Differential cross-section measurement of the *tZq* process with the ATLAS detector

Chris Boever

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

November 2019

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...
Date Signature

1. Gutachter: Prof. Dr. Ian C. Brock
2. Gutachter: Priv.-Doz. Dr. Philip Bechtle

Contents

CHAPTER 1

Introduction

„Dass ich erkenne was die Welt
Im Innersten zusammenhält, (...)“
Faust, der Tragödie Erster Teil, Johann Wolfgang von Goethe

It has been an everlong quest of humankind to understand and predict the flow of nature. An innumerable amount of theories emerged in the last centuries trying to fulfill this goal, leading to the two best physical theories known today, which have astonishing precision: General Relativity and the Standard Model of Particle Physics (SM).

In the last centuries, subsequent discoveries revealed that many objects believed to be fundamental, i.e. non-composite, turned out to be composite objects after all. Therefore, all so far discovered and by now believed fundamental particles and interactions are described by the SM, using quantum fields. However, although very successful in describing the quantum nature of particles and fields, the SM cannot explain all experimental results; e.g. the SM has been incompatible with General Relativity and also does not provide an explanation for dark matter or dark energy predicted by standard cosmological models.

A good place to search for discrepancies between data and SM predictions is the regime of the heaviest known fundamental particles: the top-quark, and W, Z and Higgs bosons. Many theories beyond the Standard Model (BSM) make predictions for these particles deviating from SM predictions.

The topic covered in this thesis is the production of a single top-quark in association with a Z boson, called tZq hereafter. The tZq process is especially interesting to challenge the SM, as it provides information on the coupling of the Z boson to the quark sector as well as to other gauge bosons. Furthermore, the tZq process is theoretically one of the most prominent backgrounds expected for searches like the production of a single top-quark in association with a Higgs boson, as well as for more exotic researches like “flavor changing neutral current” analyses. Thus, after officially being discovered, [1] the next logical and also crucial step is to fully understand the kinematics of the tZq process to see if it behaves according to predictions of the SM; this is the goal that should be achieved during this thesis by measuring a differential cross-section.

This thesis is structured as follows. An overview of theoretical concepts used in particle physics is given in chapter 2. An insight in the working principles of the ATLAS detector, which recorded the data used for this analysis, is provided in chapter 3. A detailed description of the tZq process as well as of the means developed to extract characteristic signatures from data is given in chapter 4.

The concept of unfolding, an indispensable theoretical tool needed for any differential cross-section measurement is introduced in chapter 5; furthermore, the approaches taken during development of software enabling this differential cross-section measurement are explained. Chapter 6 deals with tests performed to validate that the framework implemented for the unfolding process performs as expected. First results for differential cross-sections measurements performed on data are presented. Important key aspects are summarized in chapter 7.

CHAPTER 2

Theoretical concepts

This chapter serves the purpose of familiarizing the reader with the theory to understand particle and particle collider physics. Explanations concerning the Standard Model of Particle Physics are provided. Afterwards, concepts common for physics at particle colliders will be illustrated. In the end, the top-quark will be introduced. This particle is a focal point throughout this document and therefore requires a detailed description.

2.1 The Standard Model

The Standard Model of Particle Physics (SM) is a theory which, on microscopic scales, describes the behavior of elementary particles, fundamental particles which do not show internal structures, and the interactions between them (information in this section can be found in [2], if not explicitly stated otherwise). The mathematical background for the SM are *quantum field theories* (QFTs); the QFTs at the heart of the SM are Quantum Chromodynamics (QCD) and electroweak theory. The role of these QFTs is to link fields together. Excitations of those fields correspond to particles.

The particles described by the SM (as displayed in figure 2.1) are separated into fermions or bosons. Fermions are particles with half-integer spins, an intrinsic property of each particle which is mathematically equivalent to angular momentum. On the other hand, bosons are particles with an integer spin. The photon γ, the gluon g and the weak gauge bosons $W^{\pm}$ and Z are gauge bosons with a spin of 1, each of them associated to a specific interaction (thus also called force carriers). The Higgs boson is the only particle in the SM with a spin of 0. Each of the particles contained within the SM has a corresponding antiparticle, which has the same mass and spin, but the opposite value for charge-like quantum numbers (like electric charge).[1] The photon, Z and Higgs boson are their own antiparticles.

Fermions are subdivided into leptons and quarks, depending on whether they participate in strong interactions or not. Furthermore, they are usually grouped together into three generations. Each generation is built of two quarks and two leptons. The first of these generations contains the lightest known matter particles, namely the u- and d-quark (u and d stand for “up” and “down” respectively), the electron e^{-} and the electron-neutrino ν_e. Quarks might combine into protons and neutrons, which in turn can form nuclei. Atoms finally consist of nuclei orbited by electrons. The other generations are built from heavier particles; except for their mass, they have identical properties to their relatives

---

[1] Unless stated otherwise, the term “particle” refers to both particle and antiparticle during this thesis.

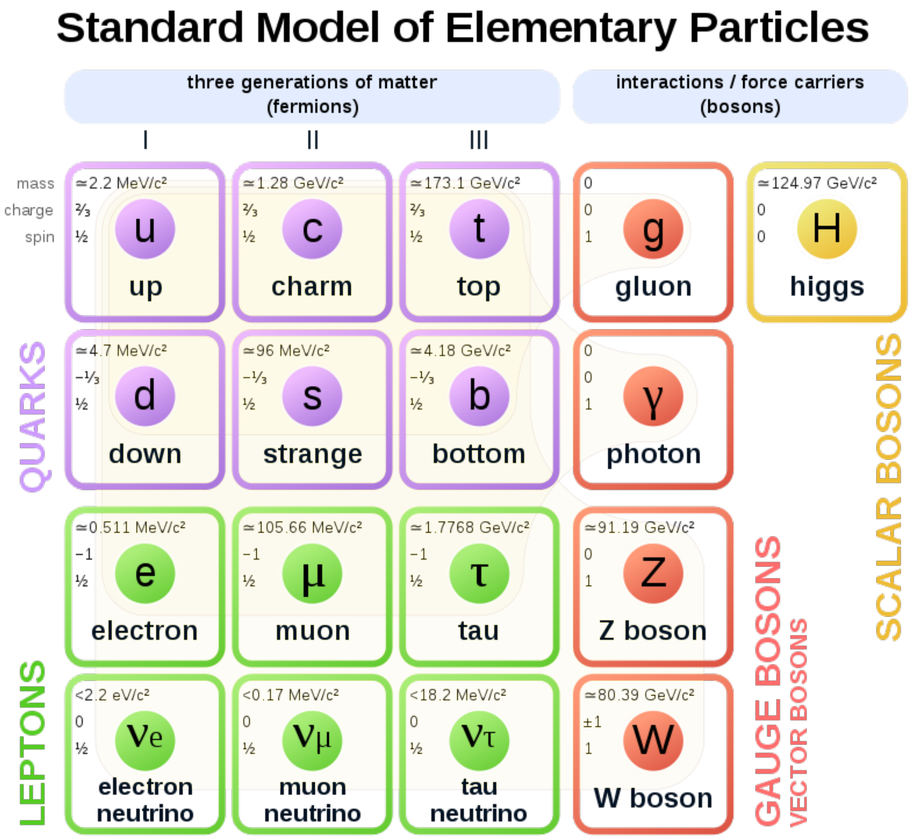


Figure 2.1: Overview of the particles that make up the SM with information of the mass, electric charge and spin of each respective particle. Furthermore, boxes are color-coded to indicate what "type" of particle they represent, as labeled in the graphic. Figure taken from [3].

in the first generation (it is still not known if neutrinos from the second and third generation have a higher mass than the electron-neutrino). The top-quark, one of the quarks of the third generation, is the heaviest particle in the SM.

A universe containing only particles without interactions between them would be impossible to explore. Fortunately, the role of force carriers is to transmit these interactions. For example, the photon, γ, is the mediator of the electromagnetic interaction, as described by Quantum Electrodynamics (QED); every particle of the SM with non-zero electric charge interacts with it. Note that the photon does not carry an electric charge itself. Self-interactions are therefore not predicted by the SM and also have not been observed. [4]

The weak force is mediated by $W^{\pm}$ and Z bosons. It is the only force in the SM in which all fermions, even neutrinos, participate. In addition, the weak force is the only interaction in the SM that allows flavor-change and also transitions between different generations. The transitions between different generations have however only been observed between different quarks, e.g. a s-quark transforming into a u-quark, by mediation of $W^{\pm}$ bosons. This is possible as weak eigenstates are not equal to mass eigenstates. The mathematical description is provided by the Cabibbo–Kobayashi–Maskawa (CKM) matrix, a unitary 3 × 3 matrix, which links the weak eigenstates q' and the mass eigenstates q:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\mathrm{CKM}} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix} \ . \tag{2.1}$$

The current values for the entries of the CKM-matrix are: [4]

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{-0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{pmatrix} \ . \tag{2.2}$$

Non-zero off-diagonal elements V_{ij} inside the CKM-matrix allow transitions between two specific quark-types; the probability for this transition is proportional to $|V_{ij}|^2$. A total of four free parameters allows for phase shifts and CP-violation in weak interactions. Note that diagonal elements have the largest values, so transitions are most likely to happen within the same generation. In most analyses, the top-quark usually is assumed to decay to b-quarks only. This aspect is important for a full reconstruction of kinematical properties of top-quarks (see sections 3.3.5 and 4.1).

Electromagnetic and weak interactions are successfully described by a joint theory called electroweak theory. It introduces the concepts of weak isospin T and hypercharge Y_W to describe electroweak interactions between particles. The third component of the weak isospin of a particle, T_3, is determining the coupling strength to $W^{\pm}$ bosons. This is important to describe weak interactions for fermions: fermion fields are classified into left- and right-handed fields (a property called chirality). Only left-handed fermions carry a weak isospin, while $T_3 = 0$ for the right-handed component (vice-versa for anti-fermions). The hypercharge is related to weak isospin and electric charge:

$$Y_W = 2\,(Q - T_3) \ . \tag{2.3}$$

However, within this theory, a mechanism called spontaneous symmetry breaking is needed to explain non-zero masses of the $W^{\pm}$ and Z bosons. This led to the prediction of a new quantum field, called the Higgs field, and an associated spin-0 particle: the Higgs boson. Fermions obtain a mass through interactions with the Higgs field; the stronger the coupling, the higher the mass of the particle. Particles that do not couple to the Higgs field, like e.g. the photon, remain massless. The Higgs boson was the last particle of the SM to be found by the ATLAS [5] and CMS [6] collaborations in 2012.

Lastly, the strong force is described by Quantum Chromodynamics, which describes the behavior of particles with color charge: the quarks and the gluons. The gluon is the force carrier of the strong interaction. It carries color charge allowing self-interactions (similar to weak gauge bosons, but in contrast to the photon). Leptons do not carry any color charge and therefore do not participate in strong interactions. The strong force is responsible for the formation of hadrons, bound states of quarks, with the restriction that the resulting particle has to be color-neutral. A composite particle is said to be color-neutral if its net color charge vanishes. This can e.g. be achieved if the color charge of the particle is created by a combination of the color charges red, blue and green (these composite particles are then called *baryons*) or by a color-anti-color combination (which are denoted as *mesons*). These are however not the only combinations known anymore. For example, pentaquarks are bound states of five quarks / anti-quarks that have been observed in 2015. [7]

Of the particles described in this section, the top-quark, t, and the Z boson are the particles of interest for this thesis.

2.2 Feynman diagrams

Many probability amplitudes for SM reactions are calculated using perturbation theory, i.e. the exact result is approximated by a sum of terms with decreasing importance. Each term of this sum can graphically be represented by a tool called *Feynman diagrams*. [8] Feynman diagrams are sketches built of lines, representing particles, and vertices, where interactions between particles occur. Time axis in this thesis is oriented from left to right, which means that the initial state is drawn on the left-hand side of the diagram, while the final state is represented by particles on the right-hand side of the diagram.

External lines correspond to observable particles, while internal lines correspond to virtual particles. Virtual particles are not observable, but are needed in quantum field theoretical perturbative calculations. A virtual particle does not need to obey the energy-momentum relation, $E^2 = m^2 + p^2$ (natural units are assumed. See section 2.3.1 for details.); a particle which does obey this relation is called *on-shell*, otherwise it is denoted as *off-shell*. However, energy and momentum conservation are not violated during the process. Straight lines with arrows pointing in the same direction as the time axis correspond to fermions, while straight lines with arrows pointing against the flow of time are anti-fermions. Wave-like lines are usually understood as one of the gauge bosons: γ, Z or $W^{\pm}$. Spring-like lines represent gluons.

Interactions between particles happen at vertices, points at which lines intersect. Which interaction is at play depends on the type of particles involved. Mathematically each vertex in a Feynman diagram is proportional to some power of a coupling constant, which is a measure how strongly particles involved in a vertex couple to each other. A vertex which involves e.g. an electron, a positron and a photon is an example of an electromagnetic interaction.

For a given process, the calculation of the probability amplitude is possible by drawing all Feynman diagrams with the same initial and final states. All the diagrams proportional to the lowest occurring power of the coupling constant are referred to as *Leading Order* (LO) diagrams. Each diagram corresponds to an amplitude $\mathcal{M}_i$. The total amplitude $\mathcal{M}$ for the process under consideration is then the sum of all possible $\mathcal{M}_i$ that have the same initial and final states. Including more diagrams of higher order (which would e.g. correspond to diagrams which contain loops) in this summation leads to more accurate results. Due to the summation, interference between different diagrams is possible. The final cross-section (compare to section 2.3.1) is proportional to $|\mathcal{M}|^2$.

An example is given in figure 2.2. The Feynman diagrams depicted correspond to LO diagrams involving electromagnetic interactions for the process $e^+e^- \rightarrow e^+e^-$. Diagram 2.2(a) shows the annihilation of an electron and a positron into a photon. The photon then goes through photon conversion, a process in which the photon decays into an e^+e^- pair. This decay process is the final state required for this example. Diagram 2.2(b) on the other hand, is the depiction of an electron-positron scattering. The scattering process is described by exchange of a virtual photon.

It is important to realize that Feynman diagrams are a helpful mathematical tool only if perturbation theory is applicable. This becomes important if the coupling constants (which are in fact not constant) become close or larger than 1; at this point, perturbation theory breaks down. Diagrams of higher order would have similar or higher contribution to the infinite sum in perturbation theory than LO

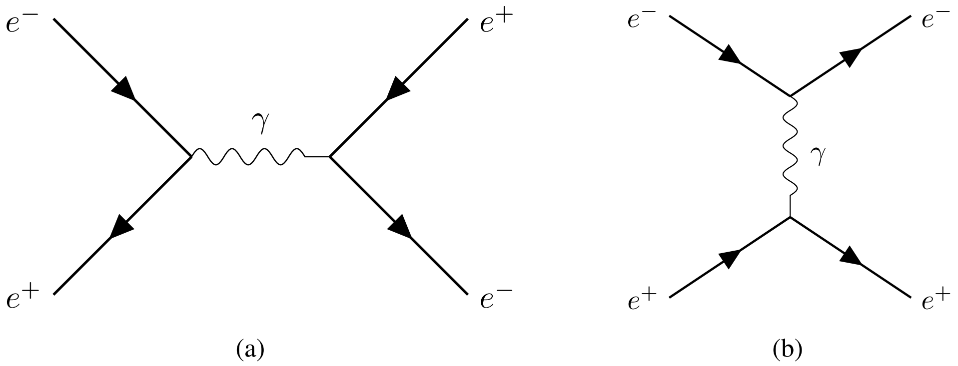


Figure 2.2: Example Feynman diagrams at LO for the process $e^+e^- \to e^+e^-$. Diagram 2.2(a) shows the annihilation of an electron and a positron into a photon, which via photon conversion results in the required final state. Diagram 2.2(b) is another Feynman diagram resulting in the same final state by exchange of a virtual photon.

diagrams, making it impossible for the final result to converge to a finite result. This plays a major role in QCD calculations, as the coupling strength α_S of the strong force increases at low momenta of the particles involved in the interaction (compare to chapter 10.5 in [2]).

2.3 Concepts from (hadron) collider physics

In order to prove or disprove the validity of the SM, particle colliders are used to accelerate particles to very high velocities and thus high energies before these particles interact with each other. If what is currently considered a fundamental particle has a substructure, this substructure should only be probeable at high energies. For this reason, large accelerators are built to reach record breaking energies. Basic concepts of particle collider physics are explained in this section.

2.3.1 Fundamental concepts

A brief overview of concepts that are widely used during this thesis follows.

Natural units

Throughout this thesis, natural units are used. In this unit system, the values for the speed of light c, the Planck constant $\hbar$ and the Boltzmann constant k_B are defined to be:

$$c = \hbar = k_B = 1 \ . \tag{2.4}$$

These constants are unitless and equal to one. As such, they are dropped from any forthcoming equations. This way, many quantities like energy or momentum can be expressed in powers of energy units. In high energy particle physics, the energy unit that is most useful is the electronvolt eV. The energy units associated to other physical quantities are as follows:

$$\begin{aligned}[\text{energy}] = [\text{momentum}] = [\text{mass}] = \text{eV}\,, \\ [\text{length}] = [\text{time}] = \text{eV}^{-1}\,. \end{aligned} \tag{2.5}$$

Cross-section, lifetime, decay rate

A cross-section, usually denoted by the symbol σ, is a measure of the probability of an interaction between particles. It can be imagined as an effective area: when the effective area of particles overlap, the reaction takes place. Consequently, the unit for a cross-section is a barn, with $1\,\text{b} = 10^{-28}\,\text{m}^2$.

Distinctions are made between inclusive and differential cross-sections. For the former, the whole phase space is considered (i.e. each variable assigned to particles can take any possible allowed value). Differential cross-sections however are measurements of cross-sections as a function of a specific observable. This can e.g. be an angle or a momentum associated to one of the final-state particles. The topic covered in this thesis is the measurement of a differential cross-section.

Another quantity specified for most particles is the lifetime τ. Due to the stochastic behavior of quantum theories, this quantity describes a time span until which a sufficiently large ensemble of particles has shrunk to $1/e$ of its initial sample size. Furthermore, Heisenberg's uncertainty relation states that time and energy are conjugate quantities. As a consequence, the decay width can be defined as:

$$\Gamma = \frac{1}{\tau}\,. \tag{2.6}$$

This quantity has therefore the unit of an energy, eV. Each possible decay mode contributes to the total decay width additively, such that:

$$\Gamma = \sum_{i=1}^{n} \Gamma_i\,, \tag{2.7}$$

where Γ_i is the decay width of a specific mode, called partial decay width. With partial decay widths known, it is also possible to define the branching fraction $\mathcal{B}_i$:

$$\mathcal{B}_i = \frac{\Gamma_i}{\Gamma}\,, \tag{2.8}$$

which is a measure for how often a particle decays into one or more particles via process i.

Center-of-mass energy

The center-of-mass energy $\sqrt{s}$ is defined as:

$$\sqrt{s} = \sqrt{\left(\sum_{i=1}^{n} E_i\right)^2 - \left(\sum_{i=1}^{n} \vec{p}_i\right)^2}\,, \tag{2.9}$$

where E_i denotes the energy and $\vec{p}_i$ the momentum of particle i. Considering this quantity in the center-of-mass reference frame $\left(\sum_{i=1}^{n} \vec{p}_i = \vec{0}\right)$, it becomes apparent that $\sqrt{s}$ describes the total energy available for particle production. In the case of two colliding beams with same energy and same particle type, the formula reduces to $\sqrt{s} = 2E_{\text{beam}}$ when neglecting the mass of the particles.

Event rates, luminosity

For collision experiments, most often the number of events measured is an important quantity. For any given process, the actual number of events is proportional to its cross-section:

$$N = \sigma \int L(t)\, \mathrm{d}t = \sigma \mathcal{L} \,. \tag{2.10}$$

The quantity L is called *instantaneous luminosity*, whereas $\mathcal{L}$ is referred to as *integrated luminosity*. These quantities can be seen as measure for the maximum available performance of a particle accelerator. For two colliding beams assumed to have a Gaussian beam profile, the instantaneous luminosity is related to the beam parameters via:

$$L = \frac{f n_1 n_2}{4\pi \sigma_x \sigma_y} \,, \tag{2.11}$$

f denotes the collision frequency of the beams, n_1 respectively n_2 the number of particles per colliding bunches (more details given in section 3.1), and σ_x and σ_y the beam sizes in horizontal respectively vertical direction.

For this thesis, the integrated luminosity plays a more important role than the instantaneous luminosity as the data used was collected over several years. $\mathcal{L}$ enters several formulas described in chapter 5, when conversions from event yields to cross-sections are required.

2.3.2 Partons and parton distribution functions

In hadron colliders, two hadron beams collide; in the case of the LHC (described in more detail in section 3.1), the colliding beams consist of protons. To understand these interactions, it is crucial to understand the structure of the protons, as they are not fundamental, but composite particles.

Protons are made up of particles which are called *partons*. This term refers to valence and sea quarks as well as gluons. According to this picture, valence quarks inside a proton are two up- and one down-quark, while sea quarks can be created by conversion from a gluon to a quark-anti-quark pair. This also means that the total momentum of a hadron is equal to the sum of momenta of the partons inside of it. This leads to the introduction of a variable x (called *Bjorken* x), which describes the momentum fraction that a parton i is carrying:

$$x_i = \frac{p_i}{p_{\text{proton}}} \,. \tag{2.12}$$

The parton model also indicates that in inelastic proton-proton collisions, not the protons as a whole are interacting, but various partons are interacting with each other. The rest of the partons are assumed to not participate in the interaction and are therefore called *spectators*. However, the type of parton and its exact momentum is not exactly known. The only restriction is that a parton has to carry less momentum than the entire proton ($x_i < 1$) to ensure momentum conservation. This is why the concept of parton distribution functions $f_i(x_i, Q^2)$ (PDFs) has been introduced. PDFs describe the dynamics inside hadrons; namely the probability density to find a specific kind of parton i with momentum fraction x_i at a given energy scale Q^2, where $Q^2 = -q^2$ can be obtained from the four-momentum q of the exchanged gauge boson.

Due to large coupling values of the strong coupling constant α_S at low energies, PDFs cannot be calculated perturbatively. They hence have to be determined experimentally (compare to [9] for a review).

If the PDFs for each parton are known, it is possible to calculate the cross-section for a process $pp \to X$ at a given center-of-mass energy $\sqrt{s}$ by equation (2.13), which is called the *factorization theorem*:

$$\sigma_{pp\to X}(s) = \sum_{i,j=\mathrm{Partons}} \int \mathrm{d}x_i \, \mathrm{d}x_j \, f_i\left(x_i, \mu^2\right) f_j\left(x_j, \mu^2\right) \sigma_{ij\to X}\left(\hat{s}, \mu\right) , \tag{2.13}$$

where μ denotes a *factorization scale* describing to what extent interactions between different partons inside the hadron can be neglected. $\hat{s}$ is the square of the center-of-mass energy of the individual partons:

$$\hat{s} = x_i x_j s \, . \tag{2.14}$$

Equation (2.13) can be split into two main components:

- The cross-section $\sigma_{ij\to X}$ is calculated from perturbation theory in a regime where two incoming partons i and j collide with a high momentum transfer involved (which means α_S is small enough to allow perturbative calculations).
- The PDFs $f_{i,j}\left(x_{i,j}, \mu^2\right)$ describe the probability to find the parton i (j) with a momentum fraction x_i (x_j) in proton p_1 (p_2).

2.3.3 Kinematic descriptions

This section deals with kinematic variables that are defined in high energy collider physics.

Transverse momentum p_T, missing transverse momentum $E_\mathrm{T}^\mathrm{miss}$

By convention, the z-axis in a collider experiment typically points along the beam direction. Therefore, if a particle in the final state has a momentum component in the transverse plane (the plane perpendicular to the z-axis), it is convenient to define a quantity called transverse momentum:

$$p_\mathrm{T} = \sqrt{p_x^2 + p_y^2} \, , \tag{2.15}$$

where p_x and p_y denote the respective momentum components. The total momentum in the transverse plane must vanish for momentum conservation to hold. Therefore, if the vectorial sum of all transverse momenta does not yield zero, one defines the missing transverse momentum as:

$$\vec{E}_\mathrm{T}^\mathrm{miss} = -\sum_i \vec{p}_\mathrm{T}^{\,i} \, , \tag{2.16}$$

where the summation is supposed to be performed for all detected particles. This quantity is used to estimate the transverse momentum of particles that typically would not interact with various detectors components, as it is e.g. the case for neutrinos (more information on particle signatures and reconstruction can be found in section 3.3).

Rapidity and pseudorapidity

In a particle collider such as the LHC, typically the center-of-mass of the beams is considered. However as mentioned in the parton model description, interactions are limited to a fraction of the proton rather than the entire particle. As a consequence, the momentum in the beam direction is not necessarily the same as the momentum of the proton as a whole. Therefore it is possible to find boosted particles along the beam direction. Of particular usefulness are *Lorentz invariant* variables, quantities which do not change when changing reference frames. One such quantity invariant under a boost in z-direction, which gives information about the angular distribution of the end-state particles, is called *rapidity* y:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) , \tag{2.17}$$

where E denotes the energy of the particle and p_z its momentum along the beam direction.

Another quantity η called *pseudorapidity* is defined as:

$$\eta = - \ln \left(\tan \frac{\theta}{2} \right) , \tag{2.18}$$

where θ is the polar angle between the beam axis and the direction the particle is traveling. It can be shown that η asymptotically approaches the rapidity in the limit of high energies.

It is also convenient to define an angular distance ΔR between two different particles. It is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} . \tag{2.19}$$

$\Delta\phi$ is the difference in the azimuthal angle between two different particles. Analogously $\Delta\eta$ is the difference in pseudorapidity between two particles.

2.3.4 Pileup

When two beams collide, more particles are created than the ones that an analysis might be interested in. Pileup (PU) is the technical term for events originating from primary interactions within the same bunch crossing or secondary interactions that mimic primary interactions. The former is referred to as in-time PU. Out-of-time PU describes the phenomenon when particles from earlier or later bunch crossings leave electronic signals inside the detector that might be mistaken for an actually interesting signature if not taken care of.

The amount of PU events is often quantified by a quantity μ describing the average number of interactions per bunch crossing. The pileup profiles recorded by the ATLAS detector during Run 2 (general information about the LHC and the ATLAS detector can be found in chapter 3) are displayed in figure 2.3. The mean number of interactions during this time interval was $\langle\mu\rangle \approx 33.7$.

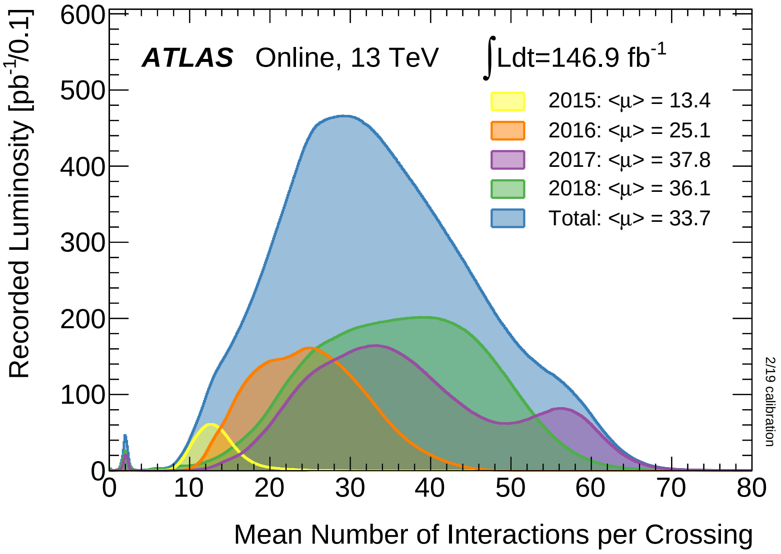


Figure 2.3: Pileup profiles during Run 2 of LHC operation, displayed for individual years, as well as a total distribution. The mean number of interactions for Run 2 was $\langle\mu\rangle \approx 33.7$. Figure taken from [10].

2.4 Top-quark physics

Initially predicted together with the b-quark in the 1970's in order to explain CP-violation in kaon-decays, the top-quark was discovered in 1995 at Tevatron. [11, 12] The top-quark has properties that make it especially interesting in comparison to other quarks.

2.4.1 General properties

The top-quark, t, is the third generation counterpart to the u-quark. It has therefore similar properties: [4] an electric charge of $+2/3$, a spin of $1/2$, and a third component of the weak isospin of $+1/2$. It is the weak isospin partner of the b-quark, and has a mass of $m_t = (173.0 \pm 0.4)$ GeV. [4] The high mass is a direct measure for its coupling to the Higgs field. Additionally, a large decay width of $\Gamma = 1.41^{+0.19}_{-0.15}$ GeV corresponds to a small lifetime of $\tau \approx 5 \times 10^{-25}$ s. [4] As a consequence, the top-quark decays before forming bound states. It differs in that way from the other quarks, being the only quark with a lifetime shorter than the typical hadronization time.

2.4.2 Top-quark production mechanisms

In proton-proton colliders like the LHC (see section 3.1), there are two main production mechanisms for top-quarks. They can either be created as a top-anti-top pair via the strong interaction, or as a single top quark as a result of weak interactions. In this thesis, the case of a top-quark being created weakly will be covered, while the creation of top-anti-top pairs will constitute one of the major backgrounds of the analysis.

Strong production modes

Top-quarks can be strongly produced either through quark-anti-quark annihilation ($q\bar{q} \rightarrow t\bar{t}$, diagram 2.4(a)) or through processes induced directly from gluon-gluon interactions ($gg \rightarrow t\bar{t}$, diagrams 2.4(b) and 2.4(c)). Which of these processes is the dominant one depends on the PDFs, the center-of-mass energies investigated and the type of particles that are colliding. At the energies reached during Run 1 and Run 2 of the LHC, approximately 90 % of $t\bar{t}$ events originated from gluon-gluon fusion processes.

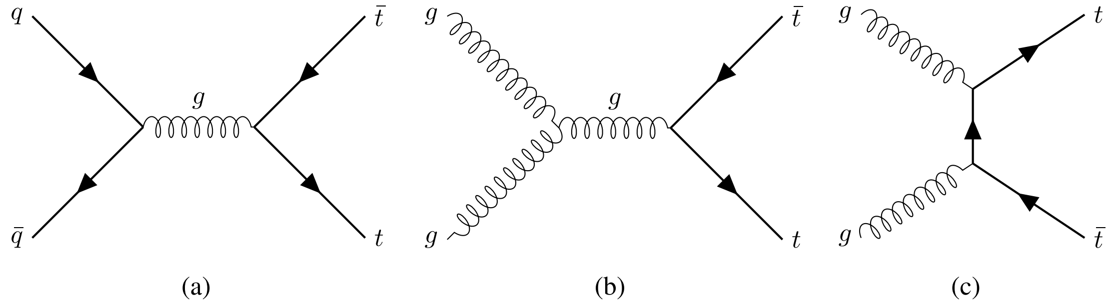


Figure 2.4: Different LO production channels for the creation of a $t\bar{t}$-pair. Diagram 2.4(a) depicts the process $q\bar{q} \rightarrow t\bar{t}$, while diagrams 2.4(b) and 2.4(c) are gluon-gluon induced processes.

The cross-section for $t\bar{t}$ production has experimentally been determined e.g. by the ATLAS collaboration: [13]

$$\sigma_{t\bar{t}} = 818 \pm 8\,(\text{stat.}) \ \pm 27\,(\text{syst.}) \ \pm 19\,(\text{lumi.}) \ \pm 12\,(\text{beam})\,\text{pb} \ . \tag{2.20}$$

Weak production modes

When the top-quark is produced alone via the weak interaction it is typically referred to as single top-quark production. LO Feynman diagrams are shown in figure 2.5. Distinctions are made between three LO channels, which are split according to the virtuality[2] of the W boson involved:

- The dominant contribution (≈ 70% for the electroweak channel) is the t-channel process. This process is shown in figure 2.5(a). A virtual W boson as exchange particle leads to the transformation of a b- into a top-quark, and from a spectator quark q to the transformation into another quark q' as well. Several studies on this process have been done within the ATLAS collaboration, with the most recent results: [14]

$$\begin{aligned}\sigma_{tq} &= 156 \pm 5\,(\text{stat.}) \ \pm 27\,(\text{syst.}) \ \pm 3\,(\text{lumi.})\,\text{pb} \\ \sigma_{\bar{t}q} &= 91 \pm 4\,(\text{stat.}) \ \pm 18\,(\text{syst.}) \ \pm 2\,(\text{lumi.})\,\text{pb}\end{aligned} \tag{2.21}$$

- The second largest electroweak contribution is depicted in figure 2.5(b). In that scenario, a single top-quark is produced in association with a real W boson. This process also has been discovered by the ATLAS collaboration. [15] The cross-section measured combining results for

---

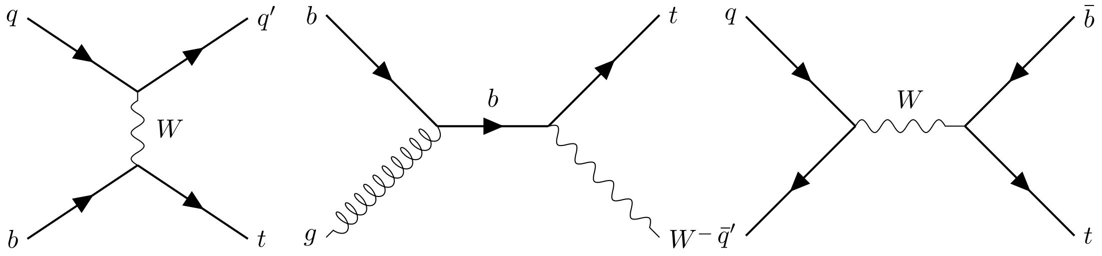
[2] Virtuality refers to off- or on-shell (with on-shell being real) particles in this case.

top- and anti-top-quark is: [15]

$$\sigma_{tW} = 94 \pm 10\,(\text{stat.})\,^{+28}_{-22}\,(\text{syst.})\ \pm 2\,(\text{lumi.})\,\text{pb}\ . \tag{2.22}$$

- The s-channel diagram has the diagram with the smallest predicted cross-section for electroweak top-quark production. It is displayed in figure 2.5(c). Evidence was reported in [16]. Theoretical predictions for the s-channel process for top- and anti-top-quarks are: [17]

$$\begin{aligned} \sigma^{t}_{s\text{-ch}}(m_t &= 173\,\text{GeV}, \sqrt{s} = 14\,\text{TeV}) = 7.93 \pm 0.14\,(\text{scale})\,^{+0.31}_{-0.28}\,(\text{PDF})\,\text{pb}\ , \\ \sigma^{\bar{t}}_{s\text{-ch}}(m_{\bar{t}} &= 173\,\text{GeV}, \sqrt{s} = 14\,\text{TeV}) = 3.99 \pm 0.05\,(\text{scale})\,^{+0.14}_{-0.21}\,(\text{PDF})\,\text{pb}\ . \end{aligned} \tag{2.23}$$



(a) (b) (c)

Figure 2.5: Overview of the dominant LO single-top production mechanisms. Diagram 2.5(a) shows t-channel production. Diagram 2.5(b) depicts the production of a single top-quark together with a real W boson. Diagram 2.5(c) corresponds to a s-channel diagram with the smallest cross-section of the displayed diagrams.

2.4.3 Rare processes involving top-quarks

With enough data collected, analyses were and are in development to search for rare processes involving the top-quark. These are processes which have comparable or even smaller cross-sections than the ones previously mentioned:

- $t\bar{t}V$ ($V = Z, W^{\pm}$): For this process, a top-anti-top pair is created in association with a heavy gauge boson. These processes have been observed by the ATLAS collaboration with a total cross-section of: [18]

$$\begin{aligned} \sigma_{t\bar{t}Z} &= (0.9 \pm 0.3)\,\text{pb} \\ \sigma_{t\bar{t}W} &= (1.5 \pm 0.8)\,\text{pb} \end{aligned} \tag{2.24}$$

- $t\bar{t}H$: Similar to $t\bar{t}V$, a top-anti-top pair is created in association with a Higgs boson. This process has been observed by the ATLAS collaboration, with a total cross-section of: [19]

$$\sigma_{t\bar{t}H} = 670 \pm 90\,(\text{stat.})\,^{+110}_{-100}\,(\text{syst.})\,\text{fb} \tag{2.25}$$

Studies of this process are especially interesting as it gives a direct measure of the Yukawa coupling (i.e. interactions between fermions to a scalar, i.e. spin-0 boson) of the Higgs boson to the top-quark.

- *tZq*: This is a process where the top-quark is produced weakly in association with a Z boson. As this is the main topic of this thesis, it will be described in more detail in section 2.4.4.
- *tHq*: A single top-quark is produced in association with a Higgs boson. It is of particular interest as it gives a direct measure of Yukawa couplings, similar to the $t\bar{t}H$ process. The *tHq* process has not been observed so far. As it has similar final states as the *tZq* process, this thesis might partially serve as a precursor for a *tHq* analysis.

2.4.4 Single top-quark in association with a Z boson

As mentioned in section 2.4.3, the main process of interest for this thesis is the production of a single top-quark in association with a Z boson. This interaction happens via a weak t-channel process, in which the Z boson can either be radiated of any of the quark-lines or from a virtual W boson, as depicted in figures 2.6(a) to 2.6(e) (4-flavor scheme[3], LO). Due to these different possible origins of the Z boson, this process is an ideal probe to test the electroweak sector of the SM. Furthermore, processes not involving a Z boson could still potentially lead to the same final state, thus including further interference effects that need to be taken into account; an example for this is shown in figure 2.6(f).

Under the assumption that particles displayed in figure 2.6 decay leptonically and requiring an invariant mass of $m_{l^+l^-} > 30\,\mathrm{GeV}$, the theoretical prediction for the cross-section in the 5FS and including processes for top- and anti-top-quarks is: [21]

$$\sigma_{tZq}^{\text{theo.}} = 102\,\mathrm{fb}\ {}^{+5.2\,\%}_{-1.3\,\%}\ (\text{scale}) \pm 1.0\,\%\ (\text{PDF})\ . \tag{2.26}$$

The *tZq* process was discovered by the CMS collaboration, in a decay channel where the Z boson decays leptonically. The measured cross-section is: [1]

$$\sigma_{\text{CMS}}\left(pp \to tZq \to tl^+l^-q\right) = 111 \pm 13\,(\text{stat.})\ {}^{+11}_{-9}\ (\text{syst.})\,\mathrm{fb}\ , \tag{2.27}$$

corresponding to a significance of 8.2σ and in agreement with the SM prediction (2.26). The cross-section measured by the ATLAS collaboration is: [21]

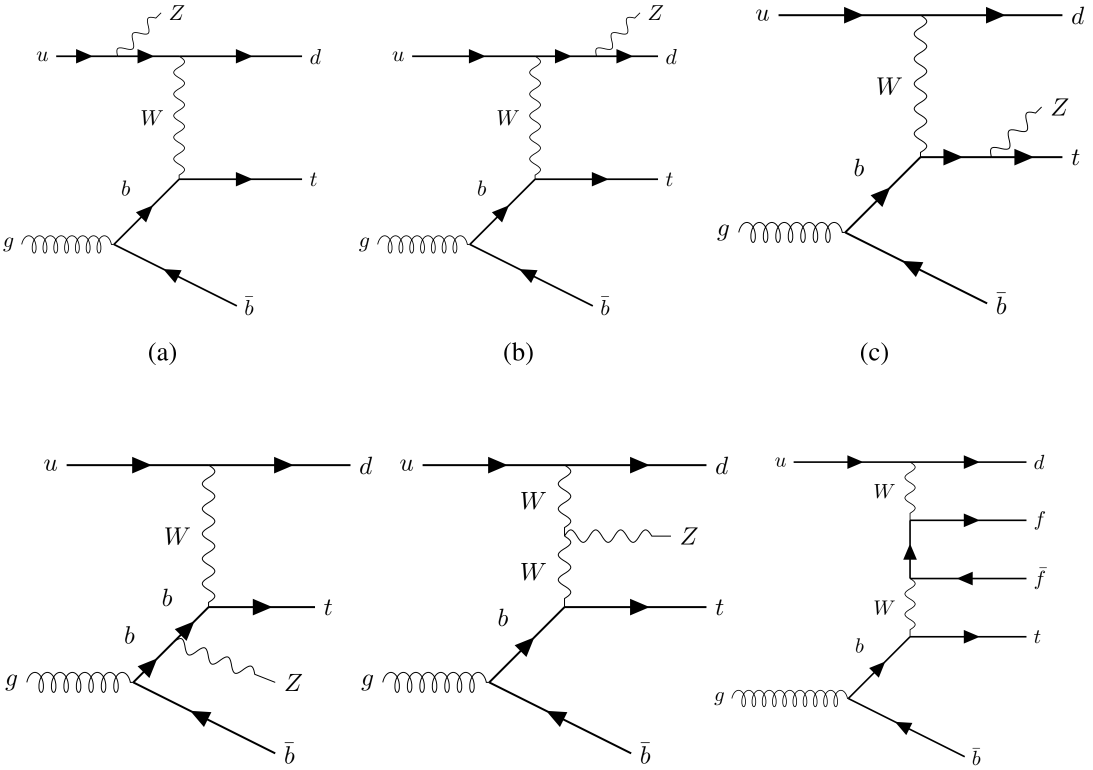
$$\sigma_{\text{ATLAS}}\left(pp \to tZq \to tl^+l^-q\right) = 97 \pm 13\,(\text{stat.}) \pm 7\,(\text{syst.})\,\mathrm{fb}\ . \tag{2.28}$$

Although these cross-sections are measured as accurately as possible with the available data, they only give a broad look at the process. This is why performing a measurement of a differential cross-section is used to get a deeper knowledge about this process.

An overview of many different processes involving the top-quark and the corresponding inclusive cross-sections are given in figure 2.7. The last process listed as *tZj* is the process denoted as *tZq* in this thesis. Its cross-section is several orders of magnitude lower than other processes involving the top-quark or processes which could lead to similar final states.

---

[3] In the 4-flavor scheme (4FS), all quarks except the b- and top-quark are assumed to be massless. Similarly, there exists a 5-flavor scheme, where the mass of the b-quark is also set to 0.



(a) (b) (c)

(d) (e) (f)

Figure 2.6: Possible LO tZq production (4FS) mechanisms, as displayed in figures 2.6(a) to 2.6(e). Figure 2.6(f) does not involve a Z boson but can lead to the same final state. As a consequence, interferences need to be taken into account.

