
**Measurement of the Ratio
of inclusive Jet Cross Sections
in Events with and without a high- p_T Z Boson**

Zur Erlangung des akademischen Grades eines

MASTER OF SCIENCE

von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

MASTERARBEIT

von

B.Sc. Bettina Schillinger
aus Welschensteinach

Tag der Abgabe: 27. April 2020

Referent: Priv.-Doz. Dr. Klaus Rabbertz
Korreferent: Asst. Prof. Dr. Mikko Voutilainen

Contents

1	Introduction	3
2	Theoretical Foundations	5
2.1	Quantum Chromodynamics	5
2.1.1	Factorisation Theorem	8
2.1.2	Parton Distribution Functions	9
2.2	Dijet Production	10
2.3	Z Boson Production at Hadron Colliders	12
2.3.1	Significance for Calibration	13
2.4	Dijet versus Z+jet: Differences and Similarities	14
2.5	Fixed-Order Theory Calculations	15
2.6	Monte Carlo Event Generators	16
2.7	Jet Clustering Algorithms	19
3	Experimental Setup	21
3.1	The Large Hadron Collider	21
3.2	CMS Experiment	21
3.2.1	Jet Reconstruction	24
3.2.2	Jet Calibration	26
3.2.3	Trigger	28
4	Cross Section Measurement	29
4.1	Observables of Triple Differential Measurement	29
4.2	General Considerations	32
4.3	Dijet Analysis	36
4.3.1	Trigger Studies	36
4.3.2	Event Selection	39
4.3.3	Comparison to Monte Carlo Simulation and Theory Predictions . .	40
4.3.4	Dijet Cross Section Results	42
4.4	Z+Jet Analysis	46
4.4.1	Z Boson Reconstruction	47
4.4.2	Event Selection	48
4.4.3	Comparison to Monte Carlo Simulation and Theory Predictions . .	48
4.4.4	Z+jet Cross Section Results	49
5	Combined Studies: Cross Section Ratios	57
5.1	Cross Section Ratios: Dijet over Z+jet	57

5.2	Double Ratios: Cross Section Ratio Υ in Data compared to Simulation Results	60
6	Conclusion and Outlook	65
A	Appendix	67
A.1	Data Sets used in the Analysis	68
A.1.1	Data Sets	68
A.1.2	Monte Carlo Simulation Sets	68
A.2	Overview of Run Periods 2018	69
A.2.1	Dijet Cross Sections: Overview of Run Periods 2018	69
A.2.2	Z+jet Cross Sections: Overview of Run Periods 2018	70
A.3	Different MC Generators	71
A.3.1	Dijet Results using Herwig7	71
A.3.2	Dijet Results using Madgraph+Pythia8	72
A.3.3	Z+jet Results using Madgraph+Herwig7	73
A.4	Additional Double Ratio Results	74
A.4.1	MadGraph+Pythia8 for Dijet and Z+jet	74
A.4.2	(MadGraph+)Herwig7 Double Ratio Results	75
A.5	Preliminary Response Matrices	76
A.6	Validation of Dijet Analysis 2017	77
A.6.1	Dijet Cross Section Results	77
A.6.2	The k-Factor	78
A.6.3	Correction Factors	78
A.6.4	Relative Uncertainties	79
A.6.5	Comparison to Simulation	80
B	List of Figures	81
C	List of Tables	83
D	Bibliography	85

Chapter 1

Introduction

Throughout history, scientists were always striving to find the tiniest building blocks of matter. Over the years, different objects, smaller and smaller in size, have been the focus of these searches. Atoms were promising candidates, nicely ordered within the periodic table of elements. But atoms were by far not the end of the journey. Quite the opposite: just as atoms were found to be composed of electrons and a nucleus, with the nucleus itself containing protons and neutrons, deep inelastic scattering experiments disclosed that these nucleons themselves hold even more secrets. The proton is not a fundamental particle, and neither is the neutron, but they both have a substructure. For the proton, in a first rough approximation, this structure consists of its three valence quarks of the flavours *up*, *up* and *down*. And that not being enough. Studying the proton structure at higher energy scales, there was more to discover than just the three valence quarks. The proton also holds a *sea* of strongly interacting quarks, antiquarks and gluons, generically referred to as *partons*. The theory for describing the strong nuclear force governing these interactions is called Quantum Chromodynamics (QCD).

Arising along with these exciting findings, the field of particle physics is exploring the smallest constituents of matter and the interactions between them. Experimentally, particle physics strongly relies on scattering experiments at particle colliders. In the case of this analysis, high energetic proton-proton collisions, where the particles in the final state of the scattering process reveal information on the proton's substructure, are subject of the presented studies.

The most powerful particle accelerator ever built is the Large Hadron Collider (LHC). Operated at unprecedented high energies, the data obtained with this machine provide insights into the nature of the subnuclear world, that have not been accessible before. Data recorded from those collisions offer a solid basis for precision measurements.

The presented analysis uses data collected with the Compact Muon Solenoid (CMS) detector, located at the LHC, in 2018.

In this thesis, the investigation of the cross section ratio of two established processes occurring in proton-proton collisions is performed. This choice is motivated by the fact, that the comparison of two separate measurements in the form of cross section ratios promises to benefit from reduced uncertainties.

The first observable of interest is the inclusive dijet cross section. Dijet events are occurring at high rates at the LHC, due to the abundant production of jets at hadron colliders. Investigation of the dijet spectrum allows for tests of QCD, as the particles within the

jets follow the laws of the strong nuclear force. Furthermore, these QCD events are the background for many other physics searches, and consequently have to be precisely understood in order to increase the accuracy of other measurements. Another advantage of the dijet event topology is the fact that it covers a wide kinematic range.

The second process of choice is the inclusive Z+jet cross section. The Z boson has been studied in various experiments and its properties are very well known. This process allows for precise measurements, because it leaves a clear signal in the detector. Together with the leading jet in the Z+jet event, this final state can be studied in an analogous manner as the dijet final state.

The comparison of the dijet and the Z+jet topology is interesting due to the fact that their production processes involve different partonic subprocesses. While some partonic scattering processes contribute to both dijet and Z+jet production, others, present in dijet already in leading order perturbation theory, only occur in higher orders for the Z+jet process. The comparison of these processes with different contributing subprocesses can be used to further constraint the proton structure.

The cross section measurements for dijet and Z+jet are carried out triple-differentially. The chosen observables are the average transverse momentum $p_{T,avg}$ of the two leading objects in the final state of the interaction, as well as the average boost y_b and the rapidity separation y^* of the final state objects. Through the binning in $p_{T,avg}$, both processes can be treated equally. Furthermore, the y_b - y^* -binning allows for separating effects caused by the proton structure from effects related to the hard parton scattering. These individual measurements are then set into relation via cross section ratios. Common systematic uncertainties of the dijet and Z+jet cross section measurement are expected to thereby partly cancel.

The presented thesis is structured as follows. First, in chapter 2 the theoretical foundations of QCD and the two processes, dijet and Z+jet, subject of the presented studies, are introduced and explained. Following this, in chapter 3 the focus lies on the experimental setup at CMS and the reconstruction of physics objects. Chapter 4 presents the basics of the triple-differential cross section measurement, followed by the individual analyses of the inclusive dijet and Z+jet cross section measurements, in section 4.3 and 4.4, respectively. Finally, in chapter 5, the results obtained for the dijet and Z+jet spectra are combined into an analysis of cross section ratios.

Chapter 2

Theoretical Foundations

This chapter describes the theoretical framework forming the basis for the research carried out in the presented analysis. In addition to the basic underlying theory, important tools for simulation are introduced.

2.1 Quantum Chromodynamics

In the standard model of particle physics (SM), the smallest building blocks of hadronic matter are *quarks* and *gluons*. Together, they are referred to as *partons*, a term introduced by R. Feynman in 1968 [1]. The fundamental force governing the interactions among these partons is the strong nuclear force. Quantum Chromodynamics (QCD) is the theory describing this force. Its interaction constant is commonly referred to as α_s , which corresponds to $\alpha_s = \frac{g_s^2}{4\pi}$, where g_s is the strong interaction coupling constant, occurring in the Lagrangian of QCD. The particle mediating this force is the *gluon*.

A summary of the particles in the standard model can be found in figure 2.1. Quarks are fermions with spin $\pm\frac{1}{2}$; they come in six different flavours, up, down, charm, strange, top and bottom. As indicated in figure 2.1, quarks are carrying electric charge $+\frac{2}{3}$ for up, charm and top and $-\frac{1}{3}$ for the down-type quarks. Additionally, quarks carry colour charge, the charge of the strong interaction. There are three different colours present in the theory. Their anti-particles correspondingly carry anti-colours. Also the massless spin-1 gluons carry colour charge, making them self-interacting in opposite to the photon in electrodynamics, which has electric charge zero and only interacts with electrically charged particles. The gluon occurs in eight versions, corresponding to the eight Gell-Mann matrices, which are the traceless hermitian generators of the special unitary group SU(3). These eight matrices emerge in quantum field theory (QFT), when describing the strong interaction in the picture of the SU(3) symmetry group.

The Lagrangian (rewritten version from the one presented in ref.[3]) describing QCD in quantum field theory reads as follows:

$$\mathcal{L}_{QCD} = \sum_q \bar{\psi}_{q,a} (i\gamma^\mu \partial_\mu \delta_{ab} - m_q \delta_{ab}) \psi_{q,b} - g_s \left(\bar{\psi}_{q,a} \left(\gamma^\mu t_{ab}^C A_\mu^C \right) \psi_{q,b} \right) - \frac{1}{4} F_{\mu\nu}^A F^{\mu\nu A} \quad (2.1)$$

In this description, the Einstein summation convention is applied, i.e. doubled indices are summed over. As usual, γ^μ are the Dirac γ -matrices. The spinors $\psi_{q,a}$ represent the

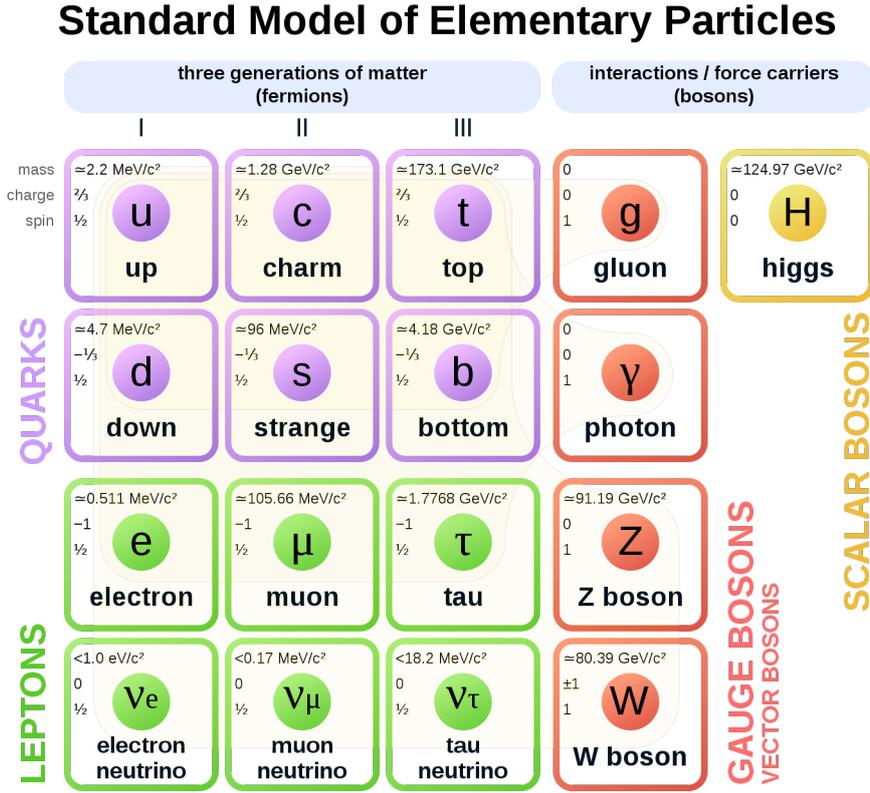


Figure 2.1: The particles making up the standard model. Figure taken from [2]

quark-fields, where the index q stands for the quark flavour and a is the colour-index, running from $a = 1$ up to $a = N_C = 3$, the number of colours in the theory. The first term of the sum in equation 2.1 refers to the Dirac equation for the quarks; it is their equation of motion, being fermions. In this context, m_q represents the mass of the corresponding quark. The middle part with the coupling g_s describes the quark-gluon interactions in QCD. The gluon fields are indicated by A_μ^C and the t_{ab}^C refer to the eight Gell-Mann matrices, generators of the SU(3). In both these quantities, the index C runs from 1 to $N_C^2 - 1 = 8$, over the gluon index. Finally, the last part with the field tensors $F_{\mu\nu}^A$ contains the gluon self-interactions and its propagation. The field tensor written out can be seen in equation 2.2, containing the structure constants f_{ABC} of the SU(3) group, see the commutator below. [3]

$$F_{\mu\nu}^A = \partial_\mu A_\nu^A - \partial_\nu A_\mu^A - g_s f_{ABC} A_\mu^B A_\nu^C \quad (2.2)$$

$$[t^A, t^B] = i f_{ABC} t^C \quad (2.3)$$

The "coupling constant" α_s in the theory of QCD is actually not a constant, but a running value depending on the energy scale ($\mu_R \simeq Q^2$) where it is probed at.

$$\mu_R^2 \frac{d\alpha_s}{d\mu_R^2} = \beta(\alpha_s) = -\left(b_0\alpha_s^2 + b_1\alpha_s^3 + b_2\alpha_s^4 + \dots\right) \quad (2.4)$$

Determined at one scale, the strong coupling constant α_s can be expanded in a series using renormalisation group equations (RGE, also called beta-function), depending on a parameter μ_R , the so-called *renormalisation scale*. The individual terms in the beta-function are containing different coefficients b_k , where $k+1$ is referred to as the loop-order of the term. The leading term in the beta-function is called 1-loop β -function coefficient, reading $b_0 = (11C_A - 4n_f T_R)/(12\pi)$ [3]. One of the interesting features of Quantum Chromodynamics is the fact, that the beta-function, describing the dependence of the renormalised coupling constant $\alpha_s(\mu_R)$ on the chosen renormalisation scale μ_R , has a negative sign, see equation 2.4 [3]. This phenomenon is the source of asymptotic freedom observed in non-abelian gauge theories, as described in ref. [4]. The current world-average of the strong coupling α_s determined at the energy scale of the squared mass of the Z boson M_Z is:

$$\alpha_s(M_Z^2) = 0.1179 \pm 0.0010 \quad (2.5)$$

The *running of the strong coupling*, as displayed in figure 2.2, gives rise to two key phenomena of QCD:

Confinement Colored particles cannot be observed as free particles. The elementary particles carrying colour, quarks and gluons, can only be observed in colour-neutral combinations (i.e. as mesons or baryons). When the distance between two quarks increases, which corresponds to lowering the energy scale, the field strength rises until it is physically more desirable to create a new quark-antiquark pair out of the field energy than upkeeping the colour-field between the separating quarks. Therefore, at high energies, more and more new particles arise out of the QCD field. This is called fragmentation and gives rise to collimated sprays of particles, referred to as hadronic jets. The occurrence of such hadronic jets is ultimately always followed by a process called *hadronisation*, which means the subsequent formation of colour-neutral particle-clusters that end up building hadrons, based on colour-confinement. The only quark being an exception from this is the top-quark, as it decays before it can become part of a composite particle, i.e. baryon or meson.

Asymptotic Freedom For small distances, which corresponds to high energy scales, the strength of the strong interaction decreases. This has as consequence that, in approximation, the coloured particles can be regarded as nearly free particles. This phenomenon is named *asymptotic freedom*. For this reason, high center-of-mass energies of the scattering process are required for studying quarks and gluons as particles of their own and probing their role in the proton.

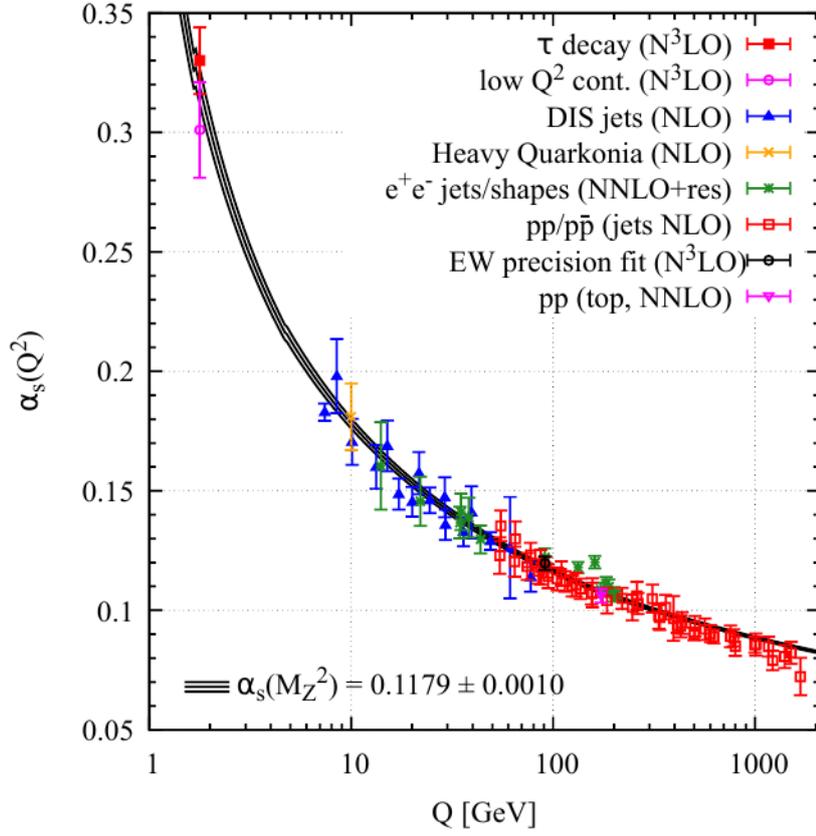


Figure 2.2: The running of the strong coupling constant α_s depending on the energy scale Q . Displayed are several measurements, carried out at different energy scales. For each of the measurements, the order of QCD perturbation theory used for extraction of α_s is noted in the legend. (Taken from ref. [3])

2.1.1 Factorisation Theorem

As can be seen from the formula 2.6 for the QCD cross section [5], the integral can be separated in a part describing the hard interaction (matrix element), here called $\hat{\sigma}$, which is perturbatively calculable, and another part, inaccessible by perturbation theory, corresponding to the parton distribution functions (PDFs) that have to be determined including input from experiments. This possible separation of the cross section formula into a perturbatively accessible and an inaccessible part is known as the factorisation theorem of QCD [6].

$$\sigma = \sum_{i,j} \int \int dx_1 dx_2 f_i(x_1, \mu_F) f_j(x_2, \mu_F) \times \hat{\sigma}_{ij}(x_1, x_2, \alpha_s(\mu_R), \mu_R, \mu_F) \quad (2.6)$$

In the cross section formula 2.6, the two functions f_i and f_j represent the PDFs, introduced in the next subsection 2.1.2. The variables μ_R and μ_F refer to the renormalisation

and factorisation scale, respectively. In principle, these scales can be chosen more or less arbitrarily and are usually set to the same value, $\mu = \mu_R = \mu_F$, often the energy scale of the hard interaction. The renormalisation scale μ_R indicates up to which order the perturbation series in α_s is continued, whereas the factorisation scale μ_F is giving a measure for where to make the separation between hard interaction and parton occurrences within the proton itself. In this picture, a gluon can for instance be part of the proton, described by its parton distribution function, or it could as well be a gluon radiated off from an initial state quark, which comes from the proton. Summation in the cross section formula 2.6 is carried out over all the possible contributing partons. The momentum fraction of proton 1 or 2 carried by the specific parton is indicated with x_1 and x_2 , respectively and integrated over. For the individual protons, after summation over all partons, the integration over the PDF, i.e. $\sum_i \int dx x f_i(x)$, should therefore yield 1.0.

2.1.2 Parton Distribution Functions

The parton distribution functions (PDFs) describe the composition of the proton at different energy scales Q by giving a probability for finding a certain parton type i carrying a fraction x_i of the proton momentum within the proton when probed at a given energy scale Q . Therefore, as shown in equation 2.6, the PDFs depend on this proton momentum fraction x_i and the factorisation scale μ_F . As briefly mentioned in the previous section, the scale choice μ_F decides where to make the cut between seeing a certain soft particle as proton-constituent, by factorising it into the PDF at a given scale μ_F [7], or as initial state radiation, reducing the incoming parton's energy. The PDFs cannot yet be derived from first principles¹, but rely on experimental input and continuous refinement. Obtained at a certain energy scale Q , the PDF corresponding to another energy scale can be derived via evolution with the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations. These are renormalisation group equations for varying the factorisation scale μ_F , at which the PDFs have been determined, to another value. [7]

As a first approximate impression, it may be mentioned, that at lower energy scales Q , the dominant constituents of the proton are its three valence quarks (up, up, down). For higher energies, it is possible to resolve more and more of the proton structure, where sea quarks and gluons start to carry more of the proton momentum fraction.

Some features of the parton PDFs' dependence on the proton momentum fraction x for different parton flavours can be read from figure 2.3, for instance that, at the chosen scale $Q^2 = 100\text{GeV}^2$, the gluon PDF becomes dominant for low values of x , while for values above $x = 0.3$ the valence quarks are dominant, as they carry the main part of the proton momentum at the chosen energy scale. There are several groups providing such PDF sets. The set used to evaluate the theory calculations in this analysis is the CT14NLO [9] PDF set.

For the determination of the PDF sets and their assigned uncertainties, different groups use different approaches. In principle, there are two main ways of how to determine

¹First attempts to do so involve lattice QCD, see for instance ref.[8].

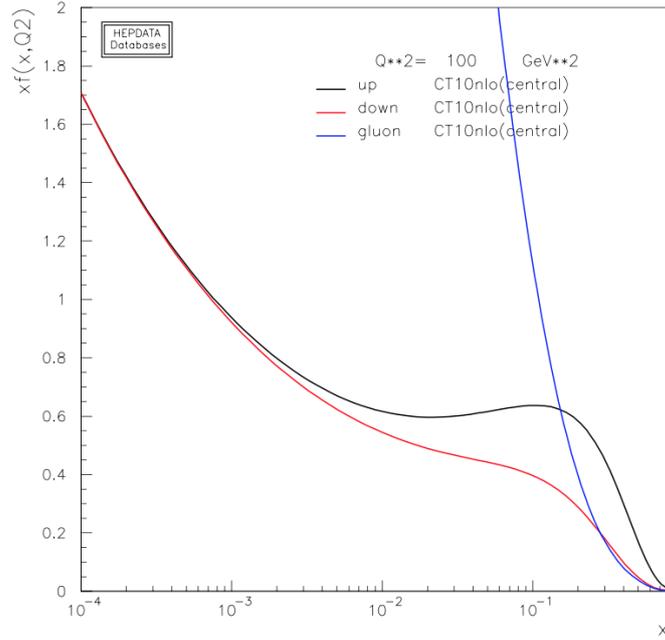


Figure 2.3: An example for the behaviour of different PDFs at an energy scale of $Q^2 = 100 \text{ GeV}^2$. Displayed are the values obtained for the CT10NLO central PDF. The bumps in the up-quark and down-quark curve when going towards higher proton momentum fractions x can be explained via the valence quarks (2x up and 1x down) of the proton. The rise of the up- and down-quark contribution in the lower x -regions are due to the sea quarks. Furthermore, it can be seen how the gluon contribution in the low x region quickly rises. The plots are created using the web tool in ref.[10]. It must be noted, that the PDF plotter tool provided by [10] relies on the LHAPDF5 library, whereas the current library to use is LHAPDF6. Nevertheless, the behaviour of the parton PDFs can be read from this figure.

uncertainties on PDF sets. The first one, which is also used for CT14 PDFs, is the Hessian eigenvector method [11]. The second method relies on uncertainties obtained via a Monte Carlo integration method [12, 13].

2.2 Dijet Production

When colliding two protons, with sufficiently high momentum, in the attempt of gaining insights on the smallest building blocks of matter, the particles actually colliding are the quarks and gluons, together referred to as partons, which make up the proton. Following the laws of QCD, partons in the final state undergo fragmentation and manifest themselves as particle bundles hitting the detector. These collimated sprays of particles, that will eventually undergo hadronisation before hitting the detector, are known as (hadronic) jets. An event topology with well established use in studies of QCD is the so-called dijet event, which, in simplified terms, stands for the $2 \rightarrow 2$ scattering process, and studies the two jets in a proton-proton collision, that carry the highest transverse momentum.

QCD processes like the (inclusive) dijet production have a large cross section and are omnipresent at hadron colliders. This makes them a good study ground for testing and investigating the theory of the strong force, the QCD. These QCD phenomena are not only interesting on their own for the sake of learning more about QCD as a theory itself. Instead, because of their pervasive occurrence, they also have to be studied in order to precisely understand the background for many searches in particle physics.

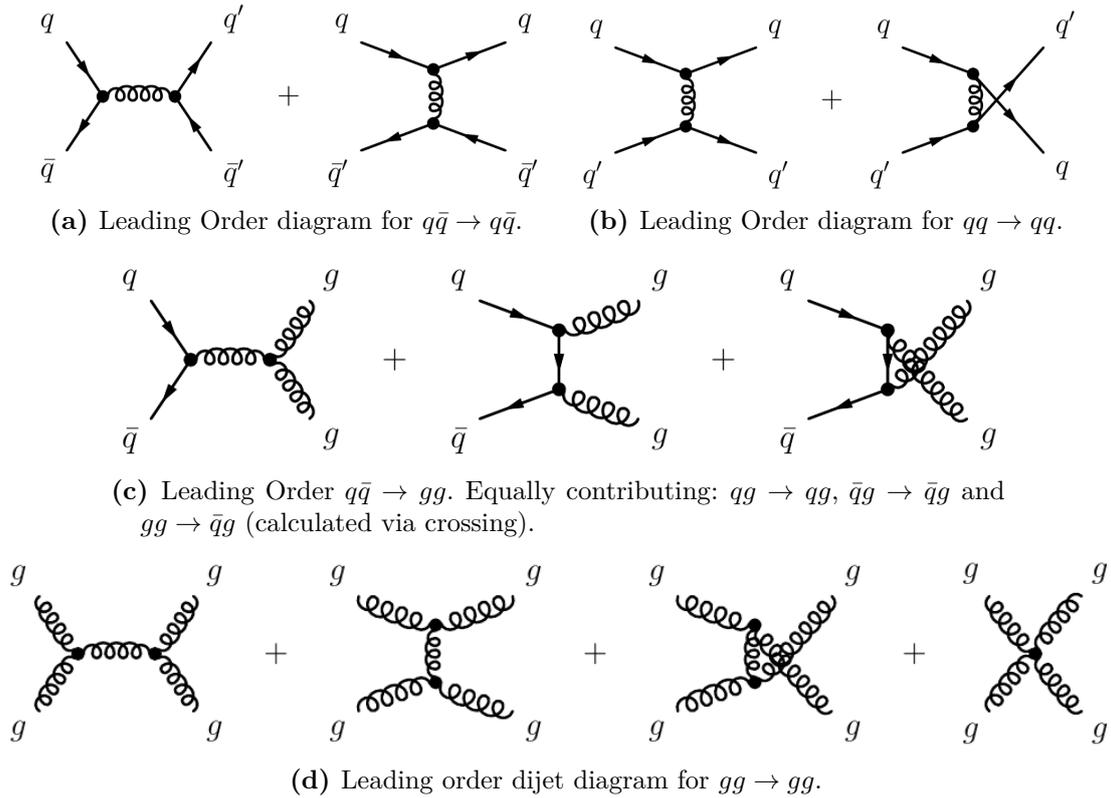


Figure 2.4: Overview of the Feynman diagrams contributing to the dijet cross section in leading order QCD. Subfigures taken from [14].

Figure 2.4 gives an overview of the Feynman graphs for dijet production in $2 \rightarrow 2$ scattering processes contributing to the leading order (LO) ($= \mathcal{O}(\alpha_s)$) cross section. In next-to-leading order (NLO) and next-to-next-to-leading order (NNLO), these graphs are complemented with, for instance, gluon radiation or with diagrams containing a gluon- or quark loop. Therefore, the matrix elements obtained from the graphs in figure 2.4 do not give a full description of the dijet cross section and shall just serve the purpose of offering a first rough impression of the process of interest in terms of its elegant representation via Feynman graphs.

2.3 Z Boson Production at Hadron Colliders

In the standard model of particle physics (SM) the Z boson is the electrically neutral force carrier particle of the weak force in the electroweak (EW) theory. It is a vector boson (spin 1) discovered in 1983 in proton-antiproton collisions using the UA1 detector at the Super Proton Synchrotron (SPS) at CERN [15]. The Z boson's properties have been studied thoroughly in precision measurements, such as the ones at the Large Electron Positron Collider (LEP) [16–18] during the 1990s. Its mass is currently determined to the world average of $m_Z = 91.1876 \pm 0.0021 \text{ GeV}$ [3].

At hadron colliders, one of the possible processes for Z boson production is the annihilation of a quark with an antiquark. Together with the Z bosons' subsequent decay in a lepton and anti-lepton pair, this process is referred to as *Drell-Yan process*. One advantage of this process over the hadronic decay of the Z into $q\bar{q}$ is the cleaner signature of leptons in the detector compared to hadronic jet signals, which are limited by jet momentum resolutions etc. Especially its decay into two opposite-sign charged muons can be measured precisely as these leave a very clean signal in the detector. A corresponding Drell-Yan Feynman diagram is depicted in figure 2.5.

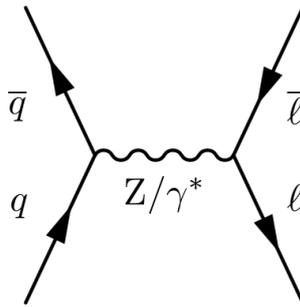


Figure 2.5: The Drell-Yan process drawn as a Feynman diagram. A Z boson is produced through quark-antiquark-annihilation. Afterwards it decays into two oppositely charged leptons. (Drawing taken from ref. [5])

In proton-antiproton collisions, this process can occur through the scattering of a valence quark coming from the proton and a valence antiquark from the antiproton. However, despite being present also in pp-collisions, the Drell-Yan process is not the dominant Z boson production channel here. Since this process is requiring an antiquark in the 2-parton scattering, it must involve a sea quark of one of the protons. The sea quark PDF gets enhanced at higher energy scales. Thus, a high center-of-mass energy of the collision is needed for the Drell-Yan process to occur in pp-collisions. After all, also the partonic center-of-mass energy $\sqrt{\hat{s}}$ has to be sufficiently high to create a Z boson of mass M_Z . In pp-collisions, instead of the Drell-Yan process, the quark-gluon scattering is the dominant process for Z+jet production [19].

The diagrams depicted in figure 2.6 show the dominant leading order (LO) diagrams contributing to the cross section for Z boson production at hadron colliders.

