

**Differential Cross Section Measurement
of Drell-Yan Production
and associated Jets with the CMS
Experiment at the LHC**

Dissertation

zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von

SAMANTHA KATHERINE DOOLING

aus Emsdetten

Hamburg
2014

Gutachter/in der Dissertation:	Dr. Hannes Jung Prof. Dr. Peter Schleper Prof. Dr. Günther Dissertori
Gutachter/in der Disputation:	Dr. Hannes Jung Prof. Dr. Peter Schleper Prof. Dr. Dieter Horns Dr. Alexei Raspereza
Datum der Disputation:	17. November 2014
Vorsitzender des Prüfungsausschusses:	Prof. Dr. Bernd Kniehl
Vorsitzende des Promotionsausschusses:	Prof. Dr. Daniela Pfannkuche
Dekan des Fachbereichs Physik:	Prof. Dr. Heinrich Graener

Abstract

A measurement of the Drell-Yan differential cross section in proton-proton collisions at a centre-of-mass energy of 7 TeV is presented. The results are obtained using $Z^0/\gamma^* \rightarrow \mu^+\mu^-$ decays in a data sample collected with the CMS detector corresponding to an integrated luminosity of 4.9 fb^{-1} . The cross section of Drell-Yan production in association with one or two jets with $p_T^{\text{jet}} > 30 \text{ GeV}$ in the range of $|\eta^{\text{jet}}| < 4.5$ is measured. The differential cross section is studied as a function of the dimuon transverse momentum $p_T^{\mu\mu}$ and of the invariant mass, covering a range of $30 < m^{\mu\mu} < 1500 \text{ GeV}$. In addition, the Drell-Yan dimuon pair is selected in the forward region with $|\eta^{\mu\mu}| > 2.5$ and the differential cross section as well as the average jet multiplicity, as a function of the rapidity separation between the leading jet and the Drell-Yan lepton pair, is presented. All measurements are compared to predictions of fixed-order perturbative QCD plus parton showers, provided by Monte Carlo simulation. An agreement of the distributions with predictions of resummed QCD combined with high fixed-order perturbative QCD is observed. Lowest-order calculations of the cross section are not able to describe the Drell-Yan + jets cross section at low $p_T^{\mu\mu}$ and large rapidity differences. The Drell-Yan plus jet topologies are sensitive to hard-parton emissions calculated at fixed-order in perturbation theory plus resummation of multi-gluon emissions.

Zusammenfassung

Im Rahmen dieser Dissertation werden Messungen des Drell-Yan Wirkungsquerschnittes in Proton-Proton Kollisionen mit einer Schwerpunktenergie von 7 TeV vorgestellt. Die Ergebnisse resultieren aus einem Datensatz von $Z^0/\gamma^* \rightarrow \mu^+\mu^-$ Ereignissen, die einer Luminosität von 4.9 fb^{-1} entsprechen. Drell-Yan Produktion in Kombination mit einem oder zwei Jets mit $p_T^{\text{jet}} > 30 \text{ GeV}$ und $|\eta^{\text{jet}}| < 4.5$, wird mit dem CMS Detektor gemessen. Der differentielle Wirkungsquerschnitt als Funktion des Transversalimpulses $p_T^{\mu\mu}$ und der invarianten Masse im Bereich von $30 < m^{\mu\mu} < 1500 \text{ GeV}$, wird untersucht. Darüber hinaus werden ausschließlich die im vorderen Bereich, mit $|\eta^{\mu\mu}| > 2.5$, produzierten Drell-Yan Paare selektiert und untersucht. Der differentielle Wirkungsquerschnitt wird als Funktion des Rapidityunterschiedes zwischen dem Drell-Yan Paar und dem harten Jet gemessen. Zusätzlich wird die Jet Multiplizität als Funktion der Rapiditydifferenz vorgestellt. Alle Messungen werden mit perturbativen Berechnungen und Parton Shower Modellen verglichen. Die Berechnungen zu hohen Ordnungen in Kombination mit Parton Shower Algorithmen beschreiben das gemessene $p_T^{\mu\mu}$ Spektrum. Berechnungen in führender Ordnung in perturbativer QCD stimmen bei kleinem $p_T^{\mu\mu}$ und großen Rapidityunterschieden nicht mit den Daten überein. Drell-Yan plus Jets Messungen sind sensiv auf harte Partonemissionen in Kombination mit Resummation von weichen Partonemissionen.

Contents

Introduction	1
1 Theory	5
1.1 Standard Model of Particle Physics	5
1.2 Proton-Proton Collisions	10
1.2.1 Structure of Protons and Factorisation	11
1.2.2 Underlying Event and Pileup Contributions	14
1.3 Theory Comparison	15
1.3.1 Monte Carlo Event Generators	16
1.3.2 Nonperturbative Effects in Shower Monte Carlo Generators .	18
1.3.3 Nonperturbative and Parton Shower Effects in NLO-matched Monte Carlo Generators	19
1.4 Drell-Yan Production	24
1.4.1 Cross Section	24
1.4.2 Perturbative QCD Corrections	25
1.4.3 Transverse Momentum Distribution	27
1.4.4 Resummation	31
1.4.5 Jets in Drell-Yan Events	33
2 The Compact Muon Solenoid at the LHC	35
2.1 The LHC Accelerator	35
2.1.1 Luminosity	36
2.1.2 Data Taking Period 2011	37
2.2 The CMS Detector	39
2.2.1 Magnet	41
2.2.2 Inner Tracking System	41
2.2.3 Electromagnetic Calorimeter	43
2.2.4 Hadronic Calorimeter	44
2.2.5 Muon System	44
2.2.6 Trigger	45
2.3 Detector Simulation with GEANT4	50
3 Event Reconstruction and Object Identification	51
3.1 Primary Vertex Reconstruction	51
3.1.1 Pileup Reweighting	52
3.2 Particle-Flow Reconstruction	53
3.3 Muons in CMS	54
3.3.1 Muon Reconstruction	55
3.3.2 Muon Identification	55
3.3.3 Particle-Flow Isolation	56

3.3.4	Muon Efficiency	58
3.4	Jets in CMS	58
3.4.1	Jet Reconstruction	59
3.4.2	Jet Identification	61
3.4.3	Jet Energy Scale and Jet Energy Resolution	61
4	Drell-Yan + Jets Analysis	65
4.1	Data and Simulated Samples	68
4.2	Event Selection	69
4.3	Monte Carlo Corrections	71
4.3.1	QED Final-State Radiation	71
4.3.2	Muon Momentum Scale Correction	72
4.4	Background Estimation	73
4.4.1	QCD Background Estimation from Data Events	74
4.4.2	Top Quark Pair Production	79
4.4.3	Diboson and $W +$ Jets Production	80
4.4.4	$Z \rightarrow \tau\tau$ Background Process	80
4.4.5	Summary of Background Estimation	81
4.5	Particle Level Correction	81
4.5.1	Migration Study	83
4.5.2	Unfolding	94
4.6	Cross Section Definition	101
4.7	Systematic Uncertainties	102
4.7.1	Experimental Uncertainties	102
4.7.2	Summary of Systematic Uncertainties	107
4.7.3	Theory Uncertainties	108
5	Drell-Yan + Jets Results	111
5.1	Drell-Yan Transverse Momentum Distribution	111
5.2	Rapidity Separation between Drell-Yan and the leading Jet	123
6	Summary	129
A	Trigger Efficiency	135
B	Primary Vertex Reweighting	139
C	Study of Top Quark Background Estimation	141
D	Purity, Stability, Acceptance and Background	145
E	Underlying Event Tunes and Double Parton Scattering	157
F	Longitudinal momentum shifts, showering, and nonperturbative corrections in matched next-to-leading-order shower event generators	195
G	Higgs as a gluon trigger	207
	List of figures	220

List of tables	222
Bibliography	228

Introduction

The 04 July 2012 was a monumental day for the world of high energy physics. A new particle was discovered, which is compatible with the theory expectations, to be the missing piece of the puzzle to confirm the theory, which explains how certain fundamental particles gain masses.

