FAKULTÄT FÜR PHYSIK
Institut für Experimentelle Teilchenphysik
ETP-KA/2022-16

Measurement of the Triple-Differential Cross-Section of Z+Jet Production with the CMS Detector at 13 TeV

Masterarbeit
von

Cedric Verstege

am Institut für Experimentelle Teilchenphysik

Reviewer:	Priv. Doz. Dr. Klaus Rabbertz
Second Reviewer:	Prof. Dr. Günter Quast

Bearbeitungszeit: 01.12.2021 – 16.11.2022

Erklärung zur Selbstständigkeit

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der gültigen Fassung vom 24.05.2018 beachtet habe.

Karlsruhe, den 16.11.2022, _____

Cedric Verstege

Als Prüfungsexemplar genehmigt von

Karlsruhe, den 16.11.2022, _____

Priv. Doz. Dr. Klaus Rabbertz

Contents

1. Introduction

During the last century, the Standard Model of particle physics (SM) of particle physics has celebrated a lot of triumphs. For the longest period, the search for new particles was the driving force behind the progress in high-energy physics. But today, another aspect rises in importance. To get a better understanding of the fundamental building blocks of matter in our universe and their interactions, high-precision measurements become more and more relevant. Modern particle accelerators, detectors, and analysis methods made it possible to probe theories on an unmet precision level. The most powerful particle accelerator today is the Large Hadron Collider (LHC) at the Conseil européen pour la recherche nucléaire (CERN). There, protons are accelerated up to 6.8 TeV as of 2022, resulting in a center-of-mass energy of 13.6 TeV for proton-proton collisions.

For understanding and interpreting particle interactions at hadron colliders, the description of the hadrons' constituents is essential, since hadrons are not fundamental particles. The constituents of a hadron, the partons, each carry a different momentum fraction of the total hadron. These momentum fractions are described by the so-called parton distribution functions (PDFs). Due to the factorization theorem of Quantum Chromodynamics (QCD), the PDFs are the same for all scattering processes. Multiple different production channels are expected to improve the precision of the experimentally obtained PDFs, including vector boson production in association with a jet [1]. The presented analysis covers this crucial ingredient in the context of the Z boson production with at least one jet.

A triple differential measurement provides an insight into the different contributions from the proton's constituents to the Z+jet production in different regions of the analyzed phase space. The triple-differential inclusive cross-section is measured as a function of three different kinematic variables, namely the transverse momentum of the Z boson p_T^Z, the difference in rapidity between the Z boson and the jet y^*, as well as the boost of the center-of-mass system y_b. Those variables allow for a suitable division of the phase space in order to obtain a better sensitivity to the partonic sub-processes and the initial state parton momenta. A similar analysis strategy was already performed for the dijet production at a center-of-mass energy of 8 TeV [2]. The presented measurement is expected to contribute to a more precise determination of the PDFs, especially the gluon PDF. In combination with further differential cross-section measurements, such as the dijet and top-antitop quark production, an even better result can be achieved [1].

The data were recorded with the Compact Muon Solenoid (CMS) detector from 2016 to 2018 at a center-of-mass energy of 13 TeV. The measurement is performed for each data-taking period separately. When combined, the full data correspond to a total integrated luminosity of $138\,\mathrm{fb}^{-1}$.

Detector effects are corrected via an unfolding procedure taking the three dimensions of the measured phase space into account. The unfolded results are then compared to theory predictions at next-to-leading order (NLO) perturbation theory. Prospectively, the obtained results can be compared to state-of-the-art perturbative theory predictions

at next-to-next-leading order (NNLO) perturbative QCD, corrected for non-perturbative effects and including NLO electroweak (EW) corrections.

The muon decay channel of the Z boson is chosen, as it provides a sufficient number of signal events and at the same time, a minimal amount of background events. Additionally, the performance of the CMS detector allows for an efficient reconstruction and identification of muons. In the central region of the detector, a precise measurement is expected, with the total uncertainty at the order of the size of the luminosity uncertainty. At low transverse momenta of the Z boson, which is strongly correlated to the transverse momentum of the jet, the uncertainty is expected to be dominated by the jet energy correction (JEC), as its precision is reduced for low transverse momenta. At a high transverse momentum of the Z boson and forward-backward signatures in the detector, fewer events are expected, resulting in a dominating statistical uncertainty.

The theoretical foundations for the physical process analyzed in this measurement are given in chapter 2. An overview of the LHC and the CMS detector is presented in chapter 3, with a focus on the parts of the detector relevant to this measurement as well as the reconstruction of particle candidates. In chapter 4, the measurement is described, including an overview of the used observables, datasets, event selection, and the unfolding procedure. The obtained results are compared to theory predictions from simulations and the measurements for each data-taking period are compared. Finally, chapter 5 summarizes the obtained results and presents the prospects for the future of this analysis.

2. Theoretical Background

The Standard Model of particle physics (SM) is a theory describing all currently established elementary particles and three of the four known fundamental forces. Except for gravity, the fundamental interactions (electromagnetic, weak, and strong) are described by the SM.

With the prediction of the Higgs boson in 1964 [3–5] and the subsequent discovery in 2012 [6, 7], the theory has been proven highly successful. However, ten years after the discovery of the final elementary particle predicted by the SM, there are still many open questions in particle physics and cosmology [8] requiring more precise measurements of the SM and searches for physics beyond the SM. In section 2.1 the key concepts of the SM are described.

One of the fundamental forces in the SM is the strong interaction, theoretically described by Quantum Chromodynamics (QCD). An overview of QCD is given in subsection 2.1.2. The strong interaction is essential to understand and predict the behavior of hadron-hadron collisions, for example happening at the Large Hadron Collider (LHC). With the start of the High Luminosity LHC on the horizon, statistical uncertainties on experimental results probing the SM will be reduced drastically as the total integrated luminosity is planned to increase by a factor of ten [9]. This will lead to measurements being limited by the systematic uncertainties in the reconstruction of data and the theoretical predictions. One of the driving factors of uncertainties in theoretical predictions are the uncertainties on parton distribution functions (PDFs). They are one of the key components in theoretical predictions of cross-sections at proton-proton colliders. Hadrons, such as protons, consist of multiple *partons*. The momentum fractions each of the partons is carrying, is described by the PDFs. An introduction to PDFs and their importance is presented in subsection 2.2.1.

In order to improve the current understanding of PDFs and subsequently reduce the uncertainties on theory predictions, suitable measurements are needed [1]. For example, the measurement of the triple-differential cross-section for Z+jet production presented in this analysis. This allows more precise measurements of the SM and searches for new physics beyond the SM.

2.1. The Standard Model of Particle Physics

This section presents a rough overview of the concepts composing the SM while focusing on the strong force as it is the most relevant force for the topics considered in this thesis. It follows the descriptions given in references [10–13], which are recommended for more in depth explanations.

The SM describes the properties and interactions of elementary particles via quantum fields. Each of the known fundamental particles can be described as a quantized excitation of the corresponding quantum field. In the SM, the interactions between particles arise from the principle of local gauge invariance. There are two different types of fields, those with integer (bosonic) and half-integer (fermionic) spin. In multi-particle environments,

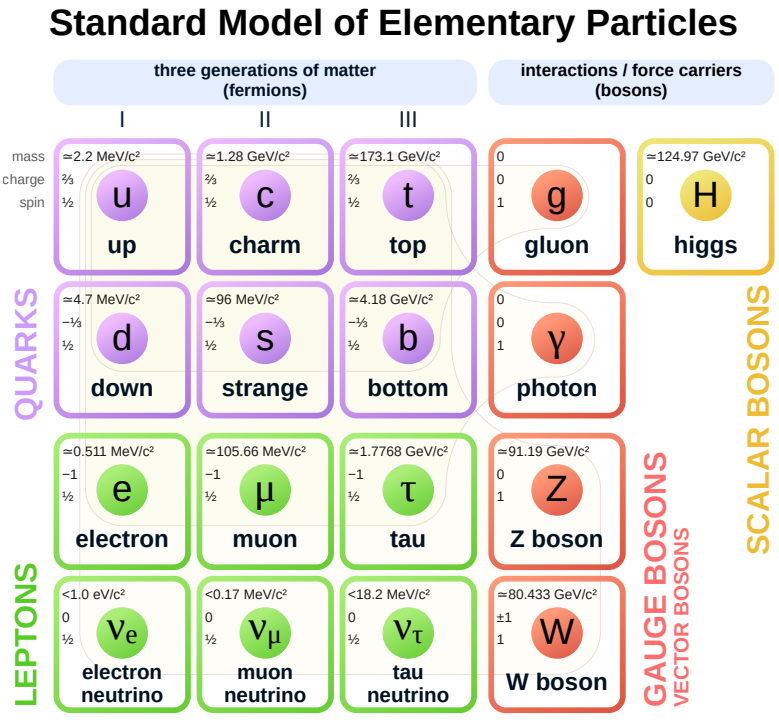


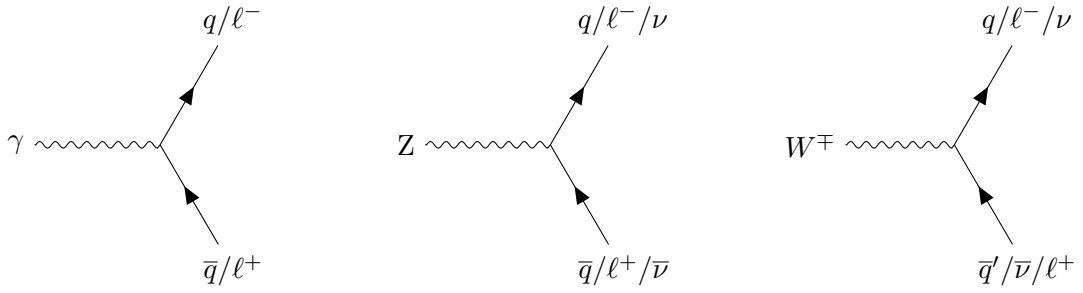
Figure 2.1.: Elementary particles of the SM and their properties. The quarks and leptons are fermions grouped in three different generations. Quarks are only observed in bound states, since they carry color charge and hence interact via the strong force and are therefore subject to color confinement. Quarks also carry electric charge and weak isospin, and therefore take part in electroweak interactions with other fermions. The leptons do not carry color charge. The electric charge carrying electrons, muons, and taus are taking part in electroweak interactions, while the charge neutral neutrinos only interact via the weak force. The fundamental interactions between particles are mediated by the bosons. Illustration taken from [14].

the quantum states for bosonic fields are symmetric under particle exchanges, while they are antisymmetric for fermionic fields. According to the SM, matter is made of fermions, while bosons mediate the fundamental forces.

Figure 2.1 provides an overview of the fundamental particles building up the SM. It consists of twelve types of fermions, six of which are leptons interacting via the electroweak (EW) force and six quarks additionally affected by the strong force. The EW interactions are mediated by the photon, the W bosons, and the Z bosons, while the strong interactions described by QCD are mediated by gluons. Through the introduction of the Higgs potential, the masses of the W and Z bosons are explained by spontaneous symmetry braking of the Higgs-field [3, 4]. Additionally, the masses of the fermions are explained through the Yukawa coupling to the Higgs field. The Higgs boson itself arises from excitation of the Higgs field.

2.1.1. Electroweak Theory

The EW theory is a unification of the electromagnetic and weak interactions invariant under transformations of the symmetry group $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. The weak hypercharge Y is the quantum number of the U(1) symmetry, while L denotes that the SU(2) symmetry only applies to left-handed particles with the third component of the weak isospin as quantum number.



(a) photon vertex **(b)** Z boson vertex **(c)** W boson vertex

Figure 2.2.: Vertices including fermions in electroweak theory. The Photon γ, as the exchange particle of the electromagnetic force, only interacts with charged leptons ℓ and quarks q. The Z and W bosons interact via the weak force with all leptons including neutrinos ν, in addition to the electromagnetic force. Interactions with the W boson also change the flavour of leptons and quarks. Additional (self-) couplings between the bosons are not shown.

With the degrees-of-freedom of the symmetry groups SU(2) and U(1) being three and one, the existence of four gauge bosons is implied. The physically observable bosons are a linear combination of the four electroweak bosons. This linear combination can be described as a rotation around the Weinberg angle $\cos(\theta_W) = \frac{m_W}{m_Z}$, where m_W and m_Z are the masses of the physically observable Z and $W^\pm$ bosons. Those masses arise due to the spontaneously broken symmetry of the Higgs field and are experimentally determined as $m_Z = (91.1876 \pm 0.0021)\,\text{GeV}$ and $m_W = (80.377 \pm 0.012)\,\text{GeV}$ for the Z and W bosons respectively [13].

The electrically neutral Z boson couples to electrically charged and weakly charged particles, with the weak charge being a linear combination of weak hypercharge and isospin. In contrast, $W^\pm$ bosons only couple to the weak isospin of a particle. The massless photon γ couples to particles with electric charge q, being a combination of weak isospin and hypercharge. In Figure 2.2 the interaction vertices of the three different electroweak gauge bosons with fermions are shown.

Photons can be converted into a quark-antiquark pair or two oppositely charged leptons of the same flavour, following the conservation rules of all quantum numbers. The same applies to the Z boson, while it can additionally interact with a neutrino-antineutrino pair via the weak force. In the presented analysis, Z boson candidates are reconstructed from two oppositely charged muons. W bosons can decay into a pair of up-type quark and down-type antiquark, or down-type quark and up-type antiquark. Furthermore, the W boson can decay into a lepton-antineutrino or antilepton-neutrino pair.

Self-couplings between the weak gauge bosons arise from the non-vanishing commutator in the field strength tensor for non-abelian gauge theories, like the SU(2). Due to the electroweak mixing, additional couplings between the gauge bosons emerge. As a consequence of the masses of the Z and W bosons, the weak interaction has a limited range, whereas the electromagnetic force carried by the massless photons has an infinite range.

2.1.2. Quantum Chromodynamics

The theory of QCD describes the strong interaction between particles. It is a Quantum Field Theory invariant under transformations of the SU(3) symmetry group. This symmetry introduces eight gauge bosons called *gluons*, which are the exchange particles of the strong force.

When using the Einstein summation convention, the QCD Lagrangian encoding the dynamics of the theory, is given by

$$\mathcal{L}_{\mathrm{QCD}} = \sum_q \overline{\psi}_{q,a} \left(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} \mathcal{A}^C_\mu - m_q \delta_{ab} \right) \psi_{q,b} - \frac{1}{4} F^{A\,\mu\nu} F^A_{\mu\nu} \,. \tag{2.1}$$

The variable $\psi_{q,a}$ denotes the quark-field spinors for a quark flavour q with mass m_q and color index a ranging from 1 to $N_c = 3$, meaning quarks exist in three colors. The fields $\mathcal{A}^C_\mu$ represent the gluon fields with C ranging from 1 to $N_c^2 - 1 = 8$ implying there are eight different kinds of gluons. While γ^μ are the Dirac matrices and δ_{ab} is the Kronecker delta, t^C_{ab} are the eight generators of the SU(3) group in fundamental representation. The generators are given via $t^C_{ab} = \lambda^C/2$ by the traceless Hermitian Gell-Mann matrices λ^C [15]. The parameter g_s is the coupling strength of the strong interaction. In particle physics, the coupling strength is often referred to as $\alpha_s = \frac{g_s^2}{4\pi}$. It is one of the fundamental parameters of the QCD next to the quark masses, which are free parameters in the electroweak Lagrangian.
The field strength tensor is given by

$$F^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu - \partial_\nu \mathcal{A}^A_\mu - g_s f_{ABC} \mathcal{A}^B_\mu \mathcal{A}^C_\nu \tag{2.2}$$

with the SU(3) structure constants f_{ABC}. They are required to attain local gauge invariance, since the SU(3) is a non-abelian group. The structure constants are given via the commutation relations

$$[t_A, t_B] = i f_{ABC} t^C \,. \tag{2.3}$$

The different terms of the Lagrangian are related to the allowed vertices. In Equation 2.1, the quark spinor fields, $\overline{\psi}_q$ and ψ_q, couple to the gluon field $\mathcal{A}$ with the coupling strength g_s. The generators t^C_{ab} mathematically describe the interaction between a quark and a gluon as a rotation of the quark's color in the SU(3) space. The last term in Equation 2.2 describes a self-coupling of the gluon field, causing cubic and quartic gluon self-interactions proportional to g_s and g_s^2.

In contrast to Quantum Electrodynamics (QED), the running of the strong coupling constant, introduced through renormalization, increases with low energies, resulting in a phenomenon called *confinement*. When bound quarks are separated, the energy density of the field of the strong interaction increases, and it is energetically favorable for additional color charged particles to be created from the vacuum. Hence, quarks appear only in colorless bound states called *hadrons*. Those can be for example *baryons* consisting of three quarks, such as protons and neutrons or *mesons* consisting of a quark-antiquark pair. Another effect of the confinement is the so-called hadronization of color charged particles in the final states of scattering processes. Isolated color charged particles create a cascade of additional color charged particles, which share the energy of the original particle. Those particles then form a coherent collection of confined particles traversing in the direction of the original particle. The resulting collimated showers of particles are called *jets*. In section 2.3, clustering methods of the showers are described combining the resulting particles of the shower to approximating the energy and momentum of the particle initially causing the jet.

At high energies, corresponding to small distances, strong interacting particles behave like free particles. This is caused as well by the dependence of the strong coupling constant on the renormalization scale. The value of the strong coupling constant is decreasing with a rising energy scale. This effect is known as *asymptotic freedom*. This means, in a first approximation, that the partons itself take part in the interaction at high energies such as at the LHC Therefore, the PDFs described in subsection 2.2.1 are needed to predict which partons of the protons colliding at the LHC contribute to the scattering processes.

2.2. Calculating and Predicting Cross-Sections

The cross-section of a scattering process contains information on the interaction strength as well as the kinematic characteristics. The interaction is encoded in the matrix element $\mathcal{M}$, while $d\phi(p_1 + p_2, P_f)$ is the differential phase space of the process. In this case, the differential phase space includes the kinematic properties of the two initial particles with momentum p_1 and p_2 as well as a set of four momenta of the particles in the final state P_f:

$$d\sigma \propto |\mathcal{M}|^2 \times d\phi(p_1 + p_2, P_f)\,. \tag{2.4}$$

In order to compute the cross-section of a specific process, its matrix element $\mathcal{M}$, describing the probability for all possible transitions from an initial state i to a final state f, has to be computed:

$$\mathcal{M}_{i\to f} \propto \langle P_f | \, \frac{1}{i}(S-1) \, |p_1, p_2\rangle \ . \tag{2.5}$$

At a collider the initial state is typically formed by two particles. Their momenta are given by p_1 and p_2. The set of four momenta in the final state are given by P_f. The scattering matrix S contains the information about the underlying interactions described by the SM. Given the time-ordering operator $\mathbf{T}$ and the Lagrangian describing the interaction $\mathcal{L}_I$, the scattering matrix can be written as

$$S = \mathbf{T} \exp\left(i \int_{-\infty}^{\infty} d^4x \mathcal{L}_I(x)\right). \tag{2.6}$$

The Taylor expansion in the coupling constant of the interaction in the exponential function in Equation 2.6 is the basis of perturbation theory. A calculation of S to the first non-vanishing series term is called a leading order (LO) prediction. Taking the next term into account, a more precise calculation at next-to-leading order (NLO) is possible. The most accurate calculations of the Z+jet cross-section to date are performed at next-to-next-leading order (NNLO) in QCD perturbation theory [16, 17].
Every term of the Taylor expansion contributes to the matrix element in Equation 2.5. Each order of the expansion can also be split into a number of processes by using Wick's theorem [18]. The different processes contributing to the overall cross-section can also be visualized graphically as Feynman diagrams [19], while each diagram can contribute to multiple orders of the expansion.

Beyond contributions from LO processes, loops can appear inside the Feynman diagrams. Loops are, for example, caused by virtual particles emitted and reabsorbed within the interaction process and are therefore not contributing to the initial and final states. Due to the unconstrained momenta in those loops, logarithmic divergences inside the integral in Equation 2.6, so-called ultraviolet divergences, occur. Those infinities are absorbed into a finite number of parameters in a process called *renormalization*. A side effect of renormalization is the introduction of the renormalization scale μ_R as a regulator for the infinities. This results in the dependence of the regularized parameters on the renormalization scale, for example the running of the strong coupling constant $\alpha_s(\mu_R^2)$.

2.2.1. Parton Distribution Functions

So far, in the SM perturbation theory can only be applied to interactions between two partons. Due to confinement, described in subsection 2.1.2, no free quarks and gluons can be observed in nature. Hence, only hadrons such as protons can be used for high energy collision, for example at the LHC. Protons are baryons consisting of two up and one down quark. These three quarks are called valence quarks. When taking a closer look into the proton and hence going to higher energy scales, the gluons carrying the strong interaction

between the valence quarks can split into two or three gluons due to the self-coupling or create new quark-antiquark pairs. Thus, not only the valence quarks but also so-called sea quarks of any flavour, as well as gluons, can be found inside a hadron.

With the factorization theorem of the QCD, the computation of the cross-section can be approximated by separating it into two parts at an energy threshold called the *factorization scale* μ_F. At energies greater than μ_F, perturbative QCD applies and the cross-section can be calculated as described before. This calculation depends on the factorization and renormalization scales μ_F and μ_R. For energies below μ_F, non-perturbative parameterizations are used to describe low energy, long distance effects. Each of the partons inside a hadron carries a part x of the total hadron momentum p with $0 < x < 1$. The PDF $f_a(x, \mu_F)$ for a parton a describes the probability for this parton to carry the momentum fraction x_a of the hadron. The PDFs are dependent on the chosen factorization scale μ_F and cannot be derived theoretically. Therefore, precise measurements of the PDFs are a crucial ingredient for enabling the theory predictions at hadron colliders. The cross-sections can be factorized as

$$\sigma_{pp\to X} = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b \; f_a(x_a, \mu_F) \cdot f_b(x_b, \mu_F) \cdot \sigma_{ab\to X}(x_a, x_b, \mu_F, \mu_R) \,. \tag{2.7}$$

Here, a and b are the initial state parton flavours or gluons with their corresponding momentum fractions x_a and x_b. The sum runs over all contributing initial state partons.

To obtain the best fit for PDFs, measurements are compared to theoretical predictions and the underlying PDFs are varied until the best agreement is found. The fit result is dependent on the measurement scale μ. The DGLAP equations [20–22] are used to evolve the PDFs to other scales and use them in theory predictions. In Figure 2.3, the PDFs obtained by the NNPDF collaboration are shown as an example at two different scales.

2.3. Jet Clustering

Due to the confinement of the strong force, final state partons of scattering processes fragment into a multitude of particles. This effect is called hadronization as described in subsection 2.1.2. Those particles are moving approximately in the same direction predetermined by the initiating parton. In order to approximate the energy and momentum of the parton initiating the shower, multiple algorithms have been developed that cluster the four-momenta of the resulting particles into a well-defined *jet*.

For a better comparison to theory predictions, jet clustering algorithms are applied to different simulation levels: parton level, particle level, and detector level. At *parton level*, a jet typically consists of only a few color charged constituents, since only the event topology of the final state is taken into account. The *particle level* describes the simulation after non-perturbative effects, such as hadronization, are taken into account. Hence, jets at particle level consist of multiple color neutral particles introduced through the non-perturbative effects. After performing the detector simulation and applying the reconstructing algorithms, the *detector level* is reached, which is the only one directly accessible via an experiment. Jets at detector level are clustered from the four-momenta of the particles reconstructed from the energy depositions in the detector.

The two common types of jet clustering algorithms are *cone* and *sequential recombination* algorithms. Cone algorithms perform a sum over all the four-momenta of particles j inside a cone Δ around a starting four-momentum of the object i

$$\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} \,. \tag{2.8}$$

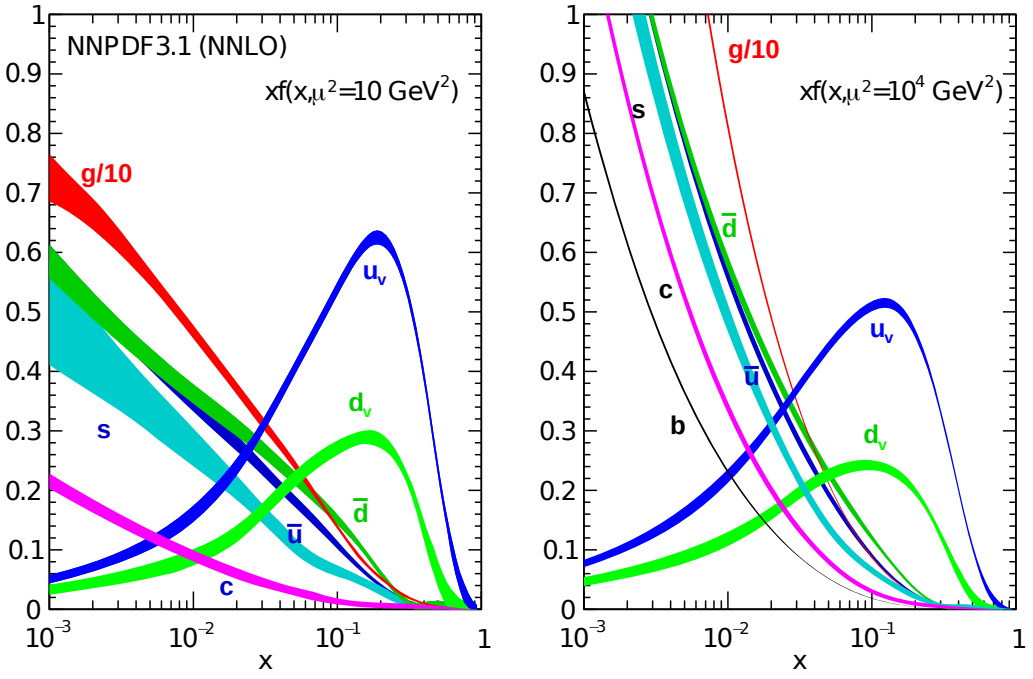


Figure 2.3.: The PDFs as a function of the momentum fraction x are shown for the up (u_v) and down (d_v) valence quarks. Addtitionally the PDFs for virtual gluons (g) and virtual quarks, also called sea quarks, are shown. Depicted are the strange (s), charm (c), and bottom (b) quarks, as well as the up ($\bar{u}$) and down ($\bar{d}$) antiquarks. The PDFs are obtained by the NNPDF3.1 fit [23] and are given at a scale of $\mu^2 = 10\,\mathrm{GeV}^2$ (left) and $\mu^2 = 10^4\,\mathrm{GeV}^2$ (right).

Here, y_i and y_j are the rapidities of the objects i and j, while ϕ_i and ϕ_j denote the azimuthal angles of the particles inside the reference frame of the detector. The coordinate system used at the Compact Muon Solenoid (CMS) detector is explained in subsection 3.2.1.

In contrast, sequential recombination algorithms cluster jets with an iterative approach. Here, a distance measure d_{ij} between two objects i and j is used, taking the transverse momentum $p_{\mathrm{T}i}$ of each particle, a distance parameter R, and an additional power parameter p into account, yielding

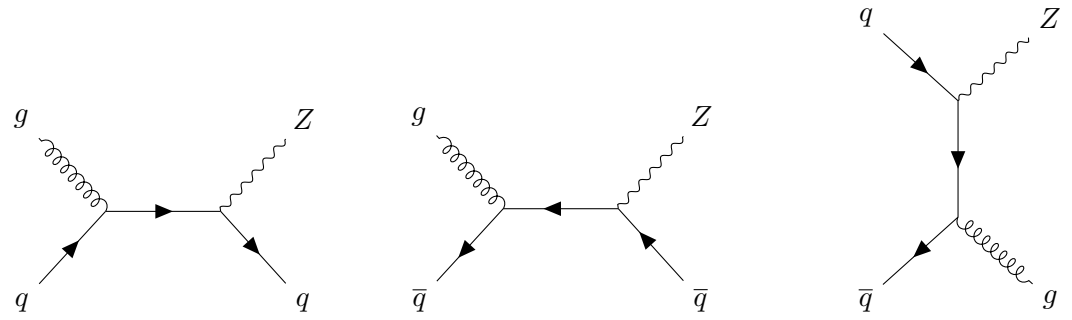
$$d_{ij} = \min\left(p_{\mathrm{T}i}^{2p}, p_{\mathrm{T}j}^{2p}\right) \frac{\Delta_{ij}^2}{R^2}\,. \tag{2.9}$$

Additionally, a distance measure with respect to the beam axis is defined as

$$d_{iB} = p_{\mathrm{T}i}^{2p}\,. \tag{2.10}$$

The sequential recombination algorithm first calculates all distances d_{ij} between entities i and j of an event, as well as the distances d_{iB} of entity i to the beam axis B. The clustering proceeds by identifying the smallest distances d_{ij} and combining the four-momenta of the closes pair i and j. If the smallest distance is d_{iB}, the corresponding entity is called a jet and removed from the list of entities. This procedure repeats until no entity is left.

In this analysis, the anti-k_t jet clustering algorithm [24], which is a sequential recombination algorithm with $p = -1$, is used with a distance parameter of $R = 0.4$. Using the anti-k_t algorithm, the shape of a jet is defined by the hard particles and not affected by soft radiation. The clustering starts around hard particles, since the d_{ij} between a hard and a soft particle is much smaller than between two soft particles. If no other hard particle is within a cone of $2R$, the resulting jet is clustered from the surrounding soft particles



(a) quark-gluon **(b)** antiquark-gluon **(c)** quark-antiquark

Figure 2.4.: Feynman diagrams for the leading order production of Z+jet events.

and has a conical shape. The jets are clipped if two hard particles are closer together, or combined into one jet if the distance Δ_{ij} is smaller than R. This results in non-conical jets with a more complex shape.

2.4. Z Boson Production With at Least One Jet at Proton Colliders

The focus of this analysis is the production of a Z boson in association with at least one jet. The inclusive Z boson production in proton-proton collision is often referred to as Drell-Yan (DY) production. In Figure 2.4, the LO Feynman diagrams of the Z+jet production at proton-proton colliders are shown.