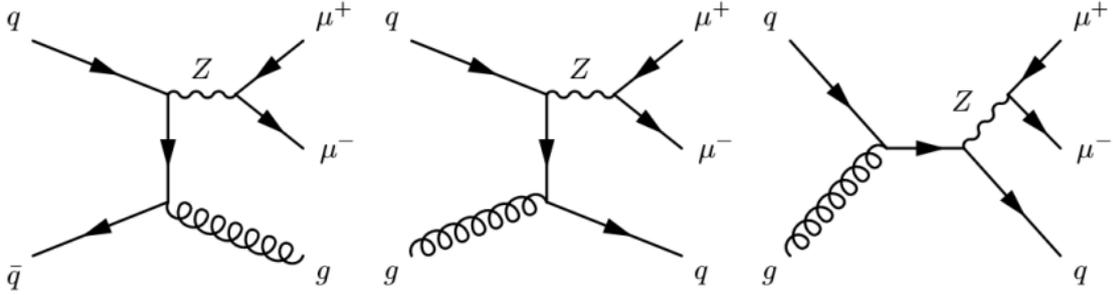


Figure 2.6: Some leading order Feynman diagrams for Z production at hadron colliders like the LHC. By inverting the quark propagator's direction, another three LO diagrams are obtained. (Drawing taken from ref. [20])

2.3.1 Significance for Calibration

As the Z boson is already very well studied and can be measured precisely (depending on its decay channel), the $Z (\rightarrow \mu\mu) + \text{jet}$ decay channel, together with the processes $Z (\rightarrow ee) + \text{jet}$ and $\gamma + \text{jet}$, offers the opportunity to calibrate jets in the medium momentum range ($30 \text{ GeV} < p_T < 800 \text{ GeV}$) by applying the p_T -balance method [21]. Relying on momentum conservation, the outgoing physics objects have to be balanced in their transverse momenta, as the incoming particles have each a transverse momentum of $p_T = \sqrt{p_x^2 + p_y^2} = 0$, being accelerated along the z-axis. The Z boson here serves as the reference object with transverse momentum $p_{T,\text{ref}}$, whereas the jet response is to be determined. This happens by comparing the reconstructed jet transverse momentum $p_{T,\text{jet}}$ to the Z boson's transverse momentum $p_{T,\text{ref}}$. The response in the p_T balance method is defined as follows [21]:

$$R_{\text{jet},p_T} = \frac{p_{T,\text{jet}}}{p_{T,\text{ref}}} \quad (2.7)$$

The second method for investigations on the jet energy response is the so-called missing transverse energy projection fraction method (MPF). The procedure itself is usually carried out including also $\gamma + \text{jet}$ events, but here explained with the Z+jet example. As Z+jet events do not contain missing transverse energy \cancel{E}_T themselves and should be balanced in transverse momentum at parton level, it is possible to state the balancing condition, given in equation 2.8. When looking at the detector level, the quantity R_Z is introduced as the detector response to the Z boson signal and R_{recoil} as the detector response to the hadronic recoil of the Z boson. Equation 2.9 then gives a picture of the amount of missing transverse energy in the Z+jet event.

$$\vec{p}_T^Z + \vec{p}_T^{\text{recoil}} = 0 \quad (2.8)$$

$$R_Z \cdot \vec{p}_T^Z + R_{\text{recoil}} \cdot \vec{p}_T^{\text{recoil}} = -\vec{\cancel{E}}_T \quad (2.9)$$

By reordering equations 2.8 and 2.9 to solve for R_{recoil} and defining this quantity to be the MPF response R_{MPF} , it is:

$$R_{\text{recoil}} = R_Z + \frac{\vec{E}_T \cdot \vec{p}_T^Z}{(p_T^Z)^2} \equiv R_{\text{MPF}} \quad (2.10)$$

Going one step further, assuming the Z response to be approximately 1.0, equation 2.10 simplifies to $R_{\text{MPF}} = 1 + \frac{\vec{E}_T \cdot \vec{p}_T^Z}{(p_T^Z)^2}$.

In good approximation it is valid that the response of the hadronic recoil is equal to the response of the leading jet in the event: $R_{j1} = R_{\text{recoil}}$. Of course this has as condition that the vast amount of recoil is clustered into this jet. For this purpose, secondary jets can be vetoed. Therefore, the balancing is carried out using the leading jet of the event, i.e. the jet with highest transverse momentum.

The MPF method is more robust in terms of systematic biases, but the p_T -balance method is equally important for cross-checking the MPF method and studying the different systematic uncertainties coming with the balancing method [22]. Thus, p_T -balance and MPF method are used complementarily.

2.4 Dijet versus Z+jet: Differences and Similarities

One of the challenges in comparing the two event topologies dijet versus Z+jet, introduced in the previous sections 2.2 and 2.3, is that they have different phase space limitations. While the dijet cross section is large enough to yield high event counts over a wide range of the leading objects' transverse momenta, the Z+jet process' cross section is several orders of magnitudes smaller over the whole kinematic phase space and hence, the probability for Z+jet events to occur with high transverse momenta of the final state objects is lower than for the dijet process. Therefore, a comparison of these processes at the highest transverse momentum values present in the dijet spectrum is limited by the obtained event counts in the Z+jet spectrum and depending on the amount of available data.

Additionally, for the Z+jet final state, the event needs to contain enough energy at parton-level to create a Z boson with mass $M_Z = 91.19$ GeV. Therefore, the number of events fulfilling all requirements decreases when imposing further momentum cuts on the final state objects. This impact of the Z boson mass diminishes with higher momentum values.

Furthermore, the momentum resolution for leptonically decaying Z bosons is better than for hadronic jets, as they can be reconstructed from two opposite charged leptons. In the case of muons, the CMS detector provides an excellent resolution for muons and their momenta, as can be found from the performance report [23] for parts of Run II. Therefore, thanks to their decay into well reconstructable muons, Z bosons leave a very clear and clean signature in the detector, unlike hadronic jets that have to be reconstructed using more complex algorithms and suffer from larger energy scale uncertainties and trigger effects.

Both processes, dijet and Z+jet, are interesting on their own, but another reason for combining these two cross section measurements is, that insights on PDFs can be gained due to the difference in partonic subprocesses that are contributing to the two topologies at various orders in perturbation theory.

In the simplest form, both processes can be described by $2 \rightarrow 2$ scattering diagrams, as indicated in figure 2.4 for dijet and 2.6 for Z+jet production. Both are occurring at hadron colliders and despite their differences, they share some phase space overlap with sufficiently high event counts allowing for combined studies of both processes, as discussed later in this thesis.

Some of the contributing diagrams can in both cases origin from the same two initial state partons, like $q\bar{q}$ scattering for instance. Again some other diagrams cannot. One example for these kinds of diagrams is gg (gluon) scattering. This scattering process strongly contributes to the dijet cross section in the low rapidity bins, which correspond to the central detector area, at average transverse momentum values of around 150 GeV, but also above (see ref. [24]), whereas it does not provide any contribution to Z boson production at leading order. For definition of the commonly used variables like rapidity and transverse momentum see section 3.2.

Furthermore, the two analyses are different in terms of momentum resolution and energy scale uncertainties. But due to the fact that both processes share some systematic uncertainties on the measurement, the common uncertainties can be reduced by taking a cross section ratio, as will be discussed in chapter 5.

2.5 Fixed-Order Theory Calculations

While the structure of the proton is not fully understood, and the PDFs occurring in equation 2.6 cannot yet be determined from first principles, the second part of the cross section formula, the matrix element describing the hard interaction of the process, can be calculated in perturbation theory.

Calculations of the cross sections up to higher orders are very time consuming, wherefore it is not desirable to repeat these. The FASTNLO project [25–27] offers an approach, in which that calculation does not have to be repeated just because of changes of the scale choices μ_R and μ_F , but "look-up" tables of the perturbative coefficients are rather stored right away, as these coefficients are independent from PDFs and the strong coupling α_s . Applying different settings for the scales μ_R and μ_F and using another PDF set, the cross section can be re-calculated much faster with the help of these interpolation tables provided by FASTNLO than by redoing the full calculation.

Providing such tables, FASTNLO is a tool for fast evaluation of cross sections based on pre-computed perturbative coefficients up to next-to-next-to-leading order (NNLO) perturbation theory. The principle idea behind this approach is relying on the factorisation theorem of QCD 2.6, within which it is possible to rewrite the cross section as a convolution of α_s in order n with the hadron PDFs indicated as f_i . The first part of the convolution represents the perturbative coefficients $c_{i,n}$, see ref. [25]:

$$\sigma(\mu_R, \mu_F) = \sum_{n,i} c_{n,i}(x_a, x_b, \mu_F, \mu_R) \otimes [\alpha_s^n(\mu_R) \cdot f_i(x_a, x_b, \mu_F)] \quad (2.11)$$

This convolution 2.11 can be rewritten into a factorised form introducing interpolation functions $e^{(k,l)}(x)$ and $b^{(m)}(\mu)$. The interpolation functions carry scale and proton momentum fraction dependencies and the $c_{n,i}$ are still the perturbative coefficients, as explained above. Here, the factorisation scale μ_F and renormalisation scale μ_R have been set to the same value μ .

$$\sigma(\mu) \simeq \sum_{n,i,k,l,m} \tilde{\sigma}_{n,i,k,l,m}(\mu) \cdot \alpha_s^n(\mu^{(m)}) \cdot f_i(x_a^{(k)}, x_b^{(l)}, \mu^{(m)}) \quad (2.12)$$

$$\text{with } \tilde{\sigma}_{n,i,k,l,m}(\mu) = c_{n,i}(x_a, x_b, \mu) \otimes [e^{(k)}(x_a) \cdot e^{(l)}(x_b) \cdot b^{(m)}(\mu)] \quad (2.13)$$

The $\tilde{\sigma}$ is computed only once and the full cross sections σ with dependencies on α_s and the PDFs can be obtained much faster by applying formula 2.12 instead of a full calculation. For the calculation of the perturbative coefficients tools like NLOJET++ [28] or NNLOJET [29] are used.

The FASTNLO toolkit comes with a wide range of ready-implemented functions for reading out cross sections and related properties. It also provides possibilities to evaluate the tables with different settings for the factorisation and renormalisation scale, μ_F and μ_R . Additionally, it allows for evaluation of the tables with different PDF sets that can be chosen by the user. Corresponding PDF uncertainties and scale uncertainties can be calculated and different orders (LO, NLO, NNLO) of the calculation can be included or excluded when evaluating the tables by switching them on or off.

2.6 Monte Carlo Event Generators

Investigating distributions of data from different perspectives, i.e. looking at different observables, it is essential to have an idea of what to expect. Nowadays, a powerful tool that scientists have at hand are Monte Carlo (MC) event generators, used for simulation of events in high energy scattering processes. The MC sample creation is based on physically motivated probability distributions.

This event generation can be carried out up to different levels of the physics process. Matrix-element (ME) event generators determine only the hard process, i.e. the perturbatively calculable matrix element, and have to be complemented with information from tools simulating parton showering and hadronisation before the resulting event can be passed on to the detector simulation. Multi-purpose MC event generators instead give the opportunity for event simulation, all the way from the initial collision of the protons up to final-state particles with their physical properties, for example mass and momenta. Afterwards, the passage of these final-state particles through the detector can be simulated by detector simulations [30]. A widely used example for such a detector simulation package is GEANT4 [31]. This toolkit is applied to the official Monte Carlo samples at CMS and widely used in high energy physics.

Some central aspects of the steps in the event simulation are briefly introduced in the following paragraphs.

Matrix Element Generation The event generation starts with simulation of the *hard subprocess* in the high-energy proton-proton collision. In this context "subprocess" means the collision of two partons, i.e. constituents of the incoming protons. This simulation is based on quantum mechanics, calculating the matrix element of the process, squaring it to obtain the probability density for this corresponding specific process and finally integrating it [32].

Monte Carlo Weighting As the inclusive jet spectrum is falling steeply, with rising transverse momentum of the leading object, roughly as p_T^{-5} , there would be many more low- p_T events generated and stored than high- p_T events. In data this issue is addressed by reducing the fraction of events selected by the trigger, i.e. applying higher prescales, but proceeding the same way in simulation would mean that an extremely high amount of events would have to be generated in order to reach sufficiently high statistical accuracy. This is a challenge, as gaining higher event numbers by simply calculating more is very expensive in terms of computing resources. In order to make sample creation more efficient, the Monte Carlo production is therefore adjusted (weighted) in a way that relatively more high- p_T events are created than there would be in the natural jet spectrum. Because this results in unphysical distributions, it is important to reweight the events contributing to the cross sections in the analysis process. For this purpose, the applied event weights are stored into the output file of the MC sample and made available for the user. Applied on each of the selected events, the physical jet spectrum is then reobtained.

Parton Shower Modelling After calculating the matrix-elements, it is necessary to account for basic features of QCD. As discussed before, particles that carry colour are undergoing fragmentation and radiate (virtual) gluons, which can themselves split again into quark-antiquark pairs or radiate further gluons. High-energetic proton-proton collisions give rise to abundant production of hadronic jets, as many of the scattering products carry colour charge and are therefore subject to these QCD phenomena. This so-called parton showering of the final state parton after the hard subprocess ($2 \rightarrow 2$ scattering) has to be simulated. In order to account for parton showering, which essentially are corrections to the hard subprocess when going to higher orders, an approximation scheme is applied. In first approximation, such corrections can be described by collinear parton splitting or emission of a soft gluon. For each of the partons taking part in the hard subprocess, incoming and outgoing, such a parton shower is developed. In the course of parton shower simulation, an evolution variable like the virtuality of the shower-partons is important [32].

Hadronisation The parton showers described above are not going on endlessly, as the energies of the involved particles drop with each splitting and at a certain point the shower stops. In simulation this is realised by checking whether the evolution variable

has already fallen below a certain threshold, the so-called hadronisation scale Q_0 . [32] What follows next is the formation of hadrons, called hadronisation. As the energy scale at this point is in a range where α_s is too high for perturbation theory to apply, the hadronisation process cannot be calculated exactly. Instead, models motivated by experimental observations have to be applied.

The two main models in use are the *string* [33] and the *cluster* [34, 35] model.

In the string model, the field strength between a parting quark and antiquark is seen as a gluonic string. Following the laws of QCD, the string will eventually break when the distance between the quark and antiquark increases, creating a new quark-antiquark pair out of the field energy. Each of these new partons is attached via a (now smaller) gluonic string to one of the initial ones. The process repeats, so that the new string segments will break again. At a certain point all of the QCD field energy has been used to produce quark-antiquark pairs, which are connected by only a short string segment. These final objects are then treated as the resulting hadrons.

The cluster model assumes that already during the parton showering, the constituents are clustered in colour-neutral groups. When the energy scale reaches the hadronisation cutoff Q_0 , the remaining gluons in the simulation must split into quark-antiquark pairs. These then participate in forming colourless clusters together with matching colour partners in the shower. Finally, these clusters decay into the final hadrons. [32]

Underlying Event and Pileup Another issue that has to be addressed is the simulation of the underlying event (UE). Basically, UE refers to everything that does not belong to the hard interaction, i.e. everything except for the $2 \rightarrow 2$ parton scattering. As well as in real proton-proton collisions, also when simulating the protons, not only one parton per proton interacts in a hard scattering event, but there are more less-energetic interactions taking place in addition to the "main" collision. This phenomenon is called multi-particle-interaction (MPI).

What occurs additionally in proton-proton collision experiments are so-called pileup events (PU). These result from the fact that when accelerating bunches of protons, instead of colliding only two protons, there are several protons interacting during the same bunch crossing. Another source is PU coming from interactions with protons remaining from a previous bunch-crossing. The PU per simulated event is usually given as an integer number μ referring to the number of pileup interactions in simulation.

In this analysis, the dijet data results are compared to MC simulation spectra obtained from samples created with the multi-purpose MC generators PYTHIA8 [36] and HERWIG7 [37]. The main difference of these two widely used generators lies in the fragmentation models used for modelling the hadronisation. While PYTHIA8 works with *Lund string fragmentation*, HERWIG7 relies on a *cluster fragmentation* model.

The matrix element creation for the MC samples used in the presented Z+jet analysis is carried out using the matrix-element generator MADGRAPH [38, 39]. Nevertheless, for parton showering and hadronisation PYTHIA8 and HERWIG7 are in use also in the Z+jet analysis. For the dijet analysis in chapter 4.3, where this step is done with PYTHIA8,

examples of the results obtained when using MADGRAPH for ME calculation are given in the appendix A.3.2.

2.7 Jet Clustering Algorithms

Having the information from the calorimeters at hand, may it be in real experimental data or from the detector simulation after the above described event generation, jets can be reconstructed. This is done in order to derive information from the measured signals on the original partons in the final state of the scattering interaction, as those are the objects of interest that shall be investigated. There are two major classes of jet reconstruction algorithms. Existing algorithms are categorised into, on one hand, cone algorithms, relying on geometric properties of the input objects and, on the other hand, algorithms using sequential recombination, iteratively combining objects based on calorimetric as well as geometric information. The algorithm mainly used at CMS to reconstruct jets is the *anti- k_T* algorithm [40], based on sequential recombination. The reason for its popularity is the fact that this algorithm copes well with two of the most important issues for jet algorithms: It provides collinear and infrared safety, while at the same time returning circular jets [41]. Collinear safe jet algorithms are robust against jet splittings that fly on into the approximately same direction as the original jet (collinear). Infrared safe algorithms can handle radiation of soft gluons and are not biased by the consequence of the soft emission lowering the original jet energy.

The anti- k_T algorithm has its origins in a generalisation of the formerly present k_T [42] - and the Cambridge-Aachen [43, 44] jet algorithms. It starts out by defining distance measures d_{ij} between objects i and j and d_{iB} between each of these entities and the beamline for the objects in the selection, as well as as a rule for combination of jets.

$$d_{ij} = \min \left(k_{Ti}^{2p}, k_{Tj}^{2p} \right) \cdot \frac{\Delta R_{ij}^2}{R^2} \quad (2.14)$$

$$d_{iB} = k_{Ti}^{2p} \quad (2.15)$$

Here, the distance in the rapidity y and azimuth Φ plane is $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$. R is the jet radius parameter. The variables k_T refer to the transverse momenta of the objects, not named as usually p_T , to avoid confusion with the power p in the exponent. The power p in equations 2.14 and 2.15 is set to $p = -1$ for the anti- k_T algorithm. [40] [41]

In this analysis the focus lies on AK4 jets, which means that the *anti- k_T* algorithm was used to reconstruct these jets and the radius parameter applied is $R = 0.4$.

Chapter 3

Experimental Setup

This chapter starts by describing the experimental setup which forms the basis for the presented analysis. Apart from the experimental setup, methods used for the reconstruction of jets are introduced.

3.1 The Large Hadron Collider

As, due to confinement, the constituents of the protons are hidden within the proton, it is required to achieve sufficiently high collision center-of-mass energies in order to resolve lengthscales below the size of the proton, i.e. below the femtometer scale ($< 10^{-15}$ m). One way to achieve such high energies is circular acceleration. The largest and most powerful particle accelerator that has ever been in use is the Large Hadron Collider (LHC) [45–47] hosted at the European Organization for Nuclear Research (CERN). It is situated near Geneva in a tunnel of 27 km circumference beneath the Franco-Swiss border. The LHC is located in the former Large Electron-Positron Collider (LEP) [16–18] tunnel and was operated at a center of mass energy of maximum 8 TeV in Run I (years 2010-2013), and at 13 TeV in Run II (years 2015-2018). Connected to other acceleration facilities at CERN for pre-acceleration, it is used for accelerating protons or heavy ions. The protons are accelerated as bunches in two separate beam pipes, clockwise in one and counterclockwise in the other, up to the desired beam energy of (in Run II) 6.5 TeV each. At four points around the LHC ring, these protons are brought into collision.

3.2 CMS Experiment

Around the 27 km long accelerator tunnel, there are four big experiments located at the interaction points. The Compact Muon Solenoid (CMS)[48, 49] is one of these main detectors, located in Cessy, France. It is a general purpose detector designed to hermetically cover as much of the 4π phase space as possible.

The coordinate system in the context of the CMS detector is the following right-handed system: The x-axis points towards the center of the LHC ring, the y-axis upwards (towards the sky) and the z-axis points westwards along the beam-axis. An illustration of this orientation is given in figure 3.1.

Some basic properties of the particles commonly used in high energy physics include their transverse momentum p_T , their rapidity y and pseudorapidity η . These quantities

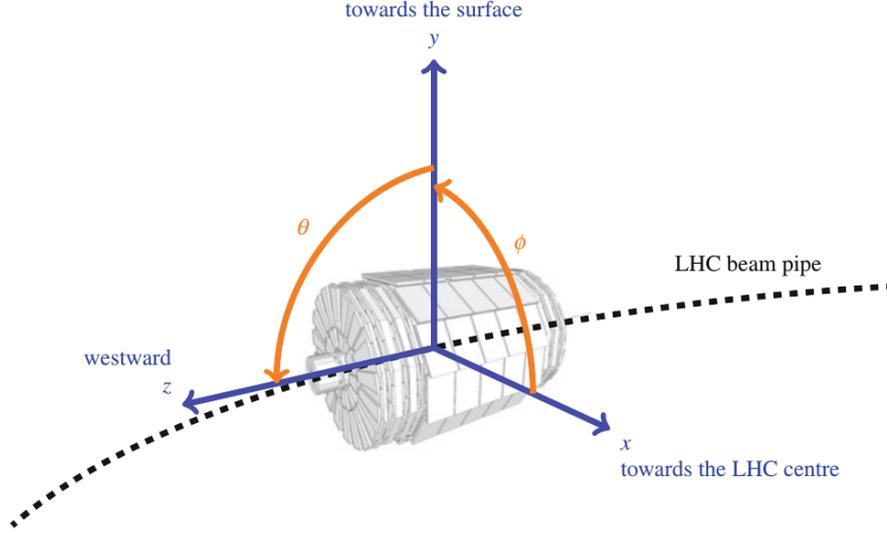


Figure 3.1: Illustration of the coordinate system of the CMS detector. The right-handed system's x-axis points towards the center of the LHC ring, its y-axis upwards and the z-axis along the beamline. (Combined graphics taken from ref. [30]; original illustration included in ref. [5], and background picture from CERN, CMS.)

are given in the following equations 3.1, 3.2 and 3.3. The three spatial components of the particle's momentum are given as p_x , p_y and p_z . The variable E stands for the particle's energy and θ for its scattering angle with respect to the z-axis, i.e. the beamline.

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (3.1)$$

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (3.2)$$

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right) \quad (3.3)$$

The detector systems of the CMS experiment are built as concentric layers around the beamline. In the following, an overview on the main detector parts is given. Figure 3.2 shows a slice of the CMS detector with its subdetectors.

Solenoid The application of a high magnetic field allows for transverse momentum measurement, as charged particles' trajectories get curved passing the magnetic field depending on their momentum, experiencing a Lorentz drift. In addition, the sign of the charge of the particle becomes visible via the direction into which its trajectory is bent. The superconductive solenoid installed in the CMS detector, covering its tracker and nearly the whole calorimetry except for the HCAL tail-catcher, generates a magnetic field of 3.8 T within the solenoid. [51]

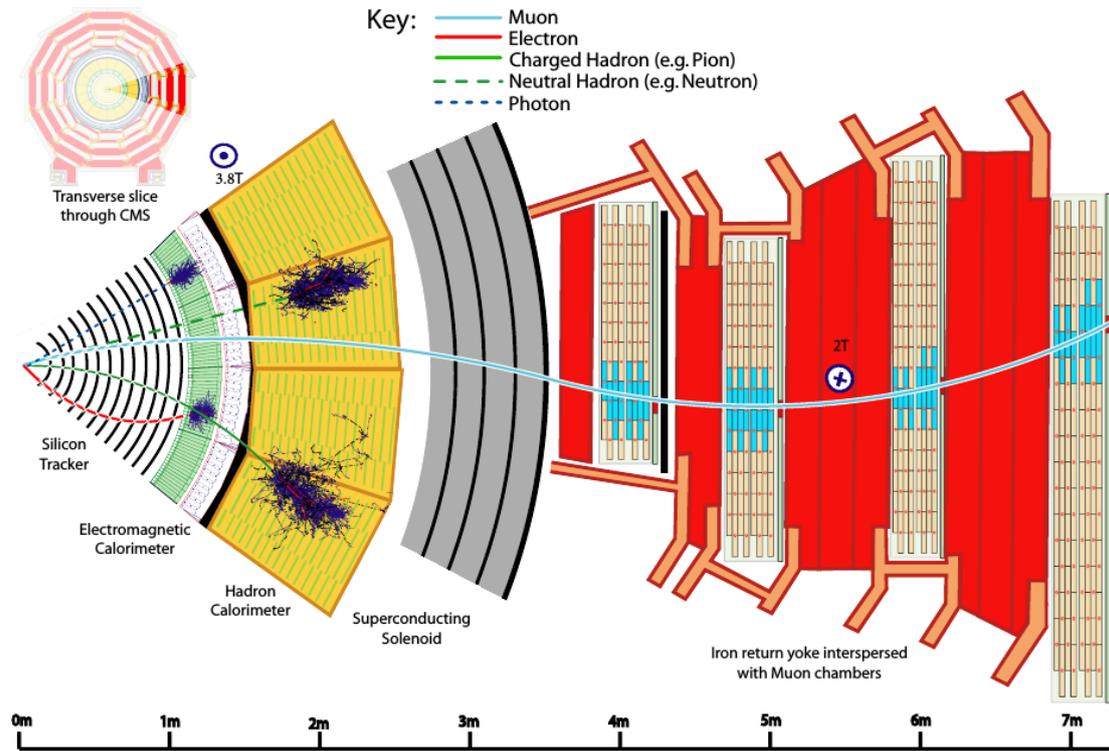


Figure 3.2: The CMS detector and its subsystems. From innermost layer to outermost: the tracker, the ECAL, the HCAL, the return yoke with the muon chambers. The trajectories within the detector are illustrated for three different exemplary particles. As can be seen from the direction of the track bending, the pion and muon are positively charged, while the electron carries negative charge. (Taken from ref. [50])

Tracker The innermost part of the detector, placed around the beamline, is the tracker. It consists of a silicon pixel and silicon strip detector parts. The task of this subdetector is to determine the particles' origin, i.e. the primary interaction vertex or pileup vertices, and to get information on their trajectory. With the combination of signals from the pixel and the strip section, the silicon tracker provides three-dimensional information on the particles' paths through the tracker. Full azimuthal coverage is given up to $|\eta| < 2.5$. [49]

Electromagnetic Calorimeter For measuring the electromagnetic energy carried by the particles hitting the detector, the first calorimeter, the electromagnetic calorimeter (ECAL), is built around the tracker. It is a homogeneous calorimeter made of lead tungstate crystals (PbWO_4) serving as both absorber and active material. Electrons, positrons and photons emerging from the collisions or further decays are stopped in the crystal and deposit their energy into it. Photodiodes in the barrel and -triodes in the endcaps detect the scintillation light. Measured in radiation lengths X_0 , the ECAL

is thicker than $25 X_0$. [49] Including endcap region the ECAL provides coverage up to $|\eta| < 3.0$ [21].

Hadronic Calorimeter For absorption of the hadronic energy of the jets emerging from high energetic hadron collisions, the hadronic calorimeter (HCAL) is the next layer for the particles to pass after the ECAL. The HCAL is a sampling calorimeter, consisting of alternating layers of plastic scintillator to measure the particles' energy and brass or steel to slow them down. [21] This kind of design of the HCAL in the central barrel region and in the endcaps reaches up to $|\eta| < 3.0$. Afterwards, in the pseudorapidity range $3.0 < |\eta| \leq 5.0$, a different technology for the energy measurement is deployed, which is relying on Cherenkov light detection via quartz readout fibres as active material and iron absorbers.

Additionally, located outside the solenoid, a *tail catcher* is installed. It detects high energetic hadronic particles that managed to pass all the previous layers without getting fully absorbed. In terms of hadronic interaction lengths λ_I , the HCAL thickness varies from 7 to 11 λ_I , depending on the direction in η . Including the tail-catcher this extends to 10 to 15 λ_I . [49]

Muon System Muons manage to pass all of the inner detector without getting absorbed or depositing all their energy. Therefore, the muon detection takes place in the outermost subdetector, located in the solenoids' iron return yoke. Drift tubes (barrel) and cathode strip chambers (endcaps), as well as resistive plate chambers, make up the muon system. This subdetector system is measuring the muons' bending angle and energy. [49]

3.2.1 Jet Reconstruction

The collimated sprays of particles emerging from high-energy proton-proton collisions are called *jets*. As described in chapter 2.6 on Monte Carlo simulations, the quarks and gluons in the final state of the collision undergo fragmentation and hadronisation following the laws of QCD, consequently forming hadronic jets. Hitting the detector layers, jets leave energy deposits in the different calorimeters, ECAL and HCAL, depending on their constituents. These different signals are then clustered into jets, in order to reconstruct the original final state physics objects for data analysis.

In the classic approach for jet reconstruction, the calorimetric detector layers are divided into grids consisting of individual calorimeter cells. The energy deposited in each of these cells has to exceed a certain threshold in order for this calorimeter cell to be counted in into the collection. This requirement is called *zero suppression*. It ensures filtering out the signals caused by pure detector noise and thereby prevents these from being counted into the jets. [30]

When relying solely on the information of energy deposits per calorimeter cells for the jet reconstruction, checking the calorimeter "towers", and clustering based on this, the resulting jets are referred to as *calorimeter jets*.

However, a more precise reconstruction method is offered by the particle flow (PF) [50]

approach, taking into account all subdetector systems of the experiment. Jets reconstructed relying on that technique are called *particle flow jets*. These are the sort of jets used in this analysis. The particle flow algorithm is briefly described in the following section.

Particle Flow Algorithm

Instead of dedicating the reconstruction of a certain physics object to separate, individual parts of the detector, using information gained from all of the subdetectors of CMS makes reconstructing the particles more precise. This is the advantage of the so-called particle flow (PF) algorithm [50], which is the approach used for combining the subdetector signals. Thus, instead of simply taking calorimeter towers as physics objects in the jet reconstruction, also tracker information is considered. This way, a more complete picture of the individual particles within the jets is gained.

The PF algorithm distinguishes between the following particle types: photons, electrons and positrons, muons, charged hadrons, neutral hadrons [52, 53]. In order to do so, the tracking information, as well as the energy deposits in both calorimeters ECAL and HCAL and the tracks obtained from the muon system, are all taken into account and combined for reconstructing the individual particles and their path through the detector. The resulting PF candidates are used for jet reconstruction, calculation of the missing transverse energy, and for reconstruction and identification of tau leptons from their decay products [52].

The algorithm starts by reconstructing the charged particles, as these provide also tracker information. The calorimeter clusters that exceed the required minimum energy threshold are then linked to the tracker information. The reconstruction procedure for the individual particle types uses the following information: Muons are reconstructed by checking the hits in the muon tracking system, the muon chambers. Depending on whether this is the only source of information, muons are categorised as either *standalone*, or if extrapolated from inner tracker information complemented with the muon system as *tracker* muons. Or, as the third option, a muon is identified as *global* muon if it is a standalone muon, whose trajectory is matched to a track from the silicon tracker [53, 54]. Electrons and positrons are identified using information obtained from the inner tracker together with data on the particles' energy from the ECAL. For charged hadrons, after removing the already identified particles, the remaining calorimeter information including the HCAL is used and linked to the trajectories gained from the tracker signals.

Afterwards, the neutral particles, being left in the collection, are either identified with photons, if they deposit their energy in the ECAL, or with neutral hadrons, if a corresponding energy deposit has been found in the HCAL [55].

