid magnet and the muon system. As a visual example some particles and their interaction in the detector are included. This graphic was taken from [53].

able to determine the position of a particle in three dimensions, and layers of silicon strip detectors for two dimensional resolution. The points where a particle was detected (hits) are combined to trajectories of particles via track-reconstruction and track-fitting algorithms (see section 3.1). In combination with the magnetic field of the solenoid magnet, the transverse momenta of charged particles can be reconstructed from the curvature of the track, following

$$p_T \propto qBR, \quad (2.7)$$

where q is the charge of the particle, B is the magnetic field of the solenoid magnet and R is the radius of the reconstructed track. The resolution σ_{p_T} of the momentum is given as

$$\frac{\sigma_{p_T}}{p_T} \propto \frac{a\sigma_x}{BL^2} \cdot p_T \oplus \frac{b}{B\sqrt{LX_0}}, \quad (2.8)$$

where the first term takes into account the hit resolution σ_x based on the Gluckstern formula [54], and the second term arises due to scattering of the observed particle with the detector material. Here, L denotes the length of the particle trajectory in the tracker and X_0 the radiation length of the tracker material. The factors a and b are material and setup dependent constants.

The Electromagnetic Calorimeter (ECAL) enables the energy measurement of photons and charged particles. At high energies, photons primarily interact by producing pairs of electrons and positrons ($\gamma \rightarrow e^+e^-$). Charged particles primarily lose their energy via bremsstrahlung (e.g. $e^- \rightarrow \gamma e^-$), while they primarily ionize the detector material at lower energies [36]. Thus, high-energy electrons and photons entering the ECAL form showers of many particles, induced by many $1 \rightarrow 2$ processes, until the energy of the shower particles is too low for the aforementioned splitting processes, and the particles ionize the detector material. Hence, the energy of the shower can be determined by counting the number of

particles in the shower, which is approximately proportional to the initial particle's energy. The energy resolution σ_E of calorimeters is given as

$$\frac{\sigma_E}{E} \propto \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \quad (2.9)$$

where the first term arises due to Poissonian counting statistics, the second term due to noise in the read-out electronics, and the third term due to leakage of the shower out of the finite calorimeter volume at higher energies. The factors a , b and c are material and setup dependent constants.

The Hadron Calorimeter (HCAL) measures the energy of particles interacting primarily via the strong force. Charged hadrons also interact in the ECAL systems, but due to the larger longitudinal extension of hadronic showers, these do not develop fully. Neutral hadrons are not detected in the ECAL at all. Hence, an HCAL, specifically designed to induce showers from strongly interacting hadrons is placed outside the ECAL system.

The Muon System is designed to detect muons that survive the otherwise destructive energy measurement of the calorimeter systems. High-energy muons, produced at the proton-proton collisions at the LHC only lose small amounts of energy via interaction with matter [36]. Particles in this energy regime are commonly referred to as minimally ionizing particles (MIPs). Thus, muons are usually not stopped in the calorimeter system and can be detected in the outer muon system, where the transverse momentum p_T is determined by measuring the curvature of the particle track in the 2 T magnetic field of the return yoke of the solenoid magnet.

The Trigger System selects events of interest using a two-tiered system [55]. The first tier (L1) is hardware based and uses information from the calorimeter systems and the muon chambers to select events with a rate of around 100 kHz. The second level, commonly referred to as high-level trigger (HLT), performs a simplified full-event reconstruction on a computer farm and reduces the event rate to around 1 kHz before data storage.

3 Object Reconstruction and Event Selection

The events recorded with the CMS detector consist of various energy-, momentum- and position measurements as described in section 2.3. In order to perform a physics analysis, these measurements need to be converted into physics objects. The object-reconstruction methods and the particle-flow (PF) algorithm are briefly described in section 3.1. Based on the reconstructed physics objects, selections are applied to the measured data to identify a subset of events where the process of interest (signal process), $t\bar{t}+Z$ production, is enriched. In section 3.2, the processes relevant for this analysis are described to motivate the selections applied, which will be discussed in section 3.3. Due to very similar kinematic properties, some other physics processes (background processes) also survive the event-selection criteria to a large degree. In order to determine the contribution of these background processes to the subset of selected events, events simulated with Monte Carlo (MC) methods are used. The general idea and procedures are briefly described in section 3.4, as well as the data sets. An excess of events, measured by the CMS experiment, relative to the Standard Model predictions of all background contributions can be translated into possible contributions of the signal process. The underlying methods for this procedure will be described in section 4.2.

3.1 Object Reconstruction

Figure 2.2 shows schematically the sub-detectors of the CMS experiment and a collection of particles (i.e. physics objects) and their response in the various detector components. Energy deposits in the tracking and muon systems are connected to form tracks of charged particles with track-reconstruction [56] and track-fitting algorithms [57]. The energy of electrons, positrons and photons is measured in the electromagnetic calorimeter, while the energy of neutral and charged hadrons is mainly measured with the hadronic calorimeter. Muons are not detected in the calorimeter systems due to their low interaction rate [36] but can be identified with the outer muon systems. Measurements of energy clusters in the calorimeter systems and track measurements are combined with the PF algorithm [58] described below in section 3.1.2, to improve particle identification efficiencies and the resolution of energy and momentum measurements.

3.1.1 Track and Vertex Reconstruction

Particle tracks in the tracker system are reconstructed using a Kálmán filter technique [57]. The hits in the detector layers are processed sequentially by starting from a seed of hits. The so-formed track is extrapolated to adjacent tracker layers, where the hit closest to the track compatible with the trajectory, within the uncertainties, is added. This algorithm is continued to the last layer, or until no compatible hits are found in a number of layers. Subsequently, track fitting algorithms estimate the parameters which best describe the track. Generally, the trajectory of a charged particle in the constant magnetic field of the solenoid magnet can be described by a helix. The track fitting algorithms also take into account the scattering of particles with the detector material and the associated energy loss.

The tracks are also extrapolated to the interaction point of the proton-proton collisions, in order to reconstruct vertices. The reconstructed vertex with the largest value of summed p_T^2 of reconstructed objects is taken to be the primary vertex. Also, secondary vertices are reconstructed, where tracks originate from particle decays.

3.1.2 Particle Flow

The particle-flow (PF) algorithm [58] aims at reconstructing and identifying each individual particle in an event. For this purpose, an optimized combination of all information of the CMS sub-detectors is used.

- The energy of **photons** is determined from ECAL energy measurements, that are not associated with a particle track in the tracking system. Also taken into account are $\gamma \rightarrow e^+e^-$ conversions in the tracker.
- The energy of **electrons** is determined from a combination of ECAL energy measurements and the electron momentum, reconstructed from tracker measurements. Electron reconstruction also takes into account photons from bremsstrahlung in the tracker.
- The energy of **muons** is determined from the curvature of the reconstructed particle track in the tracking system and the dedicated muon system.
- The energy of **charged hadrons** is determined from a combination of ECAL and HCAL energy deposits and momentum measurements of matching tracks.
- The energy of **neutral hadrons** is mainly determined from HCAL energy measurements alone, after all other particle signatures were subtracted.

3.1.3 Jets

Particles attributed to originate from the same parton of the hard interaction process are clustered in jets. This is motivated by the confinement property of QCD at lower energies, which does not allow quarks or gluons to exist alone, but rather be hadronized in color-neutral hadrons. In the hadronization process, a shower of hadrons is produced per parton in the final state of the hard sub-process.

Jet Reconstruction

The jets are reconstructed from PF candidates with the infrared and collinear safe anti- k_T algorithm [59, 60], using a jet-clustering radius of $R = 0.4$ (referred to as AK4). Infrared safety implies the invariance of an algorithm under soft, i.e. low-energy gluon radiation,

while collinear safety implies the invariance under gluon radiation at small angles. This clustering algorithm uses the particle-particle distance measure

$$d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \quad R = 0.4, \quad (3.1)$$

for each pair of particles, and the particle-beam distance

$$d_{iB} = p_{T,i}^{-2}, \quad (3.2)$$

for each particle. The particles corresponding to the minimal values of d_{ij} are combined, leading to a new pseudo-particle, for which new distances d_{ij} and d_{iB} are calculated. If the minimal value in an event is a particle-beam distance d_{iB} , the particle (or combined pseudo-particle) is declared as a jet and removed from the collection.

b-Tagging

Four of the six jets expected in a $t\bar{t}+Z$, $Z \rightarrow b\bar{b}$ event at leading-order (as shown in Figure 3.1) arise from the decay of a bottom quark. Hadrons containing bottom quarks have longer life times than other hadrons due to the small CKM matrix elements involving bottom quark decays [21], resulting in resolvable secondary decay vertices. The information can be used to identify (“b-tag”) jets from bottom quarks. In this thesis, the DEEPJET algorithm [61, 62] was used to identify b-jets. It is based on a Deep Neural Network, using various features of PF candidates and properties of secondary vertices associated with the jet in question. The jets are classified in six different classes, covering jets originating, among others, from bottom quarks, charm quarks, light flavor quarks and gluons. From this multi-classification output, a value between zero and one can be assigned to jets, where higher values indicate that the jet is more b-jet like. This value will henceforth be referred to as b-tag value. A jet will be referred to as b-tagged if the b-tag value exceeds 0.277, which corresponds to a mistag probability of approximately 1% for light flavor jets, i.e. about 1% of light flavor jets are incorrectly identified as b-tagged jets. The efficiency to correctly identify b-jets is approximately 80% [61].

3.1.4 Missing Energy

Neutrinos or other unknown particles (like typical Dark Matter candidates) escape the detector without interaction. The presence of such particles can be deduced from the momentum imbalance observed in an event. The total vectorial sum of transverse momenta in an event should amount to zero because the colliding partons do not carry a considerable amount of transverse momentum. Thus, the missing transverse energy

$$\cancel{E}_T = \left| - \sum_{\text{detected}} \vec{p}_T \right| = \left| \sum_{\text{undetected}} \vec{p}_T \right|, \quad (3.3)$$

can be used as a proxy of the energy measurement of neutrinos and other “invisible” particles.

This, however, has to be used with a caveat as the determined value of $\cancel{E}_T$ depends on the energy or momentum measurement of all particles detected in an event, and is thus prone to energy mismeasurements. Furthermore, multiple particles cannot be resolved from a single $\cancel{E}_T$ value, e.g. in events with more than one neutrino the different neutrinos cannot be resolved independently.

3.2 Event Topologies

The development of an event selection strategy, as presented in section 3.3, requires precise knowledge of the signal process and its possible final states. Furthermore, background

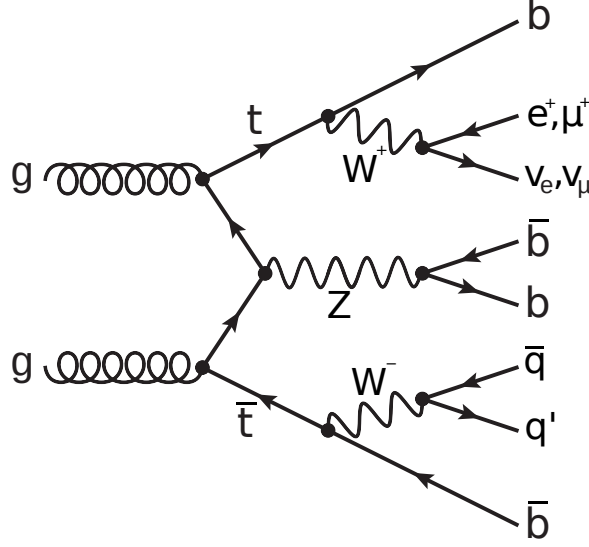


Figure 3.1: **Example LO Feynman diagram for $t\bar{t}+Z$ production including the decay of the top quarks and the Z boson.**

processes which cannot be removed with these selection criteria need to be understood. This section gives an introduction to the signal process $t\bar{t}+Z$ and its final states, followed by an overview over the different background processes considered in this thesis.

3.2.1 Signal Processes

The signal process studied in this thesis is the production of a $t\bar{t}$ pair in association with a Z boson. The analysis aims at the semileptonic final state of the $t\bar{t}$ system, i.e. the final state in which one top quark decays leptonically ($t \rightarrow Wb$, $W \rightarrow \ell\nu$), and the other top quark decays hadronically ($t \rightarrow Wb$, $W \rightarrow q\bar{q}'$). Among all $t\bar{t}$ decay modes, the best sensitivity can be reached with the semileptonic final state due to its large branching fraction of 44%, combined with a charged lepton, well-suited for selecting events and reducing QCD multi-jet backgrounds. The Z boson is analyzed in a decay to a $b\bar{b}$ pair ($Z \rightarrow b\bar{b}$), which has a branching fraction of 12.8% [36].

In summary, the final state consists of six quarks, of which four are bottom quarks, one charged lepton and a neutrino, which cannot be detected directly, but its existence can be inferred from missing transverse energy $\cancel{E}_T$ in the event. An exemplary Feynman diagram of the final state of this analysis is shown in Figure 3.1.

3.2.2 Background Processes

Several other processes can result in a final state similar to the signal process. Thus, selecting events according to this final state necessarily also selects events of the following processes. These form the backgrounds of this analysis.

$t\bar{t}$ production in association with additional jets ($t\bar{t}+\text{jets}$) is the major background for this final state. As for the signal process, the semileptonic decay of the $t\bar{t}$ system leads to four quarks, of which two originate from bottom quarks, one charged lepton and a neutrino. Associated with additional radiation of gluons, more quarks are expected in the final state. Especially the radiation of a gluon decaying into a pair of bottom quarks is an important background, as one expects four bottom quarks in the final state as well. This is referred to as $t\bar{t}+b\bar{b}$ production. An exemplary Feynman diagram of $t\bar{t}+b\bar{b}$ production

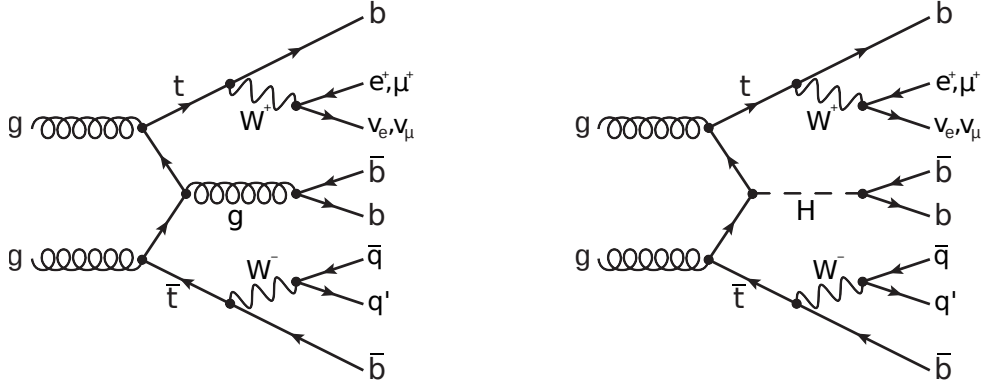


Figure 3.2: **Example LO Feynman diagrams for $t\bar{t}+b\bar{b}$ (left) and $t\bar{t}+H$ (right) production including the decay of top quarks, Higgs boson and gluon.**

is shown in Figure 3.2 (left).

To separate between $t\bar{t}+b\bar{b}$ and $t\bar{t}+Z$ events, observables involving the additional $b\bar{b}$ pair and angular differences in the $t\bar{t}$ system have been studied for the final selection. Differences in these observables hint at differences in the invariant mass of the reconstructed gluon or Z boson, as the gluon is massless, whereas the Z boson has a mass of 91 GeV, and the color-charge of the gluon, affecting angular distributions due to additional QCD interactions.

Events with additional gluon radiation and a subsequent decay into non- $b\bar{b}$ quarks can also pass the selection, as identifying jets originating from bottom quarks is prone to mistagging, as explained in section 3.1. The contributions of $t\bar{t}+\text{jets}$ events to the analysis region is split into three mutually exclusive classes to provide a model that is flexible enough to incorporate various systematic uncertainties for each process independently. The processes are defined according to the number of additional jets not originating from $t\bar{t}$ decays and the flavor of generator-level hadrons found in these jets, based on the method described in [63]:

- $t\bar{t}+b\bar{b}$: at least one additional b-jet,
- $t\bar{t}+c\bar{c}$: at least one additional c-jet, but no b-jet,
- $t\bar{t}+lf$: no additional b-jets or c-jets.

$t\bar{t}+H$ production, especially with $H \rightarrow b\bar{b}$ decays, has the same final state as the signal process. An exemplary Feynman diagram for $t\bar{t}H(bb)$ production is shown in Figure 3.2 (right). Separation of $t\bar{t}+Z$ and $t\bar{t}+H$ induced events is inherently hard as both, the Higgs boson and the Z boson have similar masses. The most notable differences are kinematic distributions of $t\bar{t}$ decay products due to the spin difference of the bosons. An example is shown in Figure 3.3, where the difference in pseudo rapidity η of both top quarks is shown for $t\bar{t}+H$, $t\bar{t}+Z$ and $t\bar{t}+\text{jets}$ events. The angular difference between both top quarks for $t\bar{t}+H$ events tends to be smaller than that of $t\bar{t}+Z$ events, which might be attributed to the different energy scales of Higgs boson or Z boson radiation, or the chirality-flip induced via Higgs boson radiation which can change the kinematics of the process.

Single Top production (single t) to a small extent also contributes to the background, especially in t-channel production, depicted in Figure 3.4 (left). At leading-order, the final state consists of three quarks, of which two are bottom quarks, one charged lepton and one neutrino. With additional radiation of gluons, similar final states can be reached,

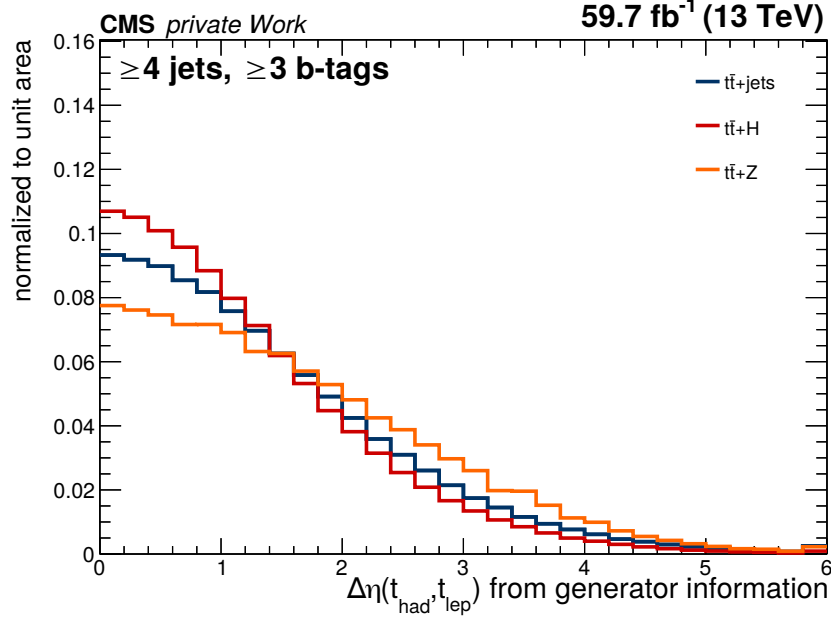


Figure 3.3: **Difference in pseudo-rapidity between both top quarks at generator level for the processes $t\bar{t}+Z$, $t\bar{t}+H$ and $t\bar{t}+jets$.** The yields of each process are normalized to unit area.

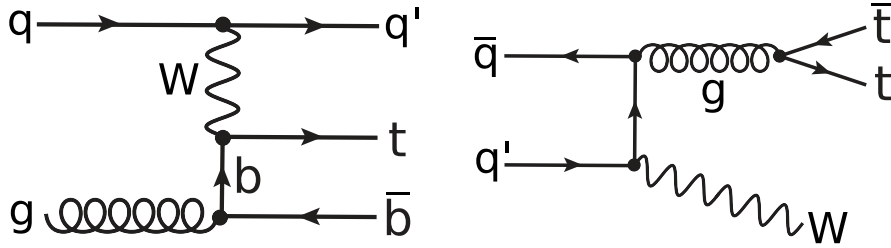


Figure 3.4: **Example LO Feynman diagrams for single t (left) and $t\bar{t}+W$ (right) production.**

however, the kinematics differ more significantly than for the other processes presented so far. Furthermore, the total cross-section of this process is smaller than for $t\bar{t}$ production.

$t\bar{t}+W$ production, as depicted in Figure 3.4 (right) can contribute to the background, e.g. from decays of the W boson involving bottom quarks. This is, however, suppressed by the CKM matrix and is thus sub-dominant.

Vector boson production in association with jets ($V + jets$), as shown in Figure 3.5 (left) has small contributions to the background. Leptonic decays of the vector bosons ($W \rightarrow \ell\nu$, $Z \rightarrow \ell\ell/\nu\nu$) can yield the required lepton and possible $\cancel{E}_T$, while additional jets might lead to final states similar to the signal final state.

Production of pairs of vector bosons (diboson), as shown in Figure 3.5 (right) is a minor contribution to the background, e.g. from $Z \rightarrow b\bar{b}$ decays or $W \rightarrow \ell\nu$ decays.

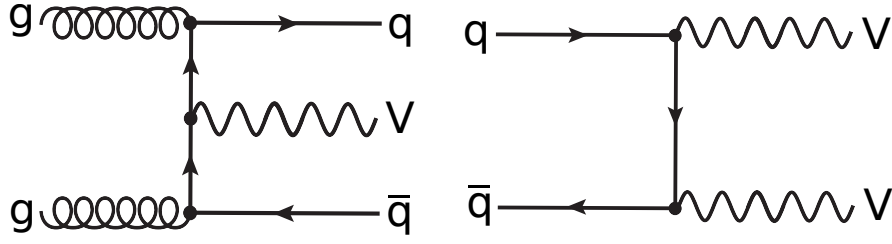


Figure 3.5: Feynman diagrams for V+jets (left) and diboson (right) production.

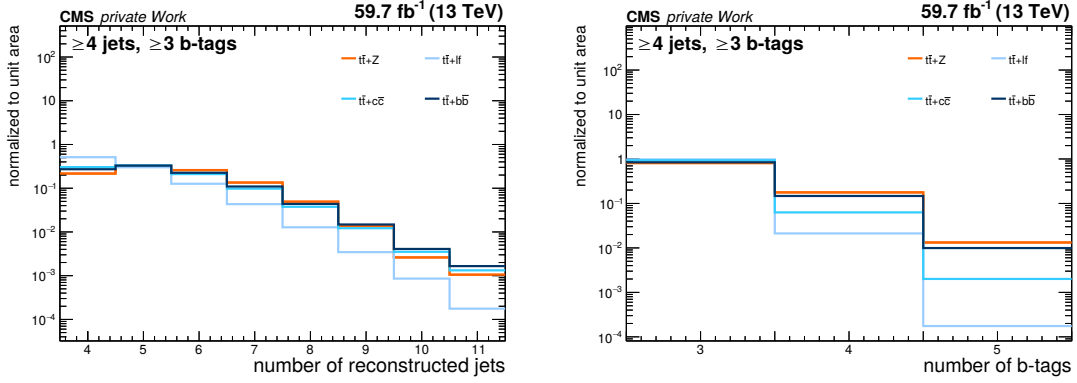


Figure 3.6: Number of reconstructed jets (left) and b-tagged jets (right). The yields of each process are normalized to unit area.

Table 3.1: **Event yields after event selection.** The number of events expected from Monte Carlo simulation per process is shown. The yields are scaled to an integrated luminosity of 59.7 fb^{-1} , corresponding to the expected event yields of the 2018 data run. The shown uncertainties are of statistical nature. Event yields for events with four, five and six or more jets and the signal-to-background ratios S/B and $S/\sqrt{B}$ are listed.

	4 jets	5 jets	≥ 6 jets	total yield
$t\bar{t}+Z$	181.6	276.7	416.8	875.2 ± 3.0
$t\bar{t}+lf$	58 071.7	34 182.3	21 245.9	$113 500.0 \pm 167.3$
$t\bar{t}+b\bar{b}$	13 902.9	16 740.4	20 100.2	$50 743.4 \pm 121.0$
$t\bar{t}+c\bar{c}$	9465.8	10 287.8	11 183.5	$30 937.0 \pm 82.7$
single t	3533.5	2194.7	1486.7	7214.9 ± 41.6
V + jets	1307.4	474.8	371.6	2153.8 ± 198.4
$t\bar{t}+H$	259.6	418.5	684.8	1362.9 ± 1.6
$t\bar{t}+W$	68.9	105.9	167.1	341.9 ± 5.0
diboson	67.5	30.7	13.7	111.9 ± 11.3
S/B	2.1×10^{-3}	4.3×10^{-3}	7.5×10^{-3}	
$S/\sqrt{B}$	0.6	1.1	1.8	
data	91 831	67 781	54 223	213 835

3.3 Event Selection

Events are selected based on reconstructed objects, i.e. leptons, jets and missing transverse energy, as defined in section 3.1. The event selection is designed according to the final state of the signal process, as introduced in the previous section. Instead of performing the event selection directly on PF objects, the selection is performed on physics objects, requiring stringent identification and reconstruction quality criteria. The calibrations and quality criteria for these physics objects are provided by the CMS collaboration and will only be discussed very briefly in this thesis.

Additionally, online event selections (triggers) are used as a base selection. To ensure compatibility between data and simulated events, the object and trigger selections need to be fulfilled in both data and simulation. However, some selections react differently in data and simulation, which is why corrections need to be applied to simulated data, as explained in section 3.4.2. The event yields of the background and signal processes passing all selection criteria are summarized in Table 3.1.

Trigger Selection

A base selection of events happens via triggers. The triggering happens live, during data taking, before data storage. One of the following high-level triggers (HLT) paths that are based on a simplified event reconstruction has to be fired for events to be considered in this analysis:

- At least one isolated electron of $p_T \geq 32$ GeV, or $p_T \geq 28$ GeV and $|\eta| \leq 2.1$ combined with a full-event transverse energy sum of $H_T \geq 150$ GeV.
(HLT_Ele32_WPTight_Gsf_vX OR HLT_Ele28_eta2p1_WPTight_Gsf_HT150_vX)
- At least one isolated muon of $p_T \geq 24$ GeV.
(HLT_IsoMu24_vX)

Jets

Events are required to have at least four AK4-jets with a momentum of $p_T \geq 30$ GeV and a pseudo rapidity of $|\eta| \leq 2.4$. The momentum threshold is introduced to reduce the contamination of low-energy background processes and pileup interactions. The criterion on η ensures a good reconstruction quality, as the instrumentation of the CMS detector is less granular at high pseudo rapidities and the tracker covers only regions of $|\eta| \leq 2.5$.

Prior to the event selection, the jet energy measurement is corrected in several ways. To reduce the impact of pileup collisions, charged PF candidates not originating from the primary vertex are discarded from the clustering (charged hadron subtraction). Jets are vetoed if charged leptons passing the event selection are found within a distance of $R = 0.4$. Furthermore, jet energy calibrations (JEC) are applied following the CMS recommendation [64]. The selection criteria for jets are developed and provided by the CMS Jet-MET physics object group [65, 66].

The number of jets required by this event selection is lower than the number of jets expected from the leading-order final state of the signal process. This is motivated by the stringent requirements on jet identification and reconstruction, jet momenta and pseudo rapidities. Figure 3.6 (left) shows the number of reconstructed jets for the signal process and some of the aforementioned background processes. As can be seen, a large fraction of signal events have fewer than the six expected jets, hence the reduced number of jets in the event selection. Furthermore, the phase space with exactly four jets is enriched in $t\bar{t}+lf$ events as these do not necessarily require additional jets apart from $t\bar{t}$ production. Thus, this region provides a well-suited control region to constrain this background, which can be transferred to the signal-enriched regions with more jets.

b-Tagged Jets

Events are selected that have at least three b-tagged jets. Similar to the number of jets, the number of required b-tagged jets is lower than expected in the LO picture, due to the mistag probabilities of the tagging algorithm and reconstruction efficiencies. The number of b-tagged jets for the signal process and some background processes are also shown in Figure 3.6 (right). A requirement of at least four b-tagged jets would reduce the number of events in the analysis region, while not increasing the signal-to-background ratio considerably.

This selection of jets and b-tagged jets will henceforth be referred to as jet-tag selection. It reduces the contribution of diboson, $t\bar{t}+W$ and $t\bar{t}+\text{jets}$ events to the analysis region.

Charged Leptons

Events are selected that have exactly one well reconstructed and identified electron or muon with a momentum of $p_T \geq 26 \text{ GeV}$ and $p_T \geq 34 \text{ GeV}$, respectively, and a pseudo rapidity of $|\eta| \leq 2.4$. The selection greatly reduces the contribution of QCD multi-jet events and to some extent the contributions of $V + \text{jets}$ and diboson events to the analysis region. This lepton selection will henceforth be referred to as single-lepton selection. The electrons and muons are required to pass the PF identification threshold [67, 68]. Furthermore, electrons and muons are required to be isolated, i.e. are vetoed in close proximity to other PF candidates (i.e. charged and neutral hadrons and photons).

To suppress the selection of non-prompt muons, requirements are set on the distance of the muon track to the primary vertex. Prompt muons are defined as muons originating from W boson decays ($W \rightarrow \ell\nu$) of the top quark. The selection criteria for muons are developed and provided by the CMS Muon physics object group [69].

Non-prompt electrons coming from photon conversions ($\gamma \rightarrow e^+e^-$) in the tracker are identified and vetoed. Due to the gap between barrel and endcap trackers at $1.444 < |\eta| < 1.556$, electrons in this region are discarded. The selection criteria for electrons are developed and provided by the CMS EGamma physics object group [70].

Final states with $W \rightarrow \tau\nu$ decays are not considered explicitly via specific selections. Tau leptons have very short life times and thus decay before reaching the detector region. The decays to electrons or muons ($\tau \rightarrow \ell\nu_\ell\nu_\tau$) are considered indirectly with the single-electron and single-muon selections.

Missing Transverse Energy

To account for the neutrino in the final state, events are required to have a missing transverse energy of at least $\cancel{E}_T \geq 20 \text{ GeV}$. This primarily reduces the contribution of QCD multi-jet events in the analysis region.

3.4 Background Determination

High-energy final-state partons form jets of stable hadronic particles which are reconstructed in the detector. Hence, the analysis needs to be conducted with these particles. After the definition of the event selection, motivated by the kinematics of the signal process, conclusions are drawn by comparing observed data to simulated data. These simulations include the simulation of particles created during the hard interaction process, their hadronization and response to the detector material. The reconstruction of simulated events is carried out in the exact same way as the observed data. This allows for a direct comparison of simulated events with data events, also at a higher level of full-event observables. As the physics process is known for simulated events, these can be categorized and thus be used to study and design analysis methods like event selections and the

construction of classifiers, as described in section 4.1 and 6.4, respectively. The description of the measured data is verified via Goodness-of-Fit tests, described in section 7.1, where the observables used for this analysis are compared in simulation and data.

In the following, a brief overview is given of the different steps of an MC-event simulation, corrections that need to be applied to the simulated events, and finally a listing of MC-based data sets used in this analysis.

3.4.1 Event Generation and Detector Simulation

The generation of simulated events starts with the calculation of matrix elements (ME) of the hard scattering process of the partons in the proton. These ME calculations can be conducted in perturbation theory either at tree-level, i.e. leading-order (LO), or by taking into account higher-order QCD or electroweak corrections. The type of partons and their momenta in the colliding protons are determined from parton distribution functions (PDF). These initial states, convoluted with the integration over the ME's final-state phase-space, yield probability distributions, from which final states are sampled. Subsequently, the fragmentation in QCD (parton showers) and additional radiation of the initial state partons are simulated. Parton showers arise from emission of gluons of the final state particles, and quark-antiquark production from gluon decays at small angles around the original parton.

In the simulations described so far, the partons are assumed to be produced at energy scales where the strong coupling constant α_S is small. After the parton showering, the energies are low enough for the partons to undergo hadronization. Hadronization describes the creation of physical color-neutral hadrons. Due to the large value of α_S at these low energy scales, calculations in the perturbation limit are no longer valid, thus requiring phenomenological models for the simulation of the hadronization. Finally, all unstable particles resulting from this simulation are decayed until only long-lived particles remain. Also considered in this simulation are additional interactions of initial partons (underlying-event).

The detector simulation finally simulates the response of the particles in the CMS detector, using a detailed model of the CMS experiment in the GEANT4 framework [71], including a simulation of the magnetic field and all detector materials and electronics. Additional proton-proton interactions (pileup) are added after detector simulation.

3.4.2 Corrections to Simulated Events

For simulated events to match the distributions observed in data, several corrections are applied after simulation.

Jet Energy Corrections are applied, as previously mentioned. The corrections and calibrations are applied following the recommendations of [64], adjusted to 13 TeV.

b-Tagging Scale Factors need to be applied to the simulation, as previous studies [72, 73] showed that the b-tagging efficiencies differ in simulations and data. The cited methods were applied to simulated data for the DEEPJET tagger and provided by CMS. In summary, scale factors ϵ dependent on p_T , η and the b-tag-value, are derived for heavy-flavor (HF) jets, in HF enriched samples containing exactly two jets, which are expected to have the same flavor (e.g. from dileptonic $t\bar{t}$ decays). One of these jets is required to pass a b-tag cut. A scale factor is derived from the difference between number of events N in data and number of events in simulation where the other jet does not pass the b-tag requirement (non-HF), relative to the number of simulated events, where the other jet passes the requirement (HF):

$$\epsilon_{\text{HF}} = \frac{N^{\text{Data}} - N_{\text{non-HF}}^{\text{MC}}}{N_{\text{HF}}^{\text{MC}}} . \quad (3.4)$$

The same applies for light-flavor jets (LF) in LF enriched samples, where scale factors ϵ_{LF} are derived. Based on these p_{T} , η and b-tag value dependent scale factors, each event in the analysis is assigned a weight according to the product of all scale factors of all reconstructed jets in the event.

Lepton Scale Factors need to be applied to simulated events, to take into account the differences in lepton identification and reconstruction efficiencies in data and simulation [67, 68]. These scale factors are derived similarly to the b-tagging scale factors via a tag-and-probe method in samples enriched with two leptons of same flavor. The scale factors are considered via weights for simulated events.

Lepton Trigger Scale Factors need to be applied due to the differences in trigger efficiencies in simulation and data, and are considered via weights applied to simulated events.

Pileup Corrections need to be applied to simulated data, as the pileup distribution in simulated events does not exactly match the one observed in data. Pileup expresses the total number of proton-proton interactions per bunch crossing. The reweighing of simulated events is performed via weights applied to simulated events.

Generator Weights need to be applied to simulated events, to account for the varying sampling densities of different phase-space regions. Phase spaces where only a small number of data events are expected need to be considered by Monte Carlo simulations, nonetheless. Thus, more events are sampled in these regions compared to the expected number (oversampling).

All corrections to simulated events are applied simultaneously. The aforementioned weights for the corrections are derived independently, thus, an event is weighted with the product of all weights. Most methods used to derive these corrections are subject to systematic uncertainties, which need to be considered when the corrections are applied. Which uncertainties are considered for the analysis of this thesis will be discussed in chapter 5.

3.4.3 Data Sets

The primary data sets used for this analysis consist of events collected in 2018 as part of the LHC Run-II. The integrated luminosity of these data sets corresponds to 59.7 fb^{-1} . Signal and background Monte Carlo events are generated at next-to-leading order accuracy of perturbation theory (NLO) with the POWHEG (v.2) [74–76] or MADGRAPH5_AMC@NLO (v.2.4.2 or v.2.6.0) [77] event generators, or at leading-order (LO) with the PYTHIA (v.8) [78] event generator. The substructure of the colliding protons is described by the parton distribution functions (PDF), where the NNPDF3.1 [40] set was used. Parton showering and hadronization are simulated with PYTHIA. The description of the shower and underlying-event corresponds to the CP5-tune [79] recommended by CMS.

To compare the modeled signal and background processes with the observed distributions in data, the simulated events are normalized to the integrated luminosity determined in data, according to their predicted cross-section. An overview over all samples used in this thesis, including their predicted cross-sections times branching fractions ($\sigma \times \mathcal{B}$) are listed in Table 3.2. All configurations correspond to the recommended CMS analysis guidelines, following [80]. The simulated $t\bar{t}+b\bar{b}$ events are scaled by a factor of 1.407 relative to its predicted cross-section. This scale factor is derived from a fit to data, as explained in section 6.2.

The $t\bar{t}+Z$ signal process is simulated with MADGRAPH5_AMC@NLO. The inclusive cross-section is normalized to $0.86^{+0.07}_{-0.08} (\text{scale}) \pm 0.03 (\text{PDF} + \alpha_S) \text{ pb}$ according to the

latest calculations from [44] with NLO and electroweak corrections and including also next-to-next-to-leading-logarithmic (NNLL) corrections.

Table 3.2: **Monte Carlo samples produced centrally by the CMS collaboration used for Standard Model signal and background prediction.** Also included are the generator, the number of events and the cross-section times branching fraction ($\sigma \times \mathcal{B}$).

Process	Generator	#Events	$\sigma \times \mathcal{B}$
Signal Samples			
$t\bar{t}+Z, Z \rightarrow q\bar{q}$	MADGRAPH5_AMC@NLO	9 641 000	0.60 pb [44]
$t\bar{t}+Z, Z \rightarrow \ell\bar{\ell}/\nu\nu$	MADGRAPH5_AMC@NLO	13 280 000	0.26 pb [44]
$t\bar{t}+Z, Z \rightarrow b\bar{b}$	MADGRAPH5_AMC@NLO	9 782 000	0.13 pb [44]
Background Samples			
$t\bar{t}$ (fully hadronic)	POWHEG (NLO)	133 808 000	377.96 pb [81]
$t\bar{t}$ (semileptonic)	POWHEG (NLO)	101 550 000	365.46 pb [81]
$t\bar{t}$ (dileptonic)	POWHEG (NLO)	64 310 000	88.34 pb [81]
single t , s-channel	MADGRAPH5_AMC@NLO	19 952 000	3.36 pb [82]
single t , t-channel	POWHEG (NLO)	154 307 600	136.02 pb [82]
single $\bar{t}$, t-channel	POWHEG (NLO)	79 090 800	80.95 pb [82]
single t , tW-channel	POWHEG (NLO)	9 598 000	35.85 pb [82]
single $\bar{t}$, tW-channel	POWHEG (NLO)	7 623 000	35.85 pb [82]
$W + \text{jets}, W \rightarrow \ell\nu, 0 \text{ jets}$	MADGRAPH5_AMC@NLO	192 288 265	50 131.98 pb [83]
$W + \text{jets}, W \rightarrow \ell\nu, 1 \text{ jets}$	MADGRAPH5_AMC@NLO	171 669 288	8426.09 pb [83]
$W + \text{jets}, W \rightarrow \ell\nu, 2 \text{ jets}$	MADGRAPH5_AMC@NLO	98 362 049	3172.96 pb [83]
$Z + \text{jets}, Z \rightarrow \ell\bar{\ell}, 0 \text{ jets}$	MADGRAPH5_AMC@NLO	93 979 507	4620.52 pb [83]
$Z + \text{jets}, Z \rightarrow \ell\bar{\ell}, 1 \text{ jets}$	MADGRAPH5_AMC@NLO	96 753 082	859.59 pb [83]
$Z + \text{jets}, Z \rightarrow \ell\bar{\ell}, 2 \text{ jets}$	MADGRAPH5_AMC@NLO	61 848 731	338.26 pb [83]
$t\bar{t}+W, W \rightarrow \ell\nu$	MADGRAPH5_AMC@NLO	4 911 941	0.18 pb [44]
$t\bar{t}+W, W \rightarrow q\bar{q}'$	MADGRAPH5_AMC@NLO	835 296	0.37 pb [44]
WW	PYTHIA (LO)	7 850 000	118.70 pb [84]
WZ	PYTHIA (LO)	3 885 000	65.54 pb [84]
ZZ	PYTHIA (LO)	1 979 000	15.83 pb [84]
$t\bar{t}+H, H \rightarrow b\bar{b}$	POWHEG (NLO)	11 835 999	0.30 pb [85]
$t\bar{t}+H, H \rightarrow \text{non-}b\bar{b}$	POWHEG (NLO)	7 525 991	0.21 pb [85]

4 Analysis Methods

The analysis presented in this thesis aims at measuring the contribution of $t\bar{t}+Z$ events in the previously defined analysis region (see section 3.3). Determining the yield of this process allows a measurement of the cross-section, i.e. the production probability of $t\bar{t}+Z$ events. The most basic approach for this signal extraction would be a so-called cut-and-count experiment, where, after applying some selections, the number of expected background events is compared to the observed number of data events. Due to the small signal cross-section the statistical significance of that measurement would not suffice for a meaningful result. The significance can be increased by analyzing discriminating observables which differ for the signal and background processes. A simple example could be the number of reconstructed jets, which is expected to peak at higher values for $t\bar{t}+Z$ events compared to the major $t\bar{t}$ background processes. The statistical methods of extracting the signal contribution from such discriminating distributions will be explained in section 4.2.

As an additional step, many of these discriminating observables can be combined with multivariate analysis (MVA) methods to construct observables with better discriminating power. In this analysis, Artificial Neural Networks (ANN) are used for this purpose.

4.1 Multivariate Analysis Methods

4.1.1 Motivation

Many high energy physics analyses perform searches of a certain signal process with a very small cross-section compared to the background. The signal and background processes have very similar kinematic properties (c.f. section 3.2). Similar, or equal, kinematic properties of signal and background processes increase the difficulty of using single observables for signal extraction as illustrated with Figure 4.1. Shown is the distribution of the invariant mass of the jet pair in the event with the mass closest to the Z boson mass of 91 GeV in the relevant phase-space for the presented analysis. Both jets are required to be b-tagged (c.f. section 3.1). Even though the only events which contain a Z boson are the signal events, the background processes also show similar peaking structures close to 91 GeV, which makes the discrimination between signal and background processes with this observable a challenge. Under this premise, multivariate analysis methods have become very common in high energy physics analyses, as these methods enable the combination of multiple weakly discriminating observables (like the one shown in Figure 4.1) to construct discriminators

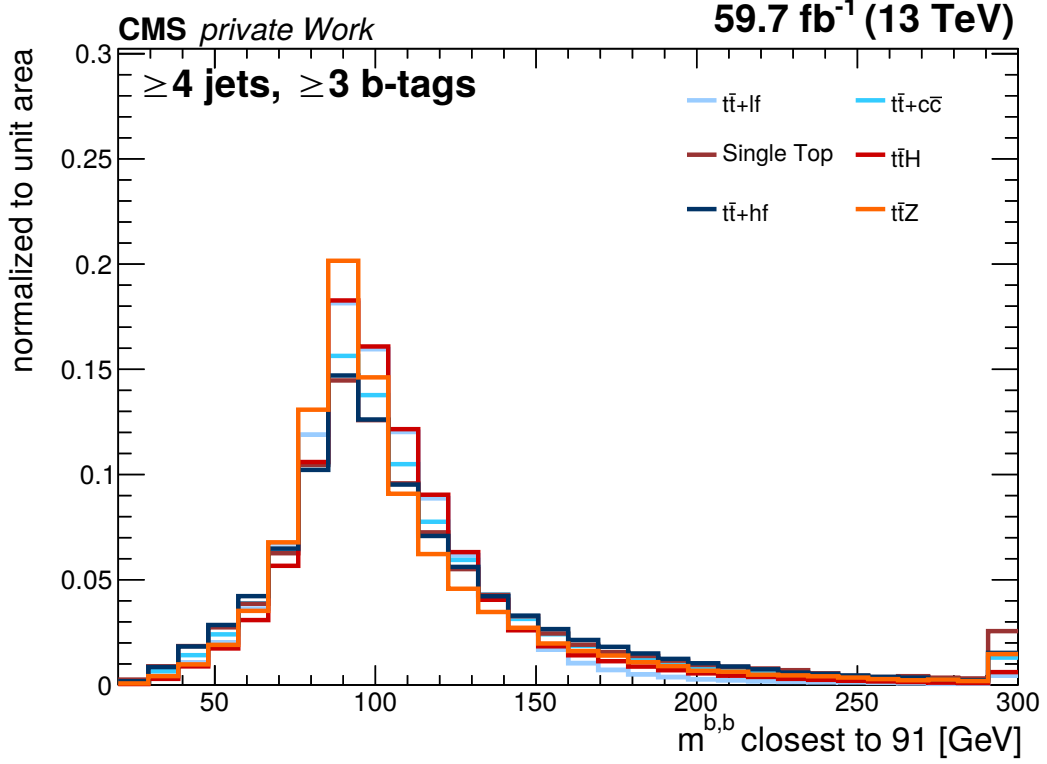


Figure 4.1: **Invariant mass of the jet pair in the event with the mass closest to the Z boson mass.** The yields of each process are normalized to unit area.

which show better discriminating powers. The combination of multiple observables enables the classifier to also use higher order correlations to increase the classification power.

In this thesis, Artificial Neural Networks (ANN) are used. Previous studies [86] in the scope of $t\bar{t}H(bb)$ analyses have shown that shallow ANNs show a sufficient level of discriminating power for this application specifically. While high classification accuracies and low misclassification rates are the main objective of classification tasks in machine learning disciplines like image recognition, this is only of secondary importance for this specific use case. The main objective in high energy physics is the maximization of the signal significance, also taking into account the impact of systematic uncertainties. Hence, the generation of regions (i.e. bins) in the discriminators with varying signal-to-background yield ratios (S/B ratios) is of capital importance. This helps identifying the signal contribution. When also taking into account systematic variations, regions which are enriched in a specific background process will also become important to constrain the contributions of these background processes in the signal-enriched regions. The reasoning for this is laid out more explicitly in section 4.2, where the fit model will be explained.