The contribution of a specific sub-process to the cross-section is dependent on the scattering angle in the center-of-mass system of the interacting partons. The total boost of the center-of-mass system is given by the momentum fractions of the partons and thus the PDFs. A division of the phase space in the scattering angle and boost of the center-of-mass system, as done in this analysis, gives an insight to the contributions in the initial state and the PDFs.

Comparing the different production channels to the already known PDFs from other analyses in Figure 2.3, some predictions can be derived. There is a high probability for a gluon or antiquark to carry a low momentum fraction, while the probability for a quark to carry a larger momentum fraction is high. The PDF of the gluon is approximately a factor of ten higher than the PDF for an antiquark at a low momentum fraction. Hence, antiquark-gluon scattering will result in events with an overall low energy and the quark-gluon sub-process will be dominant compared to the quark-antiquark sub-process.

At NLO and NNLO, further production channels are introduced, with their contribution to the total cross-section also being highly dependent on the scattering angle in the center-of-mass system. Those higher order contributions consist of quark-quark, antiquark-antiquark and gluon-gluon scattering. In order to probe different regions of the phase space in the final state and their different initial state contributions, the rapidity observables y_b and y^* are introduced in chapter 4 together with the differential cross-section measurement of this process.

3. Experimental Setup

To expedite research in fundamental physics – especially particle physics – the European Organization for Nuclear Research or Conseil européen pour la recherche nucléaire (CERN) was founded in 1954 [25]. The CERN research facilities are located at the border between Switzerland and France, close to Geneva. As of 2021, CERN is the largest research hub for particle physics worldwide, with more than 16.000 members and associates working at or with CERN [26]. The rapid development and research in particle physics during CERN's operation has led to the construction of the most powerful particle accelerator, the Large Hadron Collider (LHC).

Particle accelerators such as the LHC are vital to test the theoretical predictions derived from the SM. By colliding particles and performing scattering experiments in a controlled environment, the fundamental forces and interactions between particles can be probed. To explore particles with higher masses or interactions at smaller distances, high energies for example provided by the LHC are necessary.

This chapter gives an overview of the LHC, followed by the setup of the Compact Muon Solenoid (CMS) experiment which measured the data used in this thesis from 2016 to 2018. More detailed information on the LHC and its design can be found in ref. [27].

3.1. The Large Hadron Collider

The LHC is a proton synchrotron with a circumference of 27 km situated about 100 m below ground level at the CERN accelerator complex. It is the world's largest and most powerful particle accelerator. Figure 3.1 illustrates the CERN accelerator complex and the LHC with its four major experiments. The protons are accelerated by various precursor accelerators before they enter the LHC. Before 2020, hydrogen stripped of its electrons was injected into the first accelerator stage, the Linac2 [28]. Since 2020, the Linac4 [29] accelerates negatively ionized hydrogen, which is then stripped of the electrons before entering the next stage. After the linear accelerator, the protons are injected into the Proton Synchrotron Booster and are accelerated up to 2 GeV. Following the Booster, the particles are transferred into the Proton Synchrotron and reach an energy of 26 GeV. The last accelerator stage before the protons are injected into the LHC is the Super Proton Synchrotron, where the particles are accelerated up to 450 GeV. An overview of the current accelerator complex at CERN is found in ref. [30].

The LHC itself contains two beam pipes in which protons are accelerated in opposite directions up to an energy of 6.8 TeV as of 2022. During the second measurement period from 2015 to 2018, often referred to as Run 2, protons were accelerated up to an energy of 6.5 TeV, resulting in a center of mass energy of $\sqrt{s} = 13\,\text{TeV}$. At this energy, protons are moving at 99.999 999 % of the speed of light. An in depth overview over the operation and configuration of the LHC during Run 2 is found in ref. [32]. Protons are injected into the LHC in bunches with a nominal bunch intensity of 1.15×10^{11} protons and are

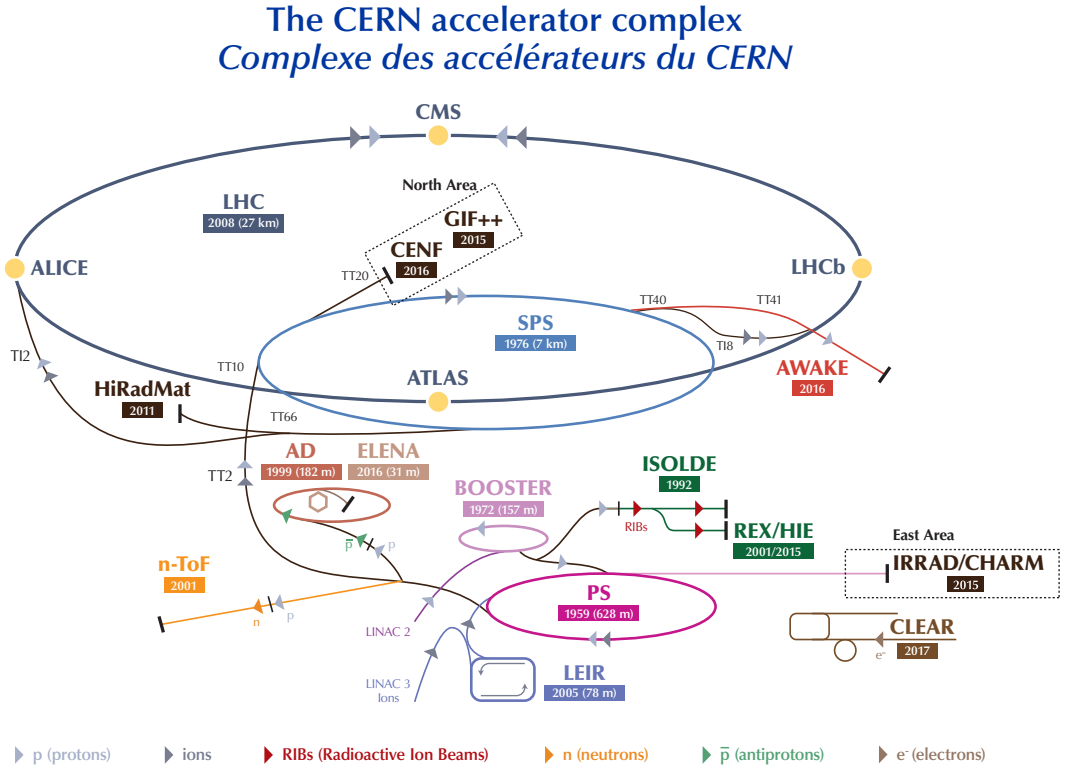


Figure 3.1.: The CERN accelerator complex as of August 2018. The main experiments CMS, ATLAS, LHCb, and ALICE are located at the LHC. Further experiments at CERN are shown as well. Hydrogen ions stripped of their electron are injected at LINAC2. The protons then enter the BOOSTER and are further accelerated by the PS and SPS before finally reaching the LHC. Illustration taken from [31].

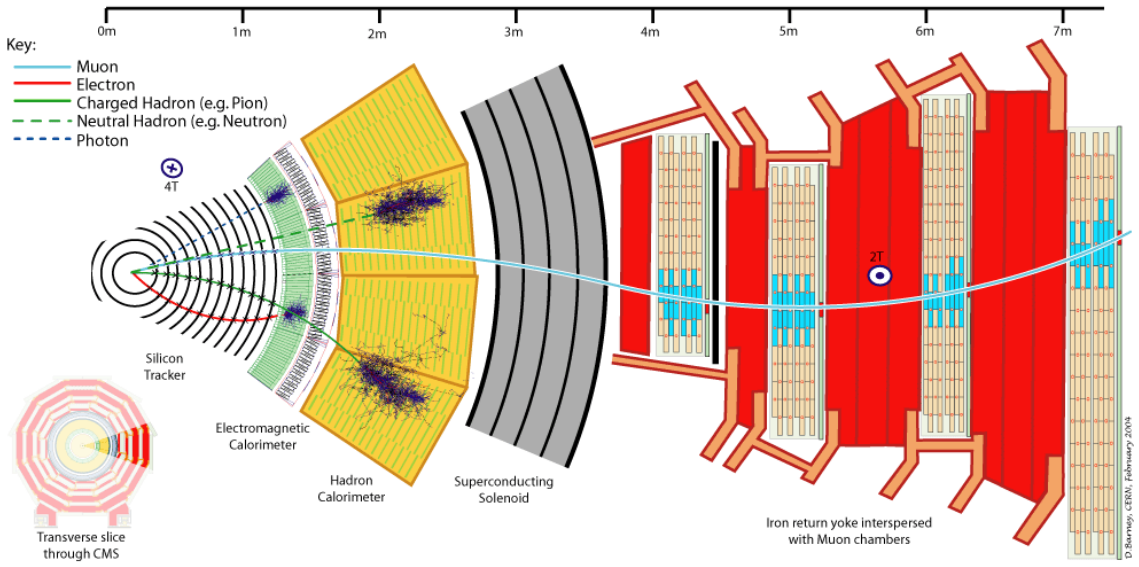


Figure 3.2.: A piece of a transverse slice through the CMS detector. All different subsystems and the interaction of selective particles, namely a positron (red), negatively charged muon (blue), a charged hadron (green), a neutral hadron (green dashed), and a photon (blue dashed) with them is depicted. The innermost part around the beam pipe is the silicon tracker, which detects the trajectory and position of the interaction point of charged particles. Outside the tracker, the electromagnetic and hadronic calorimeters measure the energy of individual or groups of particles. The superconducting magnet creates a magnetic field of 3.8 T, which bends the path of charged particles. The steel return yoke is restricting the magnetic field to the inside of the detector and the muon chambers. Combining the information from the detector subsystems enables the differentiation and reconstruction of particles. Illustration taken from [33].

spaced so that a bunch crossing occurs every 25 ns. The particles are kept on track by superconducting electromagnets, which are designed to produce nominal magnetic fields of 8.33 T. In addition to proton-proton collisions, the LHC is also used as a lead ion (Pb) collider. Proton-proton (or Pb-Pb) collisions can occur at the interaction points where the two beams intersect. The LHC houses four major experiments which are located at those interaction points. There are two general purpose experiments, CMS and A Toroidal LHC ApparatuS (ATLAS), designed for operation in high luminosity environments, as well as two specialized experiments, Large Hadron Collider beauty (LHCb) and A Large Ion Collider Experiment (ALICE). While LHCb is focusing on B-meson physics, ALICE was built to mainly investigate heavy ion collisions.

3.2. The Compact Muon Solenoid

The CMS detector is a general purpose experiment. The design allows high luminosity measurements for a broad range of particles, enabling precision studies of the SM and searches for new particles. One of the main goals was to find the Higgs boson, which was achieved in 2012 together with ATLAS [6, 7]. The central feature of the CMS detector is a superconducting solenoid of 6 m internal diameter producing a magnetic field of 3.8 T which bends the path of charged particles. Combining the information from the detector subsystems, a global event reconstruction is possible, as described in the following subsection 3.2.2. Figure 3.2 shows a small piece of a transverse slice through the CMS detector, depicting the detector subsystems.

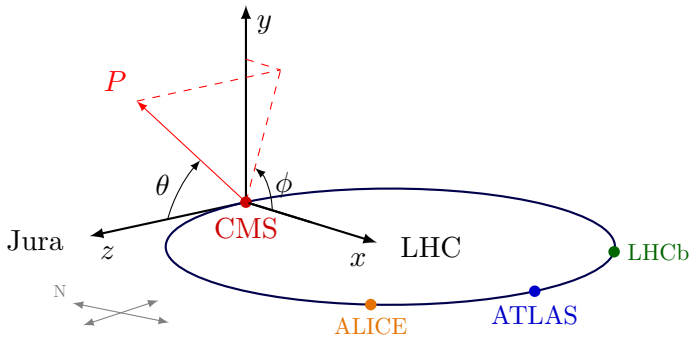


Figure 3.3.: The CMS coordinate system is shown in relation to the LHC ring and the four main experiments. The x-axis points to the center of the LHC ring, while the y-axis points upwards to the surface. The z-axis points along the beam pipe to the Jura mountains, completing a right-handed coordinate system. Additionally, the angles ϕ and θ that are used when describing momenta in spherical coordinates are depicted. Illustration taken from [34]

3.2.1. Coordinate System and Definitions

The coordinate system used at the CMS detector is right-handed and has its center at the interaction point. While the x-axis points to the center of the LHC ring, the y-axis points vertically upwards. Completing the right-handed coordinate system, the z-axis points along the beam direction to the Jura mountains.

Particle momenta are usually described in spherical coordinates. The azimuthal angle ϕ is measured in the x-y-plane starting at the x-axis, and the polar angle θ is measured from the z-axis. Figure 3.3 shows a sketch of the CMS coordinate system. The momentum and energy transverse to the beam axis are denoted as p_T and E_T, respectively. Additionally, the pseudorapidity is defined as $\eta = -\ln\tan(\theta/2)$. At high energies, if the mass of a particle is negligible compared to the momentum, the pseudorapidity equals the rapidity.

3.2.2. Detector Subsystems

In order to reliably reconstruct, track, and identify particles traversing the detector, different detection systems are necessary. Therefore, the detector is designed in several layers around the interaction point. Each of these layers is tailored to the detection of different types of particles. After the superconducting solenoid – eponymous for the letter S in the name of the experiment – is described, the detector subsystems are presented from the interaction point in the center to the outside of the detector as shown in Figure 3.2. An in-depth description of the CMS detector can be found in ref. [35].

According to the name, the magnet plays a central role in the identification of charged particles and measuring their momenta. The trajectory of charged particles is deflected through the Lorentz force $F_L = qv_TB$. Here, q denotes the electric charge, v_T describes the transverse velocity with respect to the magnetic field, and B is the magnetic field strength. Since the magnetic field inside the CMS detector is aligned along the beam axis, the total transverse momentum of the collision is conserved and extracted for each particle from the curvature in the x-y plane. The particle's transverse momentum p_T can be calculated by determining the radius r of the curvature, since the centripetal force F_c of the circular movement is induced by the Lorentz force

$$F_c = F_L \Leftrightarrow \frac{p_T v_T}{r} = qv_TB \Leftrightarrow p_T = qrB\,. \tag{3.1}$$

Therefore, a stronger magnetic field leads to a better momentum resolution, as tighter radii of the particle tracks can be measured more precisely. Hence, the CMS detector contains

a superconducting solenoid located outside the hadronic calorimeter (HCAL). Inside the solenoid, a homogenous magnetic field of 3.8 T aligned to the beam axis is created [36]. For returning the magnetic flux, a steel return yoke composed of six end cap disks and five barrel wheels is located outside the solenoid. As a consequence, the magnetic field outside the solenoid is still about 2 T strong.

If particle collisions occur, the newly created particles first pass through the inner tracking system, surrounding the interaction point at a length of 5.8 m and a diameter of 2.5 m. This region is fully contained in the homogeneous magnetic field of 3.8 T produced by the solenoid. The tracker is built using silicon detectors designed to precisely and effectively measure the trajectory of charged particles. Initially, three layers of the pixel detectors were placed inside ten layers of silicon strip trackers, enabling more precise measurements closer to the beam pipe. In forward and backward direction, the end plates consisting of two disks of pixel detectors and twelve strip trackers at each side are located. During the year-end technical stop of 2016/17, a new pixel detector with four layers and three disks was incorporated into CMS [37]. By combining information from each pixel and strip detector, the trajectory of charged particles through the tracker can be fitted. From the fit result, momentum and charge of a particle are determined. This is the basis for the particle-flow (PF) event reconstruction described in subsection 3.2.4.

During a bunch crossing, multiple proton-proton collisions can occur. The points of the interactions are called *vertices*. Usually, only the interaction with the highest momentum transfer is the one of interest. Particles originating from other collisions inside the same bunch crossing are therefore referred to as *pileup*. The track reconstruction allows for vertex matching and a classification if charged particles originate from the primary vertex or the pileup vertices.

Most of the particles created in high energy collisions pass through the tracker. They then reach the electromagnetic calorimeter (ECAL), where the energy of photons, electrons, and charged hadrons is deposited and measured. The ECAL in the CMS experiment is a hermetic homogeneous calorimeter made of lead tungstate ($PbWO_4$) crystals. In a homogeneous calorimeter, the scintillator and absorber are the same material. This results in an excellent energy resolution as the total energy of a particle is deposited inside the scintillator. Photons, electrons, and charged hadrons interact with the material and initiate a cascade of particles, a so-called electromagnetic shower, evolving within the ECAL. The shower's energy is converted into light by scintillation. The light is then captured and turned into an electric signal by avalanche photodiodes as photodetectors in the barrel region and vacuum phototriodes in the end cap. The crystals cover about 25 times the radiation length of lead tungstate [38], as the electromagnetic cascade has to be contained entirely inside the ECAL to measure the total energy of a particle. The amount of emitted light is directly proportional to the deposited energy. This allows for a precise energy measurement with a resolution from 1.6 % to 5 % [39].

After this subsystem, mostly hadrons and weakly interacting particles like muons and neutrinos survive. For measuring the energy of hadrons, the HCAL [40] is placed around the ECAL. Hadrons have a larger mass than electrons and are therefore not as much affected by bremsstrahlung. Thus, the total energy often can not be contained within a homogenous calorimeter of adequate size. Hence, CMS uses a sampling calorimeter for the HCAL, which means that the active calorimeter alternates with absorbers increasing the interaction probability and energy deposition. The HCAL in the barrel region consists of a steel plate followed by eight brass absorbers and an outer steel plate. Between the steel plates and brass absorbers, the plastic scintillators are located. The measurement principle is therefore the same as for the ECAL. The hadrons traversing the HCAL form showers in the scintillator and additionally in the absorbers. The energy deposition of the showers is

measured by the scintillation light created when the showers are traversing the scintillators. The strength and duration of the measured signal is interpreted as deposited energy and corrected for the undetectable energy depositions inside the absorbers.

Nearly all particles except for the weakly interacting muons are stopped inside the detector. An additional subsystem, the so-called muon chambers, for measuring the momentum and energy of muons is required. There are three different types of muon chambers embedded between the layers of the steel flux-return yoke. The muon system consists of different types of gaseous particle detectors. In the barrel region up to $|\eta| < 1.2$, drift tubes are used. In the end cap between $0.9 < |\eta| < 2.4$, cathode strip chambers are used. This setup allows for a full muon detector coverage up to $|\eta| < 2.4$. A more detailed description of the muon chambers can be found in ref. [41]. Combining the information from the muon detector system with the inner tracker, precise muon identification and reconstruction is possible with an efficiency greater than 96 %. Additionally, resistive plate chambers are located in the muon system, which are mainly used for a fast muon reconstruction during the triggering process described below. The single muon trigger used for this analysis exceeds an efficiency of 90 % over the full range of the muon system. When matching the muons to tracks measured in the silicon tracker, a relative resolution of the transverse momentum of 1 % in the barrel and 3 % in the end cap is achieved for muons with p_{T} up to 100 GeV. In the barrel region, the relative p_{T} resolution for muons with a transverse momentum up to 1 TeV is better than 7 %. A complete overview of the muon detector and muon reconstruction performance during Run 2 can be found in ref. [42].

With this detector setup, the momentum and energy of all known long-lived particles of the SM except for neutrinos can be reconstructed. All subsystems combined produce an enormous amount of data. Therefore, a trigger system described in the following section is needed for the data taking. The full reconstruction of final state objects in an event is described in the section thereafter.

3.2.3. Trigger and Data-Taking

In order to establish the need of a dedicated trigger system for data storage, a small estimation on the storage requirements based on ref. [43] is given. All detector systems produce about 1 MB of data per event, assuming around 20 inelastic proton-proton collisions. The average number of interactions per bunch crossing during the 2016, 2017 and 2018 data taking periods was 23, 33, and 32 respectively [44]. Since bunch crossings occur every 25 ns, resulting in a frequency of 40 MHz, around 40 PB of data would need to be stored every second if every event was kept. This data rate exceeds the capabilities of what any computing infrastructure is able to handle. In order to reduce the data rate to about 1 GB per second and select events of potential physical interest, a two-tiered trigger system is used in the CMS experiment.

First, the Level-1 (L1) trigger implemented in hardware scans for simple signatures such as energy deposition in the calorimeters or hits in the muon system. The L1 triggers reduce the rate of events accepted for further processing to less than 100 kHz. If an event is accepted by an L1 trigger, the whole detector is read out and the data is forwarded to software based high level triggers (HLTs). The HLTs perform a preliminary event reconstruction and decide if an event is permanently stored or not processed any further. This lowers the data rate even further to about 100 Hz. The combination of an L1 and an HLT is called a *trigger path*. If an event passes at least one trigger path, it is labelled accordingly and stored permanently. More detailed information on the CMS trigger system can be found in ref. [45]. A detailed overview of the L1 trigger performance during Run 2 is found in ref. [46].

In this analysis, the single muon trigger path is used. This trigger path is unprescaled and aims at storing every event where at least one muon per event above a certain p_{T} threshold is detected. Further information on the used triggers for each data taking period is found in subsection 4.3.1.

3.2.4. Event and Object Reconstruction

In order to identify and reconstruct each individual particle in an event, a global event reconstruction, also called PF event reconstruction [47], is used. It enables the identification of five different PF candidates: photons, electrons, muons, charged hadrons, and neutral hadrons. For the identification, as well as the determination of the particle trajectory and energy, information from all the sub-detectors of the CMS apparatus is combined.

Muons are reconstructed from tracks in the central tracker consistent with either a track or hits in the muon system. Standalone muons are defined as a track in the muon system only. If an extrapolated track reconstructed from the silicon tracker matches at least one hit in the muon system, the combined object is defined as a tracker muon. Global muons are defined as a tracker muon which matches the track of a standalone muon. The energy of muons is obtained from the corresponding track momentum.

Electron candidates are built from the remaining charged tracks reconstructed in the silicon tracker, matched to energy depositions in the ECAL. The electric charge is determined via the track curvature resulting from the Lorentz force. Additionally, photon emission from bremsstrahlung is taken into account during track reconstruction and the energy deposition in the ECAL matching with related photons is treated accordingly.

All remaining tracks that are neither identified as muon nor as electron are matched to energy depositions in the ECAL and HCAL and are treated as charged hadrons. Energy depositions in the ECAL which do not match any track are defined as photons, those in the HCAL as neutral hadrons.

Hadronic jets are clustered from the reconstructed PF candidates using the anti-k_t algorithm [24, 48] with a distance parameter of $R = 0.4$ as described in section 2.3. The jet momentum is calculated as the vectorial sum of all the momenta of its constituents. Additional proton-proton interactions within the same or nearby bunch crossings are classified as pileup and can contribute to the jet momentum. To mitigate this effect, a method called charged hadron subtraction (CHS) is used. All charged PF candidates not originating from the leading vertex are discarded before starting the jet clustering. Additionally, the resulting jet energy is corrected for the average energy deposition originating from neutral pileup particles. The jet momentum is found from simulation to be within 5 % to 10 % of the true momentum over the whole p_{T} spectrum and detector acceptance. To account for this difference, jet energy corrections are derived from simulation and residual differences between data and simulation are accounted for by in situ measurements of the momentum balance in $Z + \mathrm{jet}$, $\gamma + \mathrm{jet}$, dijet, and multijet events. The resulting jet energy resolution is typically around 15 % to 20 % at 30 GeV, 10 % at 100 GeV and 5 % at 1 TeV [49].

4. Measurement of the Triple-Differential Z+Jet Cross-Section

The measurement of the triple-differential cross-section for Z+jet production is a unique opportunity to study the EW and QCD parts of the SM with high precision at a hadron collider. It can be exploited to shed light on the contribution of the different sub-processes and the corresponding contribution from the PDFs, as described in section 2.4.

The first triple-differential cross-section measurement of the Z+jet production was already performed in ref. [50] for a preliminary reconstruction of the data taken with the CMS detector in 2016. This measurement was then improved and expanded to the data taken in 2017 in ref. [51]. Discrepancies bigger than the estimated uncertainties were observed between the two data-taking periods. The first preliminary reconstruction of the observed collisions is never perfect and therefore could be a source for the observed inconsistency. With a better understanding of the detector, the reconstruction of the data can be improved. The goal of this analysis is to perform the same measurement with the full Run 2 dataset taken from 2016 to 2018 using the latest reconstruction and corrections provided by the CMS Collaboration. By making use of the additional knowledge gained on the performance and efficiency of the detector, more accurate results are expected. Furthermore, the discrepancies of the cross-section between the 2016 and 2017 data taking periods observed in [51] are revisited and reevaluated.

In the following sections, the analysis procedure is described in detail, starting with the used observables and the analyzed phase space in section 4.1. An overview of the datasets and simulations used for this analysis is given in section 4.2, followed by a detailed description on the event selection in section 4.3. In section 4.4 the applied energy and efficiency corrections are described. To account for effects caused by the detector resolution and efficiency, an unfolding procedure, as outlined in section 4.5, is applied to the measurement. Multiple sources of uncertainties on the final result have to be taken into account. They are described and presented in section 4.6. In the following section 4.7, the obtained results including the estimated uncertainties are compared to theory predictions from simulations. Finally, in section 4.8, the compatibility between the different data taking periods are evaluated to verify if a combined measurement of the Run 2 data is credible.

4.1. Phase Space Description

Since the goal of this analysis is to perform a differential cross-section measurement of the Z+jet production in order to get access to the underlying PDFs and the scattering described by theory, suitable observables are used. The initial momenta of the partons are responsible for the boost of the scattering process, while the contributions from different production channels to the overall cross-section are strongly dependent on the scattering angle in the center-of-mass system of the interacting partons.

A former measurement of the triple-differential cross-section measurement of the dijet production at a center-of-mass energy $\sqrt{s} = 8\,\text{TeV}$ at CMS [2] introduced the variables y_b and y^*. There, the rapidity variables y_b and y^* represent the boost of the center-of-mass system of the dijet system and half the rapidity separation between the two jets, respectively. The benefit of using those two variables is, on the one hand, that y^* is proportional to the scattering angle in the parton center-of-mass system and therefore provides sensitivity to the flavour composition of the initial state parton luminosities in the matrix element. On the other hand, y_b is proportional to the logarithm $\ln\left(\frac{x_1}{x_2}\right)$ of the two momentum fractions x_1 and x_2 of the initial partons. Therefore, in combination, contributions to the matrix element and the PDFs can be differentiated. This sensitivity is for example beneficial in PDF fits.

In the context of the Z+jet measurement, the definitions of y_b and y^* represent the boost of the Z+jet system and the rapidity separation between the Z boson and the jet. They are specified as:

$$y_b = \frac{1}{2}|y^Z + y^{\text{jet}}| \tag{4.1}$$

and

$$y^* = \frac{1}{2}|y^Z - y^{\text{jet}}|\,. \tag{4.2}$$

Here, y^Z denotes the rapidity of the reconstructed Z boson and y^{jet} is the rapidity of the reconstructed jet with the largest transverse momentum in an event.

In an ideal scenario, the transverse momentum of the Z boson p_T^Z and the corresponding jet p_T^{jet} are perfectly balanced. In reality, this is not always the case. Experimental effects such as pileup and natural effects, for example additional QCD radiation (see subsection 2.1.2), cause additional jets and hadronic activity in the same event. In order to reduce the sensitivity to additional jet contributions, the usage of the transverse momentum of the Z boson p_T^Z is preferred over the transverse momentum of the leading jet or an averaged variable. Particles participating in EW interactions are radiated as well, but the probability is considerable lower than QCD radiation. In the presented analysis, Z boson candidates are reconstructed from the muonic decay channel ($\text{Z} \to \mu\mu$) by adding the four-momenta of two oppositely charged muons in an event. The momentum of the di-muon system can be reconstructed with higher precision than the momenta of jets using the sensitivity of the CMS detector, as described in subsection 3.2.4.

The analyzed phase space covers the rapidity range 0 to 2.5 for both, y_b and y^*. An illustration of the event topology using this binning is given in Figure 4.1. The events in each y_b-y^* bin are additionally binned in p_T^Z in three different schemes due to the limited number of events at high rapidities. This results in a *central*, *edge*, and *extra* binning shown in Table 4.1.

The illustration on the right in Figure 4.1 shows which p_T^Z binning scheme is used for each y_b-y^* bin. On the one hand, the whole event topology is boosted in bins with high y_b, resulting in the Z boson and jet being more collimated in a forward or backward direction. On the other hand, the events in high y^*-bins have a greater scattering angle, resulting in the Z boson and jet to be in a forward-backward orientation. Overall, fewer events are expected in both of these phase space regions. In addition, the reconstruction of objects is generally worse in forward and backward directions than in the central detector coverage. This results in a limited amount of events passing the selection for these phase space regions and, consequently, higher statistical uncertainties. To address this problem, the bin edges have been expanded to allow for more events per bin while simultaneously reducing the

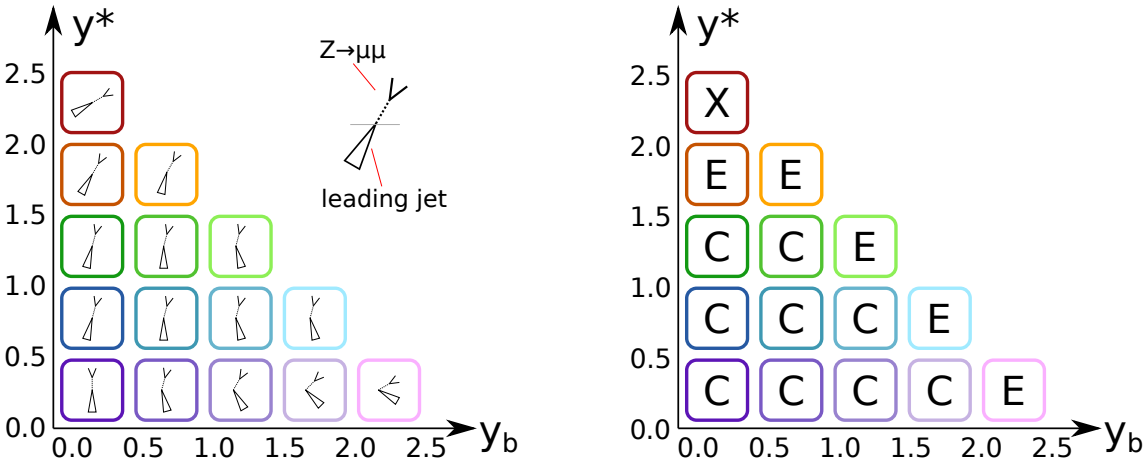


Figure 4.1.: An illustration of the event topology in each y_b-y^* bin (left) and the p_T binning scheme for each y_b-y^* bin (right) is shown. At high values of y_b, the event topology is boosted, resulting in a more collimated Z boson and jet in the forward or backward region. In the high y^* region, the scattering angle is big, resulting in a forward-backward event topology. Overall, fewer events are expected in the forward and backward region. Additionally, the detector reconstruction and resolution is worse in the forward and backward region compared to the central region. Hence, three different p_T^Z binning schemes are used. The central binning (C), the edge binning (E), and an extra binning (X). Illustrations are taken from [51].