Jets reconstructed using the Particle Flow (PF) algorithm are superior over calorimeter jets with respect to jet momentum and spatial resolution [30].

In addition to the above described method, the Charged Hadron Subtraction (CHS) algorithm removes hadrons that origin from additional primary vertices instead of the

primary vertex of the main collision from the event. This reduces pileup, especially in the low- p_T region, but does not remove neutral pileup particles. [53, 55]

Level of Reconstruction

The jets in a simulated event can be reconstructed at different levels. These include the *parton-level*, *particle-level* and *detector level* jet. Each level refers to the kind of physics input objects (4-vectors) that are given to the jet clustering algorithm: At parton-level, the jets are identified with the partons in the final state of the hard interaction, i.e. quarks and gluons. As the name suggests, particle-level jets already include composite particles, such as kaons, pions, protons or neutrons. Finally, detector level jets are jets clustered from objects obtained relying on detector information. These are reconstructed using information from the calorimeters and the tracker, as described in the previous section. [55]

At CMS, for a jet to be categorised as particle level jet, the particles inside it must have a decay length of $c\tau = 10$ mm or higher, because this is the threshold above which the particles are considered as *stable* [30].

3.2.2 Jet Calibration

Jets, measured and reconstructed from data in an experiment, have to be calibrated in order to have the correct energy scale. Many aspects cause the reconstructed jet to not have the original true jet energy, such as noise from readout-electronics or other disturbances like particles emerging from pileup, as well as underlying event. For this reason, it is important to find the relation between the reconstructed and the true jet energy by determining the jet energy scale [55]. Furthermore, there are differences in the energy scale of experiment and simulation, which have to be identified and corrected for. Other corrections introduced in the standard calibration methods at CMS account for the jet flavour and the jet's size (distance parameter R). [21]

The standard steps in the jet energy calibration can be roughly categorised and identified as follows [30]:

- Offset correction: pileup and electronic detector noise removal
- Response: determining detector jet response at particle level via MC simulation and correcting accordingly
- Residuals: data-based methods accounting for imperfections in detector simulation, i.e. for differences between data and MC
- Optional: MC-based corrections used if correcting for jet flavour

Jet Energy Scale Corrections (JES, also referred to as JEC) are determined depending on the pseudorapidity η and the transverse momentum p_T of the jet. A common graphical representation displaying the workflow of the standard procedure for jet energy scale corrections is given in figure 3.3.

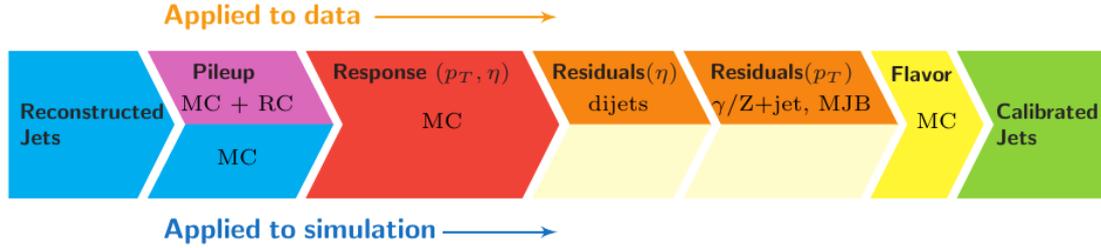


Figure 3.3: The different consecutive steps in the process of jet energy correction (JEC). While some steps are the same for jets in simulated and data events, others differ. In this figure *MC* indicates that the corresponding correction was derived from simulation, similarly the abbreviation *RC* stands for random cone, explained in ref. [21], and *MJB* for corrections gained from analysis of multijet events. The variable p_T stands for dependence of the correction on the transverse momentum of the jet, and η for the correction being dependent on the pseudorapidity of the jet in question. (Illustration taken from ref. [21])

At CMS the JEC is applied by multiplying factors, each accounting for specific effects, onto the reconstructed jet transverse momentum. The corrected transverse momentum p_T^{corr} of the jet can then be obtained as follows (adapted from ref. [55]):

$$p_T^{\text{corr}} = c_{\text{res}}(\eta, p_T'') \cdot c_{\text{mc}}(\eta, p_T') \cdot c_{\text{pileup}}(\eta, \rho, A_j, p_T^{\text{raw}}) \cdot p_T^{\text{raw}} \quad (3.4)$$

In this equation 3.4, the variable A_j stands for the jet area [56] and ρ refers to the average density of transverse momentum in an event. The density ρ is estimated using jet clustering algorithm and gives information on the amount of pileup and underlying event contained in a given event. The quantity which these corrections are applied on is the transverse momentum p_T^{raw} of the raw, uncorrected, jet. The pileup correction factor c_{pileup} depends on this initial value. The subsequent corrections derived from studies on MC samples c_{mc} , and then the residual corrections c_{res} take the output of the respectively previous multiplication as input: After correction of p_T^{raw} for pileup and detector noise effects, the MC based correction, correcting for response effects, takes p_T' as input. After these MC corrections the transverse momentum p_T'' is the input when applying the residual corrections on data, ultimately yielding the corrected transverse momentum p_T^{corr} of the jet.[55]

Accessing the Data

For a collaboration of the size of the CMS experiment, maintaining a structured overview on the data and Monte Carlo sets available for analysis is important. The experimental data as well as centrally created simulation data is stored on different sites around the world. The CMS Data Aggregation System (CMSDAS) offers the possibility to search for available data sets by entering relevant keywords [57].

The framework used for analysing the data presented in this thesis, largely relies on CERN's software toolkit ROOT [58].

A new data format in use at CMS is the *nanoAOD* data format [59, 60]. With introduction of the nanoAOD format the space consumed per event is reduced down to 1-2 kB per event. Being in order of 1kB per event, this is around 20 times smaller than the next larger format miniAOD.[59] There are specific tools designed for certain tasks to be carried out on nanoAOD files, the corresponding nanoAOD-tools git repository can be found in ref. [61]. The jets and their properties stored into the nanoAOD files are already supplied with the latest JEC factors [62]. The presented analysis was carried out relying on this new nanoAOD data format.

3.2.3 Trigger

As already mentioned in the discussion of MC event generators in section 2.6, the cross section of QCD events is much higher than the one of many other processes relevant for analyses carried out by the CMS collaboration. This is a challenge simply in terms of computing resources, because it is impossible to store all the emerging events when searching for rarer processes. Additionally, with proton bunch-crossings every 25 ns, corresponding to a frequency of around 40 MHz, also the event rates are too high to store all events. Therefore, a fast automated procedure is required for a quick decision on what events to store. The responsible system for this task is called *trigger system*.

When deciding which data to store for offline analysis, the required rejection power is too high for the triggering to be carried out in one single step [63]. To solve this issue CMS uses a multilevel trigger system, before ultimately storing the collision event data. The procedure is split into two different steps, the Level-1 (L1) trigger and High-Level-Trigger (HLT). The L1 trigger is a hardware system making decisions based on calorimeter and muon detector information. For this selection there is less than 4 μ s time per event [21]. After this first reduction of data down to potentially interesting events, the second level, the HLT implemented in software, selects events to be stored based on algorithms similar to those used in offline-analyses. Events are reconstructed using the information coming from the L1 and filtered into different HLT paths depending on predefined step-by-step applied selection criteria [64]. This procedure by the HLT reduces the event rate further from 100 kHz to less than 1 kHz before storage [21].

In order to reduce the data rate written out after the HLT even further, so-called *prescaling* is applied. This means that only a certain fraction of the events passing those trigger paths is considered when storing the data. The prescale integer X tells that every Xth event passing this trigger path is collected for storage.

Because triggering plays an important role in analyses involving jets, also this analysis treats the handling of the HLT paths. The trigger efficiencies are evaluated in order to determine when to use which HLT path. The aim is to determine the turn-on point, where the respective trigger path reaches an efficiency higher than 99 %. A more detailed description of the methods applied to determine these values is given in the corresponding dijet analysis chapter 4.3.

Chapter 4

Cross Section Measurement

In this chapter, the two measurements of inclusive jet cross sections, containing either two jets with high transverse momentum in the final state (dijet), or at least one hadronic jet together with a Z boson with high transverse momentum (Z+jet), are introduced and discussed.

4.1 Observables of Triple Differential Measurement

The presented analyses make use of three different observables in order to categorise the events into phase space bins, and measure the cross sections triple-differentially. The first of these three observables is the average transverse momentum $p_{T,\text{avg}}$ of the two leading objects, i.e. the two physics objects carrying the highest transverse momentum. In the dijet analysis $p_{T,\text{avg}}$ therefore refers to the average transverse momentum of the two so-called leading jets, whereas in the Z+jet analysis $p_{T,\text{avg}}$ refers to the average transverse momentum of the Z boson and the leading hadronic jet. The two additional observables of choice are the rapidity based variables y_b and y^* . The definition of y_b and y^* is given in equations 4.2 and 4.3, recalling the definition of a physics object's rapidity and transverse momentum given in equations 3.2 and 3.1.

$$p_{T,\text{avg}} = \frac{1}{2} (p_{T,1} + p_{T,2}) \quad (4.1)$$

$$y_b = \frac{1}{2} |y_1 + y_2| \quad (4.2)$$

$$y^* = \frac{1}{2} |y_1 - y_2| \quad (4.3)$$

Compared to using the rapidity of the two single objects under investigation, these two variables y_b and y^* prove to be advantageous [30, 55]. Choosing these two rapidity-like variables, insights on parton distribution functions (PDFs) and matrix elements can be gained more easily than by looking at the rapidity of the individual objects.

On the one hand, the variable y_b refers to the average boost of the dijet or Z+jet system. As explained in section 2.1.2, PDFs depend on the proton momentum fraction x_i that a parton i is carrying at a specific energy scale. Assuming a $2 \rightarrow 2$ parton scattering, where the initial state partons i and j were carrying the proton momentum fractions x_i and x_j before the hard interaction took place, the average boost y_b of the final state

Abbreviation	y_b range	y^* range
yb0ys0	$0.0 \leq y_b < 1.0$	$0.0 \leq y^* < 1.0$
yb0ys1	$0.0 \leq y_b < 1.0$	$1.0 \leq y^* < 2.0$
yb0ys2	$0.0 \leq y_b < 1.0$	$2.0 \leq y^* < 2.4$
yb1ys0	$1.0 \leq y_b < 2.0$	$0.0 \leq y^* < 1.0$
yb1ys1	$1.0 \leq y_b < 2.0$	$1.0 \leq y^* < 2.0$
yb2ys0	$2.0 \leq y_b < 2.4$	$0.0 \leq y^* < 2.4$

Table 4.1: An overview on the six different y_b - y^* -bins and their definition. The abbreviations shown in the first column are widely used in the labeling of the results presented in this thesis, as they give short and insightful information on the phase space bin.

system gives insights on x_i and x_j relative to each other. For instance, in order to achieve high y_b -values, one parton has to carry a very low x_i , and the one it is colliding with a very high x_j . This way, the observable y_b gives access to the PDFs. This is especially interesting for very high values of y_b , where insights on the gluon-PDF can be gained, for example through quark-gluon scattering.

On the other hand, the variable y^* represents the rapidity separation of the two leading objects in the final state. It is a measure for the scattering angle θ in the center-of-mass frame of the $2 \rightarrow 2$ scattering and therefore gives the opportunity to learn more about the matrix element of the hard interaction, as the angle θ occurs in the cross section calculations.

Binning For the binning of the resulting histograms in this analysis, the three-dimensional phase space has been divided into six y_b - y^* -bins as shown in table 4.1.

The reason for limiting the highest y_b - y^* -bins to a maximum of $y_b < 2.4$ and $y^* < 2.4$ is the tracker coverage. It results from a corresponding cut on the objects' rapidity $|y| < 2.4$. This cut is introduced for the muons in the Z boson reconstruction and is applied analogously on the two jets in the dijet analysis and the jet in the Z+jet event topology.

In figure 4.1 an illustration of the event topologies in the six different y_b - y^* -bins is given. Additionally, it indicates the phase space characteristic of the binning. The horizontal line in each of the subfigures can be seen as the beamline in the experiment (z-axis). The cones represent the two leading objects, here two jets for the dijet analysis. In the Z+jet analysis one of the cones is replaced by an outgoing Z boson. Differently from this analysis, the bins in figure 4.1 are displayed as going up until $y_b = 3.0$ and $y^* = 3.0$, but in the case of the present analysis the outermost y_b - y^* -bins are restricted to a maximum of $y_b, y^* < 2.4$. As discussed above, this is done to ensure a good tracker-coverage by applying the $|y| < 2.4$ cut on the leading objects.

The average momentum of the two leading objects in each analysis has been chosen as the third binning variable in this triple-differential cross section measurement. In the case of dijet events, this choice reduces effects of jet switching, i.e. leading jet and second leading jet changing their ordering in p_T due to, for instance, the first jet losing energy by

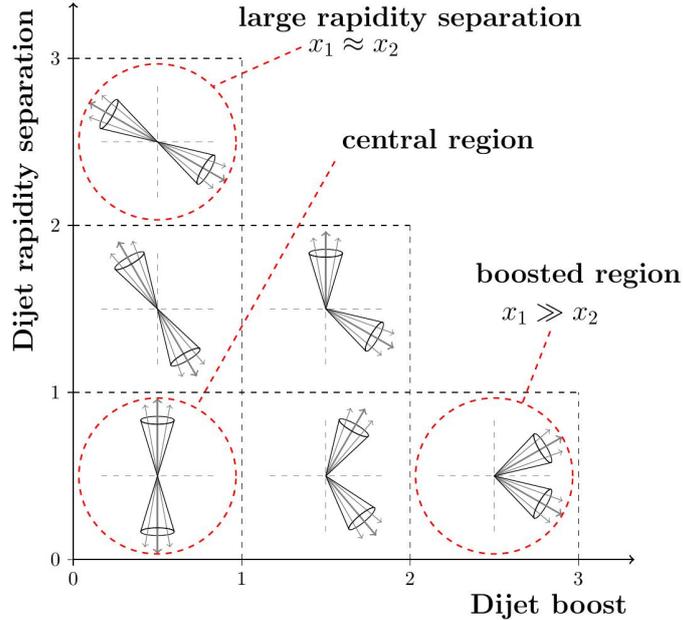


Figure 4.1: Illustration of the y_b y^* phase space. The horizontal line in each of the bins can be seen as the beamline (z-axis). The two cones stand for the two leading jets in the dijet event-topology. The approximate jet configurations are indicated to get a better impression of the phase space categorisation. The same binning is applied for the Z+jet analysis, simply replacing one of the cones with an outgoing Z boson. Instead of the common abbreviations y_b and y^* , the axes are named more descriptively by what the variables y_b and y^* stand for, namely the *dijet rapidity separation* for y^* and the *dijet boost*, usually just called y_b here. In the descriptions, the variables x_1 and x_2 denote the proton momentum fraction carried by the partons taking part in the scattering process. In this analysis, unlike the bin edges shown in the presented figure, the outermost bins are limited to a maximum of $y_b < 2.4$ and $y^* < 2.4$ due to a rapidity cut of $|y| < 2.4$ on both leading objects. (Figure taken from ref. [55].)

radiation. Additionally, the $p_{T,\text{avg}}$ binning helps with treating the two different processes, dijet and Z+jet, as similarly as possible, taking both leading objects into account in the binning. The bins in terms of $p_{T,\text{avg}}$ are used as shown in table 4.2.

Effectively, the lower bins up to $p_{T,\text{avg}} = 56$ GeV remain empty in the dijet- and ratio part (see chapters 4.3 and 5) of this analysis, which, as later explained, happens due to the trigger turn-on point of the lowest jet trigger. Similarly, the highest bins stay unfilled as soon as looking at the ratio results, because the Z+jet process' cross section in the high- $p_{T,\text{avg}}$ region is too small to yield sufficiently high event counts at the given luminosity. Thus, dijet limits the phase space from below and Z+jet from above. As a result, the $p_{T,\text{avg}}$ -range under investigation in the ratio studies in chapter 5 ranges from around 60 GeV up to roughly 1 TeV.

Observable	bin edges in GeV
$p_{T,\text{avg}}$	[32, 37, 43, 49, 56, 64, 74, 84, 97, 114, 133, 153, 174, 196, 220, 245, 272, 300, 330, 362, 395, 430, 468, 507, 548, 592, 638, 686, 737, 790, 846, 905, 967, 1032, 1101, 1172, 1248, 1327, 1410, 1497, 1588, 1684, 1784, 1890, 2000, 2116, 2238, 2366, 2500, 2640, 2787, 2941, 3103, 3273]

Table 4.2: The bin boundaries of the $p_{T,\text{avg}}$ binning used in both analyses, dijet and Z+jet, in this thesis. Due to the phase space of the Z+jet process looked at, upper bins, above around 1 TeV remain empty. The single-jet trigger with the lowest nominal value in the dijet analysis reaches 99% efficiency only at around 50 GeV. Therefore, all bins below 56 GeV stay empty in the dijet, and the combined, analysis.

4.2 General Considerations

This section discusses aspects and procedures that are concerning both analyses, dijet and Z+jet.

Triple Differential Measurement The cross sections for the processes investigated in this thesis are measured triple-differentially. For this purpose, the three chosen observables are constructed from properties of the two leading objects. The chosen observables are the average transverse momentum $p_{T,\text{avg}}$, the average boost of the system y_b and the rapidity separation y^* of the two leading objects, as explained in section 4.1. The corresponding differential cross section therefore is defined as follows:

$$\frac{d^3\sigma}{dp_{T,\text{avg}} dy_b dy^*} = \frac{1}{\epsilon \mathcal{L}_{\text{eff}}^{\text{int}}} \cdot \frac{N}{\Delta p_{T,\text{avg}} \Delta y_b \Delta y^*} \quad (4.4)$$

Here, the efficiency ϵ corresponds to the trigger-efficiency which is set to be unity. This is justified, considering that a trigger path is only used in case its efficiency in the specific $p_{T,\text{avg}}$ -bin is greater than 99%. When a trigger reaches 99% efficiency in a $p_{T,\text{avg}}$ -bin k , the next higher $p_{T,\text{avg}}$ -bin $k + 1$ is the first one filled with events coming from this trigger path. For this reason, the efficiency of the trigger can be assumed to be very close to unity in the bins where it is used. The variable N represents the event count observed in a specific $p_{T,\text{avg}}-y_b-y^*$ -bin after all selection cuts have been applied. The effective luminosity \mathcal{L}_{eff} is calculated for each of the HLT paths individually, using the BRILCALC [65] tool provided by the CMS collaboration. In the cross section calculation, the corresponding bin counts N are then divided by the integrated luminosity determined for each of the trigger paths. This procedure also accounts for the prescale values (see section 3.2.3) of the triggers with lower nominal p_T -threshold.

The variables $\Delta p_{T,\text{avg}}$, Δy_b and Δy^* correspond to the bin widths introduced in 4.1 and displayed in table 4.1 and 4.2. For each of the y_b-y^* -bins, the corresponding cross section value in the histogrammed spectrum is divided by the bin width of the $p_{T,\text{avg}}$ bin in GeV. Additionally, the outermost rapidity bins in the y_b-y^* -binning have a size of 0.4 and therefore the cross section has to be divided by 0.4 in these bins. For the other four

y_b-y^* -bins the bin width is 1.0. In those cases equation 4.4 holds already in a simplified form, only dividing by $\Delta p_{T,\text{avg}}$ and \mathcal{L}_{eff} .

Data Sets The data sets used in this analysis correspond to the data taken in year 2018 at CMS during Run II of the LHC. In a first comparison of the run periods A, B, C and D, these sets were found to be compatible within statistical uncertainties in most of the bins, see appendix A.2. However, the result plots displayed in the following chapters refer to run period D, because this data set provides the highest integrated luminosity, and hence the smallest statistical uncertainties.

As explained in section 3.2.2, the latest available JES corrections, with the exception of the optional flavour corrections, are applied to the jets stored in the nanoAOD format, which are used in this thesis. Flavour corrections are not applied to the data sets used in the presented analysis. The data is examined at detector-level, and no unfolding is performed. However, response matrices based on the Pythia8 MC sample used in the dijet analysis have been derived in the scope of this thesis and are shown in appendix A.5. These exemplary response matrices serve to give a first impression of migration effects in the $p_{T,\text{avg}}$ spectrum and suggest that unfolding of the dijet spectrum is possible without the need for regularisation.

Run period	Integrated Luminosity in fb ⁻¹	data set for dijet	data set for Z+jet
A	14.00	JetHT	DoubleMuon
B	7.10	JetHT	DoubleMuon
C	6.94	JetHT	DoubleMuon
D	31.93	JetHT	DoubleMuon

Table 4.3: The four run periods of the data taking period 2018. All four sets have been analysed and found to mostly agree within statistical uncertainties. The presented results all refer to Run D. This set has the highest integrated luminosity, which reduces statistical uncertainties on data. The full data set names can be found in the appendix A.2. The luminosity values are taken from ref. [66].

Systematic Uncertainties In the 2018 data taking period, the uncertainty on the luminosity is determined as $\pm 2.5\%$ [66] and identical for both of the individual analyses, dijet and Z+jet.

There are 26 different uncertainty sources that independently contribute to the overall JES uncertainty. A table listing all of these individual JES uncertainty sources grouped into categories is given in ref. [21]. Each of these sources is fully correlated in p_T and η , but uncorrelated with respect to every other source [55]. The contributions of some JES uncertainty sources are significantly larger than others. As an example, one of the dominant uncertainty sources for dijet is identified to be the one connected to the flavours involved in the final state. For dijet this category is represented by the *jesFlavorQCD* uncertainty source, and for Z+jet by *jesFlavorZJet*.

The recommended tool for determining the JES uncertainties is the module `JETMETHelper-RUN2.PY` provided within the official `nanoAODtools`. For an introductory description of these tools see ref.[61]. Using this module running over the events in a Monte Carlo set, in this case the Herwig7 nanoAOD set for dijet (QCD sample) and the MadGraph+Herwig7 set for Z+jet ($DYJetsToLL$ sample), the p_T spectra are re-calculated, varying all the $p_{T,jet}$ values in each event of the sample upwards and downwards according to the JES uncertainty source in question. These variations depend on the given p_T and η of the original jet in the central spectrum. For this procedure to yield more accurate estimations of the JES uncertainties, larger MC samples are required. The varied results for the specific dijet or Z+jet sample are stored for each JES uncertainty source separately into the ROOT file. Afterwards, all of these new $p_{T,jet}$ spectra, one for each of the JES uncertainty sources, are processed with the original event selection, which is applied to data in the analysis. This way, in addition to the central distribution, extra histograms are created for the respective variation upwards and downwards for every JES uncertainty source. The deviation from each of these varied spectra to the central spectrum is calculated and used to determine the JES uncertainties caused by the different sources. Added in quadrature, since the individual sources are uncorrelated, these single contributions result into the total JES uncertainty.

The values obtained for the upwards- and the downwards- uncertainty show only small differences. For simplicity, the JES uncertainties are therefore presented in a symmetrised form.

As mentioned above, the JES uncertainties are determined using the Herwig7 QCD sample for dijet and the MadGraph+Herwig7 $DYJetsToLL$ sample for Z+jet. The higher $p_{T,avg}$ -bins in these simulated samples do not always yield sufficiently high event counts for a reliable estimation of the JES uncertainty. This issue is purely caused by statistical effects and can, for example, lead to fluctuations of the obtained JES uncertainty to 0% or 100% in neighbouring bins, which does not represent the physical reality. To avoid such effects, the statistical uncertainties for the (MadGraph+)Herwig7 samples, which are used for JES uncertainty determination, are evaluated in each $p_{T,avg}$ -bin. Consequently, the $p_{T,avg}$ -bins that show a statistical uncertainty higher than $\delta_{MC, stat. unc} = 14\%$ are removed from the analysis. To obtain robust JES uncertainty estimations, this procedure is applied to both the dijet and the Z+jet sample. The threshold of 14% corresponds to removing bins with an effective event count of $N_{eff} < 50$.

As an example, figure 4.2 shows the statistical uncertainties for the three samples used in the dijet analysis in the bin `yb1ys1`. Displayed are the statistical uncertainties on data as a green uncertainty band in the background and the statistical uncertainties on the two MC simulation sets in orange for Herwig7 and blue for Pythia8. If the orange Herwig7 bar crosses the 14% threshold, the $p_{T,avg}$ -bin is removed. The dashed line drawn in red indicates this threshold. In case of the `yb1ys1` bin for example, displayed in figure 4.2, the four highest $p_{T,avg}$ -bins are removed from the final results presented in this chapter. Since there is no statistically motivated cut imposed on the considered (MadGraph+)Pythia8 MC simulations and run D data sets, the higher $p_{T,avg}$ -bins have to be treated with caution as these often come with high statistical uncertainties caused by small event counts also in data. The fact that the statistical uncertainty on data is rising around different

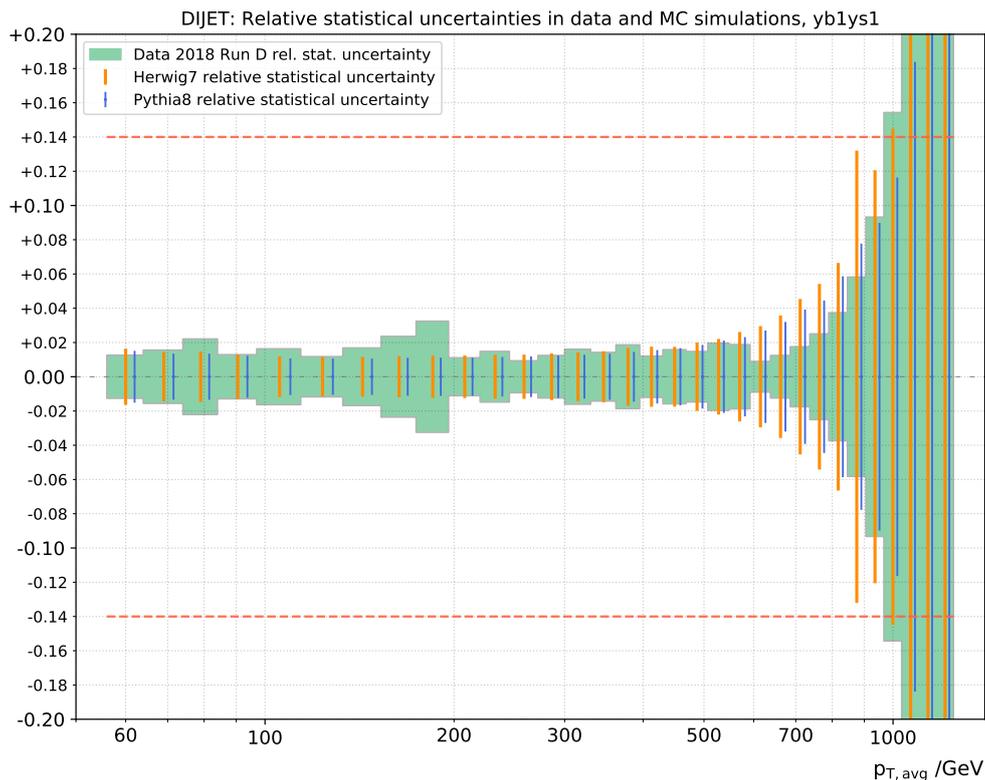


Figure 4.2: Illustration of the statistically motivated $p_{T,\text{avg}}$ -bin removal by taking the example of the yb1ys1 bin in the dijet analysis. Shown are the statistical uncertainties calculated for the data set of run D 2018 and the Pythia8 MC simulation used in the analysis, as well as for the Herwig7 sample used for JES uncertainty determination. The values of the statistical uncertainties on the Herwig7 sample are the basis for the $p_{T,\text{avg}}$ -bin removal needed to account for small event counts in the higher $p_{T,\text{avg}}$ -bins. The threshold of 14% ensures that the JES uncertainty determination via the Herwig7 sample yields statistically reliable estimations.

$p_{T,\text{avg}}$ values in the medium- $p_{T,\text{avg}}$ range and then falling again is caused by the trigger turn-ons, discussed in section 4.3.1.

4.3 Dijet Analysis

As, at hadron colliders, jets are produced in great number, the inclusive jet spectra are important to be understood. In this part of the analysis, the focus lies on events containing at least two jets, each with a transverse momentum above $p_T = 30$ GeV.

4.3.1 Trigger Studies

In order to handle the high event rates, only a limited number of events is stored to disk. As discussed in section 3.2.3, triggering is essential for dealing with high event rates in the data acquisition process. The triggers used in this analysis are the High Level Trigger (HLT) paths listed in table 4.4.

For the dijet analysis, the single-jet HLT paths are used and studied with respect to the average transverse momentum $p_{T,avg}$ of the dijet system. These single-jet triggers require at least one jet contained in the event to pass a given threshold in terms of its transverse momentum p_T . In the nanoAOD files these trigger paths are named *HLT_PFJet**, where PF stands for the particle flow algorithm used in the physics object reconstruction at CMS. Instead of the * there is the nominal turn-on point value of the trigger in GeV.

It is necessary to determine the point in $p_{T,avg}$, where each of these triggers reaches an efficiency threshold, in this analysis chosen to be 99%. There are several possible procedures to find the values in question, investigating the triggers' turn-on curves.

One way is to emulate the next higher trigger by making use of the available information on the original trigger object stored in the nanoAOD file. In this approach the basic assumption is, that the trigger $N - 1$ with the next lower p_T -threshold should observe all the events that are passing the higher trigger N , which requires a higher transverse momentum threshold.

For each event all the stored trigger objects are looped through and the two trigger objects with the highest transverse momentum p_T are determined. These are then identified to be the two leading jets. Following this procedure, it is checked whether the trigger object with the highest transverse momentum could pass the criterion to be stored in the next higher trigger path's event collection. This means the leading trigger object's p_T must be higher than the nominal value $p_{T,nom}$ of the next higher (N+1) trigger threshold. The same check is carried out for the next-to-next higher trigger (N+2).

Practically, this means taking the trigger information of trigger N and simulating the next higher trigger N+1 (and next-to-next higher trigger N+2) by applying the p_T cut corresponding to the nominal p_T -threshold $p_{T,nom}$ of the next higher (next-to-next higher) trigger on the trigger object. The amount of trigger objects passing this selection emulates the amount of events that would have passed the next higher actual HLT path. For this reason, this method is called *emulation method* within the presented analysis.

Now, a *Gauss error function*, see equation 4.5, is fitted to the data points resulting from the trigger ratio $\frac{N}{N-1}$.

$$f(p_{T,\text{avg}}) = A \cdot 0.5 \cdot \left[1 + \operatorname{erf} \left(\frac{p_{T,\text{avg}} - B}{C} \right) \right] \quad (4.5)$$

$$\text{with } \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (4.6)$$

The fitting function $f(p_{T,\text{avg}})$ in 4.5 contains three parameters A , B and C . Parameter A can be set to the expected plateau of the function $f(p_{T,\text{avg}})$ at higher $p_{T,\text{avg}}$ values. It stands for the efficiency ϵ of the trigger at saturation, in the case of trigger turn-ons this is assumed to be $\epsilon_{\text{full}} = A = 1.0$. The second parameter B is approximately the nominal value of the single-jet trigger threshold. It is left free in the first fitting round, with the initial parameter value B being set to the nominal trigger turn-on in GeV. The last parameter, C , is left free as well. In the turn-on behaviour of each trigger, $1/C$ is a measure of the width of the turn-on.