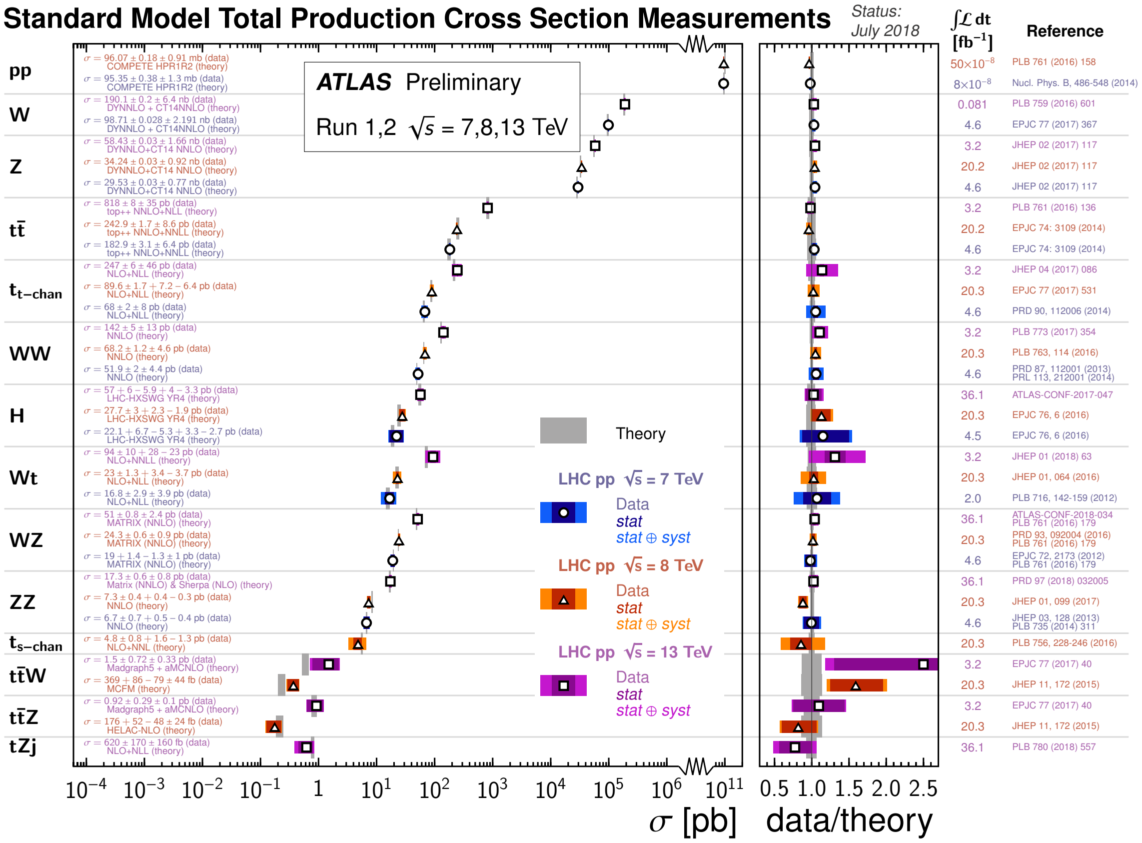


Figure 2.7: Overview of several cross-section measurements performed by the ATLAS collaboration. The central plot shows cross-section measurements for the respective process and comparison to theory. The r.h.s. plot shows the ratio for measured compared to the respective theoretical cross-section. Figure taken from [22].

CHAPTER 3

Particle production, detection and reconstruction

The Large Hadron Collider (LHC) at CERN (Conseil Européen pour la Recherche Nucléaire) is a particle collider located in the border between France and Switzerland outside Geneva. The ATLAS detector is a particle detector at the LHC which provided the data used in this thesis. This chapter first gives a general overview of the LHC and the ATLAS detector, and in the end gives more details about how particle signatures are reconstructed within the detector.

3.1 Large Hadron Collider

The Large Hadron Collider (LHC) is the most powerful hadron collider built to date ([23, 24], see [25] for a design report). It is designed to accelerate protons or heavy ions to unprecedented energies (for this analysis, data from proton-proton collisions is used). It was built in the former tunnel of the LEP collider. As such, the LHC is a circular hadron collider with a circumference of 27 km, built on average 100 m underground. The first tests were performed on 10th of September, 2008. Since November 2009, the LHC was operational for two major periods of time, called Run 1 and Run 2 respectively.[1]

3.1.1 Accelerator chain

In order to achieve energies in the range of a few TeV, several accelerating machines are needed. An overview of the main components of the accelerator complex, as illustrated in figure 3.1, is given in this section.

Protons that are injected in the LHC originate from a bottle of hydrogen gas. The hydrogen atoms are ionized in an electric field. These protons are first accelerated by a linear accelerator called “Linac2” to reach an energy of 50 MeV. Further pre-accelerators before the injection of protons into the main accelerator ring can accelerate protons to reach energies of 450 GeV. The Proton Synchrotron Booster allows to accelerate protons to energies of 1.4 GeV. It is followed by the Proton Synchrotron, where protons reach energies of 25 GeV. The last pre-accelerating complex is the Super Proton Synchrotron, where protons are accelerated to energies of 450 GeV.

---

[1] In 2019, the LHC is being upgraded during the Long-Shutdown 2, and is scheduled to be operational again by 2021: `https://lhc-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-long-term.htm`

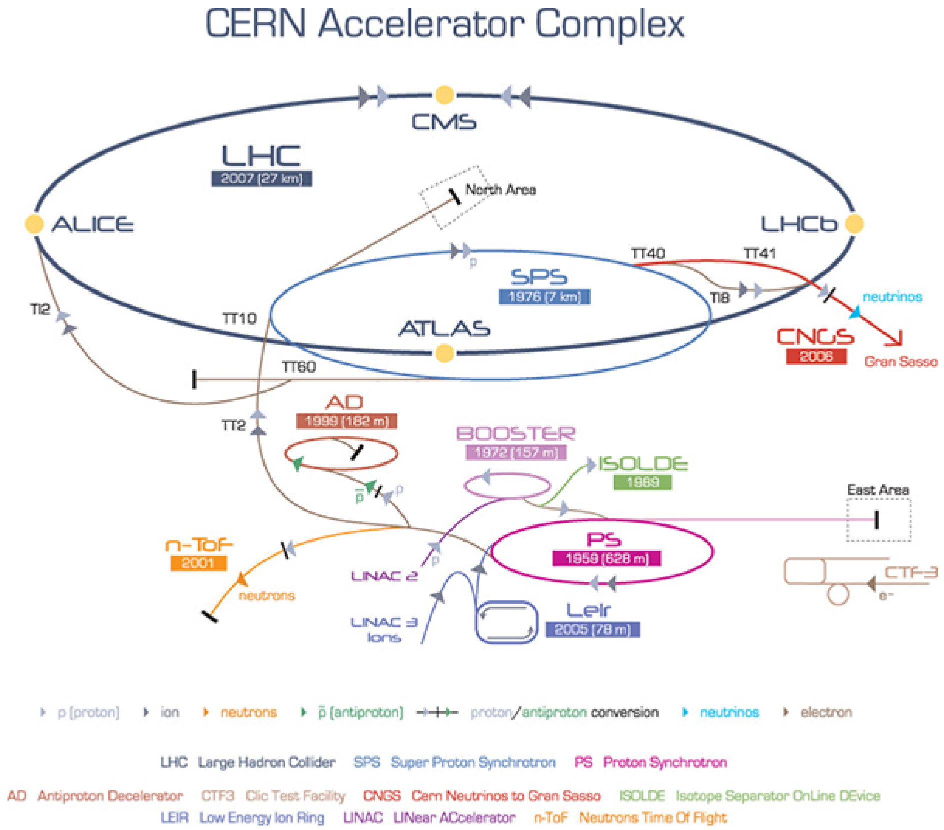


Figure 3.1: Schematic overview of the LHC and all pre-accelerators needed to guarantee proper operation. Figure taken from [23].

After reaching this energy, protons are injected into two separate beam pipes. Protons in one of these pipes circulate clock-wise, while protons in the other beam pipe travel in the opposite direction. Powerful magnets will ensure the proper operation of this circular collider: a total of 1232 dipole magnets ensure that protons stay on a circular path, while 392 quadrupole magnets are used to focus the beam. Further magnets of higher order, such as sextupoles or octapoles, are used to correct for imperfections in the magnetic field. In order to work properly, the dipole magnets need to generate a magnetic field of 8.3 T, which in turn requires a current of 11 kA. To achieve such currents without resistance losses, superconducting materials are used. These require to be cooled down to a temperature of 1.9 K, thus making the use of liquid helium necessary.

Protons are injected in the LHC in separate bunches. Each beam contains a total of 2808 bunches, each bunch containing approximately 10^{11} protons. Bunches are separated temporally by 25 ns, corresponding to a bunch crossing rate of 40 MHz.

With this setup, energies reached at the LHC during Run 2 resulted in a center-of-mass energy of 13 TeV, while also reaching the desired instantaneous luminosity of 1.0×10^{34} cm^{-2} s^{-1}. A plot of the integrated luminosity as function of time is displayed in figure 3.2. During Run 2, a total integrated luminosity of 156 fb^{-1} was delivered by the LHC; 139 fb^{-1} was deemed good enough for physics analyses.

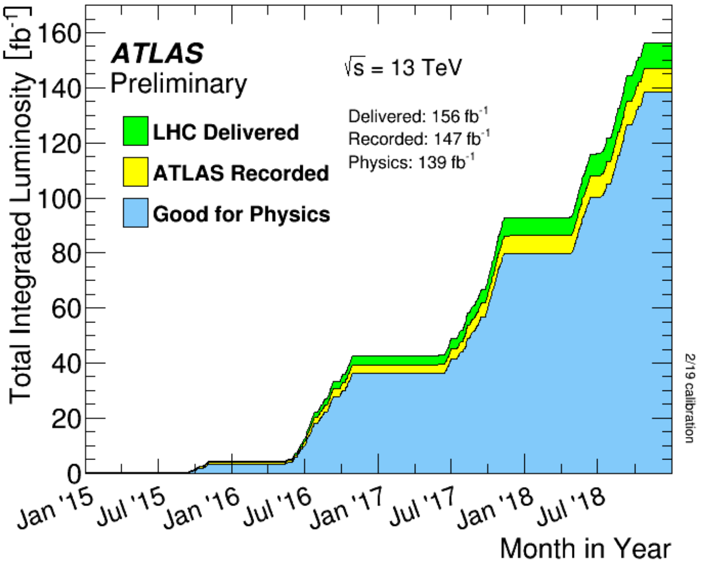


Figure 3.2: Integrated luminosity as function of time. In green the total luminosity as delivered by LHC, in blue the luminosity recorded by ATLAS and deemed good enough for physics analysis. Figure taken from [26].

3.1.2 Experiments at the LHC

Protons inside the LHC collide at four designated interaction points. To detect particles created during these collisions, specialized detectors have been built around these interaction points: the ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) detector and the CMS (**C**ompact **M**uon **S**olenoid) detector are the two general-purpose detectors at LHC. After the discovery of the Higgs boson, they are now used to search for SM and BSM processes at high energies. The main advantage of two independent general-purpose detectors is to enable validation of physics results obtained by one of the two collaborations by the other collaboration. The other two main experiments at the LHC are ALICE (**A** **L**arge **I**on **C**ollider **E**xperiment), where quark-gluon plasma generated in heavy ion collisions is studied, and the LHCb experiment (**L**arge **H**adron **C**ollider **b**eauty), which studies the properties of particles involving a b-quark in order to find asymmetries between particles and anti-particles.

Data used for this thesis was provided by the ATLAS collaboration. Therefore, the next two sections describe the design of the ATLAS detector itself as well as how particles are reconstructed with it.

3.2 The ATLAS detector

The ATLAS detector is a multi-purpose particle detector in form of a barrel (thus having cylindrical symmetry), located at Interaction Point 1, with a total length of 46 m, a height of 25 m and a total weight of 7 000 t. The coordinate system used in ATLAS is the same as described in section 2.3.3, where the x-direction is defined to be pointing towards the center of the LHC accelerator ring, the y-axis pointing upward being perpendicular to the x-direction.

The ATLAS detector consists of several layers of sub-detectors or other material; each serving a specific function in order to detect one or more properties of the particles permeating them. These layers can roughly be grouped into 4 categories: the inner detector (ID), the calorimeter system, the muon chambers and the magnet system. A schematic overview is displayed in figure 3.3. A detailed description about the general working principles of such detector components can be found in [27]. Information concerning individual sub-components is taken from [28], unless stated otherwise.

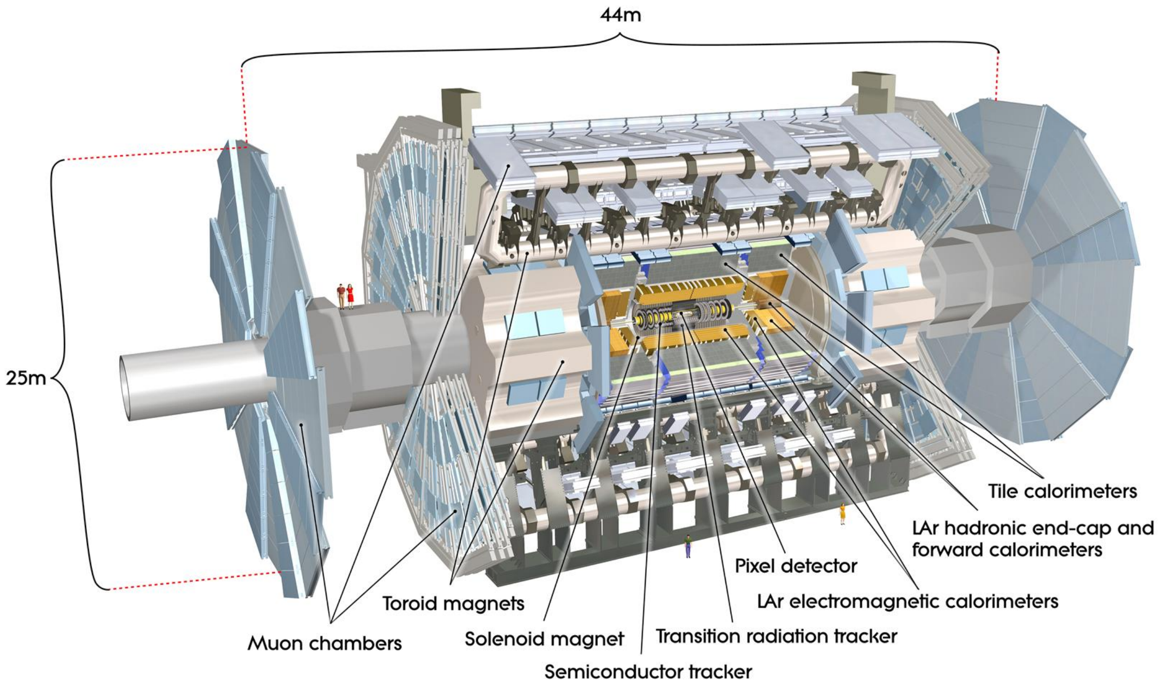


Figure 3.3: Schematic overview of the ATLAS detector. Figure taken from [28].

3.2.1 Inner detector

The task of the ID as the innermost component of the ATLAS detector system is to:

- reconstruct the primary and secondary vertices. The primary vertex is the place where the initial interaction took place, whereas secondary vertices correspond to places where particles produced in the initial interaction decay,
- measure the momentum of charged particles. For this purpose, the ID is provided with a magnetic field of 2 T along the z-direction.

To perform these tasks in the best possible way, the ID is split into three different sub-components: the pixel detector, the semiconductor tracker (SCT) and the transition radiation tracker (TRT). A particle detection by any of these subcomponents is in first instance referred to as a *hit*. Several consecutive hits may be combined to a *track* (more details can be found in section 3.3.1). The ID covers a range of $|\eta| < 2.5$.

Pixel detector

The pixel detector is the component of the ID closest to the beam pipe, distanced as close as 33 mm [29] from the beam line. Its components are four layers of silicon pixel sensors in the central part of the detector and three layers at the endcaps. The silicon pixel sensors have a dimensionality of $50 \times 400\,\mu\text{m}^2$, except for those sensors in the Insertable B-Layer (IBL; this layer was inserted in the ID of ATLAS for Run 2, see [30] for a review) which have a dimension of $50 \times 250\,\mu\text{m}^2$. [29]

Semiconductor tracker

The next tracking device of the ID is the SCT. [31] It consists of four concentric layers of silicon sensor modules in the barrel region and a total of 18 disks of silicon sensor modules in the endcaps. These sensor modules are composed of four silicon strip sensors. 2112 of these modules are used in the central part of the SCT, while 988 modules are used for each endcap. The spatial resolution realized by the SCT is 17 µm in the radial direction, while resolutions of 580 µm can be achieved along the z-direction.

Transition radiation tracker

The outermost component of the ID is the TRT. The TRT, with a total volume of $12\,\text{m}^3$, is composed of straw-tubes. These are small drift-chambers of 4 mm diameter and a total length of 144 cm (39 cm) in the barrel (endcap) region, containing a gold-plated tungsten signal wire of 0.03 mm diameter. 52 544 of these straw tubes are used in the central part of the TRT, while 122 880 straw tubes are used in each end-cap. [32]

The working principle of the TRT is the following: [27] High energy particles ionize the gas mixture contained within the straw tubes (the gas in the TRT of ATLAS consists of 70 % Xe, 27 % CO_2 and 3 % O_2, [32]). Ionized gas particles are accelerated to the anode as a high negative voltage is applied to the straw tube walls. During this process, if the primary electrons originating from those ionized gas atoms have gained enough energy, they can ionize further atoms, creating a cascade which leads to a detectable signal. Furthermore, using the effect of transition radiation occurring when a charged particle passes from one medium to another medium with different dielectric constant, the TRT can be helpful for particle identification.

3.2.2 Calorimeters

After the ID, the next layers are a magnet system (which is described further in section 3.2.4) and the calorimeter system. The general purpose of a calorimeter in high energy physics is to determine the energy of particles by exploiting the concept of particle showers: [27] once particles interact with the calorimeter material, a shower of less energetic particles is produced. These particles in turn interact with the calorimeter and shower similarly until all particles are finally absorbed. It can be shown that the maximum number of particles produced in the particle shower is proportional to the energy of the initial particle. In addition, information obtained from calorimeters is used for particle identification. Distinctions are made between electromagnetic calorimeters (ECAL) and hadron calorimeters (HCAL). This distinction depends on the main interaction process with which the aforementioned particle showers are created and therefore which particle energies are the main targets to be analyzed.

ECAL

An ECAL's main purpose is to measure the energy of particles interacting predominantly via electromagnetic interactions with the calorimeter material. These are typically electrons (via bremsstrahlung) and photons (mainly via electron-positron production). The particle shower in an ECAL is created by alternating successions of bremsstrahlung and pair-production processes. The effectiveness of this process in the ECAL material is described by the radiation length X_0, defined by the mean distance after which an initial electron has only $1/e$ of its initial energy left.

In ATLAS, a pre-sampling detector is used to estimate the energy loss of particles before reaching the main calorimeter sections and helps in the distinction between single photons and a photon-pair originating from a π^0 decay. The ATLAS ECAL itself is designed as sampling calorimeter, with lead as absorbing material and liquid argon as active detection layers. A special accordion-like design for alternating lead and argon layers enables a full coverage in ϕ over the full $|\eta|$-range up to its maximum of 3.2 without any cracks. The ECAL is separated into a barrel section, covering a range of $|\eta| < 1.475$ with a total thickness of $\approx 24X_0$, and two endcaps covering the range $1.375 < |\eta| < 3.2$ and a thickness of $\approx 26X_0$. Furthermore, it is subdivided into sections of different granularities, where finer sections in the inner parts of the ECAL are used to achieve a higher resolution, while coarser segments in the outer parts of the ECAL can e.g. be used for a separation between hadronic and electromagnetic showers (hadronic and electromagnetic showers typically show differences in their lateral extension, possibly allowing to identify their origin even in the same calorimeter). Regions with the highest granularity are divided into segments of dimension $\Delta\eta \times \Delta\phi = 0.003 \times 0.1$.

HCAL

Whereas the ECALs focus lies on electrons and photons, an HCAL is designed to determine the energy of hadrons. Hadrons, while traversing material, either ionize atoms and/or interact strongly with the nuclei of the atoms. Similar to electromagnetic showers, a hadronic shower occurs when an initial hadron leads to several secondary hadrons which in turn can interact strongly, resulting in a cascade. However, hadronic showers typically extend more than electromagnetic showers both in longitudinal and in lateral directions. Therefore, HCALs cover much more volume in a particle detector than ECALs.

The HCAL consists of several sub-components. The central section is called the Tile Calorimeter. It is composed of alternating layers of iron as passive material and plastic scintillators as active material and covering a range of $|\eta| < 1.7$. The range $1.5 < |\eta| < 3.2$ corresponds to the endcaps and is covered by another calorimeter set which uses copper as absorbing material and liquid argon as active material. A last calorimeter set, called the *Forward Calorimeter*, covers the range $3.1 < |\eta| < 4.9$. It is a region of high density, using copper in its first layer (to enable for electromagnetic showering too) while the other two layers are built of tungsten.

3.2.3 Muon spectrometer

Muons are particles that can pass the calorimeter system only losing a small fraction of their energy; other particles lose most of their energy in the calorimeters (neutrinos are a special case, they cannot be detected within the ATLAS components at all). As such, a designated system had to be installed to measure kinematic variables of the muons accurately. This is the role of the muon spectrometer (MS), which is the outermost component designed to measure the transverse momentum of muons up to 1 TeV with a precision better than 10 %.

The muon spectrometer of ATLAS is categorized into four main components: resistive plate chambers (RPCs) and thin gap chambers (TGCs) with fast time responses are used as trigger systems in the barrel and in the endcap-sections, respectively. Track reconstruction and momentum determination of muons are made possible by the use of Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs).

3.2.4 Magnet system

In order for the tracking devices to work properly, a magnetic field is provided which bends the tracks of charged particles and thus allows a momentum and charge determination via the Lorentz force.

The ID is surrounded by a superconducting solenoid magnet which is located in front of the calorimeters and is able to provide a magnetic field of 2 T along the z-direction, corresponding to a bending power of up to 8 T m. Its design was chosen to minimize the energy loss of particles traversing it in order to enable a better energy measurement by the calorimeters. The energy loss due to the ID and the solenoid magnet is estimated by the presampler, as described in section 3.2.2.

The magnetic field used for the muon spectrometer is provided by a toroidal magnet system. It is separated in a barrel and two endcap sections, each consisting of a total of eight coils. The magnetic field provided by the toroid magnet has a field strength of 4 T, which corresponds to a bending power of up to 6 T m).

3.2.5 Trigger system

As the LHC reached unprecedented energies and high luminosities especially during Run 2, it is impossible to store information of each interaction. It is the role of a trigger system to reject events which are deemed unnecessary and to only store events considered interesting for further analysis.

The ATLAS trigger system [33] used during Run 2 is split into two main parts. The Level-1 trigger is hardware-based and can reduce the event rate from 40 MHz to a rate of 100 kHz. Events that pass the Level-1 trigger are passed to the software-based High Level Trigger (HLT), which reduces the event rate even further to 1 kHz. Events that fulfill the HLT requirements are stored and can be used for different purposes: full information about the event is stored for physics analysis goals, while only some information about the event is stored for monitoring or detector calibration.

3.3 Particle reconstruction with ATLAS

Each component of the ATLAS detector has been optimized in order to identify and reconstruct various particles. This is possible as different particle types show distinctive features within the detector material. This section describes in more detail how detector signals are associated to basic properties of particles. Information from all detector components is combined to give a first consistent, low-level picture of the detected particle's basic properties; for example charge and the associated track. This information can be used to fully reconstruct particle properties further in an analysis chain. A general overview of the signatures that particle leave in the detector components is displayed in figure 3.4; details are given in the respective sections.

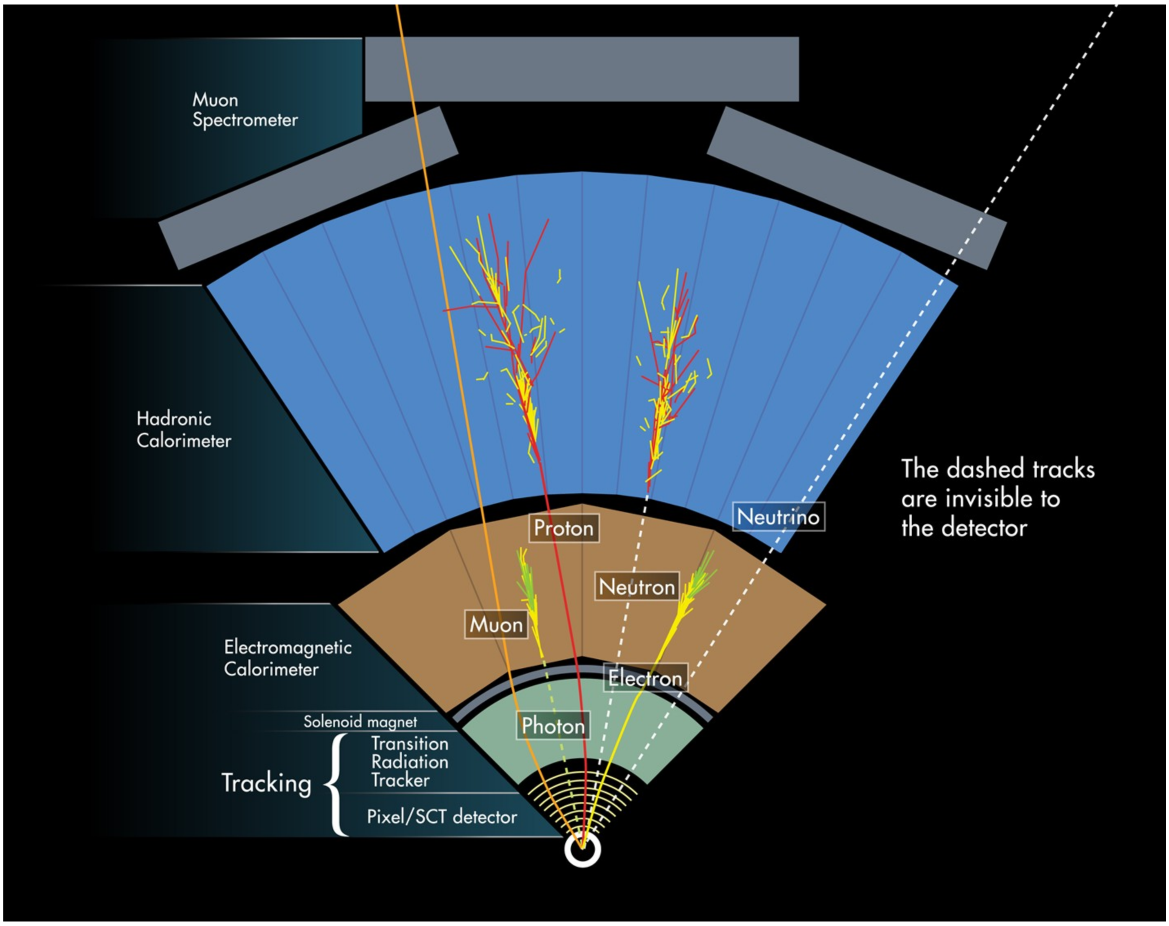


Figure 3.4: Schematic overview of typical signatures within the various ATLAS detector components that can be associated to specific particles. Figure taken from [34].

3.3.1 Low-level analysis

Before the complete reconstruction and identification of particles is possible, hits registered in the detector are combined to more useful objects which allow for particle reconstruction.

Tracks and vertices

Charged particles traversing through the ID leave a trace of consecutive hits, which are combined to tracks. These tracks give information about the charge and momentum of the particle associated to the track. Track reconstruction is possible for transverse momenta with $p_{\mathrm{T}} \geq 0.4\,\mathrm{GeV}$ with an efficiency of 70 % to 85 %. [35]

Furthermore, points in which two or more tracks meet are called vertices. The primary vertex is defined as the place where the sum of all transverse momenta squared obtained from track information takes its maximum value. Other vertices are either associated to pileup events or are labeled as secondary vertices. Secondary vertices are used to define a particle decay chain.

Calorimeter

Hits in calorimeters are combined to *clusters*. Clusters are combinations of individual calorimeter cells with energy deposits higher than the expected average noise from e.g. readout electronics. These clusters are created for both ECAL and HCAL and are used to reconstruct the energy of photons and electrons (in the ECAL) or hadrons (in the HCAL). Two main clustering algorithms are used: [36] these are the *sliding-window* and the *topological* algorithms.

For the sliding-window algorithm, energy depositions from all cells inside a rectangular frame of fixed size are added up. The position of the frame is changed until the highest value of the transverse energy (transverse energy in this context is defined as $E(\text{cluster})/\eta(\text{Associated Track})$) has been obtained. This algorithm is e.g. used for energy calibrations. For the topological cluster algorithm, a calorimeter cell is used as seed. If a neighboring cell contains a signal at least four times larger than the average noise level, this cell is added to the cluster. This procedure is then repeated for all the new cells that have been added, until the algorithm breaks down.

3.3.2 Electrons

Electrons, as the lightest electrically charged leptons, are expected to leave a track in the ID and deposit all their energy in the ECAL, as depicted in figure 3.4. The reconstruction of electrons with ATLAS requires energy depositions in the ECAL that can be related to a track in the ID. Cluster reconstruction is performed using the sliding-window approach in a range $|\eta| < 2.47$. [37] For minimum transverse energies of $E_\mathrm{T} = 15\,\mathrm{GeV}$, a reconstruction efficiency larger than 99 % can be obtained. [37] Furthermore, in order to distinguish electron from fake events, a multivariate technique exploiting information like shower shape, measured energy from the calorimeter and momentum measurement obtained from the tracking system is used to calculate a likelihood profile. Depending on the threshold value chosen, distinctions are made between *loose*, *medium* and *tight* electrons; the tight selection has the highest potential of background rejection. Additionally, a restriction on electron isolation is implied: for both tracking system and calorimeters a quantity is defined which determines quantitatively the isolation of the electron.

3.3.3 Muons

Muons as heavier counterparts of the electron are expected to leave tracks in the ID as well as in the MS. Energy losses due to bremsstrahlung are expected to be at lower levels than for electrons. Typical muon events are shown in figure 3.4. For the reconstruction of muons, tracks need to be observed in the ID as well as in the MS that can be matched with each other (i.e. tracks need to point in consistent directions, and momenta associated to both tracks need to agree). Muons reconstructed in this way are referred to as *Combined Muons*. [38] Further criteria on the track and track quality lead to the distinctions of *loose*, *medium* and *tight* muons. Isolation criteria are applied similar to the ones applied for electrons.

3.3.4 Jets

The process of hadronization, which quarks and gluons participate in, leads to structures called *jets*. These are extended objects containing mainly hadrons (and photons originating from π^0 decays), and are thus expected to deposit large amounts of their energy in the calorimeters, mainly the HCAL, as displayed in figure 3.4.

The algorithm used to reconstruct jets within the ATLAS collaboration is called anti-k_t algorithm. [39] For this procedure, a "distance" between two objects i and j is defined as:

$$d_{ij} = \min\left(\frac{1}{p_{\mathrm{T},i}^2}, \frac{1}{p_{\mathrm{T},j}^2}\right)\frac{\Delta R_y^2}{R^2}\ , \tag{3.1}$$

with the quantity $\Delta R_y^2 = \left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2$. The quantity R is used as parameter, and set to $R = 0.4$ for this analysis. A comparison distance, denoted as distance to the beam line, is defined as:

$$d_{iB} = \frac{1}{p_{\mathrm{T},i}^2}\ . \tag{3.2}$$

With these definitions, the algorithm steps are as follows: from any pair of objects, the pair with the smallest distance between them is chosen and merged into only one object. The procedure is repeated until either all objects have been merged into a single object or when for all possible distances $d_{ij} > d_{iB}$.

For the anti-k_t algorithm, jet reconstruction is performed using topological clusters. Jets built with this algorithm are required to have $p_{\mathrm{T}} > 35\,\mathrm{GeV}$ and $|\eta| < 4.5$. Furthermore, a discriminant exploiting information about the tracks, vertices and jet shapes called *Jet-Vertex-Tagger* (JVT, [40]) is used in the central region to reject pileup events. Similarly, in the forward region ($2.5 < |\eta| < 4.5$) a similar discriminant called forward Jet-Vertex-Tagger (fJVT) can be applied. [41]

3.3.5 *b*-jets

A special kind of jets needed for this analysis are called *b-tagged jets*. These are jets which originate from a hadron containing a *b*-quark. *b*-jets are especially useful as top-quarks predominantly decay into *b*-quarks, which makes them potential indicators for interactions involving top-quarks.

Hadrons containing a *b*-quark typically have, due to their weak decay chains, lifetimes long enough such that their decay vertex can be identified. Combined with the fact that their decay products typically yield large invariant masses and tertiary vertices may be related to hadrons containing *c*-quarks, [42] it is possible to introduce a multivariate discriminant with the purpose of identifying *b*-jets. [43] Cuts may then be applied to this discriminant value, which is typically specified by *working points*. A 70 % working point e.g. is defined by cutting on the discriminant value for which 70 % of jets initiated by *b*-hadrons can correctly be identified.

3.3.6 Missing transverse momentum

Not every particle interacts with the detector material. Especially neutrinos, which only interact via the weak force, are expected to leave the detector components without any measurable signal. For this purpose, the concept of missing transverse energy was introduced (see section 2.3.3). The negative vectorial sum of all transverse momenta of detected particles is usually assigned as the transverse momentum of neutrinos that might have been part of the interaction (although weakly interacting BSM particles might contribute to missing transverse momentum as well).

CHAPTER 4

Event selection, signal discrimination and extraction

An overview of the event selection for final states of interest and a summary of possible major backgrounds is provided. The used data samples and the Monte Carlo (MC) samples simulating signal and backgrounds are presented. The strategy for signal extraction is also discussed. Lastly, discrepancies between data and MC predictions after inclusion of the 2018 data set are highlighted.

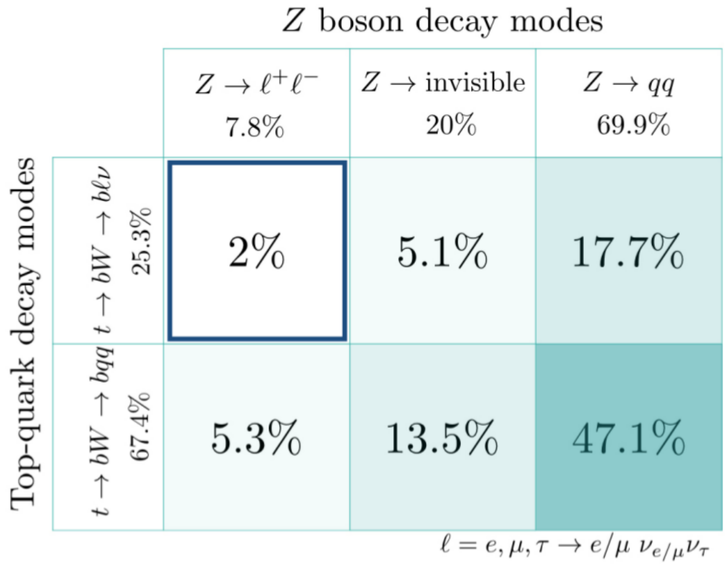
4.1 Trilepton final state of the tZq process

The final state of the tZq process investigated in this thesis is known as the trilepton final state (information described in this section can be found in [21], if not stated otherwise). As both the Z boson and the top-quark are too short-lived to be detected directly, they are reconstructed from their decay products. In the trilepton channel, some of these decay products are exactly three leptons (leptons designating electrons and muons only. τs are only considered if they decay leptonically). Although this channel has the smallest expected branching ratio ($\mathcal{B} \approx 2\%$, compare to figure 4.1), the clean signature obtained for lepton reconstruction allows this channel to have a better expected signal-to-background ratio than hadronic decays.

An example for the event topology in question is displayed in figure 4.2. The final state is expected to consist of three leptons, one neutrino, one b-tagged jet and one jet which is not b-tagged; this last jet is expected to predominantly go in the forward direction. The jet originating from the $\bar{b}$-quark shown in figure 4.2 tends to go in the same direction as the beam line and can therefore not be detected.

The Z boson is reconstructed from two leptons of same flavor, but opposite electric charge. These leptons are referred to as opposite sign, same flavor (OSSF) pairs. As three leptons are expected in the final state, four possible combinations for these leptons are possible: eee, $\mu\mu\mu$, $ee\mu$ and $\mu\mu e$. For the two latter cases, the reconstruction of the Z boson is trivial; whereas for the first two cases, the Z boson is reconstructed from the two leptons yielding an invariant mass closest to the Z boson mass. Applying restrictions to the invariant mass of the leptons is also beneficial for background rejection (see sections 4.2 and 4.4.1).

The lepton that is not associated to a Z boson is used to reconstruct the W boson originating from the top-quark decay. Missing transverse momentum is associated to the only neutrino in the event.



Z boson decay modes

Top-quark decay modes

	$Z \to \ell^+\ell^-$ 7.8%	$Z \to$ invisible 20%	$Z \to qq$ 69.9%
$t \to bW \to b\ell\nu$ 25.3%	2%	5.1%	17.7%
$t \to bW \to bqq$ 67.4%	5.3%	13.5%	47.1%

$\ell = e, \mu, \tau \to e/\mu\ \nu_{e/\mu}\nu_\tau$

Figure 4.1: Overview of branching ratios of the *tZq* process. The trilepton final state has a theoretical branching ratio of ≈ 2 %. Figure taken from [44].

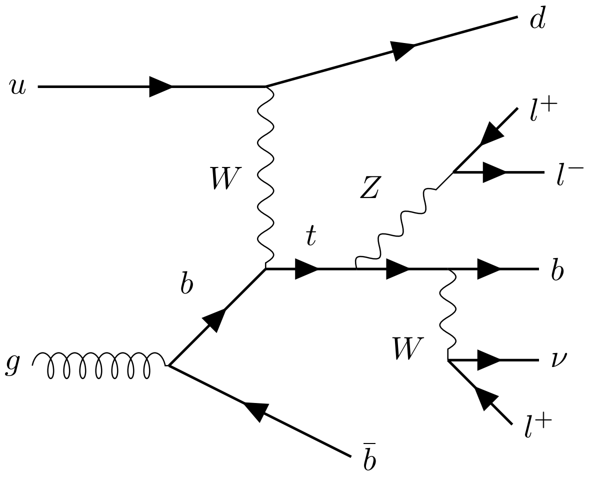


Figure 4.2: Example for a LO Feynman diagram of the *tZq* process with a fully leptonic decay chain, i.e. both Z and W bosons decay leptonically.

The full neutrino momentum is reconstructed requiring:

$$p_W^2 = (p_\ell + p_\nu)^2 = M_W^2 \, , \tag{4.1}$$

where the mass of the W boson is set to $M_W = 80.4\,\text{GeV}$. This condition allows a full reconstruction of the neutrino's kinematic distribution. Together with the information available for the lepton, this enables the estimation of kinematic properties of the W boson.

The top-quark is assumed to always decay into a b-quark, as the corresponding CKM-matrix element (see equation (2.2)) is approximately 1. Therefore, it can be fully reconstructed from the previously reconstructed W boson and the information available on a b-tagged jet.

4.2 Backgrounds

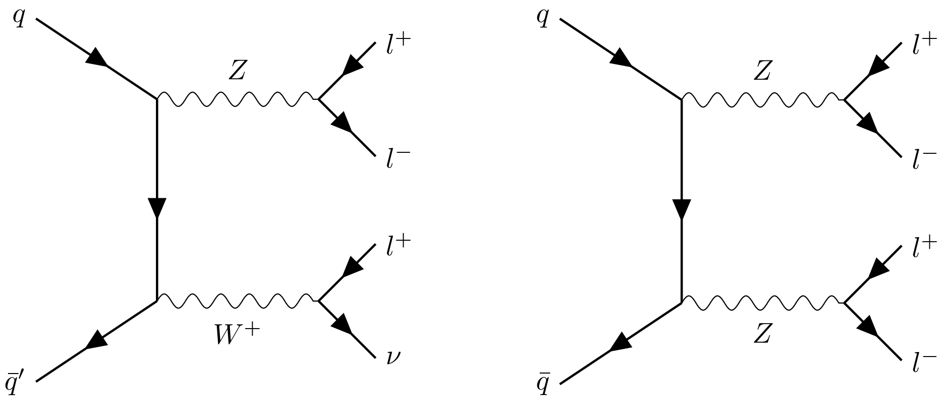
The trilepton final state is difficult to reconstruct due to contamination from several background processes. These have similar event topologies to a tZq decay. The backgrounds can be separated whether three prompt leptons are involved or at least one non-prompt or fake lepton is required to mimic the same final state. Prompt leptons originate from the decay of the particles involved in the process of interest, in this case coming from the decay of a Z boson or a W boson, which in turn originates from the top-quark decay. Non-prompt leptons denote leptons originating from the decays of hadrons containing either b- or c-quarks, leptons from the decay of kaons or pions or leptons from photon pair conversions. Fake leptons are not leptonic in nature, but are misidentified jets. Major backgrounds covering all these possibilities are discussed in the next sections, following the description given in [44].

4.2.1 Backgrounds involving real leptons

These backgrounds involve three prompt leptons. Their contribution is estimated from MC samples.

Diboson

Diboson processes are events where two bosons are produced. LO Feynman diagrams are displayed in figure 4.3. The WZ process final state, as displayed in diagram 4.3(a), can consist of three real leptons (if both bosons decay leptonically) and thus mimic the final state of the tZq process if additional jets are produced. On the other hand, the ZZ process, as shown in diagram 4.3(b), contains, for leptonically decaying bosons, four instead of three real charged leptons. Therefore, this process represents a background for the tZq process only if one of the leptons fails to be reconstructed or is outside the acceptance range in addition to separate jet production, thus being a more unlikely background than the WZ process.



(a) (b)

Figure 4.3: Two examples for diboson events, with leptonically decaying bosons. Diagram 4.3(a) corresponds to a WZ process, diagram 4.3(b) to a ZZ process.

$t\bar{t}V$, $t\bar{t}H$

$t\bar{t}V$ events consist of a top-antitop-pair produced in association with a massive gauge boson. A possible event topology is displayed in the LO Feynman diagram in figure 4.4. The final state has a similar final state to the tZq process, involving a Z boson and top-quarks, although expected to yield two b-tagged jets instead of only one. However, the case of an unidentified b-jet at the reconstruction stage can mimic the desired event topology.