4.1.2 Artificial Neural Networks

The basic elements of ANNs are based on the Rosenblatt Perceptron [87]. A perceptron takes several features x_i as inputs. These inputs are weighted with weights w_i and afterwards summed. A non-linear activation function f can be applied to the sum, such that the output value o of the perceptron yields

$$o = f\left(\sum_i x_i w_i\right). \quad (4.1)$$

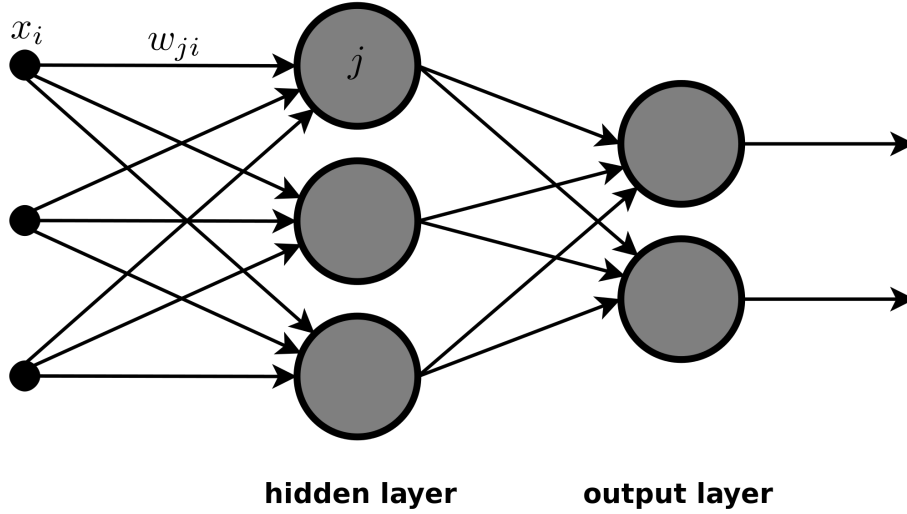


Figure 4.2: **Illustration of an Artificial Neural Network (ANN).** The black dots on the left represent the input features x_i , which are connected to the neurons j in the exemplary hidden layer via weight-connections w_{ji} . The output layer on the right produces output values which can be further processed and interpreted. Taken and adapted from [88].

The perceptron can be used for classification purposes by setting requirements on the output value o for different classes, for example $o > 0$ implying some class A and $o \leq 0$ implying some class B . With these requirements, the perceptron is able to create new decision boundaries for the classification in the multidimensional feature space of the input features x_i , based on the output value o . However, the discrimination power of the perceptron is limited due to its simple structure.

ANNs enable more complicated decision boundaries and also multi-classification boundaries. These ANNs are built from perceptron-like building blocks (neurons) that are arranged in a selected number of layers. An exemplary illustration of a simple ANN is shown in Figure 4.2. Similar to the perceptron, weighted inputs x_i are summed as $z_j = \sum_i x_i w_{ji}$, yielding the input value for the j -th neuron, where i denotes the index of the input feature. Each connection of one input feature and one neuron has a separate weight w_{ji} . The output value o_j of the neuron is calculated by applying an arbitrary, mostly non-linear activation function f to the input value z_j . These output values o_j are re-used as input values for the next layer of neurons, yielding

$$x_j^{(l+1)} = f \left(\sum_i x_i^{(l)} w_{ji}^{(l)} \right), \quad (4.2)$$

where the index l denotes the layer of the ANN.

The ANNs used in this thesis are structured such that the number of nodes in the last layer (output layer) corresponds to the number of classes (i.e. physics processes) to be categorized. For example, in the final analysis strategy presented in section 6.5, there is one output node each for the $t\bar{t}+Z$, $t\bar{t}+H$, $t\bar{t}+lf$, $t\bar{t}+c\bar{c}$ and $t\bar{t}+b\bar{b}$ process. By applying the Softmax activation function

$$o_j = \frac{e^{z_j}}{\sum_i e^{z_i}}, \quad (4.3)$$

the sum of all output values o_j is set to one, which enables the interpretation of this output as some kind of probability. A high output value, for example in the $t\bar{t}+Z$ -node, would suggest that the evaluated event with its input features x_i is $t\bar{t}+Z$ -like.

In order to reach high classification accuracies, the weights w_{ji} need to be adjusted based on the outputs o_j of single events. For this purpose, data is needed where the underlying physics process, which is supposed to be classified, is known (“labeled data”).

By propagating a labeled event through the ANN, the predictions o_j can be compared to the truth $\hat{o}_j$. For the classification purpose of this thesis, the truth of an event is always one in the output node of its true class and zero otherwise. With the help of a loss function L , the quality of the prediction can be judged. If not stated otherwise, categorical cross-entropy, defined as

$$L(o, \hat{o}) = - \sum_j [\hat{o}_j \log(o_j) + (1 - \hat{o}_j) \log(1 - o_j)] , \quad (4.4)$$

has been used as a loss function in this thesis. It is especially suited for multi-classification purposes with a probabilistic interpretation of the output values [89, 90], as it targets output values of zero or one. A loss function in general should converge to zero if the prediction of the ANN matches the underlying truth. Based on the value of the loss function, the weights $w_{ji}^{(l)}$ of the ANN are updated, following

$$w_{ji}^{(l)} \leftarrow w_{ji}^{(l)} - \lambda \frac{\partial L}{\partial w_{ji}^{(l)}} , \quad (4.5)$$

where λ is a tunable hyper-parameter of the ANN (“learning rate”). The learning rate determines the amount by which the weights are updated. This update aims at minimizing the value of the loss function and thereby maximizing the prediction accuracy. Updating the weights with this gradient is referred to as GRADIENTDESCENT algorithm. The ANNs in this thesis were trained with an adaption of this algorithm, the ADAGRAD algorithm [91]. In general, the divergence of the loss function with respect to the weights can be expanded to

$$\frac{\partial L}{\partial w_{ji}^{(l)}} = \frac{\partial L}{\partial o_k^{(N)}} \frac{\partial o_k^{(N)}}{\partial o_m^{(N-1)}} \cdots \frac{\partial o_m^{(l+1)}}{\partial o_j^{(l)}} \frac{\partial o_j^{(l)}}{\partial w_{ji}^{(l)}} , \quad (4.6)$$

where the index N denotes the maximum number of layers in the ANN. With the backpropagation-of-error algorithm [92], the divergences $\partial o^{(l+1)} / \partial o^{(l)}$ can be calculated iteratively starting from the output layer N . This reduces the high redundancy of arithmetic operations during this learning process, thereby speeding up the training of the ANNs. For a more thorough explanation of the backpropagation-of-error algorithm and the learning process of ANNs, the reader is referred to [93, 94].

During this thesis, a framework [95] based on KERAS [96] was developed to integrate the training and evaluation of ANNs into the already existing analysis workflows. This framework aims at fast and easy training, evaluation and optimization of the ANNs. It was used in the $t\bar{t}H(bb)$ -analysis in the semileptonic $t\bar{t}$ decay channel of 2017 data [97] and is used in current analysis projects in CMS.

The construction of classifiers specifically for this thesis will be further discussed in section 6.4.

4.1.3 Regularization Methods

By construction, every weight w_{ji} in an ANN is a parameter in a multi-dimensional fit. Due to the high dimensionality of the ANNs the number of weights is in the order of thousands. ANNs with many nodes, i.e. many parameters are prone to over-fitting the predictions to the events used for training. However, the purpose of an ANN is to generate a generalized classification that is also applicable for events that have not been used during the training

process itself. This is especially important for data events, which can by definition not be labeled with a certain process, but nevertheless require to be classified analogous to simulated Monte Carlo events. Thus, methods to regularize and monitor the generalization of the trained ANNs need to be applied.

Specifically, L2-REGULARIZATION [98], to counter large weights, and DROPOUT [99], to reduce the influence of single neurons, were used to regularize the training process of the ANNs used in this analysis. Furthermore, the behavior of the loss function values is monitored during the training process, separately for the data set which is used for training (train set) and an independent data set (validation set). If the loss values of the train set decrease steadily while the loss values of the validation set diverge, this is a sign for loss of generalization, as the ANN is over-fitted to the train set and is not generalizable to the validation set anymore. After the training process, the discriminator distributions of the ANN output nodes are cross-checked between the train set and another previously unseen data set (test set) in compatibility tests (see section 7.2). Only ANNs for which the signal and background distributions for test and train sets are compatible are used for the statistical interpretation.

4.2 Statistical Methods

The general statistical methods needed for the interpretation of the results are described in the following sections and are based on [100–102]. The calculations necessary for the evaluation are performed with the `combine` [103] framework of the CMS collaboration which was originally designed for Higgs analyses. In the previously defined analysis region, the background yields surpass the expected $t\bar{t}+Z$ yields (see Table 3.1) by multiple orders of magnitude. Therefore, the direct confirmation of the previously measured $t\bar{t}+Z$ cross-section [6, 8] is not possible because the needed sensitivities cannot be reached. Instead, a limit on the signal-strength is set. The signal-strength is defined as the observed cross-section relative to the theory prediction:

$$\mu(t\bar{t}+Z) = \frac{\sigma(t\bar{t}+Z)^{\text{obs.}}}{\sigma(t\bar{t}+Z)^{\text{th.}}} . \quad (4.7)$$

The procedure of generating exclusion limits is described in section 4.2.3.

4.2.1 Statistical Tests

A hypothesis containing the signal process with the cross-section predicted by theory calculations, and the background processes predicted by the Standard Model has to be tested. It is commonly referred to as signal-plus-background (S+B) hypothesis. It is customary to also test a background-only hypothesis, containing only background processes, excluding the sought-after signal process. Both hypotheses are tested by computing a probability (or p -value) under the assumption of the hypothesis under scrutiny for the observed data. The thereby calculated p -value aims at quantifying the probability to observe data which is equal or less compatible with the prediction made by the hypothesis tested. For the purpose of excluding a hypothesis, a 95% confidence level corresponding to a p -value of 0.05 is commonly used.

The p -value can be converted into a measure of significances, based on the standard-deviation intervals of the Gaussian distribution. The significance needed for a discovery of a new signal corresponds to a p -value of 2.87×10^{-7} and is referred to as a significance of 5σ .

The hypothesis tests are performed using a test statistic, which is a function of the data and the parameters of the hypothesis and quantifies their compatibility. Following the Neyman-Pearson lemma [104], the test statistic is constructed as a likelihood ratio. Commonly, the signal-strength μ (eq. 4.7) is used as the parameter of interest (POI) to directly quantify the deviation from the theoretical cross-section calculation. Typically, additional nuisance parameters (NPs) θ are introduced. Examples of such nuisance parameters are uncertainties on theory cross-section calculations or detector efficiencies. An overview of the systematic uncertainties relevant for this analysis is given in chapter 5. The nuisance parameters are additional degrees of freedom and enable the fit to properly adjust the simulated prediction to the observed data. Often, the systematic uncertainties are estimated in unrelated measurements, which yield a best estimate $\tilde{\theta}$ that can be used in the modeling of the uncertainty and its probability density function (p.d.f.) $\rho(\theta|\tilde{\theta})$.

4.2.2 Maximum Likelihood Estimation

The values of the POIs μ and nuisance parameters θ that best describe a given set of observed data are estimated via the Maximum Likelihood (ML) method. The likelihood function is constructed as a product of probability densities $\mathcal{P}$ as

$$\mathcal{L}(\mu, \theta) = \prod_i \mathcal{P}(n_i | E[n_i]) \cdot p(\tilde{\theta} | \theta), \quad (4.8)$$

where $p(\tilde{\theta} | \theta)$ are the prior probabilities of the nuisance parameters p.d.f. related to $\rho(\theta | \tilde{\theta})$ via Bayes' theorem.

The distributions observed in high energy physics are usually treated as discrete histograms. Thus, the probability densities $\mathcal{P}(n_i | E[n_i])$ express the probability to observe n_i events in a specific bin while expecting

$$E[n_i] = \mu \cdot s_i + b_i, \quad (4.9)$$

events. Here, b_i and s_i denote the expected number of background and signal events in this specific bin, respectively. The value s_i is multiplied by the signal-strength modifier μ to preserve the degree-of-freedom introduced by the POI, which allows to test various hypothesized values of μ . The number of observed events in general follows a Poisson distribution

$$\mathcal{P}(n_i | E[n_i]) = \frac{(E[n_i])^{n_i}}{n_i!} e^{-E[n_i]}. \quad (4.10)$$

From the ML function (eq. 4.8), a negative log-likelihood ratio is constructed as test statistic $\tilde{q}_\mu$:

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \text{ with } 0 \leq \hat{\mu} \leq \mu. \quad (4.11)$$

The set of parameters which maximize the likelihood $\mathcal{L}$ for a fixed value of μ are denoted as $\{\mu, \hat{\theta}_\mu\}$, while the set of parameters which globally maximize $\mathcal{L}$ are $\{\hat{\mu}, \hat{\theta}\}$. The use of the negative log-likelihood enables the use of minimization algorithms to find a value of μ which maximizes the likelihood $\mathcal{L}$, i.e. minimizes the negative log-likelihood $\tilde{q}$, and the treatment of $\mathcal{L}(\hat{\mu}, \hat{\theta})$ as a common offset for all hypotheses of μ .

4.2.3 Exclusion Limits

The results of this thesis will be interpreted as exclusion limits on possible values of signal-strength modifiers μ . The procedure to derive such exclusion limits from the negative log-likelihood test statistic $\tilde{q}$ (eq. 4.11) will be briefly discussed in this section. For a more detailed explanation, the reader is referred to [100].

By setting an upper limit μ on the signal-strength modifier, the best estimate $\hat{\mu}$ is constrained to $0 \leq \hat{\mu} \leq \mu$ with the given confidence level. The lower limit of zero is motivated by physical interpretation which requires the signal-strength to be positive. To determine the limit, a p -value is defined as

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | \text{S+B}) = \int_{\tilde{q}_\mu^{\text{obs}}}^{\infty} d\tilde{q}_\mu \cdot f(\tilde{q}_\mu | \mu, \boldsymbol{\theta}_\mu^{\text{obs}}) \equiv \text{CL}_{\text{s+b}}, \quad (4.12)$$

which is interpreted as the probability of the value of the test statistic $\tilde{q}_\mu$ to be higher than the observed value $\tilde{q}_\mu^{\text{obs}}$ under the S+B hypothesis. The p.d.f. of $\tilde{q}_\mu$ for a hypothesized μ is $f(\tilde{q}_\mu | \mu)$. Analogous, $1 - p_0 \equiv \text{CL}_b$ can be calculated, where p_0 quantifies the level of agreement for the background-only hypothesis, i.e. $\mu = 0$ with $f(\tilde{q}_\mu | 0, \boldsymbol{\theta}_0^{\text{obs}})$.

If p_μ of a given μ is smaller than a certain threshold α , the hypothesis for this μ is said to be excluded with a $(1 - \alpha) \text{CL}_{\text{s+b}}$ confidence level. It is common to quote these exclusion limits for $\alpha = 0.05$, corresponding to a confidence level of 95%. Due to the construction of the $\text{CL}_{\text{s+b}}$ confidence level, a hypothesis is said to be excluded for values smaller than the chosen threshold α , although the experiment might have no sensitivity on the signal-strength in cases where the background fluctuates to low values. Thus, a confidence level

$$\text{CL}_s = \frac{\text{CL}_{\text{s+b}}}{\text{CL}_b}, \quad (4.13)$$

is commonly used to set a more conservative limit on the signal-strength μ . If for a given μ the CL_s confidence level is smaller than the threshold α , the hypothesis corresponding to μ is said to be excluded with a $(1 - \alpha) \text{CL}_s$ confidence level.

Observed limit

The values for $\hat{\boldsymbol{\theta}}_0^{\text{obs}}$ and $\hat{\boldsymbol{\theta}}_\mu^{\text{obs}}$ which describe best the observed data under the background-only and S+B hypotheses, respectively, are determined. With these values, p.d.f.s $f(\tilde{q}_\mu | 0, \hat{\boldsymbol{\theta}}_0^{\text{obs}})$ and $f(\tilde{q}_\mu | \mu, \hat{\boldsymbol{\theta}}_\mu^{\text{obs}})$ are sampled. These p.d.f.s are then used to calculate the aforementioned p -values and 95% CL_s confidence levels for the corresponding hypotheses of μ . As the best-fit values of the nuisance parameters $\boldsymbol{\theta}$ are determined in-situ, the inclusion of signal-depleted regions, which are more sensitive to the nuisance parameters originating from background processes, can be motivated.

Expected limit

Expected limits are calculated to quantify the expectation of the observed limit under the assumption that there is no signal in data. Expected limits are evaluated on pseudo data, i.e. simulated data sets, alone. A straight forward way is to generate a set of pseudo-data consisting of background processes only. The 95% CL_s confidence levels can then be calculated with this data set by treating it as if it were real data. Via this method, one obtains a distribution of upper limits. The median of this distribution is referred to as the expected limit, while the $\pm 1\sigma(\pm 2\sigma)$ intervals are obtained from the 16%(2.5%) and 84%(97.5%) quantiles of the distribution.

Because this approach is computationally very demanding, an approximative method is commonly used which estimates the upper limit asymptotically in the limit of large data sets [102]. This is based on an Asimov data set, which is defined as the data set where all nuisance parameters $\boldsymbol{\theta}$ and observed numbers of events correspond exactly to the central values predicted by the simulated pseudo data.

5 Systematic Uncertainties

The goal of this analysis is to determine the contribution of signal events to the previously defined analysis region. Simulating the contributions of signal and background processes via Monte Carlo simulations enables the estimation of the $t\bar{t}+Z$ contribution via Maximum Likelihood (ML) fits as discussed in section 4.2. The generation of simulated events as well as the actual measurement and reconstruction of data and simulated events are subject to inherent systematic uncertainties. These uncertainties need to be considered in the statistical interpretation of this analysis. An overview of the sources of systematic uncertainties considered in this analysis and how they are treated is given in this chapter.

5.1 Treatment of Systematic Uncertainties

Uncertainties are parameterized as nuisance parameters (NPs) θ in the statistical model. These parameters are not of special interest for the result itself, but are to be considered and constrained in-situ during the fit. The consideration of these nuisance parameters, on the one hand, guarantees that the fit model is flexible enough and, on the other hand, leads to a more physically motivated estimation of the parameters of interest (POIs) as defined in section 4.2. In the statistical model, one can distinguish between two types of systematic uncertainties:

Rate uncertainties are systematic uncertainties that only have an effect on the absolute rate of a process. The probability density functions (p.d.f.) $\rho(\theta|\tilde{\theta})$ of these uncertainties are modeled with a log-normal distribution

$$\rho(\theta|\tilde{\theta}) \propto \frac{1}{\ln \kappa} \frac{1}{\theta} \exp \left[-\frac{\left(\ln(\theta/\tilde{\theta}) \right)^2}{2 (\ln \kappa)^2} \right], \quad (5.1)$$

where $\tilde{\theta}$ is the expected value and κ is the width of the distribution. For example, a width of $\kappa = 1.05$ implies that the observed value θ can be 5% smaller or larger within one standard-deviation. The log-normal distribution is zero for $\theta = 0$, which is well-suited for this purpose, as the p.d.f.s are restricted to positive values of θ . The implementation of rate uncertainties in the **combine** framework [103] also allows for asymmetric widths κ for up and down variations, i.e. increasing or decreasing the rates of a process, respectively.

Shape uncertainties are systematic uncertainties that change the distribution of events in the templates which are used for fitting. These uncertainties are modeled by two additional

templates per process that correspond to shifts of one standard-deviation of the nuisance parameter in both directions. Purely shape-changing effects are modeled on a bin-by-bin basis with a Gaussian p.d.f., where the $\pm 1\sigma$ variations correspond to the varied templates. For all other values of θ , the templates must be interpolated (template morphing). The interpolations of each bin content are assumed to be quadratic for interpolations smaller than one standard-deviation and linear for larger ones. Often, shape-changing uncertainties also change the overall rates of the processes. In such cases, only the shape effect is modeled as discussed, while the rate-changing effect is treated separately with a log-normal p.d.f. as previously described.

5.2 Sources of Systematic Uncertainties

The following sources of systematic uncertainties are considered in the estimation of the upper limits on the $t\bar{t}+Z$ signal-strength. Individual uncertainty sources are treated as completely uncorrelated. If a source affects multiple processes, the uncertainty is fully correlated between those processes, unless stated otherwise. The names of the associated nuisance parameters that will be used at a later stage are shown in brackets.

Integrated Luminosity [lumi_2018]

The uncertainty on the measurement of the integrated luminosity of the LHC is covered with a rate uncertainty of 2.5% [51].

Jet energy scale [JES]

The measurement of energies in the calorimeters is subject to systematic uncertainties, for example energy calibrations or variations in the response to different particles. The four-vectors of simulated jets are varied to encompass these detection uncertainties. The variations of the jet four-vectors are propagated to the physics observables and are treated as shape variations. Different sources of uncertainties on the measurement of the jet energy are considered as separate nuisance parameters and are treated as uncorrelated, following the recommendation in [105], based on [64]. In total, 19 sources are considered.

Jet energy resolution [JER]

Jet energy measurements are also subject to uncertainties due to the energy and momentum resolutions in calorimeter and tracker, respectively, as described in section 3.1. The jet energy is corrected accordingly. The uncertainty on this correction is covered by varying the simulated jet four-vectors via Gaussian smearing. This nuisance parameter is treated as a shape uncertainty.

Lepton (trigger) scale factors [eff_e/m, trig_e/m]

As mentioned in section 3.4.2, not all electrons and muons can be identified and reconstructed, which is why scale factors need to be applied to the simulated events to mitigate discrepancies between the predictions from simulation and the observed data. The selection efficiencies are measured using a tag-and-probe method [67, 68] in bins of lepton p_T and η . Uncertainties on these scale factors are considered for electrons and muons separately, the corresponding nuisance parameters are implemented as shapes.

Similarly, the efficiencies of triggers differ between data and simulation. Therefore, scale factors to account for the trigger efficiencies of electrons and muons, and corresponding uncertainties are also considered.

b-tagging scale factors [btag]

Similar to the aforementioned lepton scale factors, the scale factors derived for b-tagging (c.f. section 3.4.2) receive additional uncertainties from statistical effects on the scale factor derivation. Furthermore, additional uncertainties due to contamination of background processes in the control regions that are used to derive these scale factors are considered [73,

106]. These uncertainties are treated separately for the statistical and systematic effects, and separately for heavy-flavor [hf] and light-flavor [lf] induced jets. Furthermore, the statistical contribution is split into a linear contribution [stats1] covering tilts in the distributions and a quadratic contribution [stats2] covering shifts between the core and tails of the distribution. As the construction of a control sample enriched in charm-quark-induced jets is hard, uncertainties on charm-induced jets [cferr1/2] are inferred from the heavy-flavor uncertainties.

Pileup [PU]

Uncertainties in the determination of the number of pileup interactions (c.f. section 3.4.2) are covered with a shape uncertainty. It is derived by calculating predictions on the number of pileup interactions by varying the total inelastic proton-proton cross-section at LHC conditions by 4.6% [107].

PDF set [PDF, pdf_alphaS]

During the simulation of events a specific set of parton distribution functions (PDF) is used for the description of the substructure of the colliding protons. For this analysis, the NNPDF3.1 [40] set was used. The PDF sets are determined from fits to data observed by experiments at the LHC and others, resulting in systematic and statistical uncertainties, e.g. by the choice of data used to determine the PDF set. These uncertainties are covered with a nuisance parameter constructed as a shape uncertainty.

Additionally, separate rate uncertainties for gluon-gluon, quark-quark and quark-gluon induced processes are considered. These nuisance parameters cover rate variations of uncertainties on the cross-section due to the PDFs and the choice of the strong coupling constant α_S , i.e. the energy scale of hadronization, as summarized in Table 5.1.

Factorization and Renormalization scale [scaleMuR/MuF, QCDscale]

The computation of matrix elements (ME) for the calculation of cross-sections requires the introduction of a physical energy-scale (renormalization scale μ_R) to cancel out divergences in the calculations. Optimally, the ME calculations should not depend on this parameter, but due to the finite order of perturbation theory calculations this is usually not the case, and the choice of μ_R has to be considered in the evaluation. According to the factorization theorem, another physical scale μ_F has to be introduced, corresponding to the energy where the parton shower simulation takes over from the ME generator. Thus, systematic uncertainties covering the choice of μ_R and μ_F need to be considered. By convention, these uncertainties are estimated by varying the scales by factors of two up and down with respect to the nominal scale. Both, renormalization and factorization scale uncertainties are treated as uncorrelated shape uncertainties, and are also decorrelated between the $t\bar{t}+Z$, $t\bar{t}+\text{jets}$ and $t\bar{t}+H$ processes to account for the differing energy scales of the radiation of additional particles (i.e. Z, H, gluon) from the $t\bar{t}$ system.

Additionally, rate uncertainties for $t\bar{t}$ induced processes, single top, diboson, $V+\text{jets}$ and the $t\bar{t}+Z$ signal processes are considered, covering uncertainties in rate differences between theory calculations of cross-sections and the order of perturbation used in simulation. For example, the $t\bar{t}+Z$ signal process is simulated at NLO, while the assumed cross-section was calculated at NLO including NNLL corrections [44]. These uncertainties are summarized in Table 5.2.

Initial and final state radiation [ISR/FSR]

The PYTHIA parton showering process exhibits intrinsic systematic uncertainties due to the choice of the strong coupling constant α_S in the showering process, affecting the amount of additional gluon radiation. The uncertainty is split into two contributions, accounting for the parton shower uncertainties on gluons from matrix element simulations (initial-state radiation, ISR) and final-state radiation (FSR) in the parton showering process itself. The

uncertainties are approximated by varying the value of α_S as described in [108]. Both, initial and final state radiation are treated as uncorrelated shape uncertainties, and are also decorrelated for the $t\bar{t}+H$, $t\bar{t}+Z$ and all $t\bar{t}$ +jets processes separately as the description of additional radiation can affect these processes differently, especially the $t\bar{t}$ +jets processes.

Background normalization [bgnorm]

A rate uncertainty of 50% is applied to the $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$ processes separately, to provide sufficient flexibility of the heavy-flavor models. These uncertainties cover differences which may arise in different simulations of these $t\bar{t}$ +hf processes, e.g. in the four and five flavor schemes or ME and PS calculations. This follows the approach of previous $t\bar{t}H(bb)$ analyses [97, 109]. In a fraction of the $t\bar{t}+b\bar{b}$ events the two b-jets are strongly collimated, such that they are clustered within one single jet (referred to as $t\bar{t}+2b$). An additional 50% rate uncertainty is assigned to this contribution, designed to cover the additional uncertainties in the modeling of the additional gluon splitting.

Additionally, an uncertainty on the $t\bar{t}+H$ cross-section of $+24.6\%/ -20.6\%$ is considered, to cover the experimental uncertainty on the $t\bar{t}+H$ measurement from its observation [110] (see also discussion in section 7.4).

Number of simulated events [autoMCStats]

Due to the finite number of simulated events for the generation of signal and background templates, an uncertainty is considered to cover the statistical fluctuations of the signal and background predictions, following the approach of [111]. Technically, this is implemented in the `combine` framework [103], configured as `autoMCStats 5 0 1`. It introduces a single Gaussian nuisance parameter in bins where the number of events, weighted according to the corrections necessary for simulated events (c.f. section 3.4.2), exceeds a threshold of five. For bins below this threshold, a Poissonian-distributed nuisance parameter is introduced for every process, respectively.

Table 5.1: **PDF and α_S rate uncertainties.** These nuisance parameters cover rate uncertainties on the cross-section due to the PDFs and the choice of the strong coupling constant α_S . The uncertainties (columns) are treated uncorrelated, but one uncertainty can affect multiple processes (lines). The uncertainties are split according to the dominant production process and separately for the signal process $t\bar{t}+Z$. Values taken from [44, 81, 82, 84, 112].

pdf_alphaS_	t $\bar{t}$ Z	gg	qqbar	qg
t $\bar{t}$ +Z	$\pm 3.5\%$			
t $\bar{t}$		$\pm 4.2\%$		
t $\bar{t}$ +H		$\pm 3.6\%$		
t $\bar{t}$ +W			$\pm 3.6\%$	
V+jets			$\pm 3.8\%$	
diboson			$\pm 5.0\%$	
single t				$\pm 2.8\%$

Table 5.2: **Renormalization and factorization rate uncertainties.** These uncertainties cover uncertainties in rate differences between theory calculations of cross-sections and the order of perturbation used in simulation. The uncertainties (columns) are treated uncorrelated, but one uncertainty can affect multiple processes (lines). Values taken from [44, 81, 82, 84, 112].

QCDscale_	ttZ	ttbar	V	VV	singlet
t $\bar{t}$ +Z	+8.1%/-9.3%				
t $\bar{t}$		+2.4%/-3.5%			
t $\bar{t}$ +H		+5.8%/-9.2%			
t $\bar{t}$ +W		+25.5%/-16.4%			
V+jets			+0.8%/-0.4%		
diboson				$\pm 3\%$	
single t					+3.1%/-2.1%

6 Analysis Strategy

In this chapter, the strategy of the presented analysis is discussed. As an introduction, previous measurements of the $t\bar{t}+Z$ cross-section by the CMS and ATLAS collaboration are briefly reviewed. This analysis follows the $t\bar{t}H(bb)$ analysis, also described in section 6.1. Adjustments applied to the $t\bar{t}$ +jets background modeling are described in section 6.2. In section 6.3, the reconstruction of $t\bar{t}+Z$ events with a χ^2 -based method is explained. As this analysis relies on an event classification based on Artificial Neural Networks (ANN), this event-classification strategy will be discussed in section 6.4, for which the technical details have already been introduced in section 4.1. An overview of the explored analysis strategies concerning, among others, ANN setups will be given in the final section of this chapter.

6.1 Related Previous Results

Both the CMS and ATLAS experiments have published inclusive cross-section measurements of the $t\bar{t}+Z$ process obtained with data taken in the year 2016, corresponding to an integrated luminosity of 35.9 fb^{-1} .

The ATLAS and CMS experiments conducted their analyses in a final state of two, three or four charged leptons, targeting decays of the Z boson into charged leptons [6, 7]. This decay channel exhibits the highest sensitivity to the $t\bar{t}+Z$ cross-section despite contributing only 10% to the total Z boson decay rate [36]. The high sensitivity arises from the good background reduction by requiring a multitude of charged leptons, i.e. electrons or muons, which can be identified with low misidentification rates, thereby reducing QCD and $t\bar{t}$ backgrounds to a large degree. Furthermore, a good object resolution for charged leptons enables a high resolution of the Z boson mass. An inclusive cross-section of

$$\sigma(t\bar{t}+Z)^{\text{ATLAS}} = 0.95 \pm 0.08(\text{stat}) \pm 0.10(\text{syst}) \text{ pb}, \quad (6.1)$$

in a simultaneous measurement of the $t\bar{t}+Z$ and $t\bar{t}+W$ cross-sections was reported by the ATLAS collaboration. The significance of both the ATLAS and CMS measurement exceeds the 5σ threshold.

The CMS result has been updated for a combination of 2016 and 2017 data corresponding to an integrated luminosity of 77.5 fb^{-1} [8]. This new result incorporates for example enhanced lepton identification procedures based on MVA methods, improving the sensitivity of the measurement. Furthermore, differential cross-sections as a function of the transverse

momentum of the Z boson and $\cos\theta_Z^*$ are measured. The latter is the cosine of the angle between the Z boson and the negatively charged lepton from Z decay in the Z rest frame. In this analysis, the observed inclusive cross-section of $t\bar{t}+Z$ production is reported as

$$\sigma(t\bar{t}+Z)^{\text{CMS}} = 0.95 \pm 0.05(\text{stat}) \pm 0.06(\text{syst}) \text{ pb}. \quad (6.2)$$

Due to the large signal-to-background ratios in the phase spaces of these measurements and the different event kinematics and signatures compared to the $Z \rightarrow b\bar{b}$ final state, the analysis strategies applied in these measurements could not easily be transferred to the analysis presented in this thesis. Hence, the analysis was instead modeled closely to the CMS measurements of the $t\bar{t}+H$ cross-section in the $H \rightarrow b\bar{b}$ decay channel [97, 109], which are performed in a similar phase space. The definition of the phase space and event selection for the analysis presented in this thesis are modeled after the combined 2016 and 2017 data analysis [97]. There, multi-classification ANNs are used to separate between the $t\bar{t}+H$ signal process and different $t\bar{t}+\text{jets}$ background processes.

The yield of $t\bar{t}+Z$ events in this thesis is found to be approximately 62% of the $t\bar{t}+H$ yield in 2018 simulations (c.f. Table 3.1). Due to the very similar kinematic properties of $t\bar{t}+Z$ and $t\bar{t}+H$ events (see section 3.2), similar classification of events belonging to these processes is to be expected, i.e. the kinematic and ANN distributions of both processes are expected to be similar. This might reduce the sensitivity to the signal cross-section, both for $t\bar{t}+H$ or $t\bar{t}+Z$. Thus, improvements with respect to the previous $t\bar{t}H(bb)$ analysis are introduced in this thesis, to separate between $t\bar{t}+H$ and $t\bar{t}+Z$ events, e.g. by classifying both $t\bar{t}+H$ and $t\bar{t}+Z$ in the ANN approach and constructing event features specifically designed for a separation between $t\bar{t}+H$ and $t\bar{t}+Z$. A next step could be a simultaneous measurement of $t\bar{t}+H$ and $t\bar{t}+Z$ cross-sections, as briefly discussed in section 7.4.

6.2 Adjustments to the $t\bar{t}+\text{jets}$ -Background Model

Previous measurements of $t\bar{t}+b\bar{b}$ cross-sections [113, 114] show a significant offset between the prediction from theory and the measurements. In each case, the theory prediction was found to be lower than the measured cross-section. This effect might be related to the description of additional b-jets in the analyzed phase space, e.g. originating from the parton shower simulation with its current CP5-tune.

Based on these observations, a scale factor for the $t\bar{t}+b\bar{b}$ contribution is derived for this analysis, to obtain a better agreement between simulation and data. For this purpose a side-band region outside a Z boson mass window of 70 – 110 GeV is defined. A proxy for the Z boson mass is reconstructed from the invariant mass of two b-tagged jets in an event closest to the Z boson mass (see Figure 4.1). In this side-band region, a fit of the $t\bar{t}+b\bar{b}$ contribution is performed to data, without accounting for the systematic uncertainties introduced in chapter 5. The fit is performed for one bin only, containing all simulated and data events in the analysis region outside the Z boson mass veto.

The resulting best-fit value of the signal-strength of $t\bar{t}+b\bar{b}$,

$$\mu(t\bar{t}+b\bar{b}) = \frac{\sigma(t\bar{t}+b\bar{b})^{\text{obs.}}}{\sigma(t\bar{t}+b\bar{b})^{\text{th.}}}, \quad (6.3)$$

is found to be $\mu(t\bar{t}+b\bar{b}) = 1.407 \pm 0.016$, which is within the 50% rate uncertainty assumed for the $t\bar{t}+b\bar{b}$ contribution. The Z boson mass veto was introduced to reduce the number of $t\bar{t}+Z$ events in the fitted region which might bias the final measurement. However, the result was cross-checked for the full analysis region, yielding a similar scale factor. In Figure 6.1, the number of b-tagged jets per event is shown before applying the derived

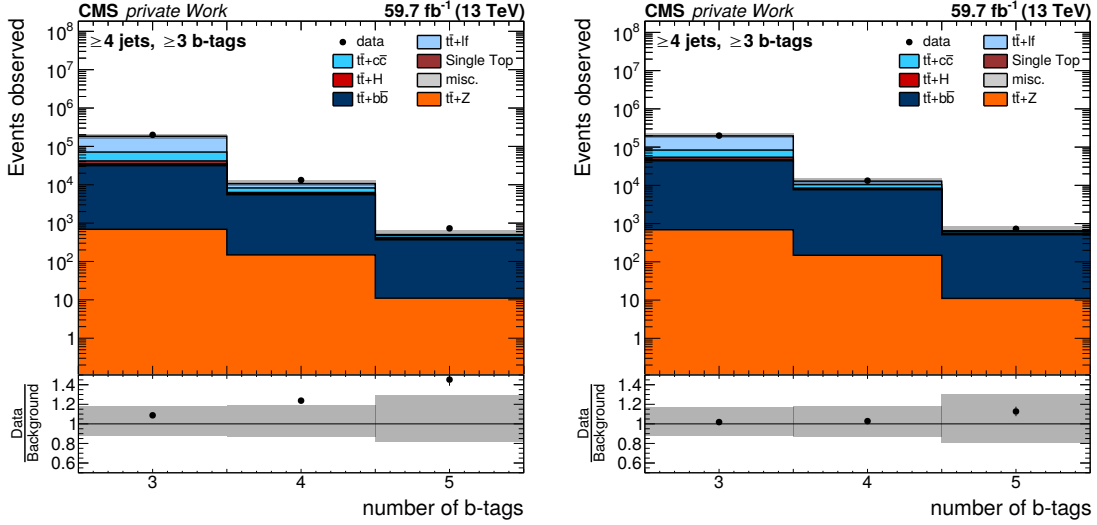


Figure 6.1: **Data/MC comparisons of the number of b-tagged jets in the analysis region before (left) and after (right) applying the scale factor for $t\bar{t}+b\bar{b}$ derived from data.** The predicted contributions of all simulated samples are stacked. Ratios between data (black dots) and all simulated samples are shown in the bottom. The error bands correspond to the systematic uncertainties of the signal and background contributions with a shape-changing effect.

scale factor of 1.407 on the left and after applying it on the right in the entire analysis region. The data/MC ratio shown in the bottom panel shows large deviations from 1, especially for a higher numbers of b-tagged jets in the beginning, which are reduced after scaling the $t\bar{t}+b\bar{b}$ contribution. All results in this thesis are produced after applying this scale factor to the $t\bar{t}+b\bar{b}$ contribution.

6.3 Event Reconstruction

Based on the kinematic properties of the data or simulated events, an event is classified as coming from a certain process via ANNs to generate templates discriminating the distinctive processes. These discriminators are designed to contain regions (i.e. bins of histograms) enriched in one certain process, such that its contribution can be constrained during the fit. To ensure a high level of classification performance, the choice of kinematic features is crucial. One group of features considered for this analysis are kinematic properties of jets and leptons assigned to hypothesized final-state partons of the observed event. Which particles are expected in the final state is based in the event topologies discussed in section 3.2. The assumptions for this assignment method are twofold:

Firstly, it is assumed that all jets and leptons expected by the event topologies are contained in the acceptance and reconstructed in the event. A leading-order $t\bar{t}+Z$, $Z \rightarrow b\bar{b}$ process as shown in Figure 3.1 is supposed to contain two jets originating from the decays of both top quarks, $t \rightarrow Wb$, two jets originating from the hadronically decaying W boson, $W \rightarrow q\bar{q}'$, one lepton originating from the leptonic decay of the other W boson, $W \rightarrow \ell\nu$, and two jets originating from the decay of the Z boson, $Z \rightarrow b\bar{b}$. This implies that for a full assignment of final-state objects of this process, six jets and one lepton are needed, which is not the case for a large fraction of events in the phase space chosen for this analysis (c.f. section 3.3). Secondly, the assignment of jets and leptons to final-state objects happens under the hypothesis of one certain process, so it is assumed that for every event a suited

assignment of jets and leptons can be found, e.g. a reconstruction of the $t\bar{t}+Z$ process is conducted both for $t\bar{t}+Z$ events and background events, like $t\bar{t}+\text{jets}$.

The method of assigning reconstructed jets and leptons to all partons and leptons expected in a $t\bar{t}+Z$ process will henceforth be referred to as “ $t\bar{t}+Z$ reconstruction”, while the same procedure for a $t\bar{t}$ system will be referred to as “ $t\bar{t}$ reconstruction”. The method used for these reconstructions is based on a χ^2 -measure computed from the masses of the distinctive particles expected in these processes.

For events with fewer than six reconstructed jets, a $t\bar{t}$ reconstruction is performed. For this purpose, every possible assignment of reconstructed jets to the four quarks expected in a $t\bar{t}$ event is considered. Assignments where the number of mistagged jets exceeds a certain threshold are not considered further. Here, a mistagged jet is a reconstructed jet that is assigned to a bottom quark, but is not b-tagged (c.f. section 3.1.3) and vice versa. The number of allowed mistags, shown in Table 6.1, depends on the number of jets and number of b-tagged jets reconstructed in the events to guarantee at least one possible assignment of jets. For each viable combination, the four-vectors of both top quarks and the hadronically decaying W boson are calculated from the jets and the lepton assigned to the corresponding decay products. For the reconstruction of the leptonically decaying top quark, two solutions for the z -component of the neutrino four-vector are considered as this component cannot be unambiguously determined from detector measurements (c.f. section 3.1.4). For events with at least six jets, also the Z boson is reconstructed by combining the four-vectors of two b-tagged jets.

From the reconstructed four-vectors, hypothetical masses $m_i^{\text{reco.}}$ of the reconstructed particles are calculated. The quality of the hypothesized assignment is then assessed with a χ^2 -measure, based on the squared difference of the reconstructed mass and the expected mass, following

$$\chi_i^2 = \frac{(m_i^{\text{reco.}} - m_i^{\text{exp.}})^2}{(\sigma_i^{\text{exp.}})^2}. \quad (6.4)$$

Further information on the derivation of the expected masses $m_i^{\text{exp.}}$ and $\sigma_i^{\text{exp.}}$ can be found in Appendix A.1. The assignment with the smallest total χ^2 -value

$$\chi_{\text{total}}^2 = \chi_{\text{t}_{\text{had}}}^2 + \chi_{\text{t}_{\text{lep}}}^2 + \chi_{\text{W}_{\text{had}}}^2 \left(+ \chi_Z^2 \right), \quad (6.5)$$

is saved and treated as the best possible $t\bar{t}$ (or $t\bar{t}+Z$) reconstruction for this event.

From this reconstruction, secondary kinematic features of the event can be derived, for example angular differences or transverse momenta, which can subsequently be used as input features of the ANNs. An evaluation of reconstruction efficiencies can be found in Table A.2.

Table 6.1: **Allowed mistags for $t\bar{t}$ and $t\bar{t}+Z$ event reconstruction.** The number of allowed mistags (i.e. b-jets misidentified as light-flavor jets and vice versa) is chosen such that for every event at least one possible assignment of jets and leptons to the final-state partons can be found.

number of reconstructed jets	$t\bar{t}$		$t\bar{t}+Z$
	4	5	≥ 6
3 b-tagged jets category	1	0	1
4 b-tagged jets category	2	1	0

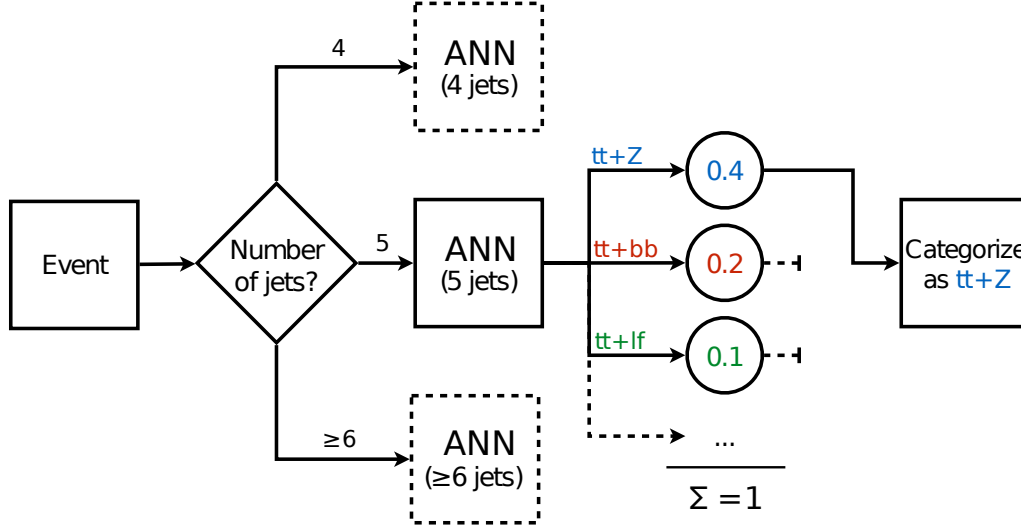


Figure 6.2: **Event classification based on ANNs.** Three separate ANNs are used for events with four, five and equal or greater than six jets. The output classes of the ANNs represent the physics processes to be distinguished. An event is categorized according to the output class with the highest value. For one event, all output values are normalized to one to allow a probability-based interpretation. [Sketch adapted from M. Rieger.]

6.4 Event Classification

Based on the explanations in chapter 4, ANNs are used to classify events. To become sensitive to the signal process, bins of ANN output distributions are needed with a large signal-to-background ratio. Furthermore, regions where one certain background process is enriched helps constraining the contributions of this process in the regions sensitive to the signal contribution. These regions also help constraining the nuisance parameters, i.e. systematic uncertainties in the fit.