When carrying out the 2-parameter fit for each trigger, the results for parameter B are stored. Afterwards, the results for parameter B are plotted as $B/p_{T,\text{nom}}$ over the nominal value of the trigger turn-on $p_{T,\text{nom}}$ and a line $h(p_{T,\text{nom}})$ is fitted to these points. Taking the values $B = p_{T,\text{nom}} \cdot h(p_{T,\text{nom}})$ from this linear fit and fixing B in function 4.5 to the value corresponding to the trigger in question, the first fit with $f(p_{T,\text{avg}})$ is repeated. Hence, C is the only free parameter left in this second turn-on fitting round. This time, the results for C are stored. These values are later used to extrapolate to the turn-on curves for the lowest triggers, where the ratio $\frac{N}{N-1}$ can not be constructed.

Getting the $p_{T,\text{avg}}$ value of this last 1-parameter fit at $f(p_{T,\text{avg}}) = 0.99$ conclusively determines the turn-on point of the trigger N under investigation. In order to increase the available points for fitting, the same ratio is taken as well to the $(N - 2)$ trigger. The above described fitting procedure with function 4.5 is carried out on a combined graph, taking the lower points in $p_{T,\text{avg}}$ from the $\frac{N}{N-2}$ ratio and the higher points from $\frac{N}{N-1}$, overlapping the graphs in the medium $p_{T,\text{avg}}$ range. Proceeding like this, more data points are available to the fit and this way uncertainties on the fit can be reduced. The following table 4.4 shows the values gained from this trigger emulation approach, i.e. the combined fit of both curves.

The values in table 4.4, obtained for the central bin yb0ys0, are used for the whole analysis. Compared to the barrel region, the outer y_b - y^* -bins are suffering from higher statistical uncertainties caused by lower event counts in these regions, which makes finding the turn-on point increasingly difficult.

After determination of the 99% efficiency point in $p_{T,\text{avg}}$ for each of the HLT paths, the next higher bin edge in the dijet $p_{T,\text{avg}}$ binning has to be identified. The first $p_{T,\text{avg}}$ histogram bin being filled with events obtained from the corresponding trigger, is indicated in column four in table 4.4. It can be concluded, that the lower bin bound of the first $p_{T,\text{avg}}$ bin filled with data from a certain trigger path lies above the 99% efficiency threshold for all the trigger paths, with the smallest difference of only 0.2 GeV being seen for the 400 GeV trigger path.

single-jet trigger path	HLT threshold in GeV	99 % efficiency point in $p_{T,\text{avg}}$ in GeV	low $p_{T,\text{avg}}$ bin boundary in GeV	effective luminosity in fb^{-1}
HLT_PFJet40	40	51.6	56	$1.196 \cdot 10^{-4}$
HLT_PFJet60	60	77.0	84	$4.955 \cdot 10^{-4}$
HLT_PFJet80	80	104.4	114	$1.865 \cdot 10^{-3}$
HLT_PFJet140	140	176.1	196	$2.676 \cdot 10^{-2}$
HLT_PFJet200	200	242.5	245	$1.147 \cdot 10^{-1}$
HLT_PFJet260	260	310.8	330	$2.485 \cdot 10^{-1}$
HLT_PFJet320	320	383.8	395	$9.917 \cdot 10^{-1}$
HLT_PFJet400	400	467.8	468	1.983
HLT_PFJet450	450	521.5	548	3.964
HLT_PFJet500	500	568.7	592	31.710
HLT_PFJet550	550	621.6	638	31.710

Table 4.4: Thresholds and turn-on points for the different single-jet HLT paths. The values presented in this table are the ones determined by using the *emulation method* used in this analysis. They correspond to the values obtained in the central y_b - y^* -bin (barrel region), but are applied to all six y_b - y^* -bins. The effective luminosity for each of the HLT paths is shown in the last column.

Figure 4.3 shows the turn-on curves for the single-jet HLT paths, obtained with the "combined" (i.e. $\frac{N}{N-1}$ and $\frac{N}{N-2}$ fitted together) emulation method described above. The same graphs are displayed in figure 4.4, zoomed-in on the y-axis.

Cross-checking the emulation method with the ratio method If the possibility to emulate the next higher trigger is not available due to missing trigger object information, another method needs to be applied in order to obtain the turn-on curves and 99% efficiency point for each of the trigger paths. The alternative is to take the event counts directly from the next higher HLT path, instead of emulating this trigger. In this case, special attention has to be paid to the fact that most of the triggers are prescaled. In order to account for this, the integrated luminosity for each trigger has to be determined. The event counts collected with each trigger are divided by the obtained luminosity. Afterwards, the procedure for determining the turn-on curves is analogous to the emulation method. Using the HLT paths, the trigger ratios $\frac{N}{N-1}$ and $\frac{N}{N-2}$ are calculated, combined, and a Gaussian error function, as given in equation 4.5, is fitted to the obtained combined graph.

Conclusively, it is observed that the two methods agree within the chosen $p_{T,\text{avg}}$ -bin widths, i.e. both methods yield the same first valid $p_{T,\text{avg}}$ bin, with efficiency $\epsilon > 99\%$, for each trigger. This indicates that the trigger emulation used to determine the turn-on points of the HLT triggers works. Therefore, the corresponding values shown in table 4.4 are used for this thesis.

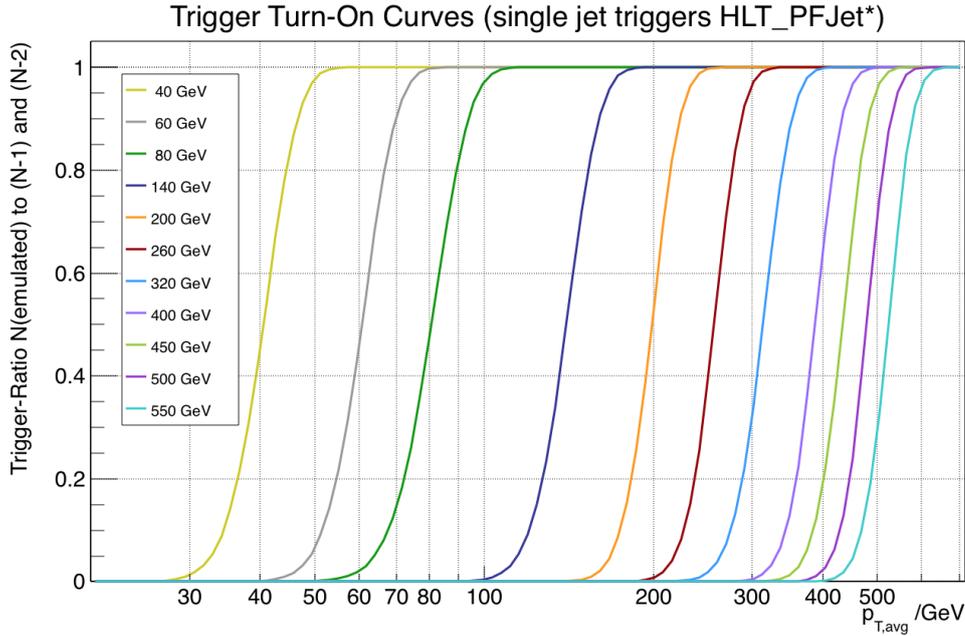


Figure 4.3: The turn-on curves for the single-jet HLT paths obtained using the emulation method, combining information from the trigger in question (N) together with the next lower (N-1) and the next-to-next lower trigger (N-2). The p_T -value in GeV given in the legend stands for the nominal turn-on point of the individual trigger.

4.3.2 Event Selection

Requiring at least two jets contained in the event, this analysis investigates the *inclusive* dijet spectrum. The cross section measurement is carried out triple-differentially, as discussed in section 4.1.

In this analysis, the jet clustering algorithm employed in the trigger decision of the single-jet *HLT_PFJet** triggers is the anti- k_T algorithm with a jet radius parameter $R = 0.4$.

When selecting the events contributing to the dijet cross section, the following phase space cuts are applied.

The first condition for the events to pass is containing at least two jets. In order to avoid contributions from pileup interactions, a minimum transverse momentum $p_T > 30$ GeV cut is applied on both the leading and subleading jet separately. The second kinematic cut is chosen in order to be in compliance with the Z+jet analysis, explained in section 4.4, where tracker coverage leads to a restriction of the rapidity of the investigated objects to $|y| < 2.4$.

If one of the triggers fired, the selected event becomes assigned to one of the six y_b - y^* -bins according to the rapidity values of its jets and filled into the corresponding $p_{T,avg}$ -bin.

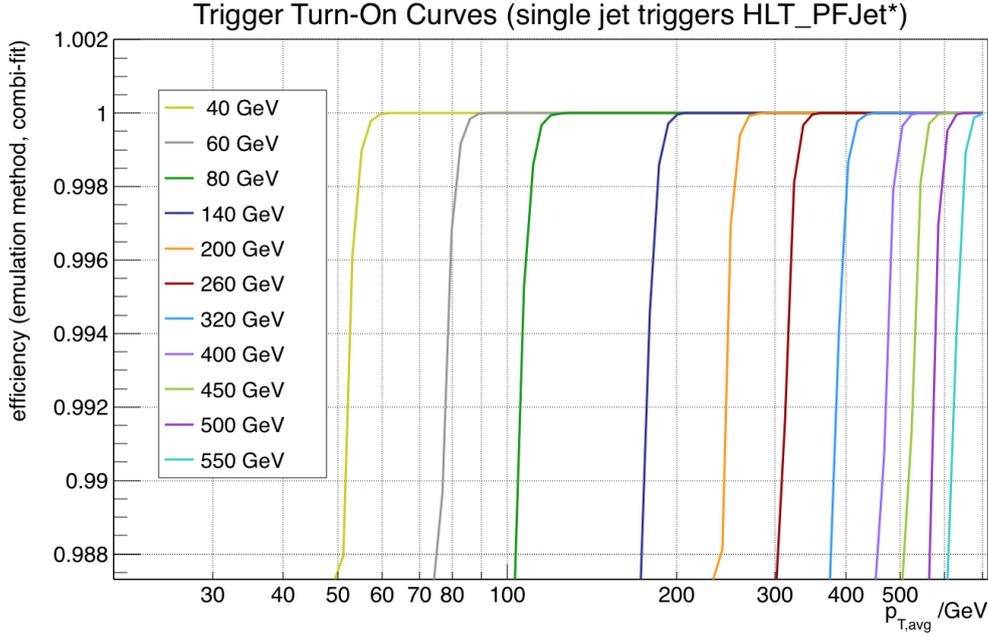


Figure 4.4: Trigger turn-on behaviour of the HLT paths used in this analysis. The y-axis is zoomed in in order to display the 99% efficiency point.

Additionally, the missing transverse energy \cancel{E}_T in the event has to be reasonably small compared to the total energy $\sum E_T$ contained in the event. This requirement is used to remove background events, as high energetic jet production usually does not include high values of \cancel{E}_T [30]. For this reason the event is required to fulfill the condition $\cancel{E}_T < 0.3 \cdot \sum E_T$.

4.3.3 Comparison to Monte Carlo Simulation and Theory Predictions

For checking the agreement between data and predictions, Monte Carlo event generators are used. Here, the dijet cross section results in data are compared to dijet cross sections extracted from simulated data sets, produced using the MC event generator PYTHIA8. Results obtained using HERWIG7 as MC generator are displayed in the appendix A.3.1. The used Monte Carlo simulation sets are listed in appendix A.1. Additionally, fixed-order theory calculations in LO and NLO precision are included in the comparison. These calculations are carried out with NNLOJET, stored into FASTNLO tables and evaluated with the CT14NLO PDF set.

Data and MC simulation results have not been unfolded and are therefore both investigated at detector level. The fixed-order theory calculations are pure theory without any detector simulation applied. These two facts have to be taken into account when comparing the different results in this chapter.

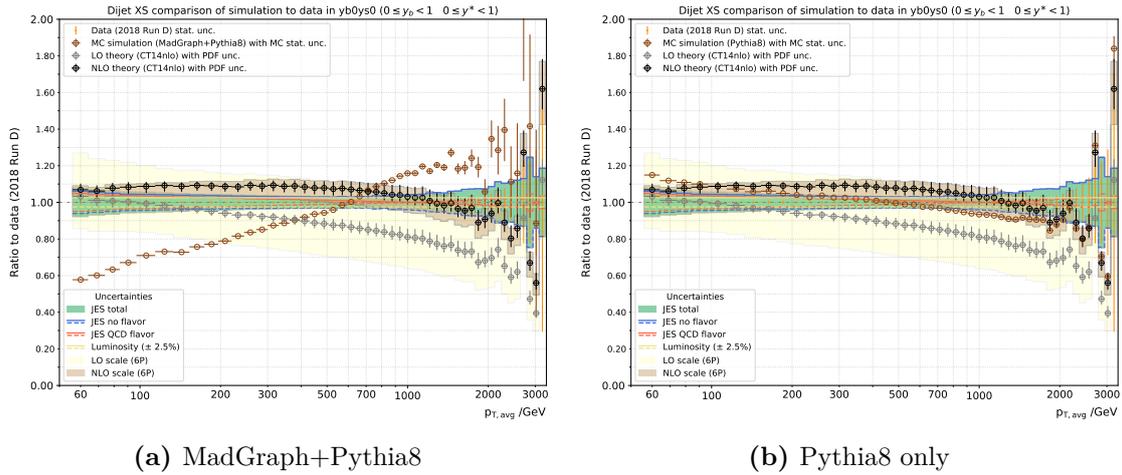


Figure 4.5: Comparison of MadGraph+Pythia8 versus Pythia8 only sample for dijet. This figure just serves illustrative purposes. The detailed contents are explained in the next section with separate graphics. Displayed in brown are the dijet spectrum obtained from the MadGraph+Pythia8 sample on the left and from the Pythia8-only sample on the right as ratio to data. Since the MC simulation that only uses Pythia8 without MadGraph describes the spectrum visibly better, it is the MC sample of choice used in the following chapters of this analysis.

Different Monte Carlo samples In the course of this analysis, also the MadGraph+Pythia8 Monte Carlo simulation set for QCD was investigated. This was necessary, as the QCD-sets used for the dijet analysis shown in this chapter are created by Pythia8, without using MadGraph. Whereas in the Z+jet analysis, MadGraph+Pythia8 was used for the Monte Carlo sample creation.

However, it is observed, that the MadGraph+Pythia8 sample does not describe the dijet cross section spectrum as well as Pythia8 alone does. This can be seen from comparing the subfigures in figure 4.5. The content of these pictures will be explained in the following section 4.3.4. Figure 4.5 just serves illustrative purposes. Focusing on the brown points, which display the ratio of the cross section spectrum obtained with the Monte Carlo sample in question over the spectrum obtained from data, it becomes evident that the MadGraph+Pythia8 sample does not provide a satisfactory description of the dijet spectrum. The MC sample on the right part of figure 4.5, using Pythia8 without MadGraph, agrees much better with the results obtained from theory calculations (in grey and black) and is generally closer to unity. This underlines that the dijet spectrum is described better by using the Pythia8 sample instead of the sample combining MadGraph and Pythia8. Therefore, all the subsequent analyses are carried out using the Pythia8-only sample for dijet. Additional result plots for the Herwig7 and MadGraph+Pythia8 case can be found in the appendix A.3.

4.3.4 Dijet Cross Section Results

At first, the dijet cross section is calculated in all of the six y_b - y^* -bins. Together with the results obtained from Monte Carlo simulation and fixed-order theory calculations (leading order and next-to-leading order), the inclusive dijet cross section is displayed in figure 4.6. The main purpose of this overview is to show that there are no major disagreements between data and the description of data by theory predictions. For getting more precise and detailed insights than can be obtained from such a double-logarithmic plot, ratio plots are much more suitable. These are presented in the next subsection.

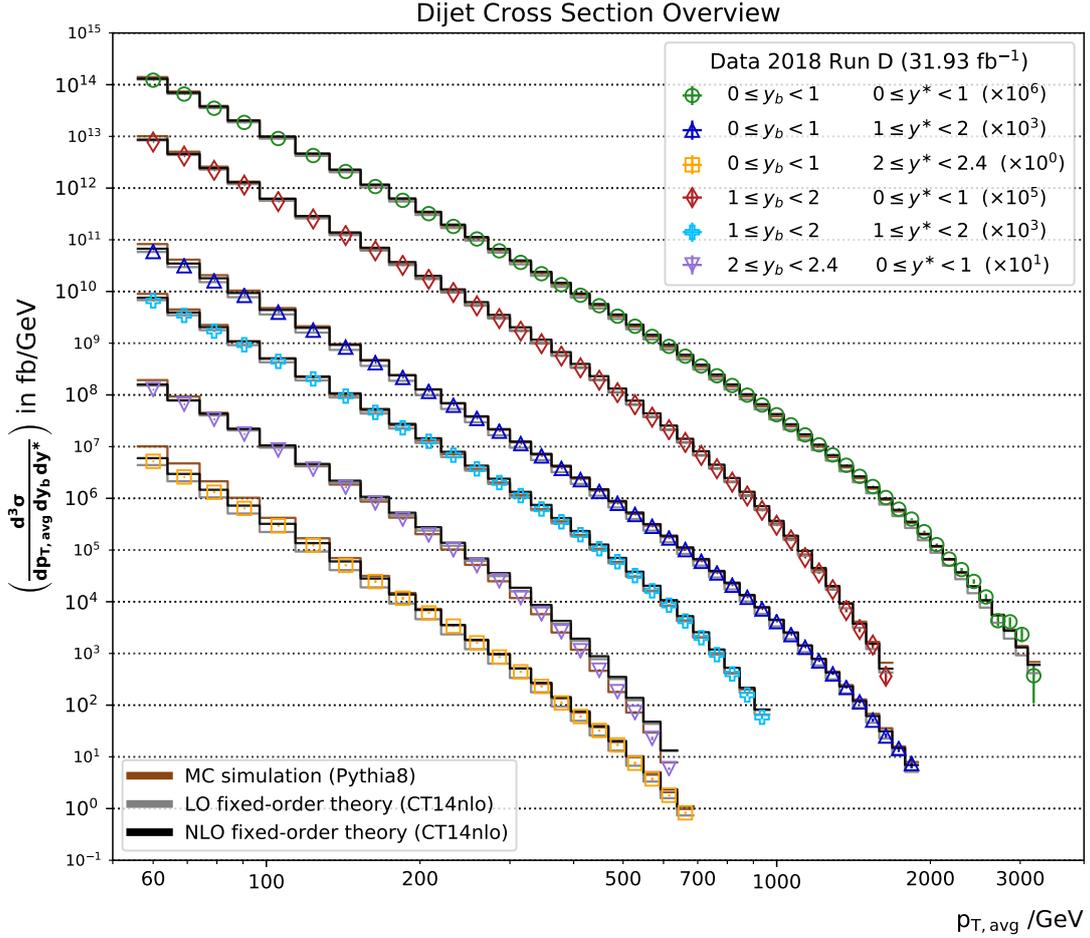


Figure 4.6: An overview of the triple-differentially measured inclusive dijet cross section in the six different y_b - y^* -bins considered in this analysis and plotted over the average transverse momentum $p_{T,avg}$ of the two leading jets. Cross section results are displayed for data (coloured), Monte Carlo simulation (brown), as well as fixed-order theory predictions (leading order in grey and next-to-leading order precision in black). The spectra are multiplied by factors 10^x as denoted in the legend to provide a better separation of the individual distributions.

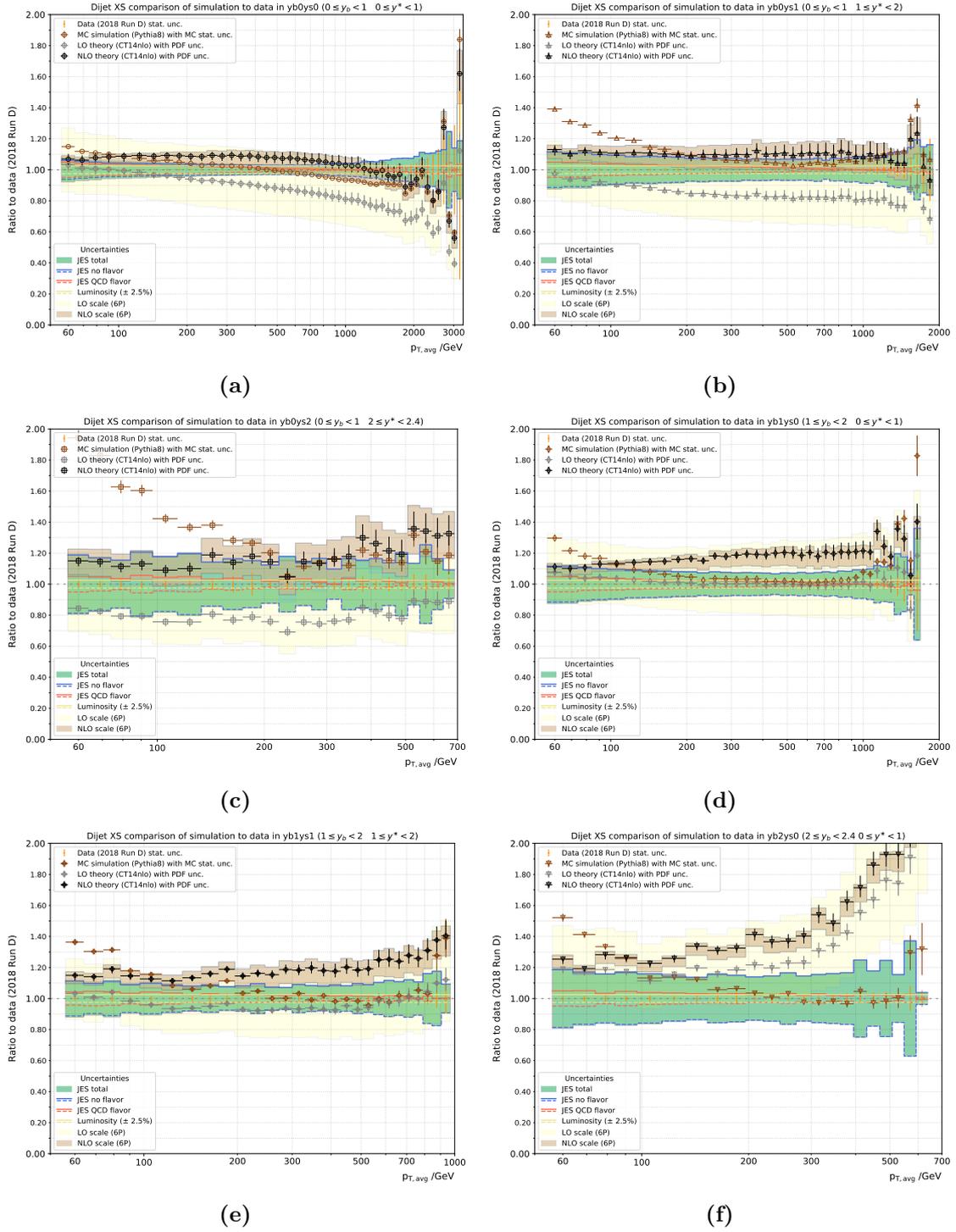


Figure 4.7: Cross section ratios comparing simulation and theory results to data as a function of the average transverse momentum of the two leading jets $p_{T,avg}$ for the six y_b - y_* -bins. Shown are the ratios for simulation with PYTHIA8 over data, as well as the fixed-order theory calculation results divided by data. The main uncertainty sources are displayed.

Displayed in figure 4.7 are six subplots corresponding to the six y_b - y^* -bins introduced in section 4.1. The dijet cross section results are displayed as ratios of fixed-order theory and MC simulation over data. Each of the six individual plots has the same structure. The brown points are the dijet spectrum obtained from Monte Carlo simulation (with the Pythia8 sample) divided by the results extracted from Run D data of the 2018 data taking at CMS. The error bars at the brown ratio points represent the statistical uncertainty of the MC simulation, scaled by the data cross section. In this sense, the brown ratio points and their error bars can be read like $\frac{\sigma_{\text{MC}} \pm \delta_{\text{MC, stat.unc.}}}{\sigma_{\text{data}}}$. For data itself, the relative statistical uncertainty in each bin is drawn in orange around unity.

The cross sections based on fixed-order theory calculations divided by the data results are displayed in grey for LO precision and in black for NLO. Assigned to those grey and black ratio points as error bars are the PDF uncertainties on the theory calculation, using the CT14NLO PDF set. Put into an equation, the theory ratios in grey (LO) and black (NLO) with error bars can be written as $\frac{\sigma_{\text{theory}} \pm \delta_{\text{PDF unc.}}}{\sigma_{\text{data}}}$. For providing an impression of the scale uncertainty $\delta_{\text{scale unc.}}$ on the fixed-order theory predictions, shaded bands are depicted behind the LO and NLO points in light yellow and brown, respectively. These are obtained by varying the factorisation scale μ_F and renormalisation scale μ_R independently by factors of 2 and $\frac{1}{2}$ and extracting the scale uncertainty by investigating the maximal deviations of the varied predictions from the nominal one. This procedure is called *6P-variation*. The factors for the variations of (μ_R, μ_F) are: $(\frac{1}{2}, \frac{1}{2})$, $(\frac{1}{2}, 1)$, $(1, \frac{1}{2})$, $(1, 2)$, $(2, 1)$, $(2, 2)$ with respect to the central scale μ_0 , see explanation in ref. [30]. Again, these uncertainties are scaled by the data and plotted as $\frac{\sigma_{\text{theory}} \pm \delta_{\text{scale unc.}}}{\sigma_{\text{data}}}$ band.

The green uncertainty band in the background shows the total Jet Energy Scale (JES) uncertainty, as explained in section 4.2. From all the contributing JES uncertainty sources, the dominant source in this case is identified to be the JES QCD flavour uncertainty and displayed separately in symmetrised form as red lines. The blue lines correspond to the rest of the contributing JES uncertainty sources. The quadratic sum of the partial JES uncertainties shown as red and blue lines yields the total JES uncertainty represented by the green band. The yellow pair of lines depicts the luminosity uncertainty estimated to $\pm 2.5\%$, as recommended in ref. [67].

It is important to note that the displayed curves for data and MC represent detector-level. Unfolding of the data back to particle-level is desirable, but was not possible anymore within the time constraints of this study. Therefore, a comparison of the theory calculations to the detector-level data or MC simulations has to be interpreted with caution. Taking this into account, the NLO predictions seem to be in reasonable agreement with the data.

In the following, some of the y_b - y^* -bins are described in more detail.

The central bin – $y_b 0 y_s 0$: It is natural to start by having a look at the central bin $y_b 0 y_s 0$ in figure 4.7, which corresponds to the barrel region of the detector. This phase space region yields the highest event counts N and therefore the highest cross section σ and lowest statistical uncertainties.

The Pythia8 MC simulation set, as well as the NLO fixed-order theory calculation show

good agreement with the cross section obtained from data in this bin. Especially in the medium- $p_{T,\text{avg}}$ region, the difference between the cross section results obtained from data and the ones obtained from MC simulation are lower than 5 %. The same observation holds for the NLO theory description for an even larger phase space region, starting already in the lowest $p_{T,\text{avg}}$ region and going up to the highest $p_{T,\text{avg}}$ values with a maximal deviation from the data of 5%. Within PDF and scale uncertainties from the NLO set, it agrees with the MC sample over almost the whole $p_{T,\text{avg}}$ -range in this central y_b - y^* -bin. The LO theory results deviate more strongly from data than NLO and MC, which is expected. Nevertheless, this deviation is covered by the uncertainties, the most dominant for LO being the scale uncertainty, displayed as light yellow band. The fixed-order theory description in LO precision is mostly underestimating the dijet cross section in this y_b - y^* -bin. This effect gets enhanced when going to higher $p_{T,\text{avg}}$ values. However, the difference from LO theory to data is lower than 30 % over a wide range.

The total Jet Energy Scale (JES) uncertainty in the central $p_{T,\text{avg}}$ bins is clearly below 5% in the barrel region. And it stays well below 10% up to the highest $p_{T,\text{avg}}$ region. In this y_b - y^* -bin, the statistical uncertainties on data, MC and theory are negligible except for the highest $p_{T,\text{avg}}$ bins, where error bars become visible. Also the systematic uncertainties increase at the high $p_{T,\text{avg}}$ -limit of the phase space.

Large rapidity separation – yb0ys2: Investigating the high- y^* region via the yb0ys2 bin, it immediately becomes clear that this bin shows increased uncertainties compared to the barrel region yb0ys0. This is expected, because the event counts in this bin are much lower than in the central yb0ys0 bin. In addition, the range in rapidity separation is smaller with $\Delta y^* = 0.4$ as compared to the central bin with $\Delta y^* = 1.0$.

While the deviation of the NLO theory description from data stays below 20% within the low and medium $p_{T,\text{avg}}$ range, the MC sample leads to much larger differences, especially in the low- $p_{T,\text{avg}}$ region. The MC simulation approaches the data results in the medium and high $p_{T,\text{avg}}$ range. For the upper half of the $p_{T,\text{avg}}$ region in this y_b - y^* -bin, MC and theory give an equally accurate description of the cross section results gained from data, staying within a maximum of 40% deviation from data. It is observed, that in the low- $p_{T,\text{avg}}$ bins even the LO theory description agrees better with the data than the Pythia8 MC sample does.

In general, the fixed-order theory results agree with the data within uncertainties. Hence, the fixed-order theory calculations without detector simulation give a good description of the data also in this extreme phase-space bin.

The remaining differences between NLO fixed-order theory and data can partly be explained by the data not being unfolded. Since the MC simulation results are not unfolded either, one would expect the MC results to agree better with data than the NLO calculation, which, however, is not the case.

Strongly boosted dijet system – yb2ys0: As a second extreme y_b - y^* phase space bin, the yb2ys0 bin, which corresponds to a high boost of the dijet system, is investigated. In

this extreme phase space bin, big deviations of the fixed-order theory calculations from data are observed. Going to higher $p_{T,avg}$ values, the difference from NLO theory to data increases from around 25% at the lowest $p_{T,avg}$ -values up to values clearly exceeding 50% in the high $p_{T,avg}$ region. The Pythia8 MC sample's difference from data instead stays below 20% in the medium $p_{T,avg}$ region. These discrepancies are covered by the systematic and statistical uncertainties in this y_b-y^* -bin. Thus, it can be concluded that the medium $p_{T,avg}$ region in the extreme y_b bin is well described by the investigated MC sample.

Medium y_b-y^* -bins and general observations: In the y_b-y^* -range between the above explained extremes, i.e. the bins yb0ys1, yb1ys0 and yb1ys1, rather low uncertainties are observed when compared to extreme bins yb0ys2, yb2ys0. Otherwise, the ratio behaviour in these medium- y_b-y^* bins is between the behaviour in the central and the extreme bins, analysed above. In those bins, over most of the $p_{T,avg}$ -range, the MC simulation is closer to the data results than the theory description is. This results at least partly from the fact that MC and data are both at detector level, whereas the theory calculations do not include any detector simulation, as identified already in the sections above. Additionally, including higher orders into the theory calculation, going to NNLO, is expected to lead to an even better description of data.