Table 4.1.: The different binning schemes for p_T^Z used in three y_b-y^* regions. In the central rapidity regions, a sufficient number of events is observed, allowing a fine central binning. The number of events is limited at the edge of the observed phase space. Therefore, a more coarse edge binning is used. In the high y^*-bin, the number of events is even more limited, resulting in an extra binning scheme.

binning scheme	bin edges of p_T^Z in GeV
central (C)	25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 130, 150, 170, 190, 220, 250, 400, 1000
edge (E)	25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 130, 150, 170, 190, 250, 1000
extra (X)	25, 30, 40, 50, 70, 90, 110, 150, 250

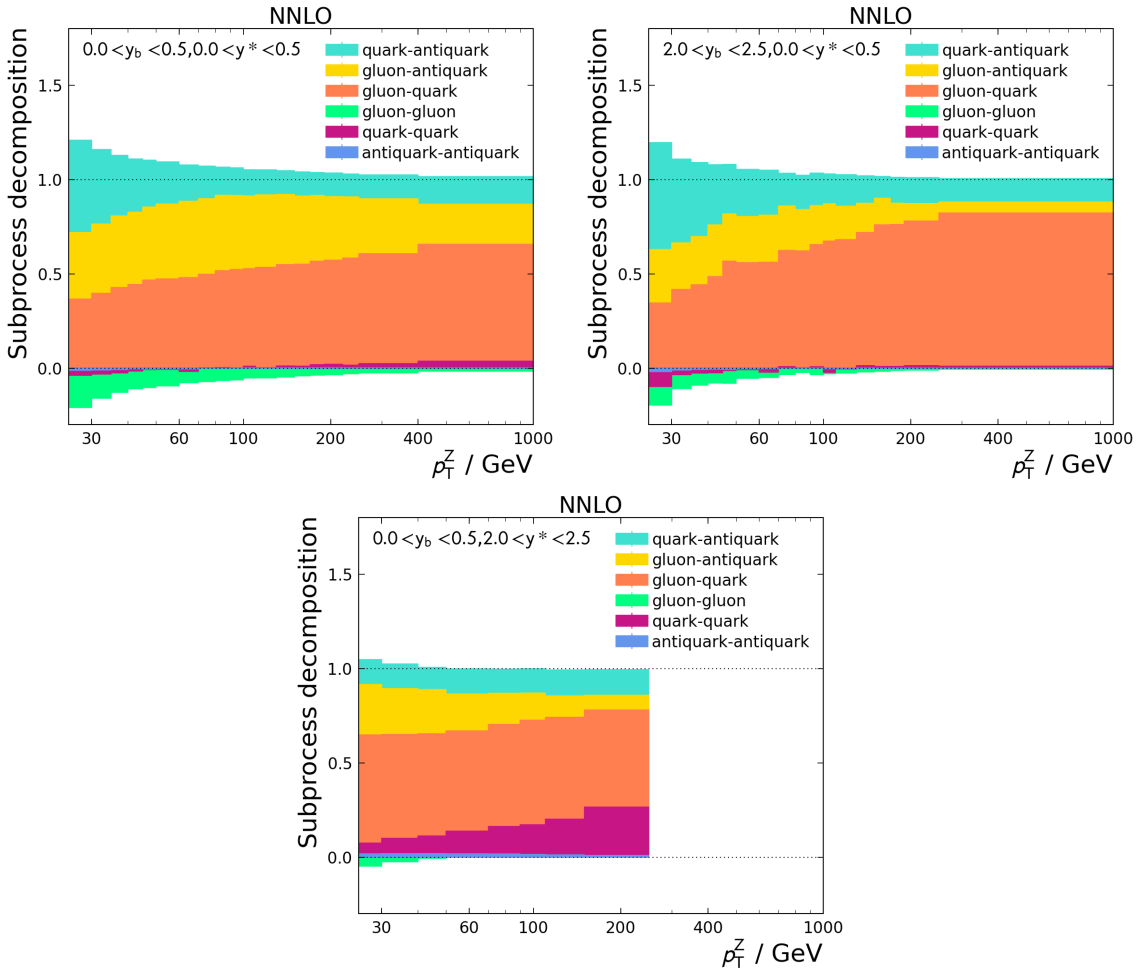


Figure 4.2.: Sub-process contributions calculated at NNLO for the Z+Jet Production in a central (top left), high y_b (top right) and high y^* (bottom) phase space region as a function of p_T^Z. Plots from [50]. Plots for all rapidity bins can be found in Figure A.1.

resolution in p_T^Z for comparisons to theory predictions. A more coarse p_T^Z binning, the so-called edge binning, is sufficient for the most bins at the edge of the analyzed phase space. Due to an even lower event count in the region with $0 < y_b < 0.5$ and $2.0 < y^* < 2.5$, an extra binning is defined, ending at a p_T^Z of 250 GeV. Bins beyond this p_T^Z range result in statistical uncertainties of more than 30 % and are therefore neglected.

The expected contributions of the underlying sub-process to the differential cross-section for NNLO predictions in selected regions of the analyzed phase space are illustrated in Figure 4.2. A full overview of the complete phase is given in Figure A.1.

For the unfolding procedure outlined in section 4.5 and comparisons over the total investigated phase space, a one dimensional representation of the three-dimensional unfolding is necessary. To obtain this, the three-dimensional binning of the analyzed phase space is represented in one dimension by a procedure called *unraveling*. Firstly, the unraveling is performed in the y^* bins, followed by the y_b bins. Lastly, the p_T^Z bins are unraveled, as depicted in Figure 4.3. This results in a one-dimensional chain of 264 bins, and all bins are indexed from 1 to 264.

4.2. Simulation and Datasets

In the presented measurement, events which contain a Z boson candidate and at least one jet are analyzed. A Z boson candidate is reconstructed from two oppositely charged

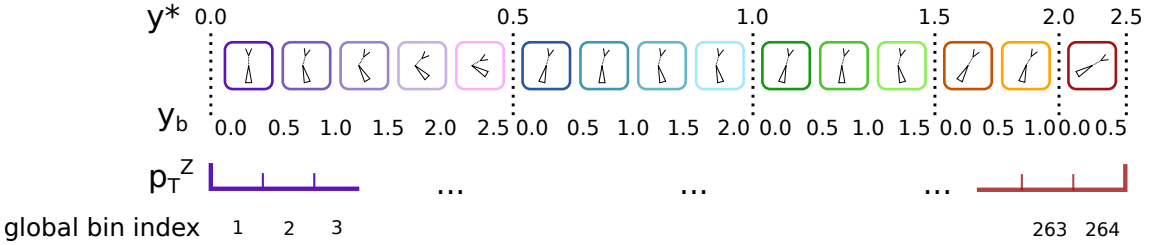


Figure 4.3.: Unraveling of the three-dimensional binning into a one dimensional form. The bins are ordered by stringing them together in one dimension. First in y^*, then y_b and lastly in p_T^Z. This results in a one-dimensional chain of 264 bins. Illustration taken from [51].

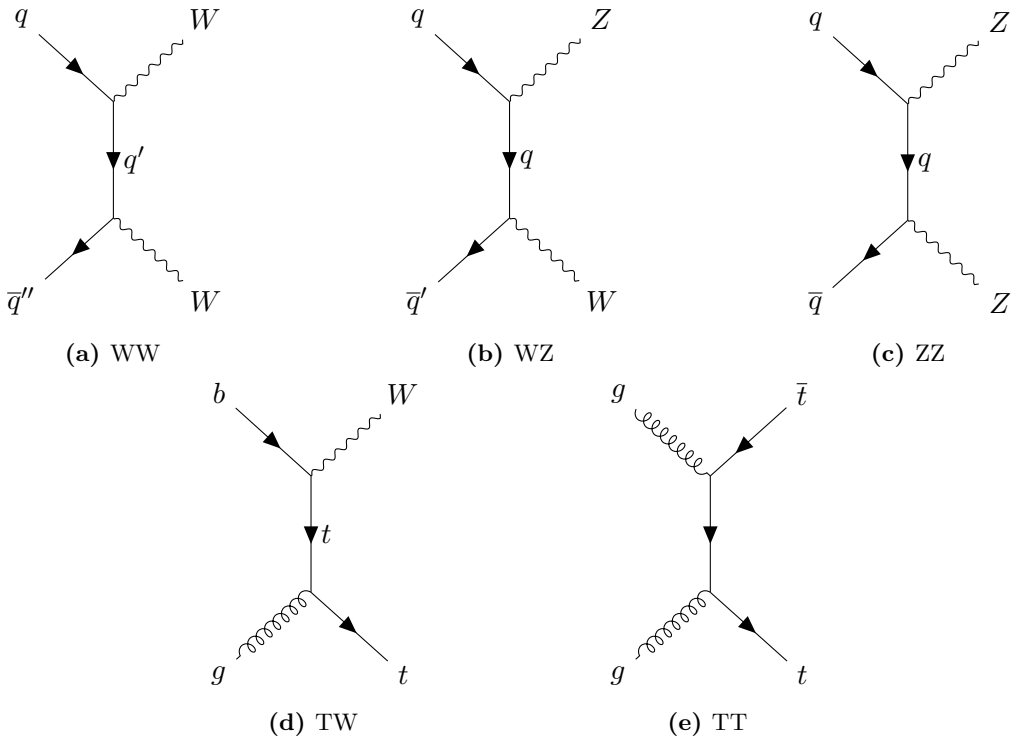
muons originating from the $Z \to \mu^+\mu^-$ decay. All events containing at least one jet and at least two oppositely charged muons creating a Z boson candidate within a mass window of $\pm 20\,\text{GeV}$ around the Z boson mass $m_Z = 91.1876\,\text{GeV}$ [13] are taken into account. Other production channels than the DY+jet process yield a similar signature in the detector and therefore pass the event selection criteria described in section 4.3. These contributions from background processes add to the overall observed number of events during data-taking. When trying to solely measure the cross-section of the signal process, the background contributions have to be taken into account and corrected for in the final result.

Processes to be misidentified as $Z(\to \mu\mu) + \text{jet}$ events are either diboson production of the Z and W bosons or top quark production processes. Selected Feynman diagrams for the production of those background events are shown in Figure 4.4.

The diboson production shown in Figure 4.4 **(a)**, **(b)**, and **(c)** can lead to a miss-identification if two oppositely charged muons are created. W bosons can decay into muon neutrino or tau neutrino pairs. Tau leptons, in turn, can decay leptonically into a tau neutrino, muon, and muon neutrino. This results in possible miss-identification if two oppositely charged W bosons create two oppositely charged muons. Furthermore, Z bosons decaying into two muons where one of the muons is not reconstructed, for example when being outside the detector coverage, can contribute muons to a possible miss-identification. Jets can arise in diboson production from hadronically decaying tau leptons, underlying event contributions, pileup, or higher order production of jets.

Top quarks almost exclusively decay into a b quark and W boson, which in turn can decay into a muon and neutrino. In case of a single top quark production in association with a W boson as shown in Figure 4.4 **(d)**, a Z boson candidate might be reconstructed from the W boson and top quark initiated W boson decaying into muons. In case of the $t\bar{t}$ production shown in Figure 4.4 **(e)**, two oppositely charged W bosons are produced in the decay of the top quarks, which in turn can result in two oppositely charged muons. Jets originate from the b quark decay, pileup, underlying event contributions, or higher order QCD emissions.

In the following, the different data-taking periods of Run 2 and the corresponding luminosities are outlined, followed by an overview of the analyzed datasets. Next, the Monte Carlo (MC) samples used to estimate the signal and background contributions are described. Those MC samples are scaled to their corresponding predicted cross-sections in order to differentiate the amount of signal and background contribution in data. The cross-sections for the scaling of the MC samples are either obtained from the generator of the sample itself or from separate higher-order theory calculations. An overview of the datasets and simulated samples used for this analysis is also found in Table B.1 to B.4.



(a) WW **(b)** WZ **(c)** ZZ **(d)** TW **(e)** TT

Figure 4.4.: Selected Feynman diagrams of background processes that might be identified as $Z(\rightarrow \mu\mu) + \text{jet}$ events. An example Feynman diagram for the diboson production for two W bosons is shown in **(a)**, a W boson and a Z boson in **(b)**, and the diboson production for two Z bosons is shown in **(c)**. An example for the single top production in association with a W boson is shown in **(d)**. Here, the bottom and top quark can be replaced by their anti-particles. The top anti-top production is shown in **(e)**.

4.2.1. Data-Taking Periods

During the 2016 data-taking period, an issue in the silicon strip tracker of the CMS detector was found. Saturation effects in the APV25 (analog pipeline voltage) readout chip [52] caused a very slow discharge and therefore loss of hits under high occupancy in the tracker. The higher-than-anticipated amount of high ionizing particles due to more severe pile-up conditions caused more heat in the tracker than expected. This, in turn, caused a decrease in the draining speed of the readout chip, which resulted in an efficiency loss of up to 10 % in the tracker [53]. After the issue was discovered, the Pre-amplifier Feedback Voltage Bias (VFP) of the readout chips was adapted and the hit efficiency of the tracker returned to more than 99 %.

As a consequence, the data-taking period of 2016 is split into two separate parts with separate MC samples aiming to simulate the found inefficiencies. For the affected dataset, a correction was applied to the offline reconstruction of the data, the so-called high-ionizing particle mitigation (HIPM). Since this mitigation could only be applied for the offline reconstruction, unknown inefficiencies in the HLT might be present for this era.

The period affected by the APV issue before the VFP fix with an amount of data recorded corresponding to $19.52\,\mathrm{fb}^{-1}$ is called 2016preVFP, 2016APV, or 2016HIPM, with all of them used interchangeably. The second part of the 2016 data-taking period with the adapted Preamp Feedback Voltage Bias is referred to as 2016postVFP or 2016nonAPV and accounts for $16.81\,\mathrm{fb}^{-1}$ of data. The total integrated luminosities for the 2017 and 2018 data-taking periods are $41.48\,\mathrm{fb}^{-1}$ and $59.83\,\mathrm{fb}^{-1}$, respectively [54–56].

4.2.2. Datasets

The datasets provided by the CMS Collaboration are classified by certain trigger requirements. For the presented analysis, the Single Muon datasets are used. Here, at least one muon was identified by the trigger algorithms in each event. The datasets have been reconstructed from raw data with improved algorithms compared to the previous analyses [50, 51], also called ReReco or Ultra Legacy datasets. The Run 2 data taking at a center-of-mass energy of 13 TeV is split in four data-taking periods. The presented analysis is performed separately for each of the data-taking periods, 2016preVFP, 2016postVFP, 2017 and 2018.

Currently, so called ReReReco datasets are being produced by the CMS Collaboration including a fix for inefficiencies in the reconstruction of high p_T muons affecting the 2017 and 2018 data-taking periods. At the time of writing this thesis, this reprocessing is still ongoing. Therefore, the ReReco samples are used throughout the presented analysis.

4.2.3. Monte Carlo Simulation

The signal sample for the production of charged lepton pairs associated with partons forming jets is calculated in NLO using `MadGraph5`. The generation is matched to the `Pythia8` [57] parton shower using the `aMC@NLO` method [58] with contributions from higher order real emissions of partons estimated using the FXFX merging method [59]. The parton shower simulation uses the Tune CP5 [60]. The DY cross-section with $m_{\ell\ell} > 50\,\mathrm{GeV}$ of $\sigma_{\mathrm{DY}} = 6077.22\,\mathrm{pb} \pm 2\,\%$ is calculated at NNLO QCD and NLO EW with `FEWZ` v3.1.b2 [61–64] making use of the NNPDF 3.1 PDF set at NNLO [65]. Since the presented analysis requires at least one jet to be produced in association with the Z boson, the predicted cross-section has only NLO accuracy for the analyzed phase space. This difference results in a worse agreement between simulation and data compared to the inclusive Z boson production.

Table 4.2.: Cross-sections for the different DY+Jet sub-processes. Due to limited statistical precision in the calculation, a difference of 0.5 % between the inclusive cross-section and summed cross-sections of the jet-binned samples is observed.

Sample	Cross-section (pb)	Fraction
DY inclusive	6427±13	1.000
DY+0 jets	5138±9	0.799
DY+1 jet	958.6 ± 4.9	0.149
DY+2 jets	358.8 ± 3.8	0.056

In addition to the inclusive signal sample, statistically independent samples at NLO precision for DY+0 jets, 1 jet, and 2 jets are provided by the CMS collaboration. Those samples can be combined with the inclusive sample in order to reduce the statistical uncertainty of the simulation. Using the cross-sections calculated directly from the MC production for the jet-binned samples σ_j and the inclusive sample σ_{inc}, the fractions for combining the jet-binned samples are obtained from the ratio $\sigma_j/\sigma_{\text{inc}}$. The cross-sections for the different sub-process obtained from the inclusive signal sample and the resulting weights are found in Table 4.2. Due to statistical fluctuations, a difference of about 0.5 % between the inclusive and summed cross-sections of the jet-binned samples is observed. This difference is negligible compared to the uncertainty of 2 % for the inclusive cross-section obtained from independent higher order theory calculations.

Using the fractions in Table 4.2, a combination of the jet-binned samples into a single sample is performed. Merging the inclusive sample with the jet binned samples is done via the effective number of events. For the combined jet-binned samples, the effective number of events is calculated as

$$N_{\text{exc}} = \sum_{j=0}^{2} \frac{\sigma_j}{\sigma_{\text{inc}}} N_j \,. \tag{4.3}$$

Here, σ_{inc} is the total cross-section obtained from the inclusive DY sample, while N_j and σ_j are the number of events and the cross-section of the exclusive DY samples with j jets. The inclusive sample is then weighted with the fraction $N_{\text{inc}}/(N_{\text{inc}} + N_{\text{exc}})$, while the combination of the jet binned samples is weighted with $N_{\text{exc}}/(N_{\text{inc}} + N_{\text{exc}})$.

The production of $t\bar{t}$ events is simulated using `POWHEG` [66] for heavy quark pair production [67]. The NLO matrix element calculation is matched to the `Pythia8` parton shower simulation with Tune CP5. The cross-section for the $t\bar{t}$ production is $\sigma_{t\bar{t}} = 831.76^{+4.8\,\%}_{-6.1\,\%}$ pb as calculated with the `Top++2.0` program to NNLO accuracy in perturbative QCD including soft-gluon resummation to next-to-next-to-leading-logarithmic (NNLL) accuracy [68] and assuming a top-quark mass $m_t = 172.5$ GeV. The total uncertainty is combined from the independent variation of the factorization and renormalization scales, m_F and m_R, and the variations in the PDF and α_S. The PDF and α_S variations are following the PDF4LHC prescription with the MSTW2008 68 % CL NNLO, CT10 NNLO and NNPDF2.3 5f FFN PDF sets [69–72]. Since the MC sample only contains final states where both W bosons decay into a lepton and a neutrino, the value is scaled with the branching ratio $BR(W \to \ell\nu) = 3 \times 0.1086$ [13] resulting in a total cross-section of $\sigma_{t\bar{t}\to 2\ell 2\nu} = 88.29^{+4.8\,\%}_{-6.1\,\%}$ pb.

The single top quark and antiquark production in association with a W boson is performed with `POWHEG` and the corresponding NLO process [73] with `Pythia8` as parton shower generator using Tune CP5. The cross-section for this process, including top and anti-top production, is calculated at NNLO with the procedure outlined in references [74, 75] for 13 TeV as $\sigma_{\text{tW}} = (71.1 \pm 3.8)$ pb. Additionally, the single top quark and antiquark

production via the t-channel is simulated with POWHEG at NLO [76]. Here, Pythia8 with Tune CP5 is used again for the parton shower simulation. The cross-section is calculated at NLO QCD using Hathor v2.1 [77, 78] for a top quark mass of $m_t = 172.5\,\text{GeV}$. The total uncertainty is the quadratic sum of the scale, the PDF, and the α_S uncertainties. They are calculated using the PDF4LHC prescription [69] with the MSTW2008 68 % CL NLO [70, 79], CT10 NLO [80], and NNPDF2.3 [72] PDF sets. The total cross-section for the top and anti-top production in the t-channel are $\sigma_t = 136.0^{+5.4}_{-4.6}\,\text{pb}$ and $\sigma_{\bar{t}} = 81.0^{+4.1}_{-3.6}\,\text{pb}$ respectively.

The diboson background processes for WW, WZ, and ZZ production are simulated in LO using Pythia8. The cross-section for the WW production at NNLO QCD is calculated in ref. [81] to be $\sigma_{\text{WW}} = 118.7^{+2.5\,\%}_{-2.2\,\%}\,\text{pb}$. The cross-sections for the WZ and ZZ production are calculated from the MC samples and found to be $\sigma_{\text{WZ}} = (27.60 \pm 0.40)\,\text{pb}$ and $\sigma_{\text{ZZ}} = (12.170 \pm 0.020)\,\text{pb}$, only taking the statistical uncertainty of the MC sample into account.

All MC simulations use Geant4 [82] for the detector simulation. After the detector simulation and digitization, the same reconstruction algorithms as used for data are applied.

4.3. Event Selection

Objects in the CMS detector are reconstructed using the PF algorithm, as described in subsection 3.2.4. The observables necessary for this analysis are described in section 4.1. In order to access these observables, one Z boson and at least one jet have to be reconstructed in a given event. With the aim of selecting events that mainly stem from the Z+jet production, the selections presented in the following paragraph need to be passed.

The background processes are contributing events that pass the selection criteria, but do not originate from Z+jet production. They are listed and explained in section 4.2.

4.3.1. Trigger

The first selection, even before storing the collision data, is based on a trigger path of L1 trigger and HLT. The CMS trigger system is briefly described in subsection 3.2.2.

Since two muons are needed to reconstruct a Z boson from its muonic decay, the events are required to contain at least one muon for further processing. Ideally, the muon is also isolated compared to other particles so that muons originating from weak decays within jets are suppressed. CMS provides a so-called IsoMu trigger that is fired for an event with a single isolated muon above a certain p_T threshold. This threshold differs between the Run 2 data-taking periods.

For the 2016 data-taking period, a muon is required to either fire the IsoMu24 or the IsoTkMu24 trigger path. The former triggers on global muons with $p_T > 24\,\text{GeV}$, while the latter triggers on tracker muons with the same p_T threshold. During the 2017 data-taking period, the IsoMu24 trigger was prescaled for a short period of time. Therefore, the IsoMu27 trigger, which was unprescaled, is used for the whole data-taking period. This trigger is similar to the IsoMu24 trigger, except that it requires a muon with a transverse momentum higher than 27 GeV. For this trigger path, the CMS Collaboration provides centrally produced efficiency corrections described in subsection 4.4.2. For the 2018 data-taking period, the IsoMu24 trigger was unprescaled for the whole year, and its usage is recommended by the CMS Collaboration.

Table 4.3.: Criteria for a muon to pass the tight identification

χ^2/ndf of the global-muon track fit	< 10
Number of hits in the muon system	≥ 2
Number of hits in the muon system included in the global muon track fit	≥ 1
Number of hits in the pixel tracker	≥ 1
Number of hits in the pixel or strip tracker	≥ 6
Transverse distance of the track to the primary vertex d_{xy}	$< 0.2\,\text{cm}$
Longitudinal distance of the track to the primary vertex d_z	$< 0.5\,\text{cm}$

4.3.2. Muons

Global muon candidates from the PF reconstruction (see subsection 3.2.4) are used. Additionally, final state radiation of photons γ is taken into account by adding the four-momenta of photons within a cone of $\Delta R(\mu, \gamma) < 0.1$ (see Equation 4.4) to the muon four-momenta.

As described in subsection 3.2.2, the muon system of the CMS detector stretches up to a pseudorapidity of $\eta = 2.4$. Therefore, all muons are required to have a pseudorapidity of less than 2.4 in order to ensure a muon reconstruction with the highest quality and sensitivity.

The lowest muon p_T requirement of the recommended isolated muon HLT during the 2017 data-taking period is 27 GeV. Efficiency corrections for the trigger are provided by the CMS Collaboration starting from 29 GeV. For the runs in 2016 and 2018, the lowest unprescaled isolated muon HLT was triggering on muons with a transverse momentum above 24 GeV. In order to compare the measurement between different data taking periods, muons are required to have a transverse momentum of at least 29 GeV, following the tightest thresholds dictated by the ones during 2017 data-taking.

Additional criteria are applied to the muons to avoid miss-identification, for example, from charged hadrons. Muon identification working points are defined using multiple variables based on the muon reconstruction. In this analysis, the tight working points are used, aiming to suppress muons from in-flight decays and hadronic punch-trough to the muon chambers. To pass the tight working point, a muon is required to be selected by the PF algorithm, being either a tracker or a global muon, and pass the criteria given in Table 4.3 [42].

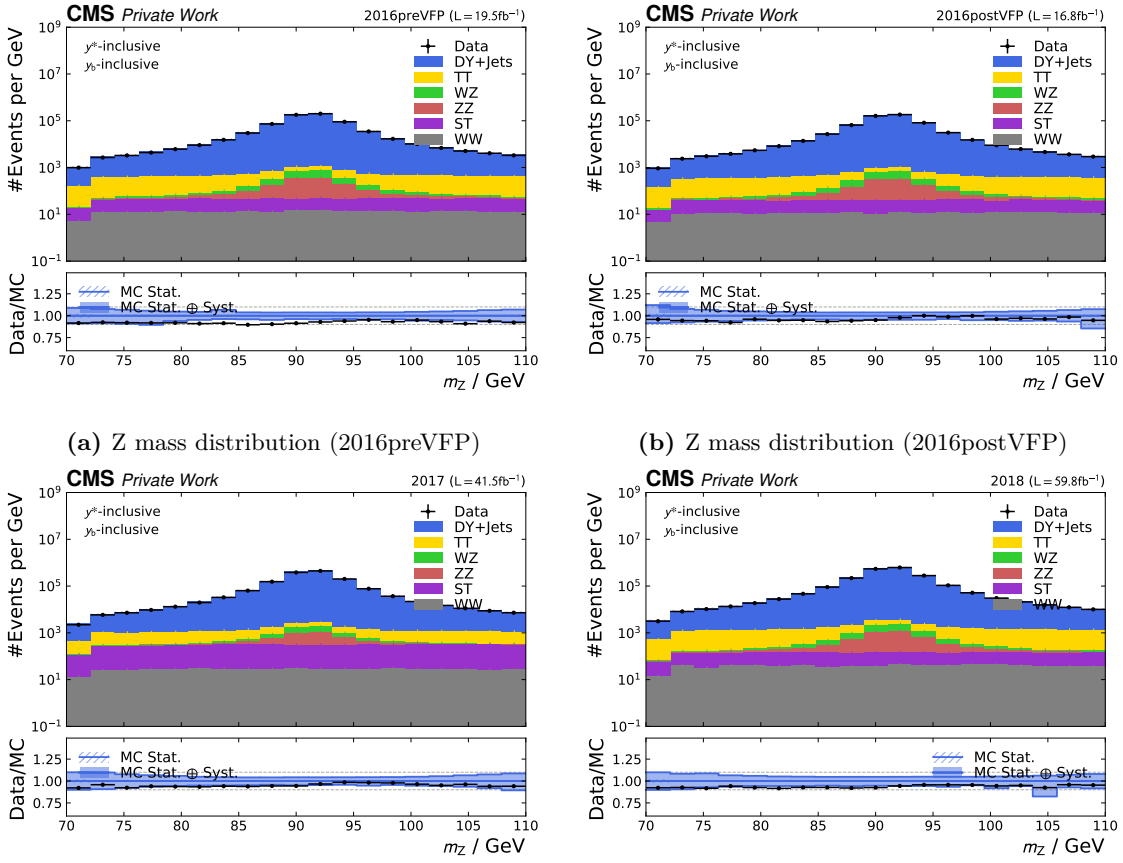
Moreover, muons originating from weak decays within jets are suppressed by requiring the muons to be isolated from other particles. The PF isolation is defined as the muon p_T relative to the sum of the p_T of charged and neutral hadrons in a cone of

$$\Delta R = \sqrt{(\Delta\Phi)^2 + (\Delta\eta)^2} \tag{4.4}$$

around the muon. The contribution from pileup of neutral particles is estimated from simulations and subtracted from the total p_T. For this analysis, the muons are required to pass the tight working point. It is defined with an efficiency of 95 %, meaning that 95 % of muons not originating from weak decays within jets pass the selection. This results in the PF isolation in a cone of $\Delta R < 0.4$ to be less than 15 %. [42]

4.3.3. Z Boson

Z bosons decay into two muons of opposite charge, as explained in subsection 2.1.1. By adding the four-momenta of two muons, a Z boson candidate can be reconstructed. If the difference of the resulting dimuon mass $m^{\mu\mu}$ to the world average for the Z boson



(a) Z mass distribution (2016preVFP)

(b) Z mass distribution (2016postVFP)

(c) Z mass distribution (2017)

(d) Z mass distribution (2018)

Figure 4.5.: Mass distributions of the reconstructed Z boson for all data-taking periods of Run 2. The shape of the distributions is well described by the simulations after the application of all corrections. The normalization however is off by less than 10 %.