Depending on the number of jets and number of b-tagged jets, separate ANNs are used. This allows for the definition of signal- and background-enriched regions. As explained in section 3.3, a selection of events with a high number of jets or b-tagged jets is signal-enriched while regions with a low number of jets or b-tagged jets are background enriched.

Templates to be used in the ML fits are built from the outputs of the ANNs. An event is evaluated with the ANN specific to the number of jets and b-tagged jets. Multi-classification ANNs are used, where the classes correspond to the signal and background processes. Dedicated classes for the processes $t\bar{t}+Z$, $t\bar{t}+H$, $t\bar{t}+b\bar{b}$, $t\bar{t}+l\nu$ and $t\bar{t}+c\bar{c}$ are considered. An event is classified as the process with the highest output value of the ANN. The classification of an event as a certain process does not imply that it will be treated as such in the fit, it merely determines to which category the event will be added. As the templates are constructed as binned distributions, the event is added to the bin corresponding to the output value of the ANN. The categorization process is illustrated in Figure 6.2.

The training of the classifiers is performed with simulated events of the five processes that are classified. Only the nominal event content is used for the training process, i.e. systematic variations like jet energy scale variations are not considered explicitly in the training. To guarantee an unbiased evaluation of the templates, only a fraction of simulated events is used for the training process. The remaining events are used in the likelihood fit. This measure is introduced, as the ANN might be able to fit its predictions to the events used

for training (over-fitting), thereby biasing the predictions for these specific events, which will not be comparable to the predictions for data events. Although the generalization of the ANNs is checked via various methods this measure is retained nonetheless.

6.5 Optimization

Different analysis strategies were explored. This includes the definition of signal and control regions, the process classes, input features and parameters of ANNs training, binning of the discriminators as well as the employment of pre-classification or binary-classification strategies. At first, the selected strategy which is used to derive the results in chapter 8, is presented. The alternative options explored in this thesis are reviewed thereafter.

6.5.1 Final Strategy

The ANNs used to produce the analysis results consist of five output classes, $t\bar{t}+Z$, $t\bar{t}+H$, $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$ and $t\bar{t}+lf$. Separate ANNs are trained and evaluated for events with four, five and more or equal than six reconstructed jets. The parameters chosen for the training process are summarized in Table 6.2. Approximately twenty input features are used per ANN. The final selection of input features is summarized in Table 6.3 and their distributions are shown in Appendix B.1. Exemplary, the distributions of two features used in the four-jet region are shown in Figure 6.3.

The resulting discriminator distributions are shown in Appendix D. For the discriminators corresponding to the outputs of the $t\bar{t}+lf$ and $t\bar{t}+c\bar{c}$ nodes only one bin is used. The discriminators corresponding to the $t\bar{t}+b\bar{b}$ and $t\bar{t}+H$ nodes consist of five bins, while the discriminators corresponding to the $t\bar{t}+Z$ nodes consist of nine bins. Exemplary, the discriminator corresponding to the $t\bar{t}+Z$ node in the ≥ 6 jet region is shown in Figure 6.4. The following sections motivate these choices.

6.5.2 Classification

It was decided to split the analysis region into three regions according to the number of jets reconstructed in an event. Studies were conducted to quantify, based on ANN performance evaluations and the calculation of expected limits, whether splitting the

Table 6.2: **Configuration of the ANNs.**

parameter	setting
number of neuron layers	2
number of neurons per layer	50
activation functions	LeakyReLU, $f(x) = 0.1x \cdot \theta(-x) + x \cdot \theta(x)$
activation function of output layer	Softmax (eq. 4.3)
loss function	categorical cross-entropy (eq. 4.4)
optimizer	ADAGRAD [91] (default settings)
number of epochs	at the maximum 1000 iterations over the train sample
batch size	200 events per weight update
DROPOUT percentage [99]	30% of neurons randomly deactivated during training
L2-REGULARIZATION [98]	add $10^{-5} \cdot w_{ji}^2$ term to the loss value
early stopping	no decrease in loss value for 10 epochs
	5% difference between validation and train loss value

Table 6.3: **Final selection of input features for the ANNs in the four-jet, five-jet and ≥ 6 jet regions.** In Appendix B.1 the features are described.

	4 jets	5 jets	≥ 6 jets
third highest b-tag value	✓	✓	—
fourth highest b-tag value	✓	✓	✓
average b-tag value	—	—	✓
average b-tag value of tagged jets	✓	✓	—
average deviation of b-tag value	✓	✓	✓
minimal b-tag value	—	✓	—
$\Delta\eta_{\text{avg}}^{\text{j,j}}$	✓	—	✓
$\Delta\eta_{\text{avg}}^{\text{b,b}}$	✓	✓	✓
$\Delta R_{\text{avg}}^{\text{j,j}}$	—	✓	—
$\Delta R_{\text{avg}}^{\text{b,b}}$	—	✓	—
$\Delta R_{\text{min}}^{\text{j,j}}$	✓	✓	—
$\Delta R_{\text{min}}^{\text{b,b}}$	✓	✓	—
$m_{\text{avg}}^{\text{b,b}}$	—	✓	✓
$m^{\text{b,b}}$ closest to 125 GeV	✓	✓	✓
$m^{\text{b,b}}$ closest to 91 GeV	✓	✓	✓
$m(\Delta R_{\text{min}}^{\text{j,j}})$	—	✓	—
$m(\Delta R_{\text{min}}^{\text{b,b}})$	—	—	✓
missing H_T	✓	—	—
$m_{\text{avg}}^{\text{b}}$	—	✓	✓
$m_{\text{tot}}^{\text{b,b}}$	—	✓	—
$p_{T,\text{avg}}^{\text{b,b}}$	✓	—	—
$p_T(\Delta R_{\text{min}}^{\text{j,j}})$	—	✓	—
$p_T(\Delta R_{\text{min}}^{\text{b,b}})$	—	—	✓
b-tag likelihood ratio	—	✓	✓
transformed b-tag likelihood ratio	✓	✓	✓
second Fox-Wolfram moment	—	✓	—
p_T of leading jet	—	—	✓
number of b-tags (medium)	—	—	✓
$t\bar{t} + Z$ reconstruction $\ln(\chi^2)$	✓	✓	—
$t\bar{t} + Z$ reconstruction $\ln(\chi^2)(W_{\text{had}})$	✓	✓	—
$t\bar{t} + Z$ reconstruction $\ln(\chi^2)(Z)$	—	—	✓
$t\bar{t} + Z$ reconstruction $m(t_{\text{had}})$	✓	—	—
$t\bar{t} + Z$ reconstruction $m(W_{\text{had}})$	✓	—	—
$t\bar{t} + Z$ reconstruction $\ln(m)(Z)$	—	—	✓
$t\bar{t} + Z$ reconstruction $p_T(Z)$	—	—	✓

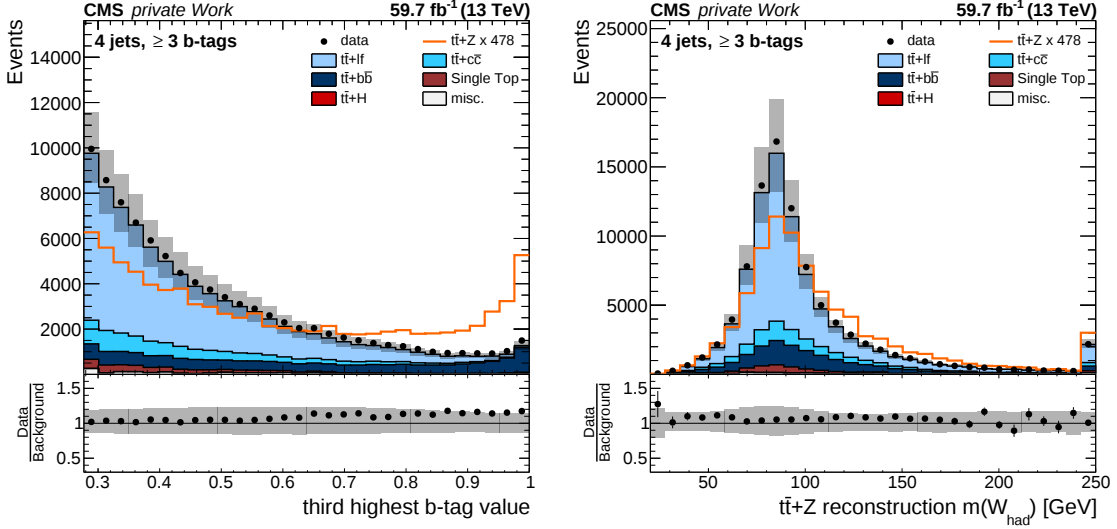


Figure 6.3: **Data/MC comparisons of two input features used in the four-jet region.** Third highest b-tag value on the left and the reconstructed hadronic top mass on the right before the fit to data. The predicted contributions of all background samples are stacked, the $t\bar{t}+Z$ contribution is overlaid as a line scaled by a factor of 478 to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape-changing effect.

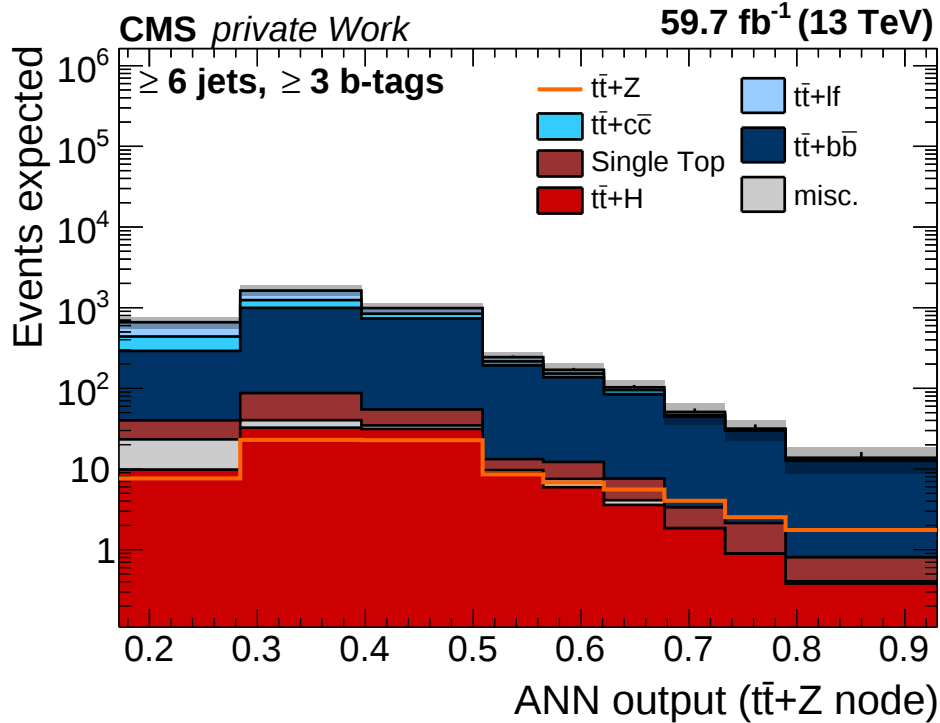


Figure 6.4: **Discriminator distribution in the ≥ 6 jet region corresponding to the $t\bar{t}+Z$ node output before the fit to data.** The predicted contributions of all background samples are stacked, the $t\bar{t}+Z$ contribution is overlaid as a line. The error bands correspond to the systematic uncertainties of the background contributions with a shape-changing effect.

analysis region according to the number of b-tagged jets (three and four or more b-tags) could be beneficial. The performance of this tag-region splitting was found to be slightly inferior to the performance of the analysis conducted with separate ANNs for jet regions. A possible reason is the use of input features from a dedicated $t\bar{t}+Z$ reconstruction, which is only conducted for events with ≥ 6 jets, and can therefore not be used in a region based on tag splitting. Other than that, the number of bins is reduced in the tag splitting approach as only two separate ANNs (i.e. tag regions) are used. This is supported by the observation that merging the four-jet and five-jet regions into one region with fewer or equal five jets also leads to a decrease in sensitivity. However, the region corresponding to ≥ 4 b-tagged jets exhibits the highest sensitivities on the $t\bar{t}+Z$ cross-section on its own, in comparison to other regions, hence it should not be discarded for future analyses from the get-go. In combination with the region corresponding to three b-tagged jets, however, this performance is surpassed by the approach using categories of jet multiplicity.

The choice of parameters for the training of classifiers is based on the evaluation of their classification performance via confusion matrices and ROC integral values. Confusion matrices evaluate the performance of an ANN based on a data set which is not used for the training process (test set). Events are categorized according to their true and predicted class, thereby quantifying the rates with which events of a certain class are classified correctly or as a wrong class (confusion). An exemplary confusion matrix for the final ANN used in the ≥ 6 jets region is shown in Figure 6.5. Events belonging to the $t\bar{t}+Z$ process are classified as such with an accuracy of 33% which leads to an enrichment of these events in the discriminator corresponding to the $t\bar{t}+Z$ node. A large fraction of approximately 26% of $t\bar{t}+Z$ events are mistaken as $t\bar{t}+H$ events due to the very similar kinematics of these processes (see section 3.2). Similarly, $t\bar{t}+lf$ and $t\bar{t}+c\bar{c}$ events are confused with a high probability. The confusion matrices for the four-jet, five-jet and ≥ 6 jet region ANNs are shown in Appendix C.1.

The ROC integral corresponds to the integral of true-positive rates versus false-positive rates for various thresholds in the discriminator, averaged over all discriminators of the multi classification. A value of 1 corresponds to flawless classification, while 0.5 corresponds to random classification.

Various combination of hyperparameters for the ANNs were tested, most of which did not change the classification power considerably. For example, a larger number of layers and neurons in the ANNs lead to a faster loss of generalization, as more weights in the ANN can be adjusted to fit the data set used for training (train set) perfectly. This effect can be countered by introducing stronger regularization policies. Hence, a low number of layers and neurons was favored as not to enforce too strong regularizations on the training process. The number of epochs (i.e. iterations over the train set) and batch size (i.e. number of events evaluated before another weight update) were adjusted to guarantee enough iterations and weight updates to allow the ANN training to converge. Early stopping methods were introduced to terminate the training process, if either the value of the loss function did not decrease for a number of iterations, or the difference between loss values of the train and validation sets exceeded some threshold. Different optimizers were tested. These exhibited various speeds of convergence, but ultimately converged to similar accuracies.

To reduce the number of $t\bar{t}+H$ events in the analysis regions, an approach was tested in which events containing a pair of b-tagged jets with an invariant mass close to the Higgs boson mass of 125 GeV were vetoed. This was aimed at being less sensitive to the $t\bar{t}+H$ cross-section, as explained in section 7.4. This strategy reduced the overall number of simulated events by approximately half, thus leading to a notable decrease in sensitivity on the $t\bar{t}+Z$ cross-section. A small decrease in sensitivity to the $t\bar{t}+H$ cross-section was also

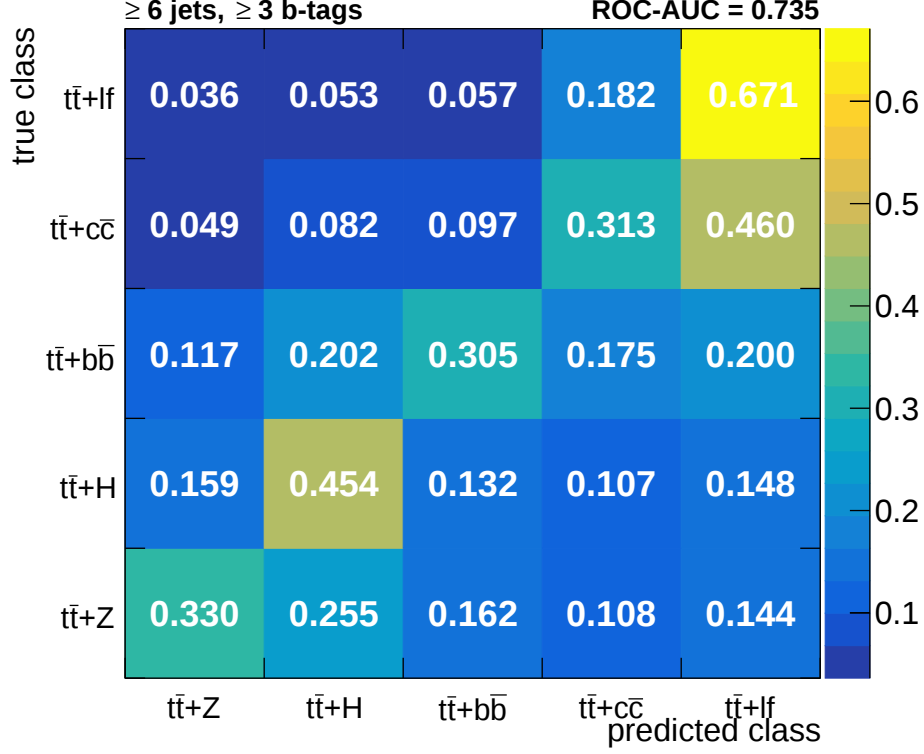


Figure 6.5: **Confusion matrix for the ANN in the ≥ 6 jet region.** Number of events belonging to one class (true class) while being predicted as another class (predicted class), normalized to unity for each true class (row). Diagonal elements express the rates with which a class is classified correctly. Off-diagonal elements express the confusion rates of events of a certain class with another class.

observed, but the strategy was ultimately discarded due to the loss in sensitivity on $\bar{t}\bar{t}+Z$ production.

6.5.3 Input Features

Input features were selected based on a heuristic measure designed to quantify the importance of one single feature in comparison to the other features used in the classification. Generally, the distributions of all features used in the ANNs are normalized to a mean value of zero and a standard-deviation of one to reduce the impact of different physical scales of features in the training process. By normalizing all features the same way, weights connecting one input feature to a node in the first layer can be compared among input features. Larger weight values hint to a larger impact of this feature on the final classification result. Thus, the measure

$$\aleph(x_i) = \frac{\sum_j^{\text{nodes in first layer}} |w_{ji}|}{\sum_k^{\text{input features}} \left(\sum_j^{\text{nodes in first layer}} |w_{jk}| \right)}, \quad (6.6)$$

for an input features x_i is used, where j corresponds to the nodes in the first layer and k to the input features.

This importance ranking has been evaluated by training each ANN several times with a selection of around 200 input features, covering kinematics of jets and leptons, angular differences between final-state objects, b-tagging related features, full-event observables like

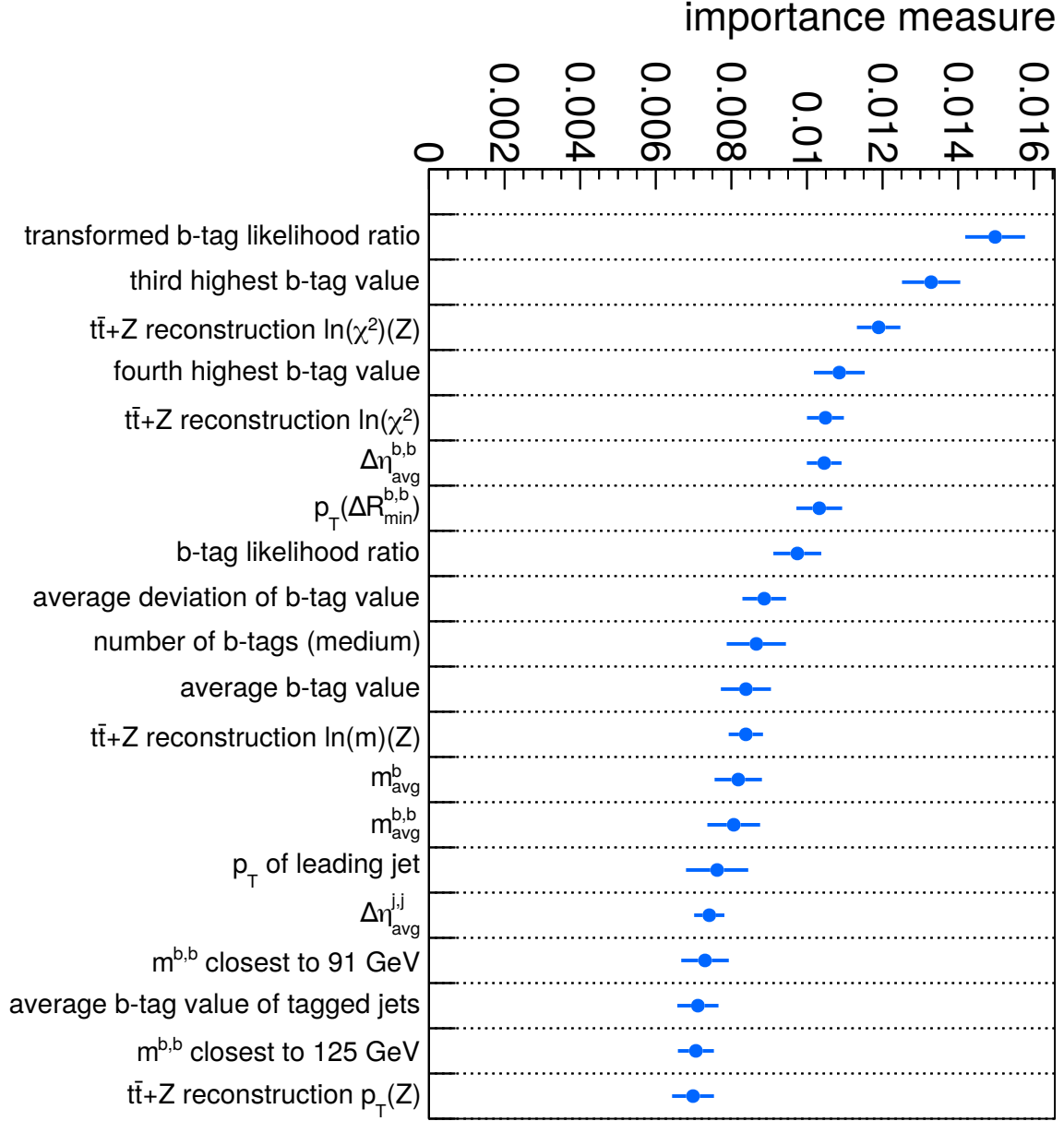


Figure 6.6: **Ranking of input features for the multi-classification ANN in the ≥ 6 jet region.** The ANNs were trained 20 times, the mean values of $\aleph$ are shown with error bars corresponding to its standard deviation. Only the 20 highest ranked features are shown.

measures of the sphericity or aplanarity of events, and features derived from the $t\bar{t}+Z$ and $t\bar{t}$ event reconstructions. The training iterations are initialized with different seeds for the initial weights, and the order of events is shuffled randomly. The same was performed for binary ANNs, trained to classify $t\bar{t}+Z$ versus $t\bar{t}+H$ events only. The superset of the twenty features with highest $\aleph$ -measures from multi-classification and the ten features with highest $\aleph$ -measures from binary classification are selected per ANN. The ranking of features for the multi-classification and binary ANNs are shown in Appendix B.2. Exemplary, the ranking of input features used in the multi-classification ANN in the ≥ 6 jet region is shown in Figure 6.6.

For each of the selected features, the agreement between simulation and data was checked via Goodness-Of-Fit tests, as further explained in section 7.1. Based on these validation

results some features were discarded from the final selection.

Additionally, the correlations of the selected input features were checked. For two features showing a correlation larger than 90%, the feature with the smaller $\aleph$ -value was discarded from the final selection. This measure aims at reducing the number of adjustable parameters (i.e. weights) in the ANNs. The correlations of all input features for the ANNs for four, five and ≥ 6 jets are visualized in Appendix B.4. The features marked in red are discarded from the final feature selection. Exemplary, the correlations of input features used in the ANN in the ≥ 6 jet region are shown in figure 6.7. The loss in classification performance of the ANNs before and after removing these features based on their correlations was found to be negligible as only few features were discarded.

Furthermore, the impact of input features on the prediction output of the ANNs was investigated by fixing all but the input feature under scrutiny at their mean values and varying the feature by two standard-deviations. This is exemplary shown in Figure 6.8 for one feature with a high $\aleph$ -ranking (left) and one with a low ranking (right). The fourth highest b-tag value, for example, leads to a strong prediction of $t\bar{t}+lf$ (i.e high ANN output value in the $t\bar{t}+lf$ node) for very small values, whereas $t\bar{t}+H$ and $t\bar{t}+Z$ are favored predictions for large values. This observation is also supported by the topologies of these processes, as $t\bar{t}+lf$ events do not contain more than two bottom quarks, thus it is less likely that the fourth highest b-tag value is large, compared to $t\bar{t}+Z$ or $t\bar{t}+H$ events, where four b-tagged jets are expected. A study of these variations also gave a possible explanation of the large unidirectional misclassifications observed in the ANN training process, e.g. visible in the confusion matrix in Figure 6.5. A large fraction of $t\bar{t}+Z$ events is misclassified as $t\bar{t}+H$, however, the vice versa confusion is significantly smaller. The same is observed for the confusion of $t\bar{t}+c\bar{c}$ and $t\bar{t}+lf$ events. Most features used in the classification show similar kinematics and thereby similar variation behavior for $t\bar{t}+Z$ and $t\bar{t}+H$ events. The kinematic features, however, are oftentimes more extreme for $t\bar{t}+H$ events in comparison to $t\bar{t}+Z$ events, e.g. a larger fraction of $t\bar{t}+H$ events exhibit third highest b-tag values close to one in comparison with $t\bar{t}+Z$ events. Thus, the variation of the output node values usually is stronger for $t\bar{t}+H$ events than for $t\bar{t}+Z$ events, leading to a large fraction of $t\bar{t}+Z$ events being misclassified as $t\bar{t}+H$ and not vice versa. This is an inherent problem in the classification of $t\bar{t}+H$ and $t\bar{t}+Z$ events and needs to be studied further to increase the classification power between these two processes. The same applies for $t\bar{t}+lf$ and $t\bar{t}+c\bar{c}$ events.

The impact of using $t\bar{t}+Z$ and $t\bar{t}$ reconstruction-based variables was assessed by comparing the results of ANNs that were trained with the best thirty features overall in comparison with the best thirty features not derived from a dedicated $t\bar{t}+Z$ and $t\bar{t}$ reconstruction. Exemplary, two confusion matrices corresponding to the ANNs for ≥ 6 jets are shown in Figure 6.9. In summary, the classification accuracies increased when also considering reconstruction-based features, especially the confusion between $t\bar{t}+Z$ and $t\bar{t}+H$ was reduced. The accuracy of correctly classifying $t\bar{t}+Z$ increased by approximately 5%, while the confusion of $t\bar{t}+Z$ events as $t\bar{t}+H$ events decreased by approximately 4%.

6.5.4 Discriminator Setup

Different classification approaches were tested. ANNs with four classes, $t\bar{t}+Z$, $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$ and $t\bar{t}+lf$ were compared to ANNs with five classes, including in addition $t\bar{t}+H$, or separated classes for $t\bar{t}+b\bar{b}$ and $t\bar{t}+2b$. The sensitivities to the $t\bar{t}+Z$ cross-section was found to be similar for all three approaches. The addition of a $t\bar{t}+2b$ class allows the 50% rate uncertainties applied to $t\bar{t}+hf$ processes to be constrained better than in other ANN setups due to the larger differences in the templates of the affected processes. Additional discriminators dedicated to $t\bar{t}+H$ were found to be helpful in separating $t\bar{t}+H$ from $t\bar{t}+Z$

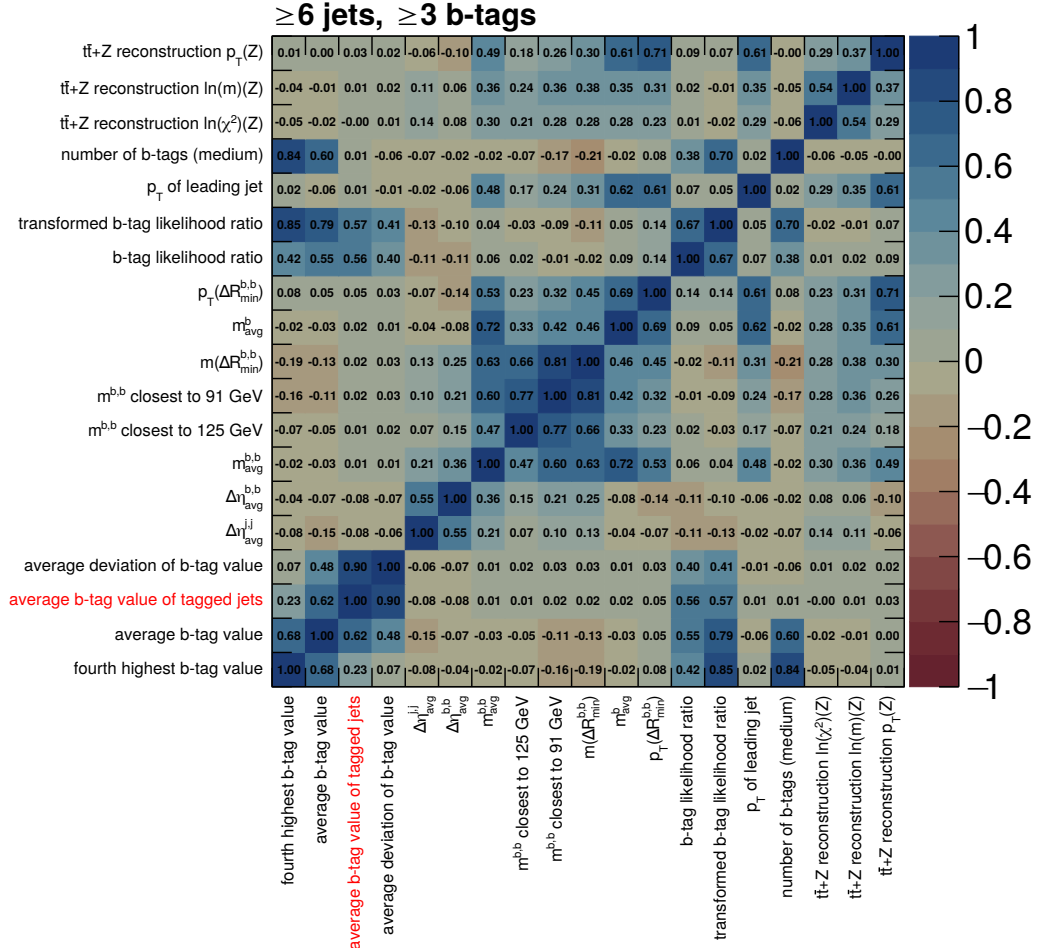


Figure 6.7: **Correlation of input features used in the ≥ 6 jet region.** Features showing correlations of larger than 90% are discarded from the final feature selection and are marked in red.

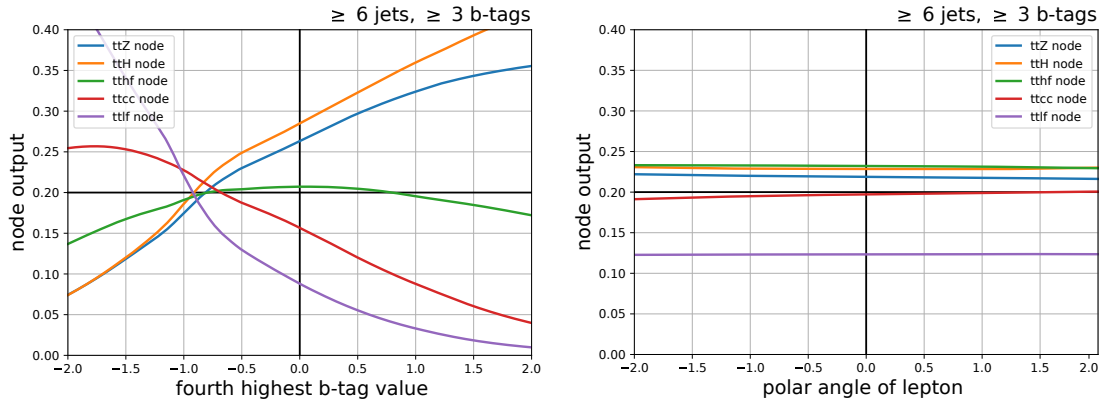


Figure 6.8: **Isolated impact of one input feature on the output values of an ANN.** All features are fixed to their mean values after the training of the ANN but one feature is varied by two standard-deviations. The impact of this variation is quantified by changes in the output predictions of the ANN in each classification node. The variation of the fourth highest b-tag value (left) and polar angle ϕ of the lepton (right) are shown.

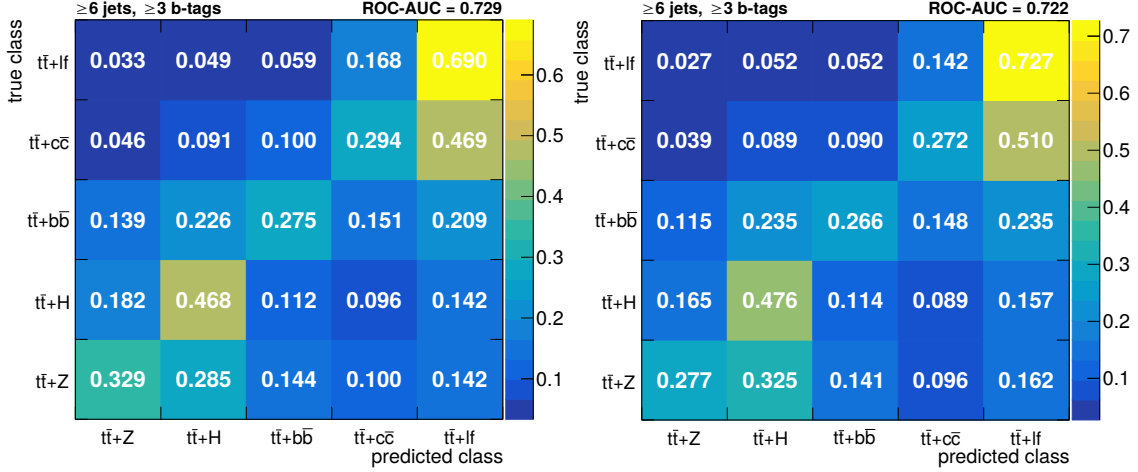


Figure 6.9: **Confusion Matrices for ANNs trained with (left) and without (right) reconstruction-based features.** For each ANN thirty input features are used, of which seven input features use information from $t\bar{t}+Z$ or $t\bar{t}$ reconstruction for the ANN on the left.

explicitly, which was assessed to be of capital importance, hence this approach was decided for (see also discussion in section 7.4).

Furthermore it was tested whether a binary classification of $t\bar{t}+Z$ events versus all backgrounds leads to similar performances compared to the multi-classification approach. It was concluded that the performance decreases considerably, as fewer signal enriched bins, or bins enriched in one specific background process, are present in the likelihood fit.

The binning of discriminator distributions was chosen based on the purpose of the discriminators. The $t\bar{t}+lf$ and $t\bar{t}+c\bar{c}$ discriminators were chosen to be constructed with only one bin each. These discriminators are mainly used to determine the contributions of these processes in the ML fit, which does not rely on shape information from various bins but mostly on the purity of this process. For example, the $t\bar{t}+lf$ node in the four jet region is predicted to consist of approximately 82% $t\bar{t}+lf$ events from simulation, while containing less than 0.1% $t\bar{t}+Z$ events. The discriminators for $t\bar{t}+H$ and $t\bar{t}+b\bar{b}$ were chosen to consist of five bins each. Several uncertainties, for example the ones related to b-tagging, can be constrained from shape information in $t\bar{t}+b\bar{b}$ distributions. Additionally, this process represents the largest background contribution in the discriminators most sensitive to the $t\bar{t}+Z$ process, e.g. the $t\bar{t}+b\bar{b}$ contribution in the $t\bar{t}+Z$ discriminator for events with ≥ 6 jet region is predicted to be approximately 58% from simulation. This contribution of $t\bar{t}+b\bar{b}$ in these signal-sensitive regions can be further constrained by a finer binning, thereby leading to a better sensitivity to the signal process. The binning in the $t\bar{t}+H$ nodes was introduced to generate differences in the shapes of templates for $t\bar{t}+Z$ and $t\bar{t}+H$, which is needed due to the similarities of these processes.

Finally, the discriminators corresponding to events classified as $t\bar{t}+Z$ consist of nine bins each. These discriminators show the largest signal-to-background ratios, especially in bins at higher ANN output values. Hence, this finer binning was chosen to retain sensitivity to the signal process.

Studies were performed with an even coarser binning, which leads to a loss of sensitivity on the $t\bar{t}+Z$ cross-section. The final binning that was chosen leads to some constraints of jet energy scale uncertainties (see section 7.3), which can be relaxed with a coarser binning. The effect, however, was judged to be small, thus, the finer binning was preferred.

7 Validation of Analysis Strategy

The analysis strategy presented in the previous chapter is validated to ensure that the methods applied to simulated data behave the same way for real data. Furthermore, it has to be ensured that the analysis yields unbiased results and the fit model is stable. Only then, the analysis strategy can be applied to real data to obtain the observed limits presented in chapter 8. For this purpose, the description of input features used in the ANN training is validated by comparing simulation to data in section 7.1. Features not described well are discarded. The training procedure of the ANNs is validated in section 7.2. The performance and behavior of the ML fit is evaluated in section 7.3, where the model is validated using pseudo data. Finally, the impact of the $t\bar{t}+H$ cross-section on the determination of the $t\bar{t}+Z$ signal-strength is discussed in section 7.4.

7.1 Validation of Input Features

All input features for the ANNs (see section 6.5.3) are validated via signal-plus-background Goodness-of-Fit tests where the signal-strength $\mu(t\bar{t}+Z)$ is fixed to its Standard Model prediction $\mu = 1$. These Goodness-of-Fit (GoF) tests aim at evaluating how well the simulation describes the observed data. At a technical level, the GoF test is performed using the so-called saturated model implemented in the `combine` framework [115]. The procedure is applied for each input feature independently. It is summarized in the following:

The templates from simulation are fit to the observed data distribution using the full uncertainty model introduced in chapter 5. The fit yields post-fit distributions of the simulated templates. From these post-fit distributions multiple pseudo experiments are generated by varying each bin content of the templates following a Poissonian p.d.f.. This type of variation aims at modeling the statistical fluctuations expected in such a counting experiment. As a consequence, each pseudo experiment consists of a new prediction of the number of data events in each bin (pseudo data). The original pre-fit templates from simulation are subsequently fit to each of the newly derived pseudo data experiments. The goodness of each fit is quantified using the negative log-likelihood ratio (c.f. eq. 4.8) as test statistics. From the distribution of these test statistic values the compatibility of the actual data distribution with the simulated data can be assessed by calculating a p -value (c.f. eq. 4.12). This corresponds to the fraction of test statistic values that are larger than the test statistic value observed in data. To assure a good description in simulation and data, input features are discarded if their p -value is lower than 0.05.

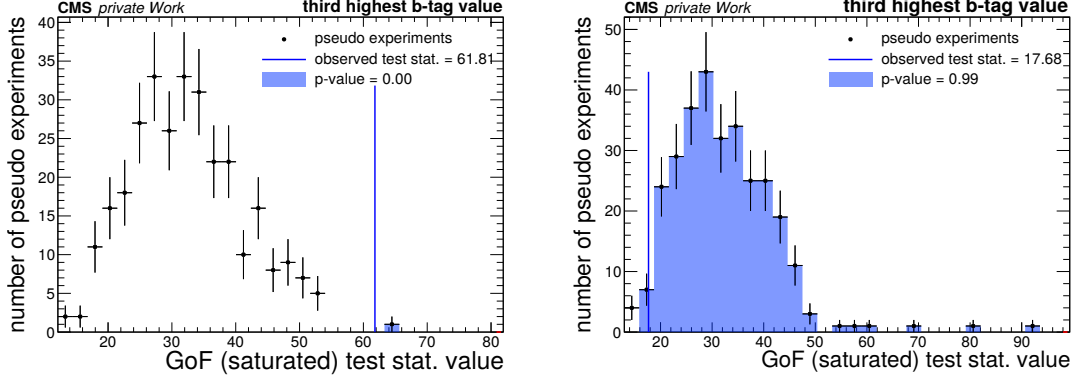


Figure 7.1: **Distribution of Goodness-of-Fit test statistic values for the third highest b-tag value in the ≥ 6 jet region (left) and four-jet region (right).** The test statistic values of the pseudo experiments are histogrammed in black. The observed test statistic value is shown as a blue line, the area corresponding to its p -value is filled in blue.

In Figure 7.1, two exemplary distributions of the test statistic values of pseudo experiments and the observed test statistic value for data are shown for the third highest b-tag value in the ≥ 6 jet region on the left, and the four-jet region on the right. The Goodness-of-Fit p -values for these features were found to be 0.00 and 0.99, for the ≥ 6 jet and four-jet region, respectively. As a consequence, the feature was discarded from the final input feature selection in the ≥ 6 jet region, as its p -value was lower than 0.05, while the simulation in the four-jet region showed a sufficiently accurate description with respect to data, hence it was not discarded.

All p -values obtained for features in the ≥ 6 jet region are shown in Figure 7.2 and for all jet regions in Appendix B.3. The final selection of input features for each jet region is summarized in Table 6.3. Short descriptions of these input features and their distributions in simulation and data are summarized in Appendix B.1. In Appendix B.5 some additional control distributions are shown in simulation and data, for example, η , ϕ and p_T distributions of electrons, muons and jets.

7.2 Validation of ANN Training

The primary aspect that needs to be validated for ANNs is its generalization, i.e. how well it generalizes what it learned on one data set to a data set not used during the training process. This simulates the behavior (i.e. classification performance) on real data. For this purpose, the data set used for the ANN development is split into three data sets: one is used for the training itself (train set), another is used to evaluate the generalization of the ANN during the training process (validation set) and the last one is used to evaluate the ANN performance after the training process is completed (test set). With the validation set, the loss function (c.f. eq. 4.4) and the accuracy of the classifier are monitored during the training. After each epoch (i.e. each iteration over the train set), the loss function is evaluated for the validation set and for the train set. Similarly, the fraction of events classified correctly is evaluated after each epoch for the validation and train sets. The histories of both metrics are shown in Figure 7.3 for the ANN used in the ≥ 6 jet region and in Appendix C.2 for all ANNs.

The constant offset in loss function values and classification accuracies between train and validation sets can be explained by the usage of DROPOUT. This disables a certain

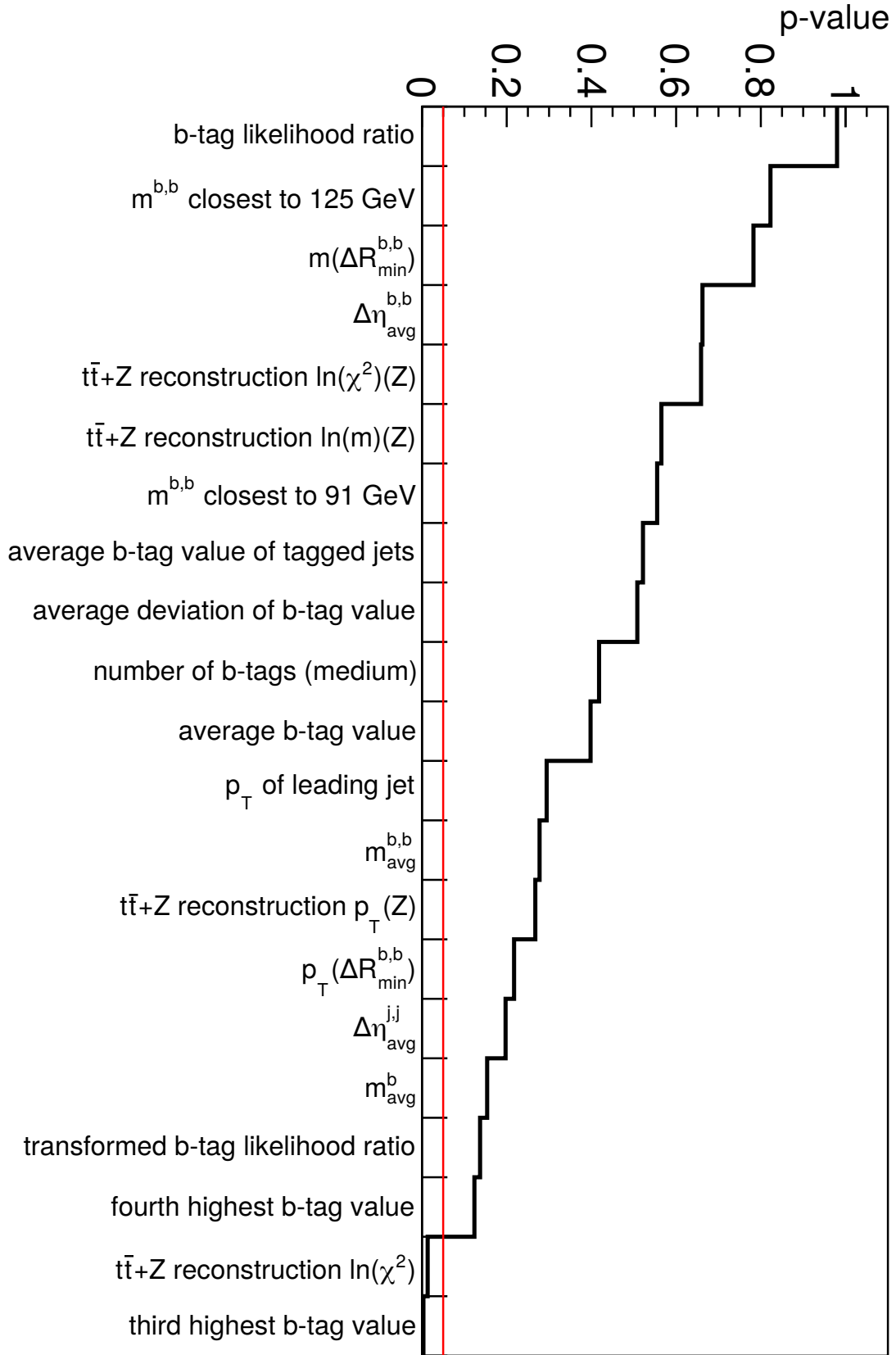


Figure 7.2: **Goodness-of-Fit p -values for all input features in the ≥ 6 jet region.** The red line indicates a p -value of 0.05. All features with a p -value lower than 0.05 are discarded from the final feature selection.