The universe and all matter that surrounds us, is constructed of a specific number of elementary building blocks governed by fundamental forces. Particle physicists devote their work to exploring the structure of matter constructed from the elementary particles and their interactions. The Standard Model (SM) is a theory that combines the fundamental particles along with three forces and successfully explains most of the experimental results and predictions so far. However, an essential constituent, theoretically predicted in 1964 by Peter Higgs, François Englert, and Robert Brout [1, 2], was still experimentally unobserved. After years of theoretical studies and experimental research, the particle physicists working in collaboration with the European Organisation for Nuclear Research, commonly known as CERN, succeeded to discover a new particle, proving the tangible existence of the Higgs particle. The discovery of the Higgs boson with a mass of 125 GeV was a big triumph for particle physics.

At the Large Hadron Collider (LHC), the world's largest and highest-energetic particle accelerator, located at CERN, protons are accelerated and made to collide, establishing physics processes at an energy scale that previously was not possible. The measurement of high energy scattering processes provide crucial information about theoretical and phenomenological concepts of particle physics. At this increased energy level, it was possible to observe the very heavy boson, compatible with the Higgs particle, by the ATLAS and CMS Collaborations.

In order to validate a physics theory, accurate theoretical calculations have to be compared to precise experimental measurements. In this way, the particle physicists are able to discover evidence for new particles or physics theories beyond the SM.

In the calculation of scattering amplitudes, simplifications have to be considered in order to be able to calculate the cross section in perturbation theory. Perturbation theory is a theoretical approach to calculate observable quantities in powers of the coupling constant of the theory. In Quantum Chromodynamics (QCD), the theory of the strong force, which describes the interactions between quarks and gluons, the perturbative terms are given in orders of the strong coupling (α_S). Ideally, the first terms of the power series yield a sufficient description of the observable. However, in some regions of phase space, some terms can be divergent and spoil the perturbative

expansion. In general, the perturbative series can be truncated and the calculations correspond to a fixed-order computation. For some observables however, the leading terms already yield a very large contribution to the perturbative expansion. Thus, a truncated power series would lead to a divergence and to a non-physical result. In this case the perturbative terms have to be resummed to all order in α_S to describe the correct distribution.

The proton-proton cross section is calculated using perturbative QCD based on the parton model. This is a simplified model, which takes into account the substructure of the protons, made of partons (quarks and gluons). The parton level cross section is calculated using perturbation theory, and the cross section of proton-proton collisions can then be written in terms of a convolution of the probability function of the incoming partons and the partonic scattering amplitude. The partons initiating the hard scattering carry a momentum fraction x of the incoming protons. Asymmetric QCD collisions, when one parton has a large value in x and another a small value in x , refer to small- x physics. These processes lead to activity in the forward region of the detector, defined by large values in rapidity y , with $y = \frac{1}{2} \ln \left(\frac{E+p_z}{E-p_z} \right)$.

An important role in the perturbative calculation of the partonic scattering process, is additional parton emissions, especially soft-gluon emissions. The parton emissions correspond to higher-order corrections to the cross section and need to be modelled in theory by evolution equations. The evolution equations rearrange the sum of the perturbative series, depending on the scale of the process Q^2 and the momentum fraction x . In some corners of phase space, large divergences appear in the calculation and the perturbative expansion fails. At large values of Q^2 , the resummation leads to a transverse momentum ordered evolution of the multi-gluon emissions, provided by logarithms of the type $\ln(Q^2)$. However, at small- x , logarithms of the type $\ln(1/x)$ dominate and other evolution equations have to be employed. An approach to simulate the higher-order emissions in the perturbative calculation is the parton shower model. The parton shower exhibits an approximation with simplified dynamics in order to obtain a complete description of QCD events.

In this thesis, the Drell-Yan process is investigated. This process is defined as a quark anti-quark annihilation creating a virtual photon or Z boson, which subsequently decays into two leptons. The theoretical approach was first developed by Sidney Drell and Tung-Mow Yan in 1970 [3]. The Drell-Yan production is a crucial process to test perturbative as well as nonperturbative QCD. Higher-order QCD corrections play a significant role in the calculation of the Drell-Yan + jets cross section. QCD emissions treated by higher-order corrections in the perturbative calculation, correspond to jets in the final state. The Drell-Yan dimuon pairs in association with jets at a centre-of-mass energy of $\sqrt{s} = 7$ TeV are measured with the CMS detector, and the results of the differential cross section are presented in this work.

The Drell-Yan transverse momentum distribution is a crucial observable to study perturbative QCD. In different regions of p_T^{DY} , different calculations contribute. At

high p_T^{DY} , for $p_T^{\text{DY}} > m^{\text{DY}}$, the transverse momentum distribution can be described by fixed-order calculations. The cross section can be computed by a truncated perturbative series in α_S . However, at small p_T^{DY} the emissions of multiple soft gluons become important and the rise of the cross section can not be described by a fixed-order computation. If $p_T^{\text{DY}} \ll m^{\text{DY}}$, large logarithms of the order of $\ln^2 \frac{m^2}{p_T^2}$ will spoil the perturbative expansion. In order to describe the cross section at small p_T^{DY} , the leading logarithms have to be resummed to all orders.

The Drell-Yan cross section is measured differentially in $p_T^{\mu\mu}$ as well as in the Drell-Yan lepton pair invariant mass, to change the scale of the process. In this thesis the effect of resummation is studied in detail by comparing the measured differential cross section to various fixed-order calculations plus parton shower simulation, produced by Monte Carlo event generators. In order to probe the impact of higher-order corrections in the resummation of soft-gluon emission (*soft-gluon resummation*), Drell-Yan production is required in association with additional jets. Drell-Yan plus jet production increases the region at low p_T^{DY} , where resummation is dominant, and enables to probe soft-gluon emissions in a larger region of phase space.

Drell-Yan production with associated jets at large rapidity provides insight into a region of phase space, where fixed-order calculations are expected to fail and resummation of multi-gluon emissions is important. In this thesis, I investigate forward Drell-Yan production and jets covering a wide range in rapidity. The differential cross section as a function of rapidity separation $|\Delta y|$ of the Drell-Yan system and the leading jet is measured. The rapidity difference is sensitive to multi-gluon emissions and thus, an appropriate observable to study multi-jet events. At large rapidities the perturbative expansion yields large logarithms in terms of $\ln(1/x)$. The fixed-order calculation is not able to describe the cross section and resummation of higher-order contributions is important. In order to resum the large logarithms at small- x the evolution of the multi-jet emissions (*multi-jet resummation*) has to be investigated in detail.

The thesis is structured as follows: In the first chapter the theory concepts, which are relevant for the understanding of the Drell-Yan cross section measurement are introduced. Secondly, the CMS detector and its sub-components are described. The third chapter is devoted to present the event reconstruction and the definition of muons and jets within the CMS Collaboration. The analysis is presented in detail in chapter 4 and the different analysis steps are introduced. Finally, the measured cross section of Drell-Yan with associated jets as a function of the transverse momentum and invariant mass is presented in section 5.1. The cross section results as a function of the rapidity separation of the Drell-Yan and the leading jet is shown in section 5.2. The results are summarised and concluded in chapter 6.

Relevant publications, which arise from this thesis, and in which my work had a significant contribution, are collected in the appendix.

1 Theory

High energy scattering processes provide crucial information about theoretical and phenomenological concepts in particle physics. The Drell-Yan process is an essential part in phenomenological studies in the Standard Model. This chapter is intended to review the theory concepts, which are relevant for the understanding of the Drell-Yan cross section measurement. The information and formulas are taken from [4].

1.1 Standard Model of Particle Physics

Elementary Particles and Interactions

After years of theoretical studies and experimental research scientists have come to the conclusion that the universe is constructed of a specific number of elementary building blocks, which are governed by four fundamental forces. These perceptions provide a deep insight into the structure of the micro cosmos and help to achieve a better understanding of the construction and coherence of matter.