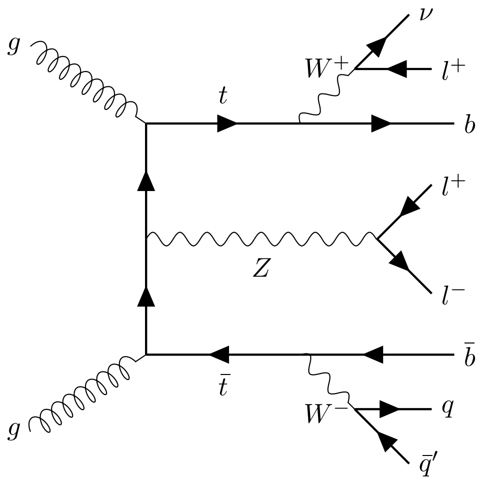


Figure 4.4: Example of a $t\bar{t}V$ process, with $V = Z$, and a final state with three real charged leptons.

$t\bar{t}H$ production can be represented by a similar Feynman diagram to the diagram displayed in figure 4.4 by replacing the Z boson by a Higgs boson. However, the contribution is expected to be considerably smaller than $t\bar{t}V$production due to the low branching ratio of a leptonically decaying Higgs boson.

tWZ

tWZ production is a potential background which has not been observed yet, but is predicted to occur with a cross-section of $\sigma = 12.2\,\text{fb}$. [45] A possible NLO Feynman diagram is displayed in figure 4.5. Similar to the tZq process, it can result in three real leptons originating from Z and W bosons. The main expected difference is a higher jet multiplicity in the tWZ process.

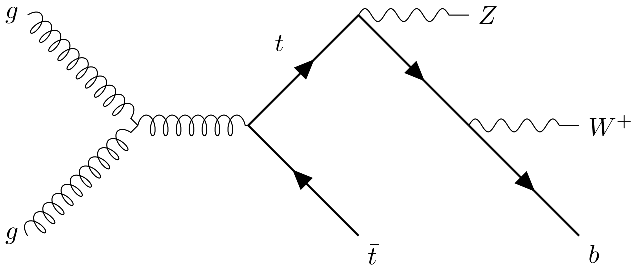


Figure 4.5: NLO Feynman diagram for a tWZ process.

4.2.2 Backgrounds involving non-prompt or fake leptons

The background processes described in this section require at least one non-prompt or fake lepton to mimic the *tZq* final state. Their contribution is estimated with a technique called *b-jet replacement method*, which is briefly described in section 4.4.2.

$t\bar{t}$

$t\bar{t}$ production, although not even having three real leptons in the final state, represents one of the major backgrounds to this analysis due to a higher cross-5section compared to *tZq* process (see figure 2.7). An example for a LO Feynman diagram is shown in figure 4.6. Three leptons in the final state are only possible if one of the leptons originates from a *b*-hadron decay.

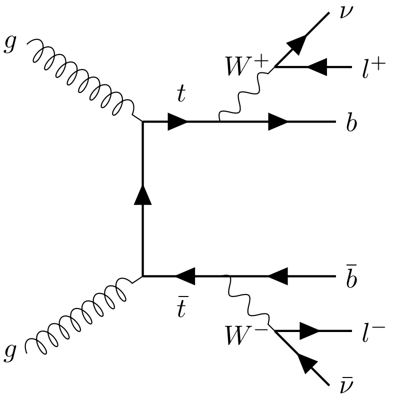


Figure 4.6: Example of a LO $t\bar{t}$ Feynman diagram with leptonically decaying *W* bosons.

Z + jets

Another potentially large background for the *tZq* analysis which does not contain 3 real leptons is the production of a *Z* boson accompanied by jets. An example for a corresponding Feynman diagram is displayed in figure 4.7. Similar to $t\bar{t}$, this process can mimic the *tZq* final state only if one of the leptons is fake or non-prompt.

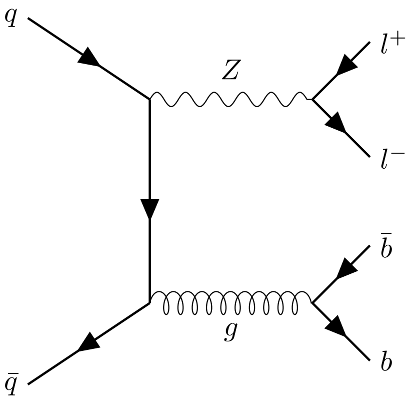


Figure 4.7: Example of a LO Feynman diagram showing a *Z*+jets events, with a leptonically decaying *Z* boson.

4.3 Datasets

After the discussion of the signal signature and possible background processes, a brief overview of the samples used is given in this section.

For both data and Monte Carlo (MC: MC is used to simulate various processes, see section 4.3.2) single-top ntuples [46, 47] form the basis for this analysis. Version 28 ntuples produced with `AnalysisTop-21.2.58` [48] have been used for this analysis (`AnalysisTop` is the framework used in the ATLAS Top Working Group to generate ntuples). The only exception is a signal sample used for the differential cross-section measurement itself: here, a MC sample is needed that contains information about particles both at reconstruction and truth-level (see section 5.2 for the definition of "reconstruction-" and "truth-level" variables). Therefore, an additional MC signal sample using `AnalysisTop-21.2.65` has been created and used for all differential cross-section measurements performed during this thesis.

4.3.1 Data samples

The data used for this analysis was collected from 2015 to 2018 with the ATLAS detector at a center-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$. The data selected for this analysis was required to be taken during stable LHC conditions and full functionality of the ATLAS detector being guaranteed (data fulfilling these conditions are collected in "Good Run Lists"). The total amount of data satisfying these conditions corresponds to a total integrated luminosity of $139\,\text{fb}^{-1}$. The integrated luminosities of data collected in individual years are summarized in table 4.1:

Year	Int. Lumi. (fb^{-1})
2015	3.2
2016	33.0
2017	44.3
2018	58.5

Table 4.1: Integrated luminosities per year of LHC Run 2 operation. Table adapted from [20].

4.3.2 Monte Carlo samples

MC samples in high-energy physics are generated using stochastic methods to simulate various processes. They can e.g. be used to predict event yields of a signal process, to estimate contributions from backgrounds processes, and to investigate final signatures in the detecting device. Predictions from MC samples can then be compared to results obtained from real data.

There are several steps of simulation that are taken into account. [49] The first level is the generation of the underlying event. Further simulation software is needed to account for parton showering, i.e. the hadronization process, as well to simulate similar pileup conditions as during actual data taking. Simulations of how all these particles are detected by the ATLAS detector are generally produced with a software called `GEANT4`. [50]

MC samples simulating Run 2 conditions are split into three categories, labeled mc16a, mc16d and mc16e. Pileup conditions simulated for mc16a samples reflect pileup conditions during data taking in 2015 and 2016. Similarly, mc16d samples are generated assuming similar pileup conditions as during 2017 data taking, while mc16e reflects the pileup conditions from 2018. As a consequence, the samples need to be scaled to the corresponding integrated luminosities, see table 4.1.

Furthermore, some physical quantities needed to be defined for the simulation process. The mass of the top-quark is set to $m_t = 172.5\,\text{GeV}$. Top-quarks are assumed to always decay into a W boson and a b-quark. τ-leptons are relevant for the analysis only if they decay leptonically.

A general overview of the simulated samples used for signal events and backgrounds is given in the respective following sections. Further details are specified in appendix A.

Signal sample

The simulation of a tZq sample (DSID:412063) is performed with the `MadGraph5_aMC@NLO` v2.3.3 generator [51] at NLO using the `NNPDF3.0NLO` [52] parton distribution function (PDF). Technically, the sample is produced to simulate $t\ell^+\ell^-q$ events, to account for interferences due to processes not involving Z bosons (see figure 2.6(f)). Off-resonant events (i.e. $m_{\ell\ell} < m_Z$ or $m_{\ell\ell} > m_Z$) are kept as long as the invariant mass of leptons fulfills $m_{\ell\ell} > 30\,\text{GeV}$. Simulations are performed using the four-flavor scheme (4FS), in which all quark masses are set to zero except for the b- and top-quark masses. The resulting distributions are rescaled to yield the five-flavor scheme (5FS) at NLO prediction of $\sigma = 102\,\text{fb}$ for the cross-section. [21]

Background samples

A summary of the simulators used for the different background processes is given in table 4.2 (a complete description can be found in appendix A). Main simulation software used were `MadGraph5_aMC@NLO`, [51] `PowhegBox`, [53] `Pythia` 8 [54] and `Sherpa`. [55]

Process	MC Generator	Parton Showering
$t\bar{t}$	`PowhegBox` v2	`Pythia` 8.230
Single-top	`PowhegBox` v2	`Pythia` 8.230
$t\bar{t}H$	`PowhegBox` v2	`Pythia` 8.230
$t\bar{t}V$	`MadGraph5_aMC@NLO` v2.3.3	`Pythia` 8.210
tWZ	`MadGraph5_aMC@NLO` v2.3.3	`Pythia` 8.212
Diboson	`Sherpa` v2.2	`Sherpa`
Z + jets	`Sherpa` v2.2	`Sherpa`

Table 4.2: Overview of generators used to generate background samples and simulate the showering process. Information taken from [21].

MC weights

Events generated by MC may not reproduce the actual conditions during which data was taken. Furthermore, selection efficiencies may be different for for different physics objects. To account for these effects, MC generated events have to be weighted. The weight that is assigned to a MC event can be written as: [46]

$$w_{\text{event}} = w_{\text{MC}} \cdot w_{\text{pileup}} \cdot w_{\text{lepton}} \cdot w_{\text{JVT}} \cdot w_{\text{trigger}} \cdot w_{b\text{-tagging}} \cdot \tag{4.2}$$

The individual weights in this product are (information taken from [44]):

- w_{MC}. The sum of w_{MC} from all generated events results in the number of events for the generated sample.
- w_{pileup} corrects for the assumed pileup conditions to match the measured pileup conditions during data-taking.
- w_{lepton} accounts for the efficiencies of lepton identification, and for lepton isolation criteria.
- w_{JVT} accounts for differences in data and MC efficiencies when cutting on the output of the JVT.
- w_{trigger} accounts for differences in data and MC efficiencies due to trigger selections.
- $w_{b\text{-tagging}}$ accounts for differences in data and MC efficiencies due to the necessity of a b-tagged jet and b-tagging efficiency.

Note that some of these weights are related to simulation only, but others, like w_{lepton}, are also related to detector or reconstruction effects. This fact will be discussed further in section 5.3.1.

The total weight for each event requires a readjustment of the w_{event} from equation (4.2), as it needs to be rescaled to match the correct integrated luminosity such that similar cross-sections between simulation and data are expected. This is done by multiplying the w_{event} with another weight to simulate the total amount of data collected. This weight is called luminosity weight, defined as $w_{\mathrm{lumi}} = \sigma_{\mathrm{process}}\mathcal{L}/N_0$, with N_0 the number of events from the original MC sample. The total weight assigned to each MC event can then be written as:

$$w_{\mathrm{total}} = w_{\mathrm{event}} \cdot w_{\mathrm{lumi}} \, . \tag{4.3}$$

4.4 Analysis strategy

An overview of the analysis strategy and framework developed for the inclusive cross-section measurement is given in this section. These concepts are important as they serve as basis for the differential cross-section measurement. First, the distinct signal and control regions are described. The signal discrimination and fitting procedures are explained afterwards.

4.4.1 Definition of signal and control regions

The investigated phase space is split into separate signal and control regions. Signal regions (SRs) should be enriched in signal (tZq) events. Similar cuts as those applied to SRs are applied to control regions (CRs) such that they are perpendicular to the SRs. CRs are designed to be enriched in background events. Both SRs and CRs are used for the extraction of signal event yields; a summary of how these regions differ is given in figure 4.8. Common selection criteria to all SRs and CRs are:

- Exactly three leptons are required. The lepton with highest p_{T} is required to have $p_{\mathrm{T}}(\ell_1) > 28\,\mathrm{GeV}$, for the other leptons the relations $p_{\mathrm{T}}(\ell_2) > 20\,\mathrm{GeV}$ and $p_{\mathrm{T}}(\ell_3) > 20\,\mathrm{GeV}$ need to be satisfied. All leptons need to satisfy $|\eta(\ell)| < 2.5$.

- Jets are required to have $p_{\mathrm{T}} > 35\,\mathrm{GeV}$ and $|\eta(\mathrm{jet})| < 4.5$. High values for $|\eta|$ account for at least one untagged[1] jet in forward direction.
- b-jets are required to pass the 70 % working point requirement. Furthermore, a stricter cut on η compared to untagged jets is implemented: $|\eta(b\text{-jet})| < 2.5$.
- The untagged jet in an event with the highest invariant mass m_{jb} between the untagged and the b-jet is called *forward jet*. In the case that no b-jet is defined, the forward jet is defined as the jet with highest $|\eta|$.

SRs

Two SRs enriched with tZq events have been defined, which are denoted as SR-$2j1b$ and SR-$3j1b$ in the following. In addition to the cuts applied to all regions, further cuts have been defined for SRs:

- At least one OSSF pair needs to be reconstructed satisfying $|m_{\ell\ell} - m_Z| < 10\,\mathrm{GeV}$. In the case that multiple combinations of leptons satisfy these conditions, the leptons whose invariant mass is closest to the Z boson mass is selected for the Z boson reconstruction.
- Exactly one b-tagged jet is required. The total number of jets is limited to either 2 or 3 jets: SR-$2j1b$ contains events with exactly 2 jets, of which exactly one is b-tagged; for SR-$3j1b$, 3 jets in total are needed of which exactly one is b-tagged.

CRs

In total, six CRs are defined. The CRs are designed to address background contributions from diboson, $t\bar{t}$ and $t\bar{t}V$ events.

- CRs to control diboson events are labelled as CR-$2j0b$ and CR-$3j0b$, following the nomenclature used for SRs. CR-$2j0b$ and CR-$3j0b$ have the following additional selection criteria compared to the common selection criteria:
  - Identically to SRs, one OSSF pair satisfying $|m_{\ell\ell} - m_Z| < 10\,\mathrm{GeV}$ is required.
  - 2 (3) jets are required for CR-$2j0b$ (CR-$3j0b$). In contrast to the SRs, none of these jets is allowed to be b-tagged.
- CRs to control $t\bar{t}V$events are labelled as CR-$3j2b$ and CR-$4j2b$. CR-$3j2b$ and CR-$4j2b$ have the following additional selection criteria compared to the common selection criteria:
  - Identically to SRs, one OSSF pair satisfying $|m_{\ell\ell} - m_Z| < 10\,\mathrm{GeV}$ is required.
  - 3 (4) jets are required for CR-$3j2b$ (CR-$4j2b$). In contrast to the SRs, 2 of these jets need to be b-tagged.
- CRs to control $t\bar{t}$ events are labelled as CR-$2j1b$ and CR-$3j1b$. CR-$2j1b$ and CR-$3j1b$ have the following additional selection criteria compared to the common selection criteria:

---

[1] “Untagged” means the jet is not b-tagged.

- For better discrimination from events containing a Z boson, these CRs are not allowed to have a OSSF pair. Instead, they are required to have at least one possible combination for an opposite sign, different flavor (OSDF) pair.
- Identically to SRs, 2 (3) jets are required for CR-$2j1b$ (CR-$3j1b$), of which one jet needs to be b-tagged.

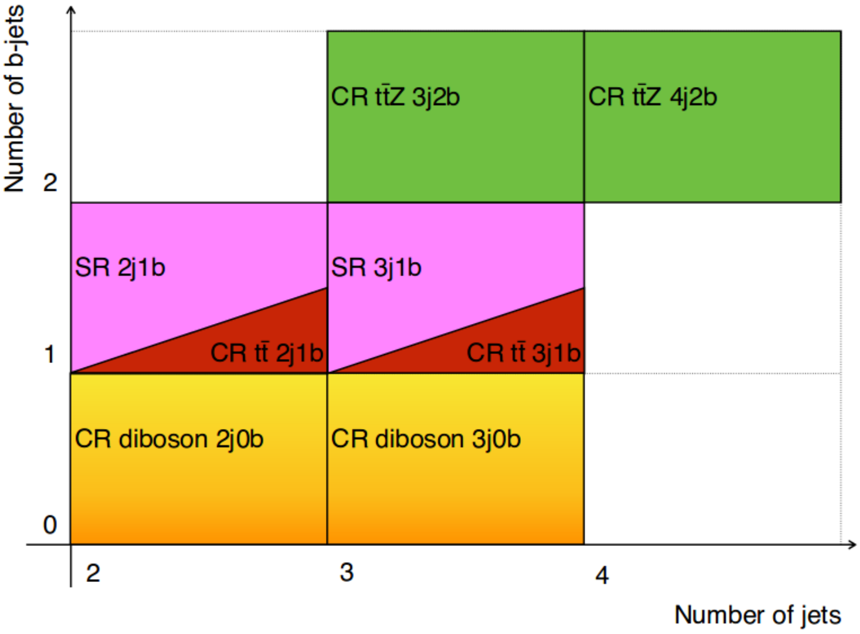


Figure 4.8: Summary of how different SRs and CRs defined for the inclusive tZq analysis differ in jet multiplicity. The SRs and $t\bar{t}$ CRs show the same jet multiplicity, but show distinctions concerning the reconstructed leptons: SRs have an OSSF pair with in invariant mass satisfying $|m_{\ell\ell} - m_Z| < 10\,\text{GeV}$, while $t\bar{t}$ CRs have an OSDF pair. Figure taken from [21].

4.4.2 b-jet replacement method

A method called *b-jet replacement method* (BJR) is used to estimate the contribution from non-prompt leptons for $t\bar{t}$ and Z+jets events. The idea behind this method is presented in an abbreviated version here. A detailed description can be found in [56].

For the application of BJR, dilepton samples with 2 reconstructed b-jets are required. In the following, one b-jet is selected randomly and assumed to decay into an electron or a muon. In the rest-frame of the b-hadron, the decay is isotropic. Thus, the azimuthal angle of the resulting non-prompt lepton is chosen randomly in the range $[0, 2\pi)$. A Lorentz transformation from the rest-frame of the b-hadron into the laboratory frame results in information about the kinematics of the non-prompt lepton. This information is used for non-prompt background estimation.

4.4.3 Machine learning, artificial neural networks

Machine learning is a concept useful in many scientific researches which have to deal with many different input variables. In this case, simple selection rules developed by human users may not be able to provide satisfactory results anymore. A typical application in high-energy particle physics is the distinction between signal and many possible background processes.

In most cases machine learning is included as an additional component to an existing analysis framework. It provides the user with a tool to develop further selections rules, deriving mathematical relations between different input variables on its own. Various concepts can allow an evaluation of the performance of the software, allowing for a change of several parameters influencing the performance of the machine learning software that is being used.

The machine learning technique used in this thesis are *artificial neural networks* (NNs), being one of the most commonly used machine learning tools. A general introduction as well as a summary to the NNs used for this thesis are explained in the following.

Neural network principles

Artificial neural networks are designed to imitate biological brains. As such, they consist of processors called *nodes* imitating neurons, which are usually arranged in *layers*, as demonstrated in figure 4.9. Nodes making up the first layer are nodes processing input variables. In the context of particle physics, input variables are most commonly kinematic variables. The nodes in the last layer represent the network response. The network response can be used for classification purposes, i.e. they are used to specify whether an event is more likely a signal or a background process. All layers in between are denoted as *hidden layers*.

Every node in one layer is related to the nodes in the previous and the next layer through a set of weights. Given n nodes in a specific layer which are assigned the values $x_1, \dots, x_n$, the value y_j of a node j in the next layer can be calculated as:

$$y_j = \mathcal{F}\left(w_{0j} + \sum_{i=1}^{n} w_{ij} x_i\right) . \tag{4.4}$$

w_{0j} is called *bias*, $\mathcal{F}$ an *activation function*.

The determination of optimal parameters w_{0j} and w_{ij} is the goal of *neural network training*. For *supervised learning*, the outcome of many given sets of input variables to the NN is known to the user. The performance of the neural network is then evaluated by a *loss function*, which relates the output of the NN to the known outcome, specifying quantitatively the disagreement. The minimization of this loss function corresponds to the case that the NN has good classification power. One way to achieve this is by tweaking the values of the weights and biases, although these are not the only possible options the user might have.

NNs for tZq analysis

For the *tZq* analysis, neural networks using the package `NeuroBayes` have been implemented. [58] The NN structure provided by `NeuroBayes` consists of three layers. The output layer consists of a single node which can take continuous values in the range [-1,1]. In the context of the *tZq* analysis,

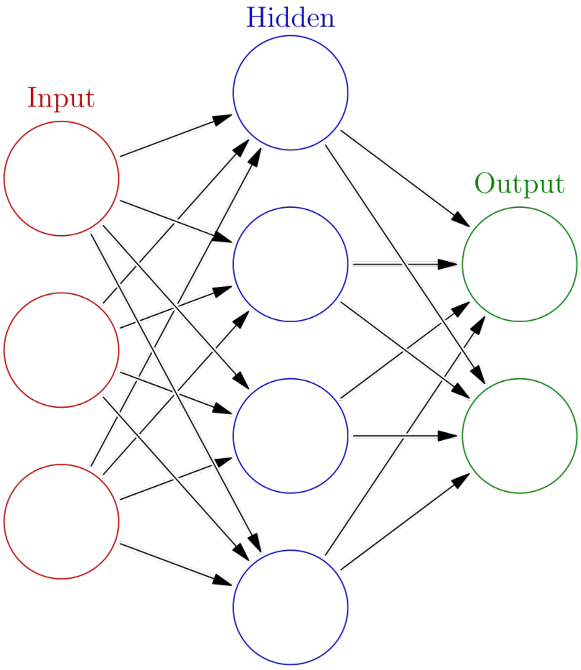


Figure 4.9: Example for a typical structure of a neural network with one hidden layer. Each node is connected to the nodes of the previous and the next layer through a set of weights. Figure taken from [57].

the node in the last layer states how “signal-like” an event is. -1 corresponds to a purely background event, +1 to a signal event.

An important feature provided by `NeuroBayes` is the preprocessing: the normalization of input variables and their ranking in terms of significance to training performance is done in addition to the the training itself. Variables which have different units or different ranges are generally not handled correctly by neural networks. Thus the need for a normalization process, during which every variable is scaled first to the range [0,1], then converted into a Gaussian distribution. This allows the neural network to not prioritize variables based on their values. Furthermore, if not choosing input variables carefully, some variables may contain nearly the same information, just displayed in different way: they are correlated. The contribution of each variable to the output is measured by a covariance matrix. By repeatedly removing one of the variables for the calculation of the NN output and subsequent computation of the correlation of the new NN output, the variable which is the least correlated with the NN output can be determined. This way, variables can be ranked according to their contribution to the NN output, potentially allowing the rejection of some unimportant variables and consequently increasing the speed of the process of NN training.

For each SR of the tZq analysis, a NN has been trained. Both NNs take 15 variables as input, and the hidden layer in each of them consists of 25 nodes. Tests on the reliability and performance of these NNs are shown in [20]. For both SRs, a list of input variables for each NN are shown in tables 4.3 and 4.4. The same NNs are used for signal discrimination on CRs. The NN responses obtained for MC signal and background processes for both SRs are shown in figure 5.1(a) (further details can be found in [20]).

Variable	Definition
$m_{b\mathrm{j_f}}$	Invariant mass of the b-jet and the forward jet
m_t	Reconstructed top-quark mass
$\lvert\eta\left(\mathrm{j_f}\right)\rvert$	Absolute value of the forward jet η
$m_\mathrm{T}\left(\ell, E_\mathrm{T}^\mathrm{miss}\right)$	Transverse mass of the W boson
b-tagging score	b-tagging score of the b-jet
H_T	Scalar sum of the p_T of the selected particles in the event
$q\left(\ell^W\right)$	Electric charge of the lepton from the W boson decay
$\left\lvert\eta\left(\ell^W\right)\right\rvert$	Absolute value of the η of the lepton from the W boson decay.
$p_\mathrm{T}(W)$	p_T of the reconstructed W boson
$p_\mathrm{T}\left(\ell^W\right)$	p_T of the lepton coming from the W boson decay
$m(\ell\ell)$	Mass of the reconstructed Z boson
$\lvert\eta(Z)\rvert$	Absolute value of the η of the reconstructed Z boson
$\Delta R\left(\mathrm{j_f}, Z\right)$	ΔR between the forward jet and the reconstructed Z boson
$E_\mathrm{T}^\mathrm{miss}$	Missing transverse momentum
$p_\mathrm{T}\left(\mathrm{j_f}\right)$	Forward jet p_T

Table 4.3: List of input variables to the NN for SR-$2j1b$, ranked according to their contribution to the NN output, with explanations for what these variables represent. Information taken from [21].

Variable	Definition
$m_{b\mathrm{j_f}}$	Invariant mass of the b-jet and the forward jet
m_t	Reconstructed top-quark mass
$\lvert\eta\left(\mathrm{j_f}\right)\rvert$	Absolute value of the forward jet η
$m_\mathrm{T}\left(\ell, E_\mathrm{T}^\mathrm{miss}\right)$	Transverse mass of the W boson
$\lvert\eta\left(\mathrm{j_r}\right)\rvert$	Absolute value of the radiation jet η
$p_\mathrm{T}(Z)$	p_T of the reconstructed Z boson.
$\Delta R\left(\mathrm{j_f}, Z\right)$	ΔR between the forward jet and the reconstructed Z boson
$q\left(\ell^W\right)$	Electric charge of the lepton from the W boson decay
$p_\mathrm{T}\left(\mathrm{j_r}\right)$	p_T of the radiation jet
$p_\mathrm{T}\left(\mathrm{j_f}\right)$	p_T of the forward jet
b-tagging score	b-tagging score of the b-jet
$\left\lvert\eta\left(\ell^W\right)\right\rvert$	Absolute value of the η of the lepton coming from the W boson decay.
$\lvert\eta(Z)\rvert$	Absolute value of the η of the reconstructed Z boson
$p_\mathrm{T}\left(\ell^W\right)$	p_T of the lepton coming from the W boson decay
$p_\mathrm{T}(W)$	p_T of the reconstructed W boson

Table 4.4: List of input variables to the NN for SR-$3j1b$, ranked according to their contribution to the NN output, with explanations for what these variables represent. For SR-$3j1b$, the radiation jet is defined as the jet which is neither b-tagged, nor the forward jet. Information taken from [21].

4.4.4 Signal extraction, fitting procedure

The NN output is used to determine the number of *tZq* events recorded in data, from which, by use of equation (2.10), the inclusive cross-section can be determined. The procedure used for the *tZq* analysis allowing this calculation is called *binned profile likelihood fit*. The theoretical introduction follows the structure given in [59]. Additional information can be found in [60].

Likelihood fit: theoretical considerations

A likelihood function is a mathematical tool used to find or verify that the model used describes the data in the best possible way. According to the concept of *maximum likelihood*, the likelihood function should be maximized by varying a set of parameters. The maximum value of the likelihood function defines the set of parameters for which the the parametrized model describes data in the probably best possible way.

The number of total events n in a considered region can be parametrized as:

$$n(\mu) = \mathcal{L}\left(\epsilon_0\sigma_0\mu + \sum_j^{\text{bkg}} \epsilon_j\sigma_j\right) . \tag{4.5}$$

$\mathcal{L}$ denotes the integrated luminosity, σ_0 $\left(\sigma_j\right)$ the theoretical cross-section for the *tZq* process (the cross-section of a particular background process j) and ϵ_0 $\left(\epsilon_j\right)$ the reconstruction and selection efficiency of the *tZq* process (of the background process j). The parameter μ is called signal strength. Determining this parameter is one of the goals of the fitting procedure.

If systematic uncertainties are neglected, the underlying distribution for the number of observed events N given a mean number of events ν follows a Poisson distribution:

$$P_\nu(N) = \frac{\nu^N}{N!}\mathrm{e}^{-\nu} . \tag{4.6}$$

Using equation (4.5), the parameter ν used in equation (4.6) can be replaced by $n(\mu)$. Furthermore, since the measurement is performed on several bins that are statistically independent from each other, equation (4.6) holds for each bin separately. This allows one to define the likelihood function as:

$$\begin{aligned} L\left(\vec{N}, \vec{n}(\mu)\right) &= \prod_i^{\text{bins}} P_{n_i(\mu)}(N_i) \\ &= \prod_i^{\text{bins}} \frac{n_i(\mu)_i^N}{N_i!}\mathrm{e}^{-n_i(\mu)} . \end{aligned} \tag{4.7}$$

$\vec{N}$ and $\vec{n}(\mu)$ are vectors of length i containing the corresponding numbers for each bin i. Instead of using this likelihood function, it is common to define the log-likelihood function as:

$$\Lambda\left(\vec{N}, \vec{n}(\mu)\right) = -2\ln\left(L\left(\vec{N}, \vec{n}(\mu)\right)\right) , \tag{4.8}$$

which allows for summation instead of multiplications. This is especially beneficial in the case of a large number of bins. The maximum value of the likelihood function is obtained for a value $\hat{\mu}$ which is used as estimator for the signal strength. $\hat{\mu}$ approximates the actual value of μ best. The uncertainty of $\hat{\mu}$ is estimated by varying $\hat{\mu}$ until the value of the log-likelihood function shifts by $+1$.

Systematic uncertainties can be included via nuisance parameters θ_k, assumed to follow a normal Gaussian distribution, deviating the total number of events by factors δ_k:

$$n_i\left(\mu, \vec{\theta}\right) = n_i(\mu)\left(1 + \sum_{k}^{\text{syst. unc.}} \delta_k \theta_k\right) . \tag{4.9}$$

This redefinition of $n(\mu)$ leads to the following expression of the log-likelihood function:

$$\Lambda\left(\vec{N}, \vec{n}(\mu)\right) = \Lambda_{\text{nom.}}\left(\vec{N}, \vec{n}(\mu)\right) + \sum_{k}^{\text{syst. unc.}} \ln\left(f_{\mathcal{N}}\left(\theta_k\right)\right) , \tag{4.10}$$

where $f_{\mathcal{N}}\left(\theta_k\right)$ denotes the aforementioned normal Gaussian distribution that the nuisance parameter θ_k follows.

Implementation for tZq analysis

The likelihood fit for the tZq analysis is performed using the software `TRExFitter`, [61] which uses histograms from all SRs and CRs defined in section 4.4.1 as input. The parameter of interest is for this particular case defined as the signal strength $\mu_{\text{SIG}} = \sigma_{tZq}/\sigma_{tZq}^{\text{theo.}}$.

The result for the inclusive cross-section obtained from the fit is: [21]

$$\sigma_{tZq} = 97 \pm 13\,(\text{stat.}) \ \pm 7\,(\text{syst.})\ \text{fb} . \tag{4.11}$$

Although systematic uncertainties are included in the inclusive cross-section measurement, time constraints limited the inclusion of experimental and modeling uncertainties in this work. Instead, only statistical uncertainties are taken into account. However, additional input to configuration files for `TRExFitter` together with new code was used to extract the estimated number of background events from MC for individual bins instead of a global number of events. This is needed for the differential cross-section measurement (see chapter 5).

4.5 The “forward jet $|\eta|$ problem”

One of the problems faced after including 2018 data in the analysis framework is described. Due to the plot where the feature described in this chapter first became apparent, it was named “forward jet $|\eta|$ problem”.

4.5.1 Description

After publication of results on the tZq process from the CMS collaboration, [1] the ATLAS analysis was extended to also include 2018 data, i.e. running over the full Run 2 dataset. However, cross-checks already at pre-fit level indicated a mismatch between MC predictions and data in the forward region, i.e. for $|\eta| > 2.5$. The discrepancy was most prominent in the control region CR-2j0b for the forward jet $|\eta|$-distribution. In that specific CR, the forward jet is defined as the jet having the highest value for

$|\eta|$. Data indicated a “bump” for $|\eta| \approx 2.5$) to $|\eta| \approx 3.2$) compared to MC predictions, as can be seen in figure 4.10.

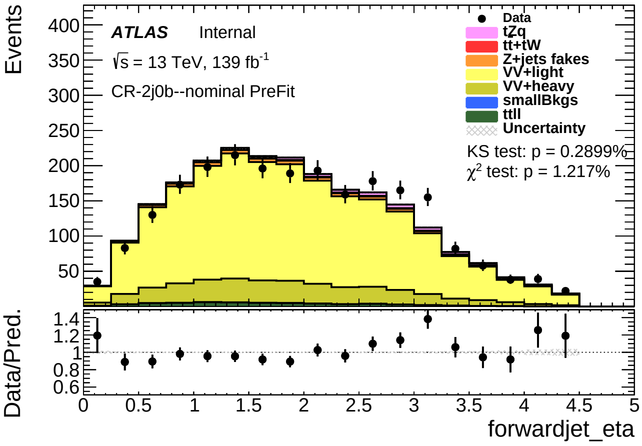


Figure 4.10: Forward jet $|\eta|$ distribution, including the data for all the years, compared to MC predictions.

4.5.2 Search for possible explanations

In order to solve this problem, some features were investigated. First, plots split according to different campaigns were generated, to see if this feature was isolated to the inclusion of 2018 data. Furthermore, the impact of the forward Jet Vertex Tagger (fJVT) was investigated. Distributions were also split according to pileup conditions. Lastly, suggestions during a Jet Calibration and Resolution meeting revealed dependences according to a specific run number marking the transition period from low- to high-pileup conditions.

2018 distribution only

The first step taken was to split the distributions according to year and see if this feature was visible if analyzing only 2018 data. Such a feature can be seen in figure 4.11.

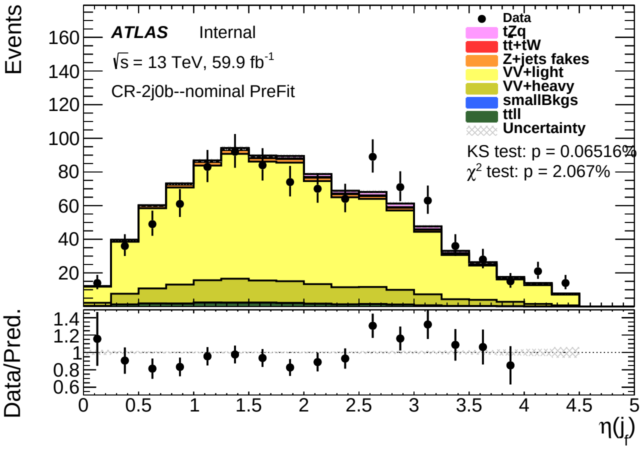
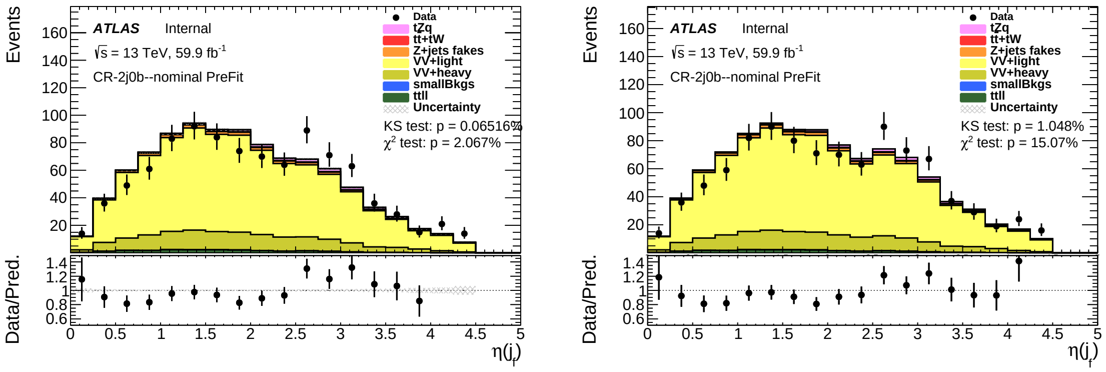


Figure 4.11: Forward jet $|\eta|$ distribution, comparing data to MC predictions for 2018 only.

Influence of the fJVT

As this “bump” is observed in the forward region, the influence of the fJVT [41] was an important element that required investigation. The results of this procedure for 2018 data can be seen in figure 4.12, applying the fJVT for figure 4.12(a), while neglecting it on for figure 4.12(b).

On the basis of the discrepancies still seen in figure 4.12, it is possible to conclude that the fJVT seems to show some impact (i.e. turning the fJVT off may reduce the discrepancy), but it is not able to explain the difference between data and MC. Therefore, the fJVT was not discarded from the analysis framework.



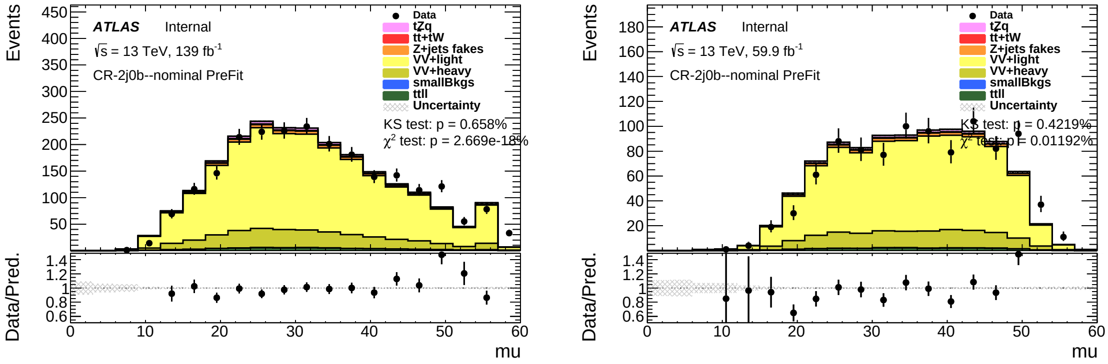
(a)

(b)

Figure 4.12: Distributions for forward jet $|\eta|$. Figure 4.12(a) shows the case of applying the fJVT, figure 4.12(b) without fJVT.

Influence of pileup conditions

Pileup conditions, described by the variable μ, were plotted and the distributions splitted into a low- respectively high-pileup regime. Low-pileup conditions in this specific case, after having a look into



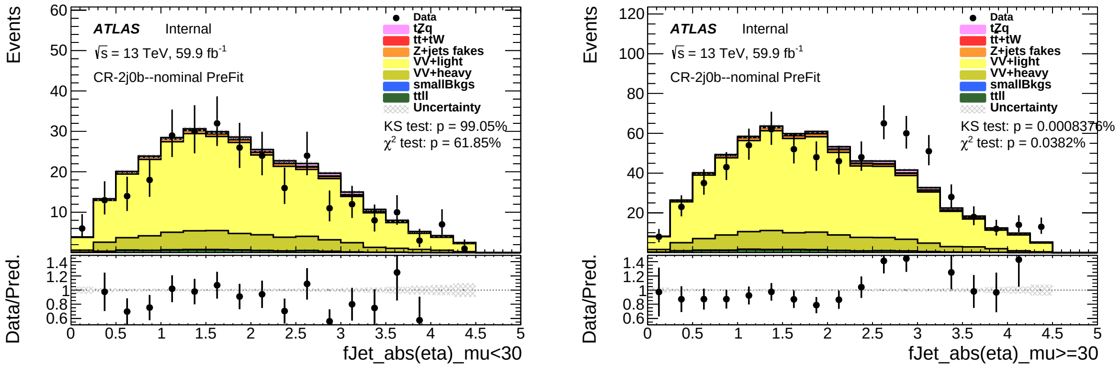
(a)

(b)

Figure 4.13: Pileup μ distributions. Figure 4.13(a) for all the years of Run 2, figure 4.13(b) for 2018 data.

the distributions shown in figure 4.13, was defined as $\mu < 30$. Anything higher than a value of 30 for μ is considered high-pileup. The value of $\mu = 30$ to split the distributions corresponds approximately to a value for which the distribution of μ shows a maximum over the years 2015-2018. This splits the events approximately equally into low- respectively high-pileup conditions.

After having defined low- and high-pileup regions, distributions for forward jet $|\eta|$ split according to those definitions were plotted (figure 4.14). While the "bump" is not visible for low-pileup conditions, it is prominent in the same $|\eta|$-region as before for high-pileup conditions, which indicates that this feature arises in high-pileup conditions only.



(a)

(b)

Figure 4.14: Forward jet $|\eta|$-distribution for 2018 data, split according to $\mu < 30$ (figure 4.14(a)) and $\mu \geq 30$ (figure 4.14(b)).

Run Number 355258

As the plots from figure 4.14 revealed a dependence on the pileup conditions, data was split according to whether it was taken before or after run number 355258. This particular run got singled out for more investigations because:

- it marked the end of low-μ data-taking in 2018,
- gains in the Liquid Argon Electromagnetic Endcap were reconfigured before and after run 355258, [62] while at the same time the LAr online data base showed problems. [63]

The plot generated after splitting 2018 data according to this recommendation is shown in figure 4.15. At $|\eta| \approx 2.5$, a prominent peak can be seen for the distribution for data after run number 355258, while it is not visible for events before that particular run number.

4.5.3 Consequences

In the meantime, other analysis teams (partially due to the request of the ATLAS *tZq* analysis team, [64]) had seen similar features. Such an effect had already been observed previously in 2015+2016 data. In that case, the problem was solved by introducing a new flag called "DFCommonJets_isBadBatman" in the `AnalysisTop` framework. [65] This was used to reject extra pileup jets in a regime of $\eta \approx 2.9$

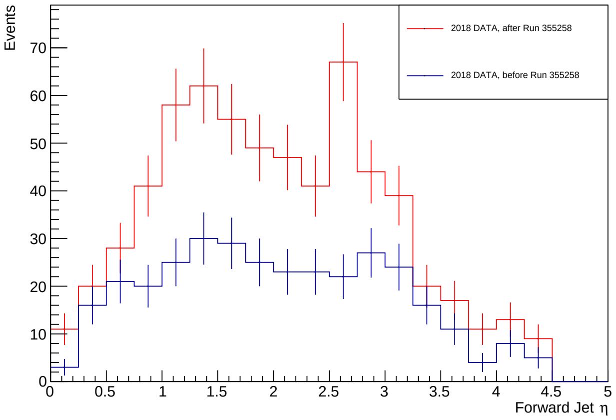


Figure 4.15: Forward jet $|\eta|$ distribution, comparing data before (blue) and after (red) run number 355258. Uncertainties shown are statistical only.

which have been modeled poorly by MC. As an immediate consequence of the studies performed during this thesis, the same flag was included as an option to `AnalysisTop` to be run for 2018 data.

Further studies revealed that the discrepancies do not have the same origin as the “isBadBatman” effect. [66] However, simulation of pileup reveals discrepancies for both 2017 and 2018 data if simulated with same train configurations (a train is a collection of several bunches, temporally spaced by 25 ns). The discrepancy is not seen in 2017 data, as simulation and running conditions were not identical for that year, which happened to move MC closer to data for 2017. The inclusion of weights to counter this effect may be added to the `AnalysisTop-22` framework. This addition is still under debate and further studies ongoing.