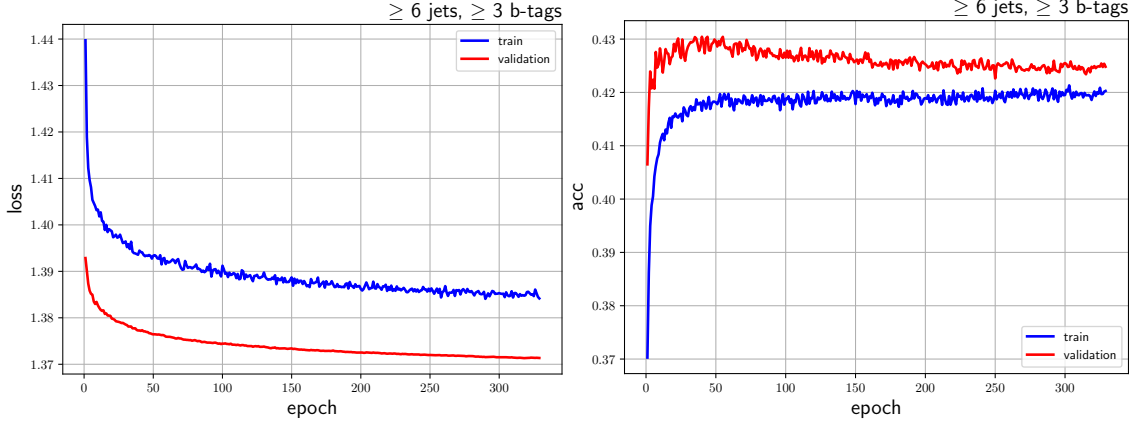


Figure 7.3: **History of loss function values (left) and classification accuracies (right) during the training process of the ANN used in the ≥ 6 jet region.** The metrics are evaluated after every pass over the train set (epoch) for the train set (blue) and validation set (red). Divergences between train and validation scores imply a loss of generalization.

number of neurons in the ANN during the training process. The metrics are evaluated for the train set with deactivated neurons, while they are activated for the validation set. The ANN without deactivated neurons is expected to provide a broader and more generalized classification and thus perform slightly better than with deactivated neurons. As a consequence, the metrics on the validation set are expected to be consistently better than the metrics on the train set, which is supported by the observations. This has also been confirmed by training an ANN without DROPOUT, in which case the validation metrics are consistently inferior to the train metrics, as expected.

The loss function values both for the train and validation set decrease steadily. Diverging loss function values would be a hint of over-fitting the ANN's weights to the train set, thereby losing the generalization to unseen data. This is not observed here. Early stopping methods are implemented to counter the loss of generalization. These terminate the training process if either the loss function value did not decrease for a given number of epochs, or the difference of loss function values between train and validation set exceeded some threshold (see Table 6.2).

The monitoring of the classification accuracies in most cases shows similar behavior. It is expected to increase during the training procedure. However, in Figure 7.3 the validation accuracy also decreases for a range of epochs. As this is only a secondary feature and not an absolute measure of the performance of the ANN, this training is judged as successful, nonetheless.

Further validation of the ANN training is performed with compatibility tests between the train and test data sets by evaluating events in both data sets with the trained classifier. The distributions of events for both data sets in the output nodes are checked for compatibility by calculating a Kolmogorov-Smirnov (KS) probability for all processes. The KS test is a statistical test of the compatibility of the shapes of two distributions [116, 117]. The KS probabilities are calculated such that they are uniformly distributed between zero and one for compatible histograms. Small values indicate a small probability of compatibility, while values close to one indicate a high probability of compatibility. The KS probabilities are evaluated for the distributions of all processes separately in each output node. An ANN is discarded if one KS probability is found to be lower than 0.05. All KS probabilities for the final setup are listed in Appendix C.3. As an example, in Figure 7.4 the distribution of $t\bar{t}+Z$ and $t\bar{t}+H$, and $t\bar{t}+\text{jets}$ background processes are shown in the discriminator corresponding

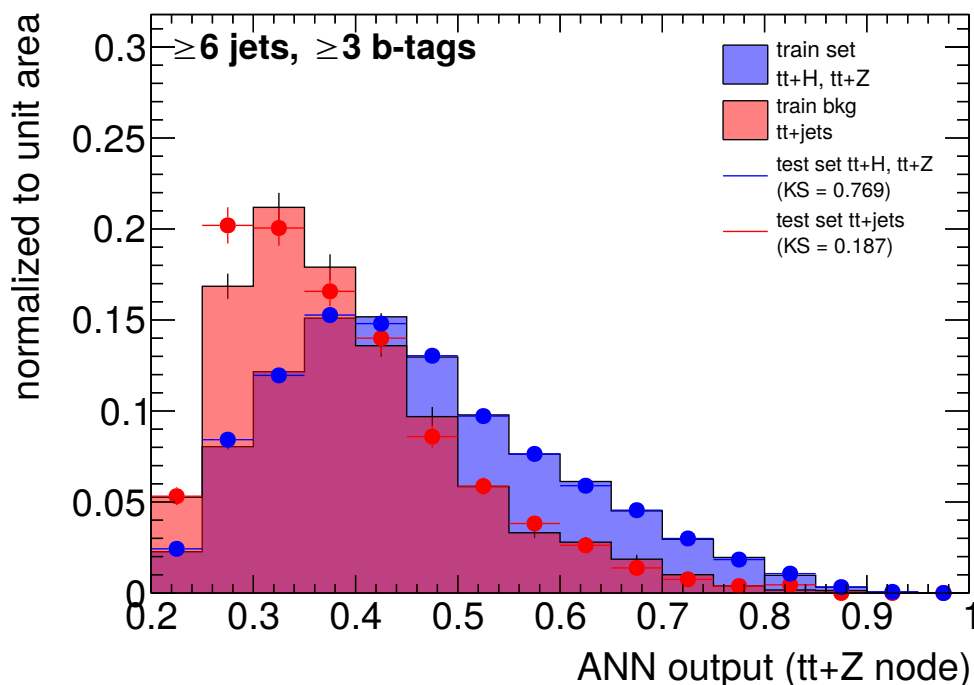


Figure 7.4: **Distribution of events in the test set (colored marks) and train set (filled histograms) for $t\bar{t}+H$ and $t\bar{t}+Z$ events (blue) and $t\bar{t}+jets$ events (red) in the $t\bar{t}+Z$ node of the ≥ 6 jet ANN.** All distributions are normalized to unit area. The KS probabilities for the compatibility of test and train distributions are shown in the legend.

to the $t\bar{t}+Z$ node of the ANN used for events with ≥ 6 jets, both for the train and the test data set.

In this example, the KS probability of the $t\bar{t}+jets$ background distributions is 0.19, which indicates sufficiently good compatibility of the distributions. Similarly, the distributions of a combination of $t\bar{t}+Z$ and $t\bar{t}+H$ events show a KS probability of 0.77, also indicating good compatibility. Hence, it can be expected that data events exhibit the same behavior in the ANN classification process as simulated events. As this criterion is fulfilled, these ANNs can be used in the evaluation of this analysis.

The described validation of the ANN performance, however, comes with a caveat. The compatibility of the ANN classification performance between a data set used for training and one not used for training is only validated in simulation. Thus, the compatibility of the descriptions of input features in simulation and data which are used for the ANNs needs to be validated separately, as previously described in section 7.1. Only in combination, these validation procedures can guarantee a comparable classification of data events and simulated events.

7.3 Fit Validation

In this section, the fit of the complete model using all systematic uncertainties is analyzed for pseudo data. In this case the fit of the signal and background templates is not performed to real data, but rather to a pseudo data set where every bin content corresponds to the nominal expectation from simulation (Asimov data set with $\mu = 1$). Hence, by construction, the best fit value for the signal-strength $\mu(t\bar{t}+Z)$ (see eq. 4.7) is always expected to be

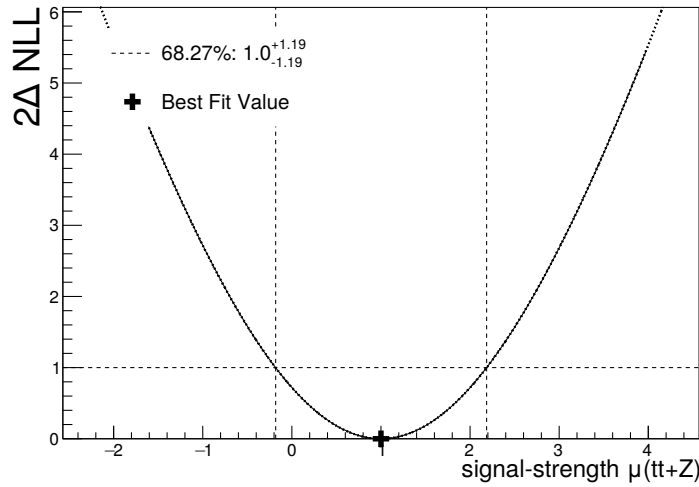


Figure 7.5: **Negative log-likelihood as a function of the signal-strength parameter.** Nuisance parameters are profiled. The vertical dashed lines indicate the 68% confidence interval.

1. The results are evaluated in terms of best fit uncertainties, expected limits, pulls of the nuisance parameters and uncertainties and impacts of systematic uncertainties and uncertainty groups. Each of these will be introduced and evaluated in the following.

It was observed that the variations of uncertainties corresponding to jet energy corrections and parton shower ISR/FSR descriptions did not propagate smoothly to the discriminator distributions. Many of the varied templates were one-sided, i.e. both the up and down variation changed the yields in a bin in the same direction. This might introduce unwanted effects or constraints to the fit which are not well motivated. Hence, it was decided to symmetrize these uncertainties. For this, the differences of both varied templates with respect to the nominal template (shift values) are calculated on a bin-by-bin basis. If the variations are one-sided, the template with the larger distance to the nominal template retains its direction, while the other template is flipped. The values of these templates (and templates in bins with two-sided variations) are symmetrized with the average of the two calculated shift values.

Furthermore, the uncertainties corresponding to jet energy corrections introduced instabilities to the fit model when also applied to minor backgrounds (explicitly $V + \text{jets}$, $t\bar{t} + W$ and diboson processes). This might occur due to the sparseness of the templates of these processes in the discriminator distributions. Hence, it was decided to omit these uncertainties for the mentioned processes to guarantee a more stable result. The effect on the fit result is expected to be negligible.

Best fit evaluation

The value for μ minimizing the negative log-likelihood ratio (see eq. 4.11) is determined via minimization algorithms. A negative log-likelihood scan of the signal-strength for the fit to the Asimov data set is shown in Figure 7.5. An uncertainty on this best fit value is derived from scanning the negative log-likelihood function and finding the values for μ where the test statistic value $\tilde{q}_\mu$ is larger than the best fit value by a value of one. This corresponds to the 68% confidence interval and expresses the significance of the determination of μ . The resulting expected best fit value is

$$\mu(t\bar{t}+Z)^{\text{exp.}} = 1.00 \pm 1.19. \quad (7.1)$$

This corresponds to an expected significance of 0.84σ .

The expected result would not justify a claim of having discovered the $t\bar{t}+Z$ production process in this phase space region and using this analysis strategy. The expected significance is small in comparison to the significance needed to claim an observation or evidence of the sought-after process. The sensitivity is limited by the large irreducible background, the (in comparison to the backgrounds) small signal cross section and the difficulties of separating the signal process from these background processes (especially $t\bar{t}+H$ production).

The systematic uncertainties considered in this analysis also reduce the expected significance of this analysis. As a comparison the determination of the best fit value of $\hat{\mu}$ and its significance is also performed in a fit without the consideration of systematic uncertainties. In this case the expected best fit value is found to be

$$\mu(t\bar{t}+Z)_{\text{stat. only}}^{\text{exp.}} = 1.00 \pm 0.36, \quad (7.2)$$

corresponding to an expected significance of 2.8σ .

Expected limits

The procedure of deriving expected limits is described in section 4.2.3. Limits on the signal-strength parameter are used for the interpretation of analysis results when the significance is not expected to be large. The expected limit is derived from a fit to background-only pseudo data, using the CL_s method, testing the hypothesis of expecting data without the sought-after signal process. From this, a limit on the signal-strength parameter can be derived, excluding all signal-strength values larger than the limit with a certain confidence level. For the full setup including all systematic uncertainties, the expected limit is found to be

$$\mu(t\bar{t}+Z)^{\text{exp.}} < 2.32^{+1.03}_{-0.67} \text{ (95\% CL)}. \quad (7.3)$$

In Figure 7.6 the expected limits are shown in the combined analysis region and separately for each jet region covered by one ANN. As expected, the ≥ 6 jet region on its own is the most sensitive to the signal process, manifesting itself via a low expected limit of $\mu(t\bar{t}+Z)_{\geq 6 \text{ jets}}^{\text{exp.}} < 3.14$. The background-enriched four-jet and five-jet regions mainly contribute to the sensitivity of the measurement by constraining the background processes (especially $t\bar{t}+lf$ and $t\bar{t}+c\bar{c}$ production), which can be transferred to the signal enriched regions. Hence, the expected limits on $t\bar{t}+Z$ production are not very low for the four-jet and five-jet regions alone, but still lower the limit when considered in a combined evaluation.

Nuisance parameter pulls and uncertainties

A fit is performed with the full set of systematic uncertainties. Before the fit each uncertainty (incorporated as nuisance parameter θ in the fit) has an a-priori value θ_0 , derived via the individual methods described in Section 5.2. In the fit, the post-fit values $\hat{\theta}$ that minimize the negative log-likelihood ratio are obtained. Similarly, for the prior widths $\Delta\theta$ of the uncertainty distributions, post-fit uncertainties $\Delta\hat{\theta}$ are obtained. For example, some systematic uncertainties, e.g. the background normalization of the $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$ processes, have large prior widths, estimated conservatively before the fit. These are expected to be constrained during the fit, as the model is more sensitive to changes in the background normalizations. The differences between nuisance parameters before and after the fit are summarized via pulls of systematic uncertainties as shown in Figure 7.7. Depicted are the differences

$$\frac{(\hat{\theta} - \theta_0)}{\Delta\theta}, \quad (7.4)$$

for each systematic uncertainty and the corresponding constraints relative to the pre-fit values $\Delta\theta$.

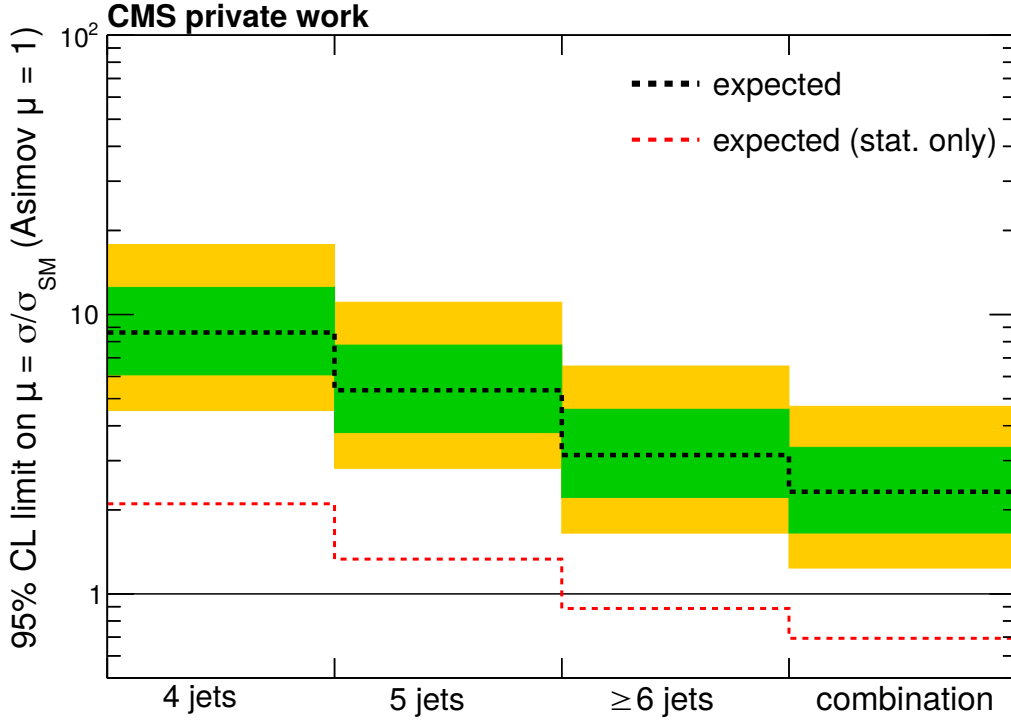


Figure 7.6: **Expected limits at 95% CL on $\mu(\bar{t}t+Z)$ with the full set of uncertainties.** The black line corresponds to the expected limit, the green and yellow bands correspond to the $\pm 1\sigma$ and $\pm 2\sigma$ quantiles of the limit distribution, respectively. The red line shows the expected limit in the case of statistical uncertainties only.

As expected, there are no differences between pre-fit and post-fit central values of nuisance parameters in the signal-plus-background fit (red), as the fit is performed to the Asimov data set with $\mu = 1$. In the background-only fit (blue) some deviations between pre-fit and post-fit central values are observed. As expected, these compensate the signal contribution in the signal-plus-background pseudo data set. None of the pulls in the background-only fit are large, as expected from the small signal cross-section.

The background normalization uncertainties for $\bar{t}t+b\bar{b}$ and $\bar{t}t+c\bar{c}$ (`bgnorm_ttbb/ttcc`), as well as the modeling uncertainty of the $\bar{t}t+2b$ contribution (`modeling_tt2b`) are constrained relative to the prior uncertainty by up to 40%. The contributions of these processes can be constrained in the fit, as their contributions to the analysis region are large. Thus, changes in the normalization of these processes are easily detectable, which in turn constrains the large 50% rate uncertainty assumed for these processes.

Uncertainties corresponding to b-tagging (e.g. `btag_hf`) are also expected to be constrained due to the high jet multiplicity and the large number of b-tagged jets and the separating power of b-tag related features. Also the uncertainties corresponding to charm quark tagging (`btag_cferr`) are constrained by approximately 50%, as their priors are estimated conservatively from the already constrainable heavy-flavor uncertainties (see section 5.1). The simulation of $\bar{t}t$ +jets processes is strongly dependent on the description of additional jet radiation from parton showers. Hence, the assumed uncertainties (e.g. `PS_ISR/FSR_ttbb`) might be constrained in the fit from the high jet multiplicity and the large irreducible background contributions of these processes to the analysis region. The contributions of

$t\bar{t}+H$ and $t\bar{t}+Z$ to the analysis region are small, thus, the corresponding parton shower uncertainties cannot be constrained beyond the prior uncertainty.

Also, uncertainties of jet energy corrections (JES/JER) are mildly constrained in the fit. Some have large prior values and can be constrained with the multi-jet final state.

In summary, the nuisance parameter pulls and constraints are evaluated for a fit to an Asimov data set. The observed pulls and constraints behave as expected. In conclusion, the fit model is expected to also perform well on real data.

Impacts of systematic uncertainties

The impacts of single nuisance parameters θ on the signal-strength parameter μ are determined by performing the fit, but fixing the nuisance parameter under scrutiny to its $\pm 1\sigma$ post-fit values. The impact $\pm\Delta\mu$ is calculated from the difference between the signal-strength value of the nominal fit, and the signal-strength obtained when fixing the nuisance parameter in question. In Figure 7.8 the twenty nuisance parameters with the highest impact on the signal-strength are shown.

The background normalization of the $t\bar{t}+b\bar{b}$ process and other background normalizations show the largest impacts on the signal-strength, with $\Delta\mu_{\text{bgnorm_ttbb}} = \pm 0.48$ being the largest. The $t\bar{t}+b\bar{b}$ process is the dominant contribution in bins sensitive to the $t\bar{t}+Z$ normalization, e.g. bins in the $t\bar{t}+Z$ discriminator of the ≥ 6 jet region. Therefore, varying this nuisance parameter is expected to change the $t\bar{t}+Z$ signal-strength prediction considerably.

Furthermore, bin-by-bin uncertainties due to the limited size of the simulated samples, especially for bins in the $t\bar{t}+Z$ discriminators, have large impacts on the best fit value. This can be inferred from the sensitivity on the best fit value in these bins explicitly. Hence, a variation of the bin content (predicted from simulation) has some impact on the fit result. Among the uncertainties with a sizable impact are also b-tagging uncertainties, parton shower ISR/FSR description uncertainties and some jet energy scale sources.

In Table 7.1 the impacts of whole nuisance parameter groups are summarized. These are derived by fixing the nuisance parameters of the group to its pre-fit values, thereby performing a fit with out them. From the squared difference between the post fit uncertainty of the fit with all uncertainties, and the fit with fixed nuisance parameters, the impact of this uncertainty group is derived.

The uncertainties are split into three groups; statistical uncertainties due to the limited number of data events, statistical uncertainties due to the limited number of simulated events and the systematic uncertainties defined in chapter 5. Of these groups the highest impact on the best fit uncertainty comes from the systematic uncertainties. The limited MC statistics also contributes significantly to the best fit uncertainty. This contribution could be reduced by simulating more events for the background modeling. Similarly, more data reduces the impact of the statistical uncertainty.

The systematic uncertainties are further divided into experimental and theory-based uncertainties. The impact of the experimental uncertainties is outweighed by the impact of the theory-based uncertainties. The major contribution of the theory uncertainties stems from background normalizations, followed by uncertainties due to the tuning of the parton shower simulation. The largest impact among experimental uncertainties originates from uncertainties from the derivation of b-tagging scale factors, followed by uncertainties from jet energy corrections.

7.4 Impact of $t\bar{t}+H$ Contribution

The analysis presented in this thesis is performed on data taken in the year 2018. This data has not yet (as the time of writing) been analyzed regarding its $t\bar{t}+H$ content. As the

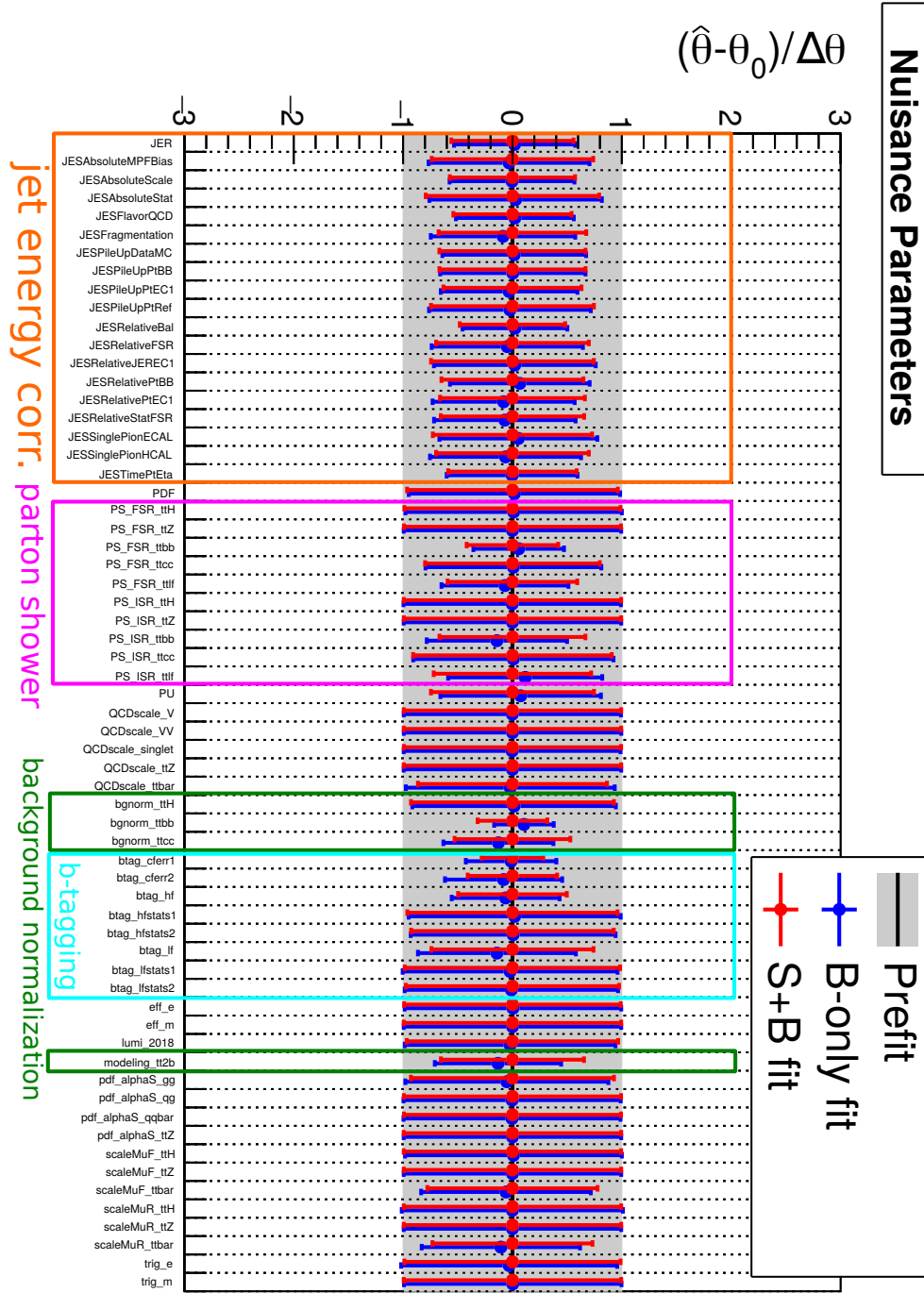


Figure 7.7: **Pulls and uncertainties of nuisance parameters from the fit to Asimov data.** The gray background corresponds to the pre-fit expectation $\Delta\theta$ of the uncertainties. The pulls (i.e. deviations) of the post-fit central values $\hat{\theta}$ relative to their pre-fit values θ_0 and uncertainties $\Delta\theta$ are shown in red for the signal-plus-background fit and in blue for the background-only fit. Some groups of nuisance parameters are highlighted in colored boxes, e.g. nuisance parameters corresponding to jet energy correction uncertainties are marked in orange.

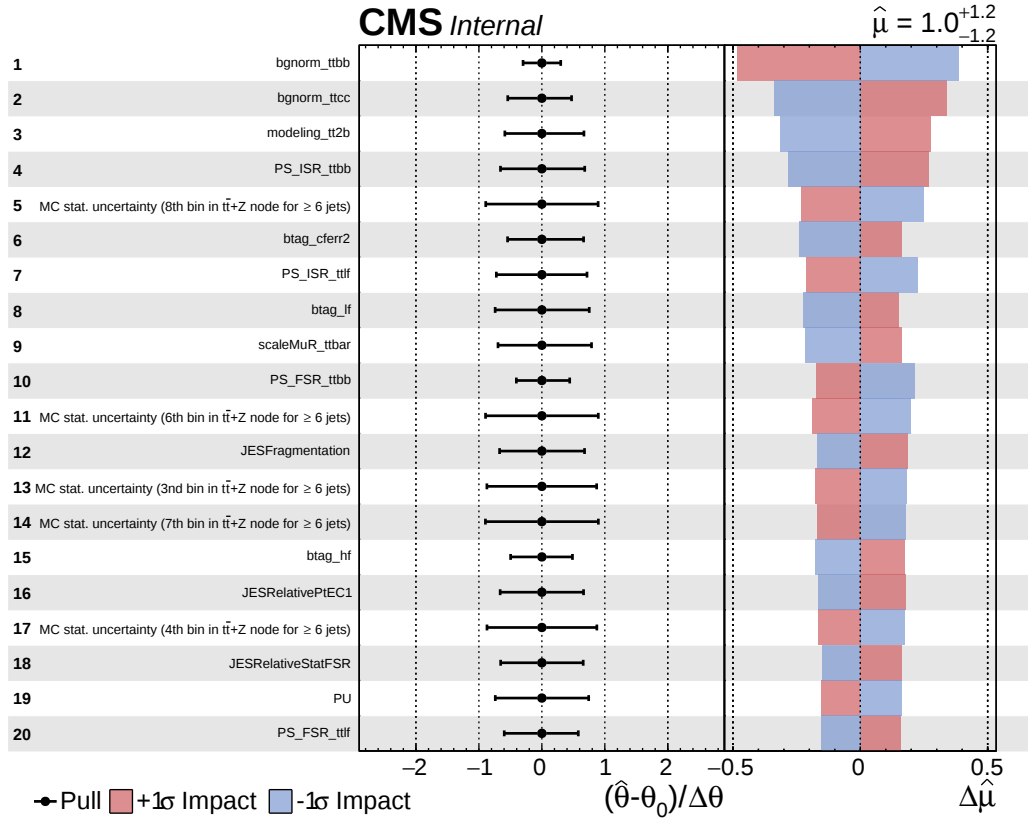


Figure 7.8: **Impacts of systematic uncertainties.** Shown are the uncertainties with the 20 highest impacts on the best fit value. The impact is computed as the difference of the nominal best fit value of μ and the best fit value obtained when fixing the nuisance parameter under scrutiny to its best fit value $\hat{\theta}$ plus (red) or minus (blue) its post fit uncertainty.

Table 7.1: **Contributions of groups of nuisance parameters on the expected best fit value $\mu(t\bar{t}+Z)$.** The contributions $\Delta\mu$ to the best fit value are obtained by fixing the group of uncertainties under scrutiny to their post fit values in the fit and subtracting the obtained best fit uncertainty in quadrature from the result of the nominal fit.

Uncertainty group	contribution to best fit value ($\Delta\mu$)
Total experimental	+0.62/−0.67
b-tagging	+0.35/−0.45
jet energy scale and resolution	+0.45/−0.45
Total theory	+0.75/−0.76
background normalization	+0.55/−0.62
parton shower	+0.42/−0.46
Total systematic	+1.05/−1.05
Size of simulated samples	+0.61/−0.64
Statistical	
Total uncertainty	+1.19/−1.19

analysis region is modeled after the analysis region of the $t\bar{t}+H$, $H \rightarrow b\bar{b}$ analysis in the semileptonic $t\bar{t}$ decay channel, it has to be verified that the fit of the $t\bar{t}+Z$ signal-strength in this analysis is not sensitive to the $t\bar{t}+H$ contribution.

Due to the similar kinematic features of the $t\bar{t}+Z$ and $t\bar{t}+H$ processes, they are expected to be classified similarly with the ANN classification approach. Thus, two distinct classes have been introduced for both processes, and input features specifically designed to separate $t\bar{t}+Z$ and $t\bar{t}+H$ events have been included in the input feature selection. This aims at constructing templates for the $t\bar{t}+Z$ and $t\bar{t}+H$ processes that differ as much as possible in shape and rate, such that the determination of the $t\bar{t}+Z$ normalization is affected by the $t\bar{t}+H$ normalization as little as possible.

As can be seen in the pre-fit distributions shown in Appendix D (left), the contributions of $t\bar{t}+H$ (red) and $t\bar{t}+Z$ (orange) differ primarily in the dedicated output discriminators. For example, the $t\bar{t}+H$ discriminator distribution of the ≥ 6 jet region is vastly enriched in $t\bar{t}+H$ in comparison to $t\bar{t}+Z$ events. The $t\bar{t}+H$ yield in this template surpasses the $t\bar{t}+Z$ yield by a factor of 3.54. Similarly, $t\bar{t}+Z$ events are enriched in the corresponding discriminator of the ≥ 6 jet region, reaching a yield of 0.87 relative to the $t\bar{t}+H$ yield in this output node. For reference, the ratio of the total number of $t\bar{t}+Z$ events relative to the total number of $t\bar{t}+H$ events is 0.57 in the ≥ 6 jet region (c.f. Table 3.1). Especially in bins with high ANN output values the yield of $t\bar{t}+Z$ is enriched by a factor of 4.61 relative to the $t\bar{t}+H$ yield.

The experimental uncertainty on the cross-section of $t\bar{t}+H$ production from the combination of measurements performed with data taken during LHC Run-I and 2017 was determined to be $+24.6\%/-20.6\%$ [110]. Hence, different hypotheses of the cross-section of $t\bar{t}+H$ production are tested with the final analysis strategy devised in this thesis. This aims at estimating the impact of a changing $t\bar{t}+H$ production cross-section on the $t\bar{t}+Z$ signal-strength determination. Optimally, the variation of the $t\bar{t}+H$ production cross-section should not have an effect on the $t\bar{t}+Z$ signal-strength parameter. This would suggest that the contributions of $t\bar{t}+H$ and $t\bar{t}+Z$ in the discriminators are distinctive enough to not be strongly correlated in the fit.

Two fits of the nominal background templates including all systematic uncertainties to pseudo data sets are performed. The predicted contributions of $t\bar{t}+H$ production are increased and decreased by 30% in the pseudo data set, to approximately cover the uncertainty on the $t\bar{t}+H$ cross-section.

For an upwards or downwards variation of the $t\bar{t}+H$ production cross-section by 30%, the best fit value of the $t\bar{t}+Z$ signal-strength changed by $+0.04$ and -0.05 , respectively. This change is small compared to the expected best fit uncertainty of ± 1.19 (see eq. 7.1) or the impact of some other nuisance parameters on the signal-strength (see Figure 7.8). In the progress of the optimization studies summarized in section 6.5, the impact of the $t\bar{t}+H$ production cross-section variation on the $t\bar{t}+Z$ signal-strength was checked for various approaches. Among these, strategies including separated ANN output nodes for $t\bar{t}+H$ and $t\bar{t}+Z$ showed the best robustness against the $t\bar{t}+H$ normalization. As a comparison, in approaches where only four output nodes are used (i.e. no $t\bar{t}+H$ node), the $t\bar{t}+Z$ signal-strength changed by up to ± 0.7 for 30% variations of $t\bar{t}+H$.

In the final setup, the signal-strength and the uncertainty corresponding to the normalization of the $t\bar{t}+H$ process are correlated by -1% . A small anti-correlation is still expected, as the separation between $t\bar{t}+H$ and $t\bar{t}+Z$ with this setup is not perfect. An increase in the normalization of one of these processes might still lead to a small decrease in the normalization of the other process.

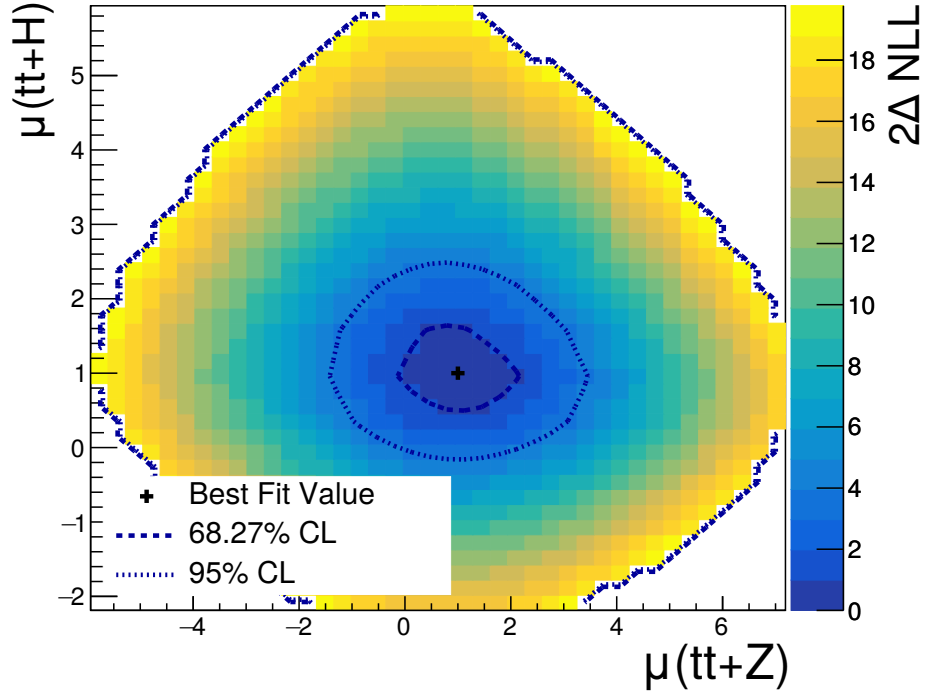


Figure 7.9: **Negative log-likelihood as a function of $\mu(t\bar{t}+Z)$ and $\mu(t\bar{t}+H)$.** The best fit value (i.e. minimum of the surface) is marked with a cross. The 1σ and 2σ contours are shown as dashed lines.

In summary, the impact of the $t\bar{t}+H$ production cross-section variation was found to be small in comparison to the sensitivity of the $t\bar{t}+Z$ signal-strength measurement itself. Thus, it is neither expected that the determination of $\mu(t\bar{t}+Z)$ is very sensitive to the normalization of the $t\bar{t}+H$ production cross-section, nor is it expected that the sensitivity to the $t\bar{t}+H$ production cross-section can be inferred from the determination of the $t\bar{t}+Z$ signal-strength.

It is important that an analysis is performed unbiased towards the expectations from real data. Thus, being sensitive to the $t\bar{t}+H$ contribution with this analysis and thereby gathering information on the $t\bar{t}+H$ content of the 2018 data set could bias the ongoing LHC Run-II analysis of $t\bar{t}+H$. Hence, a simultaneous measurement of $t\bar{t}+H$ and $t\bar{t}+Z$ is not performed on real data. Instead, a simultaneous fit of the signal-strength parameters for both $t\bar{t}+H$ and $t\bar{t}+Z$ is only performed with an Asimov data set. In Figure 7.9 the result is shown. The corresponding best fit values are

$$\mu(t\bar{t}+Z)_{2D}^{\text{exp.}} = 1.00 \pm 1.20 \quad \text{and} \quad (7.5)$$

$$\mu(t\bar{t}+H)_{2D}^{\text{exp.}} = 1.00 \pm 0.62. \quad (7.6)$$

The best fit uncertainty of the signal-strength of $t\bar{t}+Z$ does not change significantly in comparison to the previous fit (see eq. 7.1). Both signal-strength parameters are correlated by -6% .

8 Results

In this chapter the results of the analysis are presented and discussed. The signal-strength of $t\bar{t}+Z$ is determined in a likelihood fit to the data. All systematic uncertainties described in section 5 are considered via profiled nuisance parameters. This setup corresponds exactly to the setup tested with pseudo data in section 7.3. The discriminator distribution of the most sensitive region, i.e. the $t\bar{t}+Z$ node of the ≥ 6 jet region, is shown in Figure 8.1 before and after the fit. All discriminator distributions are summarized in Appendix D.

In Figure 8.2 the best fit values and uncertainties of all nuisance parameters after the fit are shown. The majority of best fit values are within one standard deviation of the

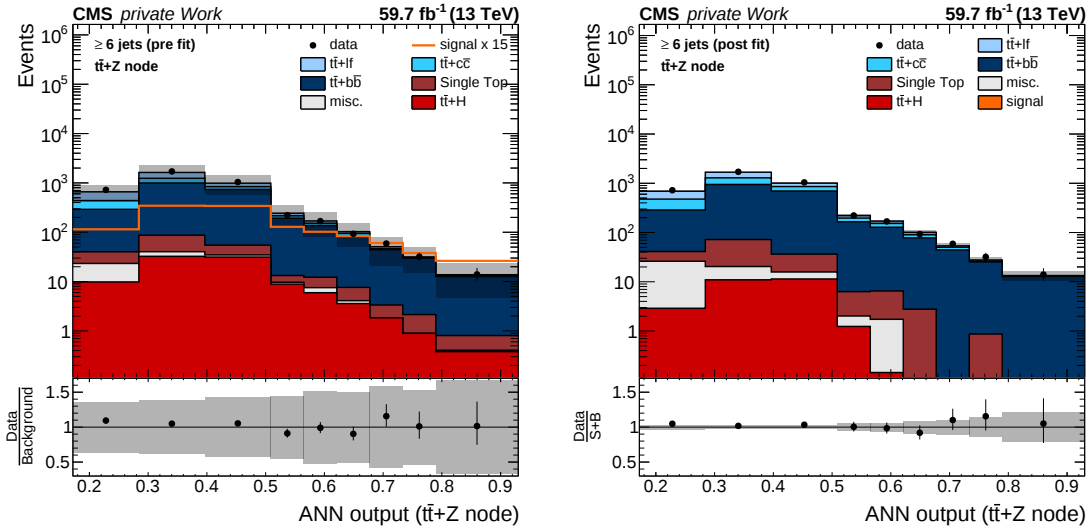


Figure 8.1: **Final discriminator distribution of the $t\bar{t}+Z$ output node in the ≥ 6 jet region before (left) and after (right) the fit to data.** The contributions of all background processes are stacked. In the pre-fit case, the signal contribution ($t\bar{t}+Z$) is scaled by a factor of 15 and overlaid as a line. In the post-fit case, the fitted signal contribution is also stacked. The uncertainty bands include the total uncertainty of the fit model. Ratios between data (black dots) and the background (pre-fit) and the sum of signal and background (post-fit) are shown in the bottom panel.

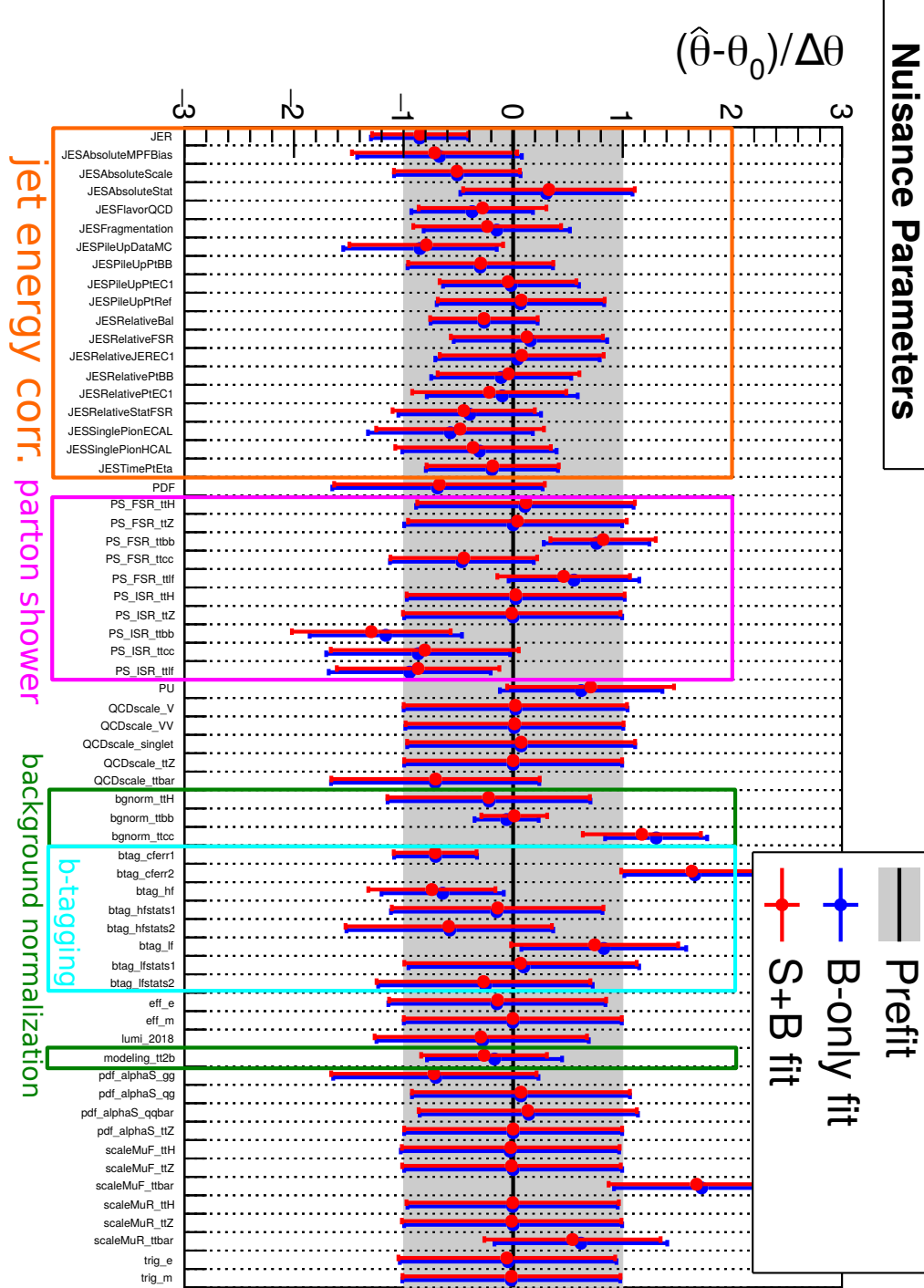


Figure 8.2: **Pulls and uncertainties of nuisance parameters from the fit to data.**

The gray background corresponds to the pre-fit expectation $\Delta\theta$ of the uncertainties. The pulls (i.e. deviations) of the post-fit central values $\hat{\theta}$ relative to their pre-fit values θ_0 and uncertainties $\Delta\theta$ are shown in red for the signal-plus-background fit and in blue for the background-only fit. Some groups of nuisance parameters are highlighted in colored boxes, e.g. nuisance parameters corresponding to jet energy correction uncertainties are marked in orange.