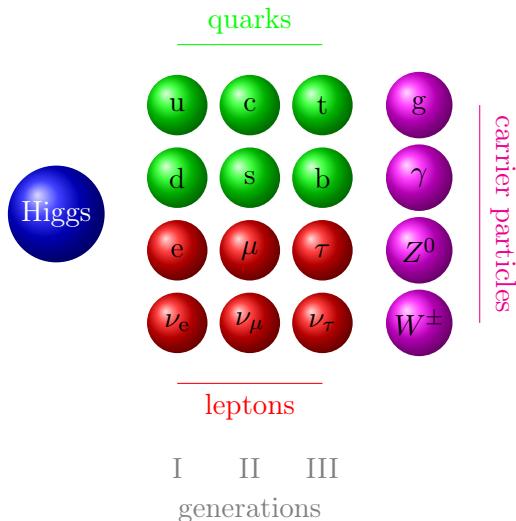


Figure 1.1: Particles of the Standard Model.

The Standard Model (SM) combines the fundamental particles, along with three forces, into a theory, which successfully explains most of the experimental results and predictions so far.

The SM includes matter and interaction particles, which are shown in Figure 1.1.

The matter particles build two groups: six *quarks* and six *leptons*, which split up into three almost identical generations, except for their masses, which increase from generation I to generation III. There are six different kinds of quark *flavours*: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). The quarks form the particle family of the *hadrons*, which can be a combination of either three or two quarks. Three quarks (or anti-quarks) form *baryons*, e.g. the neutron and proton, whereas *mesons* consist of a quark anti-quark pair. The six leptons consist of: electron (e), muon (μ), tau (τ) and their corresponding neutrinos (ν_e, ν_μ, ν_τ). For each matter particle, there exists a corresponding anti-particle, which only differs in electric charge. Leptons, along with quarks, are assumed to be elementary and are the constituents of the matter that surrounds us. Quarks and leptons are grouped together and called *fermions*, which have spin 1/2. The elementary particles can be classified by the fundamental forces. There are four fundamental interactions, which bind matter: gravitational, electromagnetic, strong and weak. They differ in range and strength, shown in Table 1.1.

Force	Range (in m)	Strength	Mediators (mass in GeV)	Participants
strong	10^{-15}	$\alpha_S \approx 1$	8 gluons (0)	quarks and gluons
weak	10^{-18}	$\alpha_W \approx 10^{-5}$	W^\pm, Z (80.4, 91.2)	quarks and leptons
electrom.	∞	$\alpha \approx \frac{1}{137}$	photon (0)	el. charged particles
gravity	∞	$\approx 10^{-39}$	graviton (0)	all particles

Table 1.1: Four fundamental forces in physics.

Each of the forces has its own carrier particles. They describe the interaction of the matter particles with each other by an exchange of mediators, by which a discrete amount of energy is transferred. The carrier particles have spin 1 and are called *bosons*. The interaction particles are gluons for the strong force, W and Z bosons for the weak force, and photons for the electromagnetic force. The electromagnetic and weak force can be described as one single force by an unified theory, called the *electroweak theory*. The strong interaction couples to particles with colour charge. The mathematical description of the theory of the strong force is called *Quantum Chromodynamics*. The electroweak and strong force are explained in more detail later. For gravity a particle called *graviton* is assumed only theoretically. The first three interactions are comprised by the SM, whereas the gravitational force can not be included into the theory successfully. The general theory of relativity can explain the macro world but can not be combined with the quantum theory yet. However, in the low energy regime of particle physics the effects of gravity are so insignificant that they can be ignored.

A further essential constituent of the SM is the *Higgs particle*. Recently, on 04 July 2012 a new boson was discovered [5, 6], which is compatible with the Higgs boson. There is a global endeavour to prove the tangible existence of the Higgs particle,

which is one possibility to provide mass to the other particles. This feature of the SM can be explained by the electroweak symmetry breaking.

The Yang-Mills theory, a local gauge theory, furnishes a unified specification of the three forces based on symmetries. The gauge symmetry group of the SM is the product of each of the three forces

$$SU(3)_c \times SU(2)_L \times U(1)_Y. \quad (1.1)$$

Electroweak Theory

The electroweak interaction is a unified description of the electromagnetic and weak force of the SM of particle physics. Although the two interactions seem to be different at low energies, beyond the unification energy of ~ 100 GeV, they are merged into one force.

Mathematically, the electroweak theory is described by a non-abelian gauge symmetry $SU(2)_L \times U(1)_Y$. Y is the hyper-charge operator, the generator of the $U(1)_Y$ symmetry, defined as $Y = 2(Q - T^3)$, where Q is the conserved charge. The weak isospin operators $T^{1,2,3}$ are the generators of the $SU(2)_L$ symmetry. The gauge field associated with the weak hyper-charge is denoted as B_μ and couples via Y . The gauge fields of the $SU(2)_L$ are $W_\mu^{1,2,3}$ and only couple to left-handed fermions, which are ordered in $SU(2)$ doublets. Right-handed fermions are ordered in $SU(2)$ singlets.

The Lagrangian of the electroweak theory is given by

$$\mathcal{L} = \bar{\psi} i\gamma_\mu D^\mu \psi - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{2} \text{Tr}(W_{\mu\nu} W^{\mu\nu}), \quad (1.2)$$

where ψ represents the Dirac field of a spin 1/2 particle. The first term describes the kinematics of the fermion field ψ . The gauge covariant derivative

$$D_\mu = \partial_\mu - igW_\mu^a T^a + ig' B_\mu Y \quad (1.3)$$

ensures local invariance under $SU(2)_L \times U(1)_Y$ symmetry transformation. The gauge fields W_μ^a , where a runs over the three generators of the group, and B_μ are introduced when requiring Lorentz invariance and are massless vector fields. Any mass term of the gauge fields in the Lagrangian would violate the induced symmetry. The gauge coupling of the $SU(2)$ is denoted by g and the coupling constant of the $U(1)$ group by g' .

The second term in eq. (1.2) describes the kinematics of the vector field B_μ . The field tensor $B_{\mu\nu}$ is gauge invariant and defined as

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (1.4)$$

The field tensors $W_{\mu\nu}^a$ of the $SU(2)_L$ symmetry are defined similarly, but contain an additional term according to the self-interaction of the gauge fields in the non-abelian group $SU(2)$,

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + gf_{abc} W_\mu^b W_\nu^c, \quad (1.5)$$

where f_{abc} are the *structure constants* defined by the commutation relation

$$[T^a, T^b] = if^{abc}T^c. \quad (1.6)$$

The gauge fields represent the mediator particles W^\pm , Z , and γ of the weak and electromagnetic interaction respectively. Thus, the Lagrangian of the electroweak theory describes the dynamics of the fermions and the gauge bosons, as well as their interactions. However, the gauge bosons and fermions are predicted to be massless in order to preserve local gauge symmetry, but experimental evidence shows that the W^\pm and Z bosons are massive, see Table 1.1.

To resolve this issue, the mathematical concept of *spontaneous symmetry breaking* is employed in order to acquire the masses of the gauge bosons. This concept is employed by the so-called *Higgs-Mechanism*, which introduces a massive scalar boson field corresponding to the Higgs particle.