CHAPTER 5

Preparation for differential cross-section measurements

In contrast to inclusive cross-sections, differential cross-sections contain more information about a process in the phase space investigated. Therefore, they are useful to test the SM with higher precision. However, enough data needs to be available to achieve an accurate differential cross-section measurement. This work is the first attempt to measure the differential cross-section of the *tZq* production process.

An overview of prerequisites needed to perform a differential cross-section measurement is given in this chapter. First, regions are defined which, by cutting on the NN output (O_{NN}), allow for a higher signal purity. The concept of unfolding is introduced and specific unfolding mechanisms are presented in the second part of this chapter. Their implementation and the procedure used to determine statistical uncertainties are described. Lastly, some explanations concerning normalized differential cross-sections are outlined.

5.1 Signal purity

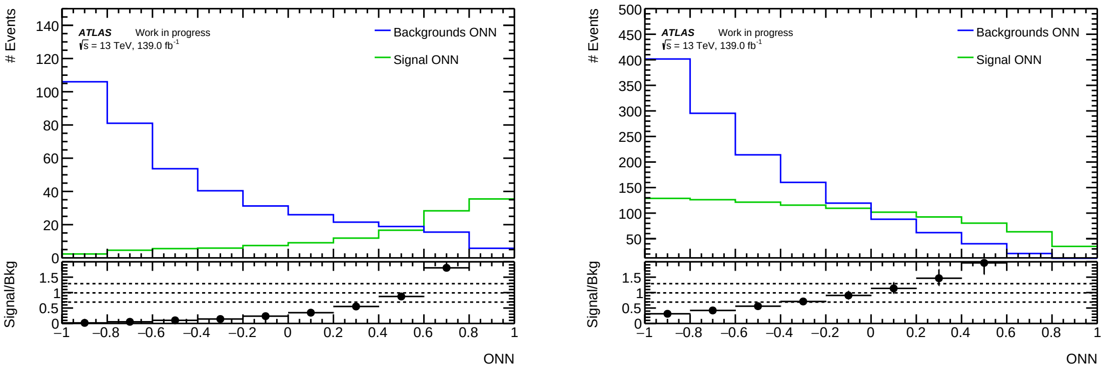
To ensure a high signal purity for the differential cross-section measurement, a cut on the output of a NN is imposed; allowing the rejection of background events. The NNs in question are the same used for the inclusive cross-section analysis, described in section 4.4.3. The NNs were not re-trained, nor were any variables added or removed. For both SRs, the NNs do not require any input variables for which differential cross-sections are measured (lists of input variables are shown in tables 4.3 and 4.4, results in chapter 6). This aspect is important to consider as to avoid obvious correlations between a cut on the NN output and the variable that is being unfolded.

Two approaches were followed to define cuts on the NN output:

1. In the first approach, event yields from both SRs are added. A cut on O_{NN} is applied such that the remaining bins each separately have a signal-to-background ratio larger than 1. This resulted in a cut of $O_{\text{NN}} > 0.6$ (see figure 5.1(a)).

2. For an alternative approach, integral event yields instead of individual bins are considered. A cut is applied as soon as the integrated number of signal events drops below the integrated

background event yield. The resulting value for a cut on O_{NN}, as depicted in figure 5.1(b), is $O_{\mathrm{NN}} > 0$.



(a) (b)

Figure 5.1: Distributions of O_{NN}. In figure 5.1(a), the event yields from both SRs are combined for each bin separately. For figure 5.1(b), event yields from both SRs have been combined as integrals: the content of each bin is the event yield from events with $O_{\mathrm{NN}} \in [\mathrm{LBE}, 1]$, where LBE is the lower bin edge of the respective bin.

5.2 Introduction to unfolding

5.2.1 Concept

When measuring differential cross-sections, reverting the impact of detector effects and selection cuts is one of the key aspects. Using technical correct terms, the result should not return quantities at *reconstruction level* (which includes detector smearing and deficiencies, selection cuts, etc.), but at *truth level*. Once the measurement is independent of these detector effects, it can be compared to theory predictions. Truth level quantities are kinematic variables of the particles involved in a reaction before the actual measurement in the detector (which, in principle, are unknown). As a consequence, problems like finite resolution and limited acceptances by the detector are not included at truth level.

Unfolding is the technique which, starting from a measurement, allows one to recover a distribution which is independent of such detector effects. As a consequence, unfolding allows one to recover a truth level distribution, as summarized in figure 5.2.

Only binned distributions have been investigated for this thesis. Therefore, the math related to unfolding is described here only for discrete distributions. Assume that a kinematic variable of interest $\vec{x}$ has been measured. The components x_i are therefore at reconstruction level. Each component of this vector contains the number of events at reconstruction level for the kinematic distribution in question for a specific bin i. The distribution at reconstruction level $\vec{x}$ can be related to its truth level counterpart $\vec{y}$ through a matrix R, generally called *response matrix*:

$$\vec{x} = R\vec{y} + \vec{b}\,, \qquad \text{component-wise:} \quad x_i = \sum_{j=1}^{\text{bins}} R_{ij} y_j + b_i\ . \tag{5.1}$$

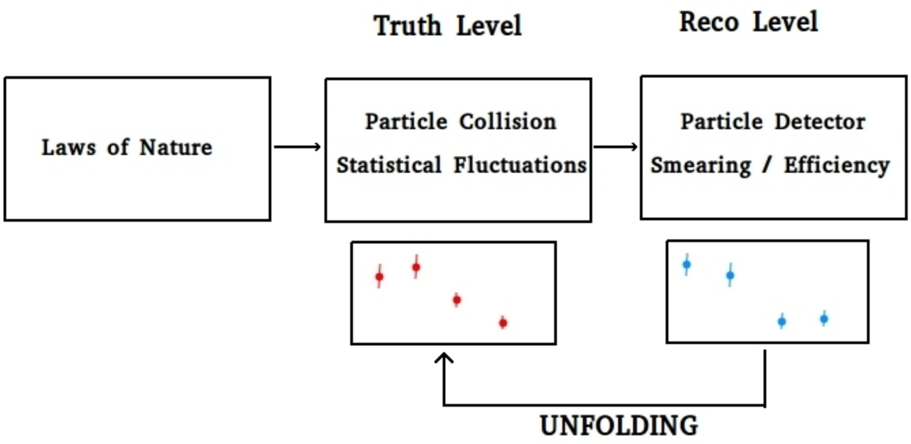


Figure 5.2: Illustration of the approach of unfolding. Laws of physics define which interactions are possible when particles collide. Due to the quantum nature of these interactions, statistical fluctuations at truth level distributions are possible. Smearing and deficiencies, e.g. due to a limited resolution of the detector, lead to reconstruction level distributions. The goal of unfolding is to recover a truth level distribution, by reverting the detector effects mentioned previously.

b_i denotes the estimated number of background events for bin i. The matrix R can be estimated from MC simulations, given that MC represents a decent description of nature.

The goal of unfolding is to invert equation (5.1). For a measured distribution $\vec{x}$ and known backgrounds $\vec{b}$, the truth level distribution $\vec{y}$ should be recovered. However, even if the description seems mathematically simple, this is a challenging task for most real-life experiments, as described in the following section.

5.2.2 Unfolding methods

Matrix inversion and its problems

A naïve guess on how to solve equation (5.1) relies on matrix inversion:

$$\vec{y} = R^{-1}\left(\vec{x} - \vec{b}\right) , \quad \text{component-wise:} \quad y_i = \sum_{j=1}^{\text{bins}} R_{ij}^{-1}\left(x_j - b_j\right) . \tag{5.2}$$

Often, equation (5.2) is split into a few components, [67] such that it can be rewritten as:

$$y_i = \frac{1}{\epsilon_i} \sum_{j=1}^{\text{bins}} M_{ij}^{-1} a_j \left(x_j - b_j\right) . \tag{5.3}$$

This way, *efficiencies* ϵ, *acceptances* a and the *migration matrix* M can be introduced. The efficiency ϵ_i describes how many of the events at truth level in bin i actually have been measured and passed all selection cuts, such that they have a counterpart at reconstruction level, and relates it to the total number of events at truth level in bin i. Similarly, the acceptances a_j account for events which have been detected, but could not be assigned to any truth level event. The migration matrix M finally describes bin-to-bin migrations for events existing both at truth and at reconstruction level.

However, equation (5.2) is ill-posed in the case of a non-invertible matrix R. Even if this is not the case, statistical fluctuations can potentially induce negative entries in the inverse matrix. In extreme cases, this can even lead to a negative number of events in some bins of the unfolded distribution. Thus, although mathematically correct, this is an unwanted feature when performing a differential cross-section measurement. An example for such an extreme case is shown in figure 5.3.

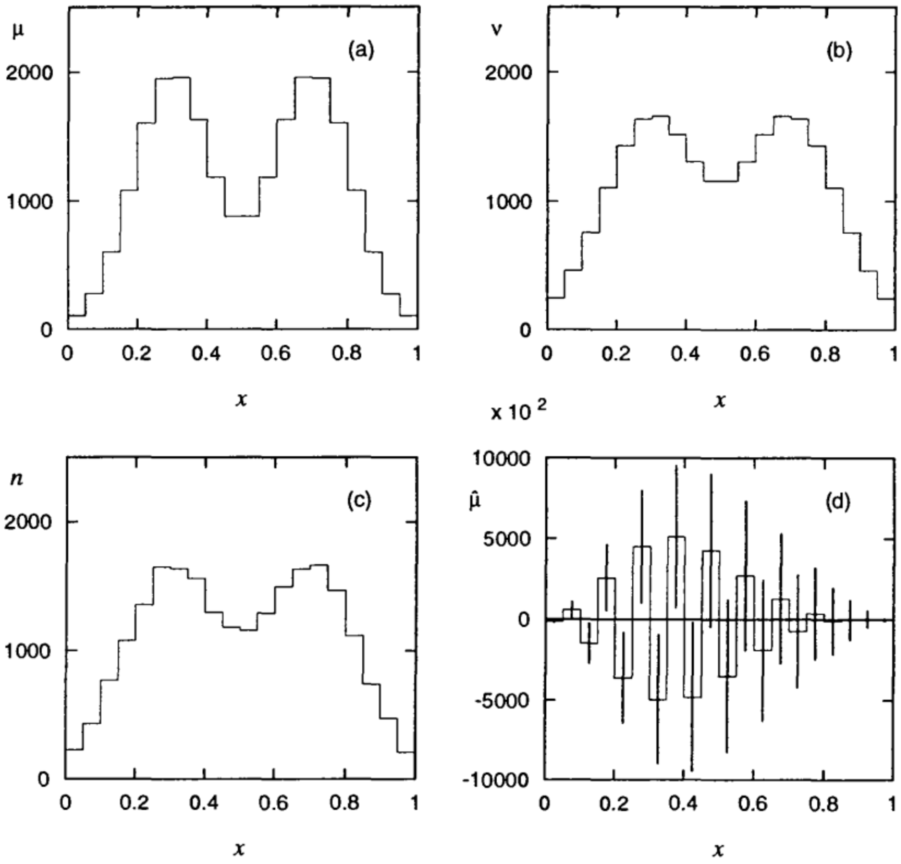


Figure 5.3: Example showing why matrix inversion should not be used when performing a differential cross-section measurement. An example for a MC truth level distribution is shown in the figure on the top left, and the corresponding MC reconstruction level distribution in the figure on the top right. The result of unfolding the measured distribution (bottom left) can be seen in the figure on the bottom right, which does deviate significantly from the MC truth level distribution. Figure taken from [60].

Several concepts have been developed to counter this problem. Two of them, implemented during this thesis, are presented in more detail.

Bin-by-bin unfolding

The simplest but also most limited method is *bin-by-bin unfolding*. For this method, a correction factor K_i for each bin i is determined: [68]

$$K_i = \frac{N^i_{\text{truth}}}{N^i_{\text{reco}}} \,, \tag{5.4}$$

where N^i_{truth} (N^i_{reco}) denotes the number of events for the truth (reconstruction) level distribution in bin i as obtained from MC. With these correction factors determined, the bin-by-bin unfolded distribution takes the form: [68]

$$N^i_{\text{unfold}} = K_i \cdot \left(N^i_{\text{obs}} - N^i_{\text{bkg}}\right) \,. \tag{5.5}$$

Here N^i_{obs} is the number of observed events in bin i, while N^i_{bkg} is the number of estimated background events.

In the most ideal case, N^i_{unfold}, the number of events obtained applying bin-by-bin unfolding, would be equal to N^i_{truth}. Indeed, when replacing $\left(N^i_{\text{obs}} - N^i_{\text{bkg}}\right)$ by N^i_{reco} in equation (5.5), the truth level distribution can be recovered. However, two important restrictions need to be taken into account:

- Like every unfolding method, bin-by-bin unfolding relies on good agreement between MC and data. If MC cannot model nature accurately, the correction factors K_i in equation (5.4) are not estimated correctly.

- This method is applicable only if all the distributions entering equations (5.4) and (5.5) have identical binning. If not, either a re-binning is required, or another unfolding method should be used.

- Correction factors are provided only for each bin separately. Events that migrate between different bins from truth compared to reconstruction level are not covered. Thus, bin-by-bin unfolding is expected to give good results only if bin-to-bin migration can be neglected. The response matrix R has approximately a diagonal form in that specific case, i.e. off-diagonal elements can be neglected.

Iterative Bayesian unfolding

Another method proposed by D'Agostini [69] is called *iterative Bayesian unfolding* (IBU), which makes use of Bayes' theorem. The argument given follows the introduction given in [70].

To apply Bayes' theorem, distinctions need to be made between *causes* C and *effects* E. Several independent causes may create the same effect. Consider truth level particles as the cause and the reconstructed particle as the effect. It is possible for many potential truth level initial states to be reconstructed into the same reconstruction level final sate, as detector effects may smear the measurement. As an example, reconstructed top-quarks with $p_{\text{T}}^{\text{reco}} = 95\,\text{GeV}$ may come from several truth level top-quarks which might have had $p_{\text{T}}^{\text{truth}}$ values in the ranges of $90 - 100\,\text{GeV}$. In general, several causes C_i ($i = 1, 2, \ldots, n_C$) can exist, which all can lead to several effects E_j ($j = 1, 2, \ldots, n_E$). Bayes' theorem provides an expression for $\mathcal{P}(C_i|E_j)$, the conditional probability that an observed quantity E_j has been provoked by a cause C_i:

$$\mathcal{P}(C_i|E_j) = \frac{\mathcal{P}(E_j|C_i)\mathcal{P}(C_i)}{\mathcal{P}(E_j)} = \frac{\mathcal{P}(E_j|C_i)\mathcal{P}(C_i)}{\sum_{i=1}^{n_c}\mathcal{P}(E_j|C_i)\mathcal{P}(C_i)} \tag{5.6}$$
$$\Rightarrow \mathcal{P}(C_i|E_j) \propto \mathcal{P}(E_j|C_i)\mathcal{P}(C_i)\ .$$

The expression $\mathcal{P}(E_j|C_i)$, the conditional probability that a cause C_i produces an effect E_j, can be estimated via MC simulations and can be identified with the response matrix. $\mathcal{P}(C_i|E_j)$ conversely represents the "inverse" matrix[1] that iterative Bayesian unfolding tries to determine. If detector efficiencies and acceptance terms are calculated separately such that only events with counterparts both at truth and at reconstruction level are considered, only the migration matrix $M_{ji} := \mathcal{P}(E_j|C_i)$ enters equation (5.6). The "inverse" matrix $M_{ij}^{-1} := \mathcal{P}(C_i|E_j)$ is called *unfolding matrix* hereafter. $\mathcal{P}(C_i)$ entering equation (5.6) is an estimate for the truth level distribution.

On the other hand, knowledge about $\mathcal{P}(C_i|E_j)$ can be used to get an estimate for the number of events due to cause C_i:

$$\hat{n}(C_i) = \frac{1}{\hat{\epsilon}_i}\sum_{j=1}^{N_E} n(E_j)\mathcal{P}(C_i|E_j)\ , \qquad \hat{\epsilon}_i \neq 0\,, \tag{5.7}$$

with $n(E_j)$ denoting the observed number of events of effect j. $\hat{\epsilon}_i$ is defined as the ratio $\hat{n}(C_i)/n(E_i)$. Dividing both sides of equation (5.7) by $\hat{N}(C) = \sum_i \hat{n}(C_i)$ and using a probabilistic interpretation:

$$\frac{\hat{n}(C_i)}{\sum_i \hat{n}(C_i)} = \frac{\hat{n}(C_i)}{\hat{N}(C)} = \frac{1}{\hat{\epsilon}_i \hat{N}(C)}\sum_{j=1}^{N_E} n(E_j)\mathcal{P}(C_i|E_j) \tag{5.8}$$
$$\Rightarrow \mathcal{P}(C_i) = \frac{1}{\hat{\epsilon}_i \hat{N}(C)}\sum_{j=1}^{N_E} n(E_j)\mathcal{P}(C_i|E_j)\ .$$

If iterative Bayesian unfolding should be applied to the migration matrix only (i.e. no efficiency and acceptance corrections are considered at this stage), the total number of events due to all effects E is equal to the number of events due to all causes C: $\sum_i \hat{n}(E_i) = \sum_i \hat{n}(C_i) = \hat{N}(C)$. Therefore, the probabilistic approach can also be applied to $n(E_j)$ when dividing by $\hat{N}(C)$, changing equation (5.8) to:

$$\mathcal{P}(C_i) = \frac{1}{\hat{\epsilon}_i}\sum_{j=1}^{N_E} \mathcal{P}(E_j)\mathcal{P}(C_i|E_j)\ . \tag{5.9}$$

Comparisons between equations (5.6) and (5.9) show that both equations link the probabilities $\mathcal{P}(C_i)$ and $\mathcal{P}(C_i|E_j)$. $\mathcal{P}(C_i)$ is related to the truth level distribution, whereas $\mathcal{P}(C_i|E_j)$ represents the unfolding matrix. Thus an iterative procedure can be adapted to solve for $\mathcal{P}(C_i|E_j)$:

---

[1] It is not the true inverse matrix, but it is used to fulfill the same purpose as the inverse matrix from equation (5.2). See [69] for a proof.

1. An initial guess $\mathcal{P}_0(C_i)$ for the probability density function is used as starting point. $\mathcal{P}_0(C_i)$ is inserted into equation (5.9) to get a first estimate of the unfolding matrix elements $\mathcal{P}(C_i|E_j)$.

2. This first estimate of $\mathcal{P}(C_i|E_j)$ is inserted into equation (5.6) to get an updated $\mathcal{P}_0(C_i)$, now denoted by $\mathcal{P}_1(C_i)$.

3. The updated $\mathcal{P}_1(C_i)$ is used again in equation (5.9). This process is repeated until some stopping criteria are met. Stopping criteria used are discussed in section 5.3.3.

Iterative Bayesian unfolding has many advantages. First, it does not require a direct matrix inversion. Because of the iterative approach, this feature makes the procedure less sensitive to statistical fluctuations. Furthermore, it can address bin-to-bin migrations, a feature that bin-by-bin unfolding is not able to handle. Lastly, although not used during this work, different binnings chosen for truth and reconstruction level distributions would still be possible to unfold using iterative Bayesian unfolding.

5.2.3 Parton and particle level unfolding

Distinctions are made between parton level and particle level unfolding:

- Parton level unfolding: the MC truth level distribution is directly available for the particle that should be investigated, even if it is not a measured final-state particle. As an example, the full kinematic information about the top-quark in the tZq process is available at parton level. Usually, the full phase-space is considered for truth level distributions, i.e. no cuts are applied. For this analysis, particles are considered at parton level after final-state radiation (e.g. emission of gluons, but before the decay of the particle) has occurred.

- Particle level unfolding: the distribution that should be unfolded needs to be determined from kinematic variables available for stable final-state particles, even when considering MC samples. As an example, the kinematic distributions of the top-quark in the tZq process need to be reconstructed first from a b-quark and a W boson, which also has to be reconstructed from final, stable particles. In addition, at particle level, event selection cuts to truth level may be applied.

5.3 Implementation of unfolding procedure

A general overview of how the unfolding procedure has been implemented is given in this section. The estimation of statistical uncertainties is highlighted. Specific problems worthwhile noting related to statistical uncertainties are discussed in appendices C and D.

5.3.1 Implementation and conversion to cross-sections

Both bin-by-bin and iterative Bayesian unfolding have been implemented as described in section 5.2.2, whereas the main focus was on iterative Bayesian unfolding. The iterative Bayesian unfolding was partially used as implemented in the software package `RooUnfold` [71, 72].

Bin-by-bin unfolding

Bin-by-bin unfolding was implemented by extending equation (5.5) to enable conversions to cross-sections:

$$\frac{\mathrm{d}\sigma_{\mathrm{BBB}}}{\mathrm{d}X^i} = \frac{K_i}{\mathcal{L} \cdot \Delta X^i} \cdot \left(N^i_{\mathrm{obs}} - N^i_{\mathrm{bkg}}\right) \;, \tag{5.10}$$

where $\mathcal{L}$ is the integrated luminosity and ΔX^i denotes the bin width of bin i for the distribution of variable X. The number of background events N^i_{bkg} for each bin i is estimated during the fitting procedure (see section 4.4.4).

Iterative Bayesian unfolding

For iterative Bayesian unfolding, unfolded event yields or differential cross-sections can be determined by extending equation (5.3): [67]

$$\begin{aligned} N^i_{\mathrm{unf}} &= \frac{1}{\epsilon_i} \cdot \sum_{j=1}^{N_{\mathrm{bins}}} M^{-1}_{ij} a_j \left(N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}}\right) \\ \Leftrightarrow \frac{\mathrm{d}\sigma_{\mathrm{IBU}}}{\mathrm{d}X^i} &= \frac{1}{\mathcal{L} \cdot \Delta X^i \cdot \epsilon_i} \cdot \sum_{j=1}^{N_{\mathrm{bins}}} M^{-1}_{ij} a_j \left(N^j_{\mathrm{obs}} - N^j_{\mathrm{bkg}}\right) \;. \end{aligned} \tag{5.11}$$

The quantities $\mathcal{L}$ and ΔX^i are defined the same way as in equation (5.10): $\mathcal{L}$ is the integrated luminosity, ΔX^i denotes the bin width of bin i for the distribution of variable X. The number of background events for each bin is again estimated from the fit.

In contrast to equation (5.3), M^{-1}_{ij} in equation (5.11) does not denote the exact inverse matrix of the migration matrix, but the matrix that can be obtained by applying iterative Bayesian unfolding. Code-wise, the migration matrix is being determined from events which have entries at truth and reconstruction level. This matrix is then given as an input to the `RooUnfold` framework.

The acceptances a_j are determined by comparing the number of events in bin j with counterparts at truth as well as at reconstruction level to the number of events at reconstruction level only: [67]

$$a_j = \frac{N^j_{\mathrm{truth} \wedge \mathrm{reco}}}{N^j_{\mathrm{reco}}} \;. \tag{5.12}$$

For the calculation of a_j, both distributions $N^j_{\mathrm{truth} \wedge \mathrm{reco}}$ and N^j_{reco} are considered at reconstruction level. In cases where $N^j_{\mathrm{truth} \wedge \mathrm{reco}} > N^j_{\mathrm{reco}}$ (which is possible due to negative total event weights), the acceptance was set to 1 by default.

The efficiencies ϵ_i are determined by comparing the number of events with counterparts at truth as well as at reconstruction level to the number of events at truth level only: [67]

$$\epsilon_i = \frac{N^i_{\mathrm{truth} \wedge \mathrm{reco}}}{N^i_{\mathrm{truth}}} \;. \tag{5.13}$$

For the calculation of ϵ_i, both distributions $N^j_{\text{truth}\wedge\text{reco}}$ and N^j_{truth} are considered at truth level. However, to yield correct unfolded event numbers (as tested on MC), the distribution of $N^i_{\text{truth}\wedge\text{reco}}$ has to be rescaled to match the total event yield of the reconstruction level distribution. The rescaling is necessary as although the same events are considered, their total weight as determined from equation (4.2) might be different. For example, there exist no weights related to b-tagging or triggering at truth level. Furthermore, the luminosity-weight w_{lumi} can be different for reconstruction and truth level events.

5.3.2 Estimation of uncertainties

An overview of steps taken to estimate uncertainties of the unfolded distributions is given. Due to time constraints, only statistical uncertainties are covered in this thesis. Systematic uncertainties need to be covered at a later stage.
Disclaimer: To avoid confusions between uncertainties and cross-sections (as both usually are abbreviated by the symbol σ), the symbol Δ is used in this chapter when referring to uncertainties.

Statistical uncertainty estimation for bin-by-bin unfolding

By construction, bin-by-bin unfolding does not yield any correlations between different bin contents. Thus, Gaussian uncertainty propagation is a viable tool without the necessity of a full covariance matrix. Under the assumption that for each bin i the number of measured events N^i_{obs} is Poisson distributed, the uncertainty associated to the unfolded distribution is:[2]

$$\Delta\left(\frac{\mathrm{d}\sigma_{\text{BBB}}}{\mathrm{d}X^i}\right) = \frac{K_i}{\mathcal{L}\cdot\Delta X^i}\cdot\sqrt{N^i_{\text{obs}}}\,. \tag{5.14}$$

An alternative procedure relying on a set of 10^5 toy experiments was also tested. Off-diagonal elements in the resulting correlation matrix were smaller than 1%, validating that different bins are not correlated. See appendix C.1 for further details.

Statistical uncertainty estimation for iterative Bayesian unfolding

In contrast to bin-by-bin unfolding, iterative Bayesian unfolding by default is constructed to allow for correlations between different bin contents. This makes the estimation of uncertainties non-trivial, as it should no longer be possible to estimate the corresponding uncertainties for each bin via Gaussian uncertainty propagation. As alternative, a set of 10^5 toy experiments is used to estimate the statistical uncertainties of the unfolded distribution.

The procedure adapted is as follows: from MC reconstruction level distributions, two random numbers m_j and b_j are generated for each bin j. b_j is generated from a Poisson distribution with the mean value being the number of background events in bin j as estimated by MC. m_j is generated from a Poisson distribution with a mean value of signal+background events in bin j. These two random numbers are then used in equation (5.11): m_j replaces N^j_{obs}, while b_j is the new input for the number

---

[2] The uncertainty was assigned $\propto\sqrt{N^i_{\text{obs}}}$ after considering extreme cases: if $N^i_{\text{bkg}} = 0$, then $N^i_{\text{sig}} = N^i_{\text{obs}}$ and the relation immediately follows. In case of $N^i_{\text{sig}} = 0$, the observed number of events is purely background $N^i_{\text{obs}} = N^i_{\text{bkg}}$, thus also being a good estimator for the uncertainty.

of background events N^j_{bkg}, returning a result for $\frac{\mathrm{d}\sigma^{\text{IBU}}}{\mathrm{d}X^i}$ for simulated input. This procedure is repeated for every bin and for the entire set of toy experiments.

The set of toy experiments obtained by this procedure enables the calculation of a covariance matrix. The covariance matrix is computed as: [60]

$$\begin{aligned} C_{ij} &= \mathbb{E}\left[\left(N^i_{\text{unf}} - \overline{N^i_{\text{unf}}}\right)\left(N^j_{\text{unf}} - \overline{N^j_{\text{unf}}}\right)\right] \\ &= \frac{1}{N_{\text{Toys}} - 1} \sum_{k=1}^{N_{\text{Toys}}} \left(N^{i,k}_{\text{unf}} - \overline{N^i_{\text{unf}}}\right)\left(N^{j,k}_{\text{unf}} - \overline{N^j_{\text{unf}}}\right) , \end{aligned} \tag{5.15}$$

where $N^{i(j),k}_{\text{unf}}$ are obtained from equation (5.11). $\overline{N^{i(j)}_{\text{unf}}}$ are the mean values in the corresponding bins i, j of all unfolded toy experiments: $\overline{N^i_{\text{unf}}} = 1/N_{\text{Toys}} \cdot \sum_{\text{toys}} N^i_{\text{unf}}$. The uncertainty associated to the unfolded event yield in each bin is obtained from the covariance matrix:

$$\Delta N^i_{\text{unf}} = \sqrt{C_{ii}} \; . \tag{5.16}$$

while off-diagonal elements allow a description of correlations between different bins. Off-diagonal elements are potentially needed for various other calculations, e.g. for normalized differential cross-sections (see appendix C.2 for further details). Conversions from event yields to differential cross-sections in the covariance matrix are possible by dividing the respective parantheses in equation (5.15) by the corresponding factor $1/\left(\mathcal{L} \cdot \Delta X\right)$.

An estimate for the covariance matrix also allows a χ^2-computation, being defined as: [60]

$$\chi^2 = \vec{v}^{\mathrm{T}} C^{-1} \vec{v} \; . \tag{5.17}$$

The components of the vector $\vec{v}$ are differences between the bin contents of the unfolded distribution and a distribution that the unfolded result should be compared to; e.g. a truth level distribution obtained from MC (see section 5.3.3).

Note that a_j, ϵ_i and M^{-1}_{ij} from equation (5.11) are not generated as random numbers when creating toy experiments. The uncertainties of a_j and ϵ_i have been determined using Clopper-Pearson intervals. [73] Clopper-Pearson intervals can be used to estimate the uncertainties on quantities that have a limited range (see appendix D for a mathematical description). However, as a_j and ϵ_i are determined from MC only, the event yields used to determine these quantities have been rescaled by $1/\langle w_{\text{lumi}} \rangle$, where $\langle w_{\text{lumi}} \rangle$ denotes the mean value of w_{lumi} from all MC events. Due to time constraints, no uncertainties were assigned to M^{-1}_{ij}. Studies on how to include uncertainties of a_j and ϵ_i have been performed, but were not completely successful. These results can be found in appendix D.

5.3.3 Choice of binning and number of iterations

An important aspect to consider in general for unfolding processes is the choice of an appropriate binning, mainly to ensure that each bin contains enough statistics to return trustworthy results. Furthermore, in the case of iterative Bayesian unfolding, it is important to have a good estimate of the number of iterations that the procedure should be applied, in order to get a reasonable estimate of M^{-1}_{ij} without picking up on statistical fluctuations. For these reasons, several criteria similar to those in previous theses [70, 74] are implemented in the developed unfolding framework:

- Both migration and unfolding matrix determined are required to have diagonal elements larger than 60 %. This threshold was set, as previous theses on differential cross-section measurements [70, 74] noted unstable behavior in the application of iterative Bayesian unfolding if diagonal elements in the respective matrices became too small.

- The binning is chosen as the one tested with the smallest χ^2-value when comparing a MC unfolded distribution to the truth level distribution.

- The number of bins needs to be specified by the user and is a priori not restricted. However, a further requirement is that the binning chosen allows for Gaussian-like pull plots (a thorough explanation is given in section 6.2.3, also see [67]). This constrains possible bin widths allowed for each bin. It can also be used to determine a maximal number of bins which still fulfill this criterion, as otherwise pull plots deviate from Gaussian behavior.

- The number of iterations is specified by two criteria which need to be fulfilled at the same time. First, a χ^2-comparison between the current and the previous iteration must satisfy $\chi^2/\mathrm{NDF} < 0.1$ (NDF corresponds to the Number of Degrees of Freedom). Note that the usual requirement of $\chi^2/\mathrm{NDF} \approx 1$ is not applied, as results from consecutive iterations are expected to asymptotically approach each other in the case of infinitely many iterations. [69] Furthermore, to avoid cases of extreme deviations in one particular bin, each bin must satisfy that the relative difference between the bin contents of the current and previous iteration is smaller than 5 %.

These steps revealed promising results both from validation tests and unfolding data for the p_T-distribution of the top-quark (see sections 6.2 and 6.3).[3]

An exception to this approach was used when unfolding the $|\eta|$-distribution of the top-quark. As specified in more detail in section 6.2.1, MC agreement between unfolded and truth level distribution was best when using one iteration only. This case can however not be covered by the previous approach due to the comparison to the unfolded result to the previous iteration. As a consequence, for the $|\eta|$-distribution of the top-quark, only one iteration was used. Furthermore, no binning could be found in which all diagonal elements of the unfolding matrix fulfilled $M^{-1}_{ii} > 0.6$. From all the binnings tested, the binning was chosen which allowed the highest values for M^{-1}_{ii}, while allowing Gaussian-like pull plots at the same time.

5.3.4 Normalized differential cross-sections

Another concept often defined is a *normalized differential cross-section*. Defining a *fiducial cross-section* as:

$$\sigma^{\mathrm{fid}} = \sum_{i=1}^{N_{\mathrm{bins}}} \frac{\mathrm{d}\sigma}{\mathrm{d}X^i} \cdot \Delta X^i = \sum_{i=1}^{N_{\mathrm{bins}}} \frac{N^i_{\mathrm{unf}}}{\mathcal{L}} , \tag{5.18}$$

---

[3] The same approach was used by my successor Nilima Akolkar, giving satisfactory results when unfolding the transverse momentum of the W boson.

normalized differential cross-sections are determined from:

$$\frac{1}{\sigma^{\text{fid}}} \cdot \frac{\mathrm{d}\sigma}{\mathrm{d}X^i} = \frac{N^i_{\text{unf}}}{\Delta X^i \cdot \sum_{j=1}^{N_{\text{bins}}} N^j_{\text{unf}}} \,. \tag{5.19}$$

Normalized differential cross-sections have two main advantages:

- They allow for shape comparisons, as by default the integral of the histograms equals 1.
- Due to the division, systematic uncertainties that enter both σ^{fid} and $\frac{\mathrm{d}\sigma}{\mathrm{d}X^i}$ do cancel. Normalized differential cross-sections are therefore a good starting point to investigate systematic effects introduced by the unfolding procedure itself.

Due to some unresolved problems with covariance matrices, the statistical uncertainties assigned to normalized cross-sections shown in section 6.3 are determined using Gaussian error propagation only. A full description of attempts to include correlations between different variables from equation (5.19) is given in appendix C.2.

CHAPTER 6

Results

Acceptances, efficiencies and unfolding matrices used to measure differential cross-sections are presented. Several validation tests are shown to demonstrate that the unfolding framework as described in section 5.3 performs as expected. Results when applying the validated framework to data are revealed afterwards. Problems faced with covariance matrices are discussed in the end.

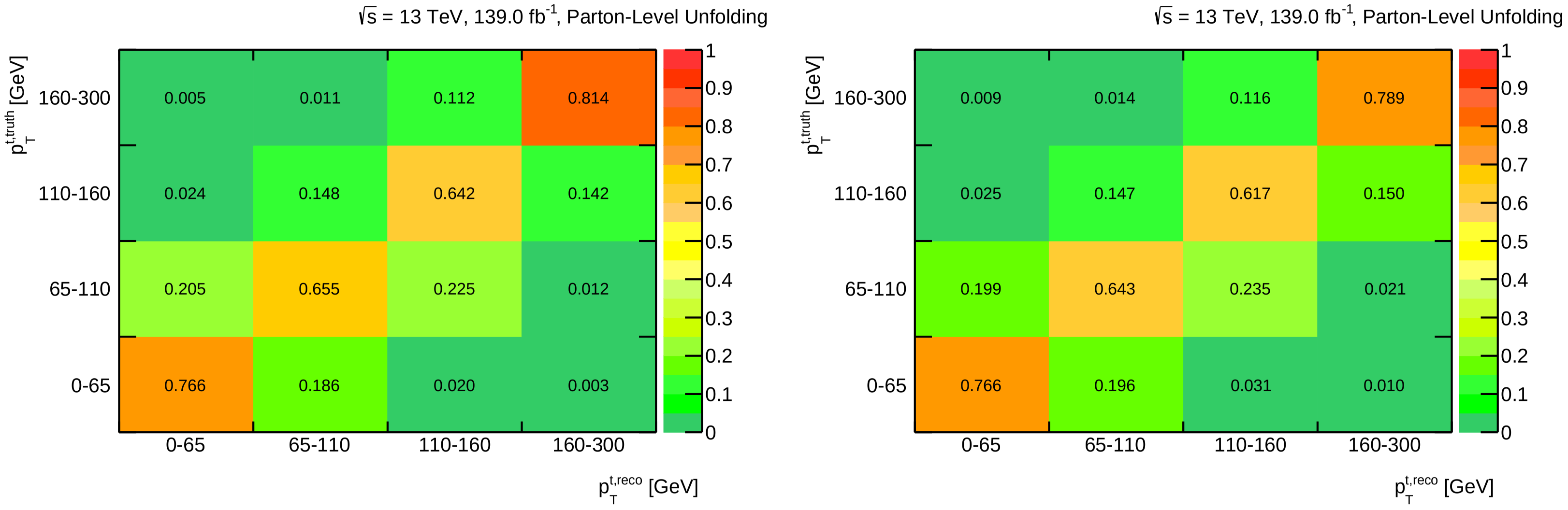
As proof of concept, p_T and $|\eta|$ of the top-quark are unfolded at parton level (these variables are denoted as p_T^t and $|\eta^t|$ hereafter). While other variables especially related to the Z boson would potentially be of interest, this information was not encoded in the ntuples used for this analysis, and would require new program code for the `AnalysisTop` framework. [48] Therefore, differential cross-sections for variables related to the Z boson have not been measured.

6.1 Determination of binning and number of iterations

The binning for the unfolded distribution and the number of iterations that are used for iterative Bayesian unfolding was determined by splitting the MC dataset in two independent samples of roughly the same size. One of these independent samples was used to determine ϵ_i, a_j and M_{ij}^{-1} as defined for equation (5.11). The other sample was used to mimic a data sample which is unfolded at a later stage for validation purposes.

The procedure described in section 5.3.3 has been followed. For p_T^t, 3 iterations for iterative Bayesian unfolding and 4 bins with a binning of $[0, 65, 110, 160, 300]$ (all values in GeV) have been chosen. A cut-off value for the highest allowed maximum value of p_T^t is required, as according to equation (5.11), the computation of a differential cross-section requires a division by the bin width ΔX^i. For this reason, no overflow bins are considered in all of the following plots. The highest value for p_T^t has been chosen to be 300 GeV, as fewer than 5 % of simulated events had a higher value of p_T^t. For $|\eta^t|$, only 1 iteration has been chosen, with a binning of $[0, 0.4, 1.0, 1.9, 4.5]$. The cut-off value corresponds to the highest value of $|\eta|$ that the ATLAS detector covers.

Migration matrices determined for these binnings are shown in figure 6.1. The corresponding unfolding matrices obtained after the given number of iterations are displayed in figure 6.2. Although off-diagonal elements of the matrices shown in figure 6.1 are smaller than diagonal elements, off-diagonal elements are non-zero, justifying the use of iterative Bayesian unfolding. Acceptances and efficiencies as defined in equation (5.11) are shown in figures 6.3 and 6.4. Confidence intervals are computed using Clopper-Pearson intervals at 68 % confidence level.

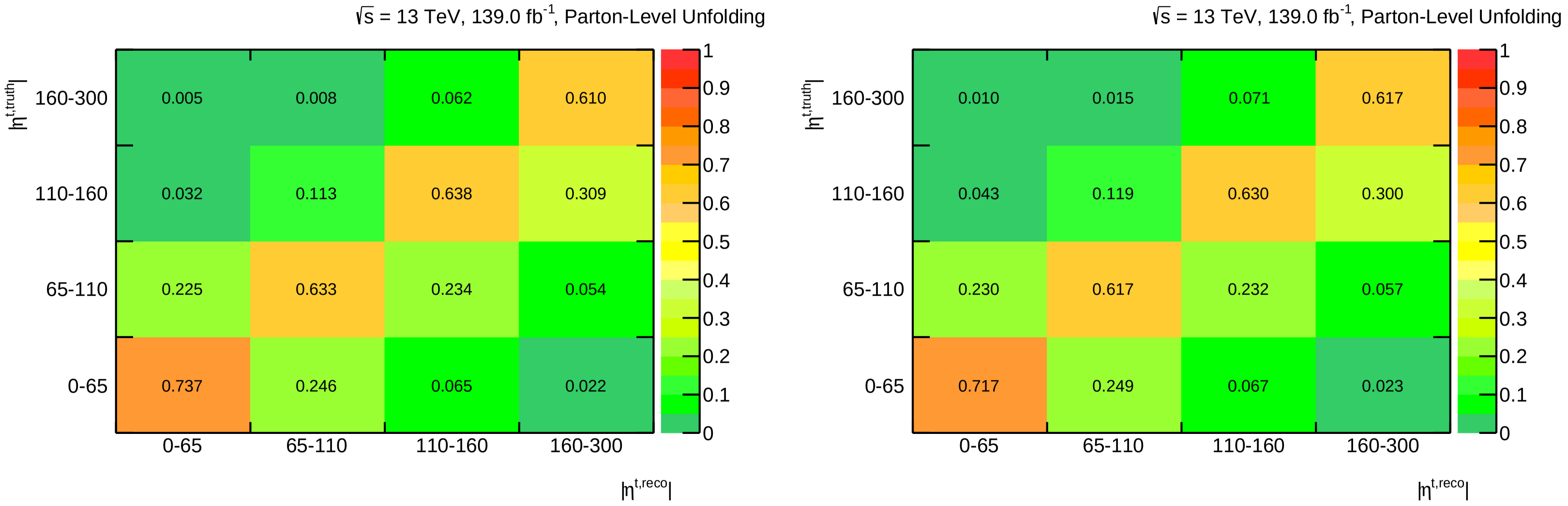


$p_T^{t,truth}$ [GeV] \ $p_T^{t,reco}$ [GeV]	0-65	65-110	110-160	160-300
160-300	0.005	0.011	0.112	0.814
110-160	0.024	0.148	0.642	0.142
65-110	0.205	0.655	0.225	0.012
0-65	0.766	0.186	0.020	0.003

$p_T^{t,truth}$ [GeV] \ $p_T^{t,reco}$ [GeV]	0-65	65-110	110-160	160-300
160-300	0.009	0.014	0.116	0.789
110-160	0.025	0.147	0.617	0.150
65-110	0.199	0.643	0.235	0.021
0-65	0.766	0.196	0.031	0.010

(a) Migration Matrix for p_T^t, applying a cut of $O_{NN} > 0.6$ to MC.

(b) Migration Matrix for p_T^t, applying a cut of $O_{NN} > 0$ to MC.