For all the y_b-y^* -bins, large statistical uncertainties are observed for data at the highest $p_{T,avg}$ -values compared to the low- and medium- $p_{T,avg}$ range. In general, the reliable $p_{T,avg}$ -range in the extreme bins does not exceed 700 GeV. The lower number of events in the high- $p_{T,avg}$ region of the high- y_b bins compared to the central y_b-y^* -bin can be explained, since in order to reach a strong boost of the dijet system, one of the incoming scattering partons has to carry a much higher proton momentum fraction x_i than its scattering partner does. Furthermore, the boosted system is propelled along the beamline. For this reason high transverse energies of the final state particles are not expected in this y_b-y^* -bin. The same holds for high y^* -bins, where the jets with large rapidity separation are also measured in the detector's forward region.

Overall, the description of the triple-differential dijet spectrum by fixed-order theory predictions and MC simulation is in agreement with the data within uncertainties. These samples are therefore used in the investigation of the double ratios in chapter 5.

4.4 Z+Jet Analysis

At CMS, Z bosons decaying into two muons can be precisely reconstructed thanks to their clear signature in the detector. In addition to the Z bosons' clear signal, their properties are very well understood, as the Z boson has been investigated in various precision measurements since its discovery in the early 1980s. A very clean observation channel for the Z boson is its decay into two muons ($Z \rightarrow \mu\mu$), because of the fact that muons can be accurately measured in the CMS detector. Therefore, the decay channel $Z \rightarrow \mu\mu$ is the decay channel of choice in this analysis. The branching fraction for this decay of the Z boson $\Gamma_{\mu\mu}/\Gamma_{tot}$ is $3.3662 \pm 0.0066\%$ [3]. This is small in comparison with

the branching fraction of a hadronic decay of the Z boson. However, compared to the hadronic final state, which is more difficult to identify, the $Z \rightarrow \mu\mu$ channel with two oppositely charged muons is much cleaner and therefore the most suitable choice for a precise measurement with CMS data.

For comparison to the dijet event topology, which is defined by two high- p_T jets in the final state, the process under investigation in this chapter is chosen to be Z+jet, hence the final state with the Z boson must contain at least one high- p_T jet in addition to the Z boson.

4.4.1 Z Boson Reconstruction

Muons used in this analysis are reconstructed using the PF algorithm, introduced in section 3.2.1. In order to be considered for the reconstruction of the Z boson, the event has to pass several conditions. The most basic requirement on the event is that it must contain at least two muons. These muons need to fulfill further criteria before being included into the muon collection used to reconstruct the Z boson candidates.

In an initial stage, muon candidates must pass a series of selection criteria defined in the CMS software as muon selectors. The first one is to pass the *Muon_mediumpromptId* selection. Firstly, muons passing this selector must pass the *medium Muon* identification, as described in more detail in ref. [54]. Secondly, this identification procedure chooses only those muons passing the *medium* working point, that additionally fulfill the requirements $|d_z| < 0.1$ cm and $|d_{xy}| < 0.02$ cm for the longitudinal and transverse distance to the primary vertex, making them *prompt* muons. This requirement of passing the boolean muon selector *Muon_mediumpromptId* removes fake muons, i.e. objects whose signatures in the detector are falsely identified as being muons, as well as non-prompt muons from other decay processes.

Next, the event is only considered for the reconstruction of possible Z boson candidates, if it passes the double-muon trigger *HLT_Mu23_Mu12*. This ensures that the leading and subleading muon's p_T is above a threshold of 23 GeV and 12 GeV, respectively. The integrated luminosity $\mathcal{L}_{\text{eff}}^{\text{int}}$ recorded by the double-muon trigger *HLT_Mu23_Mu12* is calculated with BRILCALC to $\mathcal{L}_{\text{eff}}^{\text{int}} = 1.588 \text{ fb}^{-1}$.

All remaining muons are considered in the search for good Z boson candidates, which are composed of pairs of muons. All such pairs possible with the remaining muons are subjected to further selection criteria described in the following.

The cuts on the transverse momenta $p_{T,\mu 1}$ and $p_{T,\mu 2}$ of the muons are set slightly stricter than the nominal threshold of the double-muon trigger, specifically to $p_{T,\mu 1} > 24$ GeV for the leading muon and to $p_{T,\mu 2} > 13$ GeV for the second most energetic muon in the event.

Furthermore, the isolation of the muons is checked. There are six different working points for muon isolation requirements on muons reconstructed within the PF framework. These reach from a *very loose* working point up to a *very very tight* working point. In this analysis, the minimum level of isolation for the muons is set to *loose*. For further information and detailed definition of the working points, see, for instance, ref. [50, 54, 68].

Finally, the muons must both have a pseudorapidity $|\eta_{\mu_{1,2}}|$ lower than 2.4, in order to ensure for sufficiently good tracker coverage.

4.4.2 Event Selection

The Z Boson All Z boson candidates constructed from the above muon pairs, undergo further selection steps. The transverse momentum $p_{T,Z}$ of each Z boson candidate must exceed 30 GeV. The electric charge of the Z boson candidate needs to obey $Q_{Z_{\text{cand}}} = 0$. And its mass $m_{Z_{\text{cand}}}$ must lie within a mass window set to $\Delta m_Z = |m_{Z_{\text{cand}}} - m_Z^{\text{PDG}}| < 20$ GeV, limiting the allowed deviation from the current PDG value to a maximum of 20 GeV. The current PDG value for the Z-mass is $m_Z^{\text{PDG}} = 91.1876 \pm 0.0021$ GeV [3].

Additionally, in accordance with the dijet analysis, it is again checked that $|y_Z| < 2.4$ holds for the Z boson candidate's rapidity y_Z . This cut is originally selected for muon-tracking reasons in the form of η -cuts on the individual muons and results in limiting the outer y_b - y^* -bins to maximum values below 2.4, each.

Usually, there is only a single Z boson candidate fulfilling all of the previously explained requirements. If more than one Z boson candidate in the event pass the Z boson selection cuts, the event is discarded. Like in the dijet analysis, also in the present analysis of Z+jet events, the missing transverse energy per event \cancel{E}_T must stay below a limit set to 30% of the total transverse energy $\sum E_T$ in the event under investigation. Accordingly, the corresponding condition is written as $\cancel{E}_T < 0.3 \cdot \sum E_T$.

The Leading Jet In this analysis, there is no limit set on how many jets the event may contain. This fact makes it an *inclusive* Z+jet cross section measurement, just as the dijet analysis is carried out inclusively. In the following, each event is required to contain at least one jet with transverse momentum $p_{T,j1} > 30$ GeV and absolute rapidity $|y_{j1}| < 2.4$. On the one hand, this is demanded for having symmetric cuts imposed on the leading jet and the Z boson, since the observables in this analysis are based on properties of both of these leading objects. And, on the other hand, this is necessary in order to be consistent with the cuts on the two jets in the dijet analysis. The jet which carries the highest p_T and fulfills the criteria described in this paragraph is identified as the leading jet.

Together, the leading jet and the good Z boson candidate, that pass all selection cuts, are the physics objects under investigation in this chapter. These two leading physics objects undergo analogous procedures as the previously studied dijet events. The average transverse momentum $p_{T,\text{avg}}$ is calculated from the Z boson and the leading jet's p_T and the classification into y_b - y^* -bins is made based on the observables y_b and y^* calculated from the Z boson and leading jet's rapidities y_Z and y_{j1} .

4.4.3 Comparison to Monte Carlo Simulation and Theory Predictions

The inclusive Z+jet cross sections are not only calculated from data collected during run D of the 2018 data taking period at CMS, but as well using results from MC event generators. For this purpose, MC simulation samples for the process $pp \rightarrow Z/\gamma^* \rightarrow ll(+jets)$ are

investigated. In this chapter the result plots are created using the MadGraph+Pythia8 sample for Z+jet events. This sample uses MadGraph for matrix-element generation in LO and Pythia8 for parton showering. Afterwards, the events in these MC simulation sets undergo detector simulation using Geant4. Results obtained with the MadGraph+Herwig7 sample are shown in appendix A.3.3. The full names of MC sets used in this analysis can be found in the appendix A.1.

Additionally, fixed-order theory calculations in LO and NLO precision are used for comparison with data. These theory predictions are obtained from NNLOJet and evaluated with the CT14NLO PDF set using FASTNLO.

4.4.4 Z+jet Cross Section Results

In the following, the results of the inclusive Z+jet cross section extracted from data, MC simulation and fixed-order theory calculations are presented. The cross sections are calculated for each of the y_b - y^* -bins. The structure of figures 4.8 and 4.9, displaying the Z+jet analysis results, is built analogously to the presentation of the dijet analysis results.

First of all, an overview on the cross section measurement results in the six individual y_b - y^* -bins is given in figure 4.8. Displayed are apart from the results obtained from data also the cross section results from MC simulation with MadGraph+Pythia8 and from fixed-order theory calculations in LO and NLO precision. The cross section results are plotted in fb/GeV over the average transverse momentum $p_{T,avg}$ in GeV, here calculated from the Z boson and the leading jet's kinematic properties. A minimum $p_{T,avg}$ value of 56 GeV is required for later comparison with the dijet spectrum. Moreover, the theory predictions below this $p_{T,avg}$ limit are not reliable due to the symmetric p_T cuts imposed on the final state objects, which exclude events where each of the objects carries p_T on one side of this threshold.

The Z+jet cross section values gained from data are drawn in a different colour for each of the six y_b - y^* -bins. MC simulation results are kept in brown and the theory predictions are displayed in grey for LO and black for NLO precision, respectively. In order to keep the spectra separated, the entries are multiplied with different factors of 10^x , as indicated in the data legend in brackets next to the corresponding y_b - y^* -range. As this overview is presented in a double-logarithmic manner, it is difficult to properly identify differences in the description of the Z+jet spectrum by data, MC simulation and fixed-order theory prediction. However, by comparing the grey (LO) and black (NLO) lines in figure 4.8, it can be observed that the k-factors $k_{NLO} = \frac{\sigma_{NLO}}{\sigma_{LO}}$ in the Z+jet theory prediction are large for $y^* > 1$ and particularly so for $y^* > 2$. To give an impression of the k-factor's magnitude in bin yb0ys2: At $p_{T,avg} = 60$ GeV, this k-factor is already $k_{NLO} \approx 4.6$ and rises up to $k_{NLO} \approx 10.25$ around $p_{T,avg} = 200$ GeV. This phenomenon can be traced back to the fact that some partonic subprocesses only contribute to the Z+jet cross section at NLO and are not present yet at LO, see also ref.[19]. Apart from the expected differences in the spectrum in LO compared to NLO, figure 4.8 shows that the descriptions of the inclusive Z+jet spectrum by MC simulation and theory prediction agree with the

shape observed in data. In order to achieve more precise insights on the agreement and discrepancies between data and its prediction by simulations, the next section introduces ratio plots obtained by dividing the Z+jet cross sections calculated from simulations by those gained from data.

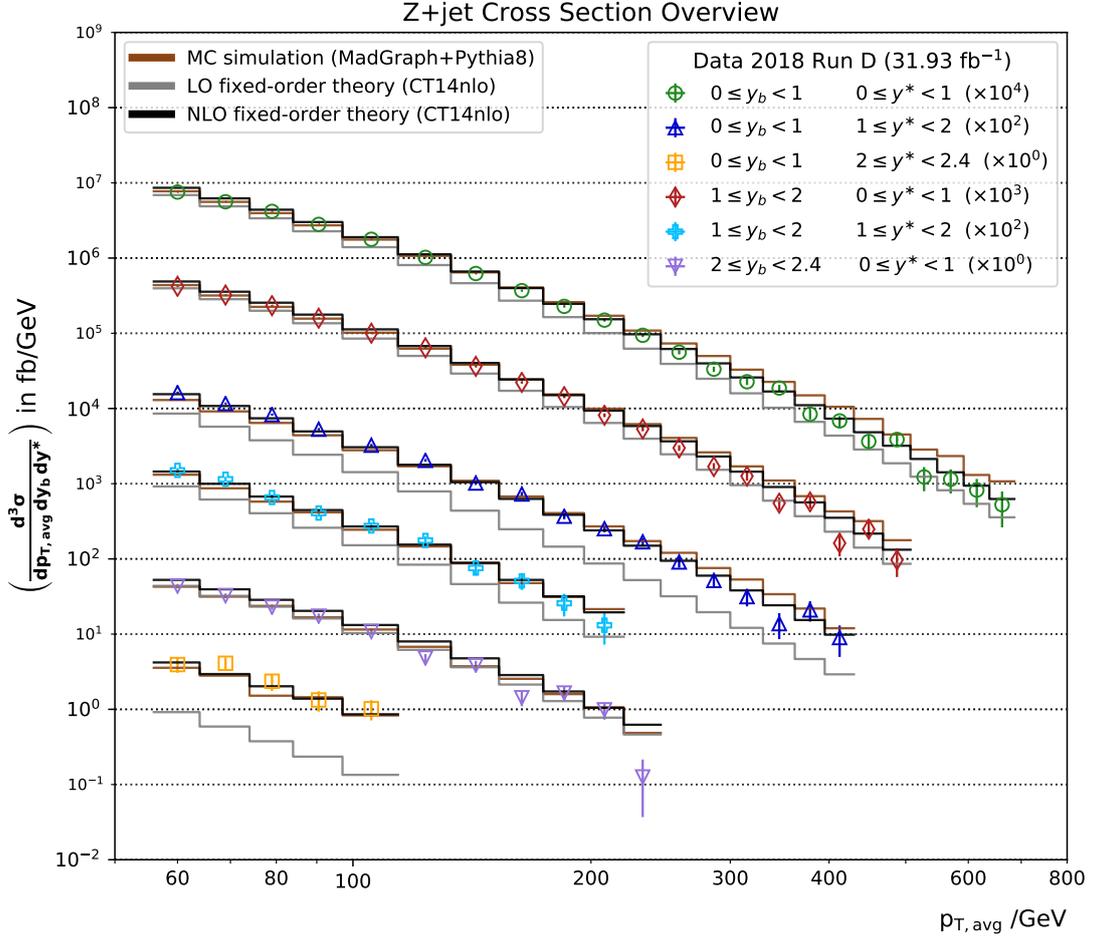


Figure 4.8: An overview of the triple-differentially measured inclusive Z+jet cross section in the six different y_b - y^* -bins considered in this analysis and plotted over the average momentum $p_{T,\text{avg}}$ of the two leading objects, here the Z boson and the leading jet in the event.

Figure 4.9 contains six subplots, one for each y_b - y^* -bin. Displayed are comparisons of simulations to data. The points in the plots are calculated by dividing the cross section results obtained from MC simulation σ_{MC} and fixed-order theory calculations σ_{theo} by the cross sections σ_{data} gained from the data set. Following the pattern chosen for the

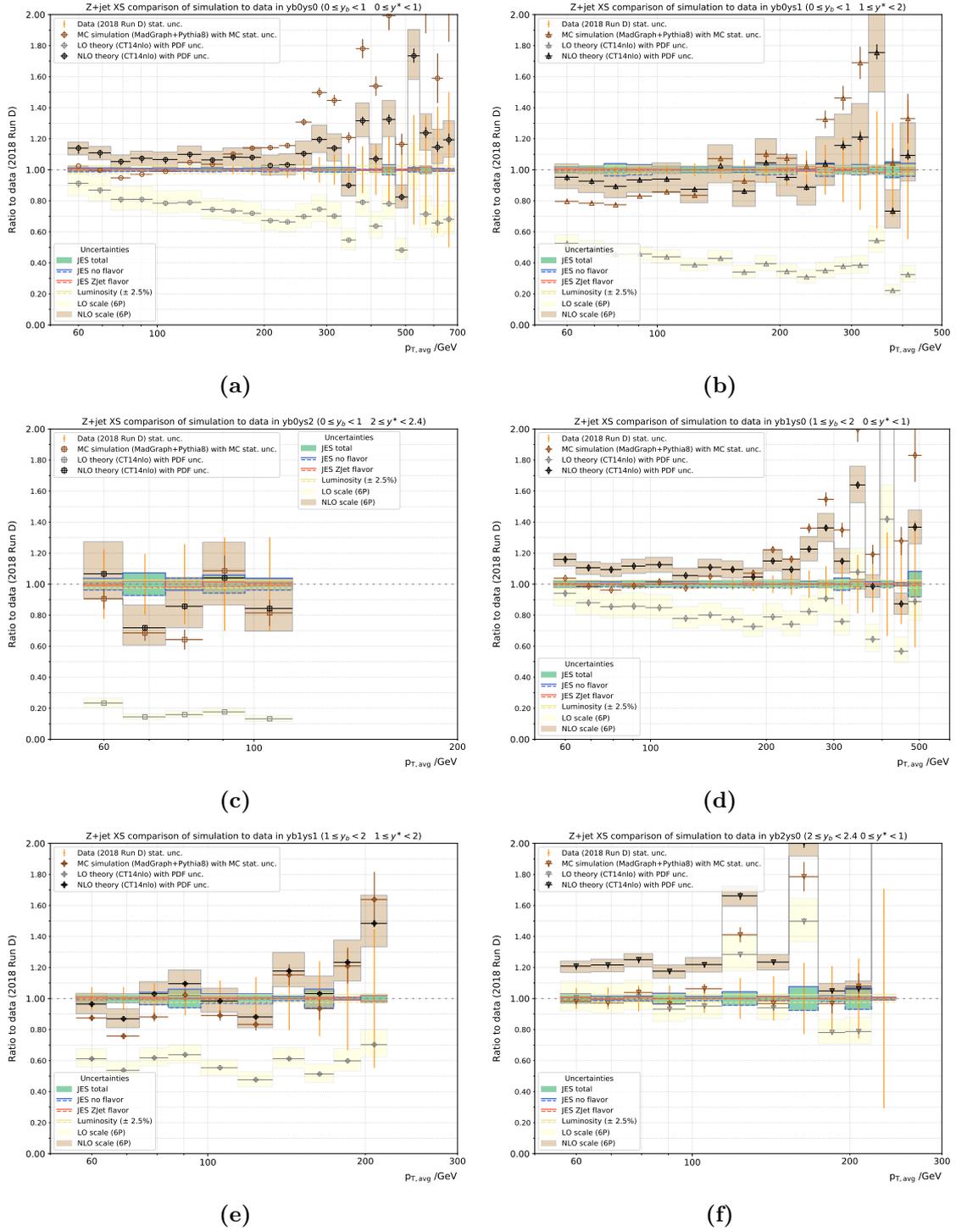


Figure 4.9: Ratios of the triple-differentially measured inclusive Z+jet cross sections obtained from MC simulation and fixed-order calculations over data. The six different plots represent the six y_b - y^* -bins. In each of these bins the ratio is plotted over the average momentum $p_{T,avg}$ of the two leading objects, here the Z boson and the leading jet in the event.

previously discussed dijet analysis, the ratio of MC simulation results to data is displayed in brown colour with the statistical uncertainty of the MC sample $\delta_{\text{MC stat. unc.}}$ assigned to the points as error bars, like $\frac{\sigma_{\text{MC}} \pm \delta_{\text{MC stat. unc.}}}{\sigma_{\text{data}}}$. The cross sections coming from fixed-order theory calculations divided by data are coloured in grey for LO and black for NLO. The error bars displayed at the theory prediction points correspond to the PDF uncertainties $\delta_{\text{PDF unc.}}$ (CT14NLO) occurring in these calculations. Together, the ratio points and error bars for the fixed-order theory calculations can be formulated as $\frac{\sigma_{\text{theo.}} \pm \delta_{\text{PDF unc.}}}{\sigma_{\text{data}}}$. Behind these distributions for LO and NLO the transparent bands in light yellow and brown, respectively, display the scale uncertainty of the theory calculation obtained from six-point (6P) asymmetric scale variations of μ_{R} and μ_{F} (by factors 2, 1/2 etc.). The uncertainty bands are drawn according to the following equation: $\frac{\sigma_{\text{theo.}} \pm \delta_{\text{scale unc.}}}{\sigma_{\text{data}}}$. The statistical uncertainty on data itself $\delta_{\text{data stat. unc.}}$ is depicted as orange error bars around unity like $\frac{\sigma_{\text{data}} \pm \delta_{\text{data stat. unc.}}}{\sigma_{\text{data}}}$. Also the JES uncertainties are shown. For the total JES uncertainty a green shaded error band is drawn around unity. It consists of the JES uncertainties coming from flavour uncertainties, here *jesFlavorZJet*, depicted as red lines around unity and the remaining JES uncertainty sources excluding flavour uncertainties, drawn as blue lines. The two yellow lines above and below unity show the luminosity uncertainty of $\pm 2.5\%$ as recommended for 2018 data in ref. [67].

Here, it is important to note that the displayed uncertainties do not represent the full set of uncertainties on the Z+jet analysis, but instead focus on the JES uncertainties and statistical uncertainties that are needed for comparison to the dijet analysis in chapter 5. As before, the fixed-order theory calculations do not include any detector simulation, while the data is displayed as results obtained at detector level, i.e. data has not been unfolded.

The central bin – yb0ys0: The bin described first is the central bin yb0ys0 which corresponds to the detector’s barrel region and is least affected by uncertainties.

In the low- to medium- $p_{\text{T,avg}}$ bins in this $y_{\text{b}}-y^*$ -bin, the description of the Z+jet cross section by the MC simulation deviates from data by less than 25%. The best agreement between MC simulation and data is observed in the lower half of the $p_{\text{T,avg}}$ -range. Going to higher $p_{\text{T,avg}}$ values, the MC sample’s behaviour differs more from data than in the low- $p_{\text{T,avg}}$ region, but also the statistical uncertainties rise. The cross section results from MC simulation reach deviations from data of more than 100% in the highest $p_{\text{T,avg}}$ bins that are still filled. However, this sloping upwards cannot be seen as a clear trend, because the points are fluctuating up and down, which is caused by statistical phenomena due to low event counts in the higher $p_{\text{T,avg}}$ -bins. For the same reason, also the statistical uncertainty $\delta_{\text{MC stat. unc.}}$ on the MC simulation result increases in these $p_{\text{T,avg}}$ -bins.

The NLO theory description of the Z+jet spectrum shows a similar behaviour as the MC sample when compared to data. Also the NLO fixed-order theory calculations lead to rather low deviations from data in the low- and medium- $p_{\text{T,avg}}$ range of the central $y_{\text{b}}-y^*$ -bin yb0ys0. It can be stated that the description of the Z+jet cross section spectrum

by the fixed-order theory calculation to NLO precision agrees with the data within uncertainties over a wide $p_{T,\text{avg}}$ -range. The theory calculation in LO precision is further off. However, the LO theory results do not deviate by more than 50% over almost the whole $p_{T,\text{avg}}$ range accessible in this y_b - y^* -bin.

In the highest $p_{T,\text{avg}}$ -range, also the statistical uncertainty on data grows significantly. Generally, the description of data by MC simulation and NLO theory calculation in this central bin yb0ys0 is more accurate in the lower $p_{T,\text{avg}}$ -range. This is expected when looking at the overall $p_{T,\text{avg}}$ -range of the Z+jet cross section spectra, see figure 4.8. Although yb0ys0 is the central bin with highest event counts, the phase space for Z+jet events does not exceed $p_{T,\text{avg}} = 1$ TeV and is mostly far below this.

The JES uncertainties in this central bin stay within the order of magnitude of the luminosity uncertainty, thus not significantly higher than 3%. For the Z+jet process this is not surprising, because the Z+jet topology as final state under investigation consists of only one jet and a Z boson. The Z boson, reconstructed from two muons, does not suffer from JES uncertainties. Therefore, these only concern the jet in the final state and are smaller than the JES uncertainties occurring in the dijet analysis.

Large rapidity separation – yb0ys2: Having a look at the y_b - y^* -bin with large rapidity separation of the Z boson and the leading jet, it immediately becomes apparent that this y_b - y^* -bin covers a smaller $p_{T,\text{avg}}$ -range than the previously investigated central bin does. Being a very sparsely filled y_b - y^* -bin, also the statistical uncertainties on data are large even in the low- $p_{T,\text{avg}}$ range. Again, it shall be stated, that the extreme y_b - y^* -bins are smaller than the central ones, here with $\Delta y^* = 0.4$.

Within the displayed $p_{T,\text{avg}}$ -range of this y_b - y^* -bin, the MC sample does not exceed a maximum difference of 40% from data. Since these $p_{T,\text{avg}}$ -bins are also otherwise affected by statistical uncertainties on data of 20% and higher, the insufficiencies of the description of the Z+jet spectrum by the MC simulation is covered by the uncertainties in this y_b - y^* -bin.

The same holds for the NLO description of the Z+jet spectrum compared to data. Over the whole accessible $p_{T,\text{avg}}$ phase space in this y_b - y^* -bin, the NLO calculation's deviation from data stays within a maximum extent of less than 30%. Furthermore, the differences between fixed-order calculation in NLO and data are covered by the shown uncertainties. The fixed-order theory predictions in LO precision are visible, but far off. This can as well be observed in the double-logarithmic cross section overview in figure 4.8. Some partonic subprocesses contributing to the Z+jet cross section only occur at NLO precision (cf. ref.[19]). The k-factors $\frac{\sigma_{\text{NLO}}}{\sigma_{\text{LO}}}$ are therefore large in this phase space region, wherefore LO cannot provide a reliable description of the physical reality in this y_b - y^* -bin.

Compared to the central y_b - y^* -bin, also the JES uncertainties are enhanced in this bin of large rapidity separation $y^* \geq 2$ of the two objects, the Z boson and the leading jet, in the final state. However, the JES uncertainties stay well below 10%. The Z+jet JES flavour uncertainty is constantly below 2.5%, i.e. lower than the luminosity uncertainty and visibly not dominating the results.

Strongly boosted Z+jet system – yb2ys0: The next y_b - y^* -bin under investigation is yb2ys0, which corresponds to a strongly boosted system with average boost $y_b \geq 2$.

In this y_b - y^* -bin, MC simulation and data agree within uncertainties with the exception of two $p_{T,\text{avg}}$ -bins. Apart from these two $p_{T,\text{avg}}$ bins, the differences between MC simulation and data stay below 10%. For data, the statistical uncertainties grow when moving towards higher $p_{T,\text{avg}}$ values.

The best agreement of the fixed-order theory prediction in NLO with data is given in the lower half of the $p_{T,\text{avg}}$ spectrum. Nevertheless, in this y_b - y^* -bin where the Z+jet system is strongly boosted, the NLO prediction is mostly overestimating the data spectrum. This offset could be caused by the missing detector simulation in theory, but can as well be linked to the choice of PDF set and depend on how well it is determined for low- x values that are involved in the boosted topologies. Detector effects are amplified in this extreme phase space region, where the detector is not understood as precisely as in the central region. The MC simulation sample offers a better description of the inclusive Z+jet cross section than the fixed-order theory calculation does in this phase space bin. For low $p_{T,\text{avg}}$ -values, also the LO prediction agrees well with data in this y_b - y^* -bin.

In comparison with the previously studied high- y^* region, the JES uncertainties in this high- y_b region are slightly smaller, but larger than in the central bin yb0ys0. Here, in yb2ys0, the JES uncertainties stay well below 10%, in most $p_{T,\text{avg}}$ -bins even below 5%. Results extracted from this y_b - y^* -bin should be limited to the low $p_{T,\text{avg}}$ -range to assure for statistical reliability. With increasing $p_{T,\text{avg}}$ -values, the cross section results become more and more imprecise, as indicated by the orange error bars for the statistical uncertainty on data.

Intermediate y_b - y^* -bins and general observations: The results of the inclusive Z+jet cross section measurement in the medium y_b - y^* -range show similarities to the observations discussed above. In the low $p_{T,\text{avg}}$ -bins, MC simulation and fixed-order theory calculation give the most reliable description of data, much better than in higher $p_{T,\text{avg}}$ -bins.

In this intermediate y_b - y^* -range, the LO theory calculations are not giving a prediction of the data spectrum as good as MC simulation and NLO do.

Uncertainties on the JES in these remaining y_b - y^* -bins are mostly lower than 5%. Statistical uncertainties on data increase when going to higher $p_{T,\text{avg}}$ -values, as for the Z+jet process, this phase space region is connected to low event counts.

Generally, when looking at all the six y_b - y^* -bins, it is observed that the Z+jet production process takes predominantly place with $p_{T,\text{avg}}$ -values far lower than 1 TeV. Especially in the extreme y_b - y^* -bins, the spectra are ending already at $p_{T,\text{avg}}$ -values between 100 and 200 GeV. This is also visible in the cross section overview in figure 4.8. Another phenomenon concerning all y_b - y^* -bins in the Z+jet studies is that the spectra of data and LO theory prediction show large differences, which is as well already seen in the cross section overview in figure 4.8.

Overall, statistical uncertainties on data are significantly higher than in the dijet analysis, whereas the JES uncertainties are noticeably smaller. One reason for this is, that in the

Z+jet final state only the jet contributes to the JES uncertainties, whereas the Z boson does not. In contrary to this, in dijet both objects under investigation are hadronic jets and therefore, the dijet results suffer from higher JES uncertainties than the Z+jet results do.

The results presented in this chapter are used in the following chapter 5 for the comparison of the dijet over Z+jet ratio Υ in data, MC simulation and fixed-order theory calculations, in the form of double ratios.

Chapter 5

Combined Studies: Cross Section Ratios

This chapter combines the individual analyses of the inclusive cross section measurements, presented in chapter 4.3 for dijet and in chapter 4.4 for Z+jet. Advantages of studying cross section ratios are known from, for instance, the ratio R_{32} investigating the ratio of the inclusive cross section for 3-jet events to the one for 2-jet events, which gives insights on the strong coupling constant α_s [30]. Various examples of such cross section ratio measurements exist, one of the first ones being published by the UA1 collaboration in 1985 [69]. Another example of this kind is the above mentioned ratio R_{32} published by the CMS collaboration in 2013 [70]. Or the cross section ratio of top-quark pair to Z boson production, published by the ATLAS collaboration in 2017 [71].

Cross section ratios prove to be advantageous, as by taking the ratio of the cross sections of two different physics processes, common systematical uncertainties are expected to partly cancel, while others even cancel completely in the combined measurement. In the following, the results obtained from cross section ratio studies in the form of $\frac{\sigma_{\text{dijet}}}{\sigma_{\text{Z+jet}}}$ are discussed.

5.1 Cross Section Ratios: Dijet over Z+jet

After carrying out the individual dijet and Z+jet analyses, ratios of cross sections are investigated. Following the studies presented in chapters 4.3 and 4.4, the ratios are taken for the cross sections obtained from data as well as for the ones gained from MC simulation and NLO theory calculation.

As discussed in the previous chapters, the MC simulation set used for studying the inclusive dijet cross section is created using Pythia8 only, while the inclusive Z+jet cross section is obtained from a MC sample including both MadGraph and Pythia8, because these two samples describe the respective data best.