$m^Z_{\text{PDG}} = (91.1876 \pm 0.0021)\,\text{GeV}$ [13] exceeds 20 GeV, the event is vetoed. If more than two muons are found in an event and multiple Z boson candidates can be reconstructed, the candidate with a mass closest to m^Z_{PDG} is chosen. After a Z boson candidate is reconstructed, its transverse momentum p^Z_T is required to be greater than 25 GeV for the event to pass the selection.

In Figure 4.5, the data and MC distributions for the mass of the reconstructed Z boson are shown after selections for each data-taking period. The shape of the data distribution is well described by the simulated samples. A slight overall normalization offset of less than 10 % for each data-taking period is observed. This offset might be introduced by the theory prediction of the DY cross-section. It is calculated at NNLO for the inclusive Z boson production, while the analyzed phase space requires at least one jet, leading to only NLO accuracy for the analyzed phase space.

4.3.4. Jets

In order to reject jets originating from pileup, CHS is used before starting the jet clustering as described in subsection 3.2.4. The jets used in this analysis are clustered from the PF candidates using the anti-k_T algorithm with the distance parameter $R = 0.4$ [24]. Jets reconstructed using these criteria are also referred to as AK4PFCHS jets in the following. Since reconstructed jets originating from pileup interactions typically carry low energy, the transverse momentum of a jet is required to be greater than 20 GeV.

Table 4.4.: Tight identification criteria, including the lepton veto for an AK4PFCHS jet with $|\eta| < 2.4$ in 2016 or $|\eta| < 2.6$ in 2017 and 2018.

Neutral Hadron Fraction	< 0.9
Neutral EM Fraction	< 0.9
Number of Constituents	> 1
Muon Fraction	< 0.8
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.8

Badly reconstructed or noisy jets are additionally rejected via a jet identification, taking the jet constituent energy fractions of the PF candidates and the respective PF candidate multiplicities into account [83]. For example, the charged hadron fraction is the energy fractions of the jet constituents identified as charged hadrons by the PF algorithm compared to the total jet energy and the charged multiplicity is the amount of charged particles in the jet constituents. The tight working points including the lepton veto with the criteria given in Table 4.4 for the data taking periods from 2016 to 2018 are used. The efficiency of this identification is more than 98 % to 99 % for all eta regions, while the background rejection is higher than 98 % for $|\eta| < 3$.

Furthermore, pileup jets are rejected by using a pileup identification (puID) discriminator [84]. This puID is derived by training a boosted decision tree on multiple variables describing the affected jet and corresponding global event. Since the boosted decision tree was only trained with jets that have a transverse momentum of less than 50 GeV, the puID is only applied to jets below this threshold. The tight puID working point corresponding to an efficiency of 80 % in the central region with $|\eta| < 2.5$ is used in this analysis.

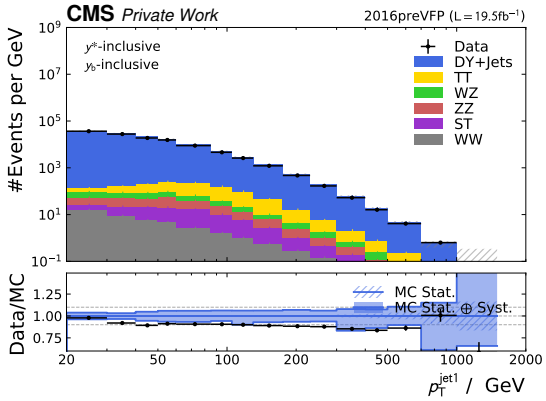
The detector simulation is never able to depict the state of the CMS experiment perfectly and, therefore, can not perfectly describe the interactions of the detector components with the collision products, as they appear during data taking. Hence, additional selection criteria are applied to each jet to remove jets that are potentially dominated by anomalous contributions from various sub-detector components or reconstruction failures. Those so-called jet veto maps for certain η-ϕ regions, where anomalous behavior has been observed, are centrally provided by the CMS Collaboration and applied to simulation and data. The inefficiencies in detector acceptance caused by this veto maps are mitigated using the unfolding method described in section 4.5.

Figure 4.6 shows the data and MC distributions for the transverse momentum of the leading jet for all data-taking periods. While the statistical uncertainties on the data are shown, the simulations also include the systematic uncertainties. The combination of statistical and systematic uncertainties for the simulation are shown as a blue band in the ratio plots below the histograms.

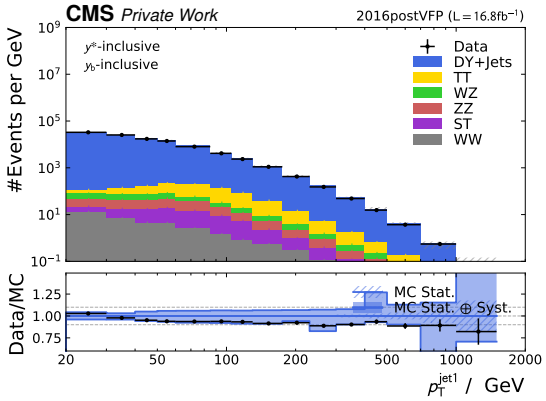
An overall good agreement between data and simulation within uncertainties is observed for each data-taking period. A constant offset in the normalization of less than 10 % can be seen. This offset at the limit of the 1σ deviation of the uncertainties might be caused by the fact that the cross-section prediction for the Z+1 jet process has only NLO accuracy, as already discussed in subsection 4.2.3.

4.3.5. Software Framework Validation

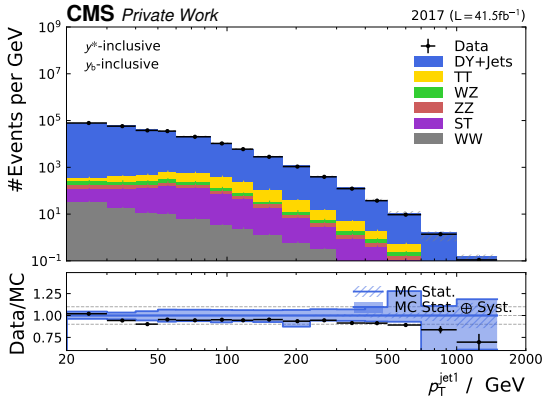
In order to minimize the effect by errors and bugs in the software used to analyze the data, the event selection produced by the software framework was verified against an independent



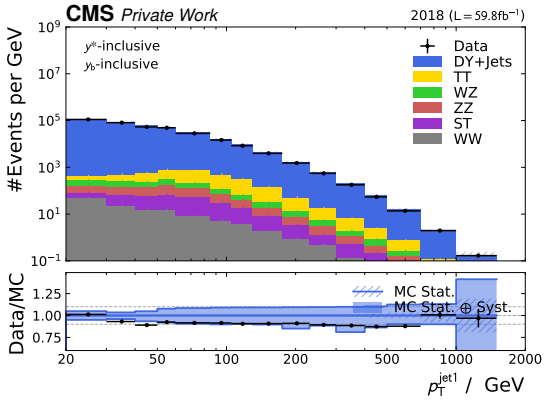
(a) Distribution of the leading jet transverse momentum for the 2016preVFP data-taking.



(b) Distribution of the leading jet transverse momentum for the 2016postVFP data-taking.



(c) Distribution of the leading jet transverse momentum for the 2017 data-taking.



(d) Distribution of the leading jet transverse momentum for the 2018 data-taking.

Figure 4.6.: Distributions of the leading jet transverse momentum for all Run 2 data-taking periods in data and simulations. An overall good agreement within uncertainties between the data and simulations is observed. However, a constant offset of less than 10 % between data and simulations in the overall normalization can be seen.

Table 4.5.: Event selection for the software framework cross-check between ULB and KIT.

Muon property	Selection
Trigger	HLT_IsoMu24
Leading $p_T^{\mu_1}$	$> 25\,\text{GeV}$
Sub-leading $p_T^{\mu_2}$	$> 20\,\text{GeV}$
ID	medium
ISO	tight
Z boson mass m_Z	$91 \pm 15\,\text{GeV}$

Jet Property	Selection
Type	AK4PF with CHS
p_T^{jet}	$> 30\,\text{GeV}$
η_{jet}	< 2.4
ID	medium
puID MVA	> -0.2
$\Delta R(\mu, \text{jet})$	> 0.4

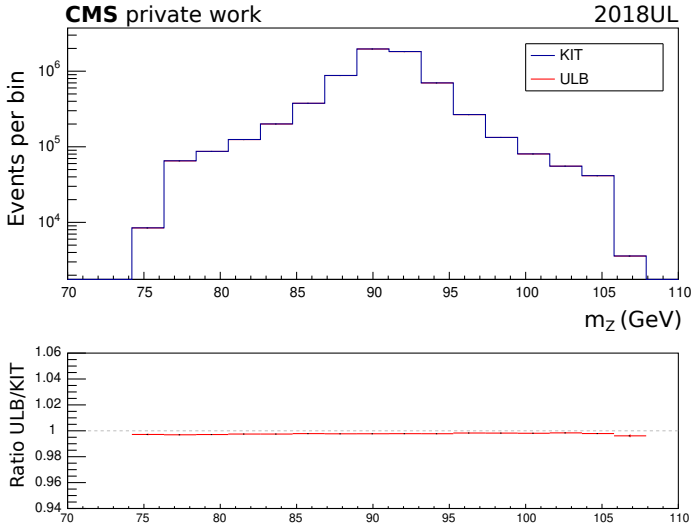


Figure 4.7.: Comparison of the Z boson mass distribution for 2018 data between the software framework used in this analysis at Karlsruhe Institute of Technology (KIT) and an independent framework used at Université Libre de Bruxelles (ULB). No significant deviation is observed.

Z boson analysis from a group at Université Libre de Bruxelles (ULB). After setting a common selection given in Table 4.5 the frameworks produced a nearly identical result as seen in Figure 4.7. Here, the mass distribution of the reconstructed Z boson with the given selection is shown. The remaining difference of less than 0.5 %, which is significantly smaller than the uncertainties, can be attributed to the differences in numerical precision between the input data formats. This analysis uses the MiniAOD data format centrally provided by the CMS Collaboration while the analysis at ULB uses the slimmed down NanoAOD data format centrally provided by the CMS Collaboration.

4.4. Energy and Efficiency Corrections

This section describes additional energy and efficiency corrections needed to improve the agreement between data and simulation, as the detector simulation used when producing the MC samples is never perfect.

Energy corrections are typically applied in both, data and MC, to improve the agreement between the true and reconstructed energy of an object. The energy resolution in MC is chosen to be slightly better than in data in order not to have a worse performance in simulation and therefore unnecessarily reducing the best possible resolution when comparing to data. Tho mitigate this effect, the resolution in data and MC is estimated and the

energy of reconstructed objects in simulation is adapted to match the one observed in data. The energy corrections and resolution smearing methods are described in subsection 4.4.1.

Additional efficiency corrections, described in subsection 4.4.2, are applied to the reconstructed events obtained from MC. Those corrections for example include the different efficiencies when applying identification criteria in MC and data.

4.4.1. Energy Corrections

The muon momentum resolution and scale is corrected for multiple effects such as detector misalignment, bias due to the software reconstruction, or uncertainties in the magnetic field [85]. Correction factors are obtained by analyzing $Z \to \mu\mu$ events in bins of charge, η, and ϕ for data and MC and then comparing the reconstructed Z boson to its precisely known mass. The derived corrections are provided by the CMS Collaboration and are applied to both simulation and data in the presented analysis.

As described in subsection 3.2.4, the momentum for each jet is corrected for the difference in true and reconstructed momentum obtained from simulation. Additionally, residual corrections accounting for remaining effects not fully modelled by simulation are applied to data calculated from in-situ measurements outlined in ref. [49]. These corrections are called jet energy correction (JEC).

Furthermore, the jet energy resolution (JER) obtained from simulation is better than the resolution observed in data. This difference between simulation and data can have a significant impact on the observables sensitive to the jet energy, especially when trying to correct for the detector effects due to the limited resolution, as described in section 4.5. Therefore, the resolution in simulation is corrected to reproduce the one observed in data.

If a reconstructed jet can be matched to a generator level jet, the four-momentum of the reconstructed jet is scaled with the factor

$$c_{\mathrm{JER}} = 1 + (s_{\mathrm{JER}} - 1)\frac{p_T^{\mathrm{reco}} - p_T^{\mathrm{gen}}}{p_T^{\mathrm{reco}}}\,. \tag{4.5}$$

The transverse momentum of the reconstructed jet is denoted as p_T^{reco} and the transverse momentum for the jet at generator level being p_T^{gen}. The data-to-simulation resolution scale factor s_{JER} is provided by the CMS Collaboration. It is derived by comparing the jet resolution obtained from simulation and the measured resolution in data after applying JEC factors [49].

The requirements for a reconstructed jet to be matched to a generator jet are

$$\Delta R < \frac{R_{\mathrm{cone}}}{2},\ |p_T^{\mathrm{reco}} - p_T^{\mathrm{gen}}| < \sigma_{\mathrm{JER}} p_T^{\mathrm{reco}}\,. \tag{4.6}$$

Here, ΔR (see Equation 4.4) is the distance between the jets at reconstruction and generator level and σ_{JER} is the relative p_T resolution obtained from simulation.

If no jet at generator level can be matched to the reconstructed jet, a stochastic smearing method is used. The scaling factor for the jet four-momentum in the smearing method is calculated as

$$c_{\mathrm{JER}} = 1 + \mathcal{N}(0, \sigma_{\mathrm{JER}})\sqrt{\max\left(s_{\mathrm{JER}}^2 - 1, 0\right)}\,. \tag{4.7}$$

Here, $\mathcal{N}(0, \sigma_{\mathrm{JER}})$ denotes a random number drawn from a Gaussian distribution.

4.4.2. Efficiency Corrections

The selections described above are not always fully efficient or have different efficiencies in data and simulation. Consequently, efficiency corrections need to be applied to the simulation to account for those differences.

Trigger efficiencies for the `IsoMu` triggers are centrally provided by the CMS Collaboration for data and simulation. Since scale factors are only applied to the simulation, the ratio of the efficiency in data over the efficiency in simulation is used as a scale factor ϵ, which is multiplied to the nominal weight in simulation:

$$\epsilon(\mu_i) = \frac{\epsilon^{\text{Data}}(\mu_i)}{\epsilon^{\text{MC}}(\mu_i)} \,. \tag{4.8}$$

As it is sufficient for only one muon in an event to fire the trigger, the correlations between the two muons used to reconstruct the Z boson need to be considered. This results in a total scale factor for the trigger of

$$\epsilon_{\text{Trigger}} = 1 - ((1 - \epsilon(\mu_1)) \cdot (1 - \epsilon(\mu_2))) \,, \tag{4.9}$$

where $\epsilon(\mu_1)$ denotes the trigger scale factor for the first muon and $\epsilon(\mu_2)$ for the second muon.

Additional scale factors for the muon reconstruction are provided. Namely, the reconstruction efficiency for tracker muons ϵ_{Reco}, the efficiency for the muon ID given a tracker muons was reconstructed ϵ_{ID}, and the scale factors for these muons' isolation criteria ϵ_{Iso}. Each of the scale factors are treated as fully correlated between the two muons constructing a Z boson candidate, meaning that the total weight for each scale factor is the product of the scale factors of the two muons. This results in a total weight applied to each event based on all muon scale factors of

$$w = \epsilon_{\text{Trigger}} \sum_{i=1,2} \epsilon_{\text{Reco}}(\mu_i)\epsilon_{\text{ID}}(\mu_i)\epsilon_{\text{Iso}}(\mu_i) \,. \tag{4.10}$$

Here, the index i refers to each of the muons of the Z boson candidate and ϵ to the respective scale factors of the muon reconstruction and trigger.

In the 2016 and 2017 data-taking periods, an additional inefficiency of the ECAL L1 trigger in the region at $|\eta| > 2.0$ caused by a gradual timing shift was observed. This timing shift mistakenly caused high η trigger primitives to be associated with the previous bunch crossing. The L1 trigger could therefore fire in two consecutive bunch crossings, once for the high η falsely associated to the previous bunch crossing and once for the actual bunch crossing where energy depositions in other regions cause the trigger to fire as well. Since the L1 trigger primitives do not allow two consecutive events to fire the trigger, an inefficiency is introduced. Therefore, all events containing an electron with p_T larger than $\approx 50\,\text{GeV}$ or a jet with $p_T \gtrapprox 100\,\text{GeV}$ suffer an efficiency loss of about ≈10–20%, highly dependent on time, p_T and η. Correction factors for the so called ECAL pre-firing were computed from unprefirable events in data and are provided by the CMS Collaboration.

Additionally, a similar effect is present in the muon system. Here a muon candidate can be assigned to the wrong bunch crossing due to the limited time resolution of the muon detectors and the event vetoes itself due to the same reason given above. This prefire issue for muons is most pronounced in the data taken during 2016, but also affects the data-taking during 2017 and 2018. The muon prefiring rate is stable for muons above $p_T > 25\,\text{GeV}$ for the whole eta range and varies between 0 % and 3 %. Here, correction factors are calculated from unprefirable events in data and provided by the CMS Collaboration as well.

The total scaling factor for the loss of detector acceptance due to prefiring is obtained as the product of the non-prefiring probability of all objects while accounting for overlapping objects in the ECAL:

$$w = 1 - P\,(\text{prefiring}) = \prod_{i=\text{photons,jets,muons}} \left(1 - \epsilon_i^{\text{pref}}(\eta, p_T^i)\right) . \tag{4.11}$$

Here, $P\,(\text{prefiring})$ refers to the probability of an event to be vetoed by this effect while ϵ_i^{pref} denotes the probability for a single particle i causing the prefiring.

The selection based on the jet identification criteria has an efficiency of more than 98 % to 99 %, as described in subsection 4.3.4. However, no efficiency corrections are required, as this effect is negligible compared to the other inefficiencies and uncertainties described in section 4.6.

In contrast, the puID for jets has an efficiency of about 80 % in the central region. Here, scale factors for simulation accounting for the inefficiencies in data and MC are provided by the CMS Collaboration for AK4PFCHS jets that pass the puID and can be geometrically matched to a jet at generator level within a distance of $\Delta R < 0.4$.

The product of all efficiency corrections is applied as a weight on a per-event basis to the reconstruction obtained from simulation. Furthermore, the uncertainties on those efficiency corrections are propagated through the analysis and their effect is evaluated in section 4.6.

4.5. Unfolding

To allow for a direct comparison to theory predictions, the experimental measurements have to be corrected for detector effects. Generally, the reduced detector acceptance caused by inefficiencies, dead zones or simple geometrical coverage for each subsystem has to be corrected. In differential measurements, additionally, migration of events between bins in the measured phase space occur due to the limited resolution of the detector. Those detector effects can only be estimated by using simulations, which in turn relies in part on random distributions. Therefore, the estimated difference between reconstructed and true values is affected by statistical uncertainties and cannot be corrected for on a per-event basis.

The effects introduced by the detector acceptance and resolution are estimated using detector simulation. In the following, the true variables obtained from the MC simulations of the matrix element are referred to as generator-level, or short generator (gen) observables. After the generator-level, particles are propagated through the detector simulation and reconstruction algorithms. The detector-level or reconstructed (reco) variables are obtained. In a discretized phase space, the transition from gen level to reco level can be represented as a linear operation:

$$\tilde{y}_i = \sum_{j=1}^{m} A_{ij}\tilde{x}_j + b_i, \text{ with } 1 \leq i \leq n\,. \tag{4.12}$$

Here, x_j represents the true distribution with m bins of the discretized generator level phase space and y_i is the average-expected distribution at detector level in n bins of the discretized phase space at reconstruction level. The observed event counts $\tilde{y}_i$ may differ from the average expected event counts due to statistical fluctuations. The response matrix A_{ij} describes the migrations between the bins at gen and reco level, as well as the acceptance of the detector in different regions of the phase space. The background contribution b_i are events from processes other than the signal process.

Without taking the statistical fluctuations into account, unfolding can be understood as a simple matrix inversion of A, in order to reconstruct the gen level x from the reco level y after subtracting the background from the observations. Statistical fluctuations in the observed spectrum can be amplified depending on the condition of the linear operation. In the case of badly conditioned matrices, this results in an ill-posed inversion problem. This effect can be mitigated by using so-called regularization methods. Regularization is unnecessary if the response matrix A is sufficiently diagonal, which means that migrations mostly appear between neighboring bins in the phase space with no huge statistical fluctuations. With the condition number, a proxy that estimates how much the result of the operation changes for small fluctuations in the input, defined as the ratio of the largest and smallest eigenvalue of a matrix, the need for regularization can be assessed. The CMS Collaboration recommends to not use regularization for unfolding if the condition number is smaller than 10.

In the presented analysis, the software package `TUnfold` [86] is used to perform the unfolding. The response matrices for the different data taking periods are presented before the unfolded results and the experimental uncertainties are shown.

4.5.1. Response Matrices

It is necessary for the unfolding algorithm to know how the detector changes the distributions from generator to reconstruction level. Those effects are described by the response matrix, which is a two-dimensional histogram filled with events from the MC simulation. Events that pass the selection criteria at gen and reco level are filled according to their selected bin at generator level on the x-axis and their selected bin at reconstruction level on the y-axis. Each column is normalized to the total number of events in the corresponding reco bin. The response matrices are created using the bin unraveling scheme described in section 4.1 resulting in a 264×264 matrix.

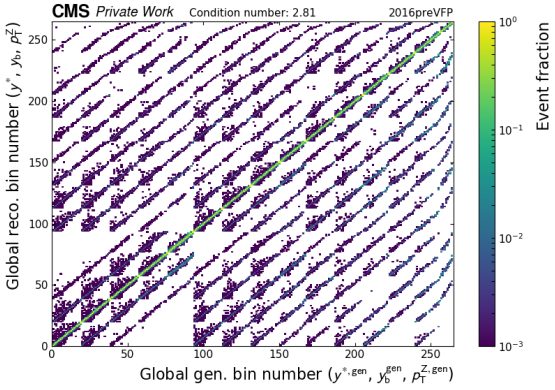
In order to increase the amount of events and, therefore, decrease the statistical uncertainty in the high y_b and y^* regions, a method called forward smearing was applied in the previous analyses [50, 51]. There, the detector resolution and acceptance of y^Z, y^{jet1}, and p_T^Z have been studied from simulations and fitted to functions aimed to describe their p_T^Z dependence. Those parametrizations are then used to generate a large amount of pseudo-events, increasing the number of events in the sparsely populated phase space regions. However, the possibility for a bias introduced by choosing the functional form of those fitted functions was never assessed.

In contrast to the previous analysis, no forward smearing was used because a sufficient number of events in the high y_b and y^* regions are provided by combining the jet binned DY MC samples with the inclusive DY MC sample, as described in subsection 4.2.3.

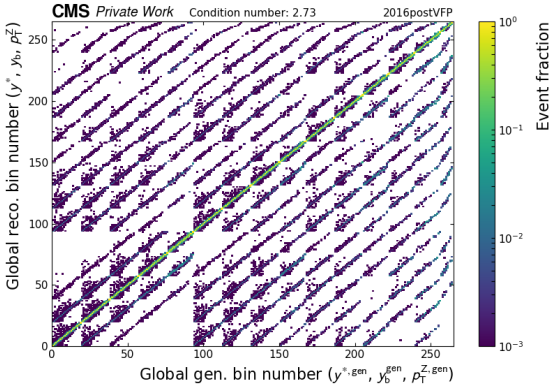
The response matrices for all data-taking periods are depicted in Figure 4.8. Entries on the main diagonal mean that no bin migrations between generator and reconstruction level appear. Since most events are at the main diagonal and migrations mostly appear between neighboring phase space bins, the matrices are well conditioned. All response matrices have a condition number much lower than 10, meaning that no regularization methods have to be used for the unfolding procedure.

4.5.2. Detector Acceptance and Fake Rate

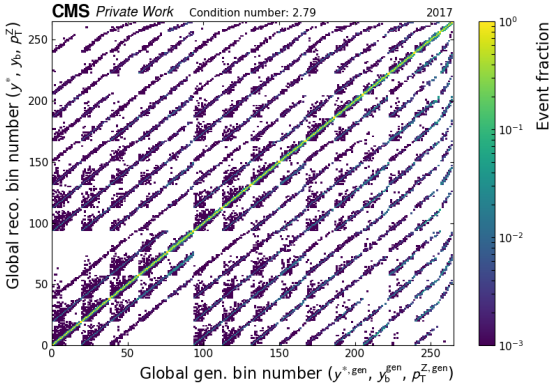
The response matrix is solely filled with events that pass the selection on both, the generator and reconstruction level. Events that are only present in either the generator level or reconstruction level have to be taken into account for the unfolding procedure. If an event is present in the selection at generator level but does not pass the selection at reconstruction



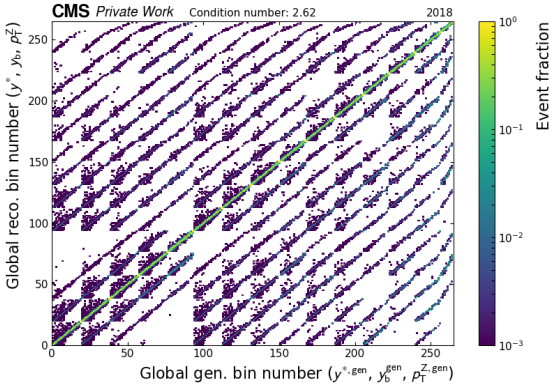
(a) Response Matrix for the 2016preVFP data-taking period.



(b) Response Matrix for the 2016postVFP data-taking period.



(c) Response Matrix for the 2017 data-taking period.



(d) Response Matrix for the 2018 data-taking period.

Figure 4.8.: Response matrices for the different data-taking periods of Run 2. The matrices are filled with events obtained from simulation, with the global bin number at generator level on the x-axis and the according reconstruction bin on the y-axis.

level, it is classified as loss. In contrast, events not present in the selection at generator level but at reconstruction level are classified as fakes.

The loss of events is described by the acceptance A, which is the fraction of events that pass the selections at both, generator and reconstruction level, compared to the number of events at the generator level:

$$A = \frac{N(\text{event in gen bin \& any reco bin})}{N(\text{event in gen bin})}\,. \tag{4.13}$$

The fake rate F describes the fraction of events that are present at reconstruction level but not at generator level:

$$F = 1 - \frac{N(\text{event in reco bin \& in any gen bin})}{N(\text{event in reco bin})}\,. \tag{4.14}$$

In Figure 4.9 **(a)**, the acceptance and fake rate is shown as function of p_T^Z in the inclusive y_b-y^* phase space. The central region with $0 < y_b < 0.5$ and $0 < y^* < 0.5$ is depicted in Figure 4.9 **(b)**. Additionally, the high y_b-region and the high y^*-region are shown in Figure 4.9 **(c)** and **(d)** respectively. The overall acceptance A when combining all y_b-y^* bins is around 50 % near the required minimum p_T^Z and rises up to about 80 % above 100 GeV, while the fake rate F is close to zero at high p_T^Z, corresponding to values of unity for $1 - F$ and about 22 % at the lower end of the p_T^Z spectrum. In the central detector region, the acceptance is about 5 % better for the whole p_T^Z spectrum and the fake rate rises to around 20 % at low transverse momenta. The detector performance gets worse in the forward and backward regions, as seen in the highly boosted phase space in **(c)** and forward-backward region in **(d)** of Figure 4.9. In both cases, the detector acceptance drops to around 40 % for low p_T^Z, while it only reaches up to 70 % in the high y^*-bin in contrast to 80 % in the high y_b-bin. In the high y^*-bin, the fake rate is as high as 35 % near the lower end of the p_T^Z spectrum, improving up to 5 % towards high p_T^Z. The fake rate in the high y_b-bin is slightly better than in the high y^*-bin, starting at 30 % and dropping to around 2.5 % for high transverse momenta of the reconstructed Z boson.

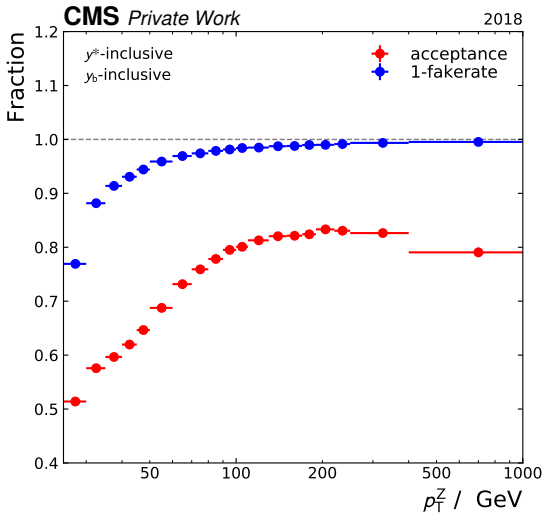
This behavior is similar for all data-taking periods, as seen in the breakdown for all y_b-y^* regions and all data-taking periods in the Figures A.2 to A.5.

The acceptance and fake rate are calculated from simulation in every phase space bin and are used as inputs for the unfolding. The `TUnfold` algorithm is able to handle loss introduced by the detector. Therefore, the calculated acceptance is used to fill the underflow bins of the response matrix, which enables `TUnfold` to correct for the loss at reconstruction level. Before unfolding, the event count in data is scaled down according to the calculated fake rate in order to disregard the amount of events originating from outside the analyzed phase space.

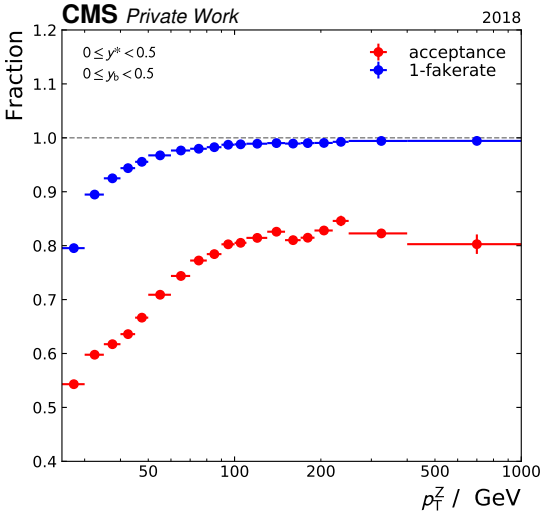
4.6. Experimental Uncertainties

The total uncertainty of the final unfolded result is constructed from multiple different sources. The statistical uncertainty arises from the limited number of events in data for each bin. The number of events filling each bin follows a Poisson distribution around the expected value, and the uncertainty is estimated with $\sqrt{N}$.