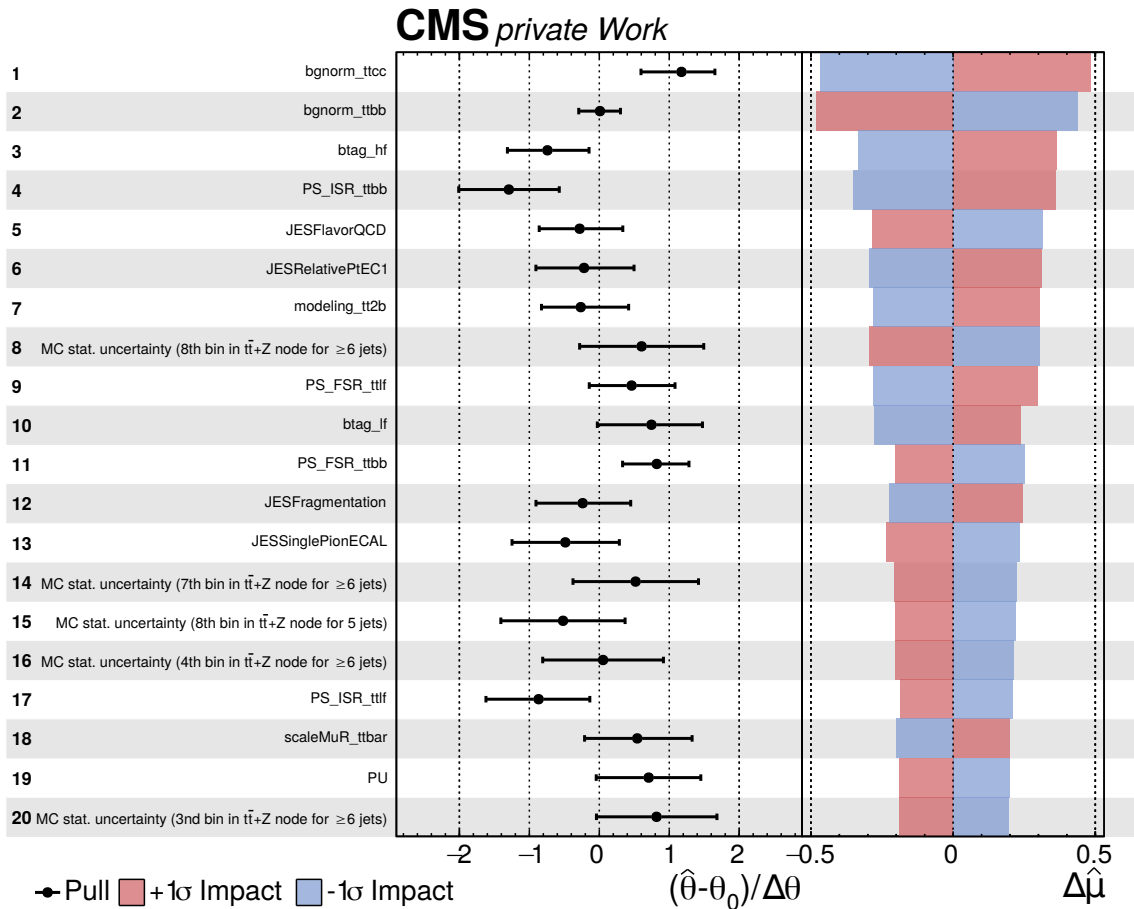


Figure 8.3: **Impacts of systematic uncertainties.** Shown are the uncertainties with the 20 highest impacts on μ in a fit to data. The impact is computed as the difference of the nominal best fit value of μ and the best fit value obtained when fixing the nuisance parameter under scrutiny to its best fit value $\hat{\theta}$ plus (red) or minus (blue) its post fit uncertainty.

pre-fit expectation. The nuisance parameters show a similar reduction of uncertainties as in the pre-fit expectation shown in Figure 7.7. The post-fit value of the background normalization of the $t\bar{t}+c\bar{c}$ process deviates by more than one standard deviation from its pre-fit value. The post-fit value of the background normalization of the $t\bar{t}+b\bar{b}$ process does not deviate from the pre-fit value at all, which is attributed to the scaling of the $t\bar{t}+b\bar{b}$ normalization before the fit to data, introduced in section 6.2. This nuisance parameter is also constrained to a level of 40% of its pre-fit uncertainty.

The 20 parameters with the highest impact on the best fit value are shown in Figure 8.3. As expected, the highest impacts are attributed to the background normalizations of the $t\bar{t}+b\bar{b}$ and $t\bar{t}+c\bar{c}$ processes. These are followed by the nuisance parameters corresponding to b-tagging uncertainties and the description of initial state radiation in the parton shower for the $t\bar{t}+b\bar{b}$ process. Both uncertainties are also expected to have a large impact on the final result, e.g. due to the high jet multiplicity. The order and size of the 20 highest impacts is consistent with the expectations depicted in Figure 7.8. 16 out of the 20 nuisance parameters with the highest impacts for a fit to data are also among the 20 nuisance parameters with the highest impacts for a fit to an Asimov data set.

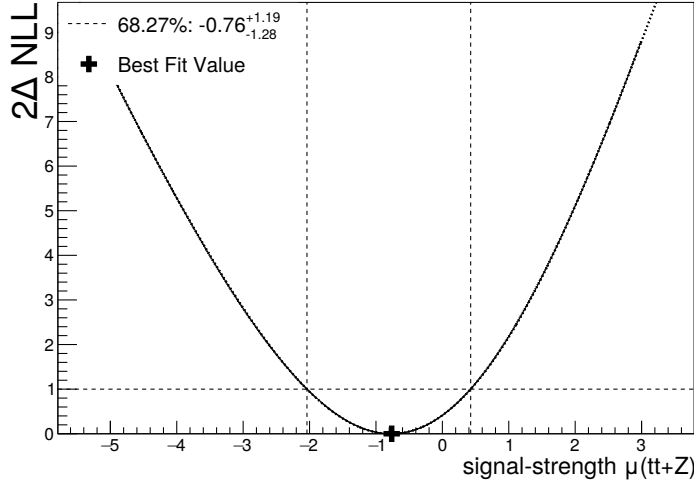


Figure 8.4: **Negative log-likelihood as a function of the μ parameter in a fit to data.** Nuisance parameters are profiled. The vertical dashed lines indicated the 68% confidence interval.

The observed best fit value of the signal-strength is

$$\mu(\bar{t}t+Z)^{\text{obs.}} = -0.77^{+1.19}_{-1.26}. \quad (8.1)$$

The observed limit on the signal-strength is

$$\mu(\bar{t}t+Z)^{\text{obs.}} < 1.90. \quad (8.2)$$

As the best fit value of the signal-strength is smaller than zero, no significance of the measurement is calculated. The best fit value is within two standard deviations of the SM expectation. The observed best fit uncertainties are of the same magnitude as the expected values. However, a larger downward uncertainty relative to the expectation is observed. A scan of the negative log-likelihood ratio as a function of the signal-strength is shown in Figure 8.4, visualizing the tilt towards negative values. The observed limit is below the expected limit (see eq. 7.3), within one standard deviation.

The results are summarized in Table 8.1 for the full combination and also per jet region. In Figure 8.5, the expected and observed limits are shown. The best fit values of the four-jet and five-jet regions also deviate by one standard deviation from the expected values. The per-category results are compatible with each other and with the combination.

In Table 8.2, the nominal event yields per process before and after the combined fit to data are summarized. The ratios between post-fit and pre-fit yields are shown in Figure 8.6. With the negative best fit value, the event yields of the signal process are negative. Due to the high post-fit value of the $\bar{t}t+c\bar{c}$ background normalization nuisance parameter, the event yields of that process are approximately 43% larger than the pre-fit yields. The post-fit yields of some minor backgrounds (V + jets and diboson) are increased by more than 10% with respect to the pre-fit yields. All other post-fit yields agree with the pre-fit yields within 10%.

In Table 8.3, the contribution of different groups of uncertainties to the total uncertainty are listed. Compared to the results of the fit to pseudo data (see Table 7.1), the contributions are slightly larger for all groups of systematic uncertainties. These uncertainty groups appear to be correlated stronger in the fit to data compared to the fit to pseudo data.

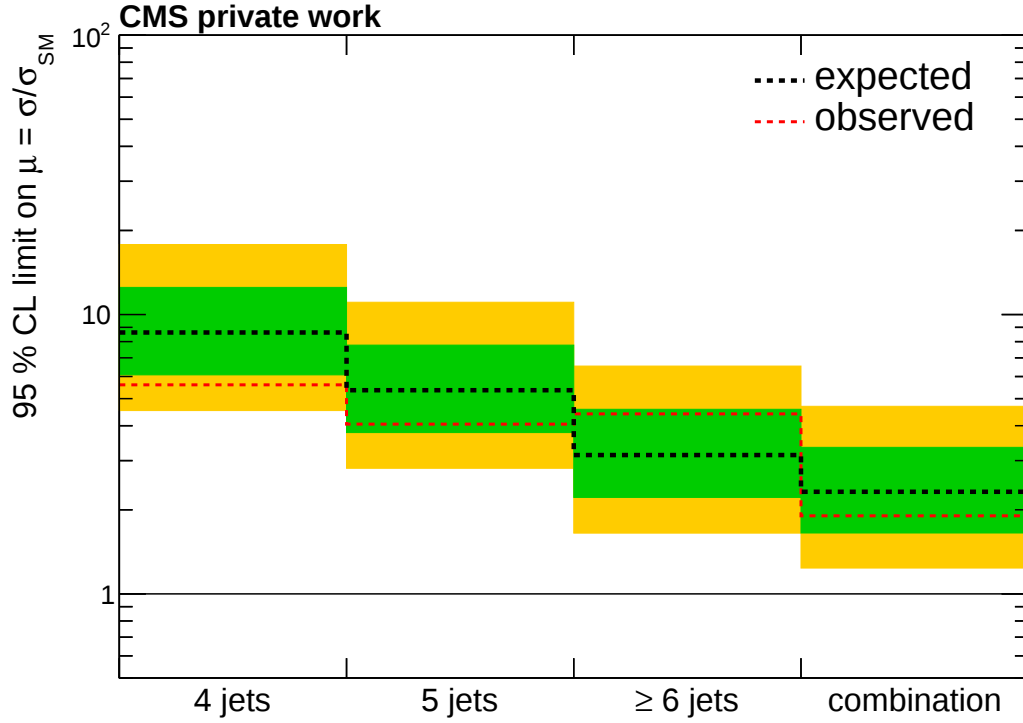


Figure 8.5: **Expected and observed limits at 95% CL on $\mu(\bar{t}t+Z)$ with the full set of uncertainties.** The black line corresponds to the expected limit, the green and yellow bands correspond to the $\pm 1\sigma$ and $\pm 2\sigma$ quantiles of the limit distribution, respectively. The red line shows the observed limit.

This is inferred from the slightly larger $\Delta\mu$ values of the systematic uncertainty groups in a fit to data compared to a fit to pseudo data, in combination with same total best fit uncertainties. The lower uncertainty of all groups of systematic uncertainties is larger than the upper uncertainty, which is also observed in the total best fit uncertainty.

Compared to the analyses of the $\bar{t}t+Z$ process in final states with two, three or four charged leptons [6–8] (see section 6.1), the observed best fit uncertainty in this thesis is very large. In these analyses, a significance of more than 5σ can be claimed. Nevertheless, the analysis of the $\bar{t}t+Z$ process in the $Z \rightarrow b\bar{b}$ final state can still contribute to a possible combined measurement, enhancing the significance of the complete $\bar{t}t+Z$ production measurement, as the analyzed phase spaces are not overlapping, or possibly constraining jet-related uncertainties due to the large jet-multiplicity in this analysis region.

Table 8.1: **Best-fit signal-strength μ , 95% CL limit on μ and significance obtained from the fit to data and expected from Asimov data in parentheses.**

	best fit obs. (exp.)	limit obs. (exp.)	significance obs. (exp.)
4 jets	$-7.50^{+4.90}_{-5.57}$ (1.00 $^{+4.27}_{-4.32}$)	$\mu < 5.60$ (8.91)	— (0.24 σ)
5 jets	$-2.95^{+2.82}_{-3.22}$ (1.00 $^{+2.69}_{-2.72}$)	$\mu < 4.04$ (5.52)	— (0.37 σ)
≥ 6 jets	$+1.27^{+1.70}_{-1.62}$ (1.00 $^{+1.60}_{-1.57}$)	$\mu < 4.40$ (3.14)	0.78 σ (0.64 σ)
combined	$-0.77^{+1.19}_{-1.26}$ (1.00 $^{+1.19}_{-1.19}$)	$\mu < 1.90$ (2.32)	— (0.84 σ)

Table 8.2: **Nominal event yields before (top row) and after (bottom row) the combined fit to data.** Event yields for events with four, five and six or more jets and the total yield are listed.

process	4 jet	5 jet	≥ 6 jets	total yield
$t\bar{t}+Z$	181.6 −146.8	276.7 −220.7	416.8 −320.1	875.2 −687.6
$t\bar{t}+lf$	58071.7 57392.2	34182.3 32814.3	21245.9 18950.6	113500.0 109157.1
$t\bar{t}+b\bar{b}$	13902.9 14932.0	16740.4 16993.3	20100.2 18279.1	50743.4 50204.5
$t\bar{t}+c\bar{c}$	9465.8 14083.6	10287.8 14689.1	11183.5 14664.2	30937.0 43436.9
single t	3533.5 3711.6	2194.7 2282.4	1486.7 1521.9	7214.9 7515.9
V + jets	1307.4 1508.1	474.8 592.4	371.6 431.2	2153.8 2531.7
$t\bar{t}+H$	259.6 251.0	418.5 382.6	684.8 602.0	1362.9 1235.5
$t\bar{t}+W$	68.9 65.2	105.9 100.0	167.1 158.3	341.9 323.1
diboson	67.5 77.3	30.7 37.0	13.7 15.4	111.9 129.7

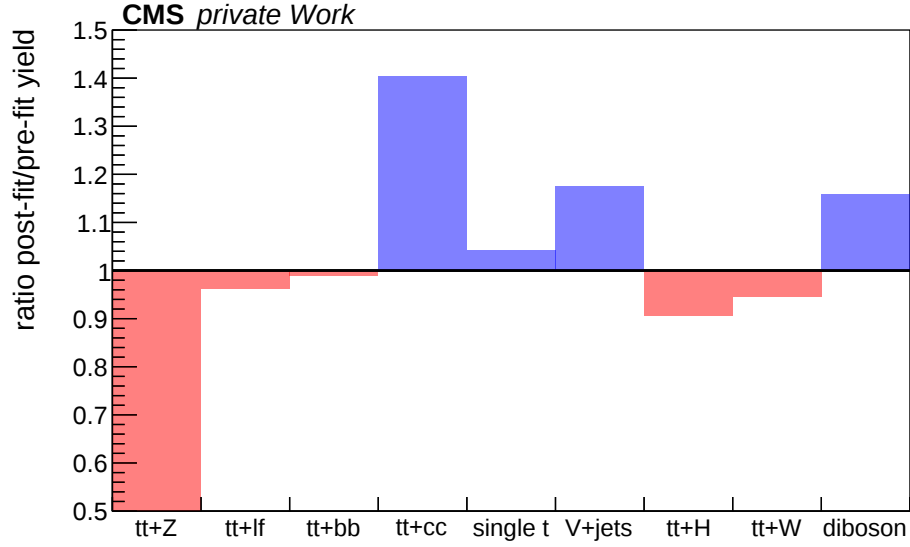


Figure 8.6: **Ratio between post-fit and pre-fit yields for each process in the combined fit to data.** Blue bars indicate an increase of the post-fit yield w.r.t. the pre-fit yield, red bars indicate a decrease.

Table 8.3: **Contributions of groups of nuisance parameters on the observed best fit value $\mu(t\bar{t}+Z)$.** The contributions $\Delta\mu$ to the best fit value are obtained by fixing the group of uncertainties under scrutiny to their post fit values in the fit and subtracting the obtained best fit uncertainty in quadrature from the result of the nominal fit.

Uncertainty group	contribution to best fit value ($\Delta\mu$)
Total experimental	+0.74/−0.81
b-tagging	+0.47/−0.55
jet energy scale and resolution	+0.62/−0.58
Total theory	+0.79/−0.91
background normalization	+0.69/−0.77
parton shower	+0.52/−0.56
Total systematic	+1.06/−1.15
Size of simulated samples	+0.69/−0.72
Statistical	+0.37/−0.37
Total uncertainty	+1.19/−1.26

9 Summary and Outlook

In this thesis, the first measurement of $t\bar{t}+Z$ production in a phase space region targeting the $Z \rightarrow b\bar{b}$ decay at the CMS experiment is documented.

The analysis strategy, derived from $t\bar{t}H(bb)$ analyses [97], was studied and developed on simulated data. Event selections are designed to select events where one of the top quarks decays leptonically and the other one decays hadronically. This provides efficient suppression of QCD multi-jet production due to the charged lepton of the leptonic top quark decay. Events are selected with a high jet multiplicity of which some are required to be tagged as jets originating from bottom quarks. The major irreducible background in this analysis region is the production of top quark-antiquark pairs with additional jets ($t\bar{t}+\text{jets}$).

Events are categorized further using multivariate classification approaches, either as the sought-after $t\bar{t}+Z$ contribution or four distinct background processes. Artificial Neural Networks (ANN) are used for the classification. Dedicated ANNs are constructed for events with four, five and equal or more than six jets. These ANNs use various event features to discriminate between the distinct processes. Among these features are b-tagging related features, angular differences of jets, mass related features and features of a χ^2 -based event reconstruction. This reconstruction targets the assignment of jets to the objects expected in a $t\bar{t}+Z$ event.

The features are chosen based on a ranking measure derived from the impact of the feature under scrutiny on the classification output. The modeling of these features is validated in a comparison between simulated events and data events.

Final discriminator distributions are obtained from the ANN output. The signal is extracted with a binned Maximum Likelihood (ML) fit of the distributions to data. Experimental uncertainties due to e.g. the derivation of b-tagging scale factors or jet energy corrections, and also theory uncertainties, e.g. from unknown background normalizations or the description of initial-state and final-state radiation in parton shower simulations are considered in the fitting procedure. The analysis strategy and fit model are thoroughly validated with simulated data, as well as in comparison between simulation and data.

The analysis is performed with data recorded in the year 2018 at the CMS experiment as part of the LHC Run-II. A best fit signal-strength of $\mu = \sigma/\sigma^{\text{SM}} = -0.77^{+1.19}_{-1.26}$ is obtained. This corresponds to an upper limit on μ of 1.90 ($2.32^{+1.03}_{-1.27}$ expected) at 95% CL. This represents the first measurement of $t\bar{t}+Z$ in the $Z \rightarrow b\bar{b}$ channel at CMS. The result complements the measurements in leptonic final states [6–8] and could be combined with these in the future.

For the purpose of this thesis, a few simplifications have been introduced to the uncertainty model. For example, systematic uncertainties corresponding to the description of the underlying-event and the tune of the parton shower simulation have been omitted. Uncertainties of jet energy corrections and the description of initial-state and final-state radiation in the parton shower simulation have been symmetrized on a bin-by-bin basis. Furthermore, the jet energy correction uncertainties have not been applied to minor backgrounds due to the lack of simulation statistics. In the future, these effects could be modeled in more detail.

A possible improvement to be considered in a future analysis is a more efficient event reconstruction. The χ^2 -based event reconstruction already enhances the classification power of the ANNs used in this analysis. In recent works in the scope of the $t\bar{t}+H$ and tH analyses, ANN-based and BDT-based jet assignment methods are developed which show higher assignment accuracies compared to the χ^2 -based method [118]. Applying these methods could increase the classification performance of the ANNs even further.

Another enhancement of the analysis could be to improve the modeling of the $t\bar{t}+b\bar{b}$ background via dedicated Monte Carlo samples [119]. In the current $t\bar{t}$ event simulation, the additional jets required for $t\bar{t}+b\bar{b}$ stem from the parton shower, while they are calculated at the matrix element in dedicated $t\bar{t}+b\bar{b}$ simulations.

As another improvement, a simultaneous measurement of the $t\bar{t}+Z$ and $t\bar{t}+H$ signal-strength modifiers, could be performed, as already exercised on pseudo data in this thesis. With the analysis strategy developed in this thesis, these modifiers are expected to be measured almost without correlation. In a future full LHC Run-II analysis, this simultaneous measurement could also be performed on real data. Furthermore, the inclusion of a simultaneous measurement of the $t\bar{t}+b\bar{b}$ signal-strength modifier can be considered, instead of determining an independent scale factor as done in this thesis.

Bibliography

- [1] The ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. eprint: 1207.7214 (hep-ex).
- [2] The CMS Collaboration. “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. eprint: 1207.7235 (hep-ex).
- [3] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.
- [4] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.
- [5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.
- [6] The ATLAS Collaboration. “Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. D* 99.7 (2019), p. 072009. DOI: 10.1103/PhysRevD.99.072009.
- [7] The CMS Collaboration. “Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *JHEP* 08 (2018), p. 011. DOI: 10.1007/JHEP08(2018)011.
- [8] The CMS Collaboration. “Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”. 2019. URL: <http://cds.cern.ch/record/2666205>.
- [9] E. Fermi. “Versuch einer Theorie der β -Strahlen. I”. In: *Zeitschrift für Physik* 88.3 (Mar. 1934), pp. 161–177. ISSN: 0044-3328. DOI: 10.1007/BF01351864.
- [10] S. Tomonaga. “On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields”. In: *Progress of Theoretical Physics* 1.2 (Aug. 1946), pp. 27–42. ISSN: 0033-068X. DOI: 10.1143/PTP.1.27.
- [11] J. Schwinger. “Quantum Electrodynamics. I. A Covariant Formulation”. In: *Phys. Rev.* 74 (10 Nov. 1948), pp. 1439–1461. DOI: 10.1103/PhysRev.74.1439.
- [12] R. P. Feynman. “The Theory of Positrons”. In: *Phys. Rev.* 76 (6 Sept. 1949), pp. 749–759. DOI: 10.1103/PhysRev.76.749.
- [13] M. Gell-Mann. *THE EIGHTFOLD WAY: A THEORY OF STRONG INTERACTION SYMMETRY*. Tech. rep. Mar. 1961. DOI: 10.2172/4008239.

- [14] S. L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [15] M. Gell-Mann. “Symmetries of Baryons and Mesons”. In: *Phys. Rev.* 125 (3 Feb. 1962), pp. 1067–1084. DOI: 10.1103/PhysRev.125.1067.
- [16] J. Goldstone, A. Salam, and S. Weinberg. “Broken Symmetries”. In: *Phys. Rev.* 127 (3 Aug. 1962), pp. 965–970. DOI: 10.1103/PhysRev.127.965.
- [17] M. Gell-Mann. “A schematic model of baryons and mesons”. In: *Physics Letters* 8.3 (1964), pp. 214–215. ISSN: 0031-9163. DOI: [https://doi.org/10.1016/S0031-9163\(64\)92001-3](https://doi.org/10.1016/S0031-9163(64)92001-3).
- [18] A. Salam and J. Ward. “Electromagnetic and weak interactions”. In: *Physics Letters* 13.2 (1964), pp. 168–171. ISSN: 0031-9163. DOI: [https://doi.org/10.1016/0031-9163\(64\)90711-5](https://doi.org/10.1016/0031-9163(64)90711-5).
- [19] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.
- [20] S. L. Glashow, J. Iliopoulos, and L. Maiani. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Phys. Rev. D* 2 (7 Oct. 1970), pp. 1285–1292. DOI: 10.1103/PhysRevD.2.1285.
- [21] M. Kobayashi and T. Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.
- [22] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.
- [23] M. Goldhaber, L. Grodzins, and A. W. Sunyar. “Helicity of Neutrinos”. In: *Phys. Rev.* 109 (3 Feb. 1958), pp. 1015–1017. DOI: 10.1103/PhysRev.109.1015.
- [24] The SLAC Collaboration. “Discovery of a Narrow Resonance in e^+e^- Annihilation”. In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406.
- [25] The BNL Collaboration. “Experimental Observation of a Heavy Particle J ”. In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404.
- [26] S. W. Herb et al. “Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions”. In: *Phys. Rev. Lett.* 39 (5 Aug. 1977), pp. 252–255. DOI: 10.1103/PhysRevLett.39.252.
- [27] The UA1 Collaboration. “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV”. In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2).
- [28] The UA2 Collaboration. “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider”. In: *Physics Letters B* 122.5 (1983), pp. 476–485. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2).
- [29] The UA1 Collaboration. “Experimental observation of lepton pairs of invariant mass around 95 GeV at the CERN SPS collider”. In: *Phys. Lett. B* 126.CERN-EP-83-073 (June 1985), 398–410. 17 p. DOI: 10.1016/0370-2693(83)90188-0.
- [30] The NA48 Collaboration. “A new measurement of direct CP violation in two pion decays of the neutral kaon”. In: *Physics Letters B* 465.1 (1999), pp. 335–348. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/S0370-2693\(99\)01030-8](https://doi.org/10.1016/S0370-2693(99)01030-8).

- [31] The CDF Collaboration. “Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab”. In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626.
- [32] The D0 Collaboration. “Observation of the Top Quark”. In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.74.2632>.
- [33] T. Lancaster and S. Blundell. *Quantum field theory for the gifted amateur*. Oxford University Press, 2014.
- [34] R. Wolf. *The Higgs Boson Discovery at the Large Hadron Collider*. Vol. 264. Springer, 2015. ISBN: 9783319185125, 9783319185118. DOI: 10.1007/978-3-319-18512-5.
- [35] H. Yukawa. “On the Interaction of Elementary Particles I”. In: *Proc. Phys. Math. Soc. Jap.* 17 (1935). [Prog. Theor. Phys. Suppl.1,1(1935)], pp. 48–57. DOI: 10.1143/PTPS.1.1.
- [36] M. Tanabashi et al. *Review of Particle Physics*. 2018. DOI: 10.1103/PhysRevD.98.030001.
- [37] The KATRIN Collaboration. *An improved upper limit on the neutrino mass from a direct kinematic method by KATRIN*. Sept. 13, 2019. arXiv: 1909.06048v1 [hep-ex].
- [38] R. P. Feynman. “Space-Time Approach to Quantum Electrodynamics”. In: *Phys. Rev.* 76 (6 Sept. 1949), pp. 769–789. DOI: 10.1103/PhysRev.76.769.
- [39] B. Madore. *A Knowledgebase for Extragalactic Astronomy and Cosmology*. (Accessed 19.09.19). URL: ned.ipac.caltech.edu/level5/March10/Garrett/Garrett4.html.
- [40] R. D. Ball et al. “Parton distributions from high-precision collider data”. In: *Eur. Phys. J. C* 77 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].
- [41] J. D. Bjorken and E. A. Paschos. “Inelastic Electron-Proton and γ -Proton Scattering and the Structure of the Nucleon”. In: *Phys. Rev.* 185 (5 Sept. 1969), pp. 1975–1982. DOI: 10.1103/PhysRev.185.1975.
- [42] O. B. Bylund et al. “Probing top quark neutral couplings in the Standard Model Effective Field Theory at NLO in QCD”. In: *Journal of High Energy Physics* 2016.5 (May 2016), p. 52. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2016)052.
- [43] B. Grzadkowski et al. “Dimension-six terms in the Standard Model Lagrangian”. In: *Journal of High Energy Physics* 2010.10 (Oct. 2010), p. 85. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2010)085.
- [44] A. Kulesza et al. “Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy”. In: *Eur. Phys. J. C* 79 (2019), p. 249. DOI: 10.1140/epjc/s10052-019-6746-z.
- [45] The ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/s08002.
- [46] The ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/s08003.
- [47] The CMS Collaboration. “The CMS Experiment at the CERN LHC”. In: *JINST* 3 (2008). DOI: 10.1088/1748-0221/3/08/S08004.

- [48] The LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008). DOI: 10.1088/1748-0221/3/08/s08005.
- [49] L. Evans and P. Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.
- [50] M. Benedikt et al. “LHC Design Report. 3. The LHC injector chain”. 2004.
- [51] The CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-LUM-18-002. Geneva: CERN, 2019. URL: <https://cds.cern.ch/record/2676164>.
- [52] E. Mobs. “The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019”. General Photo. July 2019. URL: <https://cds.cern.ch/record/2684277>.
- [53] S. R. Davis. “Interactive Slice of the CMS detector”. Aug. 2016. URL: <https://cds.cern.ch/record/2205172>.
- [54] R. Gluckstern. “Uncertainties in track momentum and direction, due to multiple scattering and measurement errors”. In: *Nuclear Instruments and Methods* 24 (1963), pp. 381–389. ISSN: 0029-554X. DOI: [https://doi.org/10.1016/0029-554X\(63\)90347-1](https://doi.org/10.1016/0029-554X(63)90347-1).
- [55] The CMS Collaboration. “The CMS trigger system”. In: *JINST* 12 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. eprint: 1609.02366 (physics.ins-det).
- [56] The CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009–P10009. DOI: 10.1088/1748-0221/9/10/p10009.
- [57] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nucl. Instrum. Methods Phys. Res., A* 262 (June 1987), 444. 19 p. URL: <https://cds.cern.ch/record/178627>.
- [58] A. M. Sirunyan et al. “Particle-flow reconstruction and global event description with the CMS detector”. In: *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].
- [59] M. Cacciari, G. P. Salam, and G. Soyez. “The anti- k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063.
- [60] M. Cacciari, G. P. Salam, and G. Soyez. “FastJet user manual”. In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2.
- [61] “DeepJet: Generic physics object based jet multiclass classification for LHC experiments”. In: *Deep Learning for Physical Sciences Workshop at the 31st Conference on Neural Information Processing Systems (NIPS)* (2017).
- [62] “Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector”. Nov. 2018. URL: <https://cds.cern.ch/record/2646773>.
- [63] *Hadron based origin identification of heavy flavour jets at generator level*. TWiki: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/GenHFHadronMatcher>.
- [64] The CMS Collaboration. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. In: *JINST* 12 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014.
- [65] *Jet Identification for the 13 TeV data Run2018*. TWiki: https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2018_r10.

- [66] *Jet identification in high pile-up environment*. TWiki: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJetID> r42.
- [67] The CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 13.06 (June 2018), P06015–P06015. DOI: 10.1088/1748-0221/13/06/p06015.
- [68] The CMS Collaboration. “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”. In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005.
- [69] *Baseline muon selections for Run-II*. Twiki: <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2> r54.
- [70] *Cut Based Electron ID for Run-II*. Twiki: <https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2> r60.
- [71] S. Agostinelli et al. “GEANT4: A Simulation toolkit”. In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.
- [72] “Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector”. June 2018. URL: <https://cds.cern.ch/record/2627468>.
- [73] A. M. Sirunyan et al. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011.
- [74] P. Nason. “A new method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040.
- [75] S. Frixione, P. Nason, and C. Oleari. “Matching QCD computations with parton shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070.
- [76] S. Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043.
- [77] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079.
- [78] T. Sjostrand et al. “An Introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024.
- [79] A. M. Sirunyan et al. “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”. Submitted to EPJC. 2019.
- [80] Physics Data And Monte Carlo Validation Group. *PPD RunII Analysis Guideline*. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PdmVAnalysisSummaryTable> r13.
- [81] *NNLO+NNLL top-quark-pair cross sections: ATLAS-CMS recommended predictions for top-quark-pair cross sections using the Top++v2.0 program*. TWiki: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/TtbarNNLO> r16.
- [82] S. Alioli et al. “NLO single-top production matched with shower in POWHEG: s - and t -channel contributions”. In: *JHEP* 09 (2009), p. 111. DOI: 10.1088/1126-6708/2009/09/111.
- [83] D. Morse, S. Cooper, and K. Long. *Corrected Cross Sections for Binned V +jet Samples*. Oct. 2017. URL: <https://indico.cern.ch/event/673253/#2-corrected-cross-sections-for>.

- [84] *Standard Model Cross Sections for CMS at 13 TeV*. TWiki: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeV> r26.
- [85] The CMS Collaboration. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. Oct. 25, 2016.
- [86] J. van der Linden. *MVA-Studien mit binären neuronalen Netzen in der Suche nach $t\bar{t}H$ -Produktion am CMS-Experiment*. Bachelor Thesis, KIT, Aug. 2017.
- [87] F. Rosenblatt. “The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain”. In: *Psychological Review* (1958), pp. 65–386.
- [88] *Artificial Neural Network*. Wiki: https://commons.wikimedia.org/wiki/Artificial_neural_network. Accessed: 08/19/2019.
- [89] S. Solla, E. Levin, and M. Fleisher. “Accelerated learning in layered neural networks”. In: *Complex Systems* 2 (Dec. 1988).
- [90] E. B. Baum and F. Wilczek. “Supervised Learning of Probability Distributions by Neural Networks”. In: *Neural Information Processing Systems*. American Institute of Physics, 1988, pp. 52–61. URL: <http://papers.nips.cc/paper/3-supervised-learning-of-probability-distributions-by-neural-networks.pdf>.
- [91] J. Duchi, E. Hazan, and Y. Singer. “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”. In: *J. Mach. Learn. Res.* 12 (July 2011), pp. 2121–2159. ISSN: 1532-4435. URL: <http://dl.acm.org/citation.cfm?id=1953048.2021068>.
- [92] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. “Learning Representations by Back-propagating Errors”. In: *Neurocomputing: Foundations of Research*. Cambridge, MA, USA: MIT Press, 1988, pp. 696–699. ISBN: 0-262-01097-6. URL: <http://dl.acm.org/citation.cfm?id=65669.104451>.
- [93] M. A. Nielsen. *Neural Networks and Deep Learning*. 2015. URL: <http://www.neuralnetworksanddeeplearning.com>.
- [94] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. URL: <http://www.deeplearningbook.org>.
- [95] J. van der Linden et al. *DRACO-MLfoy*. <https://github.com/kit-cn-cms/DRACO-MLfoy>. 2018.
- [96] F. Chollet et al. *Keras*. <https://keras.io>. 2015.
- [97] The CMS Collaboration. “Measurement of $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel in 41.5 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ ”. 2019. URL: <https://cds.cern.ch/record/2675023>.
- [98] A. Y. Ng. “Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance”. In: *Proceedings of the Twenty-first International Conference on Machine Learning*. ICML ’04. Banff, Alberta, Canada: ACM, 2004. ISBN: 1-58113-838-5. DOI: 10.1145/1015330.1015435.
- [99] N. Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15 (2014), pp. 1929–1958. URL: <http://jmlr.org/papers/v15/srivastava14a.html>.
- [100] G. Cowan. *Statistical data analysis*. Oxford University Press, USA, 1998.
- [101] The LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Aug. 2011. URL: <https://cds.cern.ch/record/1379837>.

- [102] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-013-2501-z.
- [103] The CMS Collaboration. *CMS Combine Tool*. <http://cms-analysis.github.io/HiggsAnalysis-CombinedLimit>. Accessed: 08/19/2019.
- [104] J. Neyman, P. E. S., and P. K. “On the problem of the most efficient tests of statistical hypotheses.” In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933). DOI: 10.1098/rsta.1933.0009.
- [105] *JEC Uncertainty Sources*. TWiki: <https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources> r51.
- [106] The CMS Collaboration. “Identification of b-Quark Jets with the CMS Experiment”. In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013.
- [107] *Utilities for Accessing Pileup Information for Data: Recommended cross section*. TWiki: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData> r36.
- [108] P. Skands, S. Carrazza, and J. Rojo. “Tuning PYTHIA 8.1: the Monash 2013 Tune”. In: *Eur. Phys. J. C* 74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].
- [109] The CMS Collaboration. “Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of High Energy Physics* 2019.3 (Mar. 2019), p. 26. ISSN: 1029-8479. DOI: 10.1007/JHEP03(2019)026.
- [110] The CMS Collaboration. “Observation of $t\bar{t}H$ production”. In: *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801.
- [111] R. Barlow and C. Beeston. “Fitting using finite Monte Carlo samples”. In: *Computer Physics Communications* 77.2 (1993), pp. 219–228. ISSN: 0010-4655. DOI: 10.1016/0010-4655(93)90005.
- [112] LHC Higgs Cross Section Working Group. *SM Higgs production cross sections at $\sqrt{s} = 13$ TeV*. TWiki: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV> r23.
- [113] The CMS Collaboration. “Measurement of the cross section for $t\bar{t}$ production with additional jets and b jets in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: 2019. URL: <https://cds.cern.ch/record/2684606>.
- [114] The CMS Collaboration. “Measurement of the $t\bar{t}b\bar{b}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13$ TeV”. In: 2019. URL: <https://cds.cern.ch/record/2675029>.
- [115] R. D. Cousins. “Generalization of chisquare goodness-of-fit test for binned data using saturated model, with application to histograms”. 2013. URL: http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.
- [116] A. Kolmogorov. “Sulla determinazione empirica di una legge di distribuzione”. In: *G. Ist. Ital. Attuari* 4 (1933), pp. 83–91.
- [117] N. Smirnov. “Table for Estimating the Goodness of Fit of Empirical Distributions”. In: *The Annals of Mathematical Statistics* 19.2 (June 1948), pp. 279–281. DOI: 10.1214/aoms/1177730256.
- [118] J. von den Driesch. *DNN-basierte Rekonstruktion des $t\bar{t}$ -Systems*. Bachelor Thesis, KIT, Oct. 2019.

- [119] M. Horzela. *Merging of tt and $tt+bb$ simulations for an improved background modeling for $ttH(bb)$ measurements*. Master Thesis, KIT, Sept. 2019.
- [120] G. C. Fox and S. Wolfram. “Observables for the Analysis of Event Shapes in e^+e^- Annihilation and Other Processes”. In: *Phys. Rev. Lett.* 41 (23 Dec. 1978), pp. 1581–1585. DOI: 10.1103/PhysRevLett.41.1581.
- [121] C. Bernaciak et al. “Fox-Wolfram Moments in Higgs Physics”. In: *Phys. Rev.* D87 (2013), p. 073014. DOI: 10.1103/PhysRevD.87.073014.

Appendix

A χ^2 -based Event Reconstruction

A.1 Mass Templates

For the χ^2 -based event reconstruction presented in section 6.3, mass templates of the particles to be reconstructed need to be derived. The χ^2 -value of a hypothesized particle is calculated via eq. 6.4, where $m_i^{\text{pred.}}$ is the mass of the reconstructed particle in question. The values $m_i^{\text{exp.}}$ and $\sigma_i^{\text{exp.}}$ are derived via ΔR -based matching of the generator-level particles with the reconstructed jets. For simulated events, the four-vectors of the simulated particles and their decay products are known. By searching for reconstructed jets which point in the same direction as these generator-level particles, i.e. have the same η and ϕ values, reconstructed jets can be assumed to be assigned to the correct particles. A jet candidate is considered as matched to the generator-level particle if the distance ΔR in the (η, ϕ) -plane is smaller than 0.1 with

$$\Delta R = \sqrt{(\eta_{\text{jet}} - \eta_{\text{part.}})^2 + (\phi_{\text{jet}} - \phi_{\text{part.}})^2}. \quad (9.1)$$

This assignment is based on generator-level information and can therefore not be used for the actual event reconstruction. However, the distributions of the masses $m_i^{\text{exp.}}$ of particles reconstructed matching these conditions can be treated as expected masses in the actual reconstruction. The standard deviations $\sigma_i^{\text{exp.}}$ of these distributions are used to assign relative weights to the observed mass differences $m_i^{\text{pred.}} - m_i^{\text{exp.}}$.

The derived templates for the top quark and W boson masses are shown in Figure A.1, while the templates for the Z boson mass is shown in Figure A.2. The values $m_i^{\text{exp.}}$ and $\sigma_i^{\text{exp.}}$ are summarized in Table A.1. Due to the finite energy and momentum resolution of particles in the detector and the possibility of some decay products originating from the particles to be reconstructed being out of acceptance the reconstructed masses do not necessarily match the particle masses expected from Standard Model measurements exactly.

The reconstruction quality is summarized in Table A.2 where the efficiencies for matching certain objects (i.e. top quarks or the Z boson) are summarized. In Figures

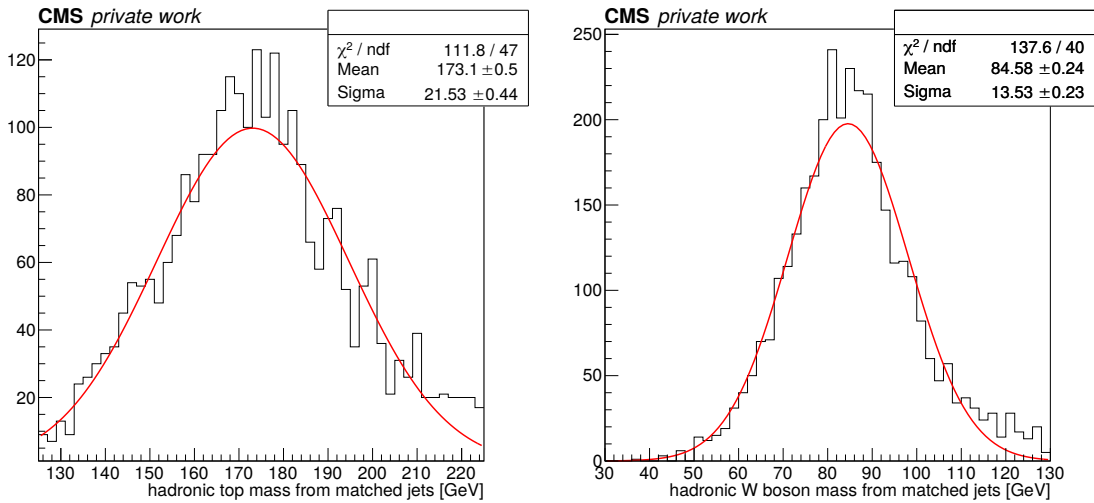


Figure A.1: Templates for the top quark (left) and hadronic W boson (left) masses.

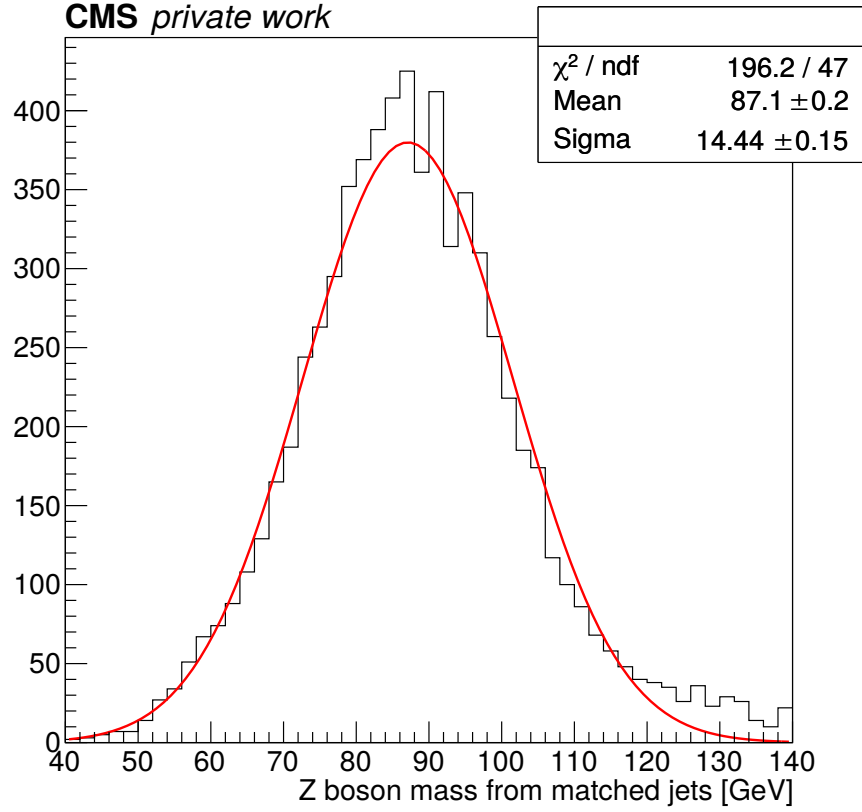


Figure A.2: Template for the Z boson mass.

Table A.1: Expected masses derived for the χ^2 -based event reconstruction.