In this combined study for dijet and Z+jet, the range in $p_{\text{T,avg}}$ is reduced compared to the individual analyses. From below, the $p_{\text{T,avg}}$ -limit is given by the lowest single-jet trigger used in the dijet analysis. Higher values in $p_{\text{T,avg}}$ are excluded in the ratio studies, because the Z+jet analysis in data does not yield sufficiently high event counts at higher $p_{\text{T,avg}}$ -values. Accordingly, these limits for the $p_{\text{T,avg}}$ -range are used for all of the ratio studies in data, MC simulation and NLO theory calculation, presented in the following. After this reduction of the cross section histograms to the common $p_{\text{T,avg}}$ -range, the cross section ratio Υ , defined in equation 5.1, with dijet over Z+jet is calculated.

$$\Upsilon = \frac{\sigma_{\text{dijet}}}{\sigma_{\text{Z+jet}}} \quad (5.1)$$

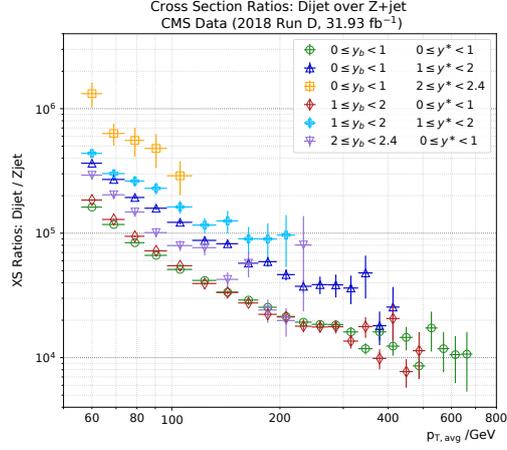
The resulting spectra are displayed in figure 5.1 separately for data, MC simulation, and theory prediction in NLO precision. The subfigures 5.1a for data, 5.1b for MC simulation and 5.1c for NLO theory are summarising the results for the "dijet over Z+jet" cross section ratio Υ yielded in the six individual y_b - y^* -bins. Additionally, the respective statistical uncertainty on the cross section ratio Υ is displayed in all of the three subfigures. The colour chosen for each of the y_b - y^* -bins is the same as used previously.

In data as well as in the simulations, the cross section ratio of the inclusive dijet over Z+jet cross section measurements lies between $4 \cdot 10^3$ and $4 \cdot 10^6$, measured in the $p_{\text{T,avg}}$ -range from 50 GeV up to 800 GeV. The strength of the dijet process' dominance decreases with increasing $p_{\text{T,avg}}$, but even at the high end of the $p_{\text{T,avg}}$ -range the inclusive dijet cross section is several orders of magnitude larger than the inclusive Z+jet cross section. This is expected due to the dijet process being a pure QCD phenomenon, which have the highest cross sections in pp-collisions, while the production of a Z boson implies a weak process and requires to produce a massive object as compared to massless partons. Since the inclusive dijet spectrum is falling more steeply than the Z+jet spectrum does, as can be seen in the cross section overviews given in figures 4.6 and 4.8, as well as in the studies presented in ref. [55] and ref. [53], the cross section ratio Υ decreases with increasing $p_{\text{T,avg}}$ -values.

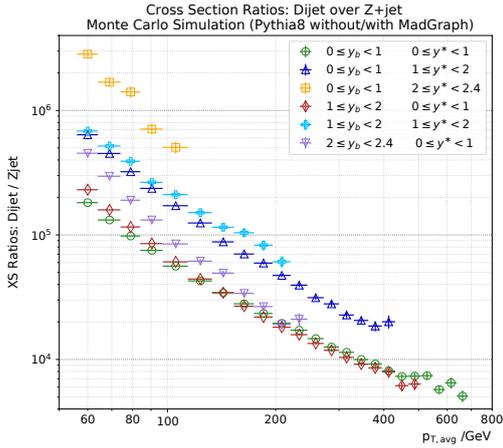
Viewing the three subfigures in figure 5.1, the very first conclusion is that the simulations give smoother spectra than data does. This is not surprising when looking at the statistical uncertainties assigned to the individual data points in the subfigures. For data, the statistical uncertainties are visibly higher than for MC simulation and NLO theory prediction. This is already seen in the individual analyses in chapter 4.3 and 4.4. The following considerations hold for all of the three cases, data, MC simulation and NLO theory.

The way in which the cross section ratio decays is similar in some of the y_b - y^* -bins. Such similarity is for instance shown for bin yb0ys0 with yb1ys0 and bin yb0ys1 with yb1ys1. These pairs of y_b - y^* -bins correspond to the same rapidity separation interval y^* with respect to each other, while the range in y_b differs. All in all, the influence of the rapidity separation y^* of the two leading objects on the cross section ratio Υ is stronger than the effect of the average boost of the system y_b . Studies on the partonic subprocesses contributing to the dijet and the Z+jet cross sections and their consequences for the ratio would provide a better understanding of the differences in the six y_b - y^* -bins. Separately, investigations of the contributing subprocesses are for instance presented in ref.[19] and ref.[53] for the Z+jet process and in ref.[55] for the dijet production.

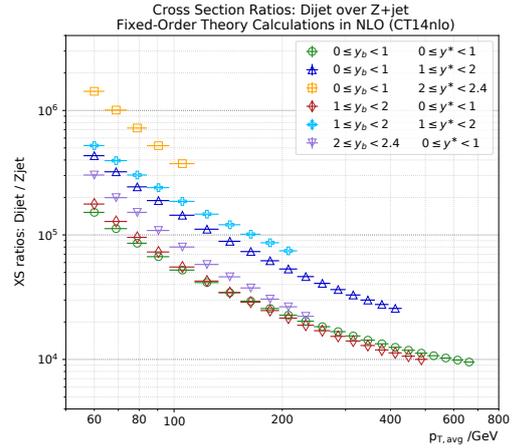
When comparing the results on the cross section ratio achieved in this dedicated study to similar ratios derived from separate previous cross section determinations, for instance ref. [55] for dijet and [53] for the $Z(\rightarrow \mu\mu)$ +jet process, the results are found to be compatible.



(a) Cross section ratio results Υ_{data} obtained from data.



(b) Cross section ratio results Υ_{MC} obtained using MC simulation.



(c) Cross section ratio $\Upsilon_{\text{NLO theo.}}$ resulting from fixed-order theory calculations in NLO precision.

Figure 5.1: Results of the ratio studies taking the ratio Υ of the triple-differentially measured dijet cross section with the Z+jet cross section measured in an analogous manner using the same observables. The subfigures correspond to the ratio Υ obtained from data, MC simulation and fixed-order theory calculation with error bars indicating the statistical uncertainty where visible. The results are shown for the six y_b - y^* -bins.

5.2 Double Ratios: Cross Section Ratio Υ in Data compared to Simulation Results

In this section, the ratio in data Υ_{data} is compared to MC simulation results Υ_{MC} and fixed-order theory calculations in NLO $\Upsilon_{\text{NLO theo.}}$.

The comparison of data and simulation is presented in this chapter in the form of double ratios $\frac{\Upsilon_{\text{simulation}}}{\Upsilon_{\text{data}}}$ that are calculated from the previously described cross section ratios in data, MC simulation and NLO predictions.

The subfigures in figure 5.2 show the results in analogy to the previous comparisons. The brown points show the cross section ratio results obtained from MC simulation divided by the results gained from data. Assigned to these brown points are the statistical uncertainties $\delta_{\text{MC stat. unc.}}$ on the cross section ratios Υ in the MC simulation, together with the double ratio plotted as $\frac{\Upsilon_{\text{MC}} \pm \delta_{\text{MC stat. unc.}}}{\Upsilon_{\text{data}}}$.

Fixed-order theory calculation results of Υ in NLO precision are displayed in black. Together with error bars corresponding to the PDF uncertainties for CT14NLO, the theory predictions are displayed as $\frac{\Upsilon_{\text{NLO theo.}} \pm \delta_{\text{corr. PDF unc.}}}{\Upsilon_{\text{data}}}$. In the calculation of the PDF uncertainties, correlations between dijet and Z+jet are taken into account. The shaded error band in light brown behind the NLO theory results corresponds to the scale uncertainty $\delta_{\text{scale unc.}}$ on $\Upsilon_{\text{NLO theo.}}$ for the case where scale uncertainties between dijet and Z+jet are assumed as uncorrelated. This uncertainty band formulated as equation reads as $\frac{\Upsilon_{\text{NLO theo.}} \pm \delta_{\text{uncorr. scale unc.}}}{\Upsilon_{\text{data}}}$. Statistical uncertainties on the fixed-order theory calculation are not displayed as these stay well below 1% in almost all phase space bins and are therefore negligibly small compared to the scale and PDF uncertainties assigned to these theory results.

The statistical uncertainty $\delta_{\text{data stat. unc.}}$ on the cross section ratio in data Υ_{data} is drawn as orange error bars around the $\Upsilon_{\text{simulation}}/\Upsilon_{\text{data}} = 1.0$ line.

The correlated JES uncertainties on the cross section ratio Υ are displayed in the same style as in chapter 4.3 and 4.4. The total JES uncertainty is depicted in light green as shaded error band around $\Upsilon_{\text{simulation}}/\Upsilon_{\text{data}} = 1.0$. Individually displayed with red and blue lines are the JES flavour uncertainty and the total JES uncertainty without the flavour component, respectively. For the calculation of the correlated JES flavour uncertainty, the *jesQCDFlavor* is used for dijet and the *jesZJetFlavor* source for the Z+jet part of the fraction.

New in the subplots in figure 5.2 is the shaded error band in light blue colour. This error band illustrates the total JES uncertainty on the cross section ratio for the case if the JES uncertainties on the dijet and Z+jet measurements would be completely uncorrelated. Hence, the light blue band corresponds to the quadratic sum of the total JES uncertainties on the dijet and on the Z+jet cross section measurement.

5.2 Double Ratios: Cross Section Ratio Υ in Data compared to Simulation Results

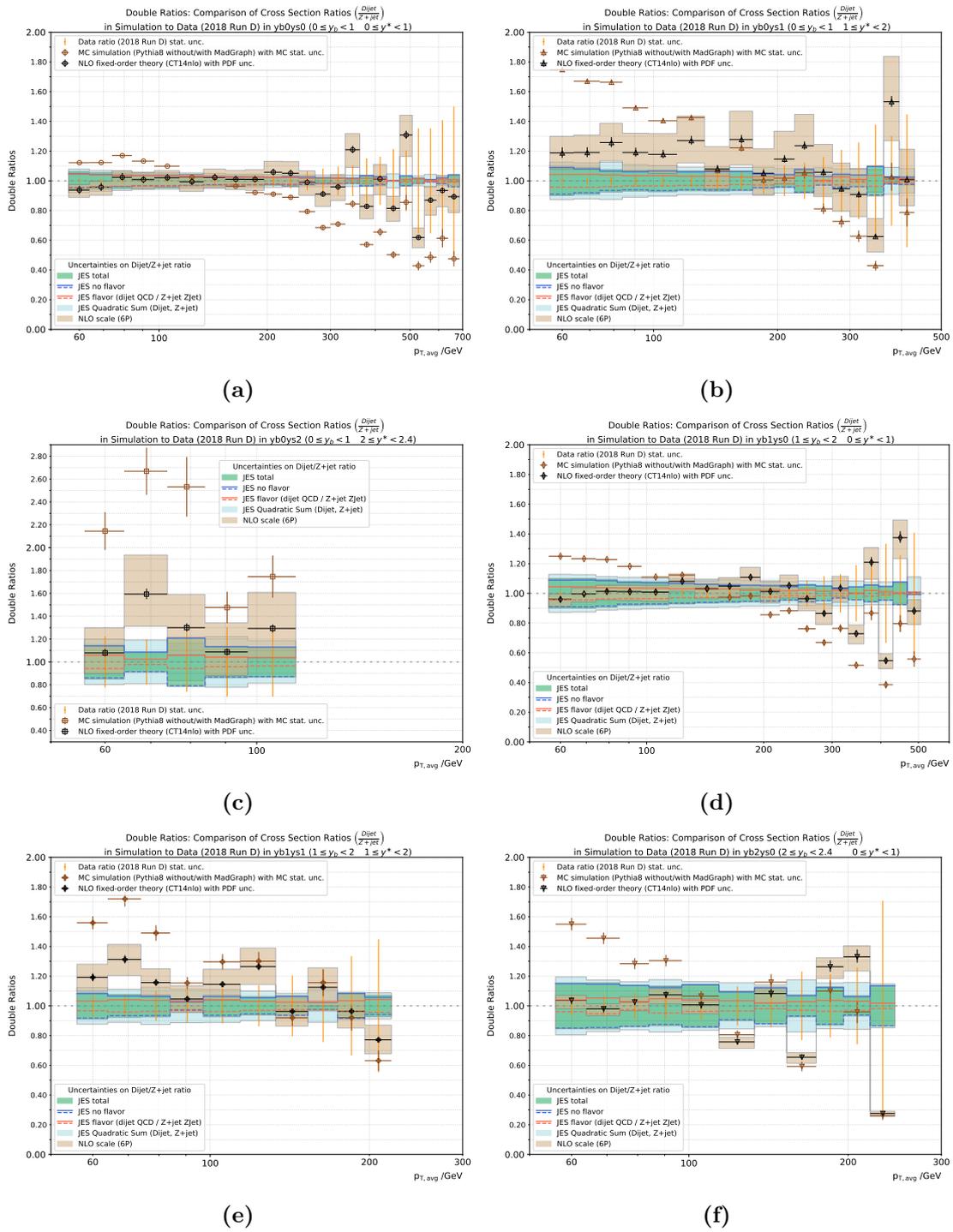


Figure 5.2: Double Ratios: Cross section ratios Υ obtained from MC simulation and fixed-order theory calculation in NLO are divided by the result for Υ gained from data. Each of the subfigures represents one of the six individual y_b - y^* -bins. The total JES uncertainty on Υ is indicated for the correlated as well as for the uncorrelated case.

The central bin – yb0ys0: The first y_b - y^* -bin under investigation is the central bin yb0ys0, displayed in subfigure 5.2a, which yields the most accurate results. It is immediately clear that the description of Υ by MC simulation shows a behaviour different from data. But up to the medium- $p_{T,\text{avg}}$ region, the deviations from Υ_{MC} to Υ_{data} stay within 20%. The results for the comparison of MC simulation to data shows a slope, which could be caused by the difference in generators used for the simulation of the dijet and the Z+jet spectrum. Also previous studies have observed such a slope in the behaviour of MadGraph samples when compared to data (cf. figure 5.14 in ref.[55]). Nevertheless, the MC simulation-based cross section ratio Υ_{MC} does not deviate from the results in data Υ_{data} by more than 60% over the whole $p_{T,\text{avg}}$ phase space analysed in yb0ys0.

The description of the cross section ratio Υ by the fixed-order theory calculation in NLO lies closer to data than the one given by MC simulation does. Differences in the comparison of $\Upsilon_{\text{NLO theo.}}$ to the ratio in data Υ_{data} stay within 10% for the low- and medium- $p_{T,\text{avg}}$ values accessible in this central y_b - y^* -bin. This holds true even though the data is not unfolded, while the NLO calculation does not include any detector simulation. For higher $p_{T,\text{avg}}$ -values, the fixed-order theory calculation results are fluctuating up and down when divided by data. This is caused by the behaviour of the data ratio Υ_{data} observed in figure 5.1a. However, these fluctuations are covered by the uncertainties $\delta_{\text{corr. PDF unc.}}$ and $\delta_{\text{uncorr. scale unc.}}$ on the theory calculation $\Upsilon_{\text{NLO theo.}}$ and the statistical uncertainties $\delta_{\text{data stat. unc.}}$ on the data ratio Υ_{data} , which are increasing quickly in this high- $p_{T,\text{avg}}$ region.

Since the detector is best understood in this central y_b - y^* -bin, the JES uncertainties in general stay small in this bin, mostly below 5% and are not exceeding 10% even for the uncorrelated case. When comparing the correlated case (light green error band) with the uncorrelated case (light blue error band), it is observed that the reduction of JES uncertainties due to partial cancellation in the ratio Υ is very small in this y_b - y^* -bin. In the case of the JES uncertainties on the two analyses being completely uncorrelated, the uncertainties are only about 1% higher. Fluctuations in the JES uncertainties observed in the highest $p_{T,\text{avg}}$ -bins are of statistical origin and do not provide a basis for reliable statements. Considering the large statistical uncertainties $\delta_{\text{data stat. unc.}}$ on data in these bins, information extracted from this high- $p_{T,\text{avg}}$ region has to be treated with caution.

Large rapidity separation – yb0ys2: Next, one of the two extreme phase space bins, namely yb0ys2 with large rapidity separation y^* of the leading objects in the final state, is investigated. This y_b - y^* -bin has shown to be exceptional already in the individual dijet- and Z+jet studies in the previous chapters.

In this high- y^* bin, the predictions Υ_{MC} made by MC simulation are differing from the data result Υ_{data} visibly stronger than in the central y_b - y^* -bin. Even in the low $p_{T,\text{avg}}$ -bins, that are otherwise well determined and least affected by uncertainties in comparison to the medium- and high- $p_{T,\text{avg}}$ range, the cross section ratio Υ_{MC} differs from data Υ_{data} by more than 110% and up to almost 170%. Also in the other $p_{T,\text{avg}}$ bins, the deviation from MC simulation results to data results for the cross section ratio Υ is large.

Despite missing detector simulation, the NLO prediction provides a more accurate description of the cross section ratio Υ than the MC simulation does. In all of the analysed $p_{T,\text{avg}}$ -bins, the result for Υ is overestimated by the fixed-order theory calculation in NLO. Generally, the theory prediction differs from data by 10% to 60% in the available $p_{T,\text{avg}}$ bins. The differences in the results $\Upsilon_{\text{NLO theo.}}$ and Υ_{data} are covered by the statistical and systematical uncertainties on these both cross section ratio results.

As expected based on the individual analysis in the previous chapters, the statistical uncertainties $\delta_{\text{data stat unc.}}$ on the cross section ratio in data Υ_{data} are high in this extreme y^* -region. Again, it has to be concluded, that the most reliable results can be gained in the low- $p_{T,\text{avg}}$ region.

Compared to the central y_b - y^* -bin, the JES uncertainties in this high- y^* region are significantly larger. At the same time, the effect of the cancellations of systematical uncertainties in the cross section ratio Υ is more visible, when comparing the respective error bands. In the uncorrelated case, the JES uncertainties on Υ are on average around 20%. When compared to the error band, where the correlations in the JES uncertainties of the dijet and Z+jet measurements are taken into account, a reduction of, on average over the $p_{T,\text{avg}}$ -range, around 5% on the total JES uncertainty can be gained.

Strongly boosted system – yb2ys0: The second extreme y_b - y^* -bin of interest is the high y_b -region, yb2ys0, shown in subfigure 5.2f. In this y_b - y^* -bin, the system of the two leading objects in the final state is strongly boosted.

In the high- y_b region, the cross section ratio Υ_{MC} obtained from MC simulation differs by a maximum of 75% from the ratio in data Υ_{data} . This means, the agreement between MC simulation and data results for Υ in this bin yb2ys0 with strong boost y_b is better than in the previously discussed extreme bin yb0ys2 with large rapidity separation y^* of the final-state objects. In the medium- $p_{T,\text{avg}}$ region, the deviation is within 30% and covered by uncertainties in all but two bins.

Again, the NLO prediction $\Upsilon_{\text{NLO theo.}}$ proves to be more compatible with data than the MC simulation Υ_{MC} is. Over the whole displayed $p_{T,\text{avg}}$ -range, the fixed-order predictions $\Upsilon_{\text{NLO theo.}}$ for the cross section ratio in NLO precision agree with Υ_{data} within uncertainties. The scale uncertainties are calculated as 6P-variations individually for dijet and Z+jet, thus being treated as uncorrelated. Despite being calculated for the completely uncorrelated case, the scale uncertainties on the theory calculation are rather small in this y_b - y^* -bin.

Even for the assumption of uncorrelated JES uncertainties, shown as light blue shaded band, their values stay below 20% in the yb2ys0 bin. When comparing these to the correlated JES uncertainties of Υ , a small reduction in the order of a few percent can be observed.

Medium y_b - y^* -bins and general observations: The remaining y_b - y^* -bins are the intermediate ones yb0ys1, yb1ys0 and yb1ys1. Also the medium y_b - y^* -bins show a negative slope in the double ratio $\frac{\Upsilon_{\text{MC}}}{\Upsilon_{\text{data}}}$ when going from low- to high- $p_{T,\text{avg}}$ bins, which can be caused by having a Pythia8-only sample for dijet in the numerator and a

Pythia8+MadGraph sample for Z+jet in the denominator of Υ_{MC} . The use of MadGraph is partly responsible for this trend. The results for the double ratio $\frac{\Upsilon_{\text{simulation}}}{\Upsilon_{\text{data}}}$ when using MadGraph for matrix-element generation and Pythia8 for parton showering in both samples, dijet and Z+jet, are shown in appendix A.4.1. These results are free from such a slope over the $p_{\text{T,avg}}$ -range. However, the Pythia8+MadGraph sample was found to give only an unsatisfying description of the inclusive dijet spectrum. For this reason, and due to missing availability of Pythia8-only samples in nanoAOD format for Z+jet, this hybrid investigation of Υ_{MC} has been chosen.

The yb1ys0 bin, with small rapidity separation, yields results comparable in accuracy to the central $y_{\text{b}}-y^*$ -bin yb0ys0. Going one bin higher in y^* , to yb0ys1 and yb1ys1, the deviation of Υ_{MC} from Υ_{data} is larger.

Below 300 GeV, the deviation of $\Upsilon_{\text{NLO theo.}}$ from Υ_{data} is below 40% for all of the three medium $y_{\text{b}}-y^*$ -bins. Again, comparing these three remaining $y_{\text{b}}-y^*$ -bins, the agreement of fixed-order calculation $\Upsilon_{\text{NLO theo.}}$ with data Υ_{data} is better in the low- y^* case, namely bin yb1ys0. In yb0ys1 and yb1ys1 the NLO prediction is mostly overestimating the cross section ratio Υ , whereas in yb1ys0 the double ratio lies around $\frac{\Upsilon_{\text{NLO theo.}}}{\Upsilon_{\text{data}}} = 1.0$. In the high- $p_{\text{T,avg}}$ region, such general statements cannot be made as easily with fluctuations of $\Upsilon_{\text{NLO theo.}}$ into both directions.

In general, the JES uncertainties in these three medium $y_{\text{b}}-y^*$ -bins are smaller than in the extreme ones of high y_{b} or high y^* . Nevertheless, even though the difference is small, it can be also seen in subfigures 5.2b, 5.2d and 5.2e that the JES uncertainties on the ratio Υ are slightly reduced when compared to the uncorrelated case of JES uncertainties.

Overall, it is observed in all six $y_{\text{b}}-y^*$ -bins that the NLO theory description $\Upsilon_{\text{NLO theo.}}$ of the cross section ratio Υ agrees better with data than Υ_{MC} does. Again it should be considered that this fixed-order calculation does not include any detector simulation and $\Upsilon_{\text{NLO theo.}}$ is therefore expected to be at least slightly off in comparison to data. However, it is possible that some of the detector effects cancel in the double ratio.

For all $y_{\text{b}}-y^*$ -bins, the uncertainty on the luminosity is given with 2.5% and the same for both individual analyses, dijet and Z+jet. For this reason, it is one of the uncertainties that cancels out when taking the cross section ratio Υ of these two. The JES uncertainties in the correlated case are slightly reduced with respect to the quadratic sum of the individual analyses' JES uncertainties.

Chapter 6

Conclusion and Outlook

In this analysis, the cross section measurements for two different processes, each of them important on its own behalf, have been carried out. Those two measurements were combined in the form of cross section ratios, in order to reduce the systematic uncertainties on the overall measurement.

The presented thesis is based on data collected with the CMS detector during run D of the 2018 data taking period at the LHC. The analysed data correspond to an integrated luminosity of 31.93 fb^{-1} and were recorded during Run II of the LHC, where proton-proton collision events took place at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$.

The first subject of this study is the inclusive dijet cross section spectrum. Due to the abundant production of hadronic jets in high-energetic pp-collisions at the LHC, this process is the background of many other processes and hence the precise understanding of the dijet spectrum is crucial also for other analyses. As second event topology, the inclusive Z+jet production has been chosen. The properties of the Z boson are well known and the cross section measurement of the Z+jet production process via the $Z \rightarrow \mu\mu$ decay channel is a suitable observable, since it provides a clear signal that can be measured precisely with the CMS detector. Allowing for precise energy measurement, the Z+jet process also plays an important role in the jet energy calibration performed within the CMS collaboration.

The two cross section measurements have been carried out triple-differentially. As observables the average transverse momentum $p_{T,\text{avg}}$, as well as the average boost y_b and the rapidity separation y^* of the two leading physics objects in the final state have been chosen. The rapidity-based variables y_b and y^* are useful for investigating effects in the cross section results coming from the hard matrix-element via y^* and coming from the PDFs via y_b .

The measurement of the cross section ratio $\Upsilon = \frac{\sigma_{\text{dijet}}}{\sigma_{\text{Z+jet}}}$, which has been performed for the first time, has been compared to the results obtained from MC simulations and fixed-order theory calculations in NLO precision in the form of double ratios $\frac{\Upsilon_{\text{simulation}}}{\Upsilon_{\text{data}}}$. The cross section ratio Υ_{MC} is calculated from the cross sections obtained using Pythia8 for dijet and MadGraph+Pythia8 for Z+jet, since these generators were found to describe the respective spectrum most accurately.

The results of the analysis presented in this thesis confirm that it is advantageous to study these two processes as described above. Thus, both measurements can be combined into a single ratio, hereby achieving a reduction of systematic uncertainties. A reduction of JES uncertainties is observed in all of the six investigated y_b - y^* -bins. The effect of cancellation

of systematic uncertainties is found to be larger, in the order of a few percent, in the high- y_b and high- y^* bins. Conclusively, this kind of studies can be beneficial especially in the phase space regions where large rapidities contribute.

With more data becoming available, statistical uncertainties, which are particularly high in the two y_b - y^* -bins of high- y_b or high- y^* , will be reduced. For further studies an improvement can therefore be made by combining all four 2018 data sets or the full Run II data, thereby minimising statistical uncertainties. In the case where the measurement is not dominated by statistical uncertainties, the presented cross section ratio Υ of the inclusive dijet and Z+jet cross section measurement can provide improved insights because of reduced systematic uncertainties. By choosing the three observables as presented in this thesis, the cross section ratio Υ offers the basis for, e.g., PDF studies and also for studies of the strong coupling constant α_s . For PDF studies it is important to understand the individual contributions of different subprocesses to the cross section of the two processes and their ratio. Differences in these production processes can be studied based on the procedures carried out in the presented analysis. For this purpose it is advantageous to define the two cross sections in an analogous way as suggested here.

For better comparability with theory predictions, it is necessary to unfold the data to account for detector effects. Furthermore, additional uncertainties relevant for the Z+jet measurement must be included in the considerations. Response matrices, which are important for estimating jet energy resolution (JER) uncertainties through unfolding, have been derived in the scope of this thesis, based on MC samples. These indicate that unfolding is possible without the need for regularisation.

Concerning MC simulations, additional generator combinations should be studied. Specifically re-performing the presented analysis with MC samples using the same generator-setup for dijet and Z+jet can help to evaluate differences caused by the choice of the MC event generator. Additionally, fixed-order theory predictions in next-to-next-to-leading order (NNLO) precision will provide a more accurate description of the cross section ratio Υ for dijet over Z+jet. Moreover, potential cancellations in theoretical uncertainties on the ratio should be investigated.

Conclusively, the feasibility of a ratio measurement of inclusive dijet and Z+jet cross sections has been demonstrated. Some reduction in systematic uncertainties has been observed. This thesis therefore serves as a starting point for future explorations of this cross section ratio and potential improvements to determinations of the PDFs and α_s .

Appendix A

Appendix

A.1 Data Sets used in the Analysis

The data sets used in this analysis are shown in table A.1. The MC simulation sets are listed in table A.2.

A.1.1 Data Sets

Process	Name	usage
dijet	/JetHT/Run2018A-Nano1June2019-v2/NANOAOD	only appendix
	/JetHT/Run2018B-Nano1June2019-v2/NANOAOD	only appendix
	/JetHT/Run2018C-Nano1June2019-v1/NANOAOD	only appendix
	/JetHT/Run2018D-Nano1June2019_ver2-v1/NANOAOD	main results
Z+jet	/DoubleMuon/Run2018A-Nano1June2019-v1/NANOAOD	only appendix
	/DoubleMuon/Run2018B-Nano1June2019-v1/NANOAOD	only appendix
	/DoubleMuon/Run2018C-Nano1June2019-v1/NANOAOD	only appendix
	/DoubleMuon/Run2018D-Nano1June2019_ver2-v1/NANOAOD	main results

Table A.1: Names of the data samples used in this analysis.

A.1.2 Monte Carlo Simulation Sets

Process	Name	usage
dijet	/QCD_Pt-15to7000_TuneCP5_Flat_13TeV_pythia8/	cross section
	RunIIAutumn18NanoAOD-102X_upgrade2018_realistic_v15_ext1-v1/NANOAODSIM	main results
dijet	/QCD_Pt-15to7000_TuneCH2_Flat_13TeV_herwig7/	JES
	RunIIAutumn18NanoAODv5-Nano1June2019_102X_upgrade2018_realistic_v19-v2/NANOAODSIM	uncertainty determination
dijet	/QCD_HT[<i>lower-HTedge</i>]to[<i>upper-HTedge</i>]_TuneCP5_13TeV-madgraphMLM-pythia8/NANOAODSIM/Nano1June2019_102X_upgrade2018_realistic_v19-v1/	cross check MadGraph
Z+jet	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/	cross section
	RunIIAutumn18NanoAODv5-Nano1June2019_102X_upgrade2018_realistic_v19-v1/NANOAODSIM	main results
Z+jet	/DYJetsToLL_M-50_TuneCH3_13TeV-madgraphMLM-herwig7/	JES
	RunIIAutumn18NanoAODv5-Nano1June2019_102X_upgrade2018_realistic_v19-v1/NANOAODSIM	uncertainty determination

Table A.2: Names of the MC simulation samples used in this analysis.

A.2 Overview of Run Periods 2018

A.2.1 Dijet Cross Sections: Overview of Run Periods 2018



Figure A.1: A first check of the compatibility of the four run periods A, B, C, D of the 2018 data taking period at CMS. The dijet cross section obtained for a run period X is divided by the cross section from run period D. Together with the statistical uncertainties this is displayed as $\frac{\sigma_X \pm \delta_{X, \text{stat. unc.}}}{\sigma_D}$.