The Lagrangian of the external scalar field ϕ is given by

$$\mathcal{L}_\phi = |D_\mu\phi|^2 - V(\phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (1.7)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor, describing the kinematics of the free gauge field A_μ . The scalar potential of the Higgs field yields

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2, \quad (1.8)$$

with a mass parameter μ and a positive dimensionless self-coupling λ . The global minimum of the Higgs potential for $\mu^2 > 0$ is given by $\phi_0 = 0$. When selecting $\mu^2 < 0$, the symmetry can be broken resulting in a non-vanishing vacuum expectation value

$$|\phi_0|^2 = -\frac{\mu^2}{2\lambda} \equiv v. \quad (1.9)$$

The scalar field ϕ can be expanded around v . Inserting the new expression representing the Higgs field in eq. (1.7) results in an expression containing a mass term with

$$m^2 = 2g^2\phi_0^2. \quad (1.10)$$

The mass eigenstates, which correspond to the measurable gauge bosons of the theory, are given as a orthogonal superposition of B_μ and W_μ^3

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \quad (1.11)$$

$$Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(g'W_\mu^3 - gB_\mu), \quad (1.12)$$

and the orthogonal expression representing the massless photon field yields

$$A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(g'W_\mu^3 + gB_\mu). \quad (1.13)$$

The masses of the gauge bosons are then given by

$$m_W = g\frac{v}{2}, \quad m_Z = \sqrt{g^2 + g'^2}\frac{v}{2} \quad \text{and} \quad m_A = 0. \quad (1.14)$$

Quantum Chromodynamics

The theory of the strong interaction is called Quantum Chromodynamics (QCD) and is a non-abelian gauge theory under the gauge group $SU(3)$. The theory describes the interaction between the quarks governed by eight gluons. In nature, $N_f = 6$ quarks are known, which are distinguished by their quantum numbers and masses. The six kinds of quarks are called *flavours f*: up, down, strange, charm, bottom and top, $f = u, d, s, c, b, t$, and their characteristics are tabulated in Table 1.2. The particles affected by the strong coupling require an additional charge to preserve the Pauli principle in QCD. Therefore, quarks and gluons have colour charge, which occurs in red, green, and blue. The colour charge is the origin of the name of the theory.

Quark	Flavour	Charge Q (in e)	Mass
u up	$I_z = \frac{1}{2}$	2/3	$2.3^{+0.7}_{-0.5}$ MeV
d down	$I_z = -\frac{1}{2}$	-1/3	$4.8^{+0.5}_{-0.3}$ MeV
c charm	$C = 1$	2/3	1.275 ± 0.025 GeV
s strange	$S = -1$	-1/3	95 ± 5 MeV
t top	$T = 1$	2/3	173.07 ± 1.24 GeV
b bottom	$B = -1$	-1/3	4.18 ± 0.03 GeV

Table 1.2: Properties of the six quark flavours. The given quantum numbers are: electric charge Q , isospin I , charmness C , strangeness S , topness T , bottomness B . The values are taken from the Particle Data Group [7].

Due to the consideration of the colour charge and flavour of the quarks, the complete QCD Lagrangian can be written, in analogy to the electroweak theory (1.2), as

$$\mathcal{L}_{\text{QCD}} = \sum_{i=1}^{N_c} \sum_{f=1}^{N_f} \bar{q}_{f_i} (i\gamma^\mu D_\mu - m_f) q_{f_i} - \frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu}. \quad (1.15)$$

The sums run over the quark flavours N_f and all colour states N_c . The quark fermion field is labelled with q_{f_i} . The last term in eq. (1.15) describes the dynamics of the gluon fields A_μ , which contain terms $\sim A^3, A^4$, due to the non-abelian structure of the $SU(3)$ group, $[A_\mu, A_\nu] \neq 0$. This term is also called Yang-Mills term and describes the self-coupling of the gluons

$$\mathcal{L}_{\text{YM}} = -\frac{1}{2} \text{Tr} (F_{\mu\nu} F^{\mu\nu}) = -\frac{1}{4} F_{\mu\nu}^a F^{a\mu\nu}. \quad (1.16)$$

The field tensors in QCD, $F_{\mu\nu}^a$, can be defined in analogy to eq. (1.5).

A characteristic property of α_S , which represents the strength of the strong force, is its different behaviour in the low and high energy regime. Due to its dependence

on the energy scale it has acquired the name *running coupling*. For high energies, i.e. short distances, the running coupling decreases. The quarks behave like free particles, which is called *asymptotic freedom*. The reason for this is the self-interaction of the gluons, which causes an anti-screening of the colour charge of the quarks. In the low energy regime, which corresponds to hadron physics, the coupling is large and the quarks are captured inside hadrons. This effect is called *confinement*.

The strong coupling depends on the energy scale Q^2 and can be written in leading order (LO) as [8]

$$\alpha_S(Q^2) = \frac{12\pi}{(33 - 2N_f) \ln \frac{Q^2}{\Lambda_{\text{QCD}}^2}}. \quad (1.17)$$

The QCD scale Λ_{QCD} is an experimentally determined parameter and obtained to be $\Lambda_{\text{QCD}} \sim 200$ MeV. In high energy collisions, the production process depends on the hard scale Q^2 . In the limit of asymptotic freedom, when $\alpha_S(Q^2 \rightarrow \infty) \rightarrow 0$, the cross section can be expanded perturbatively in powers of α_S . In the limit of confinement, $\alpha_S(Q^2 \rightarrow \Lambda_{\text{QCD}}^2) \rightarrow \infty$, e.g. by separating a quark anti-quark pair, the strong force becomes so large that new $q\bar{q}$ pairs can be produced. The partons form baryons and mesons, and collimated bunches of hadrons are measured as jets in the detector. In the low energy regime the production process can not be calculated perturbatively and the theory relies on phenomenological models.

1.2 Proton-Proton Collisions

The most powerful tests of perturbative QCD (pQCD) are scattering experiments. Prior collider experiments, such as HERA and Tevatron, performed essential measurements to improve the understanding of QCD. A summary of important results including jets measured by the H1 and ZEUS Collaborations at HERA to test pQCD can be found in [9–15]. Drell-Yan lepton pair production cross sections were also measured by the D0 and CDF Collaborations at the Tevatron and compared to pQCD calculations [16, 17]. A summary of measurements by D0 and CDF relevant for QCD is presented in [18]. CERN’s Large Hadron Collider (LHC) experiments analyse processes at seven times more energy than in previous collider experiments at Tevatron, to allow for a precise test of the theory. The high energy collisions enable to tests the structure of particles.

The parton model [19] is a model to describe the structure of hadrons (e.g. protons) in high-energy physics and how the partons interact inside the colliding hadrons. In the naive parton model, regarding to the infinite momentum frame of the proton, the mass of the proton can be neglected, thus $m_{\text{parton}} \sim 0$. The parton model, however, is a simplified description in order to calculate scattering kinematics and cross sections at tree-level, but when including QCD perturbative corrections the model has to be modified.

1.2.1 Structure of Protons and Factorisation

The proton is composed of elementary particles named partons. The constituents are three valence quarks (uud) embedded in a sea of quark anti-quark pairs (called sea quarks) and gluons. The number of partons in the proton depends on a physical scale probed by the momentum transfer Q^2 of the scattering process. Depending on the scale, the number of partons increases with larger momentum transfer. The probability density of a parton i in the proton is defined as $f_i(x, Q^2)$, depending on the scale of the process Q^2 and the parton momentum fraction x with respect to the proton momentum. These functions are called parton density functions (PDF) and can be extracted from fits to experimental data. The PDFs are universal and thus do not depend on the production process and can therefore be determined for different values of Q^2 .

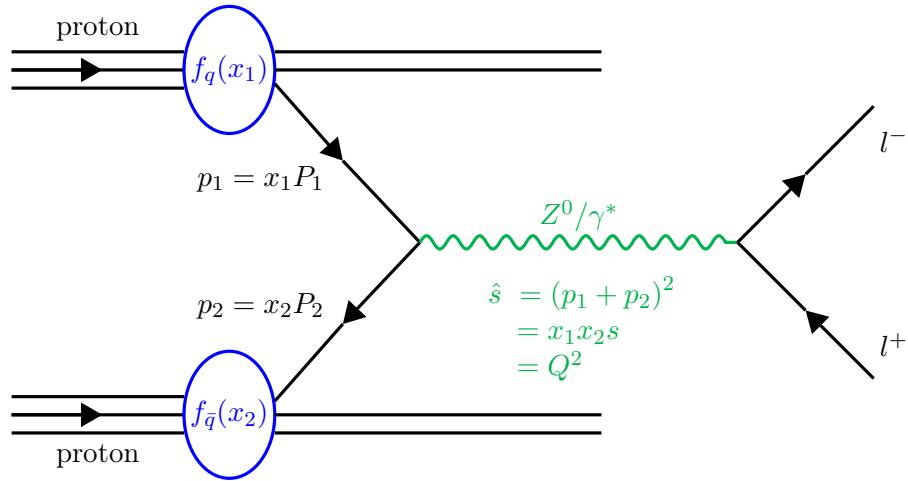


Figure 1.2: Illustration of a proton-proton collision in the parton model. The four-momenta of the protons are defined as $P_{1,2}$, giving the four-momenta of the partons initiating the hard scattering with $p_{1,2}$, depending on the momentum fractions $x_{1,2}$. The interaction of the quark and anti-quark produces a Z^0/γ^* resonance decaying into two leptons l^+l^- .