	$m^{\text{exp.}}$	$\sigma^{\text{exp.}}$
t_{had}	$(173.1 \pm 0.5) \text{ GeV}$	$(21.5 \pm 0.4) \text{ GeV}$
W_{had}	$(84.6 \pm 0.2) \text{ GeV}$	$(13.5 \pm 0.2) \text{ GeV}$
Z	$(87.1 \pm 0.2) \text{ GeV}$	$(14.4 \pm 0.1) \text{ GeV}$

Table A.2: **Quality of χ^2 -based event reconstruction.** A matching of generator level and reconstructed objects is performed. An object is said to be matched for $\Delta R \leq 0.4$. The entries for $t\bar{t}$ and $t\bar{t}+Z$ correspond to a simultaneous matching of all objects.

matched group	$t\bar{t}+Z$ events		$t\bar{t}$ events
	≥ 6 jets	≥ 4 jets	≥ 4 jets
Z boson	49.7%	—	—
t_{had}	33.4%	27.1%	33.8%
t_{lep}	38.6%	34.5%	41.0%
$t\bar{t}+Z$	12.9%	—	—
$t\bar{t}$	16.5%	13.0%	19.7%

B ANN Input Features

B.1 Input Features

In the following the input features used for the ANNs are described.

b-tagging related features

The b-tagging values show discriminating power mainly between $t\bar{t}+lf$, $t\bar{t}+c\bar{c}$ and minor background processes, and $t\bar{t}+Z$, $t\bar{t}+H$ and $t\bar{t}+b\bar{b}$ processes. Especially the third and fourth highest b-tag values in an event have high discriminating power as some background processes are not expected to have more than two b-tagged jets.

Also used are the number of b-tagged jets, the average b-tag value of all jets (and of b-tagged jets only), the minimal b-tag value in an event and the average deviation of b-tag values in an event. The latter one is calculated from the squared difference of the average b-tag value in an event and all b-tag values, normalized by the number of jets.

Angular differences

Several averaged angular differences are used. These describe kinematic features of the full event. The final feature selection includes the average $\Delta\eta$ of jets (and b-tagged jets only), and the average ΔR of jets (and b-tagged jets only).

Also included are the minimum $\Delta\eta$ distances of two jets (and b-tagged jets only) in an event. These features might be sensitive to kinematic features of di-jet systems, like the decay products of a Z boson or Higgs boson.

Mass and p_T related features

Many features in the final feature selection are mass or p_T related. Especially the mass related features show discriminating power between the signal process and most of the background processes.

With the invariant mass of two b-tagged jets closest to 91 GeV (and 125 GeV), possible Z boson (and Higgs boson) candidates are attempted to be found.

The average invariant mass of two b-tagged jets (and the average p_T of two b-tagged jets and the average mass of b-tagged jets) show discriminating power between $t\bar{t}+H$ and $t\bar{t}+Z$ events, and $t\bar{t}+\text{jets}$ events, as additional b-jet radiation in $t\bar{t}+\text{jets}$ events does usually not contain the radiation of massive particles.

Also used are the total mass in an event, the highest jet- p_T in an event and the invariant mass (and p_T) of two jets (and b-tagged jets) with the minimum ΔR distance, and missing H_T . The latter feature is the vectorial sum of transverse momenta of all jets and the charged lepton.

Features based on χ^2 -reconstruction

The χ^2 values of the complete reconstruction, the W boson reconstruction and the Z boson reconstruction are used. These values are expected to be small for a successful reconstruction.

Also used are the reconstructed top quark mass, the W boson mass, the Z boson mass and the transverse momentum of the Z boson.

Others

The b-tag likelihood ratio (blr) is a measure of whether an event has four or two real b-jets. Events with four b-jets are expected to have values close to one, events with two b-jets close to zero. This feature is derived from p_T , η and b-tag value dependent b-tagging accuracies of jets. The transformed b-tag likelihood ratio is defined as

$$\ln \left(\frac{\text{blr}}{1 - \text{blr}} \right). \quad (9.2)$$

The Fox-Wolfram moments describe angular correlations in an event via Legendre polynomials [120, 121]. Used in the final feature selection is the second Fox-Wolfram moment.

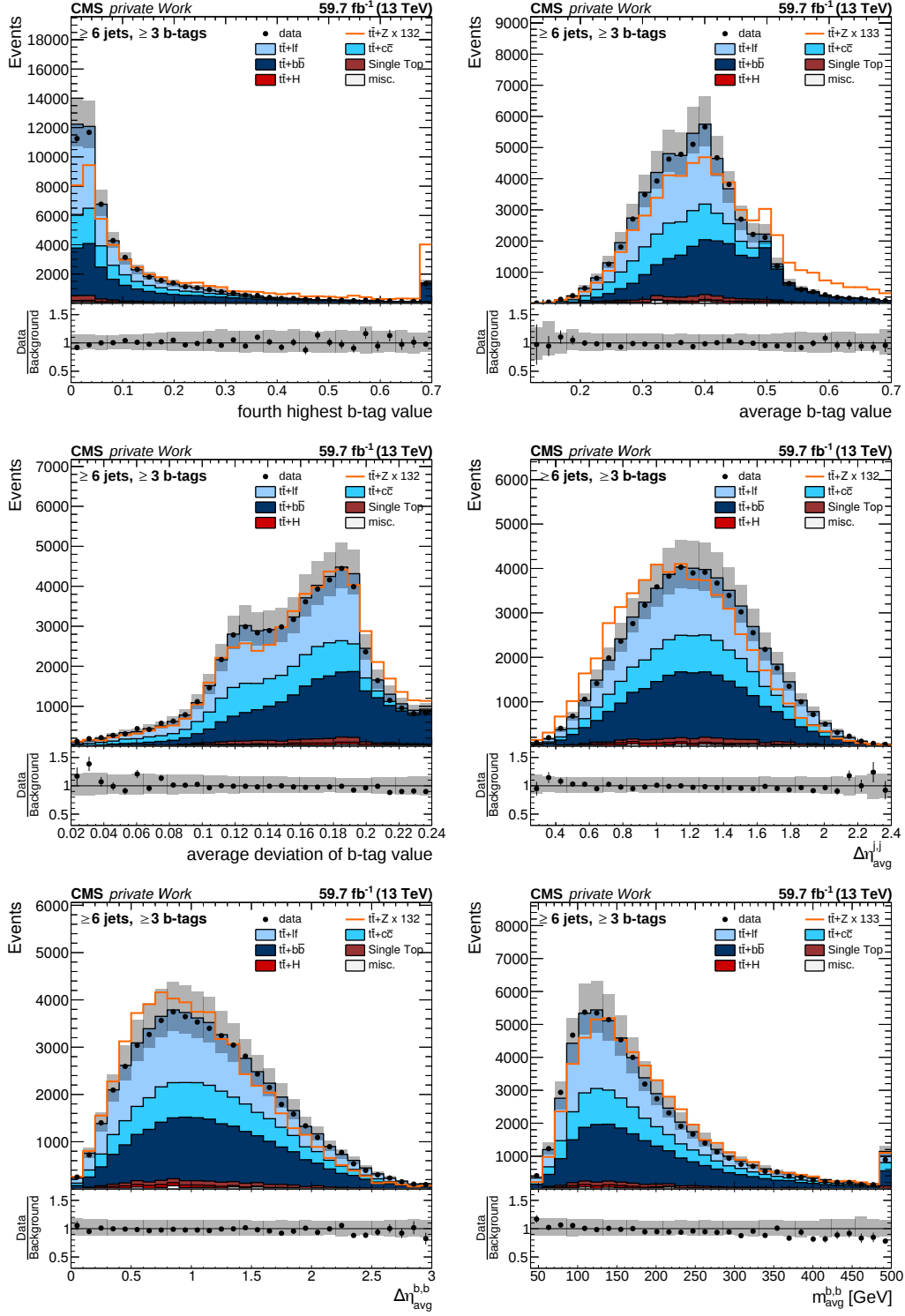


Figure B.3: **Final selection of input features for the ≥ 6 jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

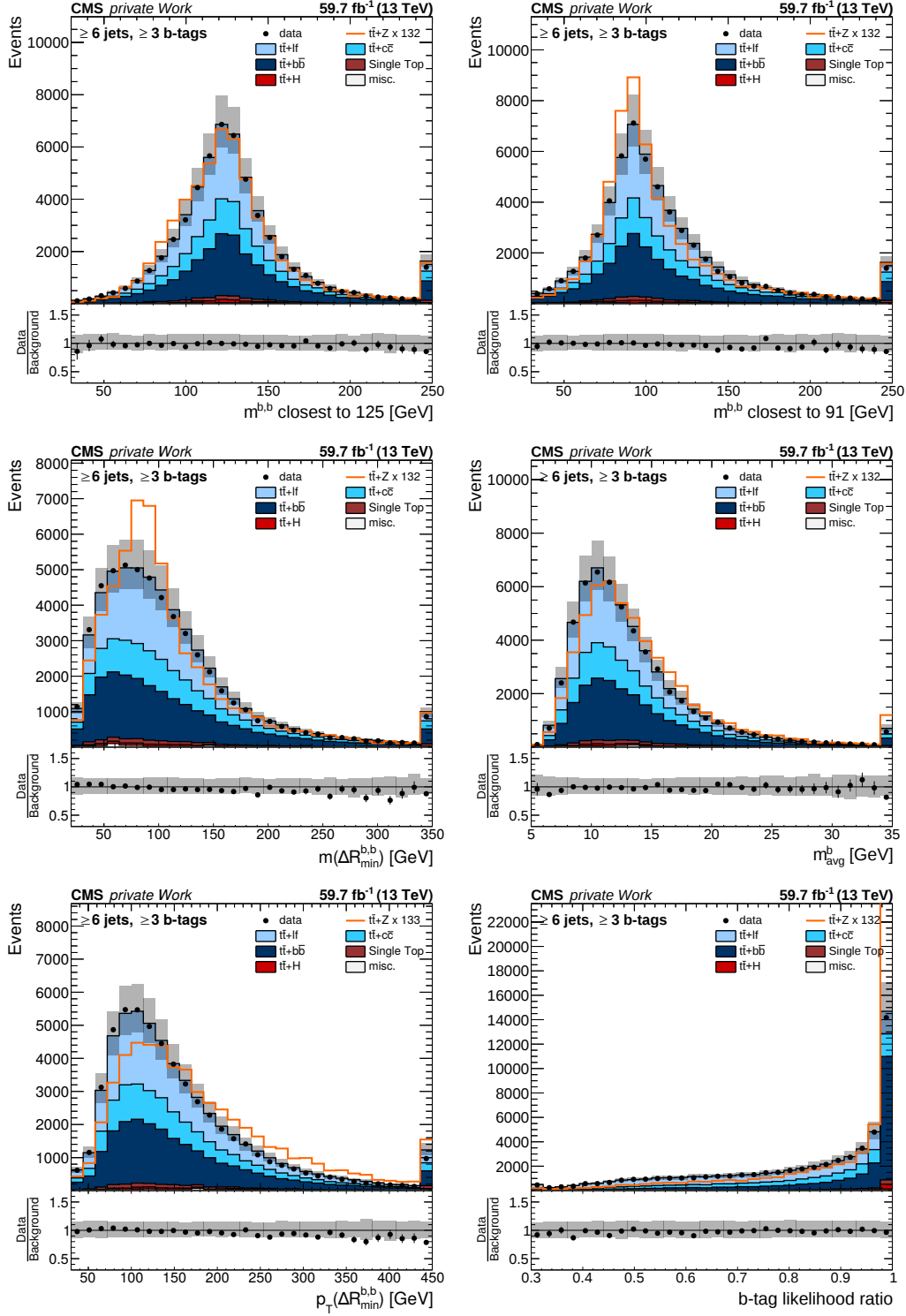


Figure B.4: **Final selection of input features for the ≥ 6 jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

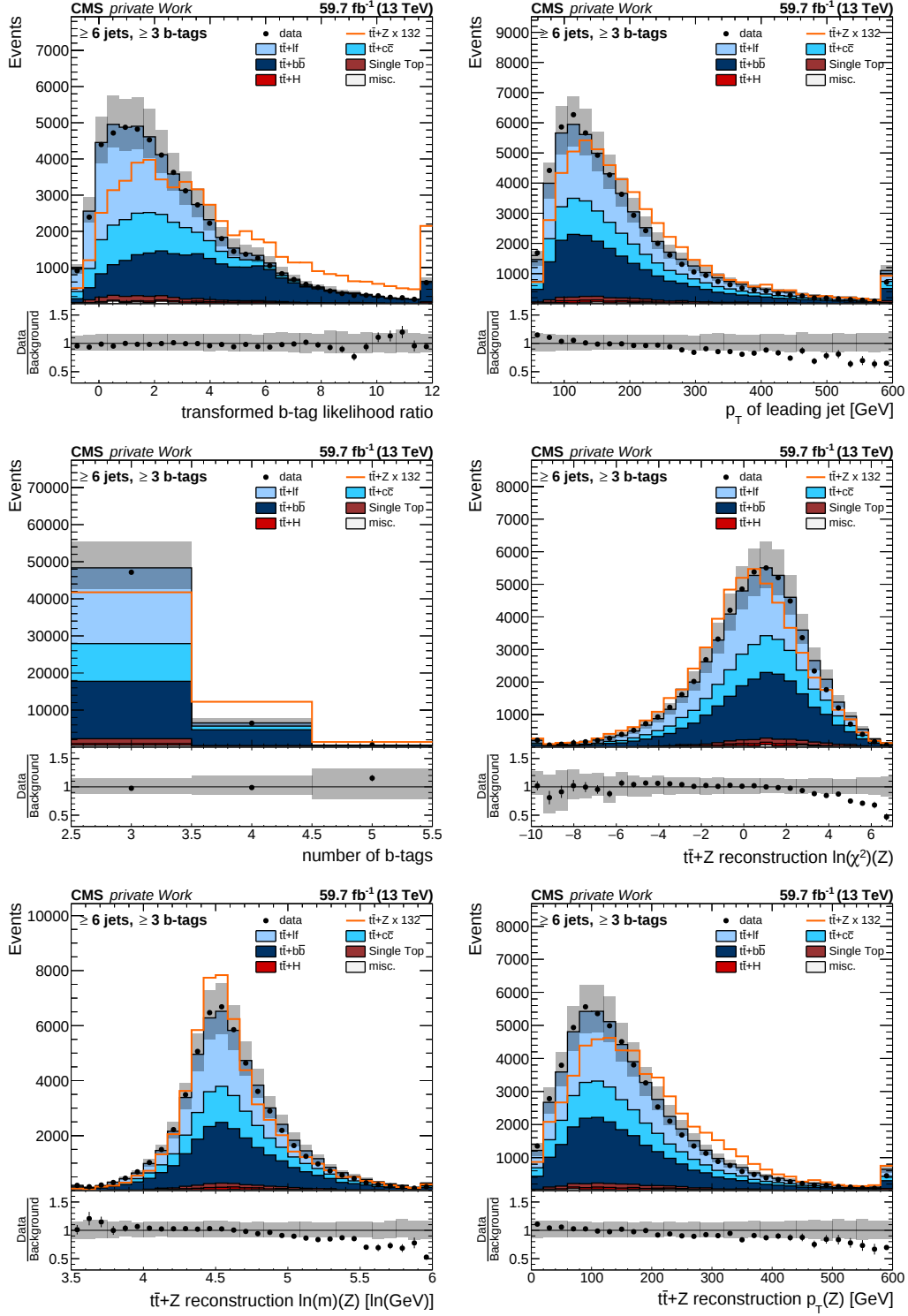


Figure B.5: **Final selection of input features for the ≥ 6 jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

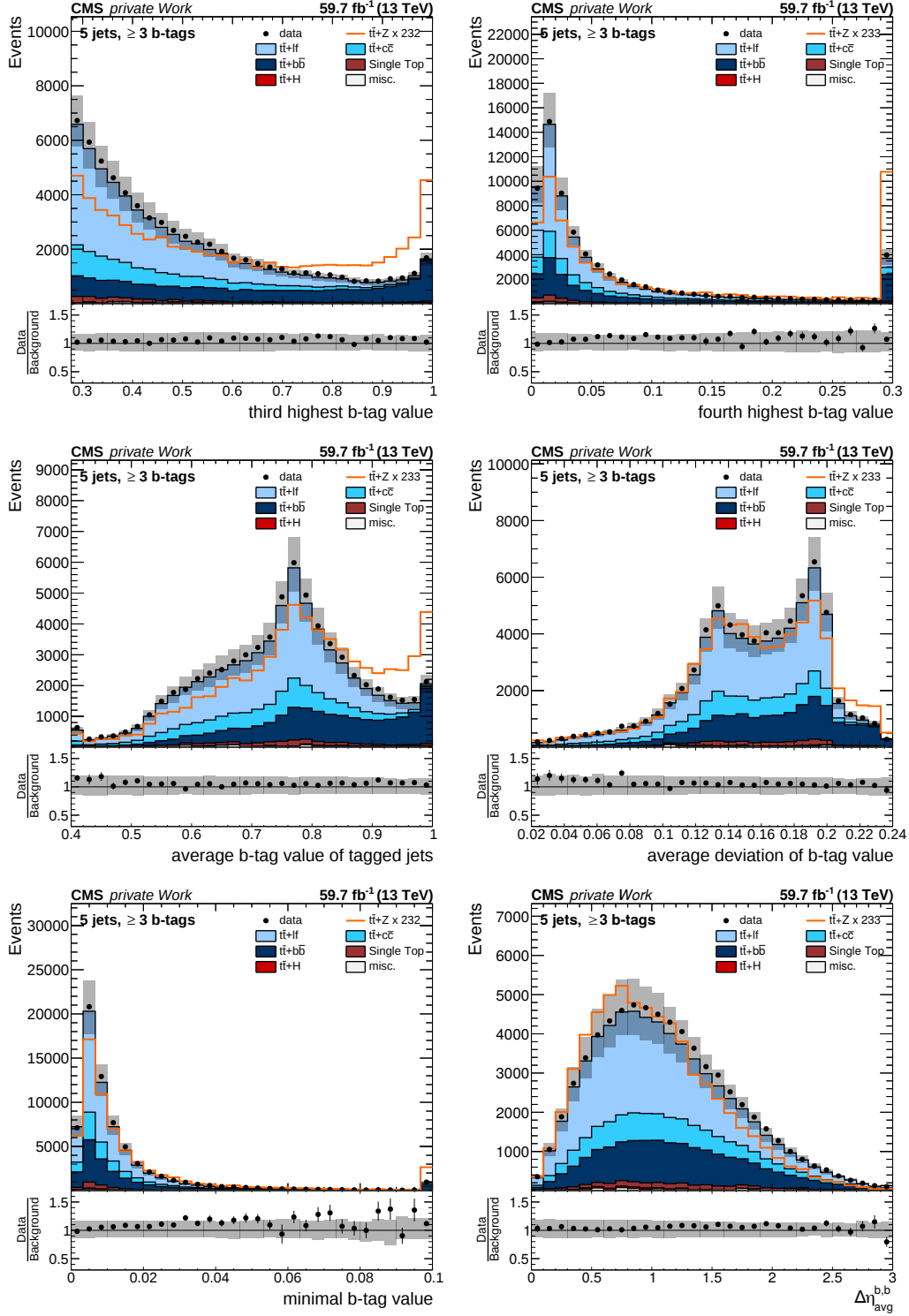


Figure B.6: **Final selection of input features for the five-jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

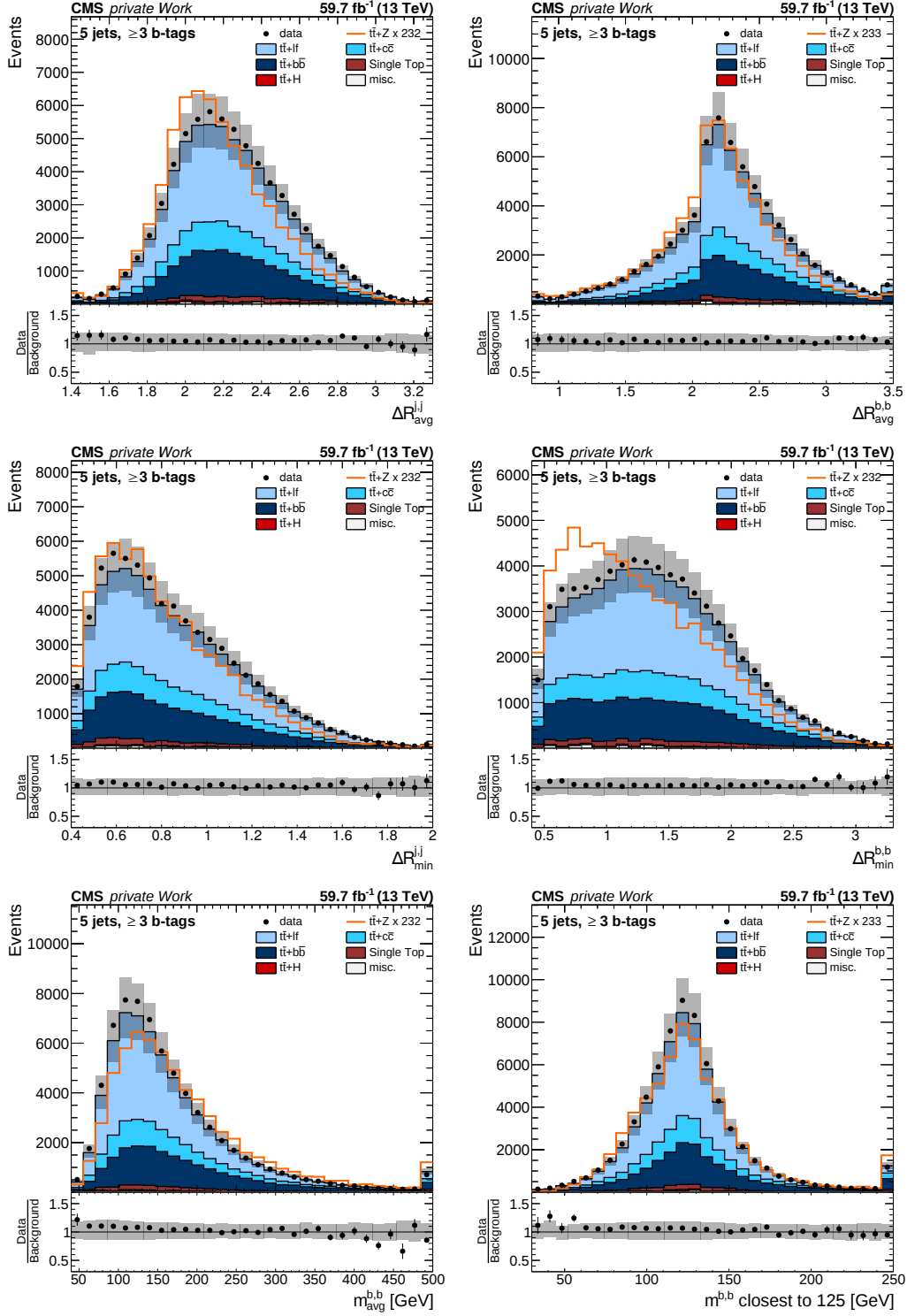


Figure B.7: **Final selection of input features for the five-jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

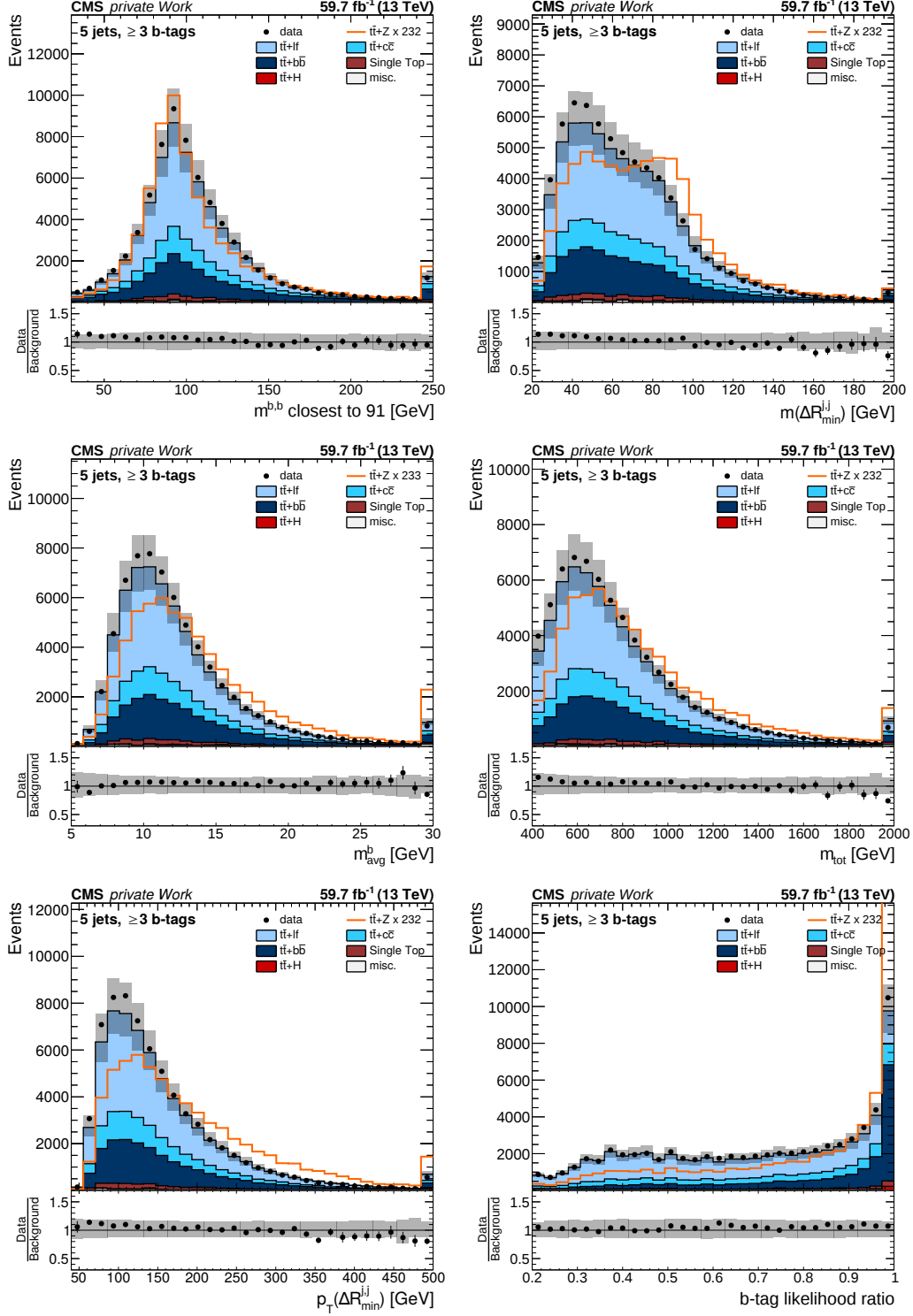


Figure B.8: **Final selection of input features for the five-jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

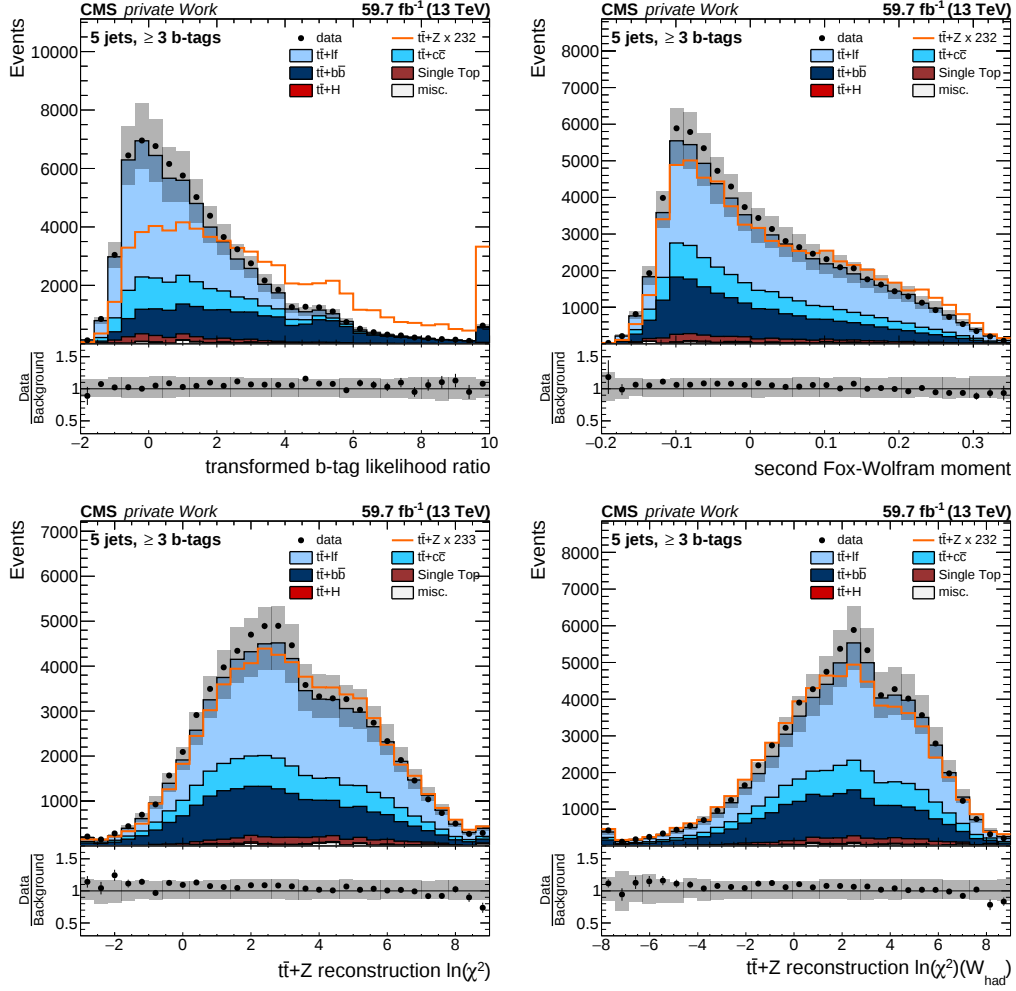


Figure B.9: **Final selection of input features for the five-jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

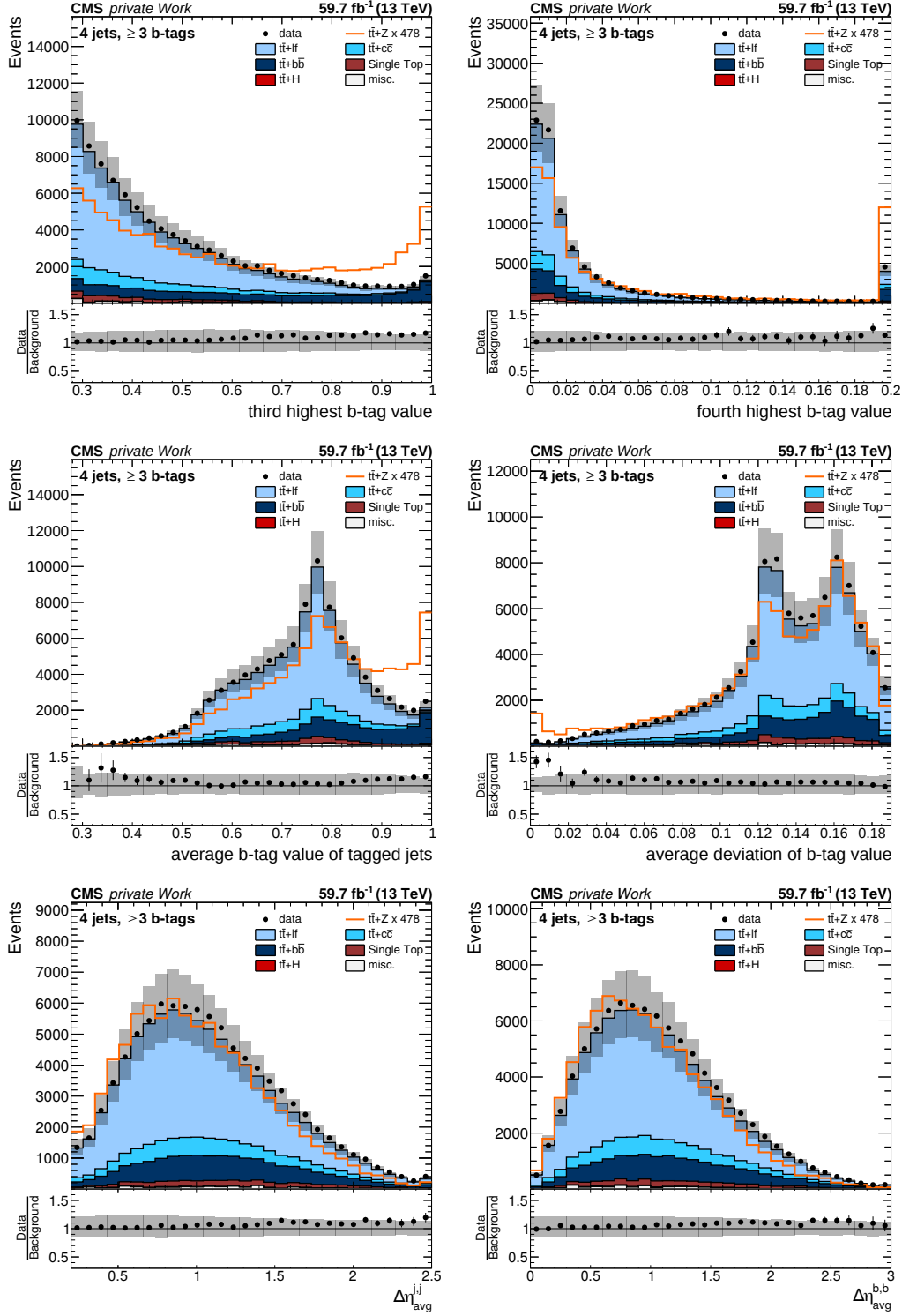


Figure B.10: **Final selection of input features for the four-jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

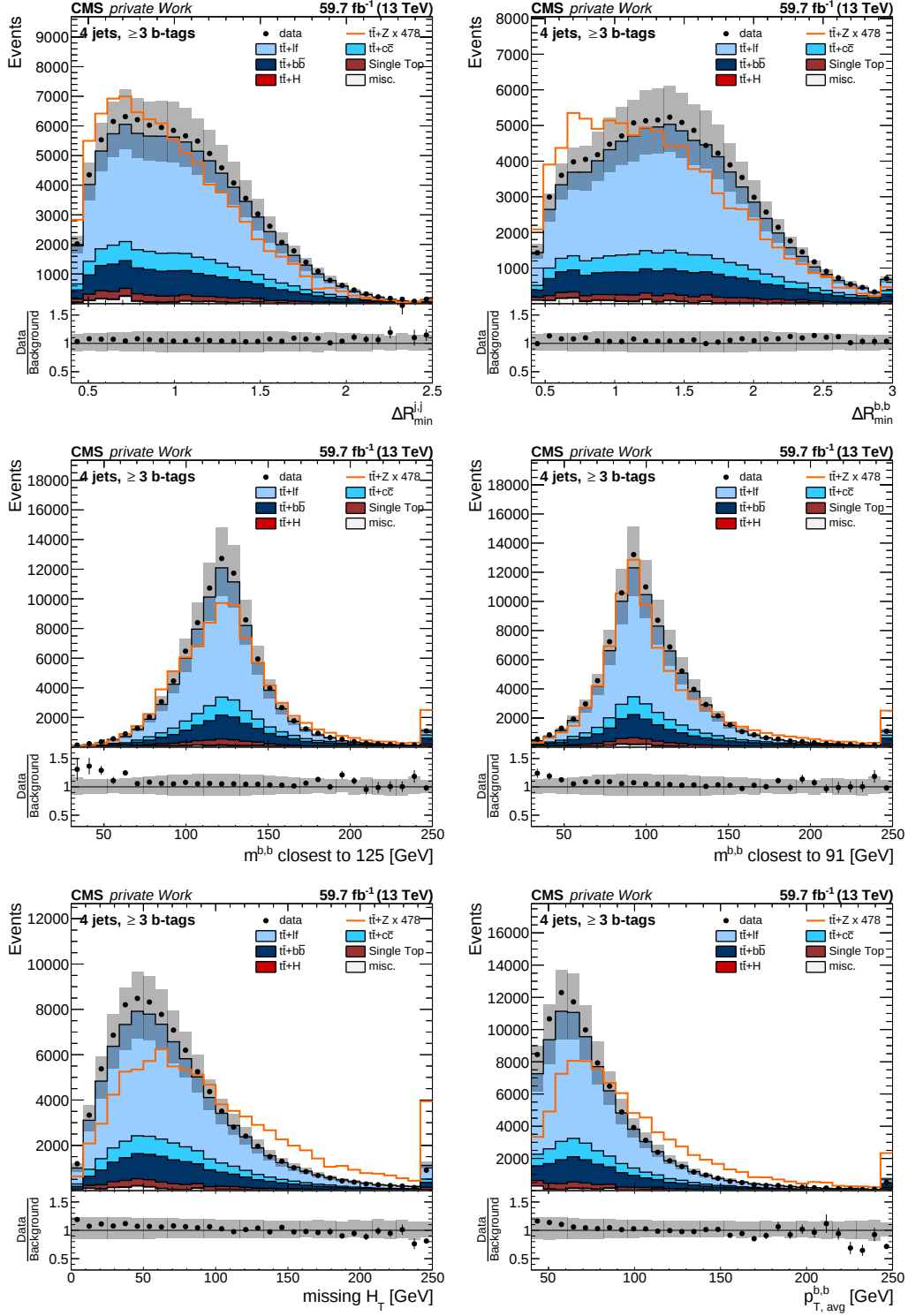


Figure B.11: **Final selection of input features for the four-jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

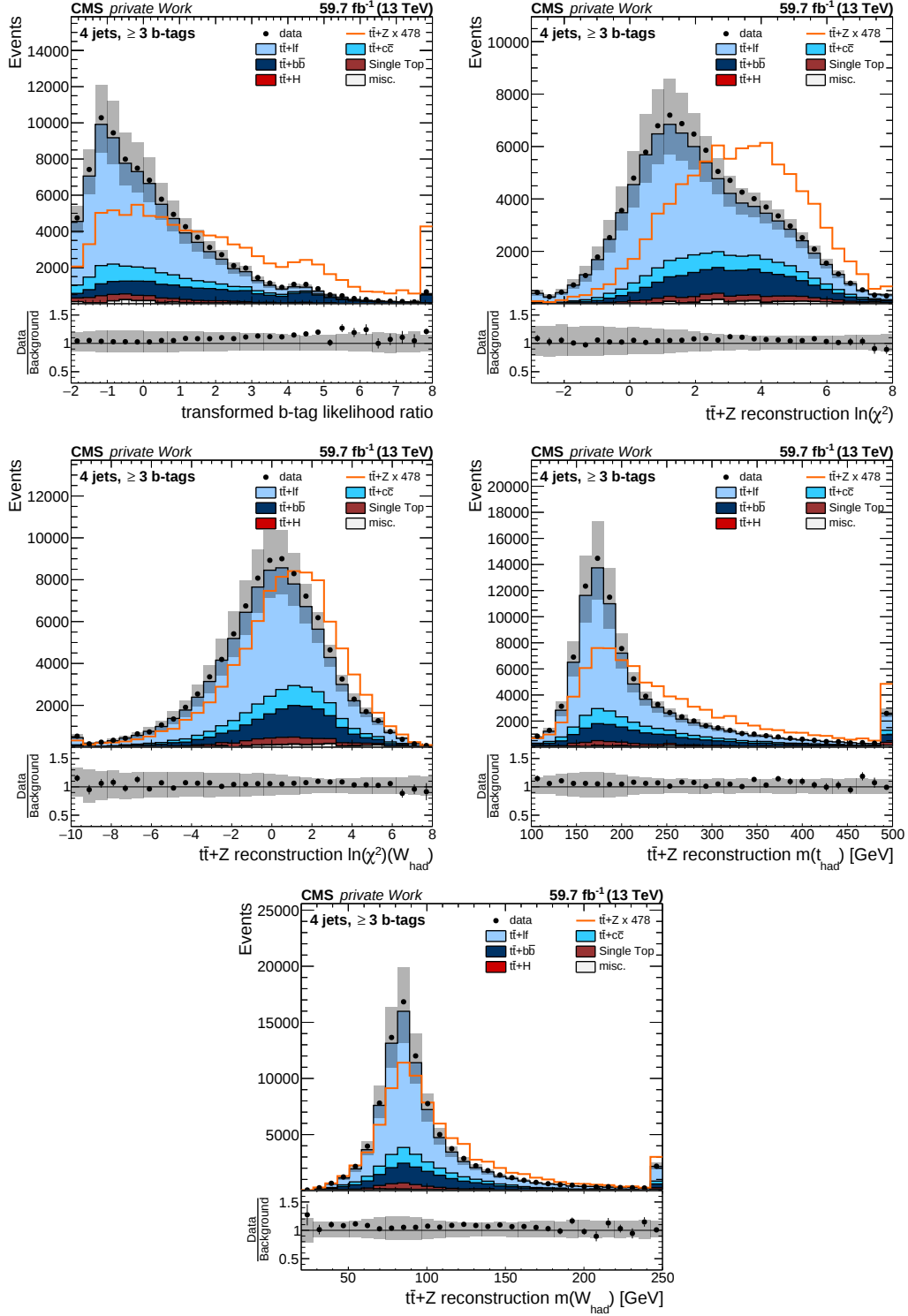


Figure B.12: **Final selection of input features for the four-jet region before the fit to data.** The predicted contributions of all background samples are stacked. The $t\bar{t}+Z$ contribution is overlaid as a line, scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom panel. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect. The last bin includes overflows.

B.2 Importance Ranking

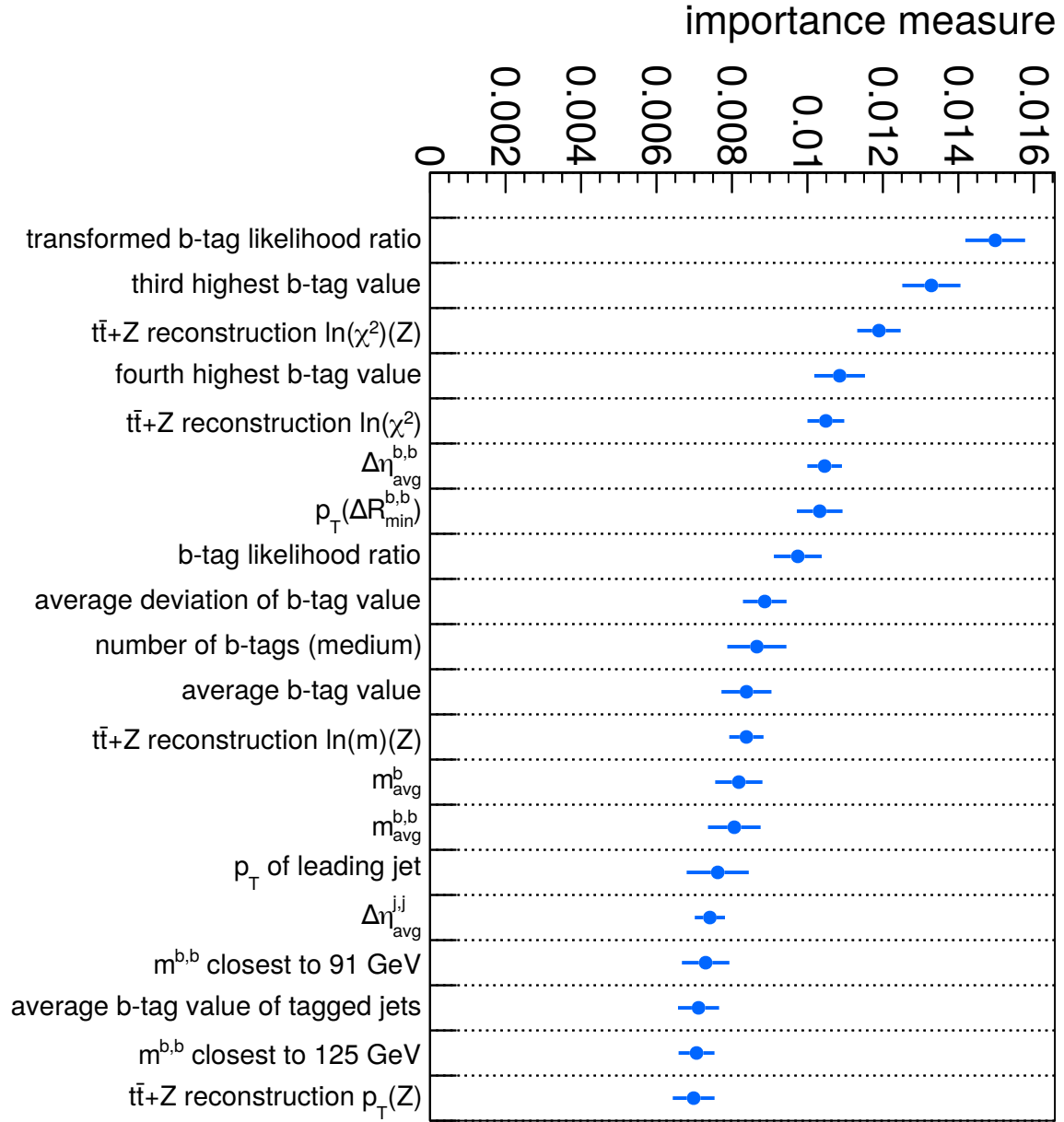


Figure B.13: **Ranking of input features for the multi-classification ANN in the ≥ 6 jet region.** The ANNs were trained 20 times, the mean values of $\aleph$ are shown with error bars corresponding to its standard deviation. Only the 20 highest ranked features are shown.