A.2.2 Z+jet Cross Sections: Overview of Run Periods 2018

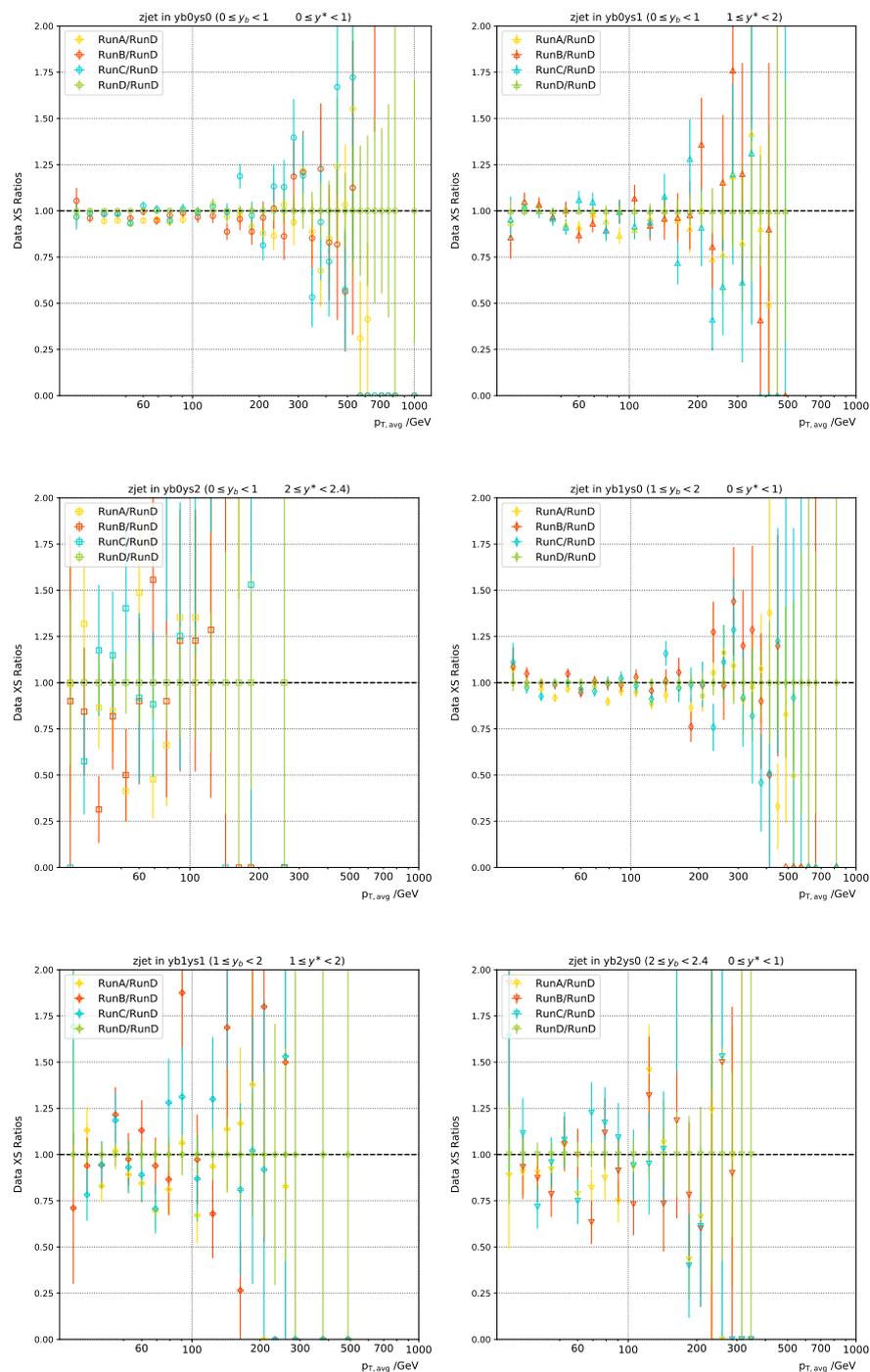


Figure A.2: A first check of the compatibility of the four run periods A, B, C, D of the 2018 data taking period at CMS. The Z+jet cross section obtained for a run period X is divided by the cross section from run period D. Together with the statistical uncertainties this is displayed as $\frac{\sigma_X \pm \delta_X^{\text{stat. unc.}}}{\sigma_D}$.

A.3 Different MC Generators

A.3.1 Dijet Results using Herwig7

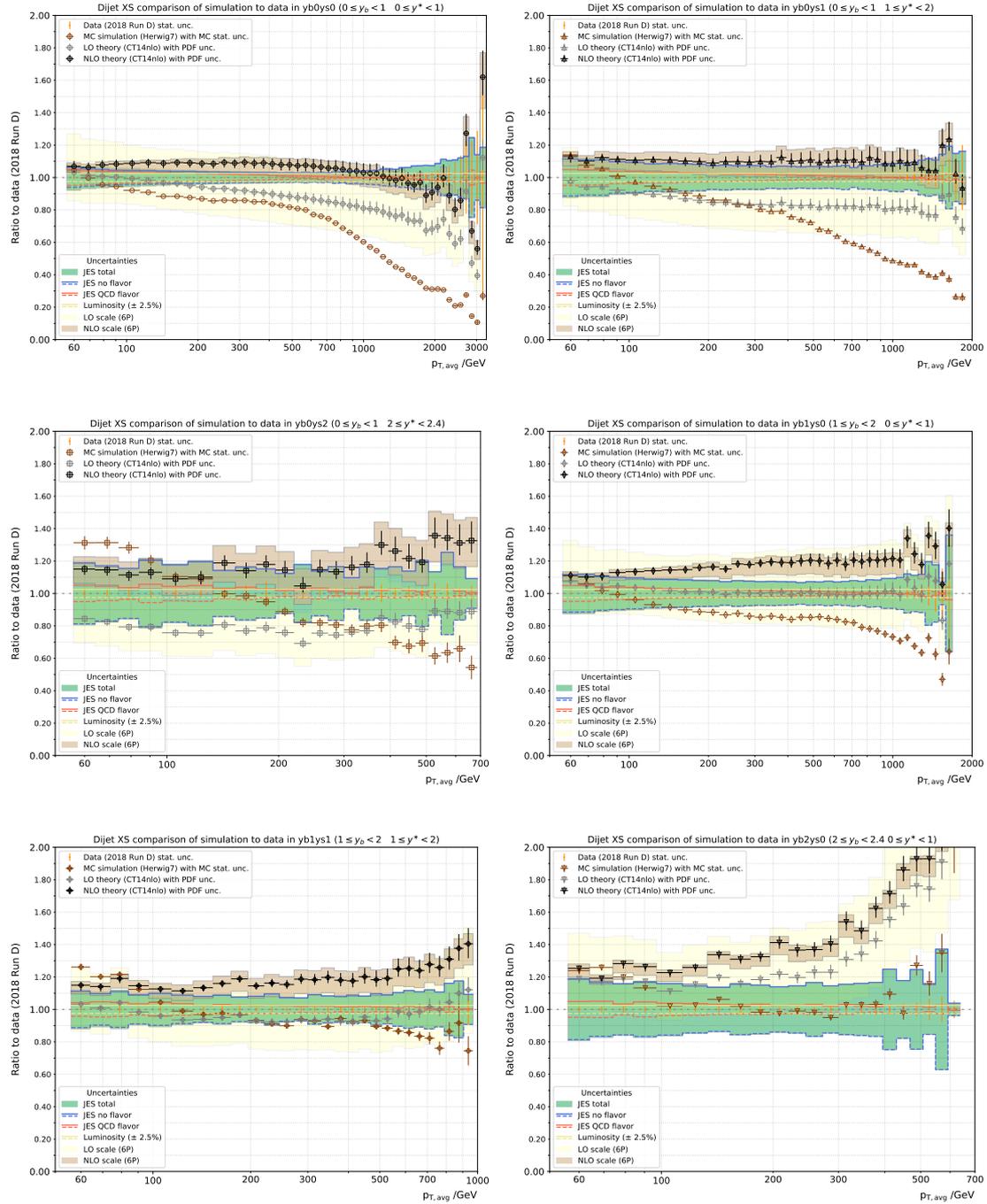


Figure A.3: Overview of the inclusive dijet cross section in the six y_b - y^* -bins. The MC simulation sample was created using Herwig7.

A.3.2 Dijet Results using Madgraph+Pythia8

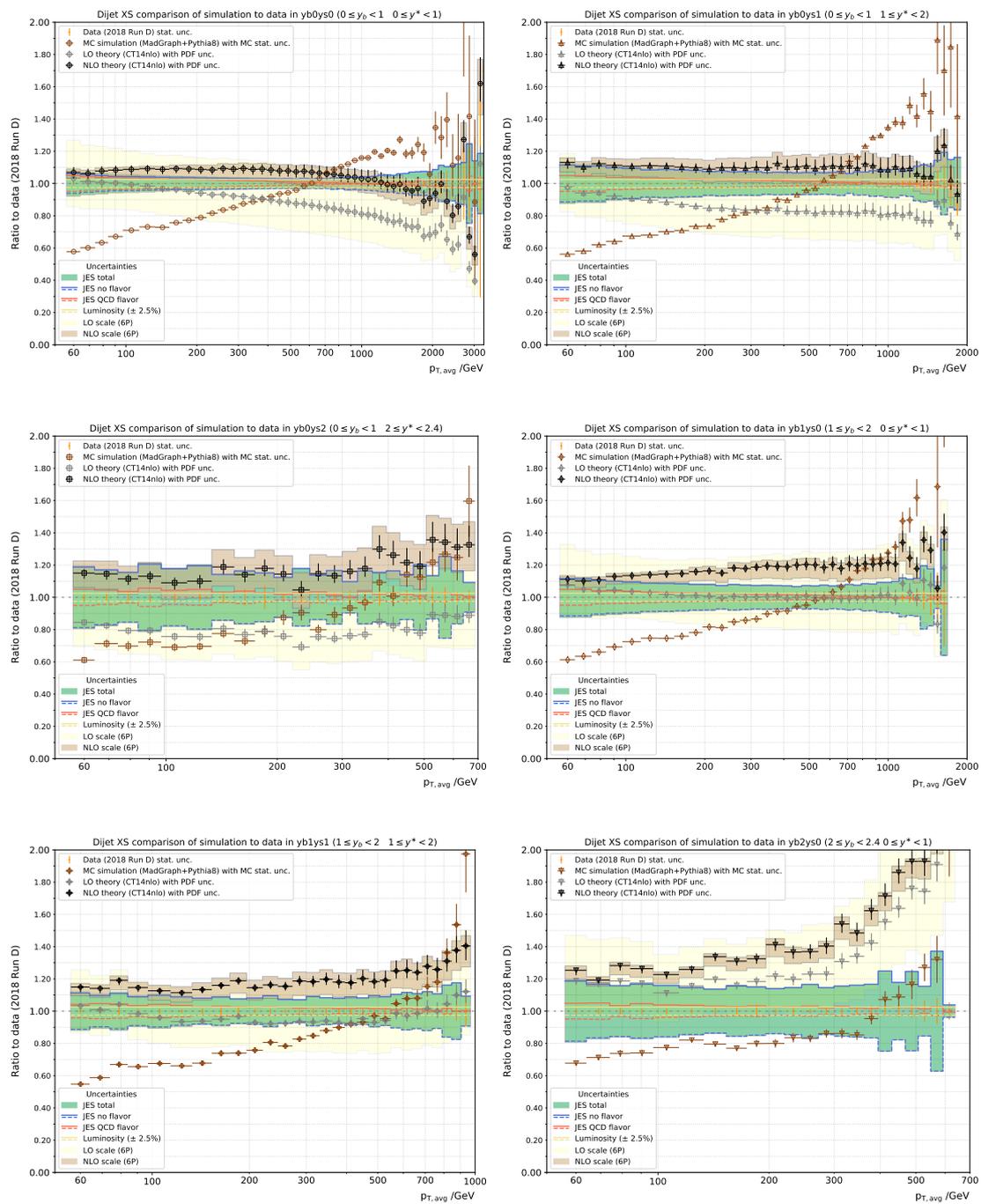


Figure A.4: Overview of the inclusive dijet cross section in the six y_b - y^* -bins. The MC simulation sample was created using MadGraph+Pythia8.

A.3.3 Z+jet Results using Madgraph+Herwig7

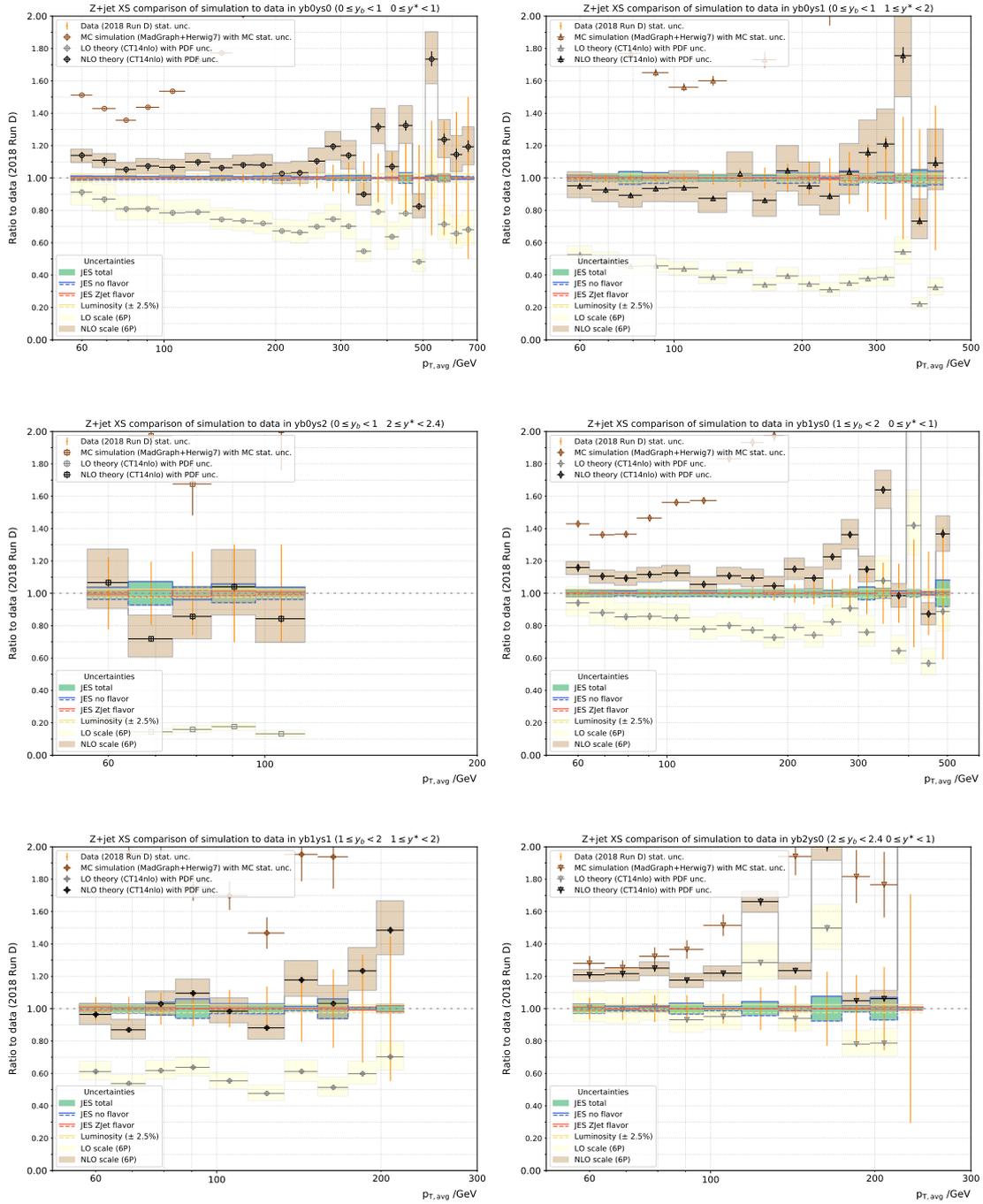


Figure A.5: Overview of the inclusive Z+jet cross section in the six y_b - y^* -bins. The MC simulation sample was created using MadGraph+Herwig7.

A.4 Additional Double Ratio Results

A.4.1 MadGraph+Pythia8 for Dijet and Z+jet

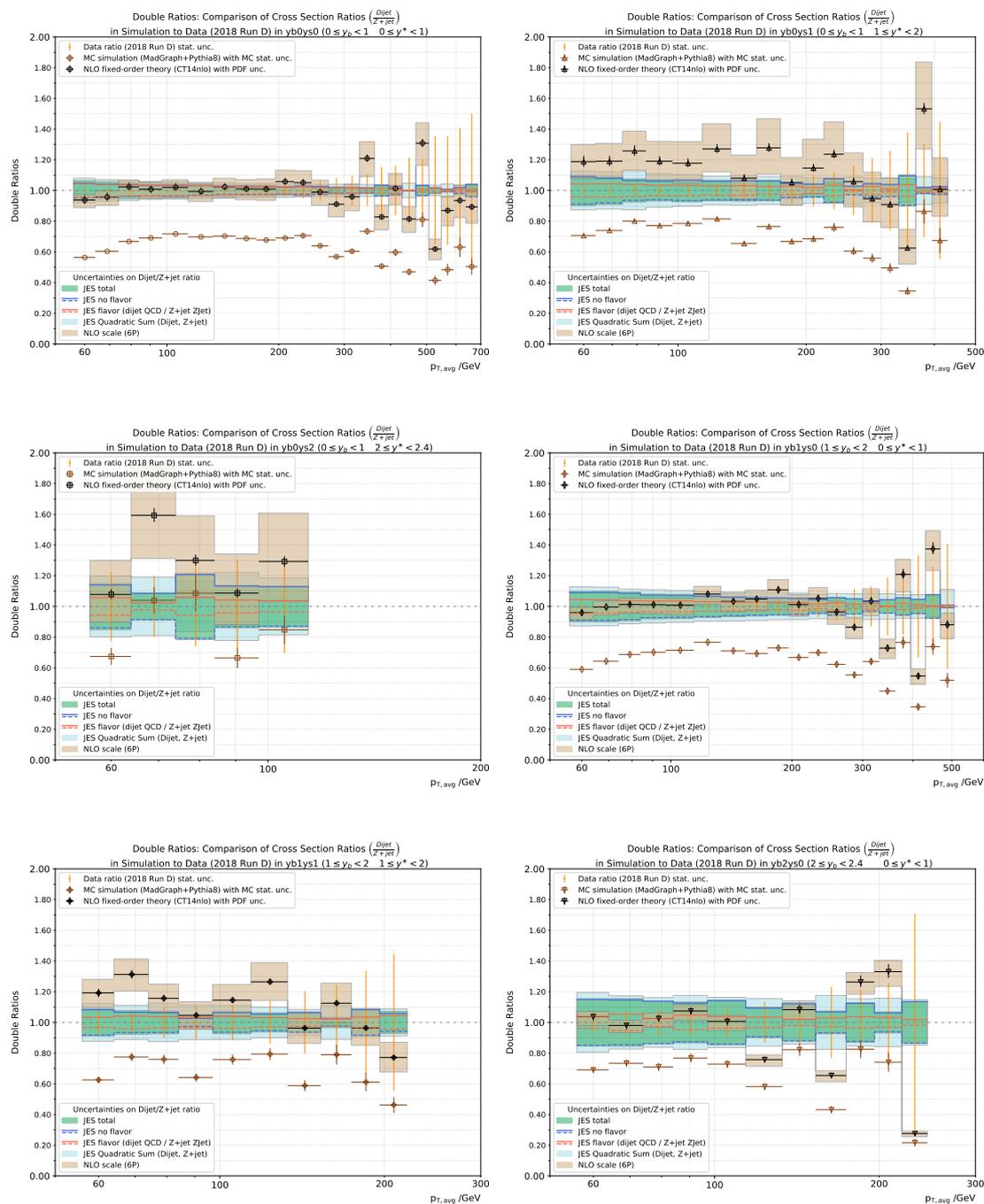


Figure A.6: Results for the double ratio when using MadGraph+Pythia8 for the simulation of the dijet as well as for the Z+jet process.

A.4.2 (MadGraph+)Herwig7 Double Ratio Results

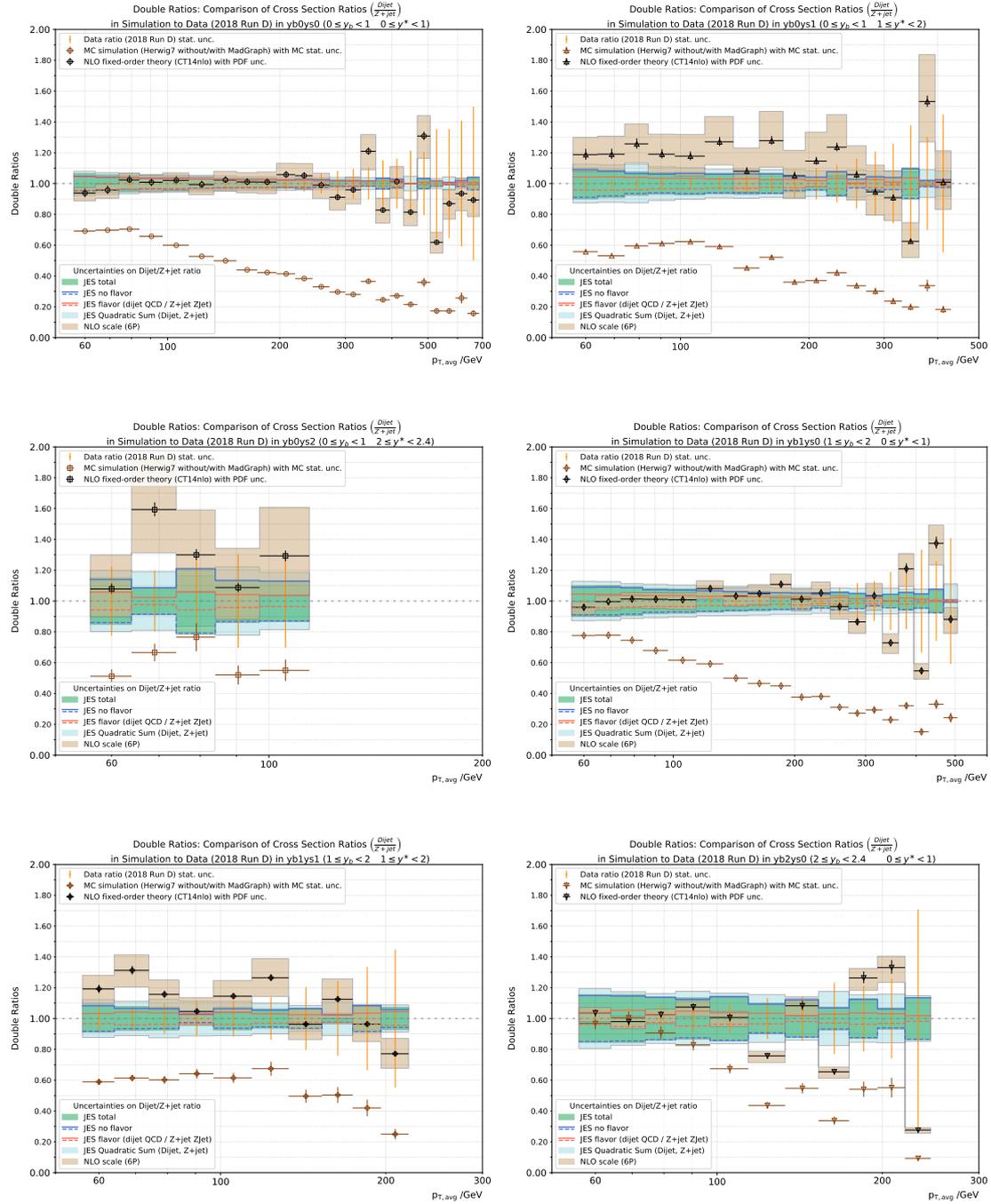


Figure A.7: The double ratio obtained using a Herwig7 sample for dijet and MadGraph+Herwig7 for Z+jet.

A.5 Preliminary Response Matrices

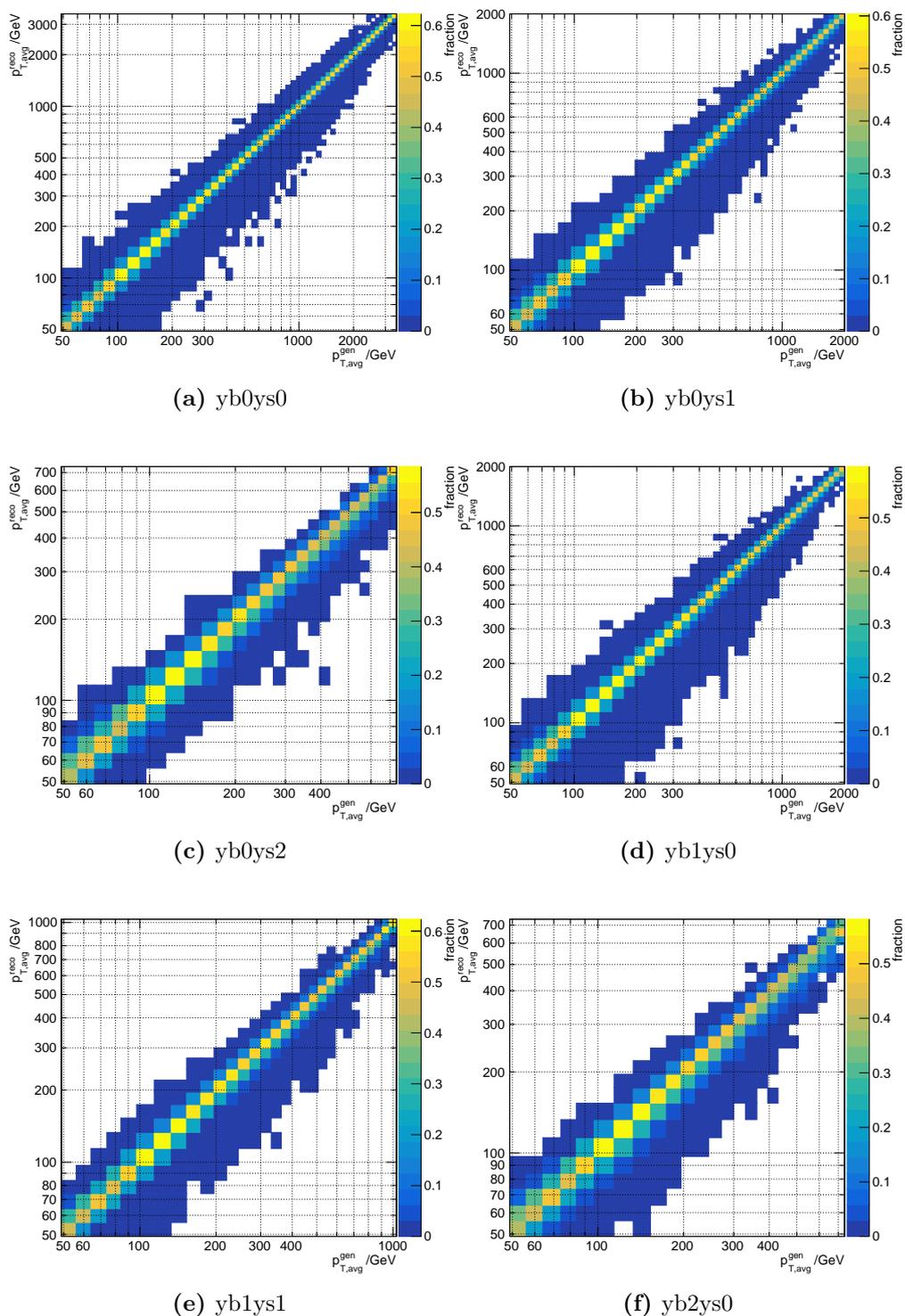


Figure A.8: Response matrices obtained from the Pythia8 dijet sample used for the analysis of the dijet process in this thesis. The subfigures indicate that unfolding is possible even without the need for regularisation.

A.6 Validation of Dijet Analysis 2017

In the course of this analysis, the results of the triple-differential cross section measurement of the dijet spectrum at $\sqrt{s} = 8$ TeV were validated by reproducing the figures in ref.[24] published by the CMS collaboration, contained in Georg Sieber's PhD thesis [55].

A.6.1 Dijet Cross Section Results

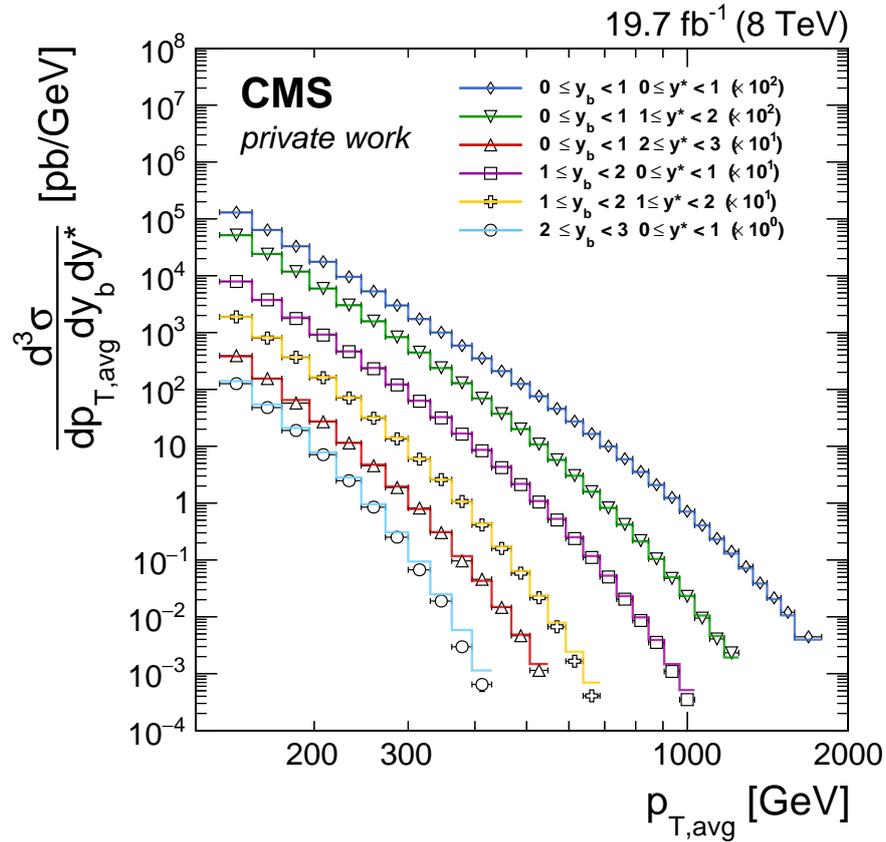


Figure A.9: Inclusive dijet cross section at $\sqrt{s} = 8$ TeV.

A.6.2 The k-Factor

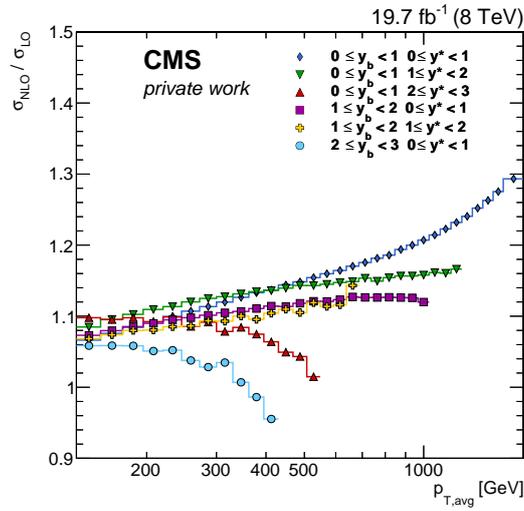
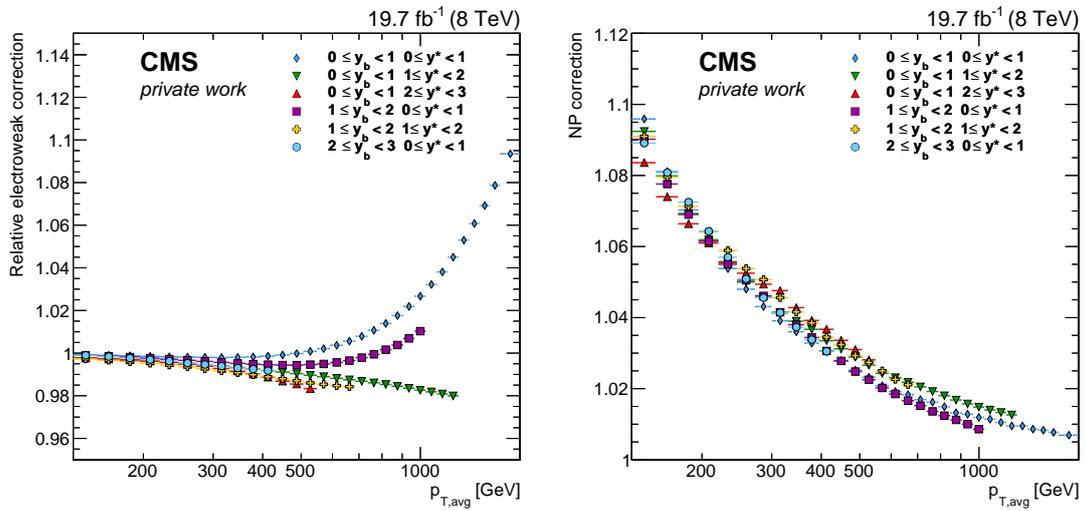


Figure A.10: The k-factor $\sigma_{\text{NLO}}/\sigma_{\text{LO}}$ evaluated using the NNPDF3.0 PDF set.