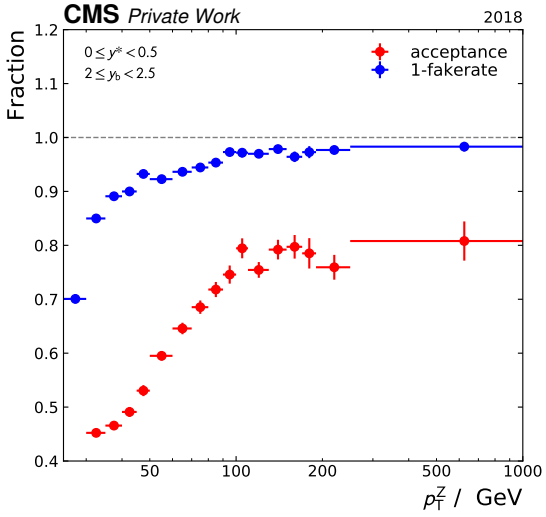
The uncertainties arising from unfolding are combined from two different sources. First, the limited number of events in the simulation and, therefore, in the response matrix have to be taken into account. This is done by the `TUnfold` algorithm that derives a full covariance matrix for the statistical unfolding uncertainties. Second, systematic uncertainties for the



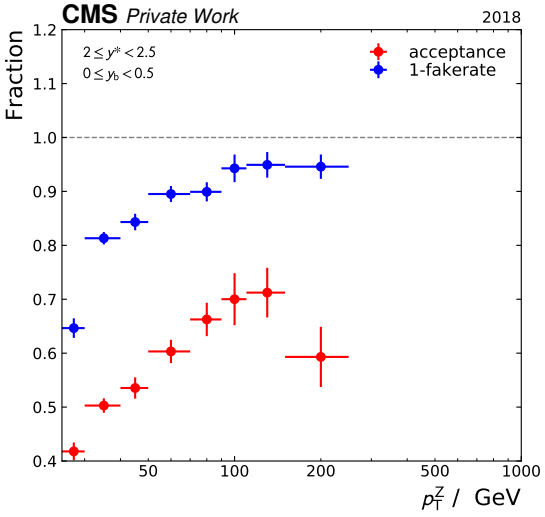
(a) Inclusive y_b-y^* phase space.



(b) Central region within $0 < y_b < 0.5$ and $0 < y^* < 0.5$.



(c) Highly boosted region within $2.0 < y_b < 2.5$ and $0 < y^* < 0.5$.



(d) Forward-backward region within $0 < y_b < 0.5$ and $2.0 < y^* < 2.5$.

Figure 4.9.: Acceptance and fake rate of the 2018 data-taking period as a function of p_T^Z in the inclusive y_b-y^* phase space **(a)**, the central region **(b)**, the highly boosted phase space **(c)**, and in the forward-backward region **(c)**. The detector performance is very good in the inclusive phase space and central region with an acceptance of up to 80 % and a fake rate F of about 0 %, resulting in values of unity for the shown $1 - F$ at high p_T^Z. In the forward and backward regions, the detector performs worse with the acceptance dropping to 40 % at low p_T^Z, as seen in **(c)** and **(d)**, and the fake rate being as high as 35 % in the high y^*-bin.

underlying simulation of the response matrix need to be taken into account. Ideally, the unfolding should be performed for MC samples produced with at least two different matrix element generators in order to estimate a possible bias introduced by the choice of the generator model. By this, effects on the unfolded result due to a difference directly at generator level can be taken into account. The additional matrix element generator should be ideally completely independent of the one used for creating the main response matrix. In the previous analysis [50], the systematic unfolding uncertainty has been estimated to be smaller than 2.5 % in the central rapidity bin. DY+0, 1, and 2 jets samples at NLO produced with `Sherpa` [87] have been requested for a central production by the CMS Collaboration. This work is currently ongoing. In the presented analysis, the total unfolding uncertainty is therefore limited to the statistical unfolding uncertainty.

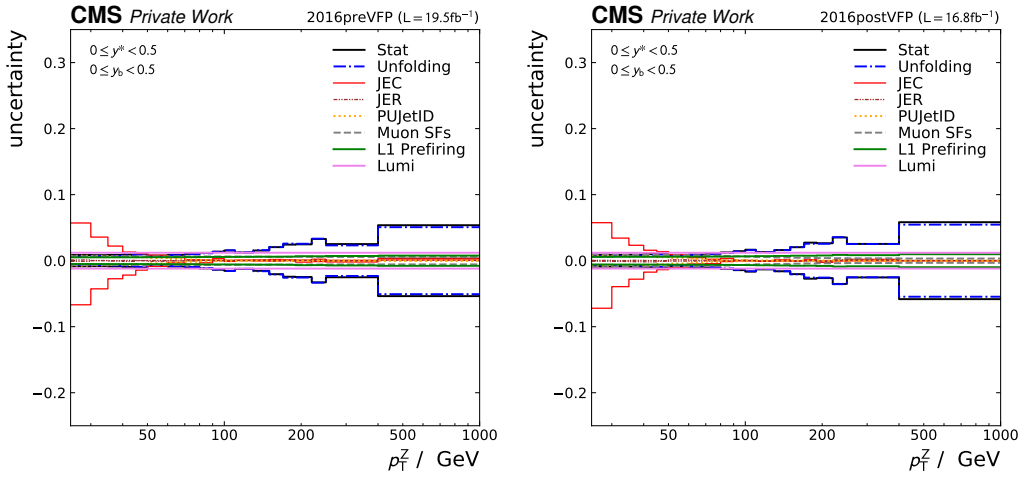
Additionally, multiple systematic variations at reconstruction level have to be taken into account, as the calculation of the energy and efficiency corrections are inherently influenced by uncertainties themselves (see section 4.4). The JEC and JER, as well as the jet puID, muon scalefactors, and L1-prefiring corrections are all subject to systematic uncertainties. For example, the JER, described in subsection 4.4.1, is subject to uncertainties that need to be propagated. By changing the energy of a jet, the transverse momentum and rapidity is changed, resulting in a possibly different bin assignment of an event or resulting in an event even failing to meet the selection criteria. All those effects need to be propagated to the final result.

The effect of each systematic uncertainty source is estimated by varying the systematic shift on the reconstruction level of the simulation and performing a new independent unfolding using the alternative response matrix. The difference in the resulting distribution to the nominal unfolding result is taken as the one-sided uncertainty caused by the according uncertainty source. A further overall normalizing uncertainty is introduced by the uncertainty on the luminosity measurement. It is 1.2 %, 2.3 %, and 2.5 % for the 2016, 2017, and 2018 data-taking periods, respectively [54–56]. This uncertainty is evaluated on the final results by varying the obtained cross-section up and down with the according uncertainty.

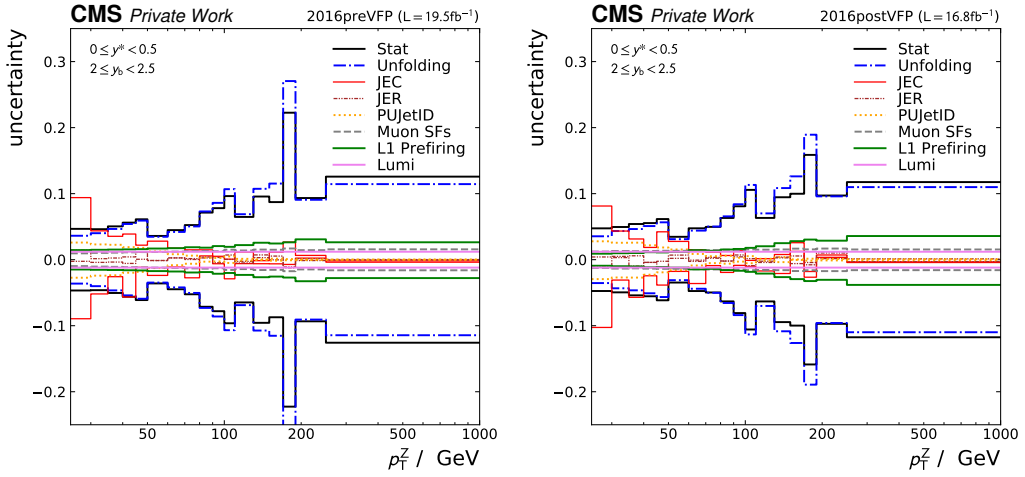
The uncertainty contribution from the background estimation has been calculated very conservatively by varying the background 50 % up and down and was estimated to be around 0.5 % in ref. [50]. It is therefore neglected. However, the reconstruction of the data and overall uncertainties could be reduced for the ReReco datasets and this contribution is briefly revisited but not propagated to the final results. In the presented analysis, the uncertainties from a background variation of 50 % range from 0.12 % to 2.1 % over all bins in the analyzed phase space. In most regions, this uncertainty is insignificant compared to the other uncertainty sources, while it can make up to 50 % of the total uncertainty in the central y^*-y_b regions at p_T^Z of around 100 GeV at detector level. Though the background variation of 50 % is very conservative, its influence on the unfolded result should be taken into account or replaced by a better approximation of the uncertainty arising from the background contributions for future analyses.

As seen in Figure 4.10 and 4.11, the JEC uncertainty is dominating for low p_T^Z across the whole phase space. For high p_T^Z, the statistical and unfolding uncertainties are the dominating contributions. In the central rapidity bins, the overall uncertainties are below 5 % for all data-taking periods. The statistical and unfolding uncertainties are rising up to 20 % in the high y^* region for 2017 and 2018 data-taking and up to 30 % for both data-taking periods in 2016. In the high y_b region, the statistical and unfolding uncertainties are the main contributing factors of around 10 % to 15 % for high p_T^Z.

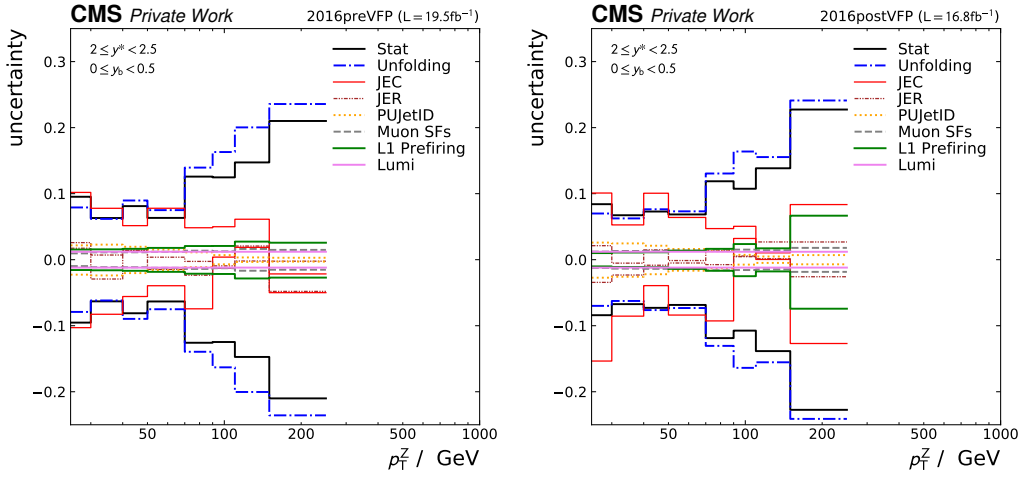
A breakdown of the uncertainties in every y_b-y^* bin for all data-taking periods is found in Figure A.10 to A.13.



(a) Uncertainty contributions in the central y^*-y_b bin.

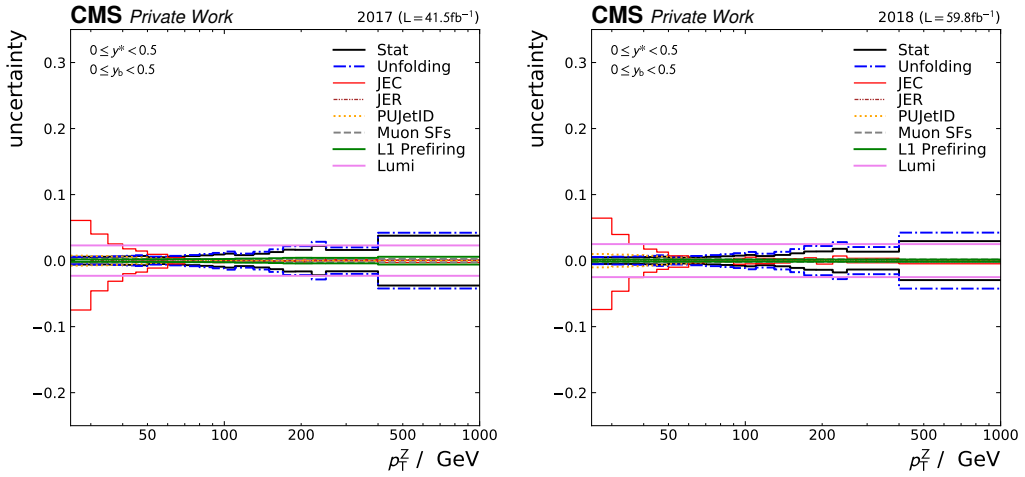


(b) Uncertainty contributions in the high y_b bin.

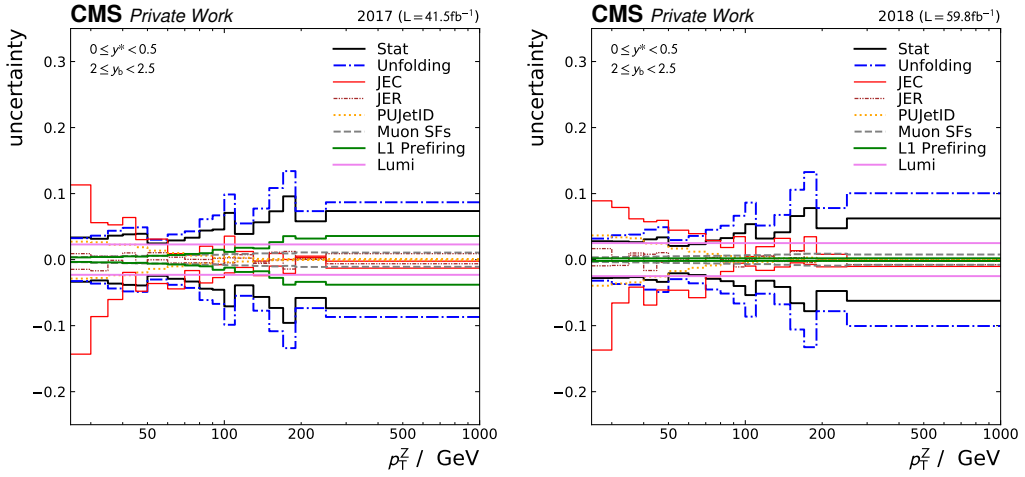


(c) Uncertainty contributions in the high y^* bin.

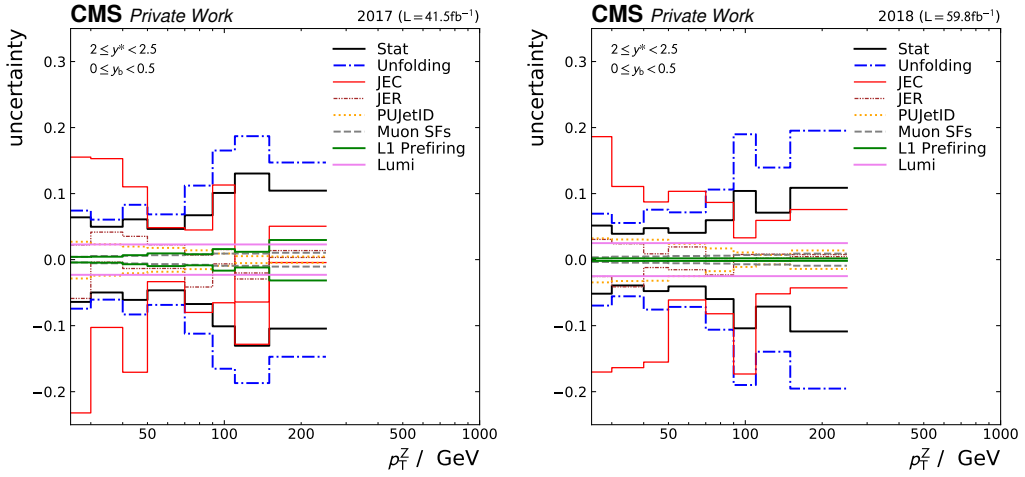
Figure 4.10.: Uncertainty contributions in the central region **(a)**, high y_b region **(b)**, and high y^* region **(a)**. The 2016preVFP data-taking period is shown on the left and 2016postVFP on the right. A similar trend with the JEC uncertainty dominating at low p_T^Z and the statistical and unfolding uncertainty dominating at high p_T^Z is observed in all data-taking periods.



(a) Uncertainty contributions in the central y^*-y_b bin.



(b) Uncertainty contributions in the high y_b bin.



(c) Uncertainty contributions in the high y^* bin.

Figure 4.11.: Uncertainty contributions in the central region **(a)**, high y_b region **(b)**, and high y^* region **(a)**. The 2017 data-taking period is shown on the left and 2018 on the right. A similar trend with the JEC uncertainty dominating at low p_T^Z and the statistical and unfolding uncertainty dominating at high p_T^Z is observed in all data-taking periods.

Table 4.7.: Correlations of the luminosity measurements for the Run 2 data-taking periods, derived from the uncorrelated and correlated uncertainties given in ref. [88].

Period	2016	2017	2018
2016	1.0	0.20	0.41
2017	0.20	1.0	0.34
2018	0.41	0.34	1.0

The combination of all uncertainty sources is performed with simple error propagation, as all uncertainty sources are treated as uncorrelated. Please note, that in the following paragraphs σ_{ij} denotes a covariance and σ_i the standard deviation of a variable subject to an uncertainty and not the cross-section of a scattering process. The total uncertainty σ_{tot} from n independent uncertainty sources is given with

$$\sigma_{\text{tot}}^2 = \sum_{i=1}^{n} \sigma_i^2 \,. \tag{4.15}$$

Here, σ_i corresponds to total uncertainty of the i-th uncertainty source. It is calculated independently for the upward and downward shift.

When deriving a new variable from multiple bins of a single measurement, the correlation between the bins introduced by the unfolding procedure has to be taken into account. Furthermore, when combining measurements of two data-taking periods, the correlations for the luminosity uncertainties have to be included. Given a covariance σ_{ij} for the correlated uncertainty sources i and j, the uncertainty on the resulting variable f depending on the different inputs x_i with uncertainty σ_i is calculated as

$$\sigma_f^2 = \sum_i \left(\frac{\partial f}{\partial x_i}\right)^2 \sigma_i^2 + 2 \sum_i \sum_{j>i} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \sigma_{ij} \,. \tag{4.16}$$

The correlation ρ_{ij} between two uncertainty sources i and j is obtained from the covariance σ_{ij} and the total uncertainties σ_i and σ_j:

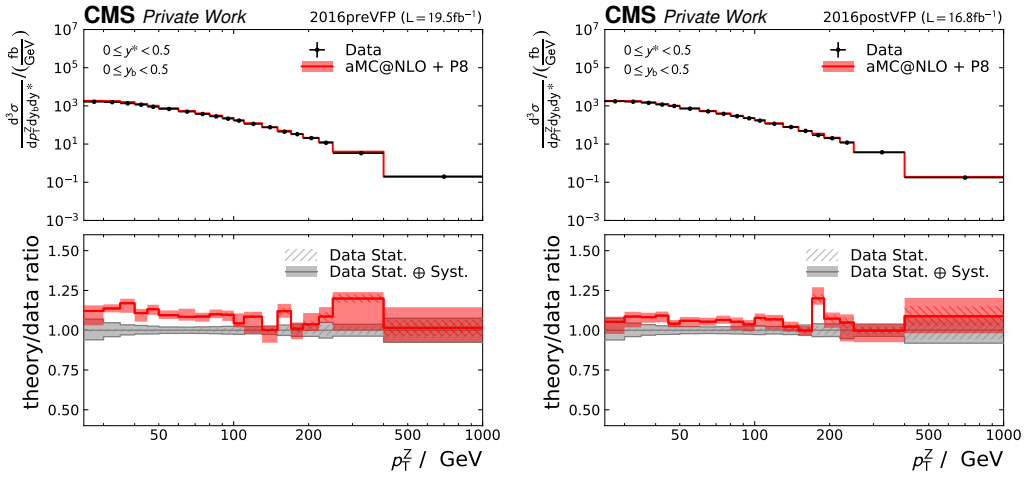
$$\rho_{ij} = \frac{\rho_i j}{\sigma_i \sigma_j} \,. \tag{4.17}$$

The correlations of the luminosity measurements for Run 2 are given in Table 4.7.

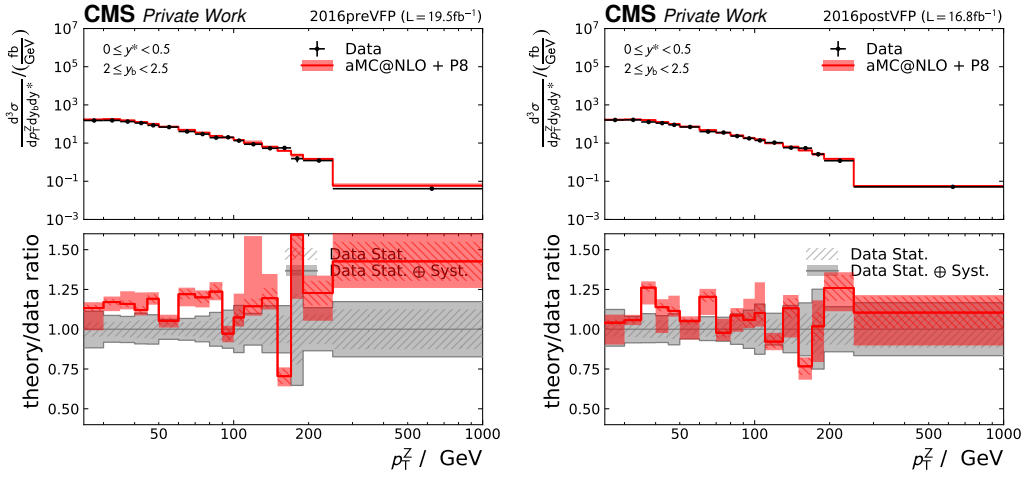
4.7. Comparison of the Measured Cross-Section to Monte Carlo Predictions

In order to obtain cross-sections after unfolding, the event counts per bin are normalized to the integrated luminosities of the corresponding input dataset. The obtained cross-sections are then compared to the prediction of the signal MC sample introduced in section 4.2 at particle level. This corresponds to a prediction for the differential cross-sections calculated at NLO for the Z+0 jet, Z+1 jet, and Z+2 jet processes with leading logarithmic (LL) resummation by the parton shower.

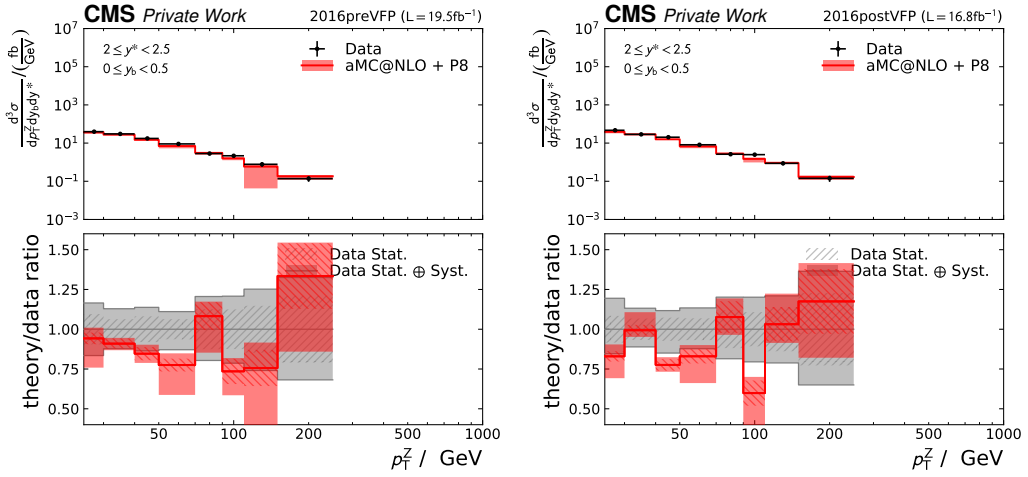
In Figure 4.12 and 4.13, the unfolded cross-section compared to the MC prediction is shown exemplary for three different y_b-y^* bins for each data-taking period. The results for the central rapidity bin (left), high y_b (middle), and high y^* (right) are shown for each period. Additionally, the ratio between the predicted and measured cross-section is plotted. Plots for all y_b-y^* bins are found in Figure A.6 to A.9.



(a) Unfolded cross section in the central y^*-y_b bin.

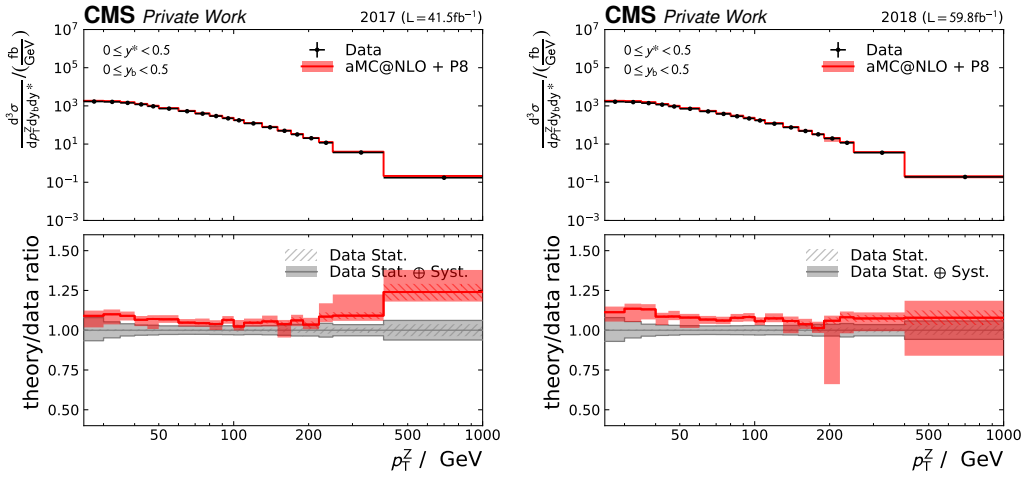


(b) Unfolded cross section in the high y_b bin.

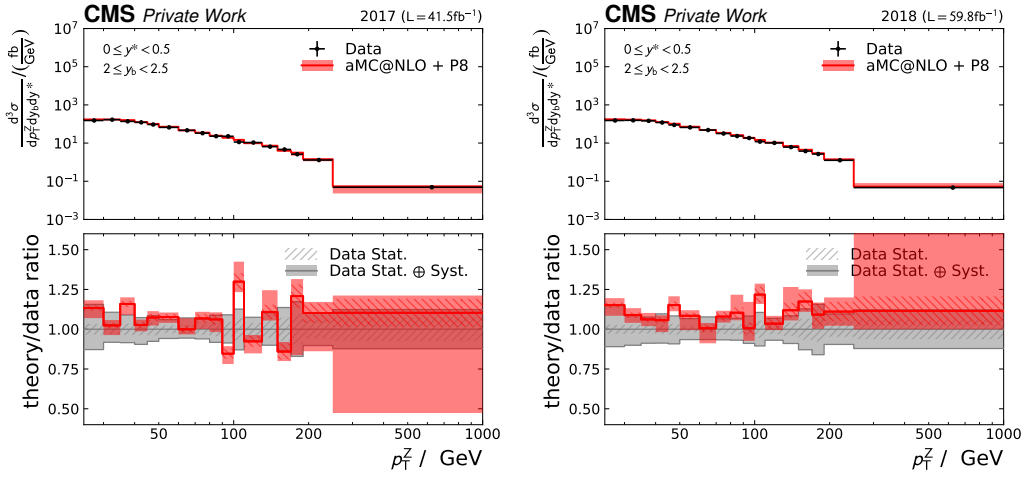


(c) Unfolded cross section in the high y^* bin.

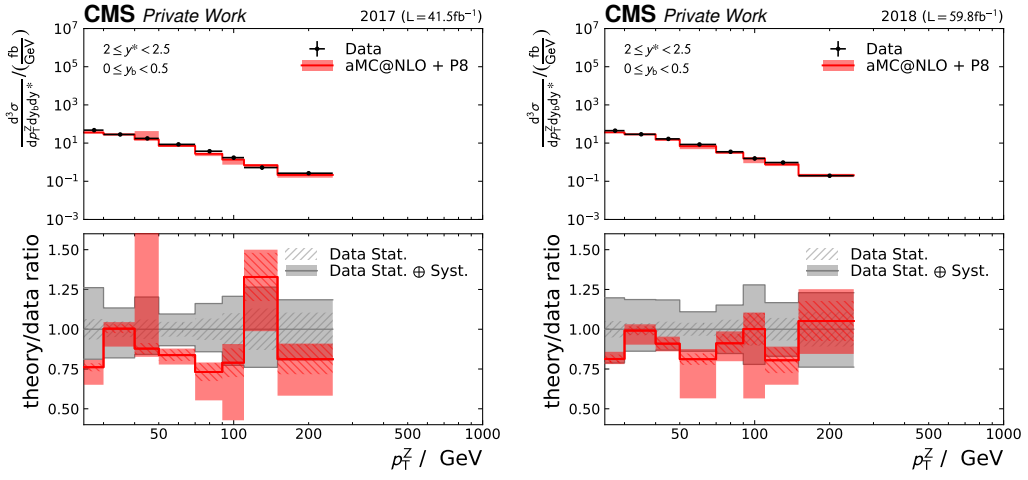
Figure 4.12.: Unfolded cross-sections compared to the theoretical predictions obtained from the MC signal sample as a function of p_T^Z. Three different y_b-y^* bins are shown, central rapidity **(a)**, high y_b **(b)** and high y^* **(c)**, with the 2016preVFP data-taking period on the left and 2016postVFP on the right.