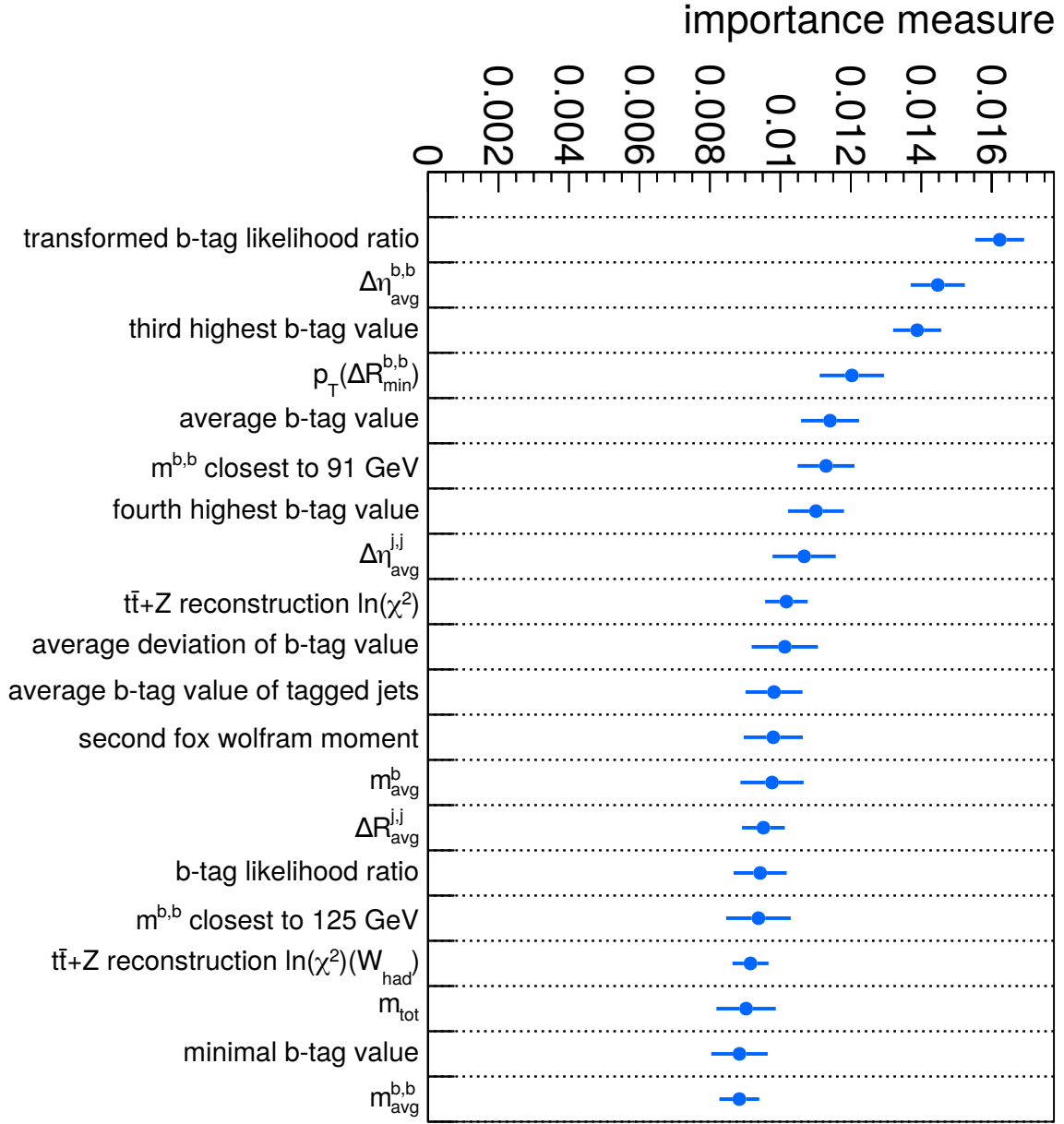


Figure B.14: **Ranking of input features for the multi-classification ANN in the five-jet region.** The ANNs were trained 20 times, the mean values of $\aleph$ are shown with error bars corresponding to its standard deviation. Only the 20 highest ranked features are shown.

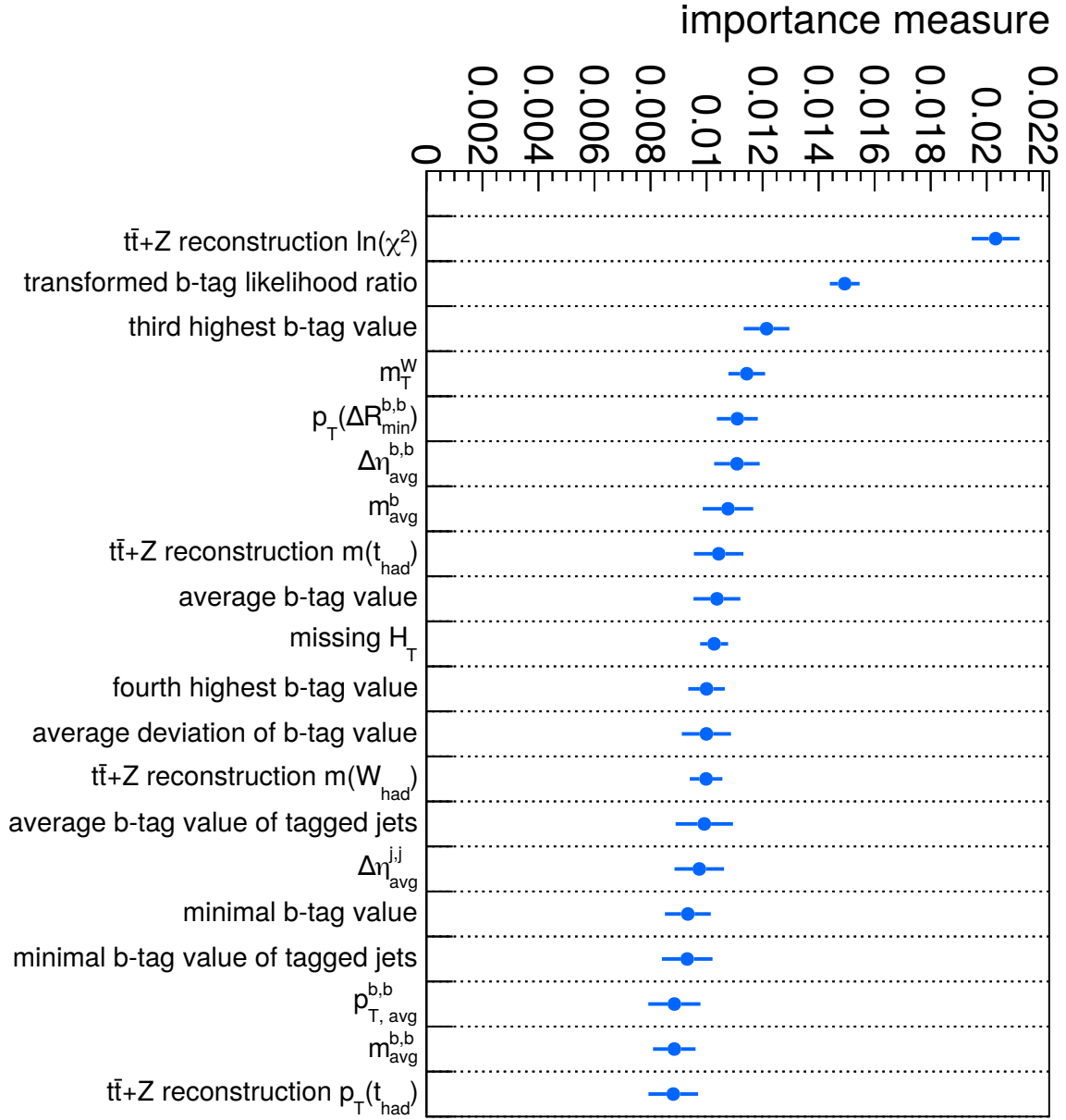


Figure B.15: **Ranking of input features for the multi-classification ANN in the four-jet region.** The ANNs were trained 20 times, the mean values of $\aleph$ are shown with error bars corresponding to its standard deviation. Only the 20 highest ranked features are shown.

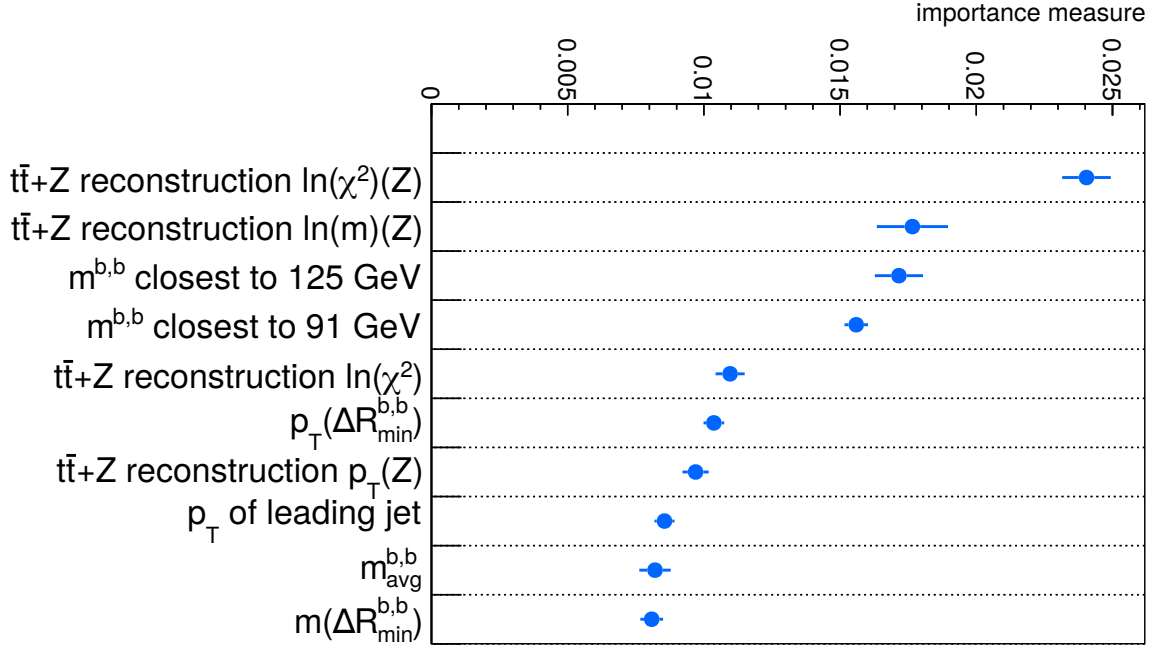


Figure B.16: **Ranking of input features for the binary-classification ANN in the ≥ 6 jet region.** The ANNs were trained 10 times, the mean values of $\aleph$ are shown with error bars corresponding to its standard deviation. Only the 10 highest ranked features are shown. The binary classification was performed between $t\bar{t}+H$ and $t\bar{t}+Z$ events.

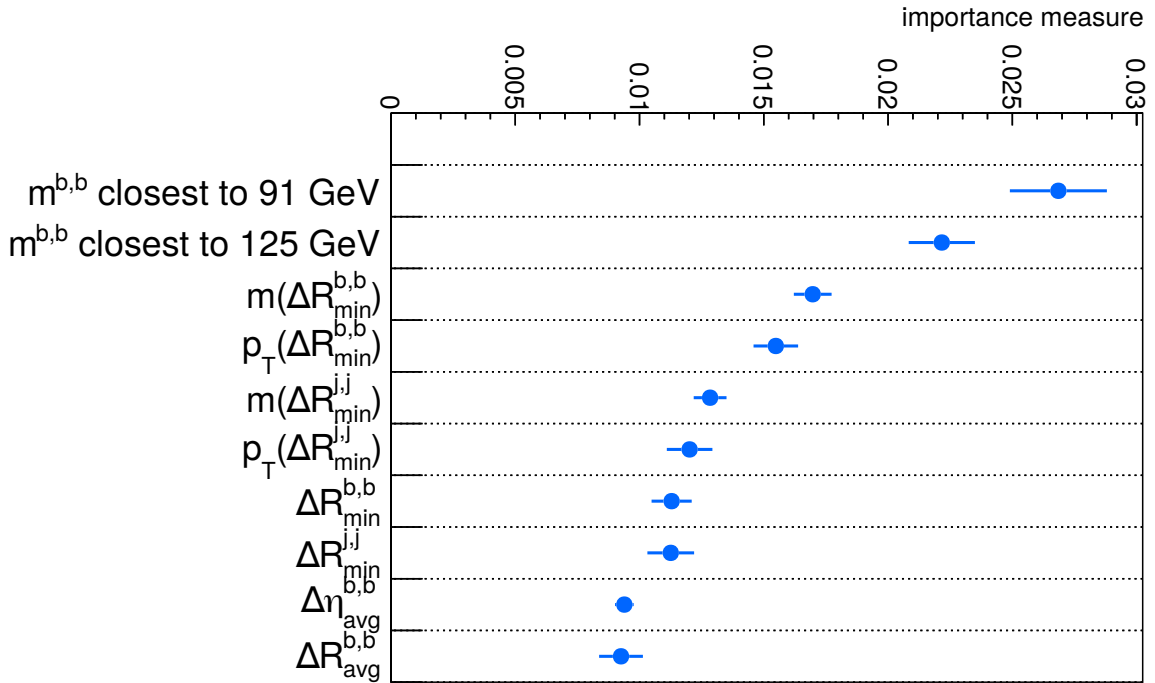


Figure B.17: **Ranking of input features for the binary-classification ANN in the five-jet region.** The ANNs were trained 10 times, the mean values of $\aleph$ are shown with error bars corresponding to its standard deviation. Only the 10 highest ranked features are shown. The binary classification was performed between $t\bar{t}+H$ and $t\bar{t}+Z$ events.

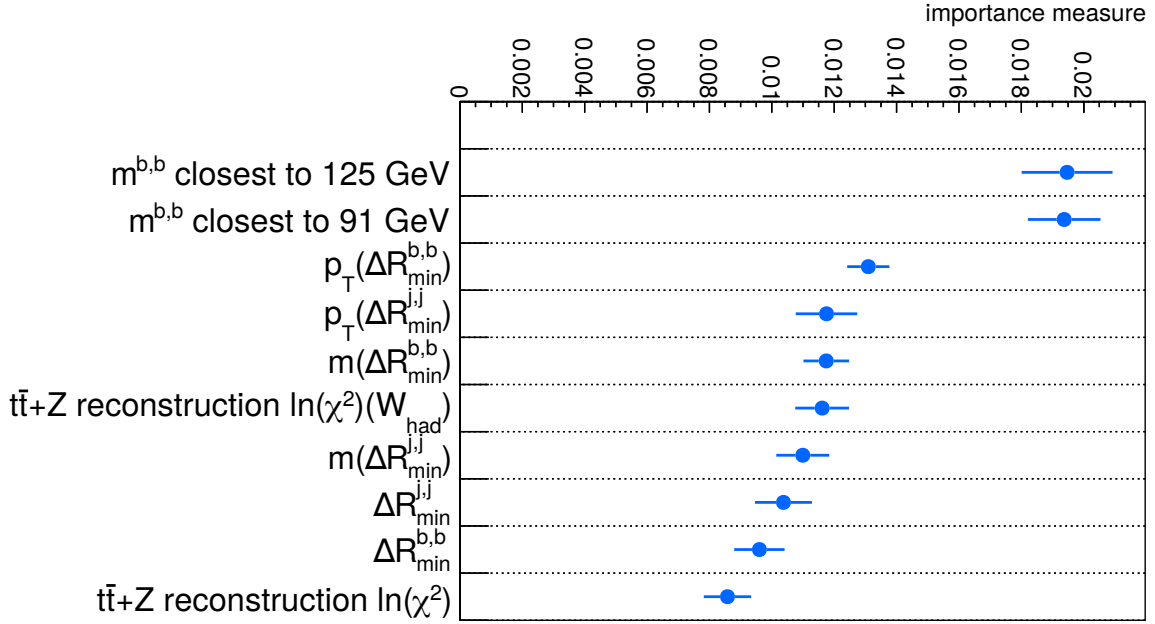


Figure B.18: **Ranking of input features for the binary-classification ANN in the four-jet region.** The ANNs were trained 10 times, the mean values of $\aleph$ are shown with error bars corresponding to its standard deviation. Only the 10 highest ranked features are shown. The binary classification was performed between $t\bar{t}+H$ and $t\bar{t}+Z$ events.

B.3 Goodness-Of-Fit Results

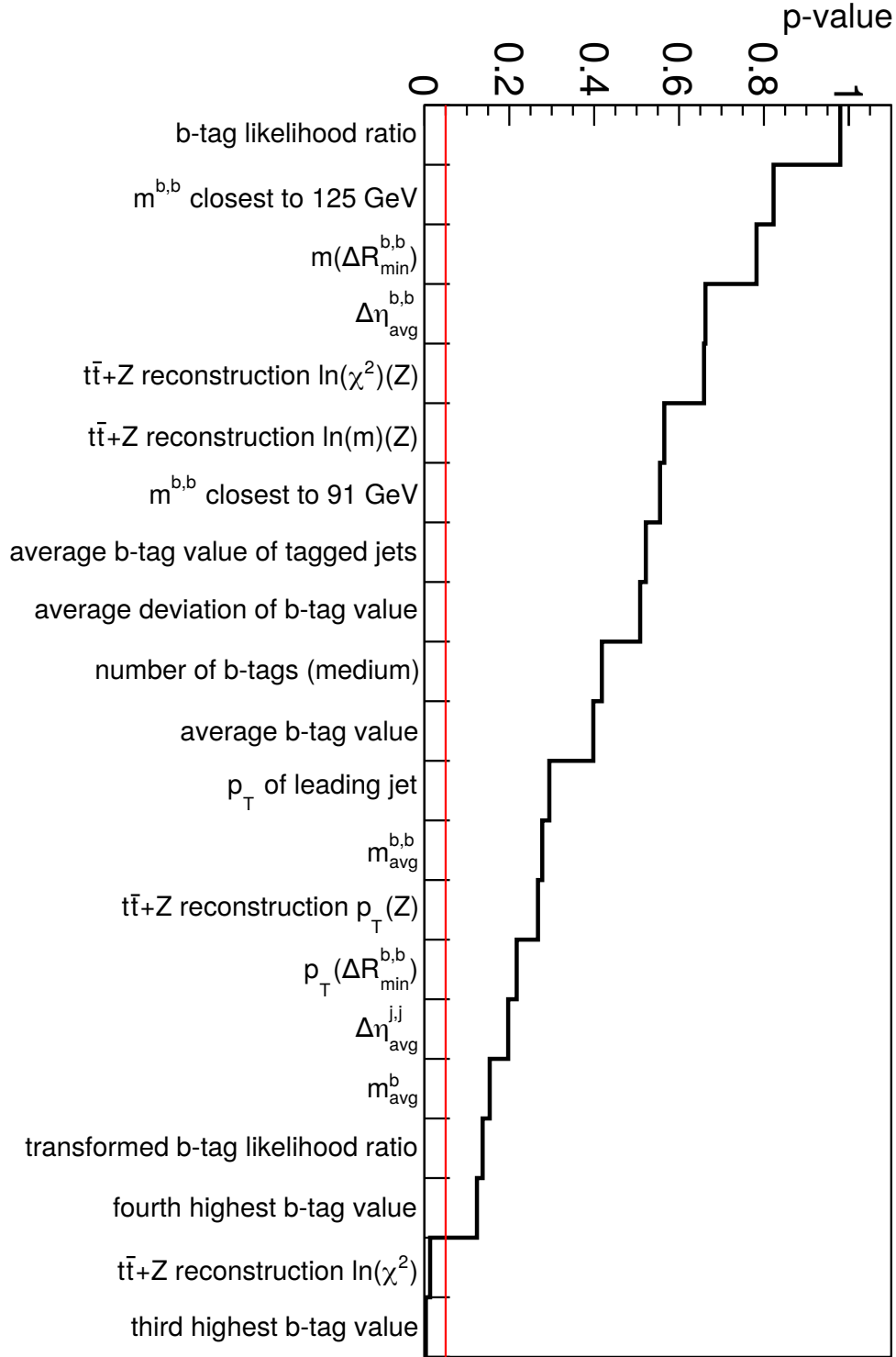


Figure B.19: **Goodness-of-Fit p -values for all input features in the ≥ 6 jet region.** The red line indicates a p -value of 0.05. All features with a p -value lower than 0.05 are discarded from the final feature selection.

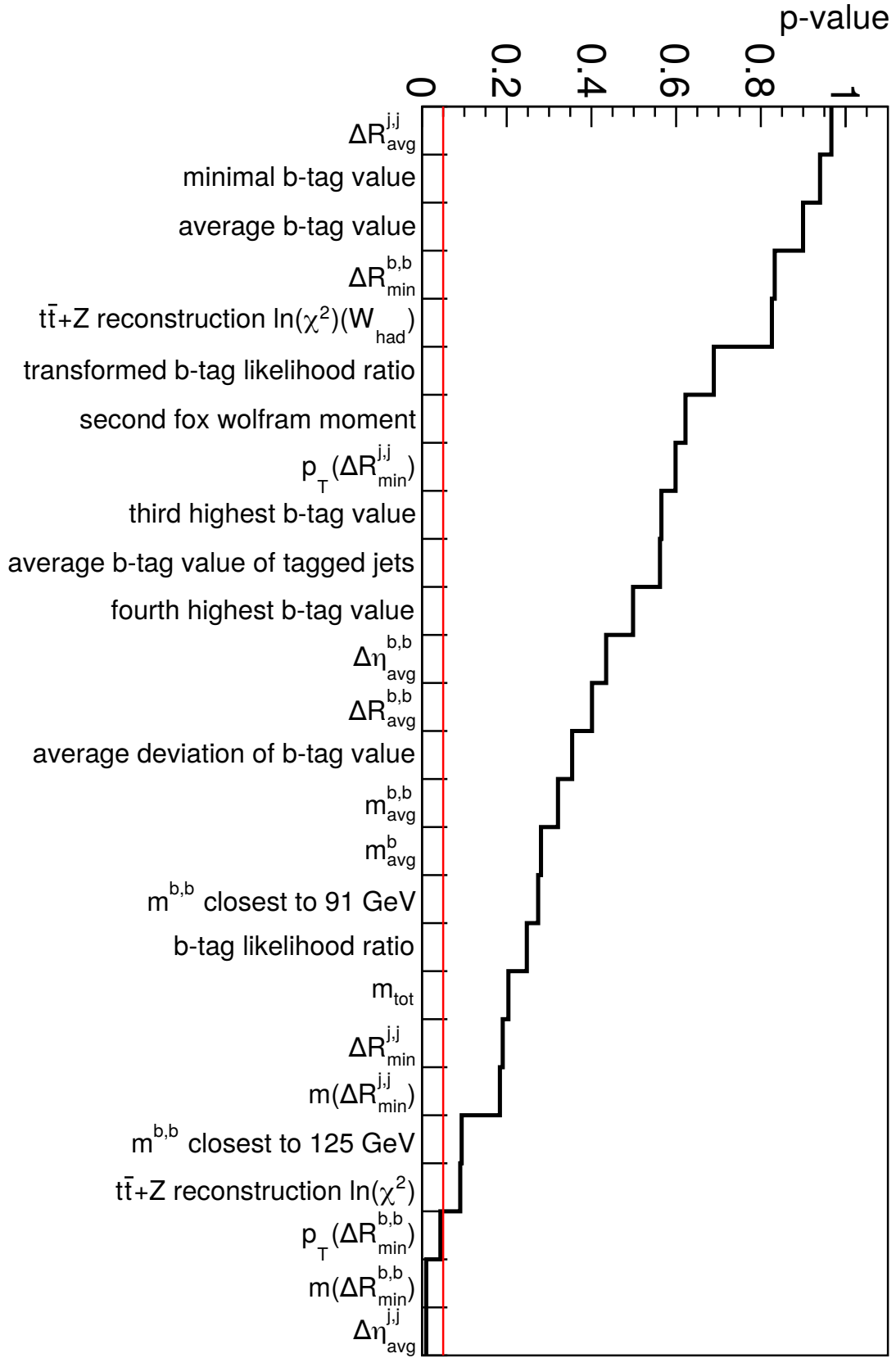


Figure B.20: **Goodness-of-Fit p -values for all input features in the five-jet region.** The red line indicates a p -value of 0.05. All features with a p -value lower than 0.05 are discarded from the final feature selection.

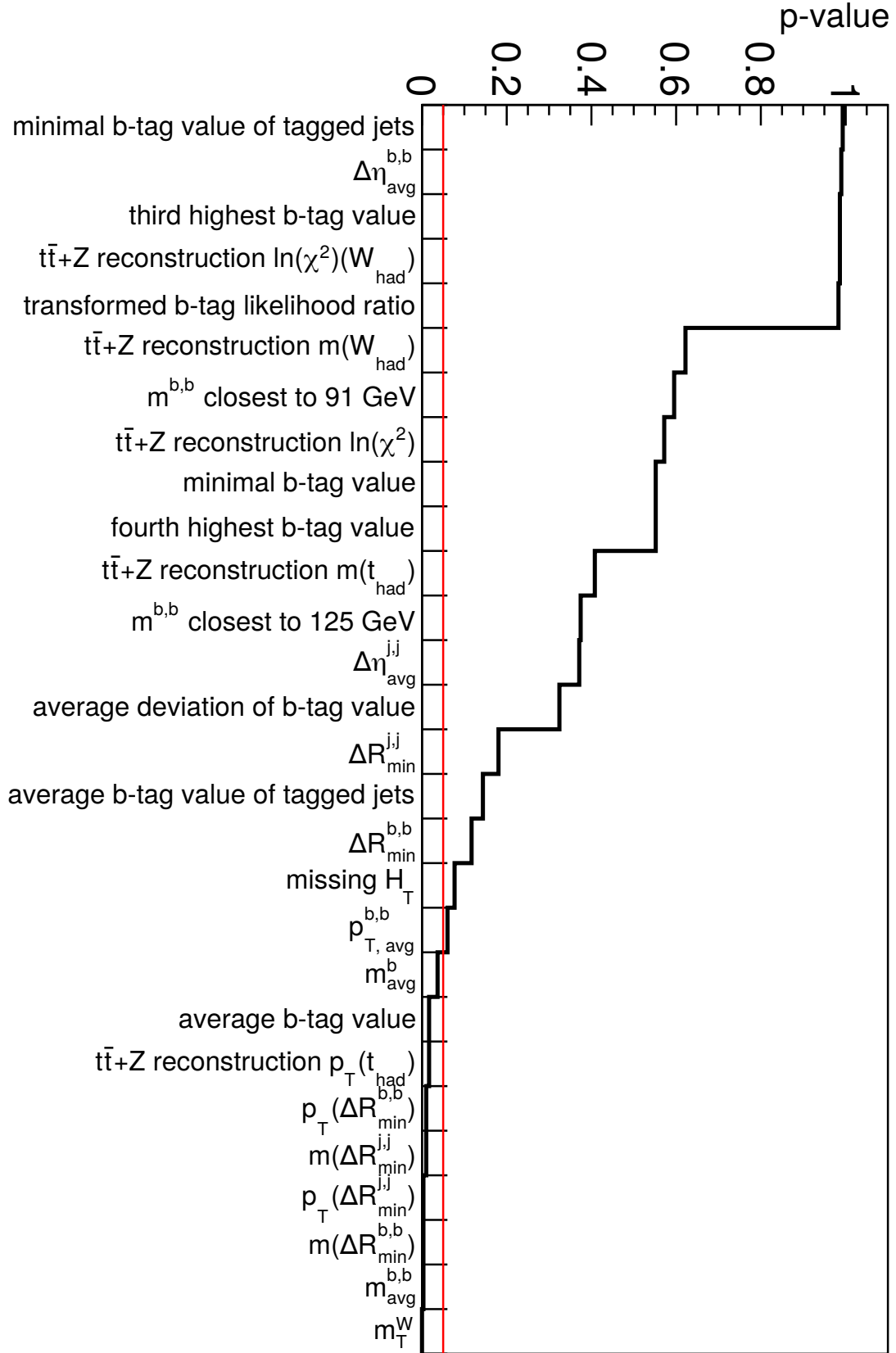


Figure B.21: **Goodness-of-Fit p -values for all input features in the four-jet region.** The red line indicates a p -value of 0.05. All features with a p -value lower than 0.05 are discarded from the final feature selection.

B.4 Feature Correlations

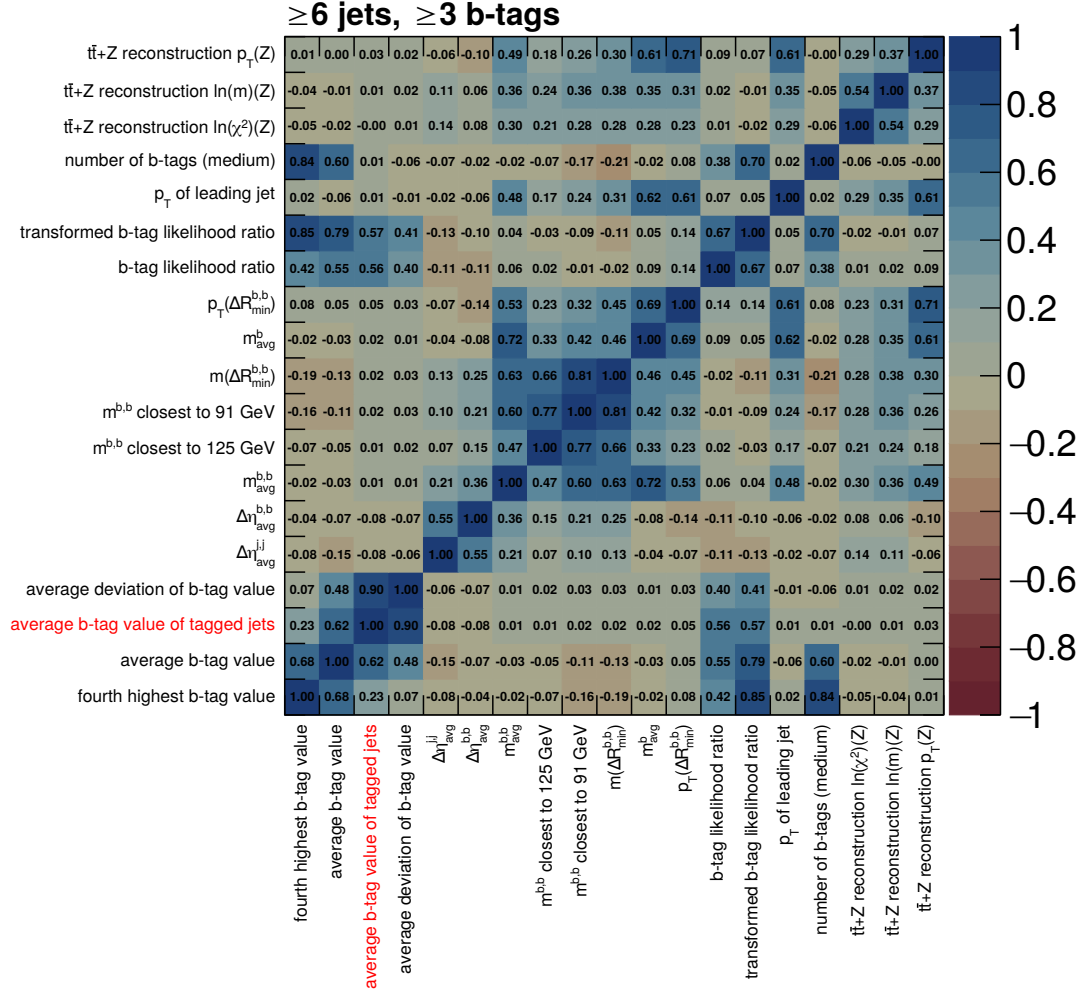


Figure B.22: **Correlation of input features used in the ≥ 6 jet region.** Features showing correlations of larger than 90% are discarded from the final feature selection and are marked in red.

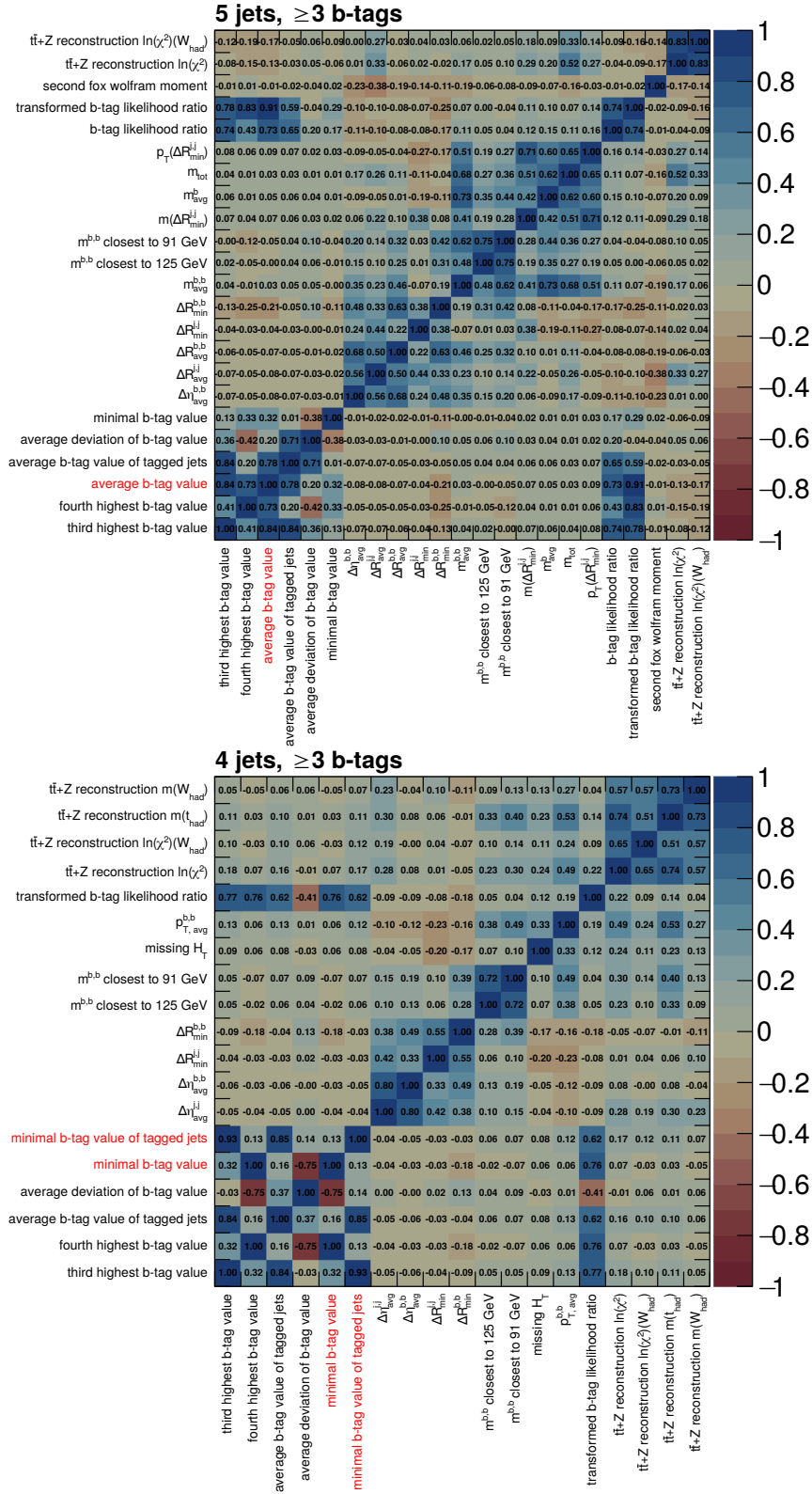


Figure B.23: Correlation of input features used in the five-jet (top) and four-jet (bottom) region. Features showing correlations of larger than 90% are discarded from the final feature selection and are marked in red.

B.5 Data/MC Comparisons

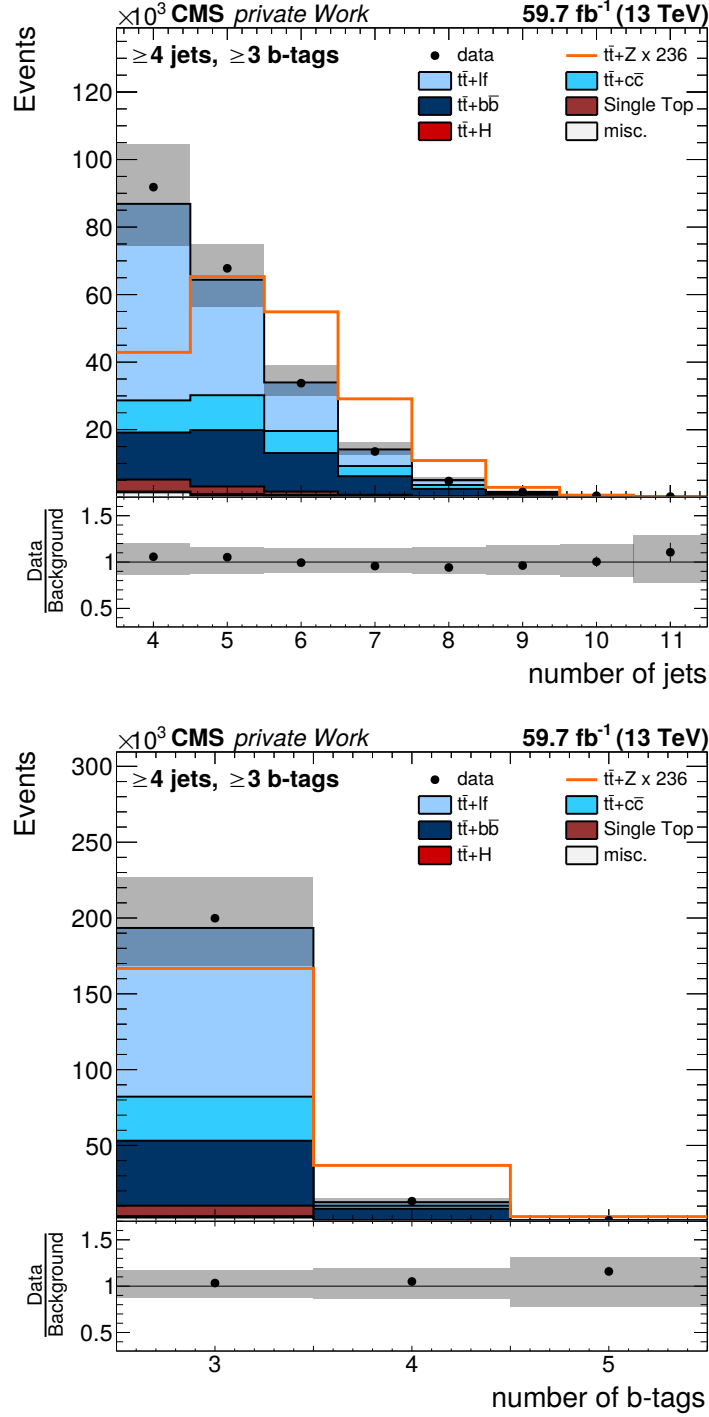


Figure B.24: **Data/MC comparison of the distribution of number of jets and number of b-tagged jets.** The predicted contributions of all background samples are stacked. the $t\bar{t}+Z$ contribution is overlaid as a line scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect.

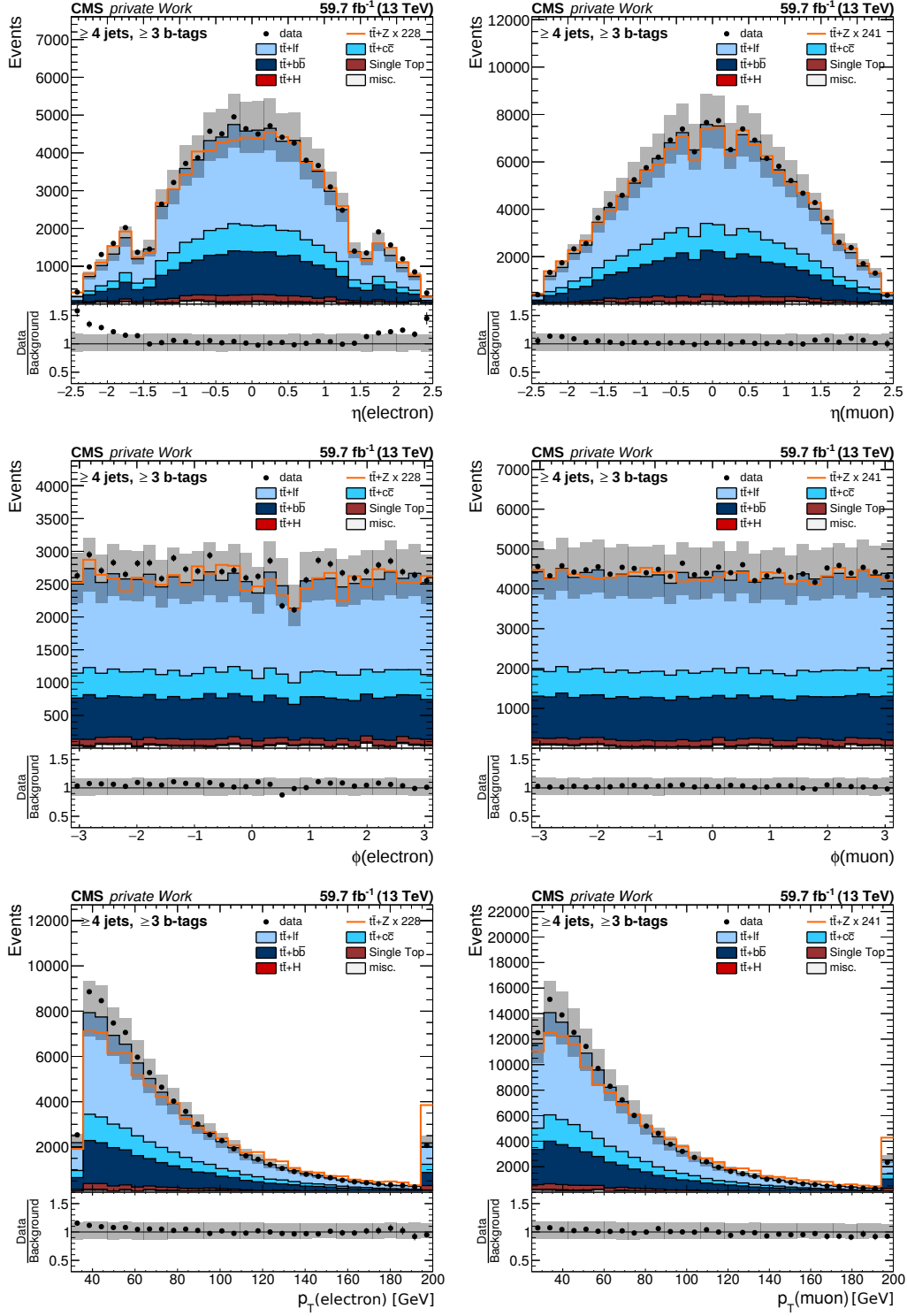


Figure B.25: **Data/MC comparison of η , ϕ and p_T distributions of electrons and muons.** The predicted contributions of all background samples are stacked. the $t\bar{t}+Z$ contribution is overlaid as a line scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect.