At high energies, the partons are assumed to be quasi free due to the principle of asymptotic freedom and the interaction between them can be neglected. In a high energy proton-proton collision, the hard scattering process (hard indicates large momentum transfer Q^2) is initiated by two partons of the two protons. Figure 1.2 shows an illustration of the production of a Z^0/γ^* resonance. The square centre-of-mass energy of the proton-proton collision is defined as

$$s = (P_1 + P_2)^2, \quad (1.18)$$

where P_i is the four-momentum of proton i . In the collinear approximation, which means neglecting the transverse momentum of the partons, the four-momenta of the

two partons can be written as

$$p_1 = \frac{\sqrt{s}}{2}(x_1, 0, 0, x_1), \quad (1.19)$$

$$p_2 = \frac{\sqrt{s}}{2}(x_2, 0, 0, -x_2), \quad (1.20)$$

with the momentum fraction x_i of parton i . The square centre-of-mass energy of the parton scattering yields

$$\hat{s} = (p_1 + p_2)^2 = x_1 x_2 s = Q^2. \quad (1.21)$$

In order to calculate the cross section of a QCD process, the collinear factorisation theorem [20] can be used. The factorisation theorem separates the short-distance terms like the partonic cross section including QCD radiation, which can be computed perturbatively, and long-distance contributions, such as hadronisation, PDF, and multiple soft interactions. The long-distance contributions can not be computed precisely in pQCD and are described by phenomenological models [21]. The cross section of a proton-proton collision can then be written as a convolution of the partonic hard cross section and the PDFs of the incoming partons. In the perturbative expansion of the partonic cross section at next-to-leading order (NLO) (or higher-orders), the real and virtual parton emissions have to be included. Due to collinear and soft emissions in the perturbative expansion of the cross section, the perturbative series breaks down. In order to cancel the divergent behaviour at small scales the factorisation scale μ_F is introduced. The collinear singularities can be absorbed in the PDFs by introducing renormalised scale-dependent PDFs $f_i(x, \mu_F)$. A detailed explanation of perturbative corrections in the Drell-Yan cross section is given in section 1.4.2.

The cross section for quark annihilation into a virtual photon or Z can be written as [8]

$$\sigma_{pp} = \sum_q \int dx_1 dx_2 f_q(x_1, \mu_F^2) f_{\bar{q}}(x_2, \mu_F^2) \hat{\sigma}_{q\bar{q} \rightarrow l^+ l^-}, \quad (1.22)$$

where the sum runs over all quark anti-quark combinations and $\hat{\sigma}_{q\bar{q} \rightarrow l^+ l^-}$ is the subprocess cross section for $q\bar{q} \rightarrow l^+ l^-$. The hard sub-process is determined by

$$\hat{\sigma}_{q\bar{q} \rightarrow l^+ l^-} = \left\{ \hat{\sigma}_{\text{LO}}(\mu_F, \mu_R) + \alpha_S(\mu_R^2) \hat{\sigma}_{\text{NLO}}(\mu_F, \mu_R) + \dots \right\}_{q\bar{q} \rightarrow l^+ l^-}. \quad (1.23)$$

Typically, the renormalisation and factorisation scales are equalised to the same order of magnitude as the momentum scale of the hard scattering Q^2 .

The scale dependence of the PDFs enables to propagate $f_i(x, Q_0^2)$, given at a defined scale Q_0^2 , to a scale $Q^2 > Q_0^2$ making use of the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) parton evolution equation [22–24]

$$Q^2 \frac{\partial f_i(x, Q^2)}{\partial Q^2} = \frac{\alpha_S(Q^2)}{2\pi} \int_x^1 \frac{d\xi}{\xi} P_{ij} \left(\frac{x}{\xi}, \alpha_S(Q^2) \right) f_j(\xi, Q^2). \quad (1.24)$$

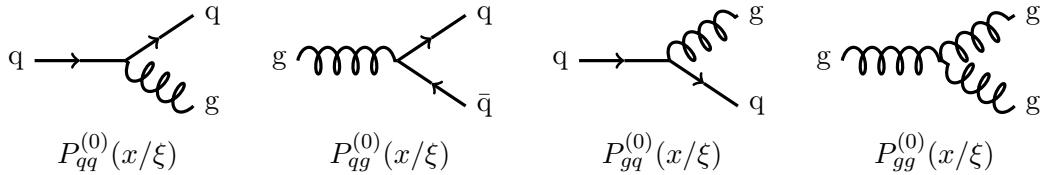


Figure 1.3: Splitting functions at leading order.

The leading order (LO) DGLAP splitting functions $P_{ij}^{(0)}$ can be interpreted as the probability of parton of kind j , with momentum fraction ξ , emitting a parton and becoming parton i with a momentum fraction x .

The splitting functions are available for all kinds of QCD radiation as shown in Figure 1.3 and are defined as

$$P_{qq}^{(0)}(z) = \frac{4}{3} \frac{1+z^2}{1-z}, \quad P_{qg}^{(0)}(z) = \frac{1}{2} (z^2 + (1-z)^2), \quad (1.25)$$

$$P_{gq}^{(0)}(z) = \frac{4}{3} \frac{1+(1-z)^2}{z}, \quad P_{gg}^{(0)}(z) = 2N_c \left[\frac{1}{z} + \frac{1}{1-z} - 2 + z(1-z) \right], \quad (1.26)$$

where $z = x/\xi$ and the number of colour states $N_c = 3$.

The relevant PDF sets for LHC physics are based on fits to data from deep inelastic scattering (DIS) and hard proton collisions. Typical PDF sets are CTEQ [25], MSTW [26], and HERAPDF [27]. Figure 1.4 shows the PDFs of the partons at two different energy scales $Q^2 = 10 \text{ GeV}^2$ (left) and $Q^2 = 10^4 \text{ GeV}^2$ (right), calculated from fits to H1 and ZEUS data. The right figure shows a scenario similar to the PDF for the Drell-Yan process, at $Q^2 \sim M_Z^2$, where M_Z is the mass of the Z boson. It is observed that, with increasing energy scale the PDFs become significantly larger at small x , leading to probe more gluons and sea quarks in the proton.

The rapidity y of the Drell-Yan lepton pair is defined as

$$\begin{aligned} y &= \frac{1}{2} \ln \left(\frac{E+p_z}{E-p_z} \right) \\ &= \frac{1}{2} \ln \left(\frac{x_1}{x_2} \right). \end{aligned} \quad (1.27)$$

The rapidity can be written in terms of the momentum fraction x_1 and x_2 . The kinematic relation of the rapidity and the momentum fraction x yields

$$x_1 = M/\sqrt{s} \exp(y), \quad x_2 = M/\sqrt{s} \exp(-y), \quad (1.28)$$

and is illustrated for different invariant masses and energies in Figure 1.5.