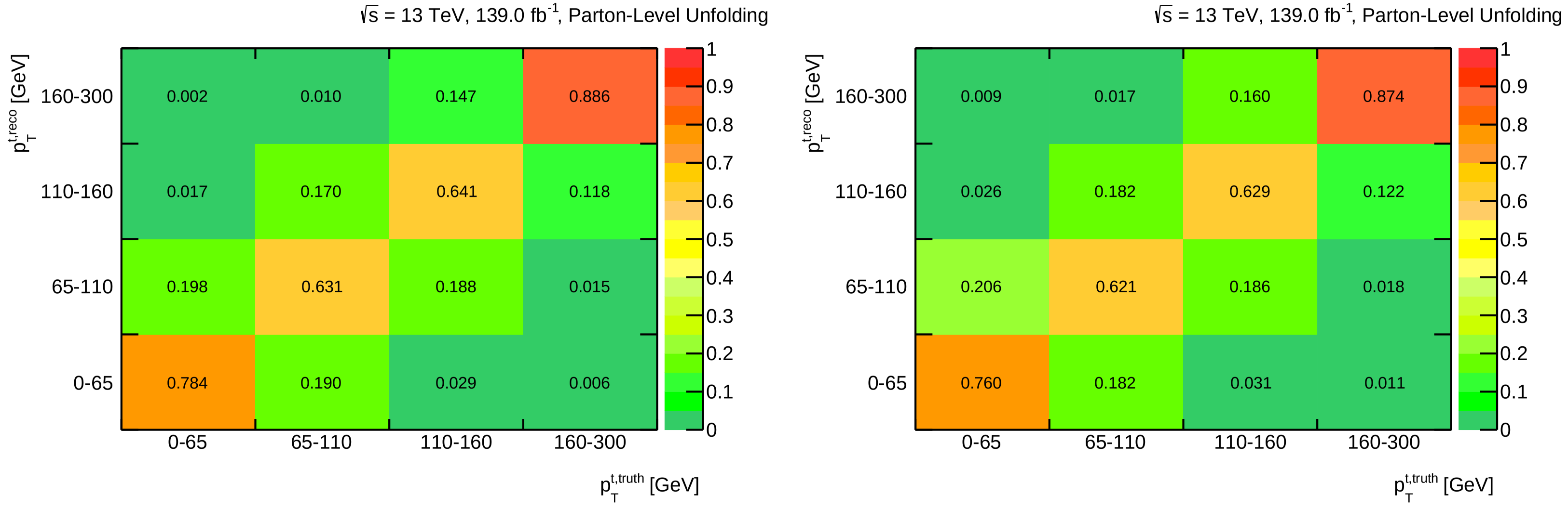
$\lvert\eta^{t,truth}\rvert$ \ $\lvert\eta^{t,reco}\rvert$	0-65	65-110	110-160	160-300
160-300	0.005	0.008	0.062	0.610
110-160	0.032	0.113	0.638	0.309
65-110	0.225	0.633	0.234	0.054
0-65	0.737	0.246	0.065	0.022

$\lvert\eta^{t,truth}\rvert$ \ $\lvert\eta^{t,reco}\rvert$	0-65	65-110	110-160	160-300
160-300	0.010	0.015	0.071	0.617
110-160	0.043	0.119	0.630	0.300
65-110	0.230	0.617	0.232	0.057
0-65	0.717	0.249	0.067	0.023

(c) Migration Matrix for $|\eta^t|$, applying a cut of $O_{NN} > 0.6$ to MC.

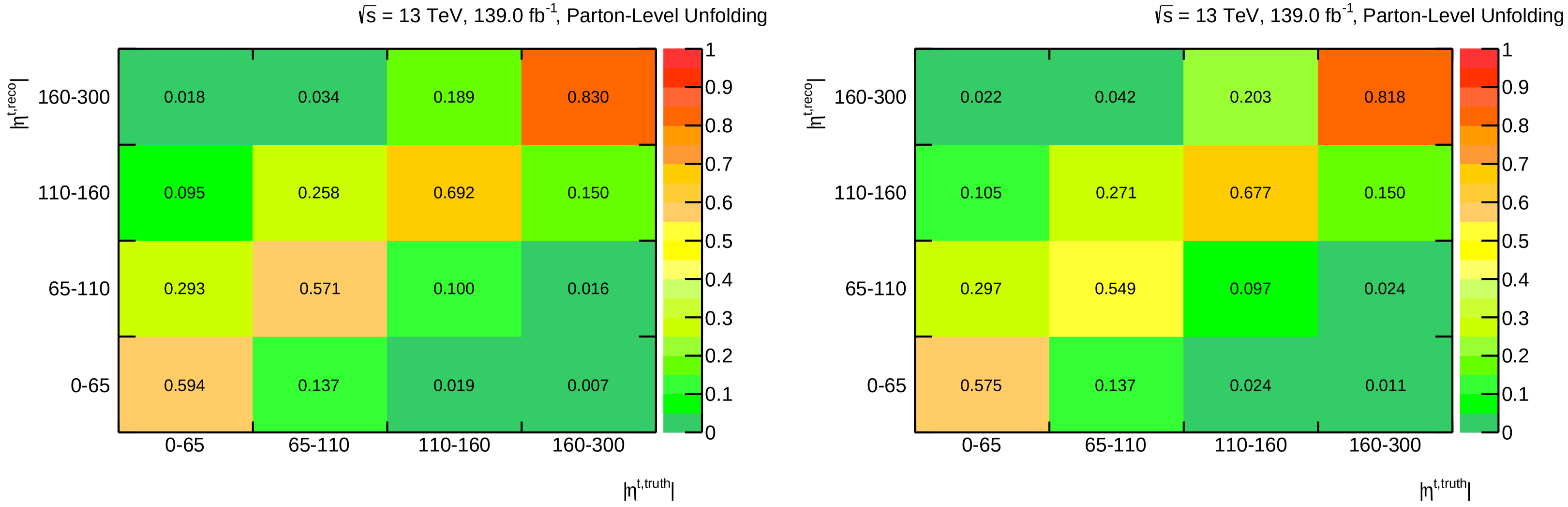
(d) Migration Matrix for $|\eta^t|$, applying a cut of $O_{NN} > 0$ to MC.

Figure 6.1: Overview of the different migration matrices used as input for the determination of unfolding matrices shown in figure 6.2. Figures 6.1(a) and 6.1(b) show migration matrices used for p_T^t. Figures 6.1(c) and 6.1(d) show migration matrices used for $|\eta^t|$. See individual captions for further details.



(a) Unfolding Matrix for p_T^t, applying a cut of $O_{NN} > 0.6$ to MC. 3 iterations were used when applying iterative Bayesian unfolding.

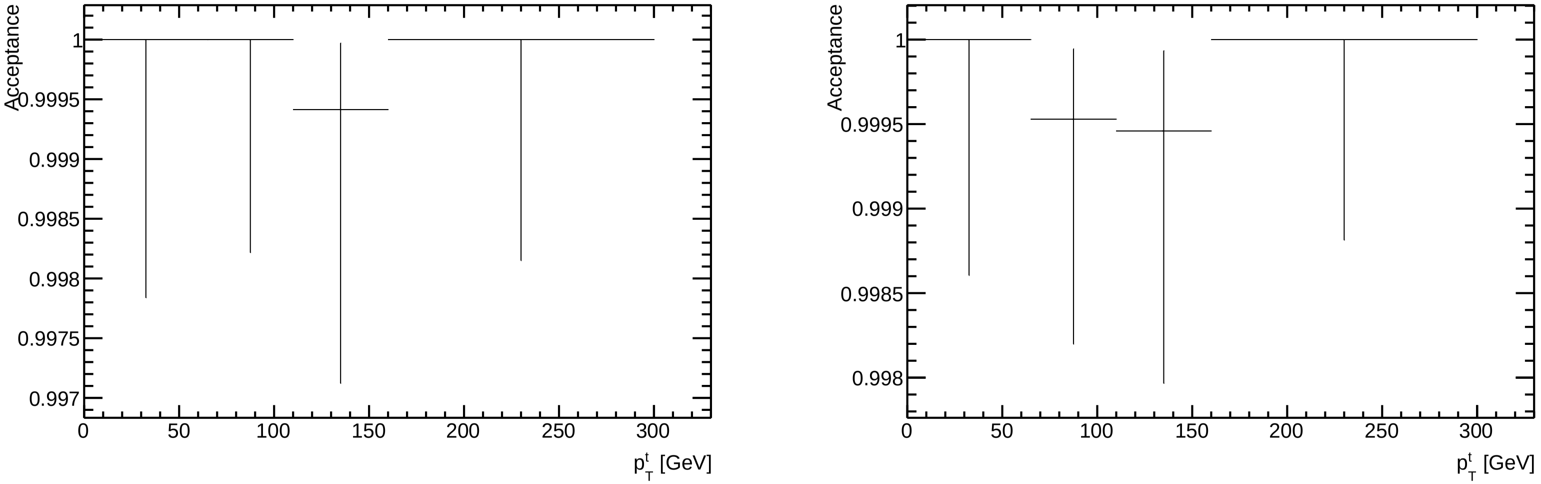
(b) Unfolding Matrix for p_T^t, applying a cut of $O_{NN} > 0$ to MC. 3 iterations were used when applying iterative Bayesian unfolding.



(c) Unfolding Matrix for $|\eta^t|$, applying a cut of $O_{NN} > 0.6$ to MC. 1 iteration was used when applying iterative Bayesian unfolding. Note that some diagonal elements are smaller than 60 %.

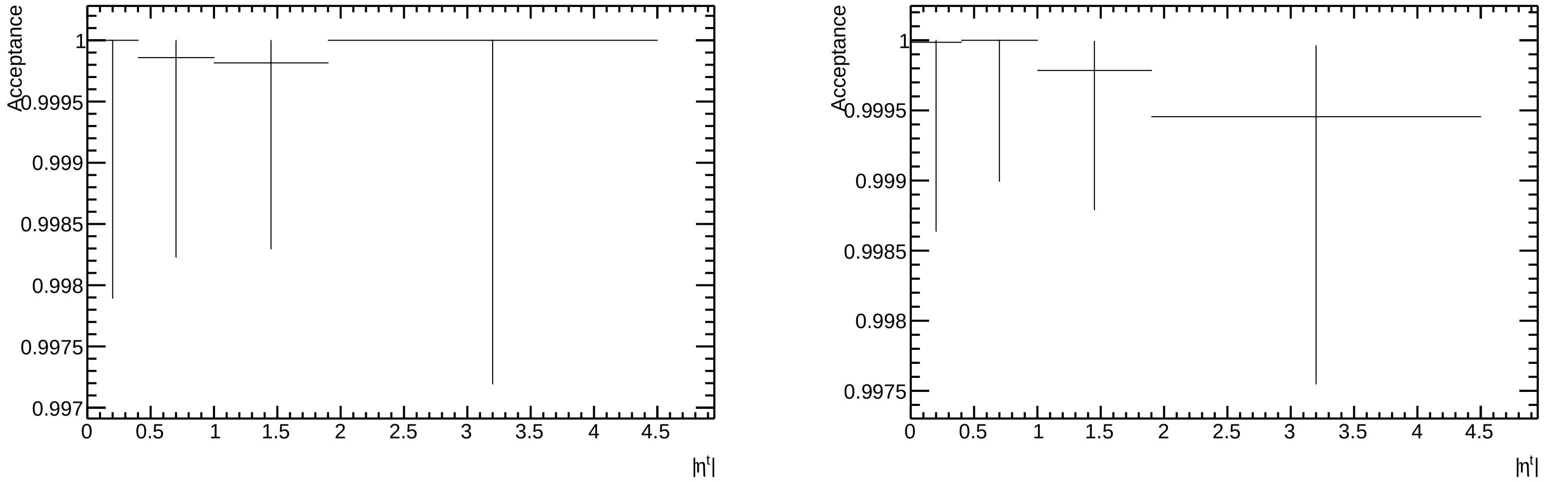
(d) Unfolding Matrix for $|\eta^t|$, applying a cut of $O_{NN} > 0$ to MC. 1 iteration was used when applying iterative Bayesian unfolding. Note that some diagonal elements are smaller than 60 %.

Figure 6.2: Overview of the different unfolding matrices used for the computation of differential cross-sections. Figures 6.2(a) and 6.2(b) show unfolding matrices used for p_T^t. Figures 6.2(c) and 6.2(d) show unfolding matrices used for $|\eta^t|$. See individual captions for further details.



(a) Acceptances for p_T^t, applying a cut of $O_{NN} > 0.6$ to MC.

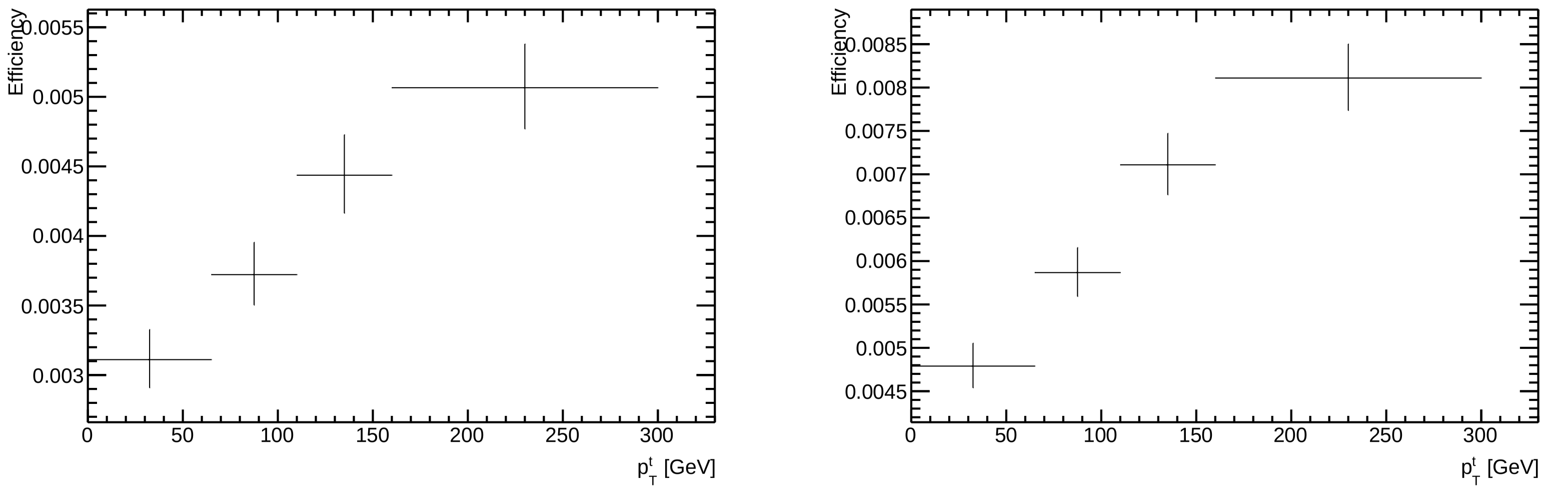
(b) Acceptances for p_T^t, applying a cut of $O_{NN} > 0$ to MC.



(c) Acceptances for $|\eta^t|$, applying a cut of $O_{NN} > 0.6$ to MC.

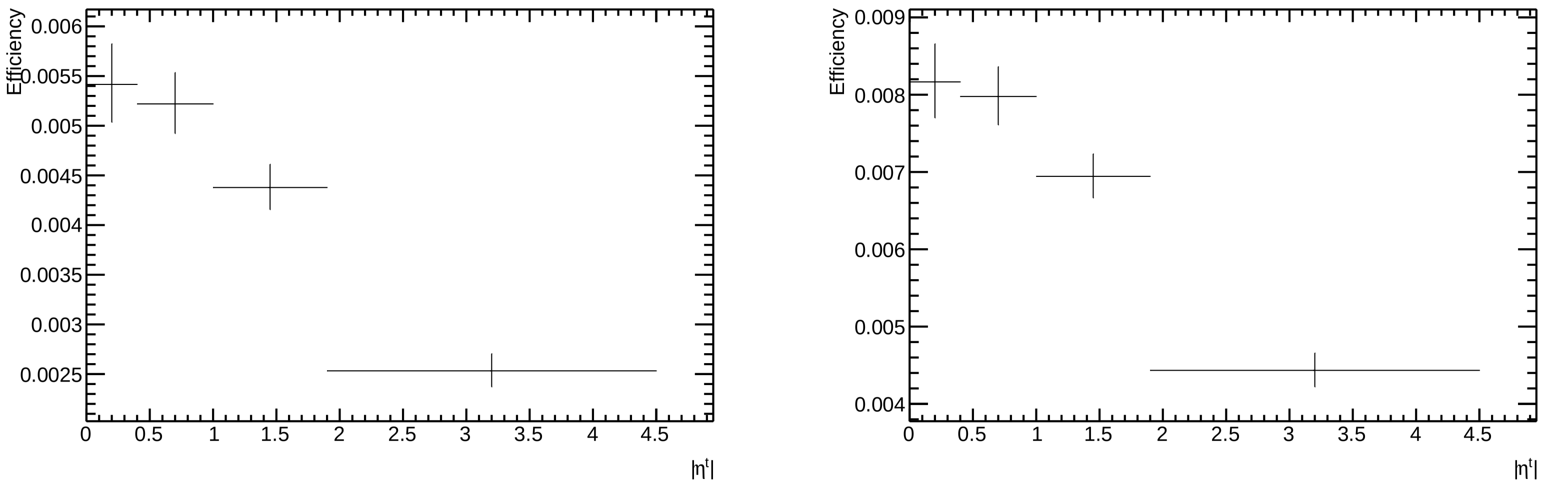
(d) Acceptances for $|\eta^t|$, applying a cut of $O_{NN} > 0$ to MC.

Figure 6.3: Overview of acceptances used for the computation of differential cross-sections. Figures 6.3(a) and 6.3(b) show acceptances used for p_T^t. Figures 6.3(c) and 6.3(d) show acceptances used for $|\eta^t|$. Uncertainties shown have been determined using Clopper-Pearson intervals at 68 % confidence level. See individual captions for further details.



(a) Efficiencies for p_{T}^{t}, applying a cut of $O_{\mathrm{NN}} > 0.6$ to MC.

(b) Efficiencies for p_{T}^{t}, applying a cut of $O_{\mathrm{NN}} > 0$ to MC.



(c) Efficiencies for $|\eta^{t}|$, applying a cut of $O_{\mathrm{NN}} > 0.6$ to MC.

(d) Efficiencies for $|\eta^{t}|$, applying a cut of $O_{\mathrm{NN}} > 0$ to MC.

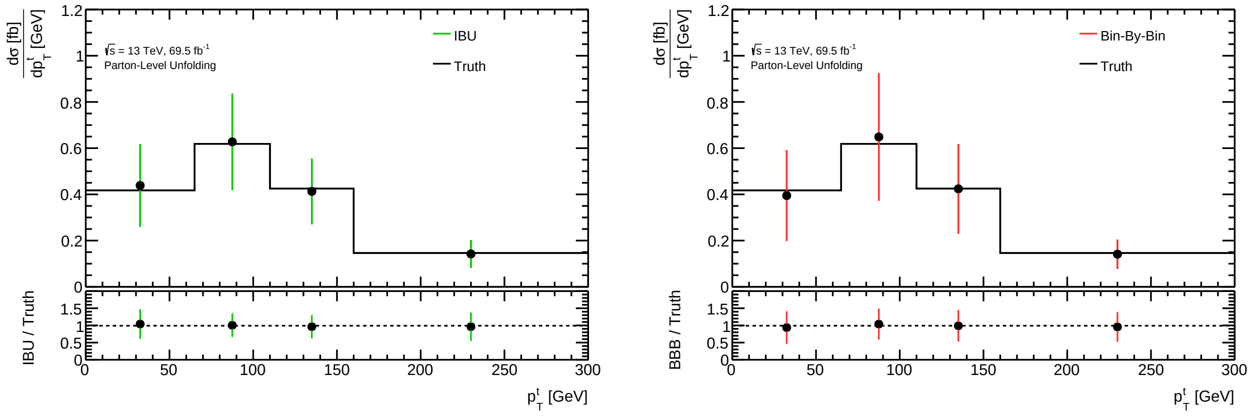
Figure 6.4: Overview of efficiencies used for the computation of differential cross-sections. Figures 6.4(a) and 6.4(b) show efficiencies used for p_{T}^{t}. Figures 6.4(c) and 6.4(d) show efficiencies used for $|\eta^{t}|$. Uncertainties shown have been determined using Clopper-Pearson intervals at 68 % confidence level. See individual captions for further details.

6.2 Validation tests

A series of tests similar to those recommended in [67] have been performed to show that the framework developed performs as anticipated. These validation tests are introduced in this section. Results obtained for p_{T}^{t} for MC events with $O_{\mathrm{NN}} > 0.6$ are shown, unless stated otherwise. Similar validation tests have been performed for $|\eta^t|$ (these results can be found in appendix B). Resulting distributions for both iterative Bayesian and bin-by-bin unfolding are shown.

6.2.1 Closure test

A closure test is performed by splitting available MC into two independent samples, labeled *iteration sample* and *test sample* in the following. Events from the iteration sample are used to determine ϵ_i, a_j and M_{ij}^{-1} as defined in equation (5.11) in case of iterative Bayesian unfolding, and to determine the correction factors K_i for bin-by-bin unfolding as defined in equation (5.5). The test sample is then used to mimic a data sample and is unfolded the same way data would to validate the method. The matrices, efficiencies and acceptances determined from the iteration sample are only valid if the unfolded distribution from the test sample matches its MC truth level distribution. The same matrices, efficiencies and acceptances determined this way are used to unfold real data afterwards. Results for a closure test are shown in figure 6.5. Good agreement between the unfolded distribution and the corresponding truth level distribution is observed for both iterative Bayesian and bin-by-bin unfolding.

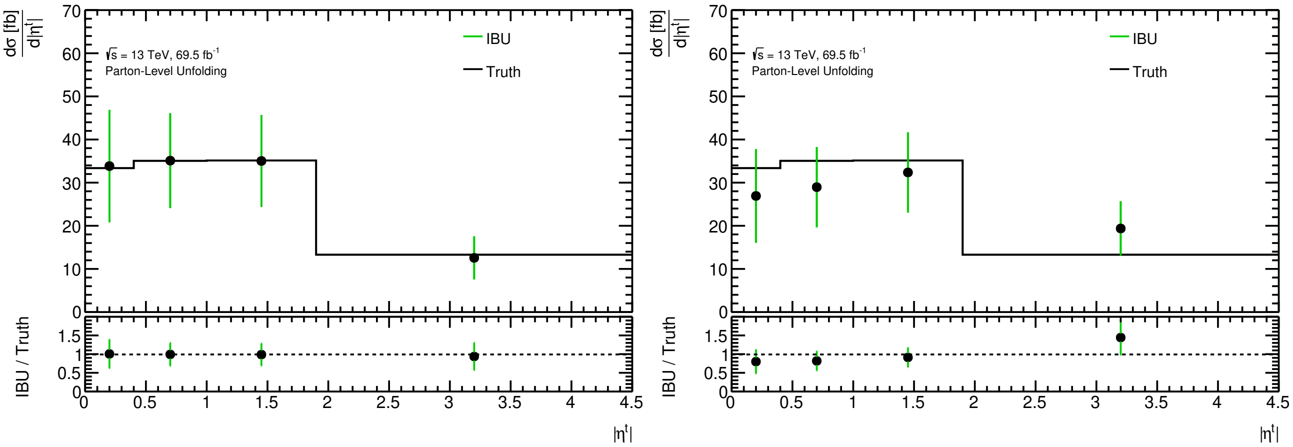


(a)

(b)

Figure 6.5: Illustration of a closure test. The resulting differential cross-section after application of iterative Bayesian unfolding is displayed in figure 6.5(a). For comparison purposes, a bin-by-bin unfolded distribution is shown in figure 6.5(b).

A closure test also revealed that only 1 iteration should be chosen for $|\eta^t|$, as only then the best agreement between MC truth level distribution and the unfolded result was obtained. For comparison purposes, the distributions obtained using iterative Bayesian unfolding with 1 and 3 iterations for $|\eta^t|$ are shown in figure 6.6. The exact reason for this behavior is unknown, but the use of more iterations when applying iterative Bayesian unfolding lead to an increased importance of matrix elements of the unfolding matrix related to the last bin. Matrix elements not related to this bin decreased with an increasing number of iterations. Such a behavior has been observed for all binnings tested.



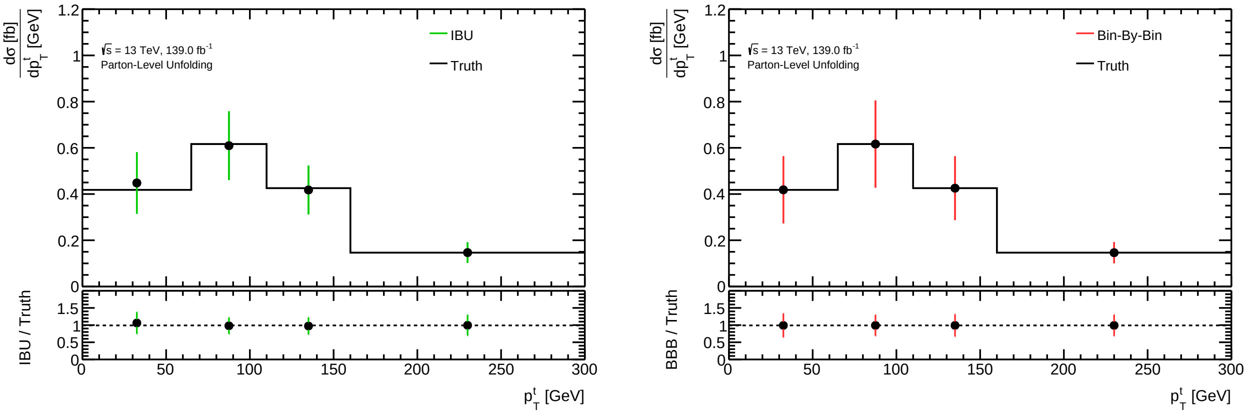
(a) (b)

Figure 6.6: Application of a closure test to $|\eta^t|$ for iterative Bayesian unfolding. Figure 6.6(a) shows the result obtained after 1 iteration, whereas figure 6.6(b) shows the resulting distribution after 3 iterations.

6.2.2 Consistency test

For a consistency test, the whole MC sample is used to determine ϵ_i, a_j and M_{ij}^{-1} instead of splitting the original MC sample. When using this same sample as data input, the unfolded distribution is expected to match the truth distribution exactly when using bin-by-bin unfolding, and nearly exactly in the case of iterative Bayesian unfolding. For iterative Bayesian unfolding, the result can deviate slightly due to the iterative nature of the method.

An example for this test is given in figure 6.7. A good agreement between unfolded and truth level distribution can be observed for the iterative Bayesian unfolding approach. By design, the bin-by-bin unfolded distribution matches the truth level distribution exactly.



(a) (b)

Figure 6.7: Illustration of a consistency test. The resulting differential cross-section after application of iterative Bayesian unfolding is displayed in figure 6.7(a). A bin-by-bin unfolded distribution is shown in figure 6.7(b).

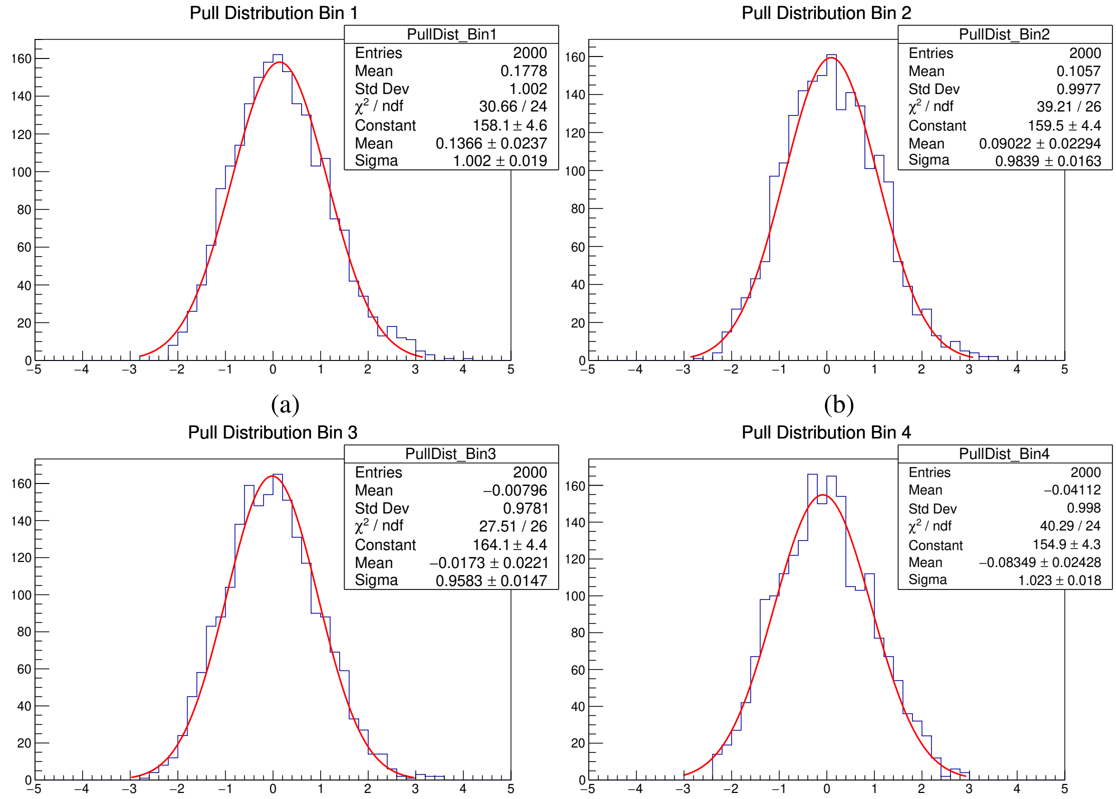
6.2.3 Pull test

The pull of bin i for an unfolded distribution is defined as: [67]

$$p_i = \frac{N^i_{\text{unf}} - N^i_{\text{truth}}}{\Delta N^i_{\text{unf}}} , \tag{6.1}$$

where ΔN^i_{unf} denotes the uncertainty assigned to the unfolded event yield.

The pull p_i for each bin i is calculated from a set of toy experiments after application of iterative Bayesian unfolding. The resulting distribution of multiple p_i's corresponds in the ideal case to a Gaussian distribution with a mean $\mu = 0$ and a width of $\sigma = 1$. [67] These requirements should be fulfilled for every bin. The uncertainty associated to each bin for each toy experiment is implemented as the uncertainty associated to the distribution as computed in section 6.1. A total of 2000 toy experiments are generated and the resulting distributions are plotted with 50 bins in a range between $[-5, 5)$. A Gaussian distribution is fitted to the plotted histogram in the range $[\mu_{\text{hist}} - 3, \mu_{\text{hist}} + 3]$, where μ_{hist} denotes the mean of the histogram. The range is limited to exclude the tails of the distribution with low statistics. No set of toy experiments was generated with $\chi^2/\text{NDF} > 5$. Therefore, there is no striking deviation from the expectation $\chi^2/\text{NDF} \approx 1$.



(a) (b) (c) (d)

Figure 6.8: Pull distributions for each bin for one set of toy experiments, with fitted Gaussian distributions.

The generation of 2000 toy experiments and subsequent fitting is repeated 10^3 times. A summary of the results of these 10^3 sets of toy experiments is given in figure 6.9. Points in this plot indicate the mean value of the mean of Gaussian distributions, while the uncertainties shown in this plot correspond to the mean value of the width of Gaussian distributions. The results are shown for each bin of the unfolded distribution. All the pulls lie inside a 1σ-region (as indicated by the region colored in green in figure 6.9), confirming that the pull test returns values as expected.

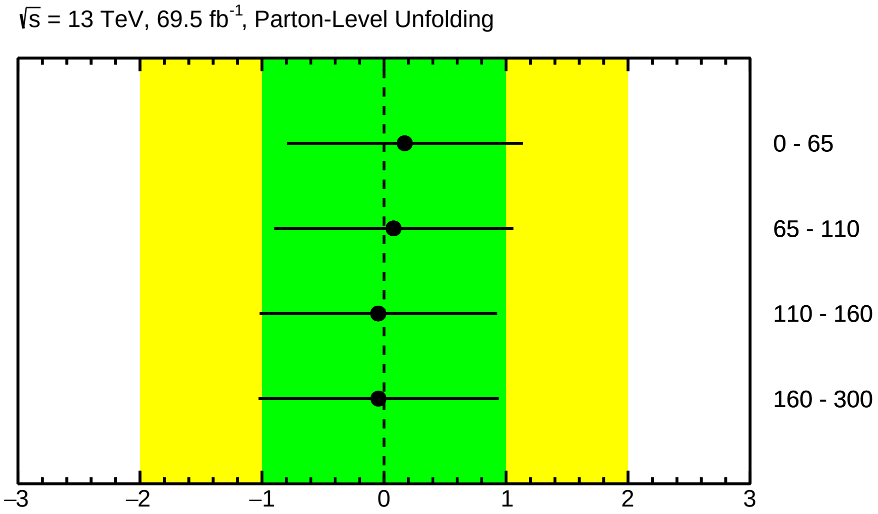


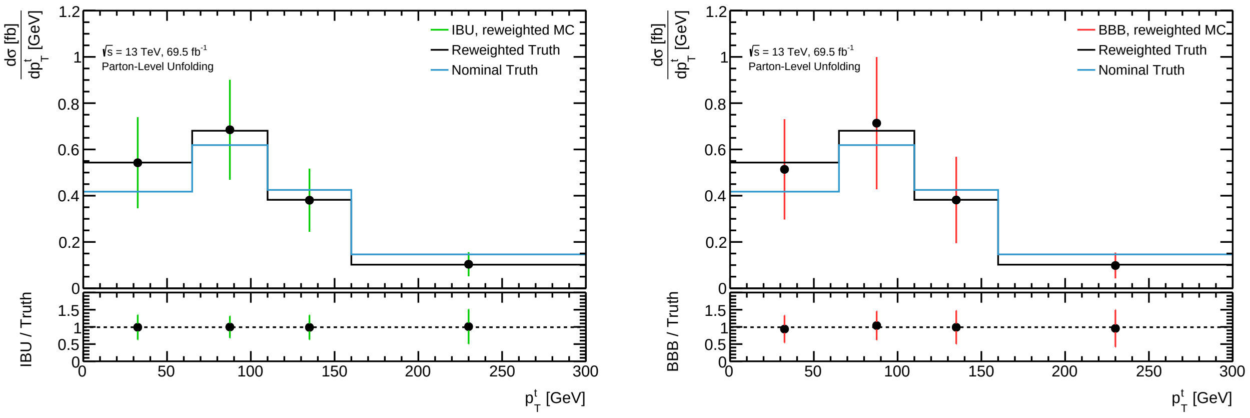
Figure 6.9: Summary of 1000 sets of toy experiments used for the generation of pull distributions, showing the result for each bin of the p_{T}^{t}-distribution (units given in GeV).

As the pull distributions are expected to follow a Gaussian distribution, the pull test can be used as a powerful tool to find an optimal binning. If the bin width has been chosen too small or if the number of bins is too large, then events would be too few to be distributed as a Gaussian distribution. A width smaller (larger) than $\sigma = 1$ can be can be interpreted as a over- (under-)estimation of the uncertainties associated to the corresponding bin.

6.2.4 Stress test

A stress test is performed to validate that the framework built based on the original MC sample is able to recover reweighted distributions or if it is biased towards the original MC distribution. To fulfill this goal, efficiencies, acceptances, and unfolding matrix determined for the original closure test are applied to a reweighted MC sample. The unfolded distribution is then compared to both the reweighted truth level distribution as well as to the original truth level distribution.

An example is shown in figure 6.10, where the MC distributions are reweighted bin-by-bin with [30 %, 10 %, −10 %, −30 %]. For both unfolding approaches, no clear statement can be made due to large uncertainties associated to the unfolded result. However, a trend that the unfolded distribution follows the reweighted truth level distribution instead of the original truth level distribution is clearly visible for both unfolding methods, validating that the framework works as expected.



(a)

(b)

Figure 6.10: Example of a stress test, where the MC distributions have been reweighted bin-by-bin with [30 %, 10 %, −10 %, −30 %]. The resulting unfolded distribution is compared to both the reweighted truth level as well as the original truth level distribution. Figure 6.10(a) shows the stress test when applying iterative Bayesian unfolding, while figure 6.10(b) shows the stress test for bin-by-bin unfolding.

6.3 Unfolding data

After several validation tests as described in section 6.2 confirmed the robustness of the framework, unfolding actual data was possible. Differential cross-sections and their normalized values are shown for p_T^t and $|\eta^t|$. Unfolding matrices, efficiencies and acceptances as described in section 6.1 have been used for their computation.

6.3.1 Top-quark p_T, $O_{NN} > 0.6$

Differential cross-sections for p_T^t after application of a cut $O_{NN} > 0.6$ are shown in figure 6.11. χ^2-comparisons between unfolded and truth level distributions yielded $\chi^2_{IBU}/NDF = 1.6/4$ for iterative Bayesian unfolding and $\chi^2_{BBB}/NDF = 2.4/4$ for bin-by-bin unfolding. Individual bin contents of the measured differential cross-sections are summarized in table 6.1.

p_T^t[GeV]	$\frac{d\sigma_{IBU}}{dp_T^t}$[fb/GeV]	$\frac{1}{\sigma}\frac{d\sigma_{IBU}}{dp_T^t}$[$10^{-3}$/GeV]	$\frac{d\sigma_{BBB}}{dp_T^t}$[fb/GeV]	$\frac{1}{\sigma}\frac{d\sigma_{BBB}}{dp_T^t}$[$10^{-3}$/GeV]
0 – 65	0.44 ± 0.13	4.30 ± 1.42	0.35 ± 0.14	3.48 ± 1.51
65 – 110	0.76 ± 0.15	7.46 ± 1.79	0.88 ± 0.19	8.83 ± 2.38
110 – 160	0.42 ± 0.10	4.14 ± 1.16	0.40 ± 0.14	3.98 ± 1.51
160 – 300	0.13 ± 0.04	1.27 ± 0.46	0.13 ± 0.05	1.27 ± 0.49

Table 6.1: Overview of the individual bin contents of the measured (normalized) differential cross-sections. Results are shown for p_T^t, after application of a cut $O_{NN} > 0.6$. Results from both iterative Bayesian and bin-by-bin unfolding are given. Only an estimate for the statistical uncertainty is provided.

6.3.2 Top-quark p_{T}, $O_{\mathrm{NN}} > 0$

Differential cross-sections for p_{T}^{t} after application of a cut $O_{\mathrm{NN}} > 0$ are shown in figure 6.12. χ^2-comparisons between unfolded and truth level distributions yielded $\chi^2_{\mathrm{IBU}}/\mathrm{NDF} = 1.1/4$ for iterative Bayesian unfolding and $\chi^2_{\mathrm{BBB}}/\mathrm{NDF} = 2.0/4$ for bin-by-bin unfolding. Individual bin contents of the measured differential cross-sections are summarized in table 6.2.

p_{T}^{t}[GeV]	$\frac{\mathrm{d}\sigma_{\mathrm{IBU}}}{\mathrm{d}p_{\mathrm{T}}^{t}}$[fb/GeV]	$\frac{1}{\sigma}\frac{\mathrm{d}\sigma_{\mathrm{IBU}}}{\mathrm{d}p_{\mathrm{T}}^{t}}$[$10^{-3}$/GeV]	$\frac{\mathrm{d}\sigma_{\mathrm{BBB}}}{\mathrm{d}p_{\mathrm{T}}^{t}}$[fb/GeV]	$\frac{1}{\sigma}\frac{\mathrm{d}\sigma_{\mathrm{BBB}}}{\mathrm{d}p_{\mathrm{T}}^{t}}$[$10^{-3}$/GeV]
0 – 65	0.34 ± 0.14	3.85 ± 1.63	0.27 ± 0.14	3.11 ± 1.68
65 – 110	0.62 ± 0.15	6.98 ± 2.02	0.71 ± 0.18	8.15 ± 2.48
110 – 160	0.37 ± 0.10	4.17 ± 1.32	0.33 ± 0.13	3.82 ± 1.59
160 – 300	0.15 ± 0.04	1.62 ± 0.56	0.15 ± 0.04	1.72 ± 0.58

Table 6.2: Overview of the individual bin contents of the measured (normalized) differential cross-sections. Results are shown for p_{T}^{t}, after application of a cut $O_{\mathrm{NN}} > 0$. Results from both iterative Bayesian and bin-by-bin unfolding are given. Only an estimate for the statistical uncertainty is provided.

6.3.3 Top-quark $|\eta|$, $O_{\mathrm{NN}} > 0.6$

Differential cross-sections for $|\eta^{t}|$ after application of a cut $O_{\mathrm{NN}} > 0.6$ are shown in figure 6.13. χ^2-comparisons between unfolded and truth level distributions yielded $\chi^2_{\mathrm{IBU}}/\mathrm{NDF} = 2.1/4$ for iterative Bayesian unfolding and $\chi^2_{\mathrm{BBB}}/\mathrm{NDF} = 2.5/4$ for bin-by-bin unfolding. Individual bin contents of the measured differential cross-sections are summarized in table 6.3.