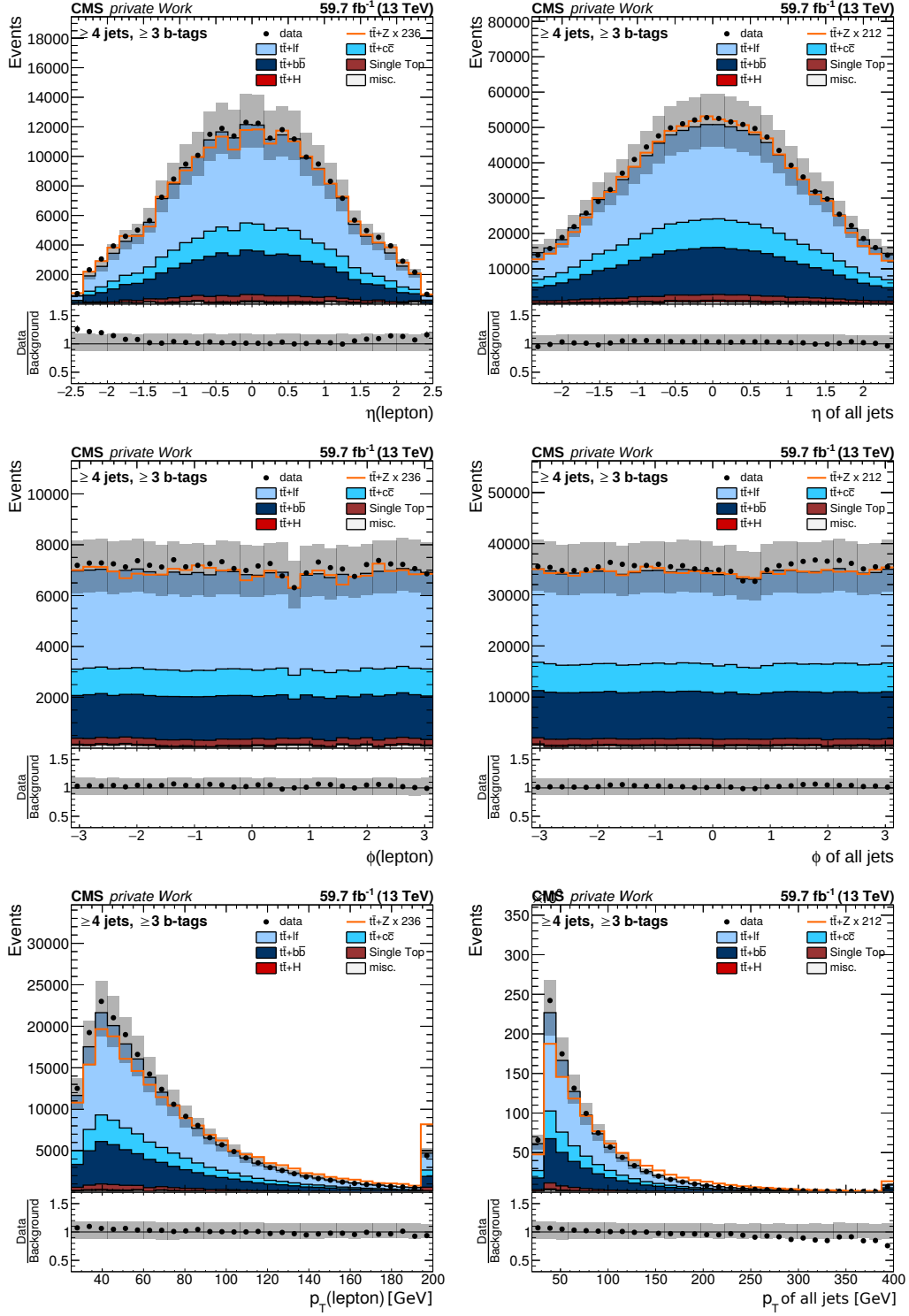


Figure B.26: **Data/MC comparison of η , ϕ and p_T distributions of charged leptons and jets.** The predicted contributions of all background samples are stacked. the $\bar{t}t+Z$ contribution is overlaid as a line scaled to match the integral of the total background. Ratios between data (black dots) and total background are shown in the bottom. The error bands correspond to the systematic uncertainties of the background contributions with a shape changing effect.

C ANN Evaluation

C.1 Confusion Matrices

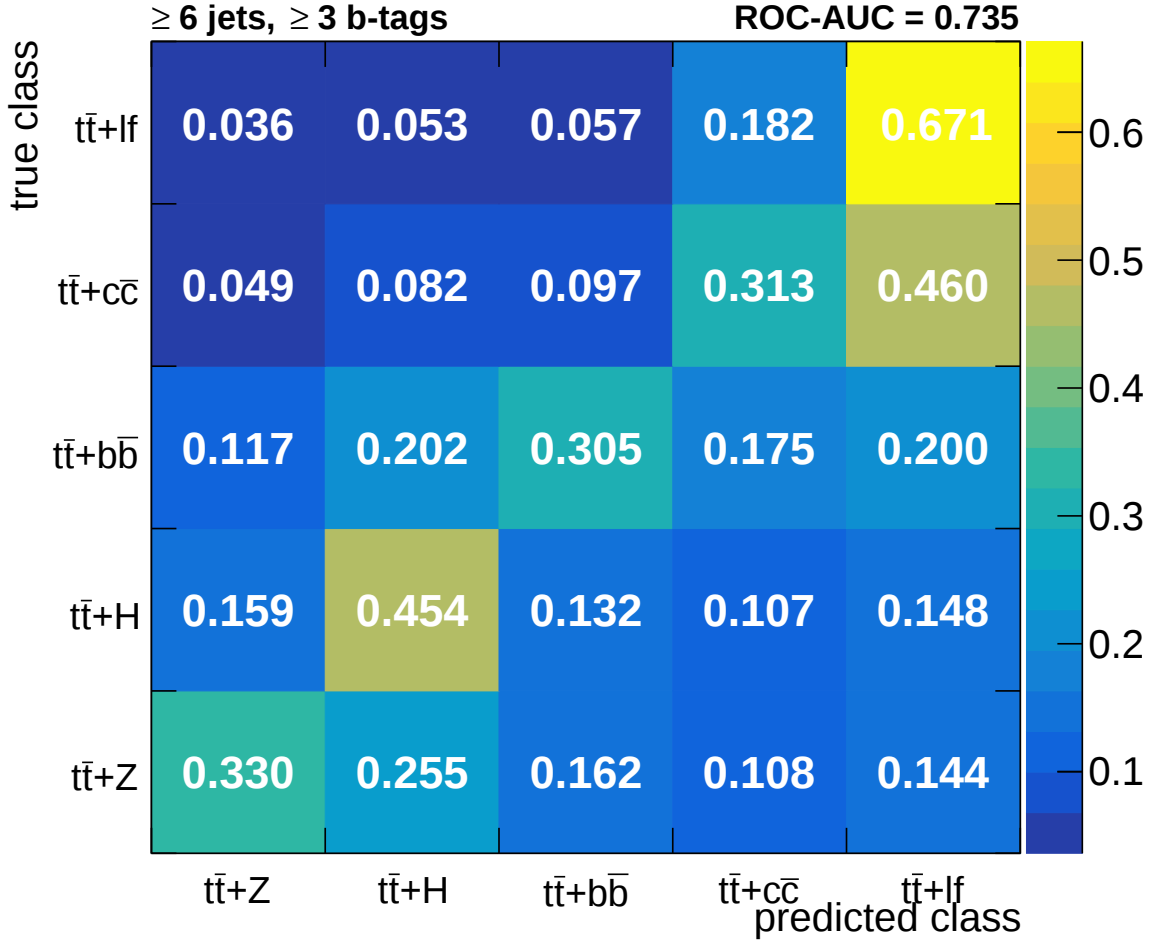


Figure C.27: **Confusion matrix for the ANN in the ≥ 6 jet region.** Number of events belonging to one class (true class) while being predicted as another class (predicted class), normalized to unity for each true class (row). Diagonal elements express the rates with which a class is classified correctly. Off-diagonal elements express the confusion rates of events of a certain class with another class.

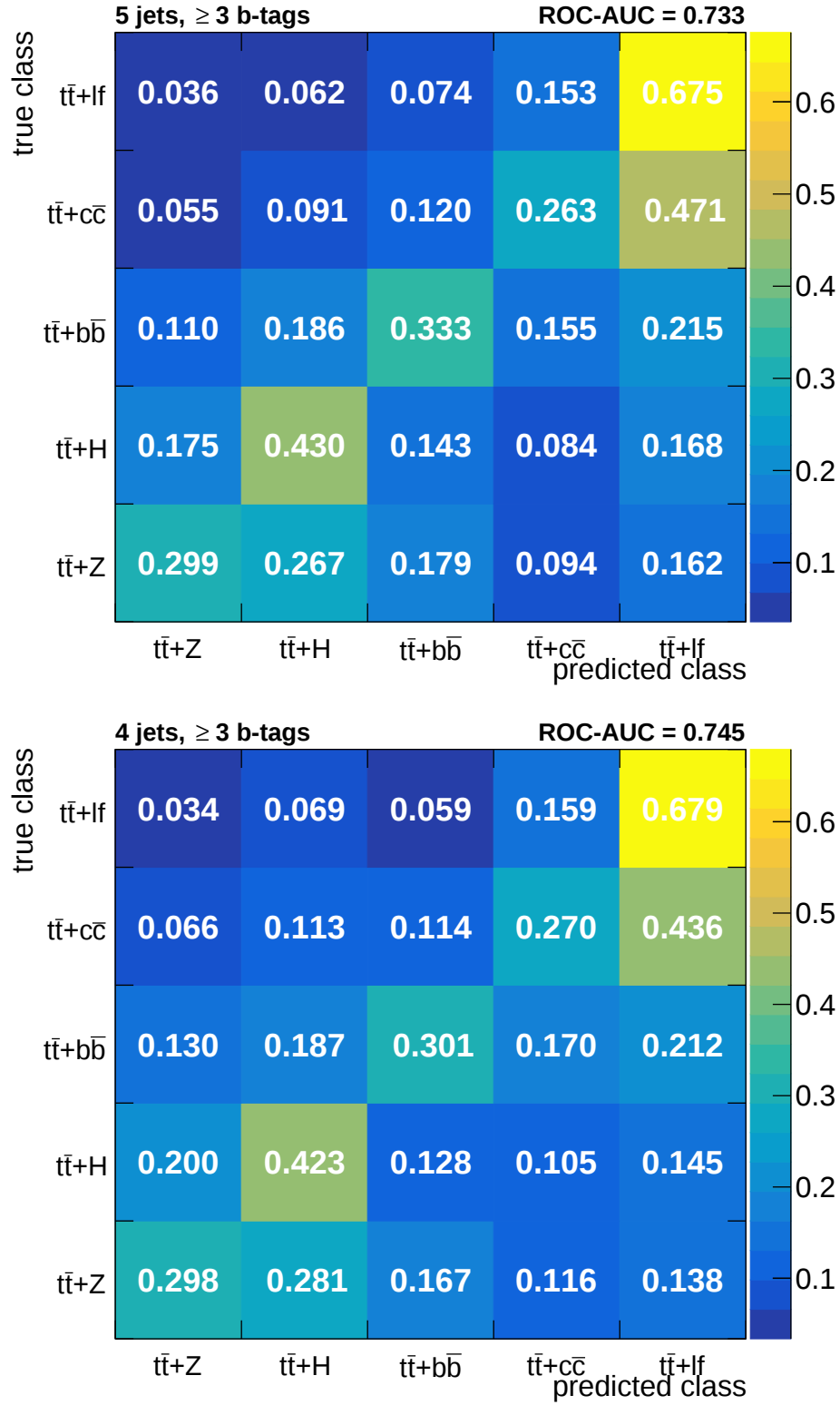


Figure C.28: **Confusion matrix for the ANN in the five-jet region (top) and the four-jet region (bottom).** Number of events belonging to one class (true class) while being predicted as another class (predicted class), normalized to unity for each true class (row). Diagonal elements express the rates with which a class is classified correctly. Off-diagonal elements express the confusion rates of events of a certain class with another class.

C.2 Performance Monitoring

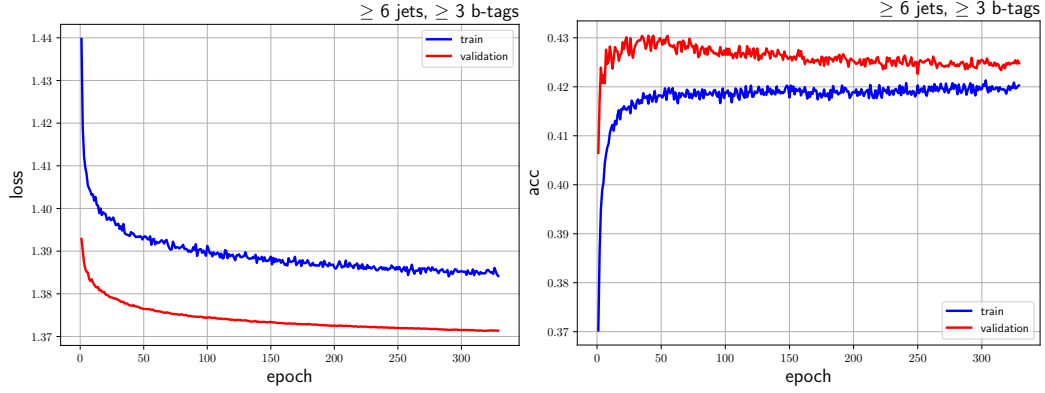


Figure C.29: **History of loss function values (left) and classification accuracies (right) during the training process of the ANN used in the ≥ 6 jet region.** The metrics are evaluated after every pass over the train set (epoch) for the train set (blue) and validation set (red).

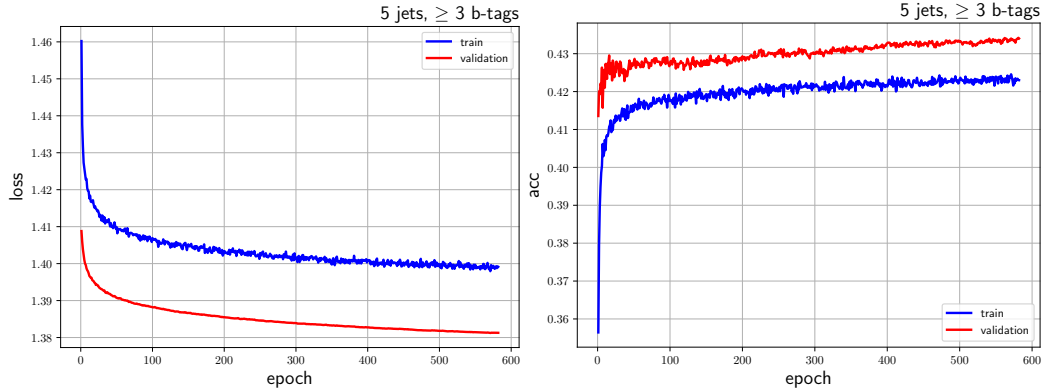


Figure C.30: **History of loss function values (left) and classification accuracies (right) during the training process of the ANN used in the five-jet region.** The metrics are evaluated after every pass over the train set (epoch) for the train set (blue) and validation set (red).

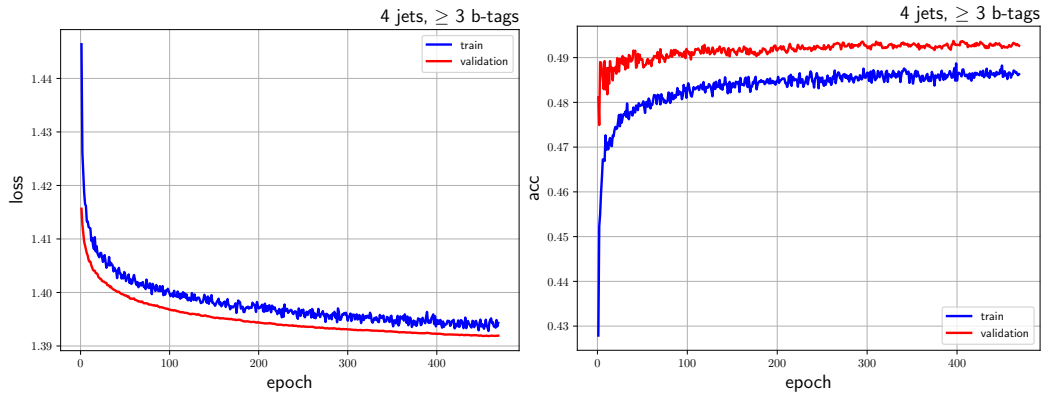


Figure C.31: **History of loss function values (left) and classification accuracies (right) during the training process of the ANN used in the four-jet region.** The metrics are evaluated after every pass over the train set (epoch) for the train set (blue) and validation set (red).

C.3 Compatibility Tests

Table C.3: **Kolmogorov-Smirnov compatibility probabilities between test data set and train data set after the ANN training.** For each ANN each process is evaluated in each output node. The discriminator distributions are compared with KS tests. Small KS probabilities values hint on bad agreement between test data set and train data set. An ANN is only used for further evaluation if no KS probability is below 0.05.

node	process	four-jet ANN	five-jet ANN	≥ 6 jet ANN
$t\bar{t}+Z$ node	$t\bar{t}+Z$	0.262	0.144	0.117
	$t\bar{t}+H$	0.827	0.636	0.981
	$t\bar{t}+b\bar{b}$	0.874	0.535	0.803
	$t\bar{t}+c\bar{c}$	0.593	0.975	0.310
	$t\bar{t}+lf$	0.407	0.460	0.783
$t\bar{t}+H$ node	$t\bar{t}+Z$	0.900	0.991	0.990
	$t\bar{t}+H$	0.969	0.887	0.997
	$t\bar{t}+b\bar{b}$	1.000	1.000	0.976
	$t\bar{t}+c\bar{c}$	0.131	0.616	0.835
	$t\bar{t}+lf$	0.835	0.721	0.777
$t\bar{t}+b\bar{b}$ node	$t\bar{t}+Z$	0.716	0.885	0.294
	$t\bar{t}+H$	0.409	0.516	0.789
	$t\bar{t}+b\bar{b}$	0.603	0.933	0.455
	$t\bar{t}+c\bar{c}$	0.449	0.970	0.876
	$t\bar{t}+lf$	0.999	0.808	0.893
$t\bar{t}+c\bar{c}$ node	$t\bar{t}+Z$	0.240	0.924	0.685
	$t\bar{t}+H$	0.538	0.994	0.419
	$t\bar{t}+b\bar{b}$	0.249	0.288	0.992
	$t\bar{t}+c\bar{c}$	1.000	0.881	0.960
	$t\bar{t}+lf$	0.609	0.885	0.998
$t\bar{t}+lf$ node	$t\bar{t}+Z$	0.703	0.331	0.178
	$t\bar{t}+H$	0.702	0.550	0.905
	$t\bar{t}+b\bar{b}$	0.856	0.986	0.775
	$t\bar{t}+c\bar{c}$	0.715	0.541	0.773
	$t\bar{t}+lf$	0.983	0.997	0.144

D Discriminator Distributions

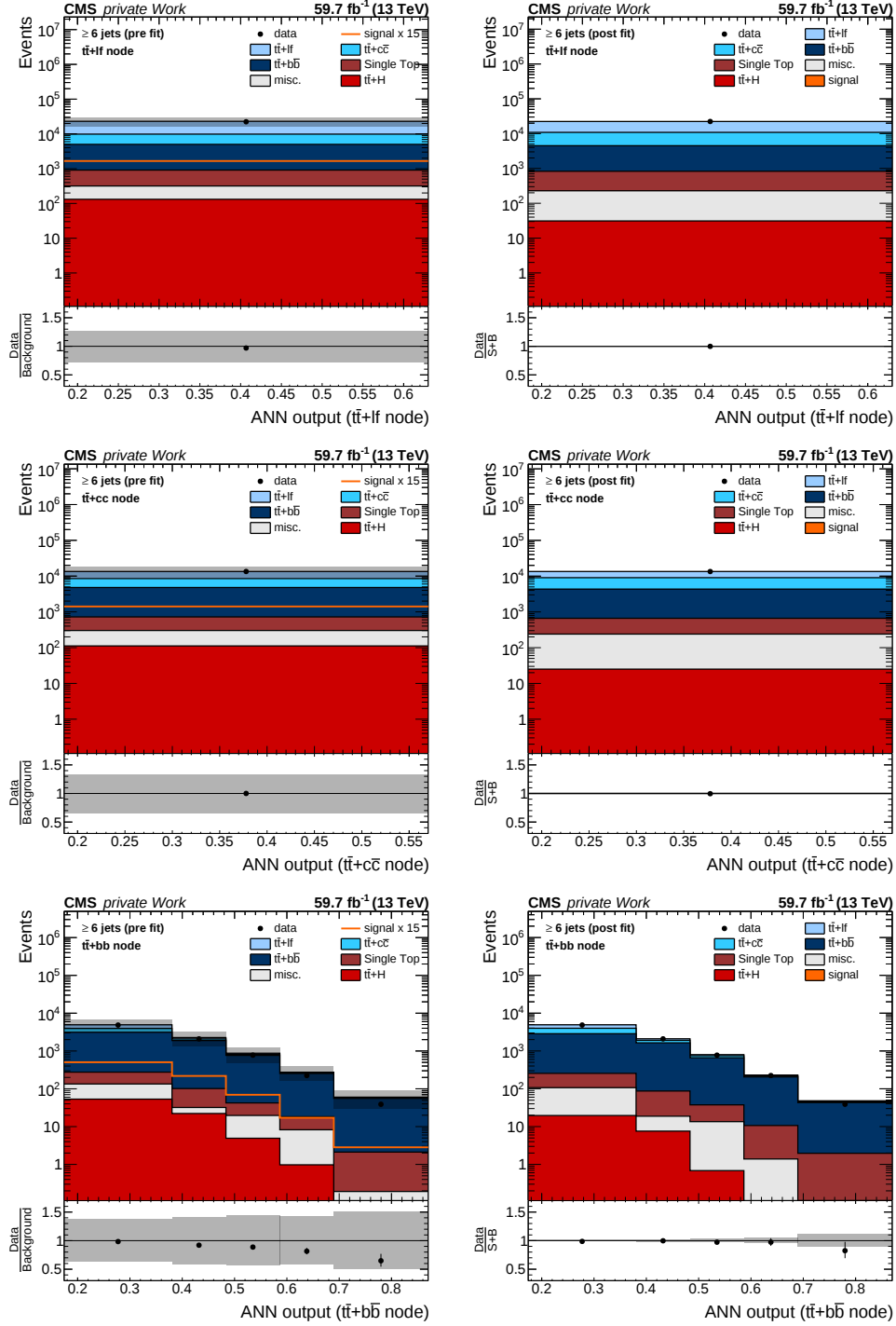


Figure D.32: **Final discriminator distributions in the ≥ 6 jet region before (left) and after (right) the fit to data.** The contributions of all background processes are stacked. In the pre-fit case, the signal contribution ($tt+Z$) is scaled by a factor of 15 and overlaid as a line. In the post-fit case, the fitted signal contribution is also stacked. The uncertainty bands include the total uncertainty of the fit model. Ratios between data (black dots) and the background (pre-fit) and the sum of signal and background (post-fit) are shown in the bottom panel.

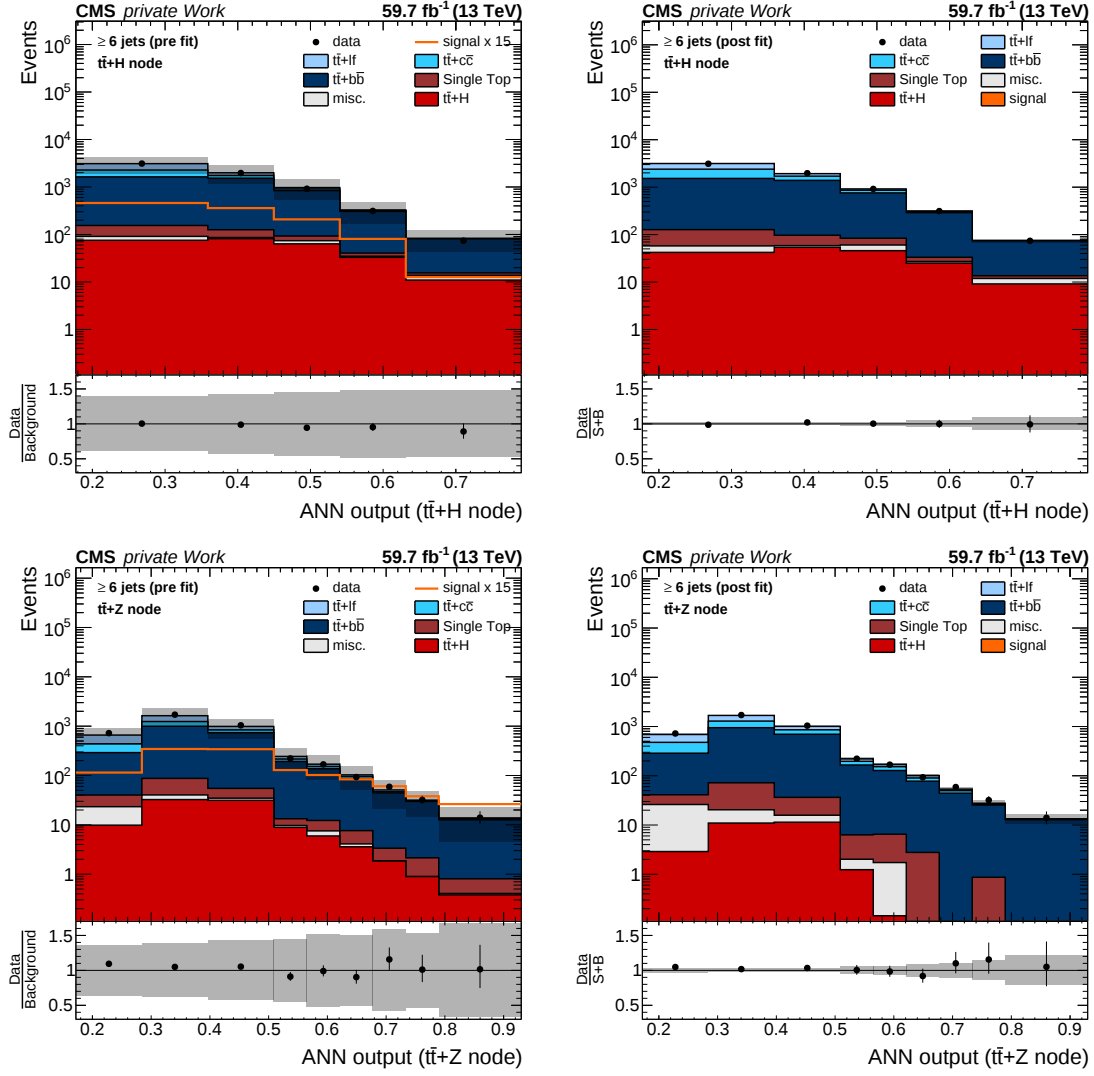


Figure D.33: **Final discriminator distributions in the ≥ 6 jet region before (left) and after (right) the fit to data.** The contributions of all background processes are stacked. In the pre-fit case, the signal contribution ($t\bar{t}+Z$) is scaled by a factor of 15 and overlaid as a line. In the post-fit case, the fitted signal contribution is also stacked. The uncertainty bands include the total uncertainty of the fit model. Ratios between data (black dots) and the background (pre-fit) and the sum of signal and background (post-fit) are shown in the bottom panel.

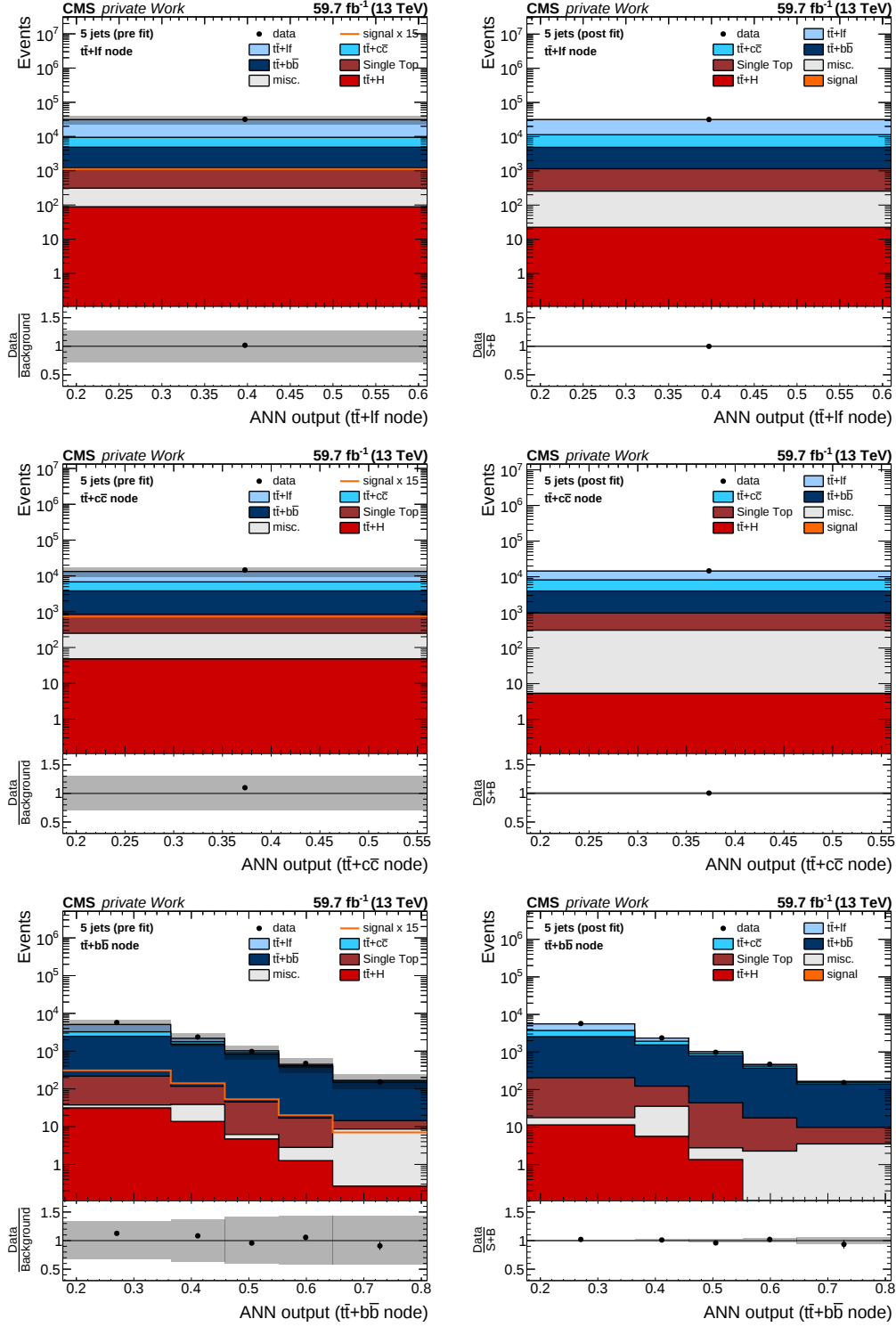


Figure D.34: **Final discriminator distributions in the five-jet region before (left) and after (right) the fit to data.** The contributions of all background processes are stacked. In the pre-fit case, the signal contribution ($t\bar{t}+Z$) is scaled by a factor of 15 and overlaid as a line. In the post-fit case, the fitted signal contribution is also stacked. The uncertainty bands include the total uncertainty of the fit model. Ratios between data (black dots) and the background (pre-fit) and the sum of signal and background (post-fit) are shown in the bottom panel.

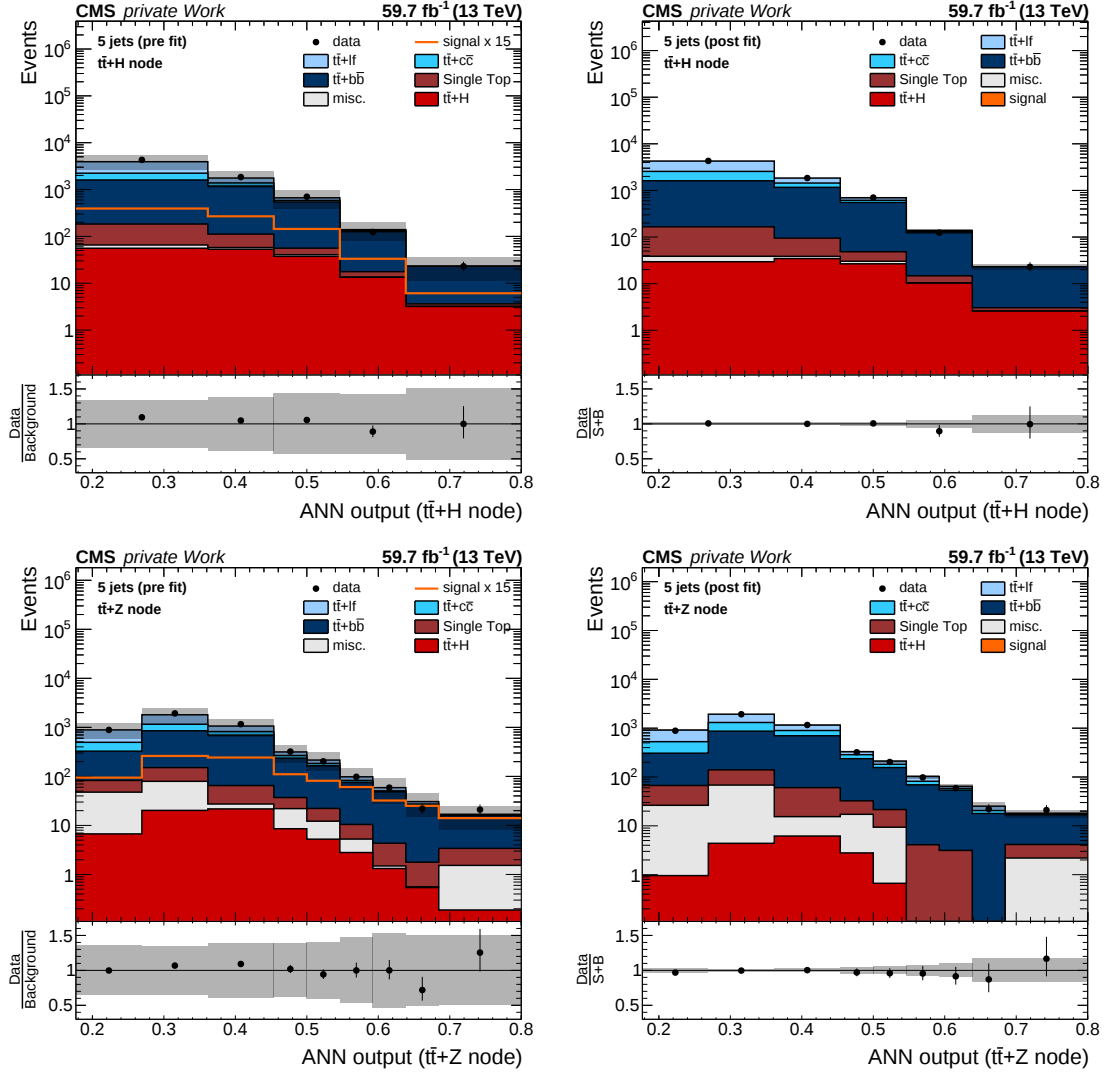


Figure D.35: **Final discriminator distributions in the five-jet region before (left) and after (right) the fit to data.** The contributions of all background processes are stacked. In the pre-fit case, the signal contribution ($t\bar{t}+Z$) is scaled by a factor of 15 and overlaid as a line. In the post-fit case, the fitted signal contribution is also stacked. The uncertainty bands include the total uncertainty of the fit model. Ratios between data (black dots) and the background (pre-fit) and the sum of signal and background (post-fit) are shown in the bottom panel.

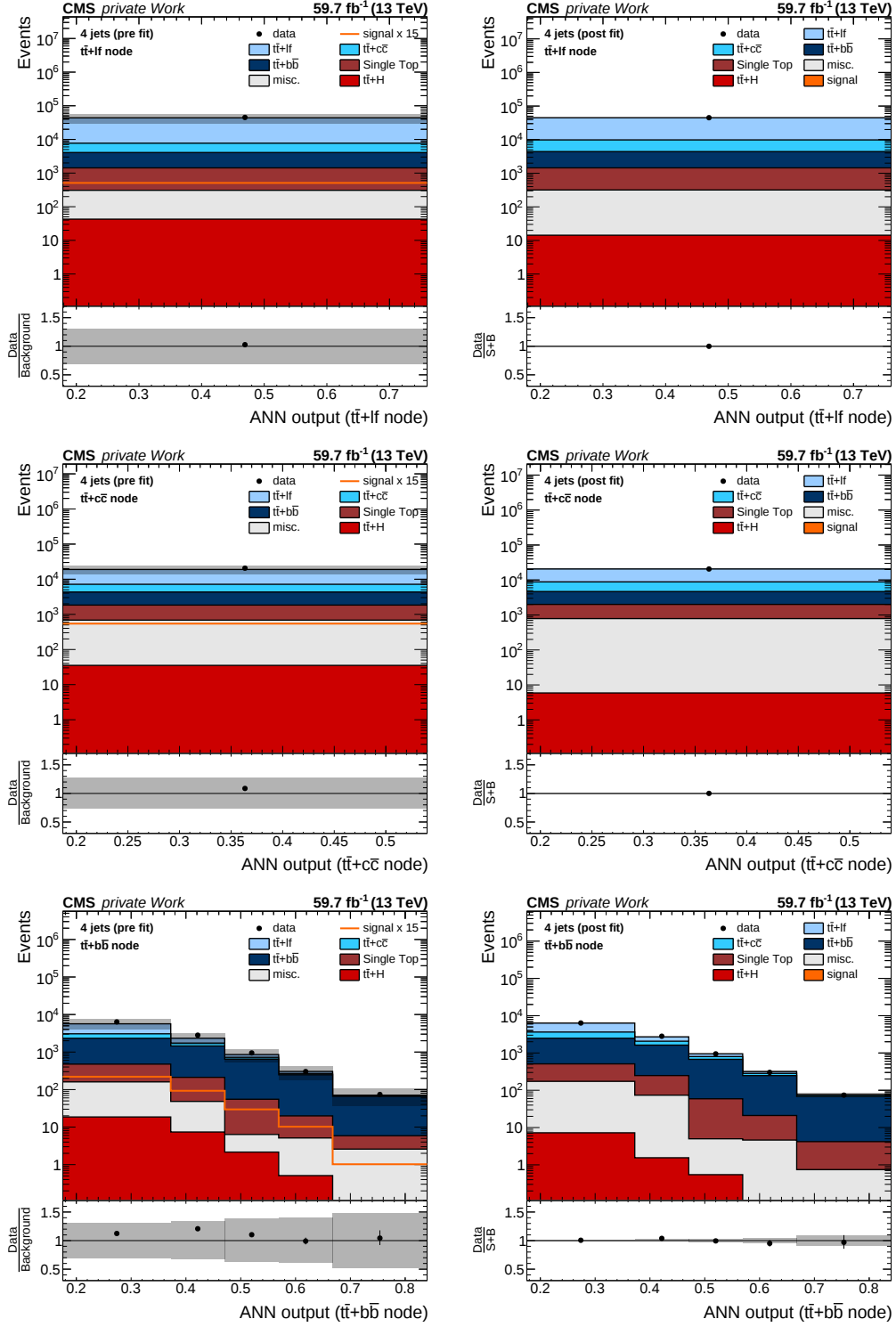


Figure D.36: **Final discriminator distributions in the four-jet region before (left) and after (right) the fit to data.** The contributions of all background processes are stacked. In the pre-fit case, the signal contribution ($t\bar{t}+Z$) is scaled by a factor of 15 and overlaid as a line. In the post-fit case, the fitted signal contribution is also stacked. The uncertainty bands include the total uncertainty of the fit model. Ratios between data (black dots) and the background (pre-fit) and the sum of signal and background (post-fit) are shown in the bottom panel.

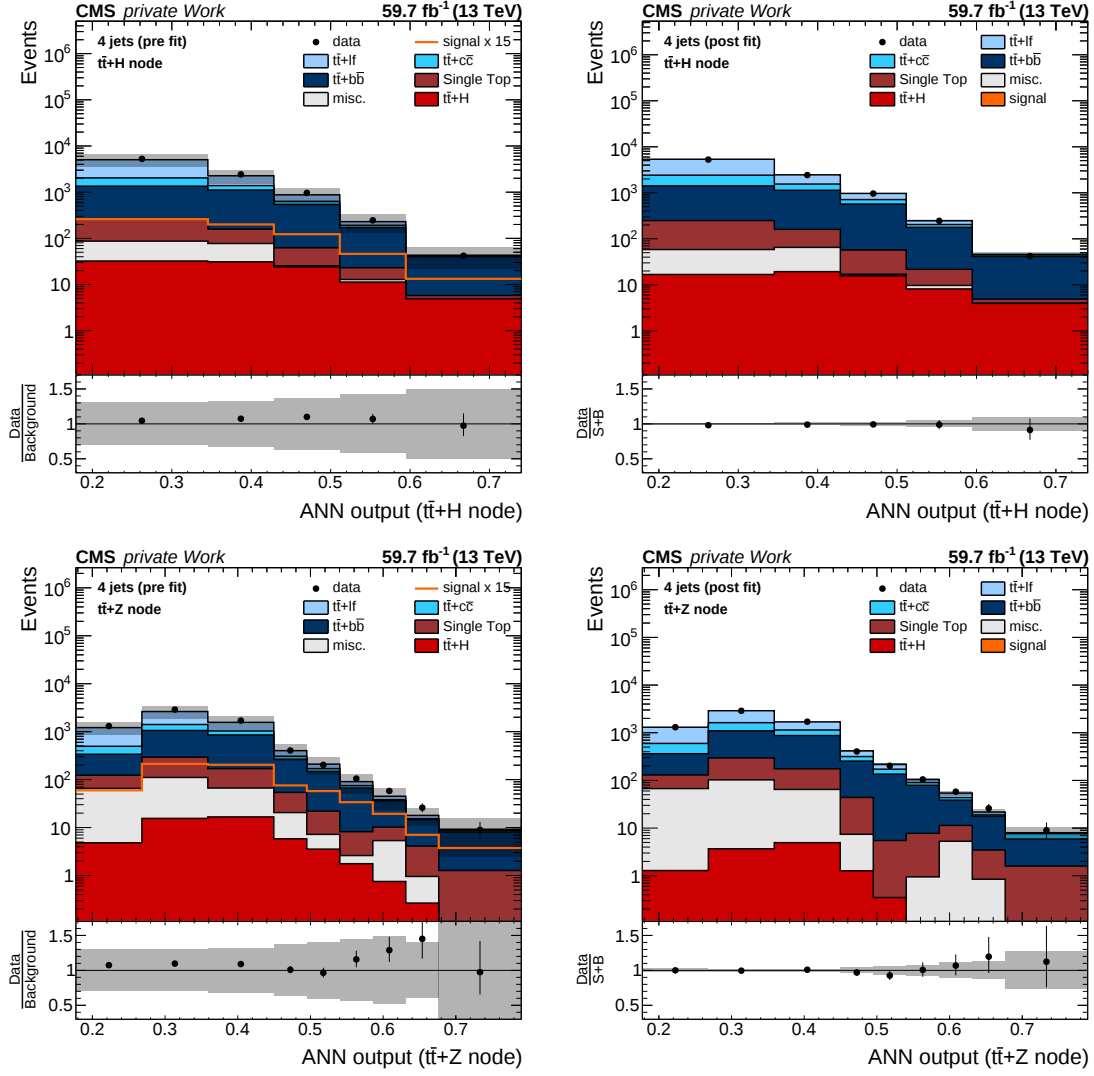


Figure D.37: **Final discriminator distributions in the four-jet region before (left) and after (right) the fit to data.** The contributions of all background processes are stacked. In the pre-fit case, the signal contribution ($t\bar{t}+Z$) is scaled by a factor of 15 and overlaid as a line. In the post-fit case, the fitted signal contribution is also stacked. The uncertainty bands include the total uncertainty of the fit model. Ratios between data (black dots) and the background (pre-fit) and the sum of signal and background (post-fit) are shown in the bottom panel.

Danksagung

Zuerst möchte ich Prof. Ulrich Husemann für die Gelegenheit danken, meine Masterarbeit in seiner Arbeitsgruppe durchführen zu dürfen. Weiterhin bedanke ich mich für die stetige Unterstützung und Hilfsbereitschaft und nicht zu letzt seinen Vorlesungen im Bachelor- und Masterstudiengang, durch welche mein Interesse an der Teilchenphysik erst geweckt wurde.

Ich möchte PD Andreas Meyer für die Übernahme meines Korreferats, der Unterstützung meiner Arbeit und der Vorlesung der Top-Physik im Masterstudiengang danken.

Ein enorm riesiger Dank gilt Dr. Matthias Schröder. Ich danke ihm für die super Beaufsichtigung und Korrektur meiner Arbeit, sowie der Beantwortung und Diskussion tausender Fragen. Sein Engagement in der Organisation der Arbeitsgruppe und in Mastervorlesungen zur Teilchenphysik sind ein erheblicher Grund, warum ich mich dazu entschieden habe in dieser Arbeitsgruppe meine Masterarbeit und nun auch meine Promotion durchzuführen. Vielen Dank für alles Matthias!

Ebenso sei allen anderen Mitgliedern der Arbeitsgruppe gedankt:

Ich möchte den Bewohnern meines Büros, Sebastian Wieland, Philip Keicher und Michael Waßmer danken, durch welche die Zeit meiner Masterarbeit sehr viel Spaß gemacht hat. Vielen Dank für die ganze fachliche Hilfe, die gute Zeit und auch die Hilfe bei der Kaktus-Zucht!

Karim El Morabit sei ein besonderer Dank gewidmet. Jegliche technischen oder physikalischen Fragen, sowie Fakten zu obskuren Tieren oder zufälligen Wikipedia-Artikeln wurden von ihm immer mit viel Enthusiasmus und Hilfsbereitschaft beantwortet und diskutiert. Dies trug nicht unerheblich zu dieser guten Zeit bei, die ich in dieser Arbeitsgruppe hatte! Lea Reuter sei für die Unterstützung in und außerhalb des Arbeitsalltages gedankt. Danke, dass du immer ein offenes Ohr für mein Rungemeckere hast!

Meinen Mitmasteranden Jörg Schindler, Maximilian Horzela, Peter Krämer, Nikita Shadskiy und Simon Ehnle sei für eine tolle Arbeitsatmosphäre und stetige gegenseitige Hilfe gedankt. Vielen Dank auch an Karim, Lea, Peter, Sebastian, Philip und Michael für das Probelesen und Verbessern meiner Arbeit!

Weiterhin möchte ich meinen Freunden danken, auch wenn sie die Teilchenphysik für nutzlosen Blödsinn halten.

Der größte Dank gilt jedoch meiner Familie. Ohne euch wäre dieser ganze Spaßausflug nicht möglich gewesen. Vielen Dank, dass ihr mich immer Unterstützt! Ihr könnt ja mal wieder zum Essen vorbei kommen.