(a) Unfolded cross section in the central y^*-y_b bin.



(b) Unfolded cross section in the high y_b bin.



(c) Unfolded cross section in the high y^* bin.

Figure 4.13.: Unfolded cross-sections compared to the theoretical predictions obtained from the MC signal sample as a function of p_T^Z. Three different y_b-y^* bins are shown, central rapidity **(a)**, high y_b **(b)** and high y^* **(c)**, with the 2017 data-taking period on the left and 2018 on the right.

The uncertainty band on the unfolded data contains the combination of the statistical and all systematic uncertainty sources listed in section 4.6, while the uncertainty band on the simulation includes the statistical uncertainties combined with the uncertainties on the cross-section, initial state radiation, and final state radiations. Further systematic uncertainties on the generation of the MC sample that will increase the total uncertainty, such as the uncertainties on the PDFs or scale variations on the coupling constants, are not included.

An overall good agreement between the MC prediction and the unfolded data within uncertainties is observed. Larger fluctuations occur in the statistically limited phase spaces, for example, at high p_T^Z, in the high y_b-bin, and high y^*-bin. In the central rapidity bin, a constant normalization offset is observed. This offset might be caused by an insufficiently accurate inclusive cross-section prediction at NLO accuracy for the Z+1 jet process (see subsection 4.2.3).

4.8. Comparison of the Different Data-Taking Periods

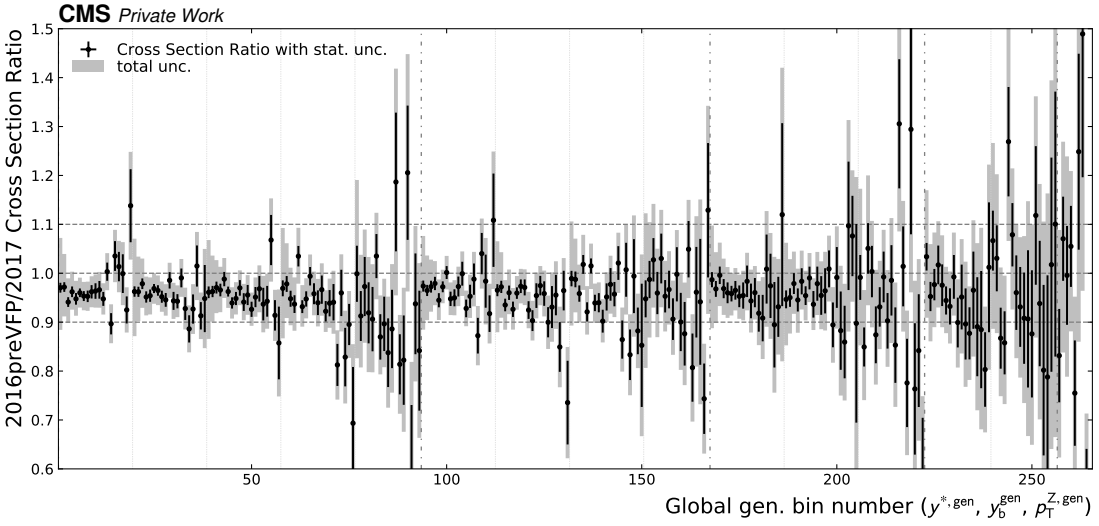
In a previous analysis of the triple-differential Z+jet cross-section measurement [51], the measured cross-section was found to be more than 5 % higher in the 2017 data-taking period than for 2016. Additionally, a small dependence on the transverse momentum of the reconstructed Z boson was observed. Since the previous analysis was performed with a prompt reconstruction of the data, not all detector effects were known and taken into account in the reconstruction and simulation. The presented analysis used the so called ReReco datasets with improved reconstruction algorithms. Furthermore, newer simulations modelling the so far understood detector effects take further sources of deviations into account.

In order to compare the cross-sections obtained for each analyzed data-taking period, the ratios between different data-taking periods are calculated. For this, the uncertainties are propagated, as described at the end of section 4.6. All uncertainty sources except for the luminosity and systematic unfolding effects are treated as uncorrelated, as they are all independently determined. However, the systematic effects of using a different generator for unfolding are not taken into account in this analysis.

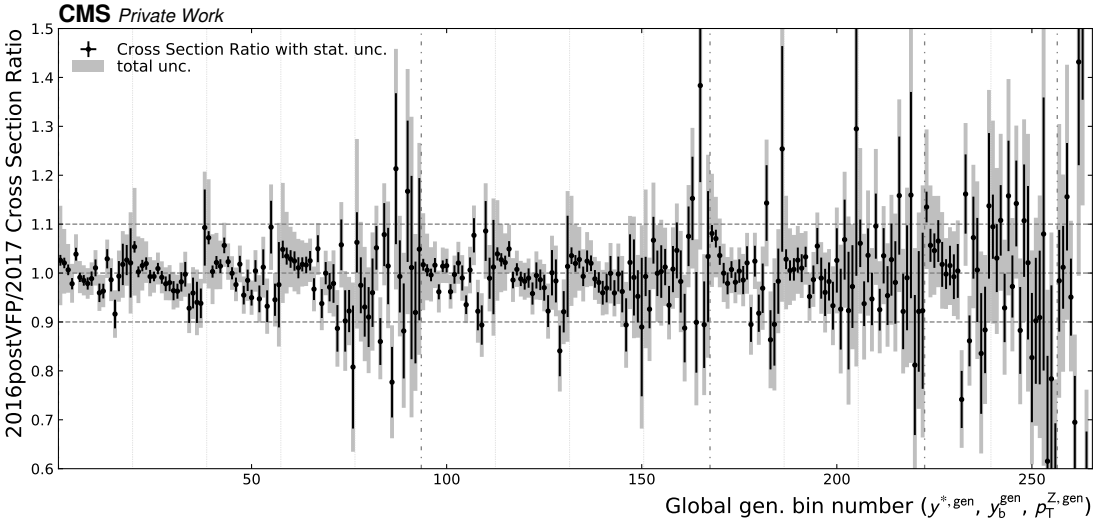
In Figure 4.14, the ratios of the measured cross-sections of the 2016preVFP and 2016postVFP data-taking periods compared to the 2017 data-taking period are shown. An overall normalization offset of around 5 %, resulting in a smaller observed cross-section in the 2016preVFP dataset, is observed. The luminosity uncertainty only accounts for 2.4 % in the ratio. This issue might originate from the issue in the APV readout chips, as described in section 4.2. Since the loss in tracking efficiency could only be recovered in offline reconstruction, an unrecoverable efficiency loss caused by the trigger system might still be present in the data. The differences in the observed cross-sections in 2016preVFP and 2017 are at the order of magnitude of about one standard deviation of the combined uncertainties.

By comparing the observed cross-sections of the 2016postVFP and 2017 data-taking periods, a dependence on the transverse momentum of the Z boson p_T^Z is observed for all y^*-y_b bins. This effect is not caused by normalization, as it is contained within uncertainties on average over all p_T^Z-bins. Nevertheless, the dependence on the transverse momentum of the Z boson is a significant deviation of the 2016postVFP era compared to the other data-taking periods.

Figure 4.15 shows the ratio of the observed cross-sections in the 2017 and 2018 data-taking periods. Here, a small overall offset resulting in a higher observed cross-section of about 2 % in the 2017 data compared to 2018 is found. This offset is within the uncertainty of the luminosity measurement, which is 2.8 % for the ratio taking correlations into account.



(a) Cross section measured in 2016preVFP over 2017.



(b) Cross section measured in 2016postVFP over 2017.

Figure 4.14.: Ratios of the measured cross-sections obtained for the data-taking periods of 2016 and 2017. The measurement of the 2017 data is chosen as a reference for comparison. The cross-section measurement of the 2016preVFP data taking period is about a constant factor of 5 % lower than the one obtained from the data taken in 2017, as shown in **(a)**. The deviation is about one standard deviation averaged over all phase space bins. Comparing the 2016postVFP data-taking period to the 2017 data-taking period as depicted in **(b)**, a significant dependence on p_T^Z is observed, while the normalization averaged over the full p_T^Z range can be attributed to the uncertainties.

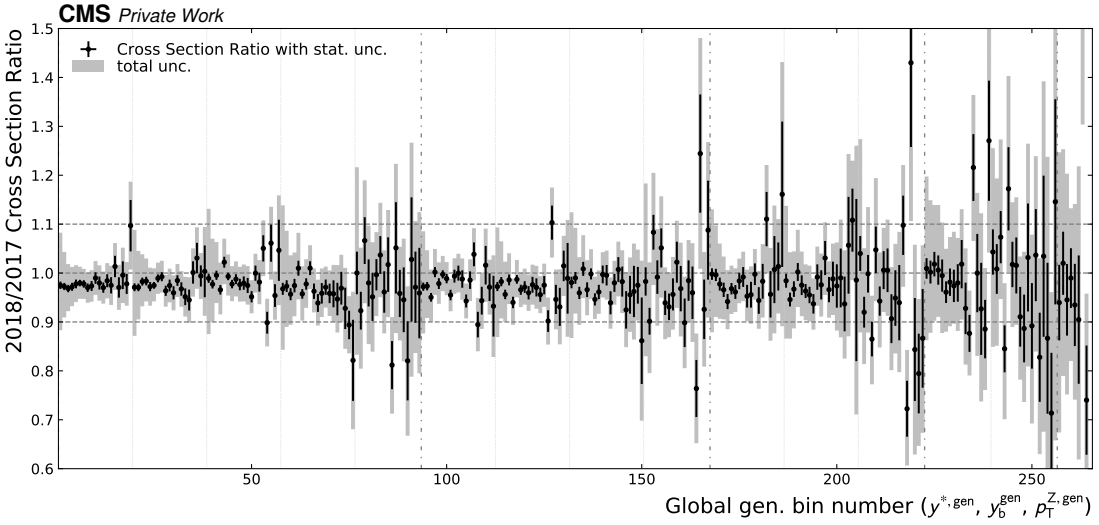


Figure 4.15.: Ratio of the measured cross-sections obtained for the data-taking periods of 2017 and 2018. The measurement of the 2017 data is chosen as a reference for comparison. For the 2018 data-taking period, an overall offset within the luminosity uncertainty compared to the 2017 data-taking period is observed.

These observations confirm the findings of the previous analysis [51], while the tension between the measurements with data taken in 2016 and 2017 is reduced due to the improved reconstruction of the ReReco datasets and updated simulations. The biggest inconsistency between the data-taking periods is found to be caused by the 2016preVFP dataset. Moreover, a significant dependence on the transverse momentum of the Z boson is observed in the 2016postVFP dataset compared to the other data-taking periods. In contrast to the 2016preVFP data-taking period, no significant discrepancy of the overall normalization is found in the 2016postVFP dataset.

Ideally, a combined measurement of the full Run 2 dataset should be performed, to reduce the statistical uncertainties on the final measurement. The measurements of both 2016 data-taking periods are in significant tensions with the other Run 2 datasets. However, a combination of the 2017 and 2018 data can be performed without risking a bias due to unknown or uncorrected detector, as the measured cross-sections are compatible within their uncertainties. The 2016 data-taking periods contribute about 26 % of the total integrated luminosity for the whole Run 2 dataset. Dropping both 2016 data-taking periods in order to perform a combined measurement would result in a significant increase in statistical uncertainty. Therefore, it is essential to investigate and understand the differences of the data-taking periods causing those effects for the goal of providing a high precision measurement of the triple-differential Z+jet cross-section at a center-of-mass energy of 13 TeV.

5. Conclusions

This thesis presents the first triple-differential cross-section measurement of the Z+jet production in proton-proton collisions for the data taken in 2018 with the CMS detector at a center-of-mass energy of 13 TeV. The datasets taken during 2016 and 2017 are analyzed as well, using the latest available reconstruction algorithms and efficiency corrections. This results in improved measurements compared to previous analyses presented in references [50, 51]. A three-dimensional unfolding procedure corrects the measurement for migrations from the true to the reconstructed spectrum, introduced due to the limited detector resolution. Additionally, the limited detector acceptance is taken into account in the unfolding procedure. In order to reduce the impact of the limited number of events in simulation on unfolding, additional MC samples are combined with the standard signal sample for Z boson production. This includes NLO simulations for the Z+0 jet, 1 jet, and 2 jet production.

Systematic uncertainties are taken into account by propagating their effect on the measured observables through the unfolding procedure. The uncertainties on the JEC dominate at a low transverse momentum of the Z boson, while the limited number of events in data and simulation are dominating the total uncertainty at high p_T^Z, as expected. The systematic uncertainty on the luminosity dominates in the central detector region for a p_T^Z range of 50 GeV to 500 GeV.

An overall good agreement within uncertainties between the measured cross-section and predictions at NLO from MC simulation is found for the 2017 and 2018 data taking periods. The prediction has an overall slightly higher normalization in the low y^*-regions, while the observed cross-section is slightly higher than the prediction for the high y^* bins. The measurements for the data taken in 2016 are also compatible with the predicted cross-sections and the according uncertainties, though the overall normalization discrepancies are greater than for the 2017 and 2018 data.

When only comparing the measurements for the different years, significant discrepancies between the data taken in 2016 compared to the data-taking periods of 2017 and 2018 are observed. The overall normalization for the 2016preVFP data-taking period is about 5 % lower compared to the measurement for the 2017 data. The luminosity uncertainty only accounts for 2.4 % in the ratio of the measured cross-sections for the data-taking periods 2016preVFP and 2017. The overall offset is at the order of magnitude of one standard deviation of the total uncertainty. When comparing the obtained cross-section of the 2016postVFP data-taking period to the data taken in 2017, a clear dependence in the ratio of those periods on the transverse momentum of the Z boson is observed. The normalization averaged over all bins is compatible with the uncertainties of the measurement.

The observation of the differences among the data-taking periods are only possible since the overall understanding of the data taken with the CMS detector has reached a new level of precision. This enabled uncovering effects at the scale of a few percent that have to be understood. These effects call for in-depth investigations of the detector performance and

improvements in the reconstruction and analysis of the data taken in 2016. Understanding those effects and subsequently correcting them allows a combination of the full Run 2 dataset and an unprecedentedly precise measurement of the triple-differential Z+jet cross-section. A full Run 2 measurement will therefore be a rigorous test of the theoretical predictions and opens the door for further improvements on the PDFs. A combination with further measurements on the full Run 2 data, such as differential cross-sections for dijet and $t\bar{t}$ production, will especially improve the description of the gluon PDF. This allows for more precise theory predictions, benefiting all analyses performed on data taken at hadron colliders. With an ongoing improvement of the experimental aspects, also more precise theory predictions are necessary. Especially for the future high luminosity LHC runs that will increase the amount of data by a factor of ten, drastically reducing the statistical uncertainties on experimental measurements. All this combined will result in a better understanding of the physics of the fundamental particles building our universe and is the foundation for future high-precision analyses.

Appendix

A. Supplementary Plots

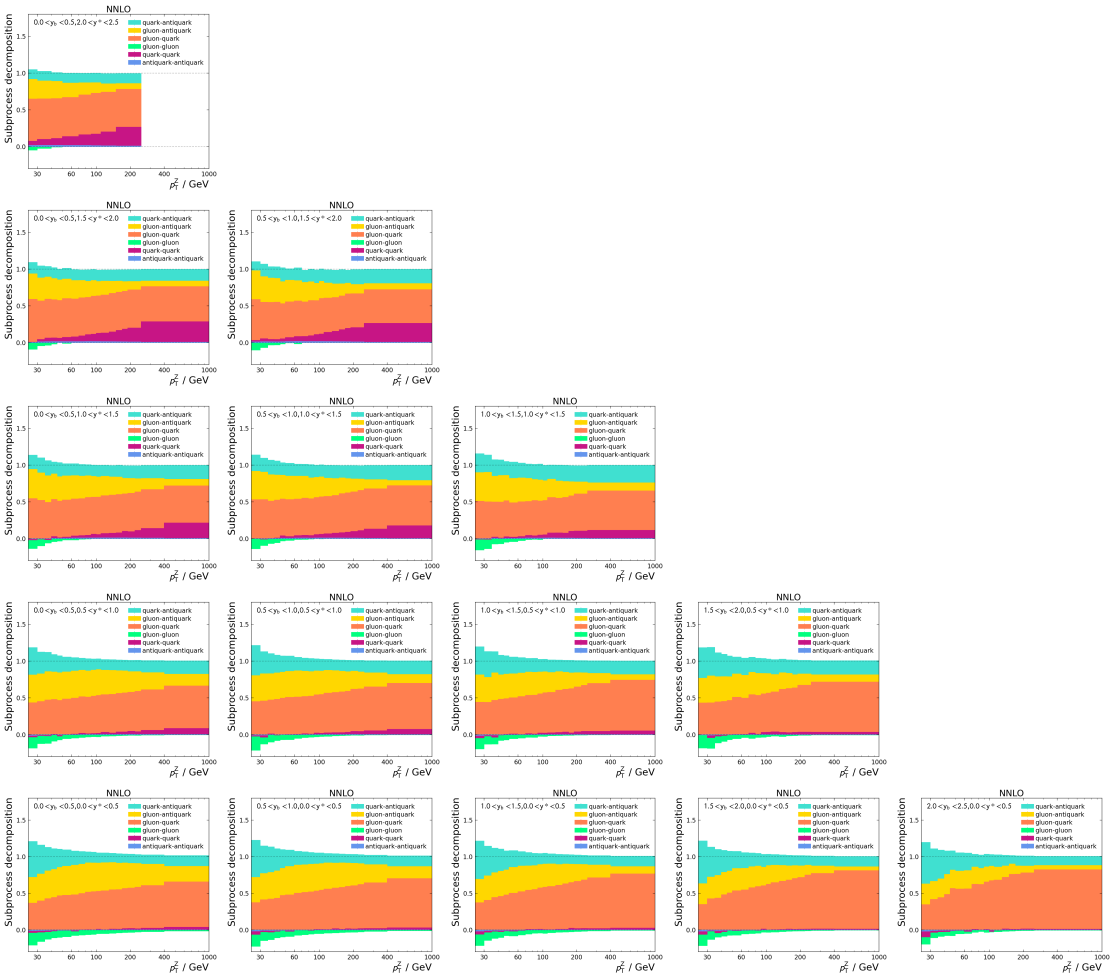


Figure A.1.: Sub-process contribution for Z + jet production at NNLO in all rapidity bins as a function of p_T^Z. Taken from [50].

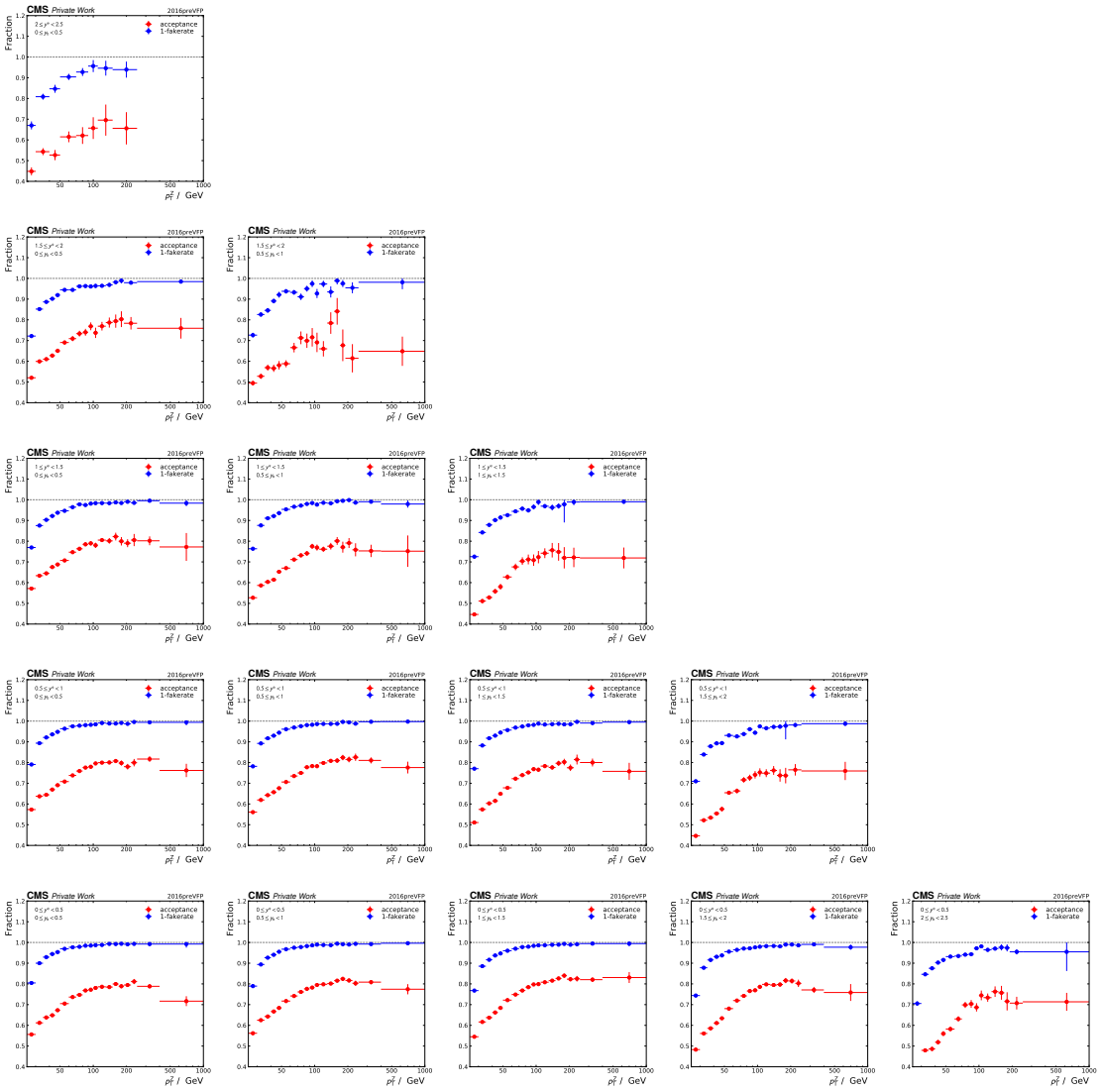


Figure A.2.: Acceptance and Fakerate as a function of p_T^Z in all y_b-y^* bins for the 2016preVFP data-taking period.

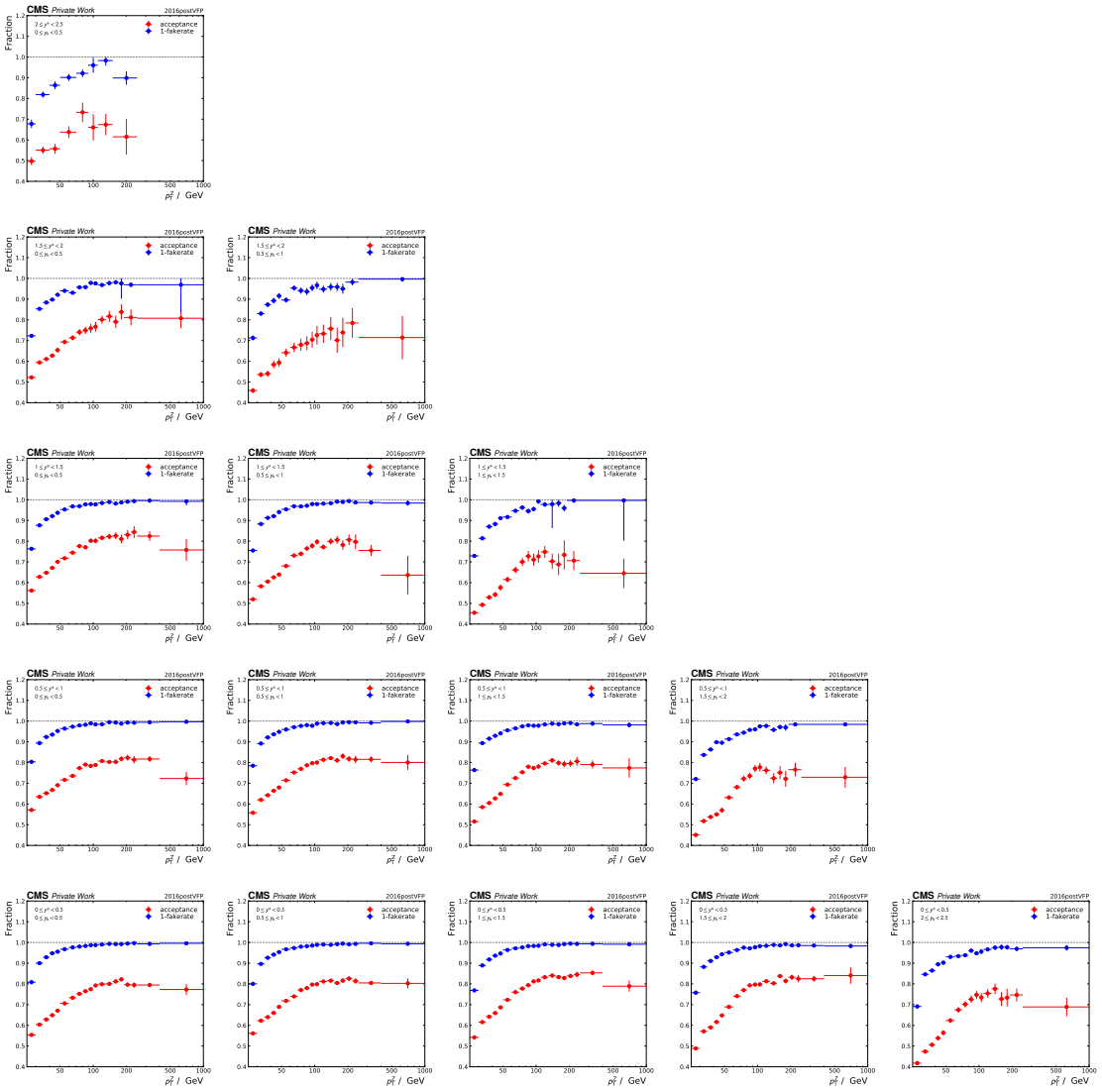


Figure A.3.: Acceptance and Fakerate as a function of p_T^Z in all y_b-y^* bins for the 2016postVFP data-taking period.

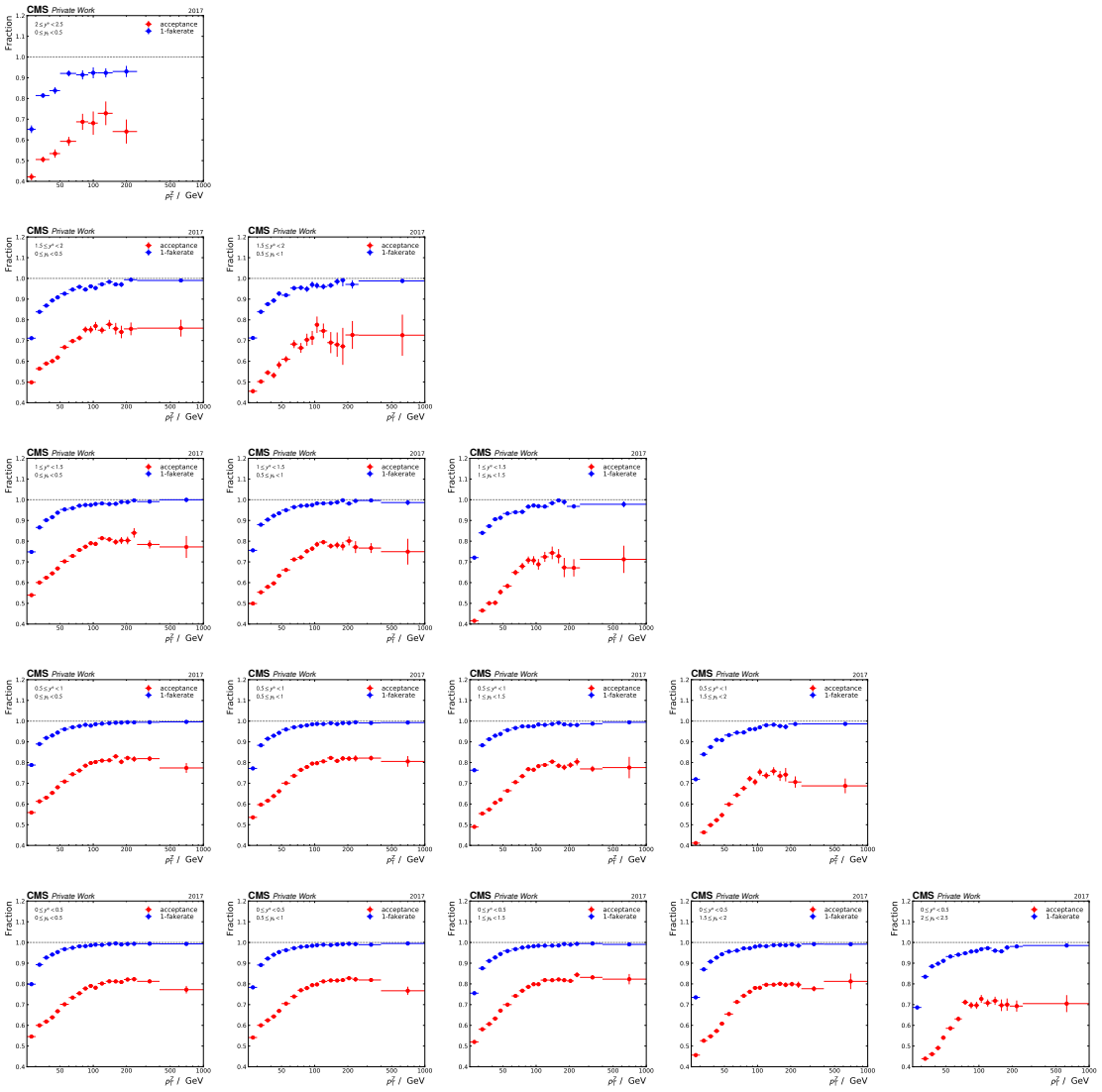


Figure A.4.: Acceptance and Fakerate as a function of p_T^Z in all y_b-y^* bins for the 2017 data-taking period.