$\lvert\eta^{t}\rvert$	$\frac{\mathrm{d}\sigma_{\mathrm{IBU}}}{\mathrm{d}\lvert\eta^{t}\rvert}$[fb]	$\frac{1}{\sigma}\frac{\mathrm{d}\sigma_{\mathrm{IBU}}}{\mathrm{d}\lvert\eta^{t}\rvert}$	$\frac{\mathrm{d}\sigma_{\mathrm{BBB}}}{\mathrm{d}\lvert\eta^{t}\rvert}$[fb]	$\frac{1}{\sigma}\frac{\mathrm{d}\sigma_{\mathrm{BBB}}}{\mathrm{d}\lvert\eta^{t}\rvert}$
0 – 0.4	29.3 ± 9.5	0.31 ± 0.11	24.9 ± 11.2	0.27 ± 0.13
0.4 – 1.0	38.8 ± 7.9	0.41 ± 0.10	44.5 ± 10.1	0.49 ± 0.15
1.0 – 1.9	37.0 ± 7.7	0.39 ± 0.10	35.9 ± 9.6	0.40 ± 0.13
1.9 – 4.5	10.2 ± 3.7	0.11 ± 0.04	8.4 ± 5.0	0.09 ± 0.06

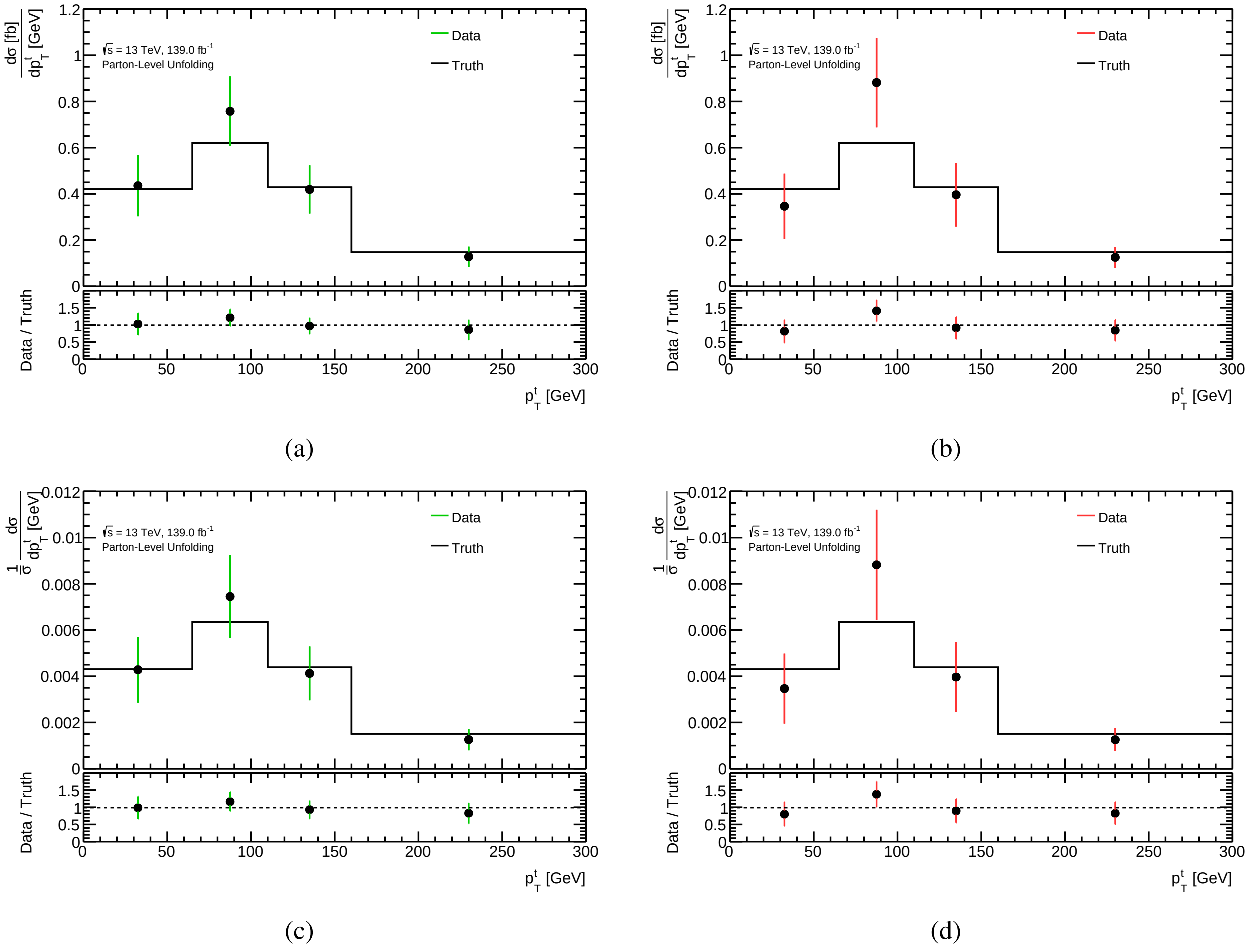
Table 6.3: Overview of the individual bin contents of the measured (normalized) differential cross-sections. Results are shown for $|\eta^{t}|$, after application of a cut $O_{\mathrm{NN}} > 0.6$. Results from both iterative Bayesian and bin-by-bin unfolding are given. Only an estimate for the statistical uncertainty is provided.

6.3.4 Top-quark $|\eta|$, $O_{\mathrm{NN}} > 0$

Differential cross-sections for $|\eta^{t}|$ after application of a cut $O_{\mathrm{NN}} > 0$ are shown in figure 6.14. χ^2-comparisons between unfolded and truth level distributions yielded $\chi^2_{\mathrm{IBU}}/\mathrm{NDF} = 1.6/4$ for iterative Bayesian unfolding and $\chi^2_{\mathrm{BBB}}/\mathrm{NDF} = 2.8/4$ for bin-by-bin unfolding. Individual bin contents of the measured differential cross-sections are summarized in table 6.4.

$\lvert\eta^t\rvert$	$\frac{d\sigma_{\text{IBU}}}{d\lvert\eta^t\rvert}$ [fb]	$\frac{1}{\sigma}\frac{d\sigma_{\text{IBU}}}{d\lvert\eta^t\rvert}$	$\frac{d\sigma_{\text{BBB}}}{d\lvert\eta^t\rvert}$ [fb]	$\frac{1}{\sigma}\frac{d\sigma_{\text{BBB}}}{d\lvert\eta^t\rvert}$
0 – 0.4	35.4 ± 9.6	0.41 ± 0.13	36.2 ± 10.5	0.42 ± 0.15
0.4 – 1.0	33.0 ± 8.0	0.38 ± 0.11	33.7 ± 9.6	0.39 ± 0.13
1.0 – 1.9	26.3 ± 7.9	0.31 ± 0.10	20.5 ± 9.2	0.24 ± 0.12
1.9 – 4.5	10.9 ± 3.6	0.13 ± 0.05	12.8 ± 4.7	0.15 ± 0.06

Table 6.4: Overview of the individual bin contents of the measured (normalized) differential cross-sections. Results are shown for $|\eta^t|$, after application of a cut $O_{\text{NN}} > 0$. Results from both iterative Bayesian and bin-by-bin unfolding are given. Only an estimate for the statistical uncertainty is provided.



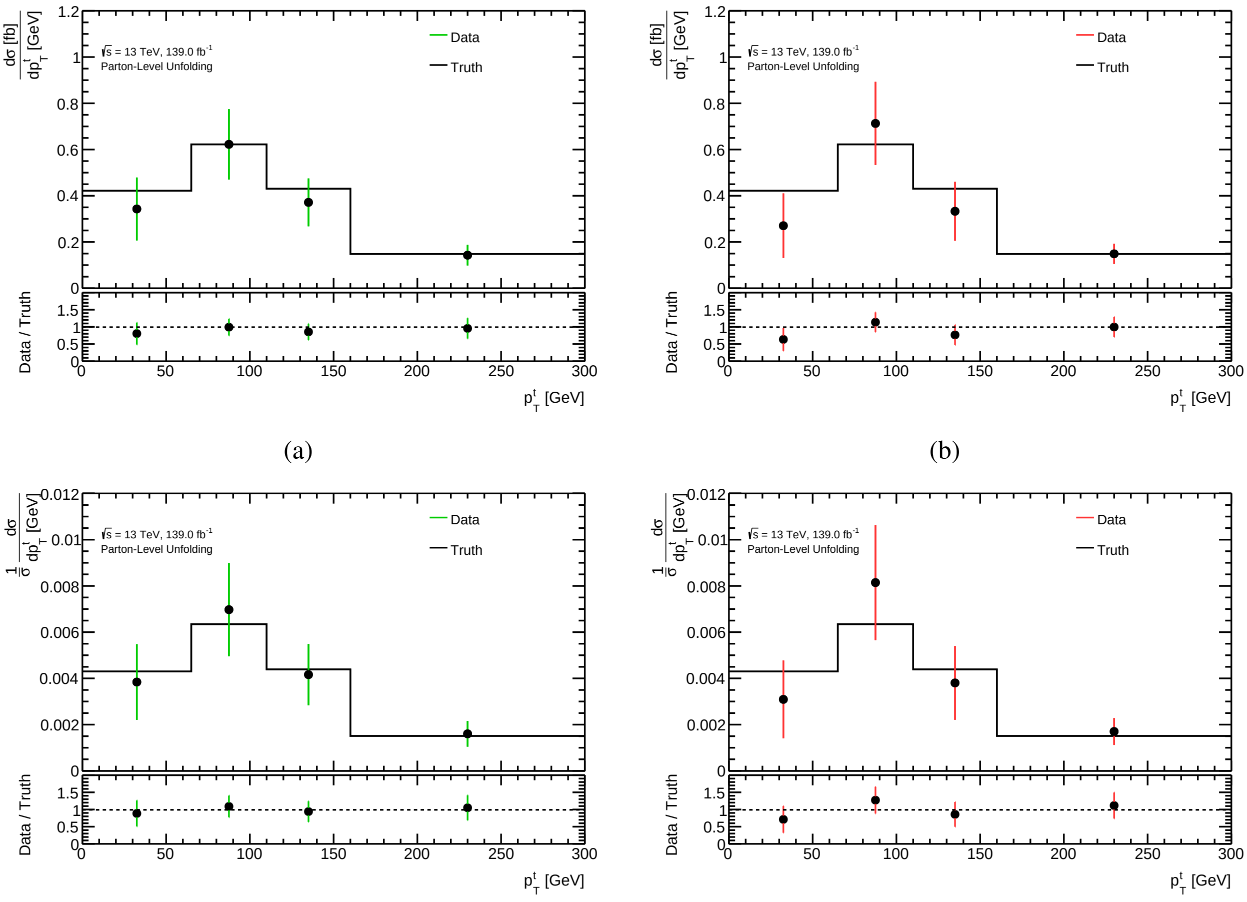
(a)

(b)

(c)

(d)

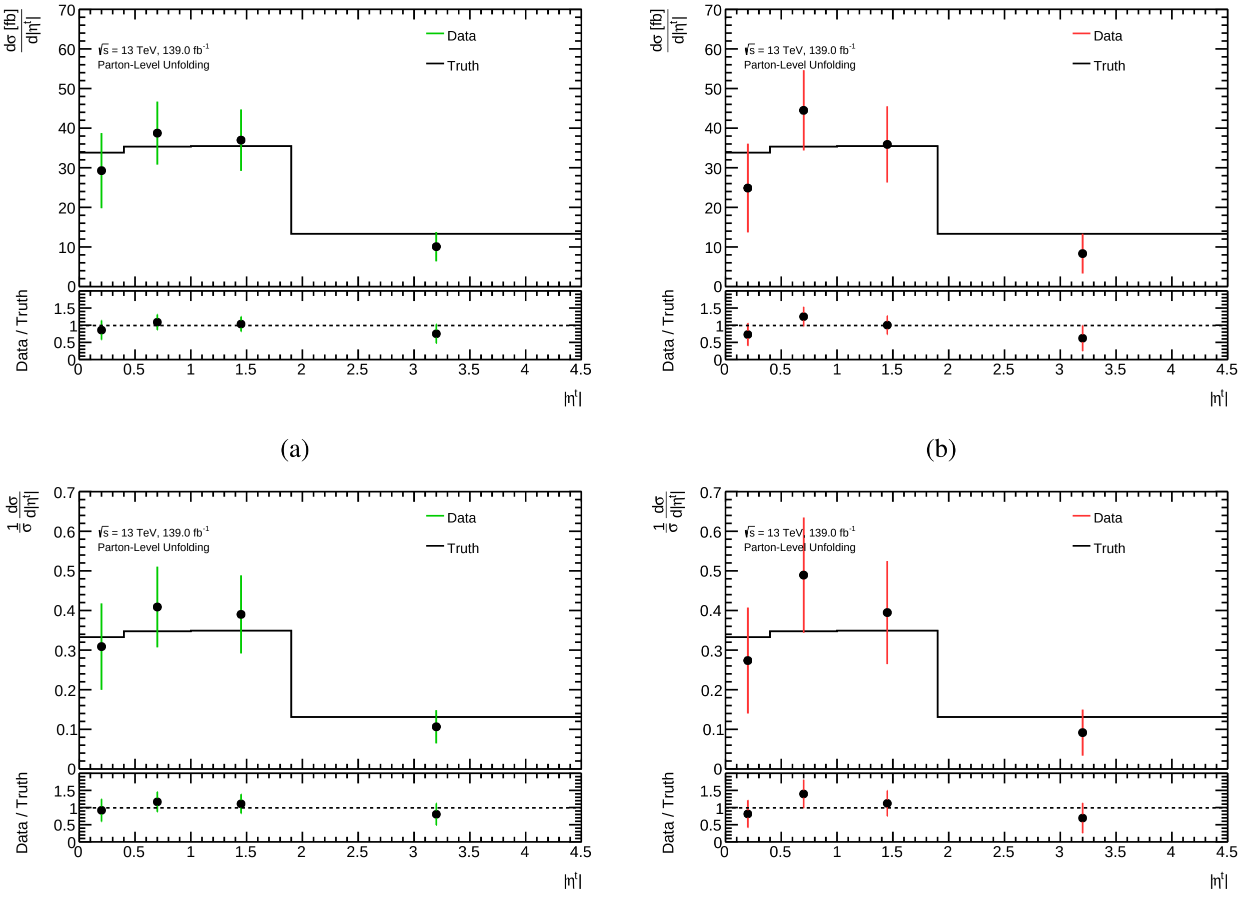
Figure 6.11: Unfolded p_T^t-distribution after application of a cut of $O_{\text{NN}} > 0.6$. Figure 6.11(a) shows the result after application of iterative Bayesian unfolding, figure 6.11(b) represents the bin-by-bin unfolded result. Figures 6.11(c) and 6.11(d) show the corresponding normalized differential cross-sections.



(a) (b)

(c) (d)

Figure 6.12: Unfolded p_T^t-distribution after application of a cut of $O_{NN} > 0$. Figure 6.12(a) shows the result after application of iterative Bayesian unfolding, figure 6.12(b) represents the bin-by-bin unfolded result. Figures 6.12(c) and 6.12(d) show the corresponding normalized differential cross-sections.



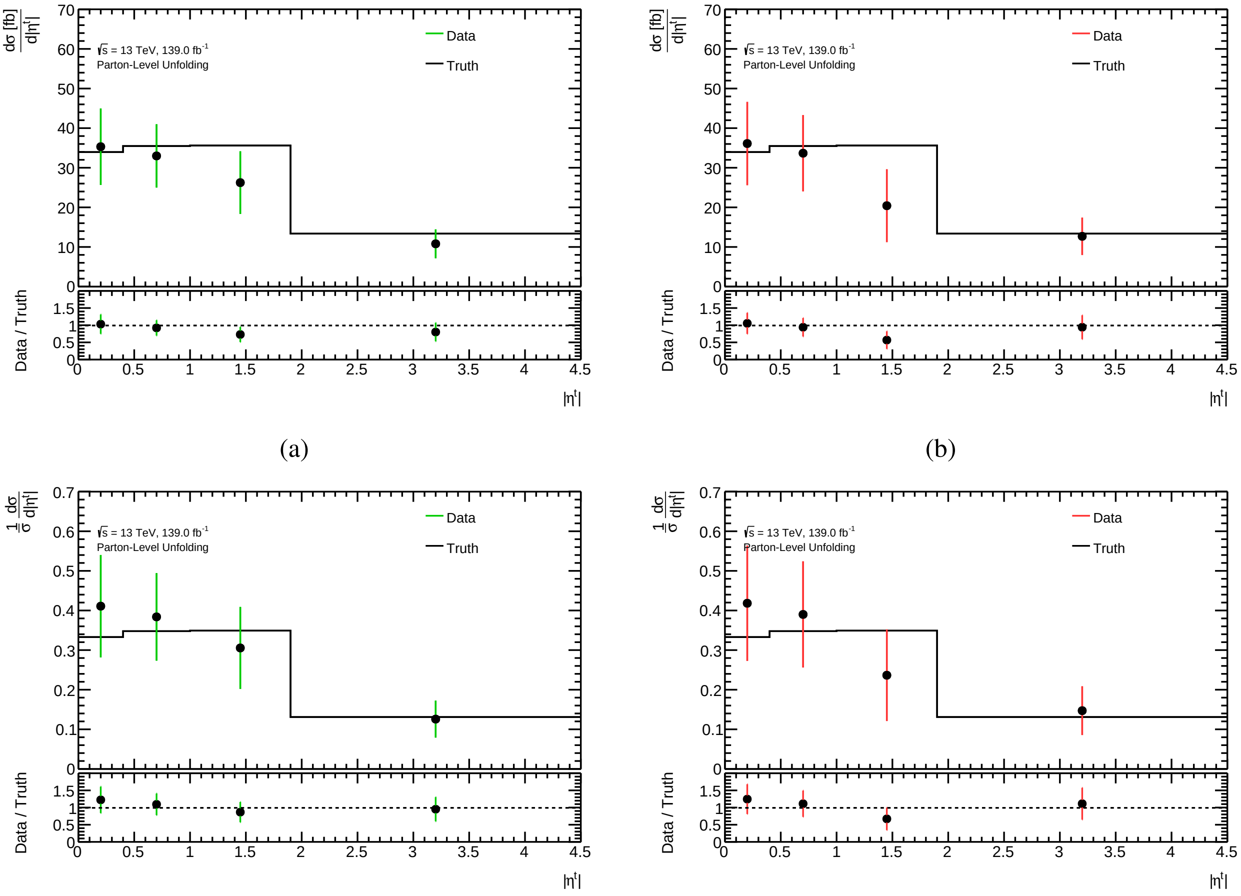
(a)

(b)

(c)

(d)

Figure 6.13: Unfolded $|\eta^t|$-distribution after application of a cut of $O_{\text{NN}} > 0.6$. Figure 6.13(a) shows the result after application of iterative Bayesian unfolding, figure 6.13(b) represents the bin-by-bin unfolded result. Figures 6.13(c) and 6.13(d) show the corresponding normalized differential cross-sections.



(a)

(b)

(c)

(d)

Figure 6.14: Unfolded $|\eta^t|$-distribution after application of a cut of $O_{\mathrm{NN}} > 0$. Figure 6.14(a) shows the result after application of iterative Bayesian unfolding, figure 6.14(b) represents the bin-by-bin unfolded result. Figures 6.14(c) and 6.14(d) show the corresponding normalized differential cross-sections.

6.4 Discussion of results

A short discussion about the results obtained from data is given. Problems not mentioned so far that were faced with covariance matrices are discussed.

6.4.1 General observations

- Unfolded distributions shown in section 6.3 agree with the MC truth level distribution for all the distributions that have been unfolded. This is in particular true for both unfolding methods that have been applied. Large uncertainties allow this agreement even for bin-by-bin unfolding, although non-zero off-diagonal elements in the migration matrices from figure 6.1 suggest that the use of bin-by-bin unfolding may not yield exact results.

- Further concrete statements are difficult to make, as statistical uncertainties (especially for bin-by-bin unfolding) can be larger than 50 % (as can e.g. be seen in the last bin from table 6.4). Note that standard deviations associated to iterative Bayesian unfolded distributions tend to be smaller than the uncertainties assigned to distributions obtained via bin-by-bin unfolding. The full information about uncertainties and correlations between bins is encoded in the covariance matrix, which has non-zero off-diagonal elements in the case of iterative Bayesian for all investigated distributions (see section 6.4.2).

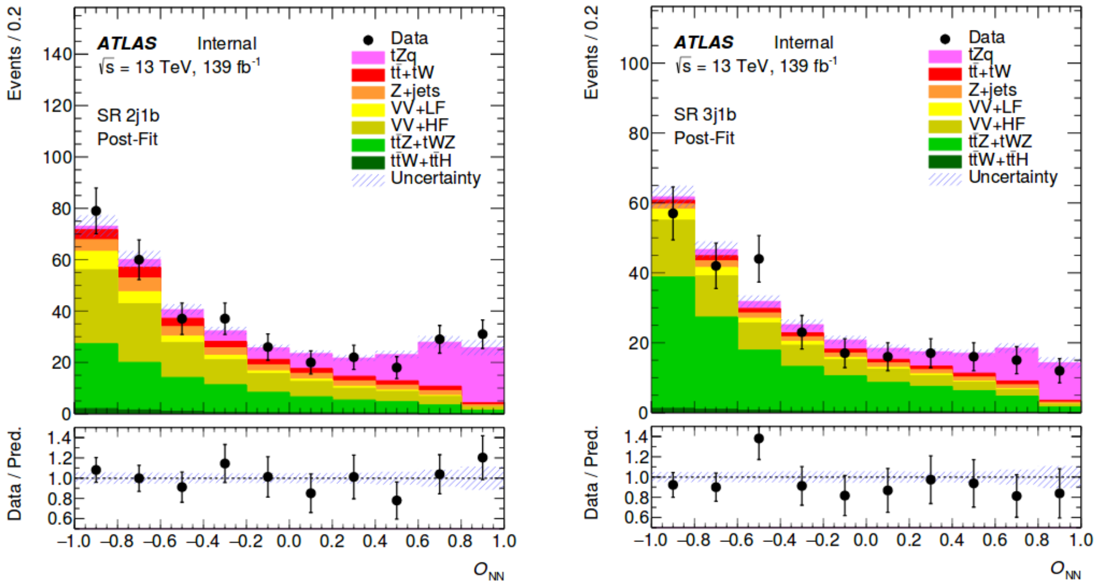
- For both variables, a trend has been observed such that the fiducial cross-section decreases when the threshold of the cut on O_{NN} was lowered. For iterative Bayesian unfolding, the fiducial cross-sections are:

  $$\begin{aligned} \sigma^{\mathrm{fid}}_{\mathrm{IBU},p_{\mathrm{T}}}\,(O_{\mathrm{NN}} > 0.6) &= (101.7 \pm 13.6)\,\mathrm{fb}\,, & \sigma^{\mathrm{fid}}_{\mathrm{IBU},p_{\mathrm{T}}}\,(O_{\mathrm{NN}} > 0) &= (89.3 \pm 13.8)\,\mathrm{fb} \\ \sigma^{\mathrm{fid}}_{\mathrm{IBU},|\eta|}\,(O_{\mathrm{NN}} > 0.6) &= (94.8 \pm 13.3)\,\mathrm{fb}\,, & \sigma^{\mathrm{fid}}_{\mathrm{IBU},|\eta|}\,(O_{\mathrm{NN}} > 0) &= (86.0 \pm 13.3)\,\mathrm{fb} \end{aligned} \tag{6.2}$$

  Note that uncertainties quoted assume that bins are uncorrelated. Further information about these correlations can be found in section 6.4.2 and appendix C. Fiducial cross-sections can also be determined from bin-by-bin unfolded distributions:

  $$\begin{aligned} \sigma^{\mathrm{fid}}_{\mathrm{BBB},p_{\mathrm{T}}}\,(O_{\mathrm{NN}} > 0.6) &= (100.0 \pm 15.7)\,\mathrm{fb}\,, & \sigma^{\mathrm{fid}}_{\mathrm{BBB},p_{\mathrm{T}}}\,(O_{\mathrm{NN}} > 0) &= (87.6 \pm 15.0)\,\mathrm{fb} \\ \sigma^{\mathrm{fid}}_{\mathrm{BBB},|\eta|}\,(O_{\mathrm{NN}} > 0.6) &= (90.9 \pm 17.4)\,\mathrm{fb}\,, & \sigma^{\mathrm{fid}}_{\mathrm{BBB},|\eta|}\,(O_{\mathrm{NN}} > 0) &= (86.4 \pm 16.4)\,\mathrm{fb} \end{aligned} \tag{6.3}$$

  It can be observed that a lower threshold for O_{NN} reduces the fiducial cross-section on average by ≈ 10 %. Further studies are required to check if this behavior is only due to statistical fluctuations or if it is a systematic effect. However, a comparison to the distributions of the neural network output (see figure 6.15) shows that data tends to be lower than MC prediction in the range of $O_{\mathrm{NN}} \in [0, 0.6]$. This behavior can be seen for both signal regions, and might therefore hint at the reason for the differences observed.



(a)

(b)

Figure 6.15: Comparison between the O_{NN}-distributions for data and MC. Figure 6.15(a) shows the distributions for SR-$2j1b$, while figure 6.15(b) shows the distributions for SR-$3j1b$. Figures taken from [21].

6.4.2 Covariance matrices

So far, only standard deviations have been shown. The calculation of the uncertainty assigned to fiducial cross-sections was performed under the assumption that different bins are uncorrelated. Full covariance matrices have been determined from toy experiments to assign uncertainties to unfolded distributions obtained via the iterative Bayesian approach. For all unfolded distributions, the covariance matrices are displayed in figure 6.16.

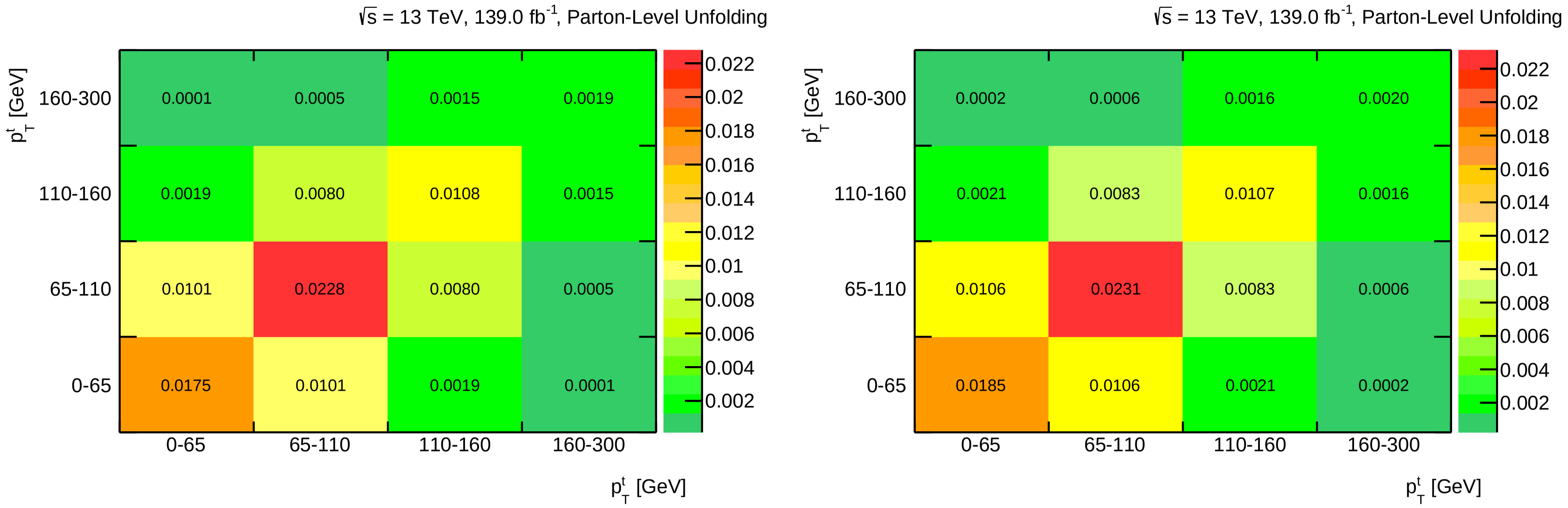
The matrices from figure 6.16 show opposite features compared to the behavior that is expected. According to [67], anti-correlations and thus negative entries are expected for off-diagonal elements of these matrices. This predicted behavior was also seen in previous differential cross-sections measurements for another process. [74] However, the matrices in figure 6.16 do not have any negative off-diagonal elements, indicating that the values of differential cross-sections between different bins are *correlated* rather than anti-correlated.

The reason for this behavior is unknown. Several tests have been performed, listed in appendix C.1. In a nutshell, all tests confirm a correct implementation of the random number generation needed for toy experiments and a correct calculation of the covariance matrix (this has been confirmed by use of toy experiments for bin-by-bin unfolding, for which no correlations are expected and have not been seen either, see appendix C.1). However, unexpected behavior has been seen for some distributions related to iterative Bayesian unfolding (especially in the distribution $(N_i^{\text{unf}} - \bar{N}_i^{\text{unf}}) \cdot (N_j^{\text{unf}} - \bar{N}_j^{\text{unf}})$ entering the computation of the covariance matrix, see equation (5.15)). Details can be found in appendix C.1. As a summary from all tests performed, two possibilities are found to be most likely:

- An undiscovered error remained in the framework built, which also did not manifest itself in any of the validation tests performed. This error would only be related to iterative Bayesian unfolding, as it was not observed for bin-by-bin unfolding.

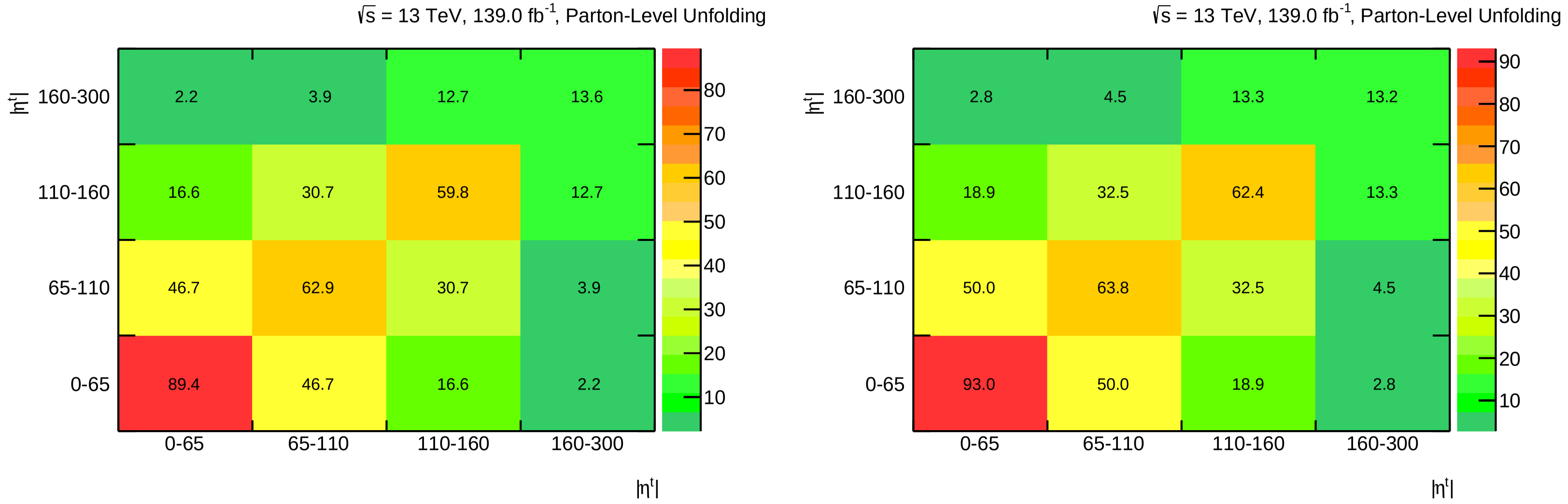
- No statement could be found in the literature if iterative Bayesian unfolding is valid for all regimes, or if it deviates from the predicted behavior in the case of low statistics. This is of particular interest for this analysis: as can be extracted from figure 5.1, only ≈ 64 events are predicted based on MC to pass a threshold of $O_{\mathrm{NN}} > 0.6$.

Further studies are required to fully understand the behavior of the covariance matrices from figure 6.16. Due to these (believed implausible) results, a full statistical uncertainty estimate of fiducial cross-sections has been postponed (results with the current covariance matrices are shown in appendix C.2). Wrong estimated uncertainties on the fiducial cross-sections according to equation (5.19) also have an impact on the uncertainties for normalized differential cross-sections (see appendix C.2 for a detailed description on how these uncertainties could fully be estimated with correct covariance matrices). Lastly, according to equation (5.17), covariance matrices also have an impact on all χ^2-values quoted in this chapter.



(a) Covariance matrix for the p_{T}^{t} differential cross-section, when applying a cut of $O_{\mathrm{NN}} > 0.6$.

(b) Covariance matrix for the p_{T}^{t} differential cross-section, when applying a cut of $O_{\mathrm{NN}} > 0$.



(c) Covariance matrix for the $|\eta^t|$ differential cross-section, when applying a cut of $O_{\mathrm{NN}} > 0.6$.

(d) Covariance matrix for the $|\eta^t|$ differential cross-section, when applying a cut of $O_{\mathrm{NN}} > 0$.

Figure 6.16: Overview of covariance matrices for different unfolded data distributions, as estimated from a set of toy experiments. None of these matrices show off-diagonal elements < 0.

CHAPTER 7

Conclusion

7.1 Summary

This thesis represents the first attempt to measure differential cross-sections of the single top-quark production in association with a Z boson, abbreviated tZq. Differential cross-sections for the transverse momentum, p_{T}^{t}, and the absolute value of the pseudorapidity, $|\eta^{t}|$, of the top-quark have been measured at parton level. The measurements presented in this thesis have been performed for an integrated luminosity of $139\,\mathrm{fb}^{-1}$ at $\sqrt{s} = 13\,\mathrm{TeV}$.

The starting point for this thesis was the framework used to measure the inclusive cross-section of the tZq process in the trilepton final state. This final state is characterized by three leptons, one b-tagged jet and one or two jets which are not b-tagged. For this thesis, neural networks (NN) are used for signal-background discrimination and for the definition of a phase space with high signal purity. The impact of the NN output on the differential cross-section has been probed by comparing the resulting distributions obtained for two different cuts on the NN output. For all binning strategies, distributions have been fitted such that one can extract the number of signal and background events from data in every bin independently.

A framework has been developed to measure differential cross-sections, relying on a technique called *unfolding*. Unfolding enables the user to revert smearing and inefficiencies introduced due to detector effects and selection cuts. Two unfolding methods have been implemented: bin-by-bin and iterative Bayesian unfolding. Normalized distributions have also been computed for both approaches.

A method to give a first estimate of statistical uncertainties has been provided as well. For bin-by-bin unfolding, no correlations between different bins are expected. The uncertainty can thus be estimated theoretically under the assumption that bin contents are distributed according to a Poisson distribution. For distributions obtained with the iterative Bayesian approach, a set of toy experiments has been used to estimate a full covariance matrix. Standard deviations obtained reached between 22 % and 60 % for bin-by-bin unfolded distributions. For iterative Bayesian unfolding, standard deviations were between 20 % and 39 %.

Several validation tests (consistency, closure, pull and stress tests) have successfully been performed to ensure that the framework works as expected. Based on knowledge gained from these tests, the binning and number of iterations for the iterative Bayesian approach have been optimized. Afterwards, data has been unfolded at parton level for p_{T}^{t} and $|\eta^{t}|$. The measured differential cross-sections are consistent with the SM predictions for both variables.

7.2 Future prospects

After this work, there are still many possible ways to improve the differential cross-section measurements of the tZq process:

- Only variables related to the top-quark have been unfolded at parton level so far. Unfolding variables related to the Z boson or unfolding at particle level could provide further insights.
- The estimation of statistical uncertainties can be refined. Some possibilities on how to include uncertainties of efficiency terms have been investigated during this thesis, but proved to be not robust. As a consequence, other ways need to be implemented to include uncertainties of efficiency terms and of the unfolding matrix due to a finite MC sample size. Furthermore, some features observed in the covariance matrices obtained by the iterative Bayesian approach have not be explained so far.
- Systematic uncertainties have not been covered. At the time of submission of this thesis, another master student has already started investigations for including these uncertainties into the framework.
- The framework built during this thesis will potentially serve as basis for future differential cross-section measurements (as planned for a $WbWb$ analysis). As a consequence, the framework will be adjusted to be as customizable yet robust as possible.

Bibliography

[1] CMS Collaboration, *Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s}$ =13 TeV*, Phys. Rev. Lett. **122** (2019) 132003, arXiv: `1812.05900 [hep-ex]` (cit. on pp. 1, 15, 43).

[2] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 9781107034266, URL: `http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013` (cit. on pp. 3, 7).

[3] *Standard Model*, `https://en.wikipedia.org/wiki/Standard_Model`, Accessed: 2019-09-22 (cit. on p. 4).

[4] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) 030001, URL: `https://link.aps.org/doi/10.1103/PhysRevD.98.030001` (cit. on pp. 4, 5, 12).

[5] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. **B716** (2012) 1, arXiv: `1207.7214 [hep-ex]` (cit. on p. 5).

[6] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, Phys. Lett. **B716** (2012) 30, arXiv: `1207.7235 [hep-ex]` (cit. on p. 5).

[7] LHCb Collaboration, *Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays*, Phys. Rev. Lett. **115** (2015) 072001, arXiv: `1507.03414 [hep-ex]` (cit. on p. 5).

[8] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*, Phys. Rev. **76** (6 1949) 769, URL: `https://link.aps.org/doi/10.1103/PhysRev.76.769` (cit. on p. 6).

[9] R. Placakyte, “Parton Distribution Functions”, *Proceedings, 31st International Conference on Physics in collisions (PIC 2011): Vancouver, Canada, August 28-September 1, 2011*, 2011, arXiv: `1111.5452 [hep-ph]` (cit. on p. 10).

[10] *Number of Interactions per Crossing*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Pileup_Interactions_and_Data_Tak`, Accessed: 2019-09-22 (cit. on p. 12).

[11] D0 Collaboration, *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632, arXiv: `hep-ex/9503003 [hep-ex]` (cit. on p. 12).

[12] CDF Collaboration, *Observation of top quark production in $\bar{p}p$ collisions*, Phys. Rev. Lett. **74** (1995) 2626, arXiv: `hep-ex/9503002 [hep-ex]` (cit. on p. 12).

[13] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, Phys. Lett. **B761** (2016) 136, [Erratum: Phys. Lett.B772,879(2017)], arXiv: `1606.02699 [hep-ex]` (cit. on p. 13).

[14] ATLAS Collaboration, *Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **04** (2017) 086, arXiv: `1609.03920 [hep-ex]` (cit. on p. 13).

[15] ATLAS Collaboration, *Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s}$ = 13 TeV with ATLAS*, JHEP **01** (2018) 063, arXiv: `1612.07231 [hep-ex]` (cit. on pp. 13, 14).

[16] ATLAS Collaboration, *Evidence for single top-quark production in the s-channel in proton-proton collisions at $\sqrt{s}$ =8 TeV with the ATLAS detector using the Matrix Element Method*, Phys. Lett. **B756** (2016) 228, arXiv: `1511.05980 [hep-ex]` (cit. on p. 14).

[17] N. Kidonakis, *NNLL resummation for s-channel single top quark production*, Phys. Rev. **D81** (2010) 054028, arXiv: `1001.5034 [hep-ph]` (cit. on p. 14).

[18] ATLAS Collaboration, *Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ production cross sections in multilepton final states using 3.2 fb^{-1} of pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 40, arXiv: `1609.01599 [hep-ex]` (cit. on p. 14).

[19] ATLAS Collaboration, *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. **B784** (2018) 173, arXiv: `1806.00425 [hep-ex]` (cit. on p. 14).

[20] M. Alhroob et al., *Observation of the associated production of a top quark and a Z boson at 13 TeV with ATLAS*, 2018, URL: `https://cds.cern.ch/record/2304824` (cit. on pp. 34, 40, 89).

[21] ATLAS collaboration, *Observation of the associated production of a top quark and a Z boson at 13 TeV with ATLAS*, tech. rep. ATL-COM-PHYS-2019-933, CERN, 2019, URL: `https://cds.cern.ch/record/2683901` (cit. on pp. 15, 29, 35, 38, 41, 43, 77).

[22] *Standard Model Total Production Cross Section Measurements*, `https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/`, Status: July 2018 (cit. on p. 17).

[23] `http://www.lhc-facts.ch/`, Accessed: 2019-09-23 (cit. on pp. 19, 20).

[24] `https://home.cern/science/`, Accessed: 2019-09-23 (cit. on p. 19).

[25] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001, URL: `https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001` (cit. on p. 19).

[26] *Total Integrated Luminosity and Data Quality in 2015-2018*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`, Accessed: 2019-09-23 (cit. on p. 21).

[27] H. Kolanoski and N. Wermes, *Teilchendetektoren Grundlagen und Anwendungen*, ISBN : 9783662453490, Springer Verlag, 2016 (cit. on pp. 22, 23).

[28] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 22).

[29] G. Ripellino, *The alignment of the ATLAS Inner Detector in Run 2*, tech. rep. ATL-INDET-PROC-2016-003, CERN, 2016, URL: `https://cds.cern.ch/record/2213441` (cit. on p. 23).

[30] A. La Rosa, *The ATLAS Insertable B-Layer: from construction to operation*, JINST **11** (2016) C12036, arXiv: `1610.01994 [physics.ins-det]` (cit. on p. 23).

[31] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker*, JINST **9** (2014) P08009, arXiv: `1404.7473 [hep-ex]` (cit. on p. 23).

[32] A. Vogel, *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*, tech. rep. ATL-INDET-PROC-2013-005, CERN, 2013, URL: `https://cds.cern.ch/record/1537991` (cit. on p. 23).

[33] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. **C77** (2017) 317, arXiv: `1611.09661 [hep-ex]` (cit. on p. 25).

[34] *ATLAS photos*, `https://atlasexperiment.org/photos/how-atlas-works.html`, Accessed: 2019-09-23 (cit. on p. 26).

[35] ATLAS Collaboration, *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment*, (2012) (cit. on p. 26).

[36] ATLAS Collaboration, *Calorimeter Clustering Algorithms: Description and Performance*, tech. rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, 2008, URL: `https://cds.cern.ch/record/1099735` (cit. on p. 27).

[37] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at* $\sqrt{s}$ *= 13 TeV*, Eur. Phys. J. **C79** (2019) 639, arXiv: `1902.04655 [physics.ins-det]` (cit. on p. 27).

[38] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s}$ *=13 TeV*, Eur. Phys. J. **C76** (2016) 292, arXiv: `1603.05598 [hep-ex]` (cit. on p. 27).

[39] M. Cacciari, G. P. Salam and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: `0802.1189 [hep-ph]` (cit. on p. 28).

[40] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, tech. rep. ATLAS-CONF-2014-018, CERN, 2014, URL: `https://cds.cern.ch/record/1700870` (cit. on p. 28).

[41] ATLAS Collaboration, *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 580, [Erratum: Eur. Phys. J.C77,no.10,712(2017)], arXiv: `1705.02211 [hep-ex]` (cit. on pp. 28, 45).

[42] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11** (2016) P04008, arXiv: `1512.01094 [hep-ex]` (cit. on p. 28).

[43] ATLAS Collaboration,
Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run,
tech. rep. ATL-PHYS-PUB-2016-012, CERN, 2016,
URL: `https://cds.cern.ch/record/2160731` (cit. on p. 28).

[44] I. A. Cioară, *Associated Production of a Top Quark and a Z Boson in pp Collisions at* $\sqrt{s} = 13\,\text{TeV}$ *Using the ATLAS Detector*, urn:nbn:de:hbz:5n-51941,
PhD thesis: University of Bonn, 2018 (cit. on pp. 30, 31, 36).

[45] O. Bessidskaia Bylund, "Modelling Wt and tWZ production at NLO for ATLAS analyses", *Proceedings, 9th International Workshop on Top Quark Physics (TOP 2016): Olomouc, Czech Republic, September 19-23, 2016*, 2016, arXiv: `1612.00440 [hep-ph]` (cit. on p. 32).