The kinematic region of QCD, where the hard scale of the process is large compared to the QCD scale and small compared to the total centre-of-mass energy, i.e.

$$\Lambda_{\text{QCD}} \ll \mu \ll \sqrt{s}, \quad (1.29)$$

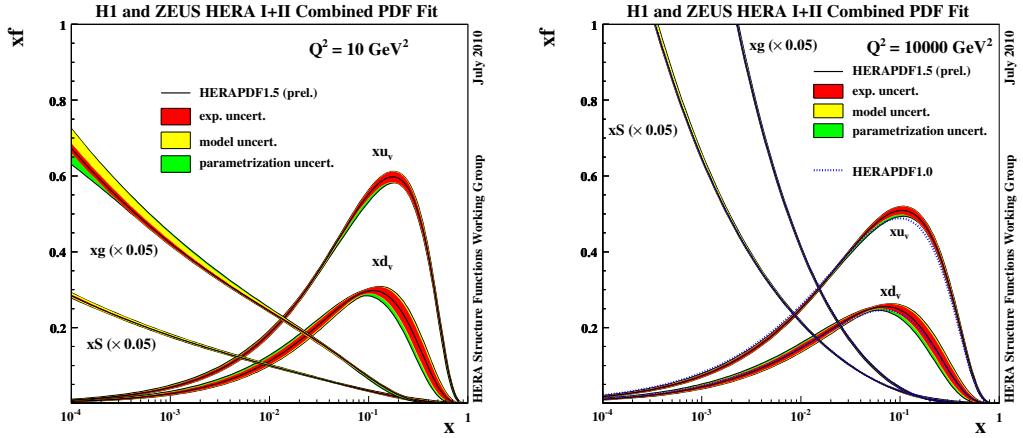


Figure 1.4: PDFs for quarks, anti-quarks and gluons for energy scale $Q^2 = 10 \text{ GeV}^2$ (left) and $Q^2 = 10^4 \text{ GeV}^2$ (right). The distributions are determined from the H1 and ZEUS Collaborations from global fits to data [28].

is defined as the *small-x region* [29]. In this corner of phase space, different physics effects are not completely understood, e.g. the strong increase of the PDFs or parton saturation. The phenomenology of small- x QCD can be studied by measuring asymmetric hard QCD collisions [30]. In this configuration one parton has a large value in x and the other a small value in x , which leads to activity in the *forward region* (large rapidity values y) of the detector, cf. Figure 1.5.

1.2.2 Underlying Event and Pileup Contributions

The hadronic final states of hadron-hadron collisions are composed of an overlay of several contributions [31]. Thus, final states can also be caused by contributions not resulting from the hard scattering process. These effects are called Underlying Event (UE). It is a combination of soft parton radiation, additional parton scatterings, and additional scatters of the proton beam remnants. Partons initiating the hard process are able to emit additional partons. This is called initial-state radiation (ISR). Similar radiation is possible for the final-state partons after the hard scattering, and from partons emerging from multiple soft scatters called final-state radiation (FSR).

In the case of Drell-Yan lepton pair production, the hadronic final states are not affected by the QCD FSR, this provides a suitable study of parton shower effects due to ISR. The importance of the parton shower effects in the case of inclusive jet production, especially in the forward region of large rapidity values, is discussed in section 1.3.3.

Furthermore, multiple parton scatterings of softer partons, also called Multi Parton Interaction (MPI), which do not interact with the hard process, can occur. This phenomena is nonperturbative and has to be modelled. A detailed description on available tools used for the modelling of the nonperturbative effects, like MPI but

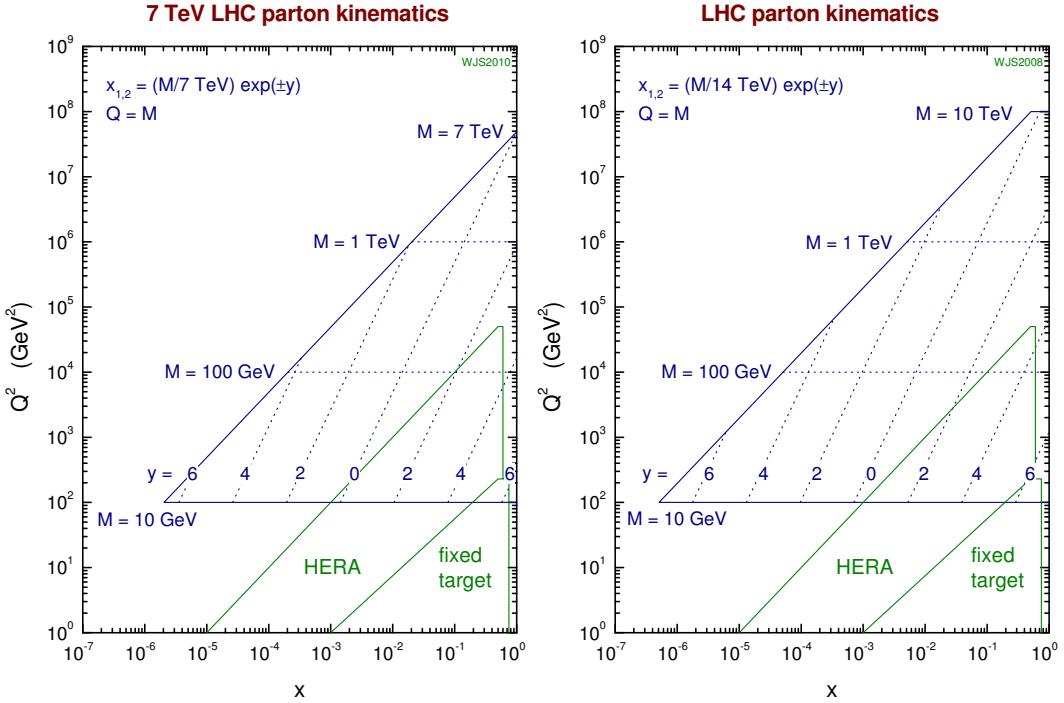


Figure 1.5: Kinematic relation between partonic quantities: momentum fraction x , hard scale Q^2 and the final-state quantities: invariant mass M and rapidity y , for different centre-of-mass energies [21].

also hadronisation, is provided in section 1.3.2.

Due to the fact that high density proton bunches are collided, the probability to have more than one proton-proton collision increases with increasing luminosity. These events, emerging from additional proton-proton collisions, are called pileup (PU) events and influence the measurement.

1.3 Theory Comparison

Monte Carlo (MC) event generators contain a large library of hard SM and Beyond Standard Model (BSM) processes, which are interfaced with different models of parton shower, hadronisation, and hadronic decays. There are different uses of MC generators: In one way, they provide accurate predictions of pQCD theory calculations in combination with different phenomenological models of nonperturbative effects. Thus, they can be used to provide predictions for collider experiments or set limits on certain parameter space regions. In the other way, they are used to obtain corrections applied to measured data in order to account for detector effects, and corrections applied to fixed-order theory calculations, to account for nonperturbative effects.

In this section, first the MC method is introduced and a summary of the MC generators, used in this analysis, is presented. Further, the treatment of nonpertur-

bative and parton shower effects obtained by Shower Monte Carlo (SMC) generators, in processes containing hadronic jets, are discussed.

1.3.1 Monte Carlo Event Generators

In high energy physics, data analyses rely on MC event generators. A MC event generator simulates high energy collision processes based on a procedure using random number generation and probability statistics. The MC simulation is performed starting with initial conditions and subsequent random sampling of processes and final states on an event by event basis. There are different MC event generators available in high energy physics, which implement different models in QCD and contain several choices of beam particles.

First, the hard scattering process is simulated, which can be determined at LO or NLO depending on the MC generator. Afterwards, parton emissions emerging from initial and final states are included. The initial-state parton shower is performed via the backward-evolution, which means the shower evolution starts at the scale of the hard sub-process and decreases with virtuality to a cut-off scale Q_0 , at which the parton is taken from the hadron. The final-state parton emissions evolve with decreasing scale until perturbation theory breaks down and the hadronisation has to be modelled.

The parton emissions correspond to higher-order corrections in the perturbative series. These corrections are accounted for by evolution equations in SMC programs. The perturbative series in α_S consists of terms like $\alpha_S^n \ln^k(Q^2)$ and $\alpha_S^n \ln^k(1/x)$ and depend on the scale Q^2 and the momentum fraction x of the parton [32]. The evolution equations rearrange the sum of the perturbative series and employ the leading-log approximation (LLA)¹, in which only single logarithmic terms are used. Different evolution equations are used in SMC programs. The most commonly used evolution schemes are DGLAP [22–24], BFKL [33–35], and CCFM [36–38]. They differ in the ordering of the parton emissions depending on the virtuality k_i , momentum fraction x_i , or the splitting angle θ_i of parton i , respectively. Diverse MC generators comprise different parton shower and hadronisation phenomena. The most common fragmentation models are the Lund string [39–41] and cluster model [42].