A.6.3 Correction Factors



(a) EW correction factors

(b) NP correction factors

Figure A.11: Electroweak and non-perturbative correction factors.

A.6.4 Relative Uncertainties

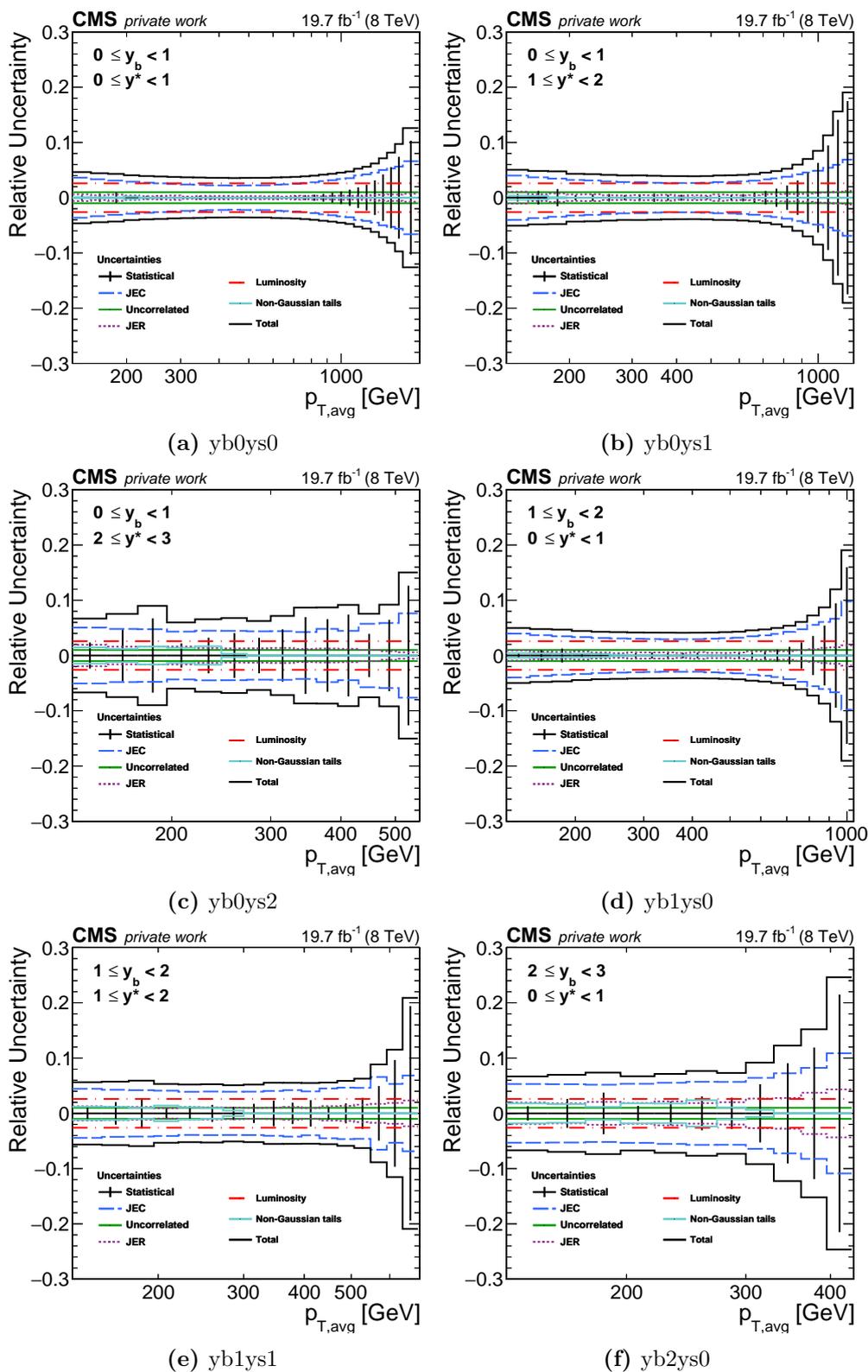


Figure A.12: Relative uncertainties on the measurement of the inclusive dijet cross section, displayed for the six y_b - y^* -bins. 79

A.6.5 Comparison to Simulation

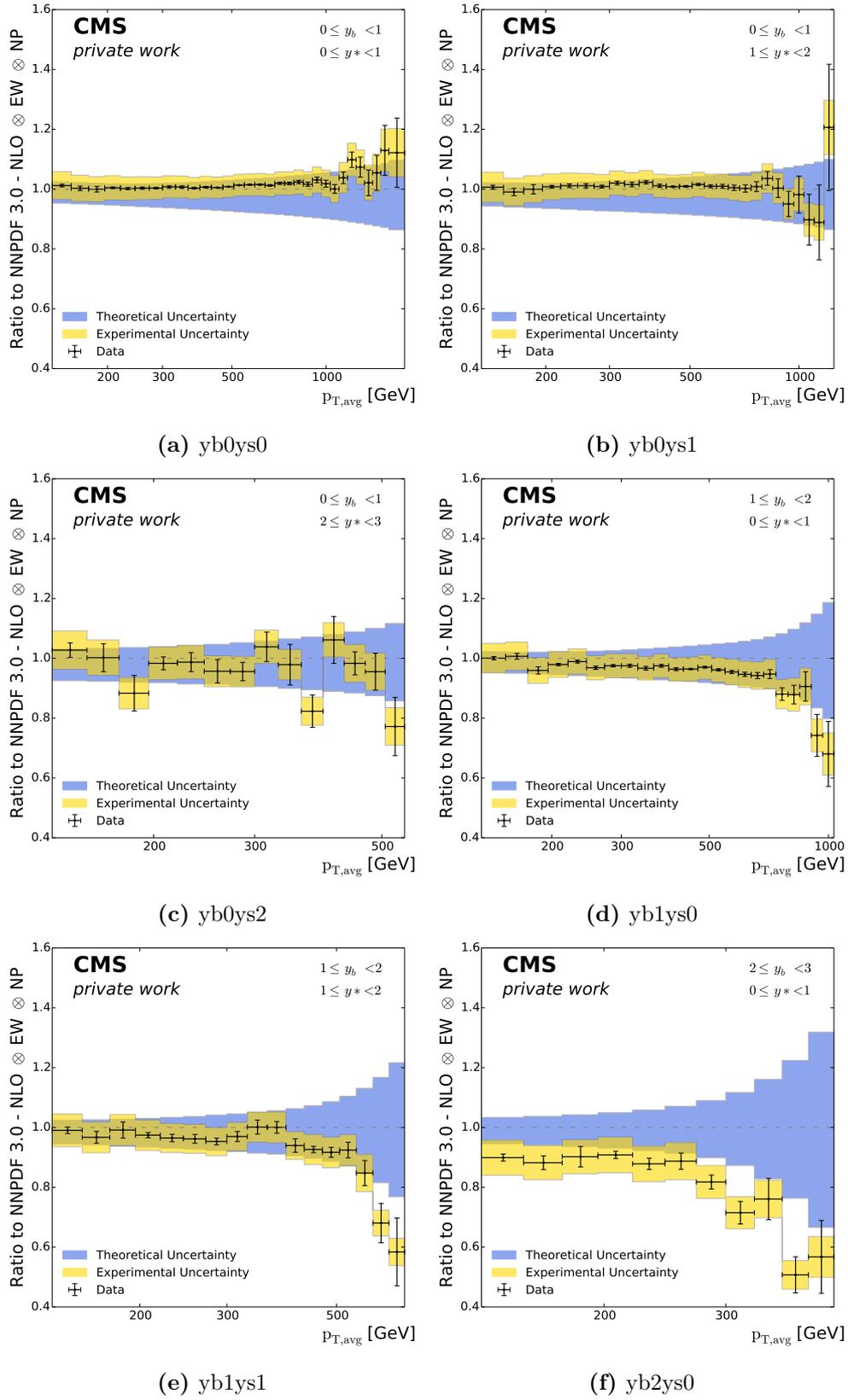


Figure A.13: Comparison of cross section results obtained from data to results obtained from NLO theory calculations. Displayed are the six y_b - y^* -bins.

Appendix B

List of Figures

2.1	Particles in the Standard Model	6
2.2	Running of the strong coupling	8
2.3	Example of PDF behaviour for different partons.	10
2.4	Dijet Feynman diagrams in leading order QCD	11
2.5	Feynman diagram of the Drell-Yan process	12
2.6	Dominant Feynman diagrams of Z production in LO.	13
3.1	Illustration of the coordinate system for the CMS detector.	22
3.2	Subsystems of the CMS detector, view of a slice.	23
3.3	Consecutive steps in Jet Energy Correction.	27
4.1	Illustration of the y_b - y^* -binning and corresponding event topologies.	31
4.2	Statistical uncertainties in JES uncertainty determination	35
4.3	Trigger turn-on curves obtained with the emulation method	39
4.4	Trigger turn-on curves zoomed in	40
4.5	Comparison of MadGraph+Pythia8 versus Pythia8-only sample for dijet.	41
4.6	Dijet cross section overview in data, theory and MC simulation.	42
4.7	Dijet Analysis: Comparison of simulation and theory to data results.	43
4.8	Z+jet cross section overview in data, theory and MC simulation.	50
4.9	Z+jet Analysis: Comparison of simulation and theory to data results.	51
5.1	Cross section ratio results in data, MC simulation and theory.	59
5.2	Double ratios: Comparison of simulation and theory to data	61
A.1	Dijet: Compatibility check of the four run periods in 2018.	69
A.2	Z+jet: Compatibility check of the four run periods in 2018.	70
A.3	Inclusive dijet cross section: Herwig7 sample	71
A.4	Inclusive dijet cross section: MadGraph+Pythia8 sample	72
A.5	Inclusive Z+jet cross section: MadGraph+Herwig7 sample	73
A.6	Double ratio using MadGraph+Pythia8	74
A.7	Double ratio using (MadGraph+)Herwig7	75
A.8	Response matrices for dijet	76
A.9	Dijet analysis 2017: Cross section results	77
A.10	Dijet analysis 2017: k-factors	78

A.11	Dijet analysis 2017: EW and NP correction factors	78
A.12	Dijet analysis 2017: Relative uncertainties	79
A.13	Dijet analysis 2017: Comparison of data to theory	80

Appendix C

List of Tables

4.1	Definition of the y_b - y^* -binning.	30
4.2	Binning in average transverse momentum $p_{T,avg}$	32
4.3	Data sets of the 2018 data taking period.	33
4.4	Single Jet Trigger Turn-On Points	38
A.1	Data sets used in this analysis.	68
A.2	Monte Carlo simulation samples used in this analysis.	68

Appendix D

Bibliography

- [1] Tung-Mow Yan and Sidney D. Drell. “The Parton Model and its Applications”. *Int. J. Mod. Phys. A*29 (2014), p. 0071.
DOI: [10.1142/S0217751X14300713](https://doi.org/10.1142/S0217751X14300713). arXiv: [1409.0051](https://arxiv.org/abs/1409.0051) [[hep-ph](#)].
- [2] Standard Model of Elementary Particles. URL: https://commons.wikimedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg (visited on 04/23/2020).
- [3] Particle Data Group Collaboration. “Review of Particle Physics”. *Phys. Rev. D*98.3 (2018).
DOI: [10.1103/PhysRevD.98.030001](https://doi.org/10.1103/PhysRevD.98.030001).
- [4] David J. Gross and Frank Wilczek. “Ultraviolet Behavior of Non-Abelian Gauge Theories”. *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1343–1346.
DOI: [10.1103/PhysRevLett.30.1343](https://doi.org/10.1103/PhysRevLett.30.1343).
- [5] Dominik Haitz. “Precision Studies of Proton Structure and Jet Energy Scale with the CMS Detector at the LHC”. PhD thesis. Karlsruher Institut für Technologie, 2016.
DOI: [10.5445/IR/1000055680](https://doi.org/10.5445/IR/1000055680).
- [6] John C. Collins, Davison E. Soper, and George Sterman. *Factorization of Hard Processes in QCD*. 2004. arXiv: [hep-ph/0409313](https://arxiv.org/abs/hep-ph/0409313) [[hep-ph](#)].
- [7] Gavin P. Salam. *Elements of QCD for hadron colliders*. 2010. arXiv: [1011.5131](https://arxiv.org/abs/1011.5131) [[hep-ph](#)].
- [8] Joseph Karpie et al. “Parton distribution functions on the lattice and in the continuum”. *EPJ Web of Conferences* 175 (2018). Ed. by M. Della Morte et al., p. 06032. ISSN: 2100-014X.
DOI: [10.1051/epjconf/201817506032](https://doi.org/10.1051/epjconf/201817506032). arXiv: [1710.08288](https://arxiv.org/abs/1710.08288) [[hep-lat](#)].
- [9] Sayipjamal Dulat et al. “New parton distribution functions from a global analysis of quantum chromodynamics”. *Physical Review D* 93.3 (Feb. 2016). ISSN: 2470-0029.
DOI: [10.1103/physrevd.93.033006](https://doi.org/10.1103/physrevd.93.033006). arXiv: [1506.07443](https://arxiv.org/abs/1506.07443) [[hep-ph](#)].
- [10] The Durham HepData Project. PDF Plotter. URL: <http://hepdata.cedar.ac.uk/pdf/pdf3.html> (visited on 03/26/2020).

- [11] J. Pumplin et al. “Uncertainties of predictions from parton distribution functions. II. The Hessian method”. *Physical Review D* 65.1 (Dec. 2001). ISSN: 1089-4918. DOI: [10.1103/physrevd.65.014013](https://doi.org/10.1103/physrevd.65.014013).
- [12] Walter T. Giele and Stephane Keller. “Implications of hadron collider observables on parton distribution function uncertainties”. *Physical Review D* 58.9 (Oct. 1998). ISSN: 1089-4918. DOI: [10.1103/physrevd.58.094023](https://doi.org/10.1103/physrevd.58.094023).
- [13] Walter T. Giele, Stephane A. Keller, and David A. Kosower. *Parton Distribution Function Uncertainties*. 2001. arXiv: [hep-ph/0104052](https://arxiv.org/abs/hep-ph/0104052) [[hep-ph](#)].
- [14] Christian Schwinn. *Lecture Notes: Modern Methods of Quantum Chromodynamics*. Mar. 2015.
- [15] UA1 Collaboration. “Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider”. *Phys. Lett. B* 126 (1983), pp. 398–410. DOI: [10.1016/0370-2693\(83\)90188-0](https://doi.org/10.1016/0370-2693(83)90188-0).
- [16] “LEP design report”. Vol. 1. By the LEP Injector Study Group. Geneva: CERN, 1983.
- [17] “LEP design report”. Vol. 2. Copies shelved as reports in LEP, PS and SPS libraries. Geneva: CERN, 1984.
- [18] Carlo Wyss. “LEP design report, v.3: LEP2”. Vol. 3. Vol. 1-2 publ. in 1983-84. Geneva: CERN, 1996.
- [19] Bettina Schillinger. “Untersuchung des dreifach differentiellen Wirkungsquerschnitts der Z+Jet Produktion auf PDF- Sensitivität”. Bachelor’s Thesis. Karlsruhe Institute of Technology (KIT), 2016.
- [20] Miguel Santos Correa. “Automated Production of Interpolation Grids at NNLO for the Triple-Differential Z+Jet Cross Section Measurement at the LHC”. Master’s Thesis. Karlsruhe Institute of Technology (KIT), 2018.
- [21] CMS Collaboration. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. ISSN: 1748-0221. DOI: [10.1088/1748-0221/12/02/p02014](https://doi.org/10.1088/1748-0221/12/02/p02014). arXiv: [1607.03663](https://arxiv.org/abs/1607.03663) [[hep-ex](#)].
- [22] The CMS collaboration. “Determination of jet energy calibration and transverse momentum resolution in CMS”. *Journal of Instrumentation* 6.11 (Nov. 2011), P11002–P11002. ISSN: 1748-0221. DOI: [10.1088/1748-0221/6/11/p11002](https://doi.org/10.1088/1748-0221/6/11/p11002). arXiv: [1107.4277](https://arxiv.org/abs/1107.4277) [[physics.ins-det](#)].
- [23] CMS Collaboration. *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV*. 2018. arXiv: [1804.04528](https://arxiv.org/abs/1804.04528) [[physics.ins-det](#)].

-
- [24] A. M. Sirunyan et al. “Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV and constraints on parton distribution functions”. *The European Physical Journal C* 77.11 (Nov. 2017). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-5286-7](https://doi.org/10.1140/epjc/s10052-017-5286-7). arXiv: [1705.02628](https://arxiv.org/abs/1705.02628) [hep-ex].
- [25] T. Kluge, K. Rabbertz, and M. Wobisch. “fastNLO: Fast pQCD Calculations for PDF Fits”. *Deep Inelastic Scattering DIS 2006* (Jan. 2007). DOI: [10.1142/9789812706706_0110](https://doi.org/10.1142/9789812706706_0110). arXiv: [hep-ph/0609285](https://arxiv.org/abs/hep-ph/0609285) [hep-ph].
- [26] M. Wobisch et al. *Theory-Data Comparisons for Jet Measurements in Hadron-Induced Processes*. 2011. arXiv: [1109.1310](https://arxiv.org/abs/1109.1310) [hep-ph].
- [27] Daniel Britzger et al. *New features in version 2 of the fastNLO project*. 2012. arXiv: [1208.3641](https://arxiv.org/abs/1208.3641) [hep-ph].
- [28] Zoltán Nagy and Zoltán Trócsányi. “Multijet Cross Sections in Deep Inelastic Scattering at Next-to-Leading Order”. *Phys. Rev. Lett.* 87 (8 Aug. 2001), p. 082001. DOI: [10.1103/PhysRevLett.87.082001](https://doi.org/10.1103/PhysRevLett.87.082001).
- [29] James Currie et al. *Jet cross sections at the LHC with NNLOJET*. 2018. arXiv: [1807.06057](https://arxiv.org/abs/1807.06057) [hep-ph].
- [30] Klaus Rabbertz. “Jet Physics at the LHC”. *Springer Tracts Mod. Phys.* 268 (2017). DOI: [10.1007/978-3-319-42115-5](https://doi.org/10.1007/978-3-319-42115-5).
- [31] S. Agostinelli et al. “Geant4—a simulation toolkit”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [32] B. Webber. “Parton shower Monte Carlo event generators”. *Scholarpedia* 6.12 (2011). revision #128236, p. 10662. DOI: [10.4249/scholarpedia.10662](https://doi.org/10.4249/scholarpedia.10662).
- [33] Torbjörn Sjöstrand. “Jet fragmentation of multiparton configurations in a string framework”. *Nuclear Physics B* 248.2 (1984), pp. 469–502. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(84\)90607-2](https://doi.org/10.1016/0550-3213(84)90607-2).
- [34] B.R. Webber. “A QCD model for jet fragmentation including soft gluon interference”. *Nuclear Physics B* 238.3 (1984), pp. 492–528. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(84\)90333-X](https://doi.org/10.1016/0550-3213(84)90333-X).
- [35] G. Marchesini and B.R. Webber. “Monte Carlo simulation of general hard processes with coherent QCD radiation”. *Nuclear Physics B* 310.3 (1988), pp. 461–526. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(88\)90089-2](https://doi.org/10.1016/0550-3213(88)90089-2).
- [36] Torbjörn Sjöstrand. “Pythia 8”. Pythia 8 online manual. 2019. URL: <http://home.thep.lu.se/~torbjorn/pythia82html/Welcome.html> (visited on 02/01/2020).
- [37] The Herwig Collaboration. “The Herwig Event Generator”. Herwig 7 Documentation web page. 2019. URL: <https://herwig.hepforge.org/> (visited on 02/01/2020).

- [38] Fabio Maltoni and Tim Stelzer. “MadEvent: automatic event generation with MadGraph”. *Journal of High Energy Physics* 2003.02 (Feb. 2003), pp. 027–027. ISSN: 1029-8479.
DOI: [10.1088/1126-6708/2003/02/027](https://doi.org/10.1088/1126-6708/2003/02/027).
- [39] Johan Alwall et al. “MadGraph 5: going beyond”. *Journal of High Energy Physics* 2011.6 (June 2011). ISSN: 1029-8479.
DOI: [10.1007/jhep06\(2011\)128](https://doi.org/10.1007/jhep06(2011)128).
- [40] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The anti-kt jet clustering algorithm”. *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479.
DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [41] Gavin P. Salam. “Towards jetography”. *The European Physical Journal C* 67.3-4 (May 2010), pp. 637–686. ISSN: 1434-6052.
DOI: [10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6).
- [42] Stephen D. Ellis and Davison E. Soper. “Successive combination jet algorithm for hadron collisions”. *Physical Review D* 48.7 (Oct. 1993), pp. 3160–3166. ISSN: 0556-2821.
DOI: [10.1103/physrevd.48.3160](https://doi.org/10.1103/physrevd.48.3160).
- [43] Yu.L Dokshitzer et al. “Better jet clustering algorithms”. *Journal of High Energy Physics* 1997.08 (Aug. 1997), pp. 001–001. ISSN: 1029-8479.
DOI: [10.1088/1126-6708/1997/08/001](https://doi.org/10.1088/1126-6708/1997/08/001).
- [44] M. Wobisch and T. Wengler. *Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering*. 1999. arXiv: [hep-ph/9907280](https://arxiv.org/abs/hep-ph/9907280) [[hep-ph](https://arxiv.org/abs/hep-ph)].
- [45] Oliver S. Bruning et al. “LHC Design Report Vol.1: The LHC Main Ring”. CERN Yellow Reports: Monographs (2004).
DOI: [10.5170/CERN-2004-003-V-1](https://doi.org/10.5170/CERN-2004-003-V-1).
- [46] O. Buning et al. “LHC Design Report. 2. The LHC infrastructure and general services”. CERN Yellow Reports: Monographs (2004).
DOI: [10.5170/CERN-2004-003-V-2](https://doi.org/10.5170/CERN-2004-003-V-2).
- [47] M. Benedikt et al. “LHC Design Report. 3. The LHC injector chain”. CERN Yellow Reports: Monographs (2004).
DOI: [10.5170/CERN-2004-003-V-3](https://doi.org/10.5170/CERN-2004-003-V-3).
- [48] C.J. Seez. “CMS, a general purpose detector for the LHC”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 344.1 (1994), pp. 1–10. ISSN: 0168-9002.
DOI: [https://doi.org/10.1016/0168-9002\(94\)90645-9](https://doi.org/10.1016/0168-9002(94)90645-9).
- [49] CMS Collaboration. “The CMS experiment at the CERN LHC”. *JINST* 3.S08004 (2008).
DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).

-
- [50] A.M. Sirunyan et al. “Particle-flow reconstruction and global event description with the CMS detector”. *Journal of Instrumentation* 12.10 (Oct. 2017), P10003–P10003. ISSN: 1748-0221.
DOI: [10.1088/1748-0221/12/10/p10003](https://doi.org/10.1088/1748-0221/12/10/p10003). arXiv: [1706.04965](https://arxiv.org/abs/1706.04965) [[physics.ins-det](#)].
- [51] CMS Collaboration. “CMS Detector”. CMS web page, detector page. URL: <https://cms.cern/detector> (visited on 02/02/2020).
- [52] *The CMS Particle Flow Algorithm*. 2013, pp. 295–304. arXiv: [1401.8155](https://arxiv.org/abs/1401.8155) [[hep-ex](#)].
- [53] Thomas Berger. “Jet energy calibration and triple differential inclusive cross section measurements with $Z (\rightarrow \mu\mu) + \text{jet}$ events at 13 TeV recorded by the CMS detector”. PhD thesis. Karlsruhe Institute of Technology (KIT), 2019.
DOI: [10.5445/IR/1000104286](https://doi.org/10.5445/IR/1000104286).
- [54] CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. *Journal of Instrumentation* 13.06 (June 2018), P06015–P06015. ISSN: 1748-0221.
DOI: [10.1088/1748-0221/13/06/p06015](https://doi.org/10.1088/1748-0221/13/06/p06015). arXiv: [1804.04528](https://arxiv.org/abs/1804.04528) [[physics.ins-det](#)].
- [55] Georg Sieber. “Measurement of Triple-Differential Dijet Cross Sections with the CMS Detector at 8 TeV and PDF Constraints”. PhD thesis. Karlsruher Institut für Technologie, 2016.
DOI: [10.5445/IR/1000055723](https://doi.org/10.5445/IR/1000055723).
- [56] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The catchment area of jets”. *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 005–005. ISSN: 1029-8479.
DOI: [10.1088/1126-6708/2008/04/005](https://doi.org/10.1088/1126-6708/2008/04/005).
- [57] Valentin Kuznetsov, Dave Evans, and Simon Metson. “The CMS data aggregation system”. *Procedia Computer Science* 1.1 (2010). ICCS 2010, pp. 1535–1543. ISSN: 1877-0509.
DOI: <https://doi.org/10.1016/j.procs.2010.04.172>.
- [58] Rene Brun and Fons Rademakers. “ROOT — An object oriented data analysis framework”. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86. ISSN: 0168-9002.
DOI: [https://doi.org/10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X).
- [59] Rizzi, Andrea, Petrucciani, Giovanni, and Peruzzi, Marco. “A further reduction in CMS event data for analysis: the NANO AOD format”. *EPJ Web Conf.* 214 (2019), p. 06021.
DOI: [10.1051/epjconf/201921406021](https://doi.org/10.1051/epjconf/201921406021).
- [60] CMS Collaboration and Karl Ehataht. *NANO AOD: a new compact event data format in CMS*. Nov. 2019.
DOI: [10.5281/zenodo.3598830](https://doi.org/10.5281/zenodo.3598830).
- [61] CMS nanoAOD development. “nanoAOD-tools: Tools for working with NanoAOD”. URL: <https://github.com/cms-nanoAOD/nanoAOD-tools> (visited on 03/13/2020).

- [62] CMS Collaboration. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAOD> (visited on 03/13/2020).
- [63] CMS Collaboration Collaboration. “CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project”. Technical Design Report CMS. Geneva: CERN, 2002.
- [64] CMS Collaboration. “The CMS trigger system”. *Journal of Instrumentation* 12.01 (Jan. 2017), P01020–P01020. ISSN: 1748-0221. DOI: [10.1088/1748-0221/12/01/p01020](https://doi.org/10.1088/1748-0221/12/01/p01020). arXiv: [1609.02366](https://arxiv.org/abs/1609.02366) [physics.ins-det].
- [65] CMS Collaboration. “BRIL Work Suite: Documentation”. URL: <https://cms-service-lumi.web.cern.ch/cms-service-lumi/brilwsdoc-35x.html> (visited on 03/18/2020).
- [66] CMS Collaboration. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/PdmV2018Analysis> (visited on 03/18/2020).
- [67] CMS Collaboration 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.
- [68] CMS Collaboration. “TWiki: Muon Isolation”. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2> (visited on 04/17/2020).
- [69] UA1 Collaboration. “Comparison of three-jet and two-jet cross sections in $p\bar{p}$ collisions at the CERN SPS $p\bar{p}$ collider”. *Physics Letters B* 158.6 (1985), pp. 494–504. ISSN: 0370-2693. DOI: [https://doi.org/10.1016/0370-2693\(85\)90801-9](https://doi.org/10.1016/0370-2693(85)90801-9).
- [70] S. Chatrchyan et al. “Measurement of the ratio of the inclusive 3-jet cross section to the inclusive 2-jet cross section in pp collisions at $\sqrt{s} = 7$ TeV and first determination of the strong coupling constant in the TeV range”. *The European Physical Journal C* 73.10 (Oct. 2013). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-013-2604-6](https://doi.org/10.1140/epjc/s10052-013-2604-6). arXiv: [1304.7498](https://arxiv.org/abs/1304.7498) [hep-ex].
- [71] ATLAS Collaboration. “Measurements of top-quark pair to Z-boson cross-section ratios at $\sqrt{s} = 13, 8, 7$ TeV with the ATLAS detector”. *Journal of High Energy Physics* 2017.2 (Feb. 2017). ISSN: 1029-8479. DOI: [10.1007/jhep02\(2017\)117](https://doi.org/10.1007/jhep02(2017)117). arXiv: [1612.03636](https://arxiv.org/abs/1612.03636) [hep-ex].

Acknowledgements

In this chapter, I would like to thank the persons who supported this thesis project in one or another way.

First of all, I sincerely thank my supervisor Priv.-Doz. Dr. Klaus Rabbertz, who originally introduced me to the field of jet physics. Thank you for your guidance and answering my questions over the past years.

To Asst. Prof. Dr. Mikko Voutilainen I am very grateful for granting me the opportunity to work with him and his group in Helsinki. I am thankful for our numerous discussions during this thesis project and your patience whenever I needed to ask again and again. A special thank you goes to Prof. Dr. Günter Quast for welcoming me in his group at the ETP about four years ago. Thank you for your advice and encouragement.

I want to express my gratitude to the three of you for offering me the opportunity to work on an interesting topic and for your open-mindedness towards a joint thesis project. For support in the early stage of this project I would like to thank Dr. Thomas Berger, Dr. Henning Kirschenmann and Daniel Savoiu.

For feedback and comments while writing down this thesis, I am thankful to Dr. René Caspart, Christoph Heidecker, Daniel Savoiu and Matthias Schnepf.

In particular to René I am indebted for his support. Thank you for keeping up our "Remote-Office" routine together and helping me to keep - more or less - calm during the final weeks.

I am happy about my nice colleagues at the Institute of Experimental Particle Physics (ETP) in Karlsruhe as well as at the Helsinki Institute of Physics (HIP) in Helsinki and want to thank both research groups for the good working atmosphere.

A big thank you also to all of my dear friends and fellow students.

Furthermore, I would like to thank my family, especially my parents Erika and Markus Schillinger for their support throughout my studies. Thank you for believing in me also during difficult times.

Finally, I want to thank Pekka Lehtelä for his love and support during quite eventful years. Thank you for always being there for me.

Erklärung der selbstständigen Anfertigung der Masterarbeit

Hiermit versichere ich, die vorliegende Masterarbeit mit dem Titel

*»Measurement of the Ratio of inclusive Jet Cross Sections in Events with and without a
high- p_T Z Boson«*

selbstständig und unter ausschließlicher Verwendung der angegebenen Quellen
angerfertigt zu haben.

Bettina Schillinger
Espoo, den 26. April 2020