[46] *SgTopRun2NtuplesContents*, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/SgTopRun2NtuplesContents`, Accessed: 2019-09-23
(cit. on pp. 34, 35).

[47] *SgTopRun2NtuplesSamples*, `https://twiki.cern.ch/twiki/bin/view/AtlasProtected/SgTopRun2NtuplesSamples`,
Accessed: 2019-09-23 (cit. on p. 34).

[48] *AnalysisTop*,
`https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/AnalysisTop`,
Accessed: 2019-10-24 (cit. on pp. 34, 61).

[49] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823,
arXiv: `1005.4568 [physics.ins-det]` (cit. on p. 34).

[50] S. Agostinelli et al., *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250
(cit. on p. 34).

[51] 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*, JHEP **07** (2014) 079,
arXiv: `1405.0301 [hep-ph]` (cit. on p. 35).

[52] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040,
arXiv: `1410.8849 [hep-ph]` (cit. on p. 35).

[53] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: `1002.2581 [hep-ph]`
(cit. on p. 35).

[54] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159,
arXiv: `1410.3012 [hep-ph]` (cit. on p. 35).

[55] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007,
arXiv: `0811.4622 [hep-ph]` (cit. on p. 35).

[56] M. Blaut, *Non-prompt Lepton Background Estimation for Associated Production of a Top Quark and a Z Boson in pp Collisions at* $\sqrt{s} = 13\,\text{TeV}$ *at ATLAS*,
MA thesis: University of Bonn, 2018 (cit. on p. 38).

[57] *Artificial Neural Network*,
`https://en.wikipedia.org/wiki/Artificial_neural_network`,
Accessed: 2019-10-13 (cit. on p. 40).

[58] M. Feindt and U. Kerzel, *The NeuroBayes neural network package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **559** (2006) 190, Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research, ISSN: 0168-9002, URL: `http://www.sciencedirect.com/science/article/pii/S0168900205022679` (cit. on p. 39).

[59] S. Mergelmeyer, *Measurement of the Associated Production of a Single Top Quark and a W Boson in Single-Lepton Events with the ATLAS Detector*, urn:nbn:de:hbz:5n-43949, PhD thesis: University of Bonn, 2016 (cit. on p. 42).

[60] G. Cowan, *Statistical Data Analysis*, Clarendon (Oxford), 1998 (ISBN: 0-19-850156-0 or 0-19-850155-2 in paperback) (cit. on pp. 42, 52, 58, 105).

[61] *TRExFitter*, `https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TtHFitter`, Accessed: 2019-09-23 (cit. on p. 43).

[62] *Studies on the forward Jet* $|\eta|$ *problem.* `https://atlasop.cern.ch/elisa/display/386055` (cit. on p. 46).

[63] *Studies on the forward Jet* $|\eta|$ *problem.* `https://atlasop.cern.ch/elisa/display/386069` (cit. on p. 46).

[64] *Studies on forward Jet* $|\eta|$*, as performed by the* tW *analysis group.* `http://webhome.phy.duke.edu/~ddavis/tW/for_tZq/`, Readme at `http://webhome.phy.duke.edu/~ddavis/tW/for_tZq/README.txt` (cit. on p. 46).

[65] C. Young, *Batman Ears III*, `https://indico.cern.ch/event/623054/contributions/2514810/attachments/1427603/2191039/Presentation.pdf` (cit. on p. 46).

[66] C. Young, *Forward Pile-up Jets Over the Years*, `https://indico.cern.ch/event/784293/contributions/3448199/attachments/1856112/3048923/Presentation.pdf` (cit. on p. 47).

[67] J. Howard, *Unfolding Tutorial, Top Workshop*, `https://indico.cern.ch/event/795477/contributions/3378783/attachments/1848586/3033879/JayHowarth_ATLAS_TopWorkshop_UnfoldingTutorial.pdf`, Talk at the Top Working Group Workshop 2019, 21-23 May 2019 (cit. on pp. 51, 56, 59, 66, 68, 77).

[68] G. Cowan, *A survey of unfolding methods for particle physics*, Conf. Proc. **C0203181** (2002) 248, [,248(2002)] (cit. on p. 53).

[69] G. D'Agostini, *A multidimensional unfolding method based on Bayes' theorem*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **362** (1995) 487, ISSN: 0168-9002, URL: `http://www.sciencedirect.com/science/article/pii/016890029500274X` (cit. on pp. 53, 54, 59).

[70] R. Zhang, *Inclusive and differential cross-section measurements of tW single top-quark production at* $\sqrt{s} = 13$ TeV *with the ATLAS detector*, urn:nbn:de:hbz:5n-54880, PhD thesis: University of Bonn, 2019 (cit. on pp. 53, 58, 59).

[71] `RooUnfold`, `http://hepunx.rl.ac.uk/~adye/software/unfold/RooUnfold.html`, `RooUnfold` package and documentation. Accessed: 2019-09-25 (cit. on p. 55).

[72] T. Adye, “Unfolding algorithms and tests using RooUnfold”, *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, CERN, CERN, 2011 313, arXiv: `1105.1160 [physics.data-an]` (cit. on p. 55).

[73] M. Thulin, *The cost of using exact confidence intervals for a binomial proportion*, arXiv e-prints, arXiv:1303.1288 (2013) arXiv:1303.1288, arXiv: `1303.1288 [math.ST]` (cit. on pp. 58, 111).

[74] P. Seema-Mergelmeyer, *Measurements of differential t-channel single top-quark production cross-sections in protpn-proton collisions at a centre-of-mass energy of* 8 TeV *with the ATLAS detector*, urn:nbn:de:hbz:5n-50124, PhD thesis: University of Bonn, 2018 (cit. on pp. 58, 59, 77, 105).

[75] S. L. Eun and F. Ronald N., *Strategies for Variance Estimation*, `https://www.sagepub.com/sites/default/files/upm-binaries/6427_Chapter_4__Lee_(Analyzing)_I_PDF_6.pdf`, Accessed: 2019-11-01 (cit. on p. 109).

[76] ROOT Reference Guide, *TEfficiency Class Reference*, `https://root.cern.ch/doc/master/classTEfficiency.html`, Accessed: 2019-10-22 (cit. on pp. 111, 113).

[77] WolframMathWorld, *Beta Distribution*, `http://mathworld.wolfram.com/BetaDistribution.html`, Accessed: 2019-10-22 (cit. on p. 111).

[78] *Beta Distribution*, `https://valelab4.ucsf.edu/svn/3rdpartypublic/boost/libs/math/doc/sf_and_dist/html/math_toolkit/dist/dist_ref/dists/beta_dist.html`, Accessed: 2019-10-22 (cit. on p. 112).

APPENDIX A

List of Monte Carlo samples

An overview of the MC samples that were used for this analysis is provided. As systematic uncertainties were not covered, only nominal samples are contained in this list. A complete description of the samples can be found in [20].

Table A.1: Overview of nominal signal and background MC samples (full simulation) used for this thesis. Samples used correspond to version 28 ntuples. Table taken from [20].

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
tZq	412063	MadGraphPy8Ev–A14–tllq–NLO	0.0300	1.00	mc16a: 4.99 M mc16d: 6.23 M mc16e: 8.26 M
$t\bar{t}$	410472	PowPy8Ev–A14–ttbar–hdamp258p75–dil	77.0	1.14	mc16a: 79.83 M mc16d: 44.88 M * mc16e: 99.25 M
tW	410648	PowPy8Ev–A14–Wt–t–dil	4.00	0.94	mc16a: 1.00 M mc16d: 1.25 M mc16e: 1.66 M
	410649	PowPy8Ev–A14–Wt–tbar–dil	3.99	0.94	mc16a: 1.00 M mc16d: 1.25 M mc16e: 1.65 M
$t\bar{t}H$	346343	PowPy8Ev–A14–ttH125–allhad	0.0534	1.00	mc16a: 4.98 M mc16d: 6.49 M mc16e: 8.25 M
	346344	PowPy8Ev–A14–ttH125–sl	0.223	1.00	mc16a: 4.99 M mc16d: 6.50 M mc16e: 8.26 M
	346345	PowPy8Ev–A14–ttH125–dilep	0.231	1.00	mc16a: 4.99 M mc16d: 6.49 M

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
					mc16e: 8.28 M
$t\bar{t}H$	410155	aMcAtNloPy8Ev–A14–ttW	0.548	1.10	mc16a: 7.50 M mc16d: 7.50 M mc16e: 12.04 M
	410156	aMcAtNloPy8Ev–A14–ttZnunu	0.155	1.11	mc16a: 1.50 M mc16d: 1.50 M mc16e: 2.00 M
	410157	aMcAtNloPy8Ev–A14–ttZqq	0.528	1.11	mc16a: 3.00 M mc16d: 3.00 M mc16e: 3.59 M
	410218	aMcAtNloPy8Ev–A14–ttee	0.0369	1.12	mc16a: 1.41 M mc16d: 1.34 M mc16e: 2.17 M
	410219	aMcAtNloPy8Ev–A14–ttmumu	0.0369	1.12	mc16a: 1.41 M mc16d: 1.34 M mc16e: 2.17 M
	410220	aMcAtNloPy8Ev–A14–tttautau	0.0365	1.12	mc16a: 0.94 M mc16d: 0.90 M mc16e: 0.96 M
tWZ	410408	aMcAtNloPy8Ev–A14–tWZ–Ztoll–DR1	0.0200	1.00	mc16a: 0.10 M mc16d: 0.12 M mc16e: 0.16 M
Diboson	363356	Sherpa221–ZqqZll	15.6	0.14	mc16a: 5.40 M mc16d: 5.40 M mc16e: 8.95 M
	363358	Sherpa221–WqqZll	3.44	1.00	mc16a: 5.40 M mc16d: 26.91 M mc16e: 8.96 M
	364250	Sherpa222–llll	1.25	1.00	mc16a: 17.84 M mc16d: 36.00 M mc16e: 25.68 M
	364253	Sherpa222–lllv	4.58	1.00	mc16a: 15.54 M mc16d: 32.11 M mc16e: 26.79 M
	364254	Sherpa222–llvv	12.5	1.00	mc16a: 15.00 M mc16d: 29.98 M mc16e: 24.89 M
Z+jets	364114	Sherpa221–Zee–maxHtPtV0_70–L	1 630	0.98	mc16a: 8.00 M mc16d: 10.00 M

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
	364115	Sherpa221–Zee–maxHtPtV0_70–C	224	0.98	mc16e: 13.27 M mc16a: 5.00 M mc16d: 6.24 M
	364116	Sherpa221–Zee–maxHtPtV0_70–B	126	0.98	mc16e: 8.31 M mc16a: 8.00 M mc16d: 9.99 M
	364117	Sherpa221–Zee–maxHtPtV70_140–L	76.3	0.98	mc16e: 13.28 M mc16a: 5.96 M mc16d: 7.38 M
	364118	Sherpa221–Zee–maxHtPtV70_140–C	20.3	0.98	mc16e: 9.95 M mc16a: 2.00 M mc16d: 2.50 M
	364119	Sherpa221–Zee–maxHtPtV70_140–B	12.6	0.98	mc16e: 3.33 M mc16a: 5.97 M mc16d: 7.49 M
	364120	Sherpa221–Zee–maxHtPtV140_280–L	25.0	0.98	mc16e: 9.91 M mc16a: 5.00 M mc16d: 6.25 M
	364121	Sherpa221–Zee–maxHtPtV140_280–C	9.37	0.98	mc16e: 8.36 M mc16a: 3.00 M mc16d: 3.75 M
	364122	Sherpa221–Zee–maxHtPtV140_280–B	6.08	0.98	mc16e: 4.99 M mc16a: 12.44 M mc16d: 15.66 M
	364123	Sherpa221–Zee–maxHtPtV280_500–L	4.87	0.98	mc16e: 20.74 M mc16a: 2.00 M mc16d: 2.50 M
	364124	Sherpa221–Zee–maxHtPtV280_500–C	2.28	0.98	mc16e: 2.78 M mc16a: 1.00 M mc16d: 1.25 M
	364125	Sherpa221–Zee–maxHtPtV280_500–B	1.49	0.98	mc16e: 1.74 M mc16a: 2.00 M mc16d: 2.50 M
	364126	Sherpa221–Zee–maxHtPtV500_1000	1.81	0.98	mc16e: 3.33 M mc16a: 3.00 M mc16d: 3.71 M
	364127	Sherpa221–Zee–maxHtPtV1000_Ecms	0.150	0.98	mc16e: 4.98 M mc16a: 1.00 M mc16d: 1.25 M

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
	364100	Sherpa221–Zmumu–maxHtPtV0_70–L	1 630	0.98	mc16e: 1.67 M mc16a: 7.97 M mc16d: 9.91 M
	364101	Sherpa221–Zmumu–maxHtPtV0_70–C	224	0.98	mc16e: 13.26 M mc16a: 4.98 M mc16d: 6.20 M
	364102	Sherpa221–Zmumu–maxHtPtV0_70–B	127	0.98	mc16e: 8.28 M mc16a: 7.98 M mc16d: 9.26 M
	364103	Sherpa221–Zmumu–maxHtPtV70_140–L	75.0	0.98	mc16e: 13.24 M mc16a: 5.98 M mc16d: 7.48 M
	364104	Sherpa221–Zmumu–maxHtPtV70_140–C	20.4	0.98	mc16e: 9.94 M mc16a: 2.00 M mc16d: 2.49 M
	364105	Sherpa221–Zmumu–maxHtPtV70_140–B	12.4	0.98	mc16e: 3.31 M mc16a: 5.98 M mc16d: 7.47 M
	364106	Sherpa221–Zmumu–maxHtPtV140_280–L	24.3	0.98	mc16e: 9.94 M mc16a: 5.00 M mc16d: 6.24 M
	364107	Sherpa221–Zmumu–maxHtPtV140_280–C	9.28	0.98	mc16e: 8.29 M mc16a: 3.00 M mc16d: 3.75 M
	364108	Sherpa221–Zmumu–maxHtPtV140_280–B	6.01	0.98	mc16e: 4.99 M mc16a: 12.46 M mc16d: 15.63 M
	364109	Sherpa221–Zmumu–maxHtPtV280_500–L	4.77	0.98	mc16e: 20.74 M mc16a: 2.00 M mc16d: 2.46 M
	364110	Sherpa221–Zmumu–maxHtPtV280_500–C	2.27	0.98	mc16e: 3.32 M mc16a: 1.00 M mc16d: 1.25 M
	364111	Sherpa221–Zmumu–maxHtPtV280_500–B	1.49	0.98	mc16e: 1.67 M mc16a: 2.00 M mc16d: 2.50 M
	364112	Sherpa221–Zmumu–maxHtPtV500_1000	1.79	0.98	mc16e: 3.33 M mc16a: 3.00 M mc16d: 3.75 M

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
	364113	Sherpa221–Zmumu–maxHtPtV1000_Ecms	0.150	0.98	mc16e: 5.09 M mc16a: 1.00 M mc16d: 1.25 M
	364128	Sherpa221–Ztautau–maxHtPtV0_70–L	1 630	0.98	mc16e: 1.67 M mc16a: 7.99 M mc16d: 10.00 M
	364129	Sherpa221–Ztautau–maxHtPtV0_70–C	224	0.98	mc16e: 13.27 M mc16a: 4.98 M mc16d: 6.14 M
	364130	Sherpa221–Ztautau–maxHtPtV0_70–B	128	0.98	mc16e: 8.27 M mc16a: 8.00 M mc16d: 9.99 M
	364131	Sherpa221–Ztautau–maxHtPtV70_140–L	76.0	0.98	mc16e: 13.28 M mc16a: 6.00 M mc16d: 7.50 M
	364132	Sherpa221–Ztautau–maxHtPtV70_140–C	20.2	0.98	mc16e: 9.97 M mc16a: 2.00 M mc16d: 2.50 M
	364133	Sherpa221–Ztautau–maxHtPtV70_140–B	12.3	0.98	mc16e: 3.33 M mc16a: 5.97 M mc16d: 7.50 M
	364134	Sherpa221–Ztautau–maxHtPtV140_280–L	24.8	0.98	mc16e: 9.96 M mc16a: 4.94 M mc16d: 6.23 M
	364135	Sherpa221–Ztautau–maxHtPtV140_280–C	9.33	0.98	mc16e: 8.30 M mc16a: 3.00 M mc16d: 3.75 M
	364136	Sherpa221–Ztautau–maxHtPtV140_280–B	5.48	0.98	mc16e: 4.99 M mc16a: 4.99 M mc16d: 6.21 M
	364137	Sherpa221–Ztautau–maxHtPtV280_500–L	4.79	0.98	mc16e: 8.29 M mc16a: 2.00 M mc16d: 2.50 M
	364138	Sherpa221–Ztautau–maxHtPtV280_500–C	2.28	0.98	mc16e: 3.32 M mc16a: 1.00 M mc16d: 1.23 M
	364139	Sherpa221–Ztautau–maxHtPtV280_500–B	1.50	0.98	mc16e: 1.67 M mc16a: 1.98 M mc16d: 2.50 M

Process	Sample ID	Generator	σ [pb]	k-factor	N-generated
					mc16e: 3.32 M
	364140	Sherpa221–Ztautau–maxHtPtV500_1000	1.81	0.98	mc16a: 3.00 M mc16d: 3.75 M mc16e: 4.96 M
	364141	Sherpa221–Ztautau–maxHtPtV1000_Ecms	0.150	0.98	mc16a: 1.00 M mc16d: 1.25 M mc16e: 1.67 M

APPENDIX B

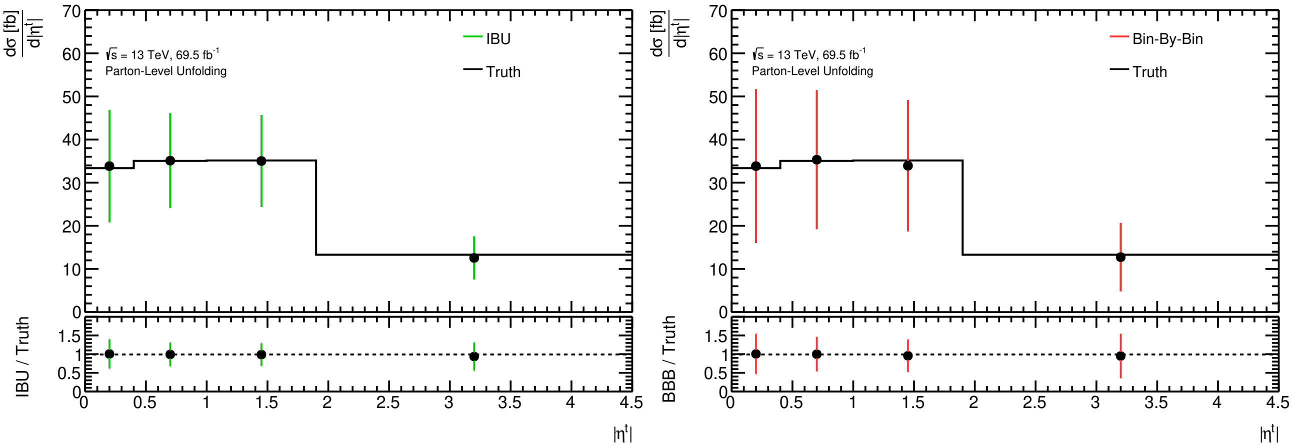
Further validation tests

An overview of results from validation tests not shown in chapter 6 is provided.

B.1 Top-quark $|\eta|$, $O_{\text{NN}} > 0.6$

Validation tests are shown for $|\eta^t|$ with a cut of $O_{\text{NN}} > 0.6$. One iteration is used when applying iterative Bayesian unfolding.

Closure test

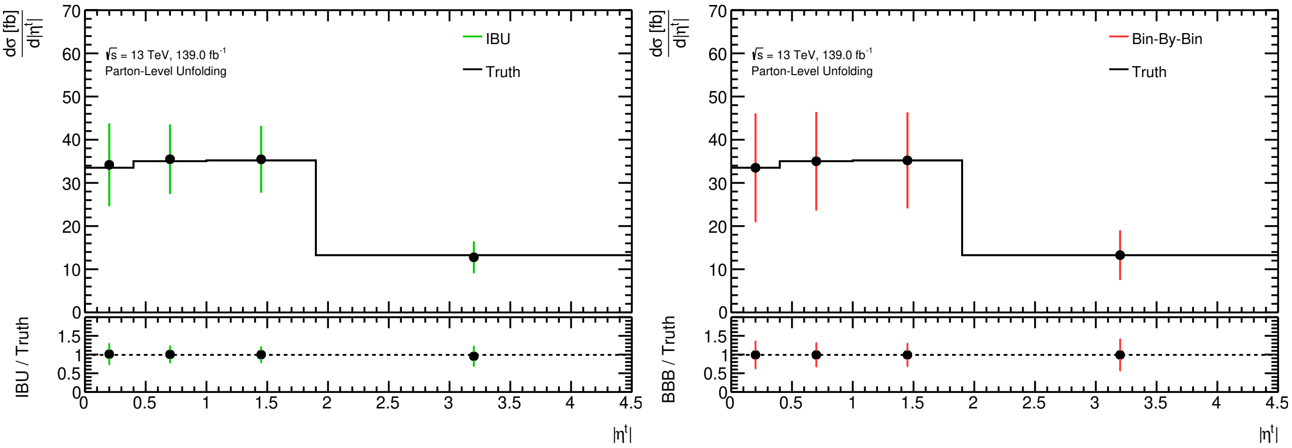


(a)

(b)

Figure B.1: Results from a closure test for $|\eta^t|$, for figure B.1(a) after application of iterative Bayesian unfolding, for figure B.1(b) after application of bin-by-bin unfolding.

Consistency test



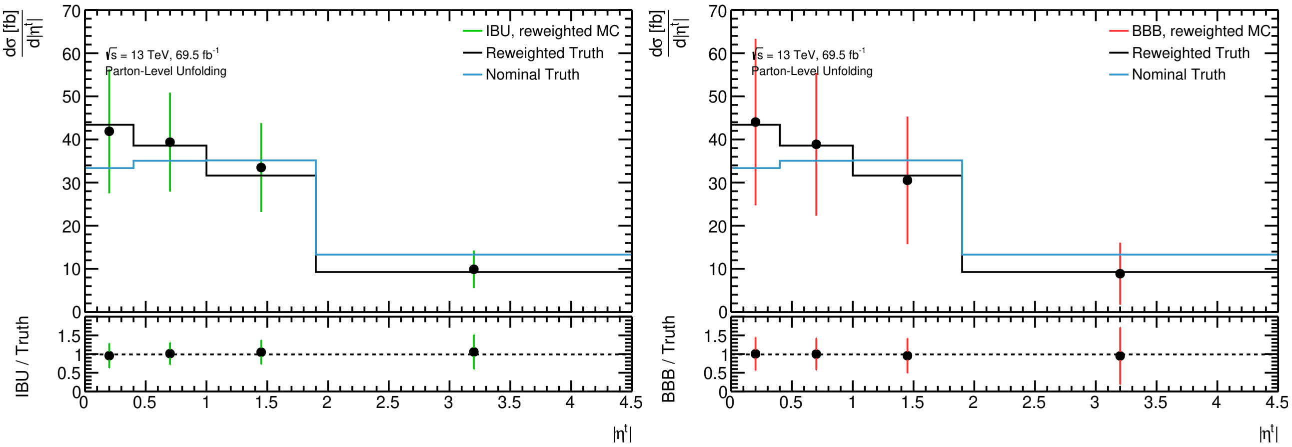
(a)

(b)

Figure B.2: Results from a consistency test for $|\eta^t|$, for figure B.2(a) after application of iterative Bayesian unfolding, for figure B.2(b) after application of bin-by-bin unfolding.

Stress test

For the stress test, MC distributions have been reweighted bin-by-bin with [30 %, 10 %, −10 %, −30 %].



(a)

(b)

Figure B.3: Results from a stress test for $|\eta^t|$, for figure B.3(a) after application of iterative Bayesian unfolding, for figure B.3(b) after application of bin-by-bin unfolding.

Pull test

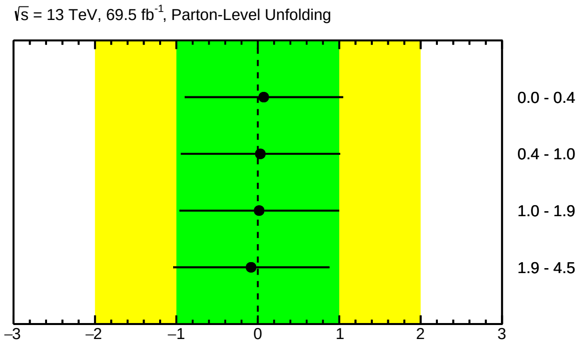


Figure B.4: Overview of pulls as obtained from 10^3 sets of toy experiments. Pulls obtained for $|\eta^t|$, $O_{\mathrm{NN}} > 0.6$.

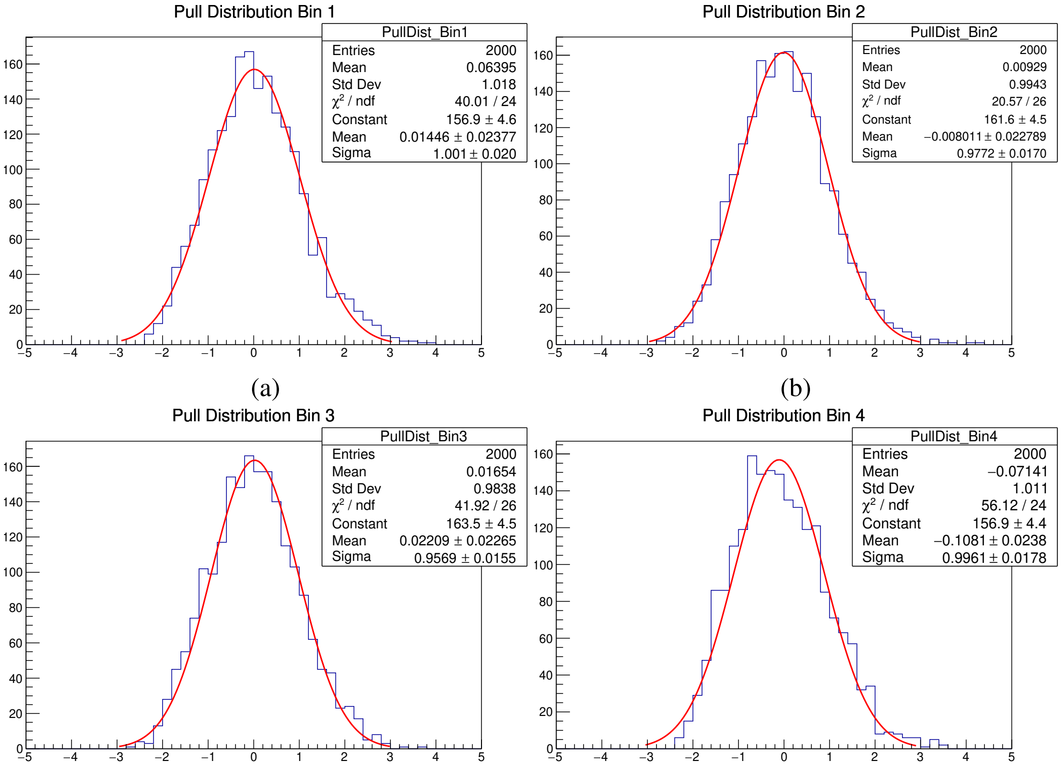
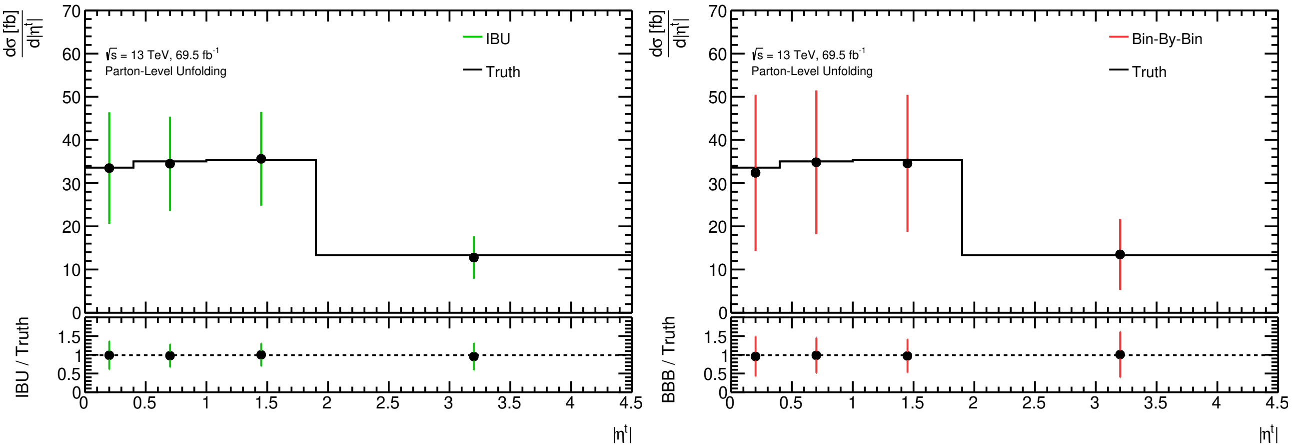


Figure B.5: Examples of pull distributions from one set of toy experiments, calculated for $|\eta^t|$.

B.2 Top-quark $|\eta|$, $O_{\mathrm{NN}} > 0$

Validation tests are shown for $|\eta^t|$ with a cut of $O_{\mathrm{NN}} > 0$. One iteration is used when applying iterative Bayesian unfolding.

Closure test

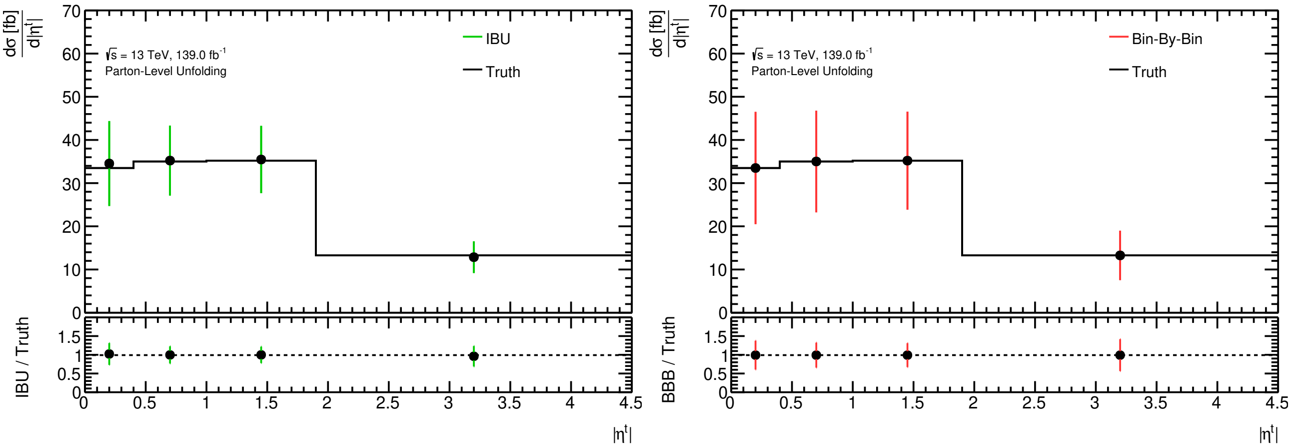


(a)

(b)

Figure B.6: Results from a closure test for $|\eta^t|$, for figure B.6(a) after application of iterative Bayesian unfolding, for figure B.6(b) after application of bin-by-bin unfolding.

Consistency test



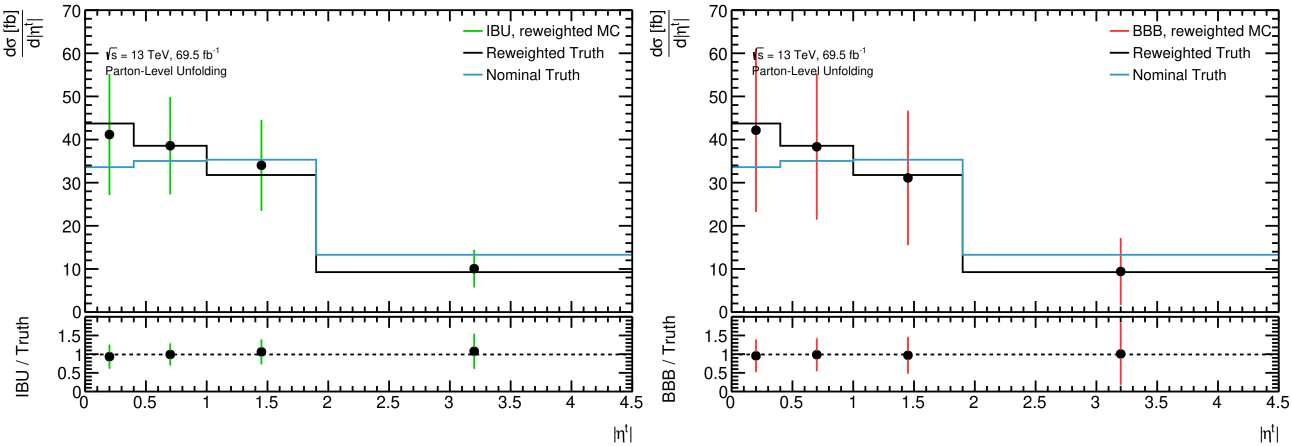
(a)

(b)

Figure B.7: Results from a consistency test for $|\eta^t|$, for figure B.7(a) after application of iterative Bayesian unfolding, for figure B.7(b) after application of bin-by-bin unfolding.

Stress test

For the stress test, MC distributions have been reweighted bin-by-bin with [30 %, 10 %, −10 %, −30 %].



(a) (b)

Figure B.8: Results from a stress test for $|\eta^t|$, for figure B.3(a) after application of iterative Bayesian unfolding, for figure B.3(b) after application of bin-by-bin unfolding.

Pull test

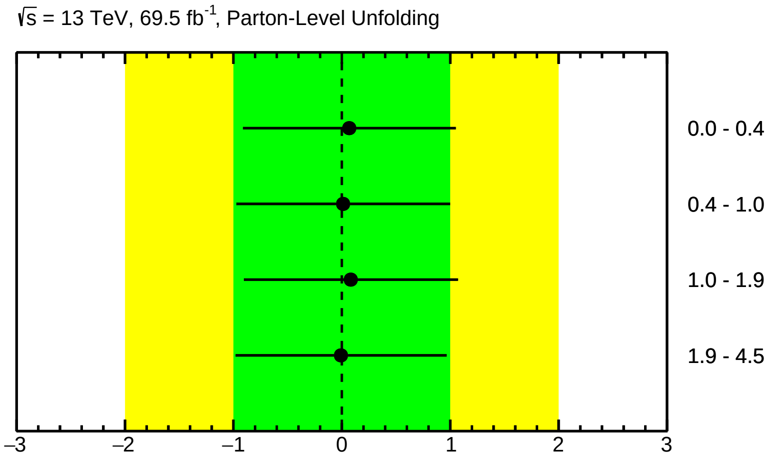
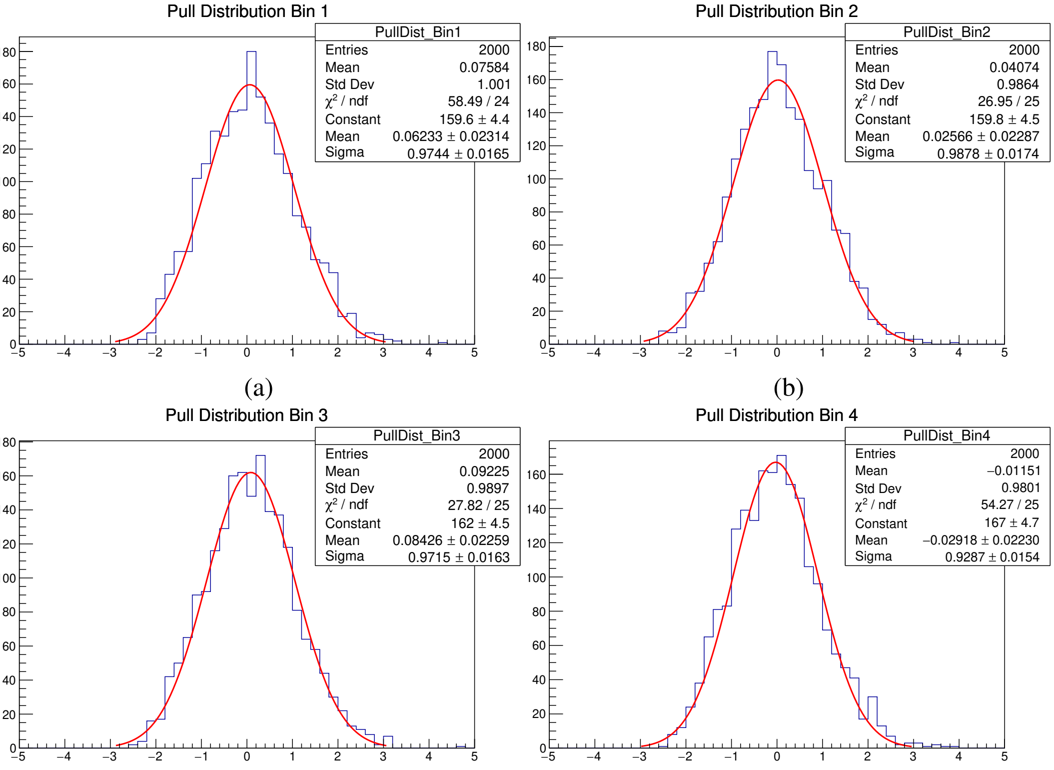


Figure B.9: Overview of pulls as obtained from 10^3 sets of toy experiments. Pulls obtained for $|\eta^t|$, $O_{\mathrm{NN}} > 0$.



(a) (b)

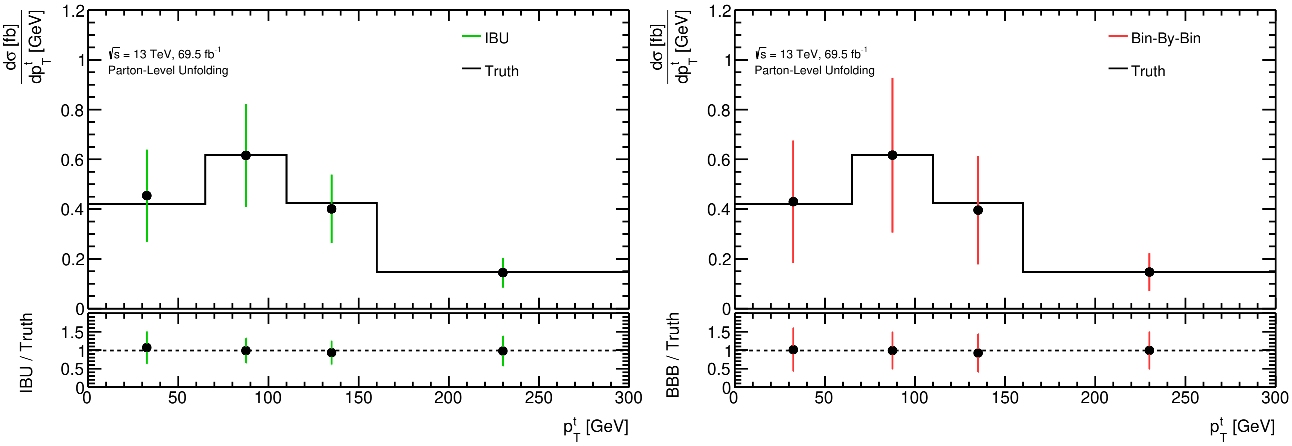
(c) (d)

Figure B.10: Examples of pull distributions from one set of toy experiments, calculated for $|\eta^t|$.

B.3 Top-quark p_{T}, $O_{\mathrm{NN}} > 0$

Validation tests are shown for p_{T}^{t} with a cut of $O_{\mathrm{NN}} > 0$. Three iterations are used when applying iterative Bayesian unfolding.

Closure test

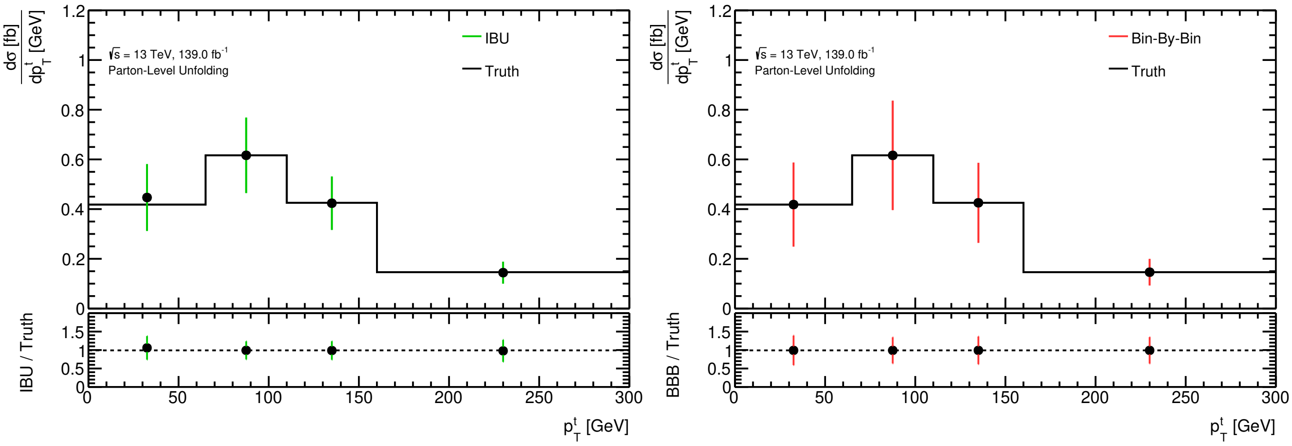


(a)

(b)

Figure B.11: Results from a closure test for p_{T}^{t}, for figure B.11(a) after application of iterative Bayesian unfolding, for figure B.11(b) after application of bin-by-bin unfolding.