There are different choices of physics parameters in the MC generator. A set of parameters is called *tune*. In order to yield a reasonable comparison to data, the MC parameters are adjusted to describe the measurement. There are several tunes available, which differ in PDF sets or UE parameters, e.g. probabilities for colour reconnection, parameters to determine the matter overlap between the incoming protons, scale parameters for MPI or parton shower, etc..

The MC generators, which are used in this work are illustrated here:

PYTHIA6

PYTHIA6 [43] is a parton shower based MC event generator. The hard matrix-element

¹The expression *leading-log* refers to leading logarithms.

is calculated at LO and higher-order corrections are modelled by parton showers in LLA. The parton evolution obeys the DGLAP evolution equation and the emissions are ordered in transverse momentum. The hadronisation process is supplemented by the Lund String model. In this analysis the version 6.4.33 of PYTHIA6 is used.

The tune Z2 [44] was tuned to UE data measured with CDF. The tune Z2* is derived from the earlier tune Z2, where the PYTHIA6 parameters PARP(82) and PARP(90) are retuned. The parameters PARP(82) and PARP(90) correspond to the MPI cut-off scale and the MPI energy extrapolation parameter, respectively. Both tunes Z2 and Z2* employ the PDF set CTEQ6L [25]. The energy dependence of MPI is taken into account by tuning to CDF data measured at $\sqrt{s} = 300, 900$ and 1960 GeV. Recently a new tune CUETP6S1 including CMS data at $\sqrt{s} = 7$ TeV became available [45] (see appendix E).

PYTHIA8

PYTHIA8 is the successor of PYTHIA6 and employs an improved MPI model, which is interleaved by parton shower. The PDF set CTEQ6L and the tune 4C [46] is used. In the case of PYTHIA8 also an improved UE tune was recently published by us (see appendix E), including CDF and CMS data at different centre-of-mass energies, named CUETP8S1 [45].

HERWIG++

HERWIG++ 2.5.0 [47] is a parton shower based LO and NLO event generator. The QCD parton emissions from the initial and final states are taken into account via an angular ordering in θ , where θ is the angle of the radiated parton. For the hadronisation, the cluster model, based on nonperturbative gluon splitting, is adopted. In this work the PDF set CTEQ6L1 and the UE tune EE-3C [48] are used.

MADGRAPH

MADGRAPH [49] calculates the hard matrix-element for high energy $2 \rightarrow n$ collisions at LO. In this work, MADGRAPH generates ≤ 4 partons in the final state. The generation of Drell-Yan + n partons is done separately for $n = 0, \dots, 4$, which leads to a precise description of the multi-jet topology. The LO diagrams are supplemented with parton shower programs to model the parton emissions and the UE. In order to avoid double counting of parton emissions produced by MADGRAPH and the supplemented parton shower, a phase space threshold is used. This matching procedure of matrix-element calculation and parton shower is described in [50]. In this thesis, MADGRAPH is interfaced with PYTHIA6.

POWHEG

POWHEG stands for POsitive Weight Hard Emission Generator [51–53]. It generates the hard matrix-element at NLO. One feature of POWHEG is that the hardest emission is also generated at NLO. The modelling of the parton shower and nonperturbative effects is obtained by interfacing the POWHEG output with any SMC program. In order not to double count the first hard emission when interfacing with the parton shower, the SMC generator must use a p_T ordered parton shower algorithm or be able to apply a p_T veto. In this thesis, POWHEG is interfaced with PYTHIA6.

CASCADE

In CASCADE [54, 55], the hard matrix-element is generated off-shell and at LO QCD. The parton emissions follow the CCFM evolution algorithm and require an unintegrated PDF, depending on the transverse momentum of the propagator. The transverse momentum dependent (TMD) distribution set JH-2013-set2 [56] is used. The initial-state parton cascade is modelled by a pure gluon chain. For the fragmentation, the Lund string model is adopted.

1.3.2 Nonperturbative Effects in Shower Monte Carlo Generators

Cross sections for high energy processes measured at the LHC are compared to theory calculations in order to test pQCD. To obtain a meaningful comparison between data and MC predictions, the calculation has to be known to the highest accuracy. Precise predictions for jet measurements are achieved by fixed-order pQCD calculations and are available at NLO [57–59]. In order to compare measured data to fixed-order calculations, both sides have to refer to the same level of measured particles. Therefore, measured observables (*detector level*) have to be corrected for detector effects in order to refer to stable-particles, i.e. colour neutral particles, with mean decay length of $c\tau > 10$ mm (*stable-particle level*). The NLO theory calculation, however, refers to particles on *parton level* and has to be corrected to account for nonperturbative effects, such as MPI and hadronisation.

Hadronisation (HAD)

At the scale where pQCD breaks down, the hadronisation of the partons begins. Partons of the incoming protons interact in the hard scattering and the outgoing partons carry colour charge. A strong colour field arises between the partons and the rest of the proton. Increasing the distance between two partons of opposite charge increases the colour field. The rising of α_S at low scales makes it impossible to separate two partons, but, the gain in energy will create a new pair of partons. With every creation of a new pair of partons, the original parton loses some of its energy and momentum until there is no energy left to create new partons and the colour charge becomes neutral. At the end of the hadronisation, the original partons are confined in hadrons, which form jets. There are two common phenomenological models of the hadronisation implemented in SMC generators: the Lund string fragmentation, which is adapted by PYTHIA, and the cluster fragmentation used by HERWIG++.

Multi Parton Interaction (MPI)

In addition to the hard interaction in proton-proton collisions, soft interactions can occur and create particles, which also leave an energy deposit in the detector. The origin of the additional soft particle production evolves from the proton remnants or from ISR. The proton remnants can interact with the outgoing partons of the hard process or interact with the other proton remnant. Furthermore, the parton shower can be a source of MPI (but this feature is not included in MC simulation). Before the hard partons interact they emit QCD radiation in the form of soft gluons or quarks, which can lead to additional interactions with the remnants or the outgoing partons. The MPI model of PYTHIA and HERWIG++ is explained in [60, 61].

The theoretical predictions for inclusive jet production comprises fixed-order pQCD calculations and a nonperturbative correction factor to account for MPI and hadronisation [59]. In previous inclusive jet analyses [57, 58], the nonperturbative correction was derived from the average predictions of the two LO SMC event generators PYTHIA6 and HERWIG++. In [62] (see appendix F) we propose an alternative method, in which we use a NLO hybrid MC event generator, POWHEG for the generation of the hard sub-process complemented with hadronisation and showering from SMC program of PYTHIA6.

1.3.3 Nonperturbative and Parton Shower Effects in NLO-matched Monte Carlo Generators

Available SMC generators are LO MC generators (PYTHIA, HERWIG++), which treat the hard process with LO accuracy and the emissions of additional partons are accounted for in LLA. The high p_T emissions are not treated correctly in LLA and problems can arise when comparing fixed-order theory calculations to the measurement. Furthermore, when combining the NLO parton level calculations and the nonperturbative correction derived from a LO MC generator, a potential inconsistency results from treating the first radiative correction differently in the two parts of the calculation [62]. It is necessary to provide the QCD corrections with the best accuracy, therefore, an alternative method is performed by using a NLO MC generator to determine the hard scattering. The events are generated within the POWHEG BOX framework [63, 64] and are then showered with PYTHIA.

Parton Shower (PS)

The parton shower represent the higher-order corrections to the fixed-order hard matrix-element of the scattering process. The shower evolves from collinear parton splittings and soft-gluon emissions. The initial-state radiation evolves from the hard process starting at a high space-like virtuality via a backward-evolution to lower virtuality. During the final-state showering the virtual scale Q^2 decreases until reaching a minimum value, where the evolution is stopped and nonperturbative models have to be considered.