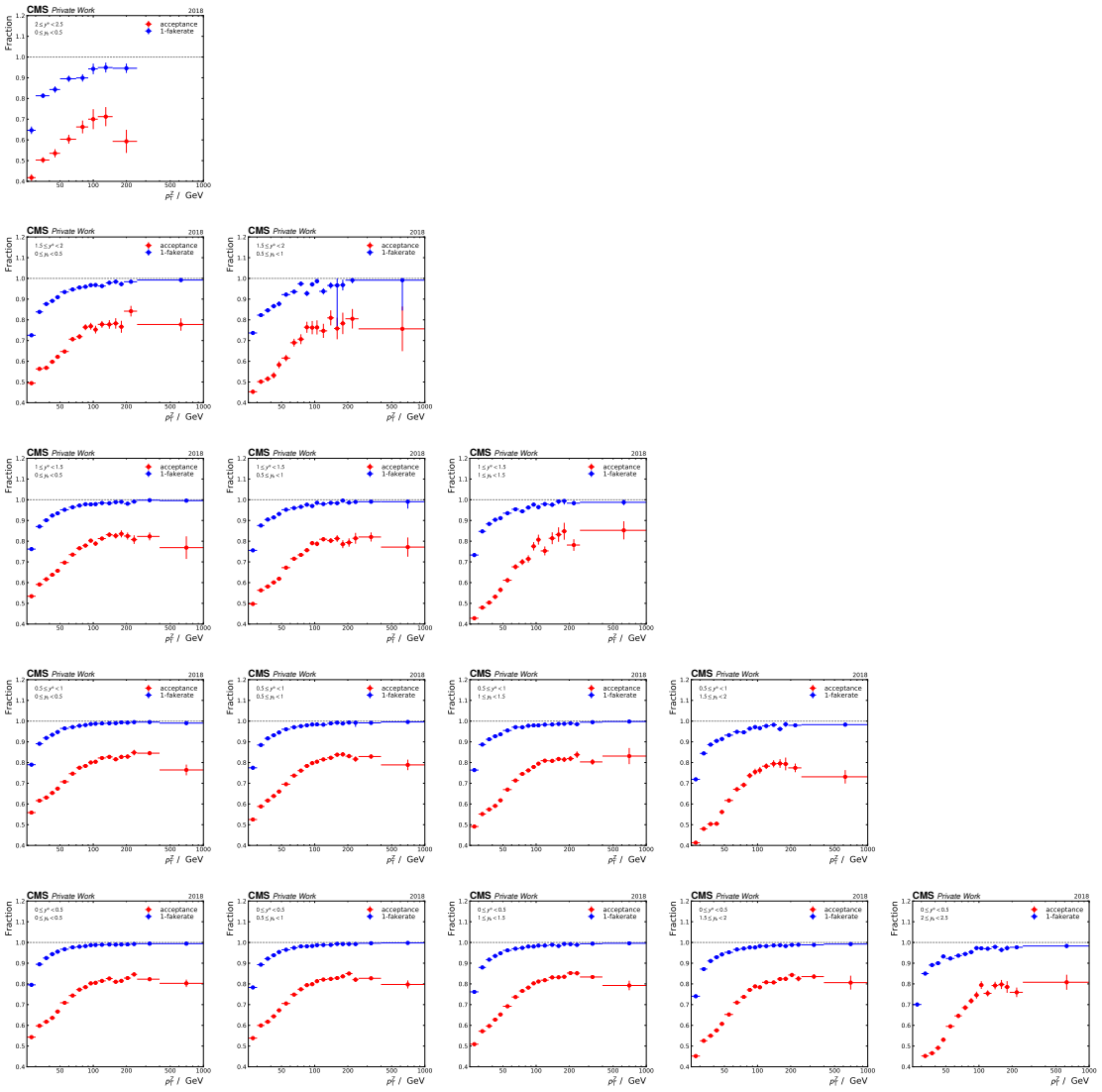


Figure A.5.: Acceptance and Fakerate as a function of p_T^Z in all y_b-y^* bins for the 2018 data-taking period.

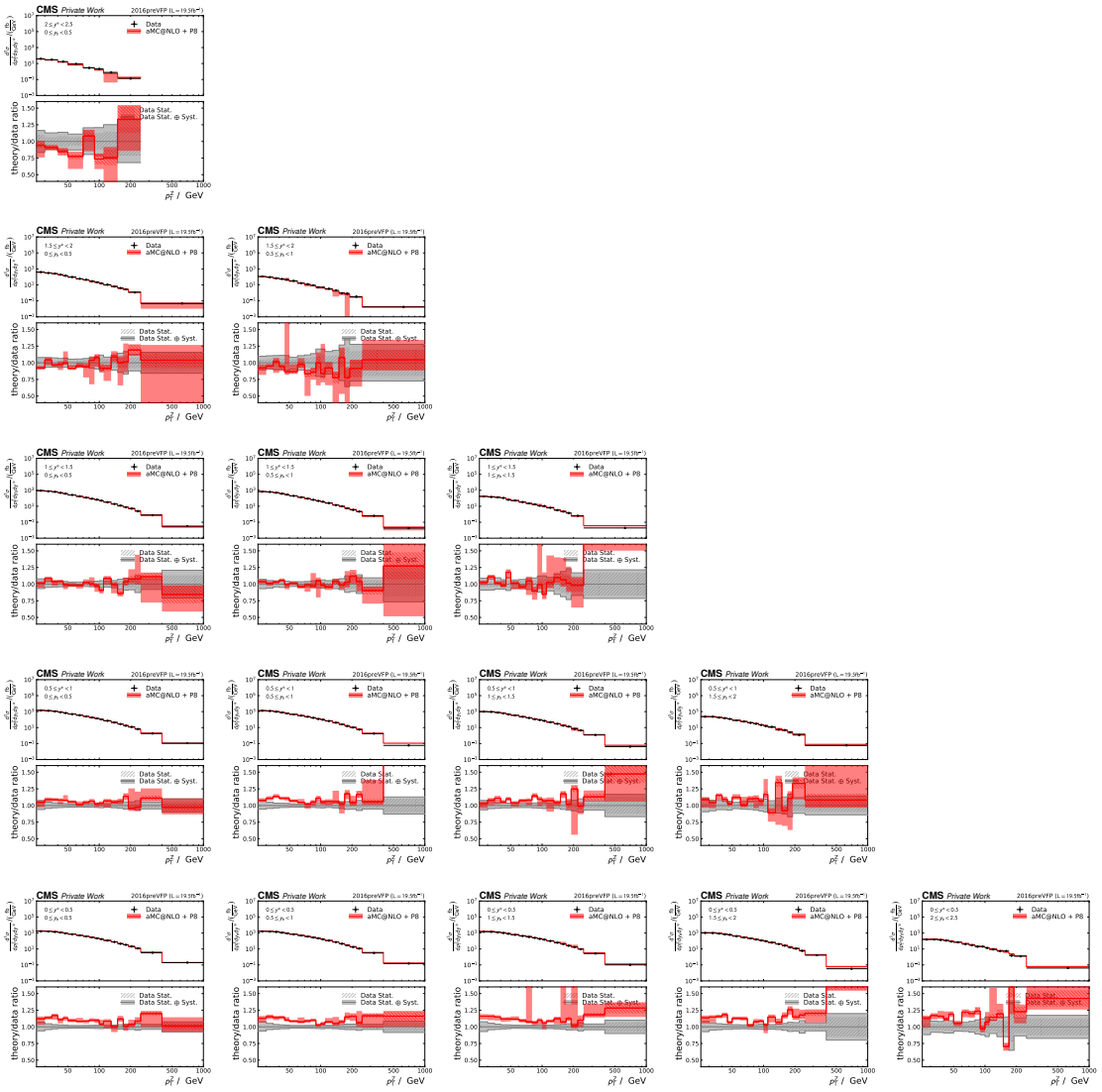


Figure A.6.: Unfolded cross-section compared to the NLO prediction from the aMC@NLO MC sample for the 2016preVFP data-taking period.

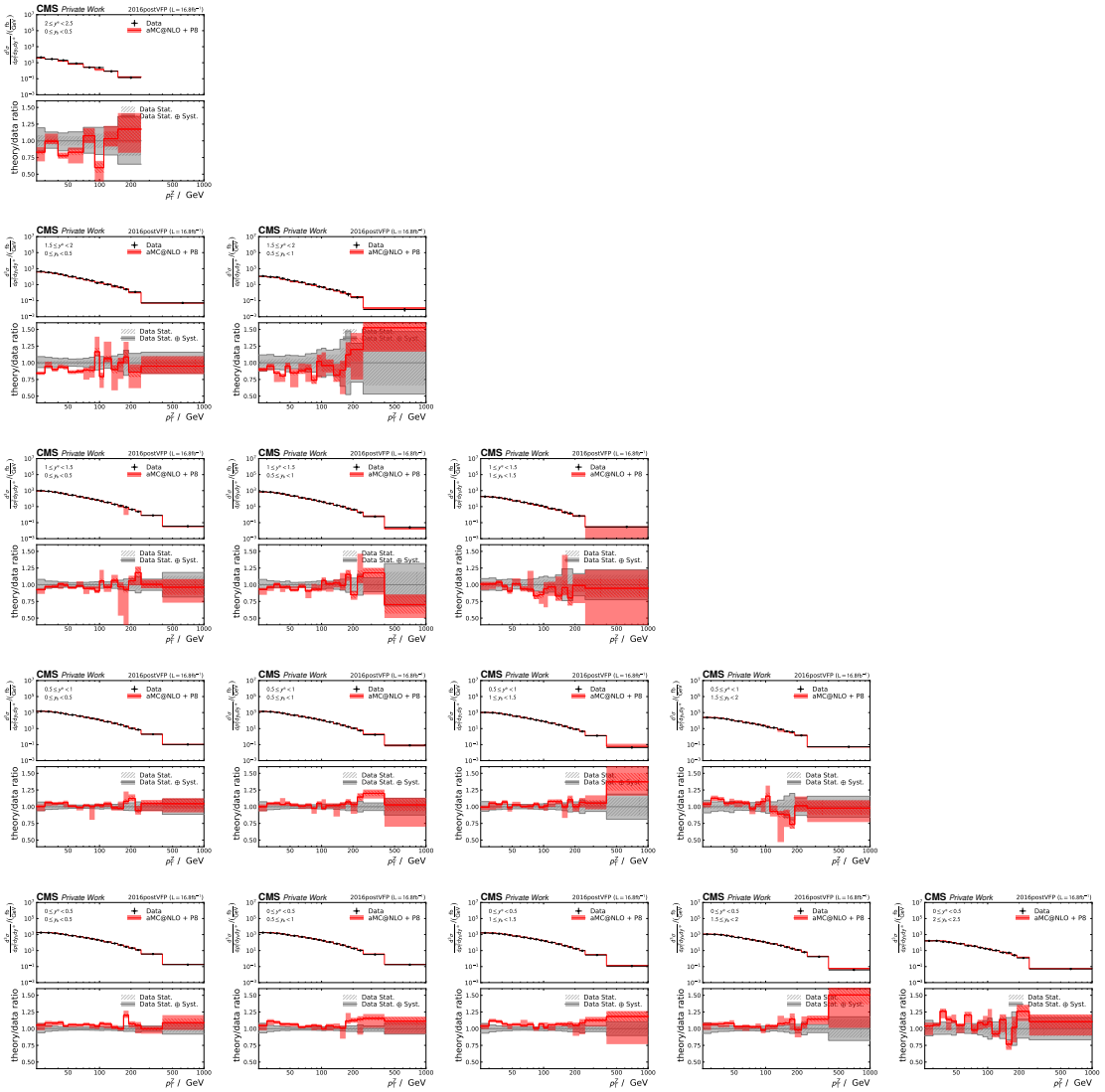


Figure A.7.: Unfolded cross-section compared to the NLO prediction from the aMC@NLO MC sample for the 2016postVFP data-taking period.

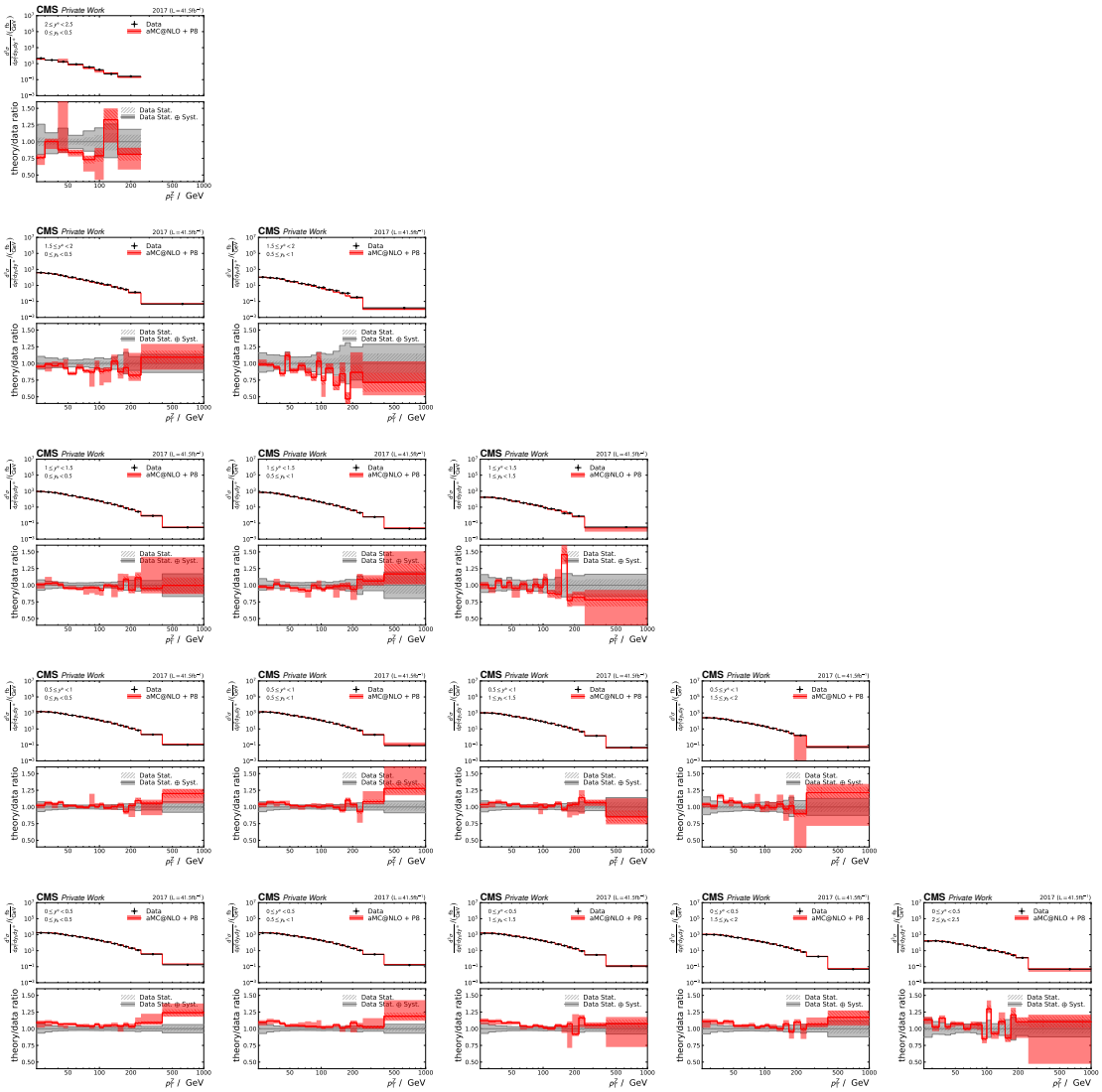


Figure A.8.: Unfolded cross-section compared to the NLO prediction from the aMC@NLO MC sample for the 2017 data-taking period.

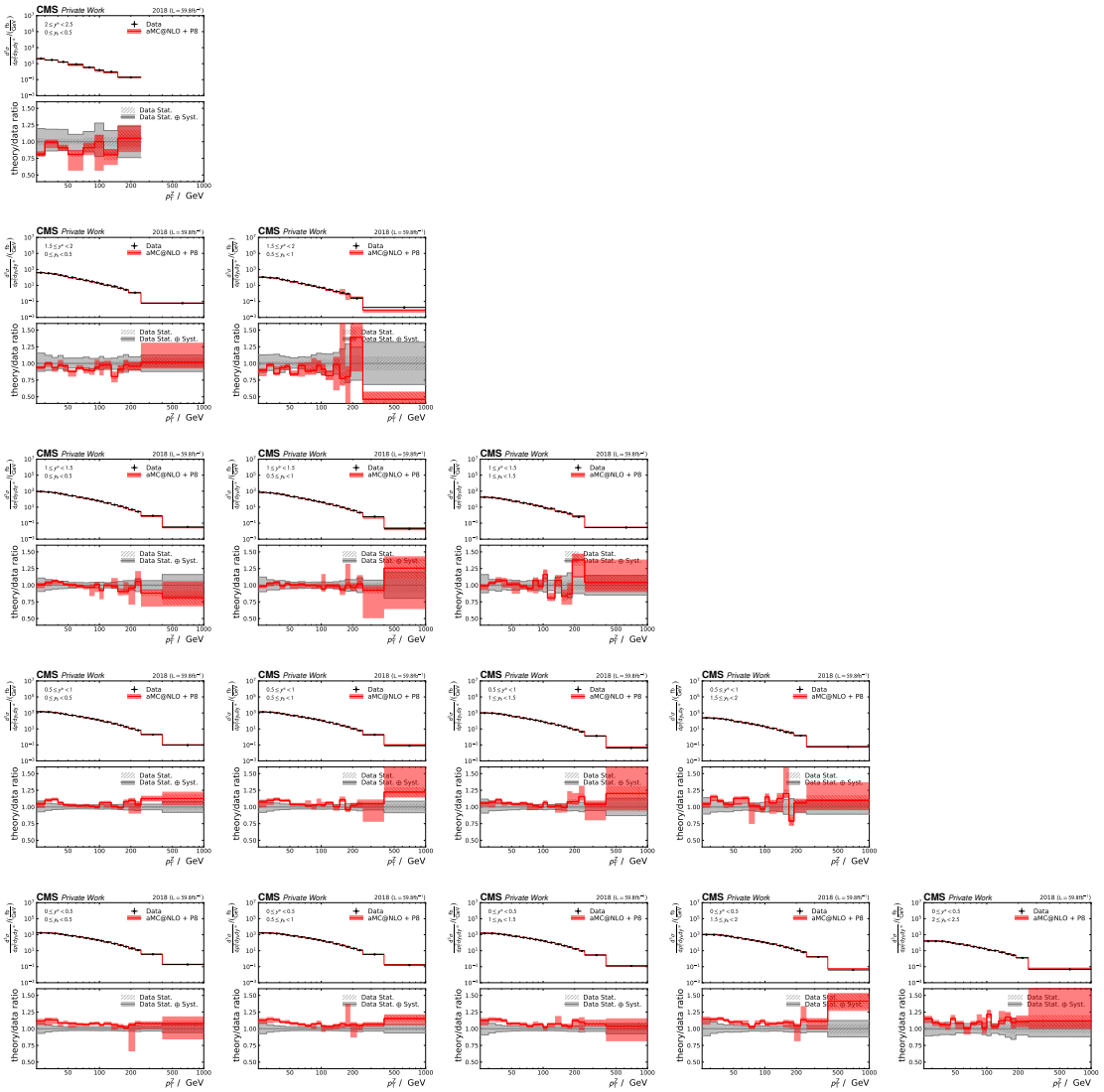


Figure A.9.: Unfolded cross-section compared to the NLO prediction from the aMC@NLO MC sample for the 2018 data-taking period.

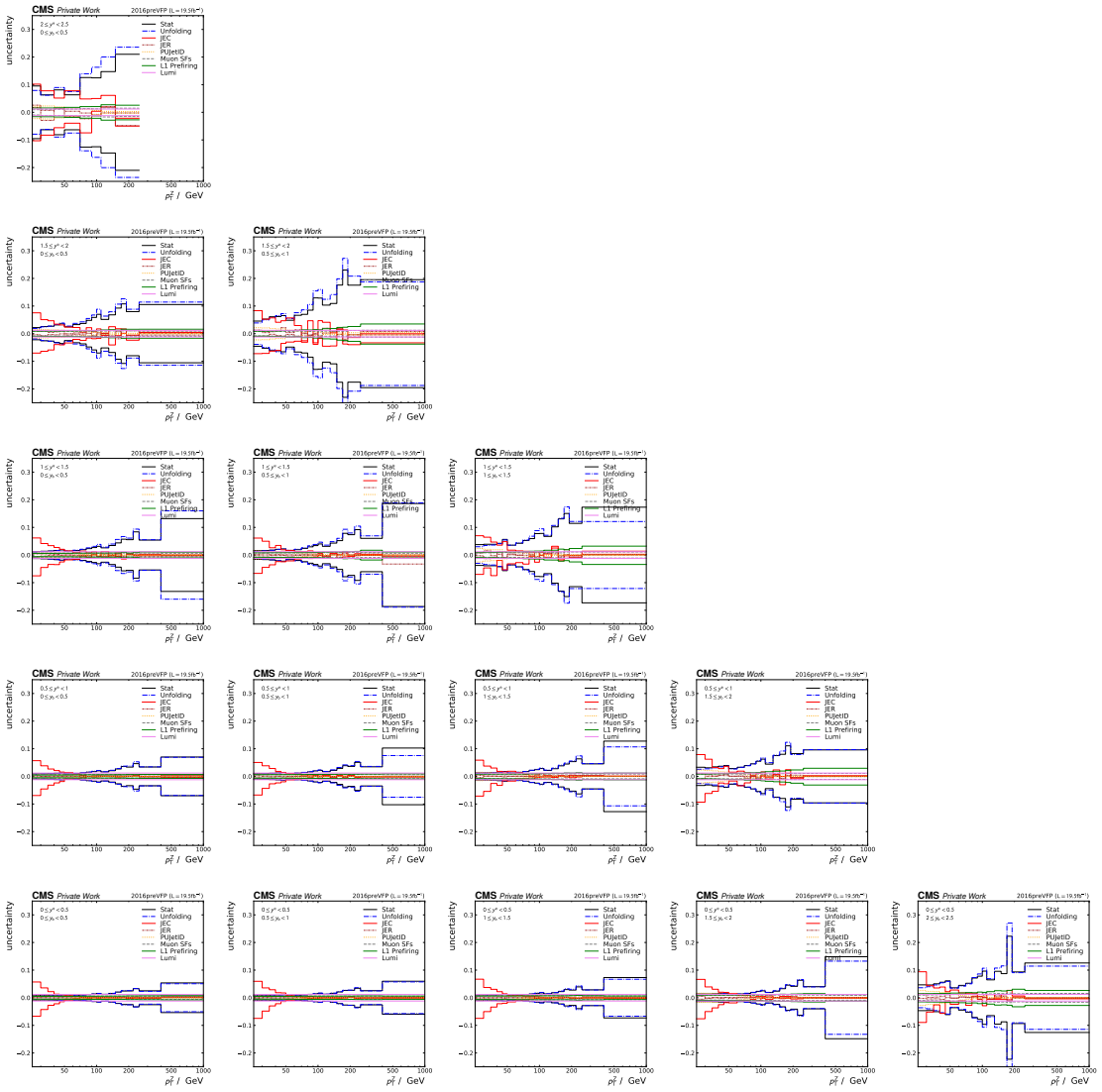


Figure A.10.: Breakdown of the different uncertainty contributions in all y_b-y^* bins for the 2016preVFP data-taking period.

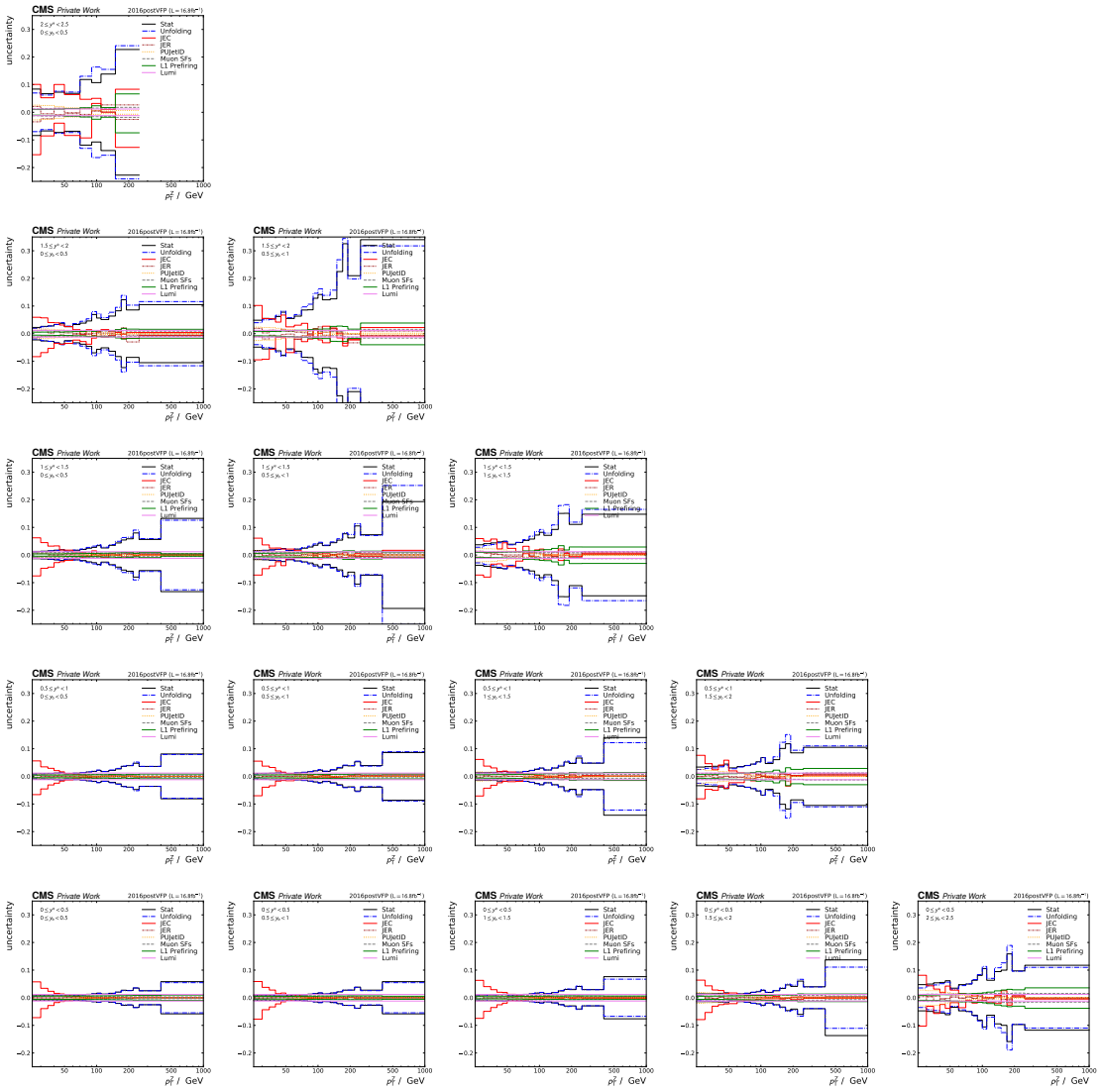


Figure A.11.: Breakdown of the different uncertainty contributions in all y_b-y^* bins for the 2016postVFP data-taking period.

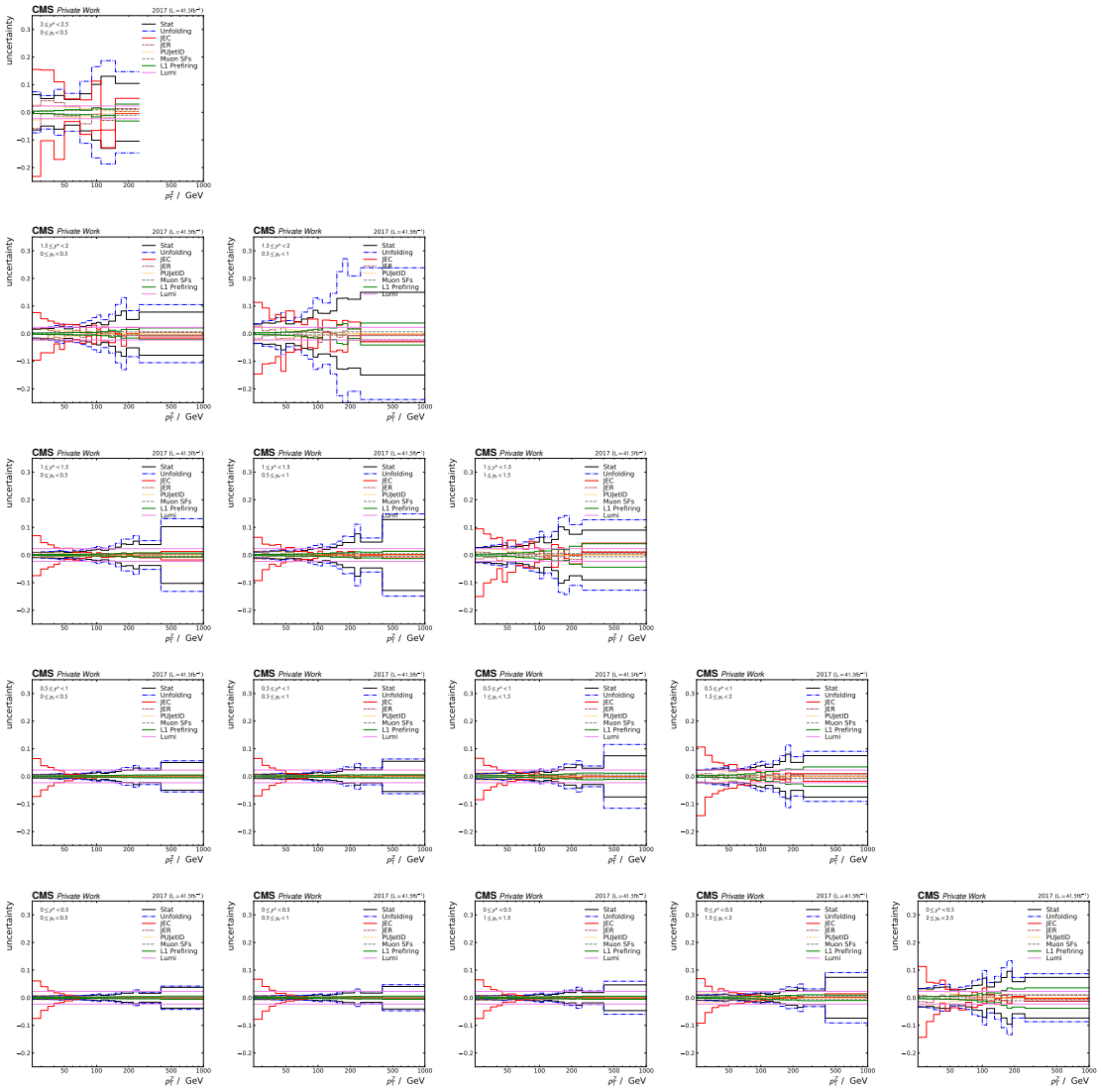


Figure A.12.: Breakdown of the different uncertainty contributions in all y_b-y^* bins for the 2017 data-taking period.