Consistency test



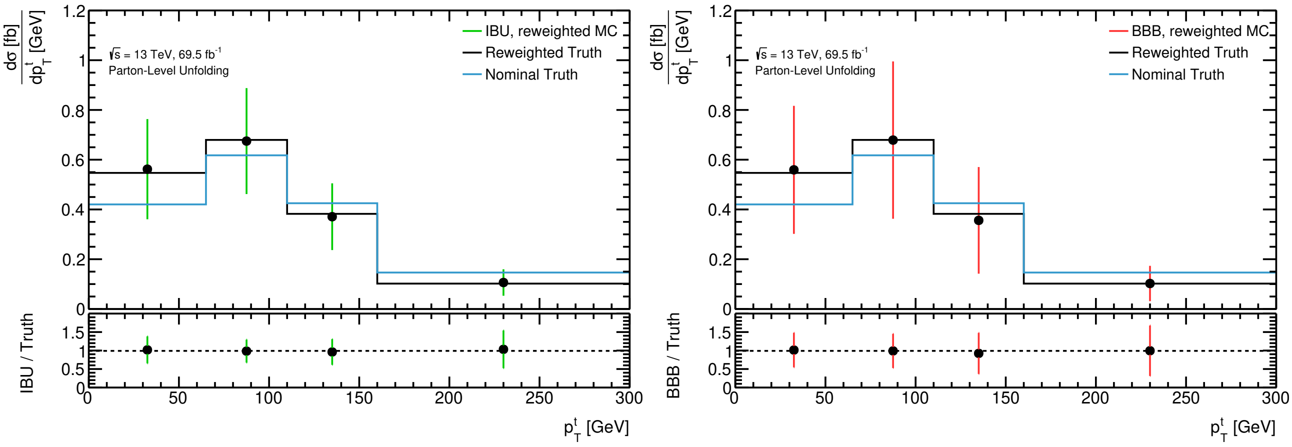
(a)

(b)

Figure B.12: Results from a consistency test for p_{T}^{t}, for figure B.12(a) after application of iterative Bayesian unfolding, for figure B.12(b) after application of bin-by-bin unfolding.

Stress test

For the stress test, MC distributions have been reweighted bin-by-bin with [30 %, 10 %, −10 %, −30 %].



(a)

(b)

Figure B.13: Results from a stress test for p_T^t, for figure B.13(a) after application of iterative Bayesian unfolding, for figure B.13(b) after application of bin-by-bin unfolding.

Pull test

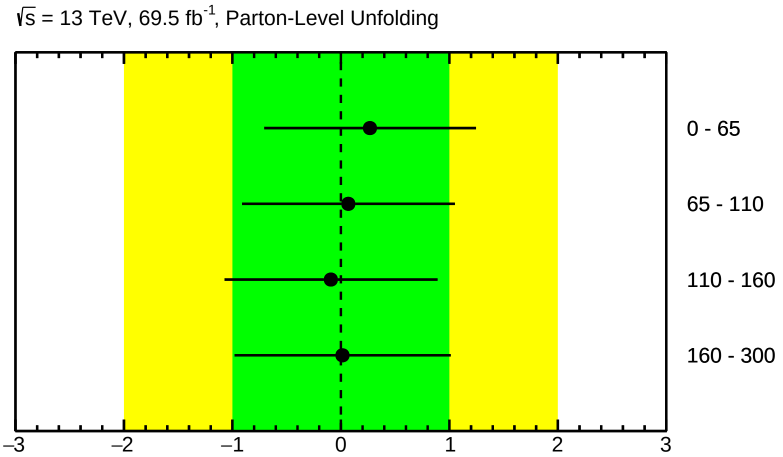
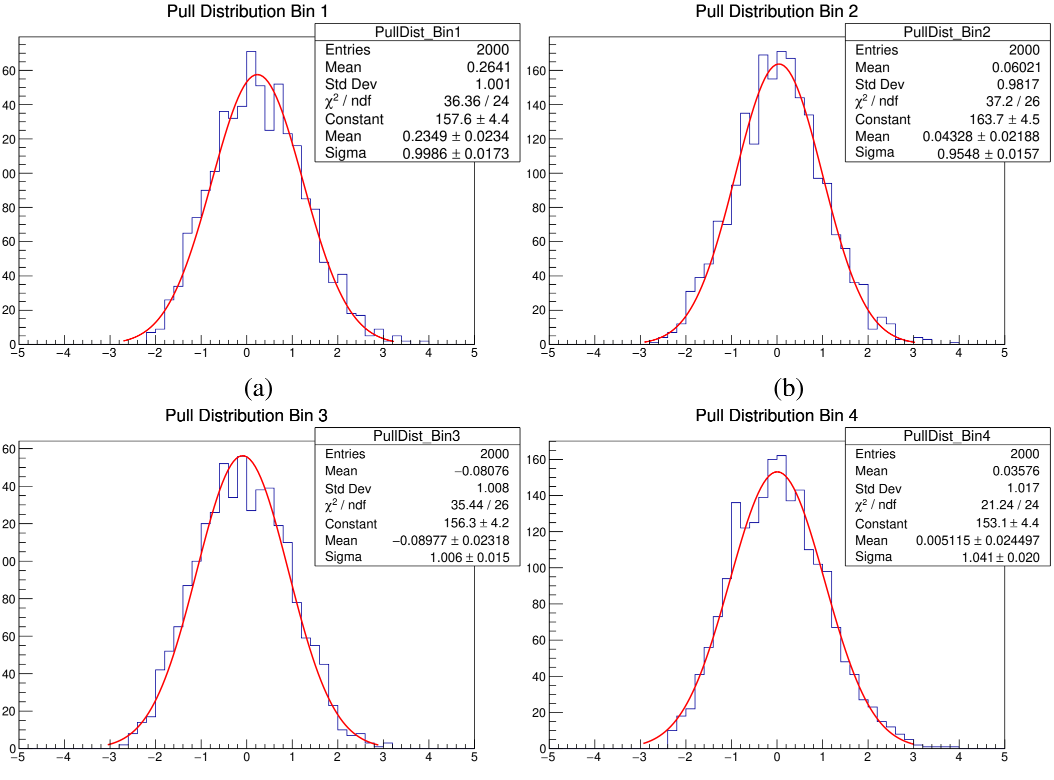


Figure B.14: Overview of pulls as obtained from 10^3 sets of toy experiments. Pulls obtained for p_T^t, $O_{NN} > 0$. Bin boundaries given in GeV.



(a)

(b)

(c)

(d)

Figure B.15: Examples of pull distributions from one set of toy experiments, calculated for p_{T}^{t}.

APPENDIX C

Discussions about correlations

Unexpected behavior observed for the covariance matrices (see section 6.4.2) is discussed in further detail. The influence of the currently obtained covariance matrices on uncertainties assigned to normalized differential cross-sections is investigated.

C.1 Further studies on covariance matrices

As described in section 6.4.2, the covariance matrices that have been obtained do not behave according to expectations, as they show large correlations between different bins instead of anti-correlations (which e.g. [74] reported for another differential cross-section measurement).
For the remainder of this section, correlation matrices instead of covariance matrices are considered. A correlation matrix ρ_{ij} is related to the covariance matrix C_{ij} via: [60]

$$\rho_{ij} = \frac{C_{ij}}{\sigma_i \sigma_j}\,, \tag{C.1}$$

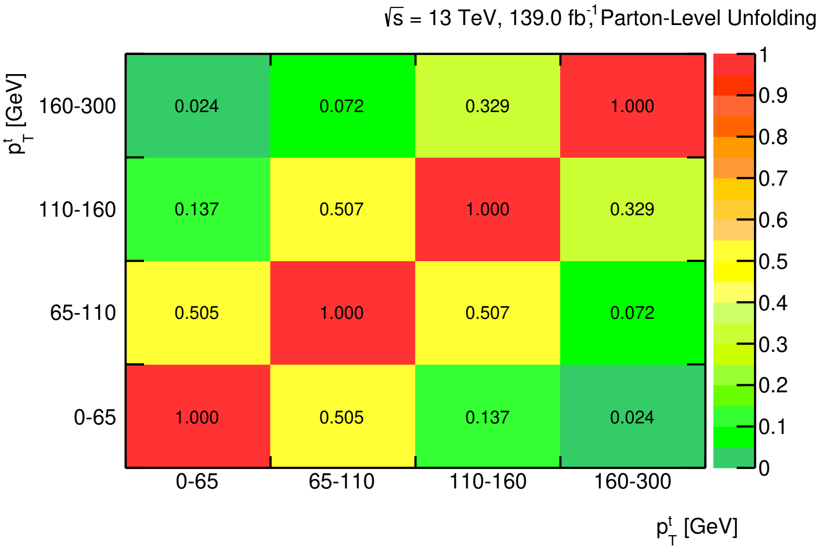
where $\sigma_{i(j)}$ denote standard deviations assigned to the corresponding bins. An example for the correlation matrix as obtained for p_{T}^{t} with $O_{\mathrm{NN}} > 0.6$ is shown in figure C.1.

Normalizing distributions

An attempt to solve the problem of only positive diagonal elements was a *normalization* procedure of the distribution before entering the computation of the covariance matrix. This attempt was tried as to directly be able to compare the shapes of the distributions: for iterative Bayesian unfolding, it is expected that an increase of the event yield in one bin leads to a decrease in another bin. Mathematically, the covariance matrix for this approach can be calculated as:

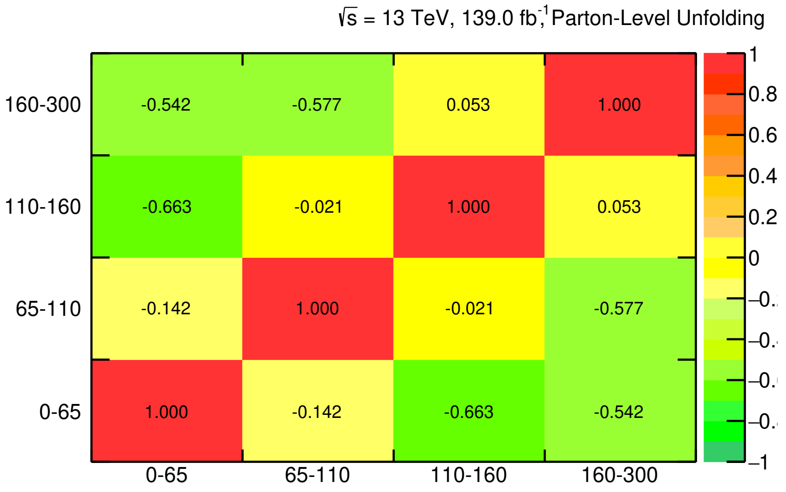
$$C_{ij} = \frac{1}{N_{\mathrm{Toys}} - 1} \sum_{k=1}^{N_{\mathrm{Toys}}} \left(\frac{N_{\mathrm{unf}}^{i,k}}{\sum_{m=1}^{N_{\mathrm{bins}}} N_{\mathrm{unf}}^{m,k}} - \overline{\frac{\overline{N_{\mathrm{unf}}^{i}}}{\sum_{m=1}^{N_{\mathrm{bins}}} N_{\mathrm{unf}}^{m}}} \right) \left(\frac{N_{\mathrm{unf}}^{j,k}}{\sum_{m=1}^{N_{\mathrm{bins}}} N_{\mathrm{unf}}^{m,k}} - \overline{\frac{\overline{N_{\mathrm{unf}}^{j}}}{\sum_{m=1}^{N_{\mathrm{bins}}} N_{\mathrm{unf}}^{m}}} \right)\,. \tag{C.2}$$

An example for the resulting covariance matrix is shown in figure C.2. Although this matrix did allow for negative off-diagonal elements, the behavior was still not as expected: direct off-diagonal



p_T^t [GeV]	0-65	65-110	110-160	160-300
160-300	0.024	0.072	0.329	1.000
110-160	0.137	0.507	1.000	0.329
65-110	0.505	1.000	0.507	0.072
0-65	1.000	0.505	0.137	0.024

Figure C.1: Correlation matrix obtained for p_T^t, $O_{NN} > 0.6$. In contrast to the expectations, there are no negative off-diagonal elements.



	0-65	65-110	110-160	160-300
160-300	-0.542	-0.577	0.053	1.000
110-160	-0.663	-0.021	1.000	0.053
65-110	-0.142	1.000	-0.021	-0.577
0-65	1.000	-0.142	-0.663	-0.542

Figure C.2: Correlation matrix obtained for p_T^t, $O_{NN} > 0.6$, when applying a normalization procedure as described in equation (C.2). Bin boundaries given in GeV.

elements are closer to 0 than matrix elements relating bins far away from each other. This means that non-adjacent bins are heavily anti-correlated (according to figure C.2 by approximately 60 %) while correlations between adjacent bins are less important. Similar tests have been performed for a set of toy experiments by application of bin-by-bin unfolding. The correlation matrix obtained for bin-by-bin unfolding (see next section) is expected to show diagonal elements only, whereas off-diagonal elements should be negligible. However, the matrix observed showed similar behavior to the one showed in figure C.2, meaning that the process of normalization artificially introduces new correlations. As a consequence, this approach was rejected.

Toy experiments for bin-by-bin unfolding

To ensure that the generation of random numbers for toy experiments and the implementation of the covariance matrix according to equation (5.15) is correct, a set of toy experiments have been unfolded using bin-by-bin unfolding. Figure C.3 shows the correlation matrix obtained for p_T^t, $O_{NN} > 0.6$. It

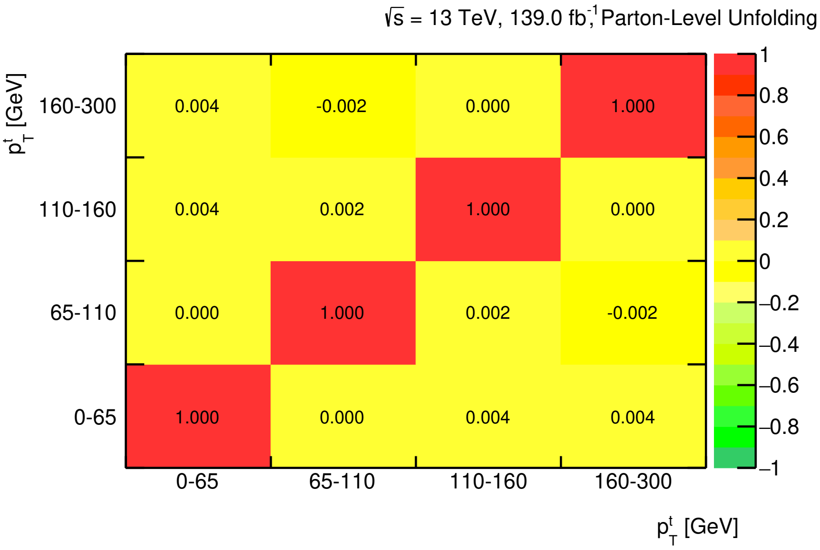


Figure C.3: Correlation matrix for p_T^t, $O_\mathrm{NN} > 0.6$, obtained from a set of toy experiments for the determination of the covariance matrix using bin-by-bin unfolding. Off-diagonal elements are negligible (< 0.5 ‰).

can be seen that all off-diagonal elements are smaller than 0.5 ‰, meeting the expectation of negligible correlations for bin-by-bin unfolding. However, this also confirms that the implementation of the covariance matrix is correct. As a consequence, the focus shifted towards investigating the outcome of the iterative Bayesian procedure itself.

Investigation of distributions $1/\left(\mathcal{L}\Delta X^i\right)\cdot\left(N_i - \overline{N}_i\right)$

As bin-by-bin unfolding returned results as expected, the distributions $\frac{1}{\mathcal{L}\Delta X^i}\left(N_i - \overline{N}_i\right)$ and $\frac{1}{\mathcal{L}^2\Delta X^i\Delta X^j}\left(N_i - \overline{N}_i\right)\left(N_j - \overline{N}_j\right)$ that enter equation (5.15) have been plotted.
The distributions $\frac{1}{\mathcal{L}\Delta X^i}\left(N_i - \overline{N}_i\right)$ for all 4 bins that have been used for p_T^t, $O_\mathrm{NN} > 0.6$ are shown in figure C.4. For all the bins, the distribution is centered around 0 as expected. Slight tails towards positive values are potentially recognizable, which could be a remnant from the generated Poisson distributed random number the toy experiment started with.
The distributions $\frac{1}{\mathcal{L}^2\Delta X^i\Delta X^j}\left(N_i - \overline{N}_i\right)\left(N_j - \overline{N}_j\right)$ are especially interesting to investigate for those combinations of i and j which showed large positive correlations. A comparison with figure C.1 shows that this is e.g. the case for $i = 1, j = 2$: $\rho_{12} = 0.505$. The corresponding distribution as generated from toy experiments is shown in figure C.5(a). The distribution from figure C.5(a) is asymmetric, with values shifted towards positive values, thus explaining a positive correlation. A similar behavior has been observed for the distribution with $i = 2, j = 3$, which also had large correlations: $\rho_{13} = 0.507$. On the other hand, the distribution obtained for $i = 1, j = 4$ (the corresponding matrix element is $\rho_{14} = 0.024$) is nearly symmetric around 0.
The asymmetry of the distribution $\frac{1}{\mathcal{L}^2\Delta X^i\Delta X^j}\left(N_i - \overline{N}_i\right)\left(N_j - \overline{N}_j\right)$ is able to explain the large positive correlations factors obtained for some ρ_{ij}'s. However, the reason for the asymmetric behavior of the distribution itself is still unclear, although all the checks performed suggest that the reason for the discrepancy lies within the iterative Bayesian unfolding approach itself, and not in the computation of the correlation matrix. Further investigations are required to find the reason for the discrepancies observed.

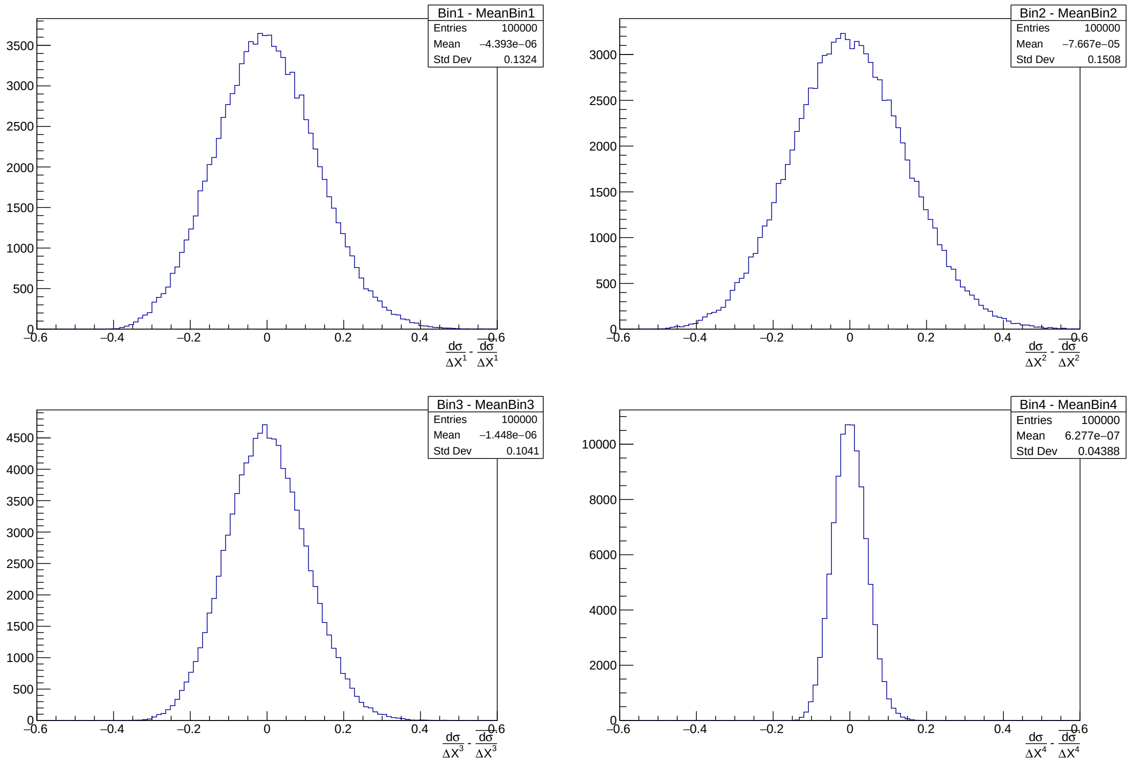
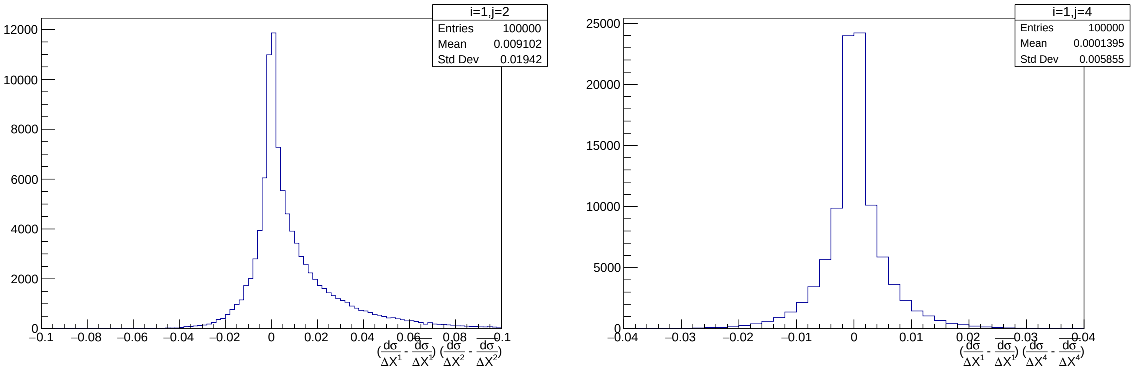


Figure C.4: Distributions $\frac{1}{\mathcal{L}\Delta X^i}\left(N_i - \overline{N}_i\right)$ obtained from toy experiments for each bin of p_{T}^t, $O_{\mathrm{NN}} > 0.6$.



(a) (b)

Figure C.5: Two examples for distributions $\frac{1}{\mathcal{L}^2\Delta X^i\Delta X^j}\left(N_i - \overline{N}_i\right)\left(N_j - \overline{N}_j\right)$ obtained from a set of toy experiments, for p_{T}^t, $O_{\mathrm{NN}} > 0.6$. Figure C.5(a) shows the distribution for $i = 1$, $j = 2$, figure C.5(b) shows the distribution for $i = 1$, $j = 4$.

C.2 Uncertainties of normalized differential cross-sections

A problem not covered in section 5.3.4 on normalized differential cross-section measurements are correlations between σ_{fid} and $\frac{\mathrm{d}\sigma}{\mathrm{d}X^i}$ entering equation (5.19). As a consequence, error propagation should not be treated as for the uncorrelated case, but correlations need to be taken into account. σ^{fid} takes the same functional form as $g = \sum_i a_i x_i$, resulting in the same expression for uncertainties:

$$g = \sum_i a_i x_i \qquad \Rightarrow \qquad \sigma_g^2 = \sum_i \sum_j a_i a_j C_{ij} \text{ with } C_{ii} = \sigma_i^2 \,. \tag{C.3}$$

C_{ij} in the case of σ^{fid} are the respective covariance matrix entries that have already been determined for the uncertainty estimation of the differential cross-section in the first place. For comparison purposes, table C.1 shows the results obtained for the fiducial cross-sections and the uncertainties, once while considering correlations, once without. As all the covariance matrices shown in figure 6.16 have positive matrix elements only, the uncertainty assigned to the fiducial cross-section increases after implementation of these correlations.

	σ^{fid}[fb]	Uncertainty without correlation [fb]	Uncertainty with correlation [fb]
$\sigma^{\mathrm{fid}}_{\mathrm{IBU},p_{\mathrm{T}}}$ $(O_{\mathrm{NN}} > 0.6)$	101.7	13.6	17.9
$\sigma^{\mathrm{fid}}_{\mathrm{IBU},p_{\mathrm{T}}}$ $(O_{\mathrm{NN}} > 0)$	89.3	13.8	18.4
$\sigma^{\mathrm{fid}}_{\mathrm{IBU},\lvert\eta\rvert}$ $(O_{\mathrm{NN}} > 0.6)$	94.8	13.3	17.9
$\sigma^{\mathrm{fid}}_{\mathrm{IBU},\lvert\eta\rvert}$ $(O_{\mathrm{NN}} > 0)$	86.0	13.3	18.2

Table C.1: Comparison of uncertainties assigned to the fiducial cross-section obtained for different variables, while investigating the impact of correlations. Estimated uncertainties while treating all correlations have been computed with the help of covariance matrices shown in figure 6.16.

The normalization process takes the functional form and uncertainty: [75]

$$g = \frac{a}{b} \qquad \Rightarrow \qquad \sigma_g \approx |g| \sqrt{\left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{\sigma_b}{b}\right)^2 - 2 \cdot \frac{C_{ab}}{ab}}\,, \tag{C.4}$$

where the uncertainty σ_g is not an exact formula, but provides higher accuracy than treating the normalization uncorrelated. For the case of normalized differential cross-sections, a can be assigned to $\frac{\mathrm{d}\sigma}{\mathrm{d}X^i}$ and b to σ^{fid}; therefore σ_a and σ_b are known to be the standard deviations associated to σ^{fid} respectively $\frac{\mathrm{d}\sigma}{\mathrm{d}X^i}$. However, the quantity C_{ab} still needs to be determined. For this purpose, another set of 10^5 toy experiments has been generated. For each bin, a 2×2 dimensional correlation matrix[1] correlating $\frac{\mathrm{d}\sigma}{\mathrm{d}X^i}$ and σ^{fid} will be generated. The off-diagonal element in this matrix corresponds to C_{ab} in equation (C.4).

---

[1] Instead of creating N_{bins} correlation matrices of dimensions 2×2, it should in principle also be possible to create a single $\left(N_{\mathrm{bins}} + 1\right) \times \left(N_{\mathrm{bins}} + 1\right)$ dimensional correlation matrix which contains the same information about correlations between σ^{fid} and all the $\frac{\mathrm{d}\sigma}{\mathrm{d}X^i}$s. The approach using N_{bins} correlation matrices of dimensions 2×2 was kept after its implementation was shown to produce accurate results. However, the other approach should computationally be more efficient and might be worth implementing instead.

The results obtained for the normalized differential cross-sections for p_{T}^{t} with $O_{\mathrm{NN}} > 0.6$ are summarized in table C.2. Two uncertainties are compared: the first uncertainty is calculated while assuming that the uncertainty of σ^{fid} and the resulting normalized differential cross-section can be estimated from Gaussian error propagation. For the second approach, all correlations are considered, as described in equations (C.3) and (C.4). A decrease of the standard deviation assigned to each bin can be observed in table C.2.

p_{T}^{t}[GeV]	$\frac{1}{\sigma^{\mathrm{fid}}}\frac{\mathrm{d}\sigma}{\mathrm{d}X^{i}}[10^{-3}/\mathrm{GeV}]$	Uncertainty without correlation $[10^{-3}/\mathrm{GeV}]$	Uncertainty with correlation $[10^{-3}/\mathrm{GeV}]$
0 – 65	4.30	1.42	0.92
65 – 110	7.46	1.79	0.92
110 – 160	4.14	1.16	0.77
160 – 300	1.27	0.46	0.38

Table C.2: Comparison of uncertainties assigned to the normalized differential cross-section obtained with iterative Bayesian unfolding for p_{T}^{t}, $O_{\mathrm{NN}} > 0.6$, while investigating the impact of correlations.

APPENDIX D

Further information on statistical uncertainty estimation

Features that were planned to be included in the main unfolding framework are described. Unresolved problems however meant that these features are not completely trustworthy. As a consequence they are only included as optional computation.

D.1 Clopper-Pearson intervals, theoretical considerations

Uncertainties assigned to efficiencies and acceptances[1] due to the finite MC sample size shown in figures 6.4 and 6.3 are determined using *Clopper-Pearson intervals* (see e.g. section 2 in [73]). The computation of confidence intervals is performed by using the class `TEfficiency` [76] as implemented in `ROOT`.
The range of efficiencies is limited to [0, 1]. This poses a problem for values close to 0 or 1, as in general the efficiency is no longer Gauss distributed. Instead, a Beta-distribution, a probability distribution closely linked to the Binomial distribution, will be assumed. [77] The underlying probability density function is given as:

$$f(x;\alpha,\beta) = \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,, \qquad 0 \leq x \leq 1\,. \tag{D.1}$$

The factor $\frac{1}{\mathrm{B}(\alpha,\beta)}$ is used as normalization constant. α and β are called *shape factors*.
Assume n_t events were generated in total and n_p events passed the selection cuts for the calculation of an efficiency. Using the Beta-distribution, the lower limit of a confidence interval is: [76]

$$\begin{aligned} \mathcal{P}\left(N \geq n_p; n_t\right) &= 1 - \mathcal{P}\left(N \leq n_p - 1; n_t\right) \\ &= 1 - \frac{1}{\mathrm{norm}} \int_0^{1-\epsilon} t^{n_t - n_p}(1-t)^{n_p - 1}\mathrm{d}t \\ &= I_\epsilon\left(n_p, n_t - n_p + 1\right)\,. \end{aligned} \tag{D.2}$$

$I_\epsilon(\alpha,\beta)$ is called regularized incomplete beta function and describes quantiles of the Beta-distribution.

---

[1] In this chapter, the term "efficiency" is used for both efficiencies and acceptances unless stated otherwise.

On the other hand, for a given confidence level CL:

$$\mathcal{P}\left(N \geq n_p; n_t\right) = \frac{1 - \mathrm{CL}}{2} \; . \tag{D.3}$$

Equaling equation (D.2) and equation (D.3) should therefore for the lower boundary of the confidence interval at given CL result in:

$$I_\epsilon\left(n_p, n_t - n_p + 1\right) = \frac{1 - \mathrm{CL}}{2} \; . \tag{D.4}$$

A similar expression can be found for the upper boundary of the confidence interval:

$$I_\epsilon\left(n_p + 1, n_t - n_p\right) = \frac{1 + \mathrm{CL}}{2} \; . \tag{D.5}$$

D.2 Clopper-Pearson intervals, implementation for toy experiments

Given 68 % confidence intervals for efficiencies in each bin and the known underlying probability density function being a Beta-distribution, a set of random numbers could (in theory) be produced to include uncertainties on efficiencies in toy experiments for the total uncertainty estimation. However, the shape factors need to be determined first. Assuming that the actual value of the efficiency in bin j corresponds to the mean value μ and the variance σ^2 can also be approximated, the shape factors can be estimated as: [78]

$$\begin{aligned} \alpha &= \mu \left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right) \; , \\ \beta &= (1-\mu) \left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right) \; . \end{aligned} \tag{D.6}$$

For actual computations, the variance is approximated as:

$$\sigma^2 = \left(\frac{\mathrm{UL} - \mathrm{LL}}{2}\right)^2 \; , \tag{D.7}$$

where UL (LL) denotes the upper limit (lower limit) of the confidence interval.

Several tests have been performed to understand the behavior of the Beta-distribution when used as underlying probability density function. An example with the parameters $\mu = 0.8$, $\sigma = 0.1$ is shown in figure D.1. As anticipated, ≈ 68 % of events lie within the confidence interval $[\mu - \sigma, \mu + \sigma]$.

However, some combinations for mean values and variances have proven to be troublesome for the implementation in `ROOT`. The distributions shown in figure D.2 were generated from $\mu = 0.9, \sigma = 0.1$. In this case, one of the shape factors was smaller than 1. Although σ should be identifiable with a 68 % confidence interval, ≈ 85 % of the random numbers generated according to the underlying Beta-distribution lied within this confidence interval, thus more than expected. This is of special relevance for the random number generation of the acceptances (as can be seen from figure 6.3), where values for μ are close or equal to 1. According to equation (D.6), the case of $\mu = 1$ leads to $\beta = 0$ and cannot be handled properly.

Furthermore, small values for σ were not handled properly by `ROOT`, potentially even resulting in

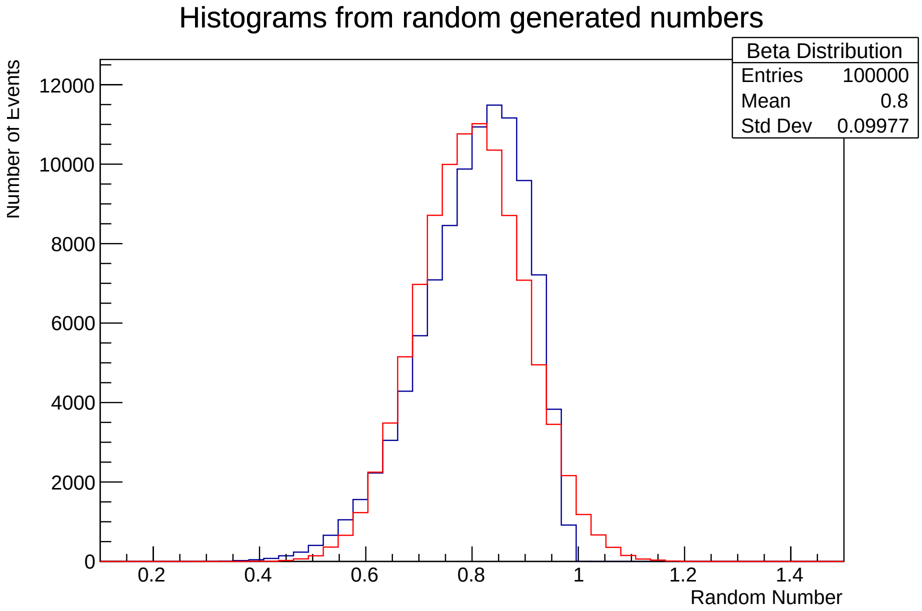


Figure D.1: Example of a Beta-distribution (in blue) with the parameters $\mu = 0.8, \sigma = 0.1$, with a distribution as expected. For comparison purposes a Gaussian distribution with same parameters is displayed in red.

error messages. An example with the parameters $\mu = 0.005$ and $\sigma = 0.001$ is depicted in figure D.3. For this case, only 20 % lie within a 68 % confidence interval, thus less than expected. This is especially problematic for the correct implementation of random number generation for efficiencies, as uncertainties assigned to efficiencies are typically small (displayed in figure 6.4).
Due to these problems, the random number generation for efficiencies and acceptances based on an underlying Beta-distribution was included as option only and is not recommended.

D.3 Random number generation with Gauss approximation

The impact of including the random number generation for efficiencies in toy experiments from a Gaussian distribution was studied as well. However, defining an efficiency as:

$$\epsilon = \frac{n_\mathrm{p}}{n_\mathrm{t}}\,, \tag{D.8}$$

where n_p defines the number of events that pass selection cuts and n_t the total number of events, it is known that this approximation is not valid in the case of $\epsilon \approx 0$ or $\epsilon \approx 1$ (which can already be seen e.g. in figures D.1 and D.2, as efficiencies > 1 were generated). If the approximation holds, the uncertainty assigned to ϵ is: [76]

$$\sigma_\epsilon = \sqrt{\frac{\epsilon(1-\epsilon)}{n_\mathrm{t}}}\,. \tag{D.9}$$

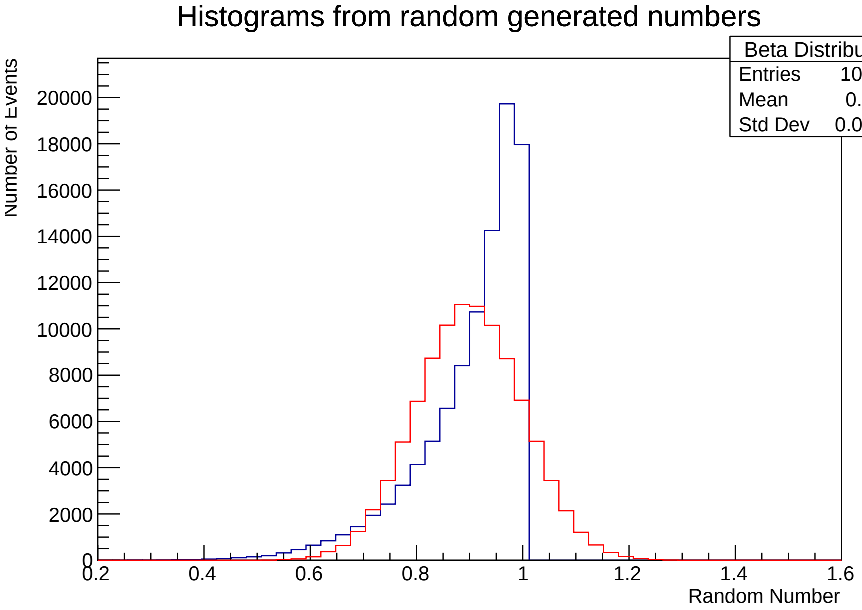


Figure D.2: Example of a Beta-distribution (in blue) with the parameters $\mu = 0.9, \sigma = 0.1$. More than 68 % of events lie within a 68 % confidence interval. For comparison purposes a Gaussian distribution with same parameters is displayed in red.

The impact of the uncertainties of efficiencies and acceptances to the total statistical uncertainty of the unfolded result was implemented by generating Gaussian distributed random numbers with $\mu = \epsilon$ and $\sigma = \sigma_\epsilon$ for each toy experiment in each bin. The process of random number generation was repeated if the generated number was smaller than 0. Applying this procedure to both acceptances and efficiencies lead to an increase of the total statistical uncertainty of the unfolded event yield of ≈ 3 % for the distributions tested. An example for the uncertainties associated to unfolded p_T^t bin contents when applying a cut of $O_\mathrm{NN} > 0.6$ (results shown in section 6.3.1), is shown in table D.1.

p_T^t[GeV]	$\frac{\mathrm{d}\sigma}{\mathrm{d}p_\mathrm{T}^t}$[fb/GeV]	Original Uncertainty [fb/GeV]	Uncertainty with Gauss [fb/GeV]	Relative Increase of Uncertainty [%]
0 – 65	0.437	0.132	0.137	3.3
65 – 110	0.758	0.151	0.156	3.3
110 – 160	0.420	0.104	0.108	3.3
160 – 300	0.130	0.044	0.045	2.0

Table D.1: Comparison of uncertainties without and with efficiency variation during random toy generation assigned to each bin of the unfolded data-distribution of p_T^t, with $O_\mathrm{NN} > 0.6$. For both cases, the random seed was set to the same value for accurate comparison.

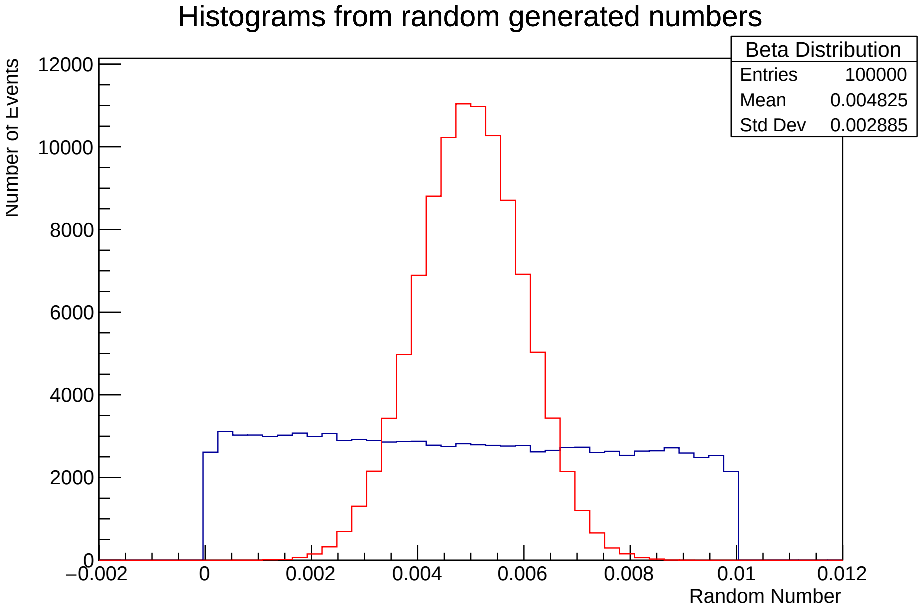


Figure D.3: Example of a Beta-distribution (in blue) with the parameters $\mu = 0.005, \sigma = 0.001$. Less than 68 % of randomly generated events lie within a 68 % confidence interval. Furthermore, the standard deviation as computed by `Root` deviates by a factor ≈ 3 from the input parameter. For comparison purposes a Gaussian distribution with same parameters is displayed in red.

D.4 Summary and conclusion

Tests were performed to see if Gaussian or Beta-distributions can be used to include statistical uncertainties assigned to efficiencies and acceptances to the bin contents of the unfolded distribution. Typical ranges that would be required for efficiencies and acceptances proved to be troublesome for random number generation via an underlying Beta-distribution. Results could be provided for efficiencies and acceptances generated according to Gaussian distributions. However, this result can at most be seen as an approximation, as the assumption of Gaussian distributions as probability density functions does not hold for $\epsilon \approx 0$ or $\epsilon \approx 1$. As a consequence, these concepts were included as options only in the framework.

The approach recommended for future analyses to include statistical uncertainties due to the size of MC samples is the following: based on the MC sample, a set of pseudo-data should be created. This pseudo-data is allowed to fluctuate according to statistical uncertainties. The determination of the unfolding matrix, the efficiencies and the acceptances should be performed for each set of pseudo-data individually. Calculating the variance when considering all pseudo-data samples can give an accurate estimate for the uncertainty associated to MC. Due to time constraints however, this approach was not implemented during this thesis.

List of Figures

List of Tables

Acknowledgements

The work presented in this thesis would not have been possible without the help of many different people. I am deeply grateful for all their help.

First, I would like to thank Prof. Dr. Ian C. Brock for the possibility to write a thesis about this interesting topic in your group. Thank you for all your advice and the interesting discussions when I ran into trouble for the N-th time. I am also happy that Priv.-Doz. Dr. Philip Bechtle agreed to be the second referee for this thesis.

Furthermore, the tZq process has been analyzed by a group of people from all over the globe. I want to especially thank Lidia, Dylan and Muhammad for all your patience during the initial phase of my work, and all your effort to help me understanding the tZq process and the software that we used.

In addition, I would like to thank many people in the Brock research group. Thank you, Tanja, Regina and Rico, for all your help when I had questions related to physics or programming. I am also very grateful that Rico decided to proofread my first draft of this thesis. Thank you, Christian, for your help related to machine learning techniques. Thank you, Rui, for explaining me everything one needs to know about unfolding and differential cross-section measurements. Thank you, Anji, for spending countless hours discussing statistics and correlations. Thank you, Maxx, for always having a positive attitude, always reminding me to keep a good mood. To the entire Brock research group (also those I have not mentioned already), thank you for all the interesting discussions and pleasant moments that make life enjoyable.

Lastly, I want to thank my family and friends for their support. Without their help and many words of encouragement, it would have been way more difficult for me to pursue and finish my studies.

It has been six amazing years living in Bonn. I am looking forward to new challenges in life, but would be glad to come back to Bonn once in a while and visit all the amazing people I have met during that time.