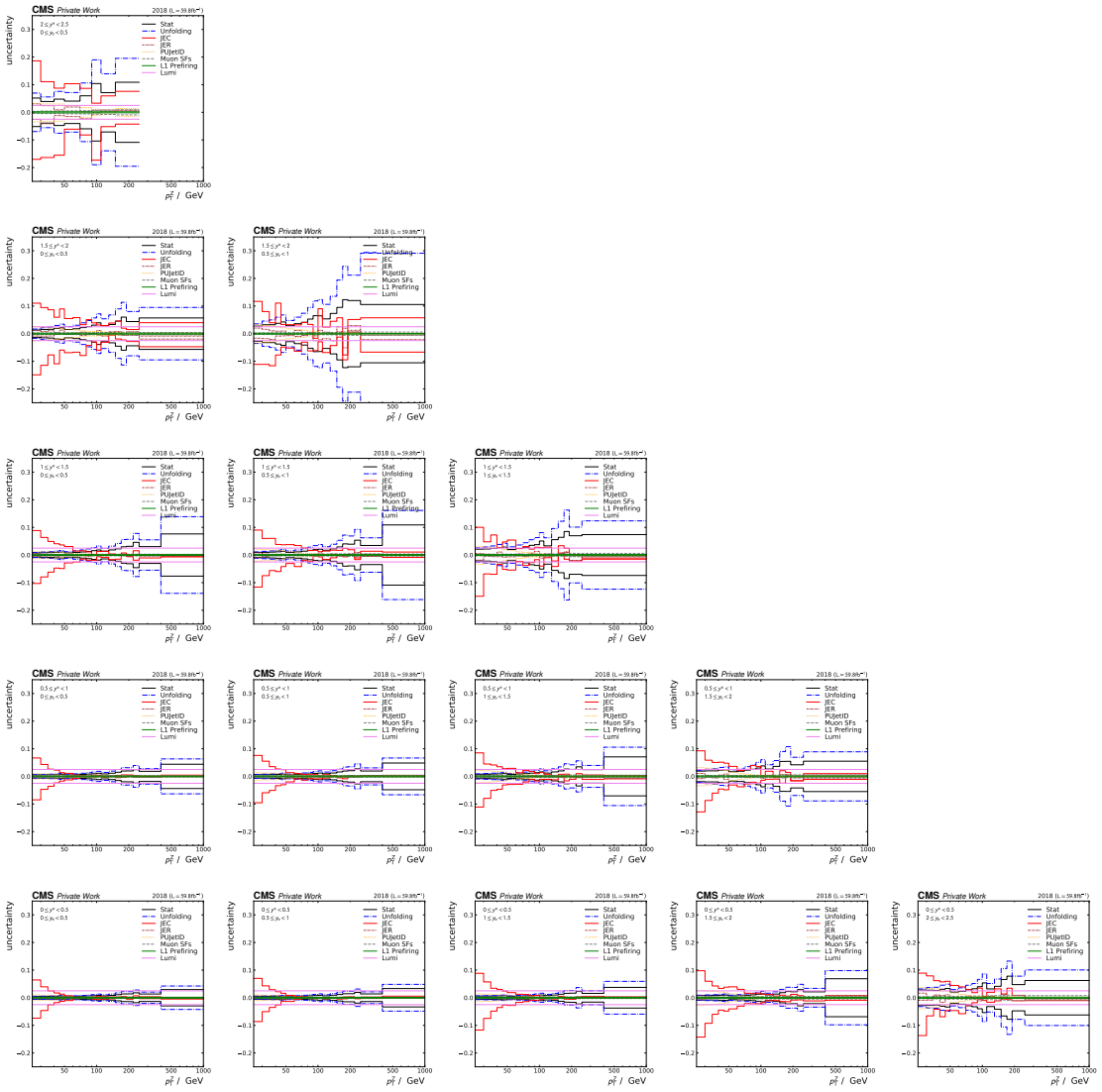


Figure A.13.: Breakdown of the different uncertainty contributions in all y_b-y^* bins for the 2018 data-taking period.

B. Datasets and Monte Carlo Samples

Table B.1.: List of datasets used in the analysis for the 2016preVFP data taking period.

Dataset	# events
Data	
/SingleMuon/Run2016B-ver1_HIPM_UL2016_MiniAODv2-v2/MINIAOD	2 789 243
/SingleMuon/Run2016B-ver2_HIPM_UL2016_MiniAODv2-v2/MINIAOD	158 145 722
/SingleMuon/Run2016C-HIPM_UL2016_MiniAODv2-v2/MINIAOD	67 441 308
/SingleMuon/Run2016D-HIPM_UL2016_MiniAODv2-v2/MINIAOD	98 017 996
/SingleMuon/Run2016E-HIPM_UL2016_MiniAODv2-v2/MINIAOD	90 984 718
/SingleMuon/Run2016F-HIPM_UL2016_MiniAODv2-v2/MINIAOD	57 465 359
Signal MC	
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	90 947 213
/DYJetsToLL_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	79 751 592
/DYJetsToLL_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	88 364 975
/DYJetsToLL_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	42 948 118
Background MC	
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[...]-v1/MINIAODSIM	37 505 000
/WW_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	15 859 000
/WZ_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	7 934 000
/ZZ_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	1 282 000
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[...]-v3/MINIAODSIM	31 024 000
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[...]-v3/MINIAODSIM	55 961 000
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[...]-v1/MINIAODSIM	2 300 000
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[...]-v1/MINIAODSIM	2 300 000

[...] = RunIISummer20UL16MiniAODAPVv2-106X_mcRun2_asymptotic_preVFP_v11

Table B.2.: List of datasets used in the analysis for the 2016postVFP data taking period.

Dataset	# events
Data	
/SingleMuon/Run2016F-UL2016_MiniAODv2-v2/MINIAOD	8 024 195
/SingleMuon/Run2016G-UL2016_MiniAODv2-v2/MINIAOD	149 916 849
/SingleMuon/Run2016H-UL2016_MiniAODv2-v2/MINIAOD	174 035 164
Signal MC	
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	73 859 224
/DYJetsToLL_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	76 450 315
/DYJetsToLL_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	82 259 479
/DYJetsToLL_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	41 908 027
Background MC	
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[...]-v1/MINIAODSIM	43 630 000
/WW_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	15 821 000
/WZ_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	7 584 000
/ZZ_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	1 151 000
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[...]-v3/MINIAODSIM	30 609 000
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[...]-v3/MINIAODSIM	63 073 000
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[...]-v2/MINIAODSIM	2 554 000
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[...]-v1/MINIAODSIM	2 491 000

[...] = RunIISummer20UL16MiniAODv2-106X_mcRun2_asymptotic_v17

Table B.3.: List of datasets used in the analysis for the 2017 data taking period.

Dataset	# events
Data	
/SingleMuon/Run2017B-UL2017_MiniAODv2-v1/MINIAOD	136 300 266
/SingleMuon/Run2017C-UL2017_MiniAODv2-v1/MINIAOD	165 652 756
/SingleMuon/Run2017D-UL2017_MiniAODv2-v1/MINIAOD	70 361 660
/SingleMuon/Run2017E-UL2017_MiniAODv2-v1/MINIAOD	154 618 774
/SingleMuon/Run2017F-UL2017_MiniAODv2-v1/MINIAOD	242 140 980
Signal MC	
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v2/MINIAODSIM	196 329 377
/DYJetsToLL_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v2/MINIAODSIM	78 448 070
/DYJetsToLL_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v2/MINIAODSIM	84 321 311
/DYJetsToLL_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	47 006 742
Background MC	
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[...]-v1/MINIAODSIM	106 724 000
/WW_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	15 634 000
/WZ_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	7 889 000
/ZZ_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	2 706 000
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[...]-v2/MINIAODSIM	5 674 000
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[...]-v2/MINIAODSIM	5 649 000
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[...]-v2/MINIAODSIM	5 674 000
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[...]-v2/MINIAODSIM	5 649 000

[...] = RunIISummer20UL17MiniAODv2-106X_mc2017_realistic_v9

Table B.4.: List of datasets used in the analysis for the 2018 data taking period.

Dataset	# events
Data	
/SingleMuon/Run2018A-UL2018_MiniAODv2-v2/MINIAOD	241 613 524
/SingleMuon/Run2018B-UL2018_MiniAODv2-v2/MINIAOD	119 918 017
/SingleMuon/Run2018C-UL2018_MiniAODv2-v2/MINIAOD	109 986 009
/SingleMuon/Run2018D-UL2018_MiniAODv2-v3/MINIAOD	513 909 894
Signal MC	
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v2/MINIAODSIM	196 626 007
/DYJetsToLL_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	86 443 196
/DYJetsToLL_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	92 267 238
/DYJetsToLL_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8/[...]-v1/MINIAODSIM	44 735 479
Background MC	
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[...]-v1/MINIAODSIM	146 058 000
/WW_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	15 679 000
/WZ_TuneCP5_13TeV-pythia8/[...]-v1/MINIAODSIM	7 940 000
/ZZ_TuneCP5_13TeV-pythia8/[...]-v2/MINIAODSIM	3 526 000
/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[...]-v3/MINIAODSIM	95 833 000
/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[...]-v3/MINIAODSIM	178 756 000
/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[...]-v2/MINIAODSIM	7 749 000
/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[...]-v2/MINIAODSIM	7 956 000

[...] = RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1

C. Software Tools

For the efficient evaluation of the huge amounts of data recorded by the CMS detector, potent software frameworks have to be used. In the presented analysis, multiple software packages are used. The main component for all CMS analyses is the CMS Software Framework (CMSSW). For the evaluation of the ultra legacy Run 2 data, CMSSW in version 10.6 is used. This version is found on GitHub at `https://github.com/cms-sw/cmssw/tree/CMSSW_10_6_X`. Additional software packages making use of the framework provided by CMSSW are utilized in the workflow of this analysis

As a first step in the analysis, a skim of the data provided by the CMS Collaboration is performed with `appa`, to reduce the size of the events to be stored. The code is found at `https://github.com/KIT-CMS/Kappa`. The skimmed datasets are then stored at the local dCMS disk cache for further processing.

The software framework `law` is used for managing the workflow and can be found at `https://github.com/riga/law`. The workflow configuration for all steps after the skimming has been performed is located at `https://gitlab.etp.kit.edu/cverstege/zjet-analysis`. Flat n-tuples are created with the help of `Excalibur` (`https://github.com/KIT-CMS/Excalibur`) and `Artus` (`https://github.com/KIT-CMS/Artus`). Afterwards, histograms are created from the analyzed events by using the tool `Lumberjack` from the `Karma` software package (`https://github.com/dsavoiu/Karma`). Further analysis and plots are created with `Palisade` from the same software package. The used configuration for `Karma` is located at `https://gitlab.etp.kit.edu/cverstege/zjet-postprocessing`.

List of Figures

List of Tables

Bibliography

[1] Juan Rojo et al. „The PDF4LHC report on PDFs and LHC data: Results from Run I and preparation for Run II“. *J. Phys. G* 42 (2015), p. 103103.
DOI: 10.1088/0954-3899/42/10/103103. arXiv: 1507.00556 [hep-ph].

[2] CMS Collaboration. „Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ and constraints on parton distribution functions“. *Eur. Phys. J. C* 77.11 (2017), p. 746.
DOI: 10.1140/epjc/s10052-017-5286-7. arXiv: 1705.02628 [hep-ex].

[3] F. Englert and R. Brout. „Broken Symmetry and the Mass of Gauge Vector Mesons“. *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323.
DOI: 10.1103/PhysRevLett.13.321.

[4] Peter W. Higgs. „Broken Symmetries and the Masses of Gauge Bosons“. *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509.
DOI: 10.1103/PhysRevLett.13.508.

[5] Peter W. Higgs. „Broken symmetries, massless particles and gauge fields“. *Phys. Lett.* 12 (1964), pp. 132–133.
DOI: 10.1016/0031-9163(64)91136-9.

[6] ATLAS Collaboration. „Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC“. *Phys. Lett. B* 716 (2012), pp. 1–29.
DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[7] CMS Collaboration. „Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC“. *Phys. Lett. B* 716 (2012), pp. 30–61.
DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[8] Gavin P. Salam, Lian-Tao Wang, and Giulia Zanderighi. „The Higgs boson turns ten“. *Nature* 607.7917 (2022), pp. 41–47.
DOI: 10.1038/s41586-022-04899-4. arXiv: 2207.00478 [hep-ph].

[9] Oliver Brüning and Lucio Rossi, eds. „The High Luminosity Large Hadron Collider: the new machine for illuminating the mysteries of Universe“. Vol. 24. 2015. ISBN: 978-981-4675-46-8.
DOI: 10.1142/9581.

[10] John Campbell, Joey Huston, and Frank Krauss. „The Black Book of Quantum Chromodynamics: A Primer for the LHC Era“. Oxford University Press, Dec. 2017. ISBN: 9780199652747.
DOI: 10.1093/oso/9780199652747.001.0001.

[11] Klaus Rabbertz. „Jet Physics at the LHC: The Strong Force beyond the TeV Scale“. Vol. 268. Springer Tracts in Modern Physics. Berlin: Springer, 2017. ISBN: 978-3-319-42113-1.
DOI: 10.1007/978-3-319-42115-5.

[12] Roger Wolf. „The Higgs Boson Discovery at the Large Hadron Collider“. Springer, 2015. ISBN: 978-3-319-18512-5.
DOI: 10.1007/978-3-319-18512-5.

[13] Particle Data Group Collaboration. „Review of Particle Physics“. *PTEP* 2022 (2022), p. 083C01.
DOI: 10.1093/ptep/ptac097.

[14] „Standard Model of Elementary Particles“. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (visited on 10/12/2022).

[15] Murray Gell-Mann. „Symmetries of baryons and mesons“. *Phys. Rev.* 125 (1962), pp. 1067–1084.
DOI: 10.1103/PhysRev.125.1067.

[16] A. Gehrmann-De Ridder et al. „Precise QCD predictions for the production of a Z boson in association with a hadronic jet“. *Phys. Rev. Lett.* 117.2 (2016), p. 022001.
DOI: 10.1103/PhysRevLett.117.022001. arXiv: 1507.02850 [hep-ph].

[17] Radja Boughezal et al. „Z-boson production in association with a jet at next-to-next-to-leading order in perturbative QCD“. *Phys. Rev. Lett.* 116.15 (2016), p. 152001.
DOI: 10.1103/PhysRevLett.116.152001. arXiv: 1512.01291 [hep-ph].

[18] G. C. Wick. „The Evaluation of the Collision Matrix“. *Phys. Rev.* 80 (1950), pp. 268–272.
DOI: 10.1103/PhysRev.80.268.

[19] R. P. Feynman. „Space - time approach to quantum electrodynamics“. *Phys. Rev.* 76 (1949). Ed. by L. M. Brown, pp. 769–789.
DOI: 10.1103/PhysRev.76.769.

[20] Yuri L. Dokshitzer. „Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.“ *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[21] V. N. Gribov and L. N. Lipatov. „Deep inelastic e p scattering in perturbation theory“. *Sov. J. Nucl. Phys.* 15 (1972), pp. 438–450.

[22] Guido Altarelli and G. Parisi. „Asymptotic Freedom in Parton Language“. *Nucl. Phys. B* 126 (1977), pp. 298–318.
DOI: 10.1016/0550-3213(77)90384-4.

[23] NNPDF Collaboration. „Parton distributions from high-precision collider data“. *Eur. Phys. J. C* 77.10 (2017), p. 663.
DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[24] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. „The anti-k_t jet clustering algorithm“. *JHEP* 04 (2008), p. 063.
DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[25] „Where did it all begin?“ URL: https://home.web.cern.ch/about/who-we-are/our-history (visited on 08/31/2022).

[26] „CERN Annual Personnel Statistics 2021“. 2021. URL: https://cds.cern.ch/record/2809746.

[27] „LHC Design Report Vol.1: The LHC Main Ring“ (June 2004). Ed. by Oliver S. Bruning et al.
DOI: 10.5170/CERN-2004-003-V-1.

[28] „Linear accelerator 2“. URL: https://home.cern/science/accelerators/linear-accelerator-2 (visited on 08/31/2022).

[29] „Linear accelerator 4“. URL: https://home.cern/science/accelerators/linear-accelerator-4 (visited on 08/31/2022).

[30] „CERN's accelerator complex“. URL: https://home.cern/science/accelerators/accelerator-complex (visited on 08/31/2022).

[31] Esma Mobs. „The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018“. General Photo. 2018. URL: https://cds.cern.ch/record/2636343.

[32] Jorg Wenninger. „Operation and Configuration of the LHC in Run 2“. 2019. URL: https://cds.cern.ch/record/2668326.

[33] CMS Collaboration. „Interactive Slice of the CMS detector“. 2016. URL: https://cds.cern.ch/record/2205172.

[34] Izaak Neutelings. June 2017. URL: https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates (visited on 09/06/2022).

[35] CMS Collaboration. „The CMS Experiment at the CERN LHC“. *JINST* 3 (2008), S08004.
DOI: 10.1088/1748-0221/3/08/S08004.

[36] CMS Collaboration. „CMS, the magnet project: Technical design report“ (May 1997).

[37] „CMS Technical Design Report for the Pixel Detector Upgrade“ (Sept. 2012). Ed. by David Aaron Matzner Dominguez et al.
DOI: 10.2172/1151650.

[38] CMS Collaboration. „Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in *pp* Collisions at $\sqrt{s} = 7$ TeV“. *JINST* 8 (2013), P09009.
DOI: 10.1088/1748-0221/8/09/P09009. arXiv: 1306.2016 [hep-ex].

[39] CMS Collaboration. „ECAL 2016 refined calibration and Run2 summary plots“. CMS Detector Performance Summary CMS-DP-2020-021. 2020.

[40] „The CMS hadron calorimeter project: Technical Design Report“ (1997).

[41] CMS Collaboration. „The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC“. *JINST* 8 (2013), P11002.
DOI: 10.1088/1748-0221/8/11/P11002. arXiv: 1306.6905 [physics.ins-det].

[42] CMS Collaboration. „Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV“. *JINST* 13.06 (2018), P06015.
DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[43] „CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger“ (Dec. 2002). Ed. by P. Sphicas.

[44] „Public CMS Luminosity Information“. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults (visited on 09/12/2022).

[45] CMS Collaboration. „The CMS trigger system“. *JINST* 12.01 (2017), P01020.
DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[46] CMS Collaboration. „Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV“. *JINST* 15.10 (2020), P10017.
DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[47] CMS Collaboration. „Particle-flow reconstruction and global event description with the CMS detector“. *JINST* 12.10 (2017), P10003.
DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[48] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. „FastJet User Manual“. *Eur. Phys. J. C* 72 (2012), p. 1896.
DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[49] CMS Collaboration. „Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV“. *JINST* 12.02 (2017), P02014.
DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[50] Thomas Berger. „Jet energy calibration and triple differential inclusive cross section measurements with Z ($\rightarrow\mu\mu$) + jet events at 13 TeV recorded by the CMS detector“. PhD thesis. Karlsruher Institut für Technologie (KIT), 2019. 139 pp.
DOI: 10.5445/IR/1000104286.

[51] Matthias Jochen Schnepf. „Dynamic Provision of Heterogeneous Computing Resources for Computation- and Data-intensive Particle Physics Analyses“. PhD thesis. Karlsruher Institut für Technologie (KIT), 2022. 129 pp.
DOI: 10.5445/IR/1000143165.

[52] M. J. French et al. „Design and results from the APV25, a deep sub-micron CMOS front-end chip for the CMS tracker“. *Nucl. Instrum. Meth. A* 466 (2001). Ed. by T. Ohsugi et al., pp. 359–365.
DOI: 10.1016/S0168-9002(01)00589-7.

[53] „CMS Silicon Strip Performance Results 2016“. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/StripsOfflinePlots2016 (visited on 10/25/2022).

[54] CMS Collaboration. „Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS“. *Eur. Phys. J. C* 81.9 (2021), p. 800.
DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].

[55] CMS Collaboration. „CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV“. Tech. rep. CMS-PAS-LUM-17-004. Geneva: CERN, 2018.

[56] CMS Collaboration. „CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV“. Tech. rep. CMS-PAS-LUM-18-002. Geneva: CERN, 2019.

[57] Torbjörn Sjöstrand et al. „An introduction to PYTHIA 8.2“. *Comput. Phys. Commun.* 191 (2015), pp. 159–177.
DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[58] 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), p. 079.
DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[59] Rikkert Frederix and Stefano Frixione. „Merging meets matching in MC@NLO“. *JHEP* 12 (2012), p. 061.
DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph].

[60] CMS Collaboration. „Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements“. *Eur. Phys. J. C* 80.1 (2020), p. 4.
DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[61] Kirill Melnikov and Frank Petriello. „Electroweak gauge boson production at hadron colliders through $O(\alpha_s^2)$“. *Phys. Rev. D* 74 (2006), p. 114017.
DOI: 10.1103/PhysRevD.74.114017. arXiv: hep-ph/0609070.

[62] Ryan Gavin et al. „FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order“. *Comput. Phys. Commun.* 182 (2011), pp. 2388–2403.
DOI: 10.1016/j.cpc.2011.06.008. arXiv: 1011.3540 [hep-ph].

[63] Ryan Gavin et al. „W Physics at the LHC with FEWZ 2.1“. *Comput. Phys. Commun.* 184 (2013), pp. 208–214.
DOI: 10.1016/j.cpc.2012.09.005. arXiv: 1201.5896 [hep-ph].

[64] Ye Li and Frank Petriello. „Combining QCD and electroweak corrections to dilepton production in FEWZ“. *Phys. Rev. D* 86 (2012), p. 094034.
DOI: 10.1103/PhysRevD.86.094034. arXiv: 1208.5967 [hep-ph].

[65] NNPDF Collaboration. „Parton distributions for the LHC Run II“. *JHEP* 04 (2015), p. 040.
DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[66] Simone Alioli et al. „A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX“. *JHEP* 06 (2010), p. 043.
DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[67] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. „A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction“. *JHEP* 09 (2007), p. 126.
DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

[68] Michal Czakon and Alexander Mitov. „Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders“. *Comput. Phys. Commun.* 185 (2014), p. 2930.
DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].

[69] Michiel Botje et al. „The PDF4LHC Working Group Interim Recommendations“. Jan. 2011. arXiv: 1101.0538 [hep-ph].

[70] A. D. Martin et al. „Uncertainties on alpha(S) in global PDF analyses and implications for predicted hadronic cross sections“. *Eur. Phys. J. C* 64 (2009), p. 653.
DOI: 10.1140/epjc/s10052-009-1164-2. arXiv: 0905.3531 [hep-ph].

[71] Jun Gao et al. „CT10 next-to-next-to-leading order global analysis of QCD“. *Phys. Rev. D* 89.3 (2014), p. 033009.
DOI: 10.1103/PhysRevD.89.033009. arXiv: 1302.6246 [hep-ph].

[72] Richard D. Ball et al. „Parton distributions with LHC data“. *Nucl. Phys. B* 867 (2013), p. 244.
DOI: 10.1016/j.nuclphysb.2012.10.003. arXiv: 1207.1303 [hep-ph].

[73] Emanuele Re. „Single-top Wt-channel production matched with parton showers using the POWHEG method“. *Eur. Phys. J. C* 71 (2011), p. 1547.
DOI: 10.1140/epjc/s10052-011-1547-z. arXiv: 1009.2450 [hep-ph].

[74] Nikolaos Kidonakis. „Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-“. *Phys. Rev. D* 82 (2010), p. 054018.
DOI: 10.1103/PhysRevD.82.054018. arXiv: 1005.4451 [hep-ph].

[75] Nikolaos Kidonakis. „Top Quark Production“. *Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons*. 2014, pp. 139–168.
DOI: 10.3204/DESY-PROC-2013-03/Kidonakis. arXiv: 1311.0283 [hep-ph].

[76] Simone Alioli et al. „NLO single-top production matched with shower in POWHEG: s- and t-channel contributions“. *JHEP* 09 (2009). [Erratum: JHEP 02, 011 (2010)], p. 111.
DOI: 10.1088/1126-6708/2009/09/111. arXiv: 0907.4076 [hep-ph].

[77] P. Kant et al. „HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions“. *Comput. Phys. Commun.* 191 (2015), pp. 74–89.
DOI: 10.1016/j.cpc.2015.02.001. arXiv: 1406.4403 [hep-ph].

[78] M. Aliev et al. „HATHOR: HAdronic Top and Heavy quarks crOss section calculatoR“. *Comput. Phys. Commun.* 182 (2011), pp. 1034–1046.
DOI: 10.1016/j.cpc.2010.12.040. arXiv: 1007.1327 [hep-ph].

[79] A. D. Martin et al. „Parton distributions for the LHC“. *Eur. Phys. J. C* 63 (2009), pp. 189–285.
DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[80] Hung-Liang Lai et al. „New parton distributions for collider physics“. *Phys. Rev. D* 82 (2010), p. 074024.
DOI: 10.1103/PhysRevD.82.074024. arXiv: 1007.2241 [hep-ph].

[81] T. Gehrmann et al. „W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD“. *Phys. Rev. Lett.* 113.21 (2014), p. 212001.
DOI: 10.1103/PhysRevLett.113.212001. arXiv: 1408.5243 [hep-ph].

[82] GEANT4 Collaboration. „GEANT4–a simulation toolkit“. *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303.
DOI: 10.1016/S0168-9002(03)01368-8.

[83] CMS Collaboration. „Performance of the Particle-Flow jet identification criteria using proton-proton collisions at 13 TeV of the ultra-legacy 2016, 2017 and 2018 datasets“. Tech. rep. CMS AN-21-055. 2021.

[84] CMS Collaboration. „Performance of the pile up jet identification in CMS for Run 2“. Tech. rep. CMS-DP-2020-020. 2020.

[85] A. Bodek et al. „Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments“. *Eur. Phys. J. C* 72 (2012), p. 2194.
DOI: 10.1140/epjc/s10052-012-2194-8. arXiv: 1208.3710 [hep-ex].

[86] Stefan Schmitt. „TUnfold: an algorithm for correcting migration effects in high energy physics“. *JINST* 7 (2012), T10003.
DOI: 10.1088/1748-0221/7/10/T10003. arXiv: 1205.6201 [physics.data-an].

[87] Sherpa Collaboration. „Event Generation with Sherpa 2.2“. *SciPost Phys.* 7.3 (2019), p. 034.
DOI: 10.21468/SciPostPhys.7.3.034. arXiv: 1905.09127 [hep-ph].

[88] „Luminosity recommendations for Run 2 analyses“. URL: https://twiki.cern.ch/twiki/bin/view/CMS/LumiRecommendationsRun2 (visited on 10/30/2022).

Acronyms

ALICE A Large Ion Collider Experiment. 13

ATLAS A Toroidal LHC ApparatuS. 13

CERN Conseil européen pour la recherche nucléaire. 1, 11

CHS charged hadron subtraction. 17, 29, 32

CMS Compact Muon Solenoid. 1, 2, 9, 11, 13–17, 19, 20, 25–28, 30, 32–36, 40, 49, 69

DY Drell-Yan. 10, 23, 25, 26, 29, 36, 40, 73

ECAL electromagnetic calorimeter. 15, 17, 34, 35

EW electroweak. 2, 4, 19, 20, 25

HCAL hadronic calorimeter. 15, 17

HLT high level trigger. 16, 25, 27, 28

JEC jet energy correction. 2, 33, 40–42, 49

JER jet energy resolution. 33, 40

L1 Level-1. 16, 27, 34, 40

LHC Large Hadron Collider. 1–3, 6, 7, 11, 13, 14, 50

LHCb Large Hadron Collider beauty. 13

LL leading logarithmic. 43

LO leading order. 7, 10, 27

MC Monte Carlo. 23, 25–27, 29, 30, 32, 33, 35, 36, 40, 43–46, 49, 71

NLO next-to-leading order. 1, 2, 7, 10, 25–27, 29, 30, 40, 43, 46, 49

NNLL next-to-next-to-leading-logarithmic. 26

NNLO next-to-next-leading order. 2, 7, 10, 22, 25–27, 29, 51, 71

PDF parton distribution function. 1, 3, 6, 8–10, 19, 20, 25–27, 46, 50, 71

PF particle-flow. 15, 17, 27–30

puID pileup identification. 30, 32, 35, 40

QCD Quantum Chromodynamics. 1–8, 19, 20, 23, 25, 27

QED Quantum Electrodynamics. 6

SM Standard Model of particle physics. 1, 3, 4, 7, 11, 13, 16, 19, 71