The impact of collinear partons, treated as low-order approximations in the SMC generators, give a non-negligible contribution, especially in the forward rapidity region. Kinematic effects are observed in longitudinal momentum distributions when combining the collinear approximation with energy momentum conservation [62, 65]. The longitudinal momentum distributions are shifted towards larger values at large rapidities. Further studies on kinematic effects due to non-collinear parton emissions are given in [66–68]. This dynamical impact of the treatment of parton shower in SMC generators can influence the determination of the PDFs. It is of great importance to also study the parton shower effects in terms of TMD distributions [69–72] in formulating non-collinear momentum components using unintegrated initial-state distributions. TMD branching algorithms are employed for the evolution of QCD initial-state radiation in the case of vector boson + jet production in [73].

Nonperturbative and Parton Shower Corrections

Estimating the correction factors by a NLO MC event generator interfaced with the SMC generator PYTHIA motivates the study of two separate factors to the fixed-order calculation due to nonperturbative and parton shower effects. We define in [62] the nonperturbative and parton shower corrections from a NLO-matched SMC event generator as the cross section ratios with different generator settings as

$$K_{\text{NLO}}^{\text{NP}} = \frac{N_{\text{NLO}}^{\text{PS+HAD+MPI}}}{N_{\text{NLO}}^{\text{PS}}}, \quad (1.30)$$

$$K_{\text{NLO}}^{\text{PS}} = \frac{N_{\text{NLO}}^{\text{PS}}}{N_{\text{NLO}}}. \quad (1.31)$$

The numerator of the nonperturbative correction, eq. (1.30), is defined by a simulation including parton shower (PS), hadronisation (HAD), and MPI and the denominator by simulation at NLO only including PS. The parton shower correction, eq. (1.31), is given as the ratio of the generator settings including PS divided by the pure NLO cross section without any showering or nonperturbative effects applied. By estimating the corrections from a NLO-matched SMC event generator, the fixed NLO calculation can be corrected for nonperturbative and parton shower effects by multiplying two individual correction factors. A NLO calculation for inclusive jet cross section as a function of jet p_T and rapidity y , can then be corrected by [59]

$$\frac{d^2\sigma_{\text{theo}}}{dp_T dy} = \frac{d^2\sigma_{\text{NLO}}}{dp_T dy} \cdot K_{\text{NLO}}^{\text{NP}} \cdot K_{\text{NLO}}^{\text{PS}}. \quad (1.32)$$

The nonperturbative and parton shower corrections obtained from a NLO-matched MC generator are applied for the first time in inclusive jet measurements of CMS [59].

I calculated the nonperturbative corrections with POWHEG +PYTHIA6, as shown in Figure 1.6, in comparison to the correction factors obtained by LO generators PYTHIA6 and HERWIG++, as derived in [58]. I calculated the corrections as a function of the jet p_T in five bins of the absolute rapidity $|y|$ from 0 to 2.5. In order to estimate the impact of the UE, I used two different tunes, P11 [74] and Z2*, for PYTHIA6. The average of the two predictions of POWHEG +PYTHIA6 with tune Z2* and P11 is shown. In all cases, I employ a parametrisation in form of a fit using a functional form of $a_0 + a_1/p_T^{a_2}$.

Comparing the factors derived from LO and NLO-matched MC event generators, it can be observed that they decrease with increasing jet p_T and approach one at very high p_T . Differences at low jet p_T come from the matching of MPI to the NLO calculation. The scale of MPI is typically smaller than the scale of the hard subprocess, which is defined as the average of the transverse momentum of the generated partons in the MC generator. The average transverse momentum of the generated partons is different in the LO and NLO generation.

For Figure 1.7, I have derived the parton shower corrections by POWHEG +PYTHIA6 using the tune Z2. To examine the impact of the parton shower, I changed the upper scale limit μ_{PS} of the evolution. In PYTHIA6 this corresponds to the parameter PARP(67), which is varied by factors of 0.5 and 1.5 as shown in Figure 1.7. Again, I use a parametrisation of a functional form of $a_0 + a_1/p_T^{a_2}$.

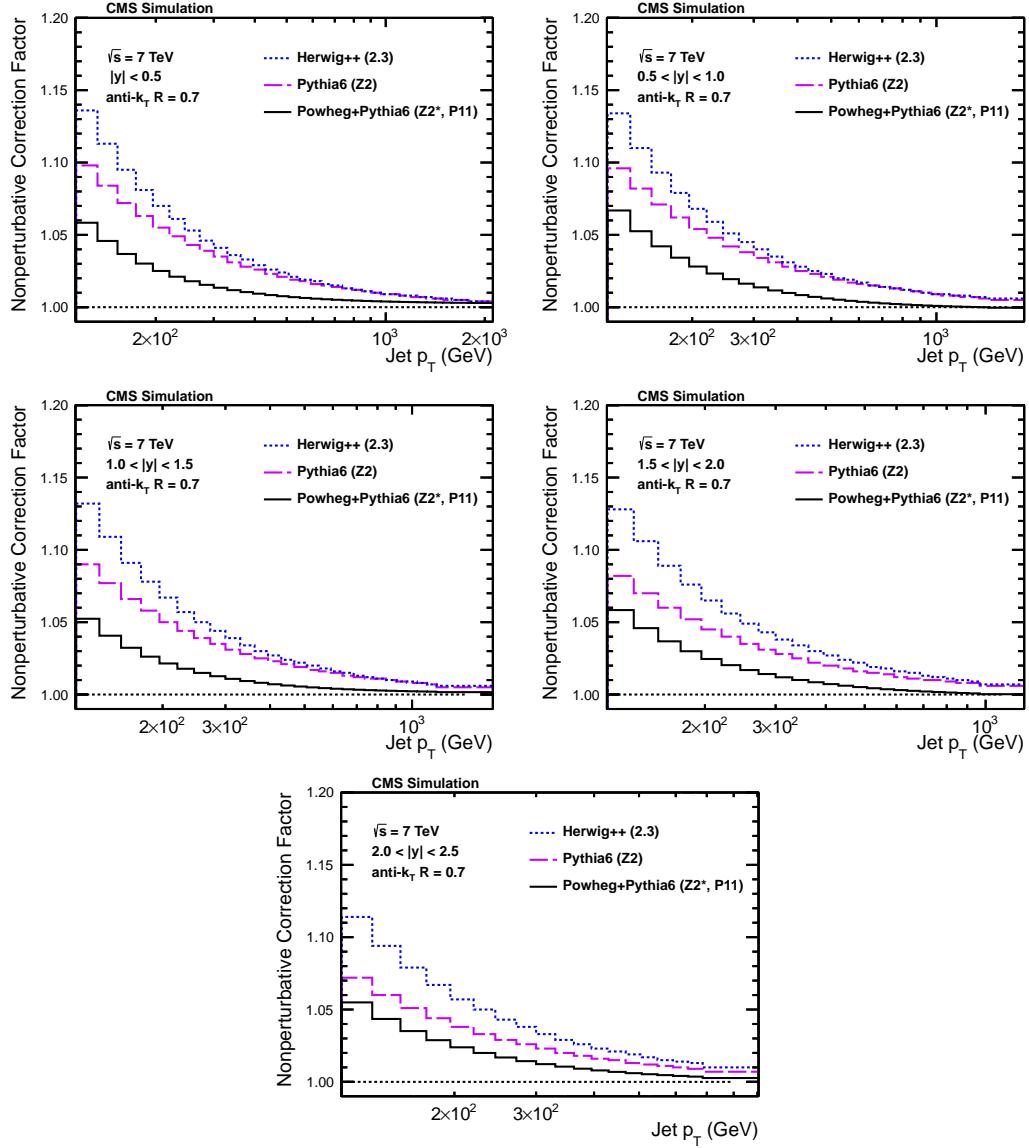


Figure 1.6: Nonperturbative corrections as derived in [58] from PYTHIA6 tune Z2 and HERWIG++ with the default tune of version 2.3 in comparison to corrections as derived from POWHEG +PYTHIA6 with the two different underlying event tunes P11 and Z2* [59].

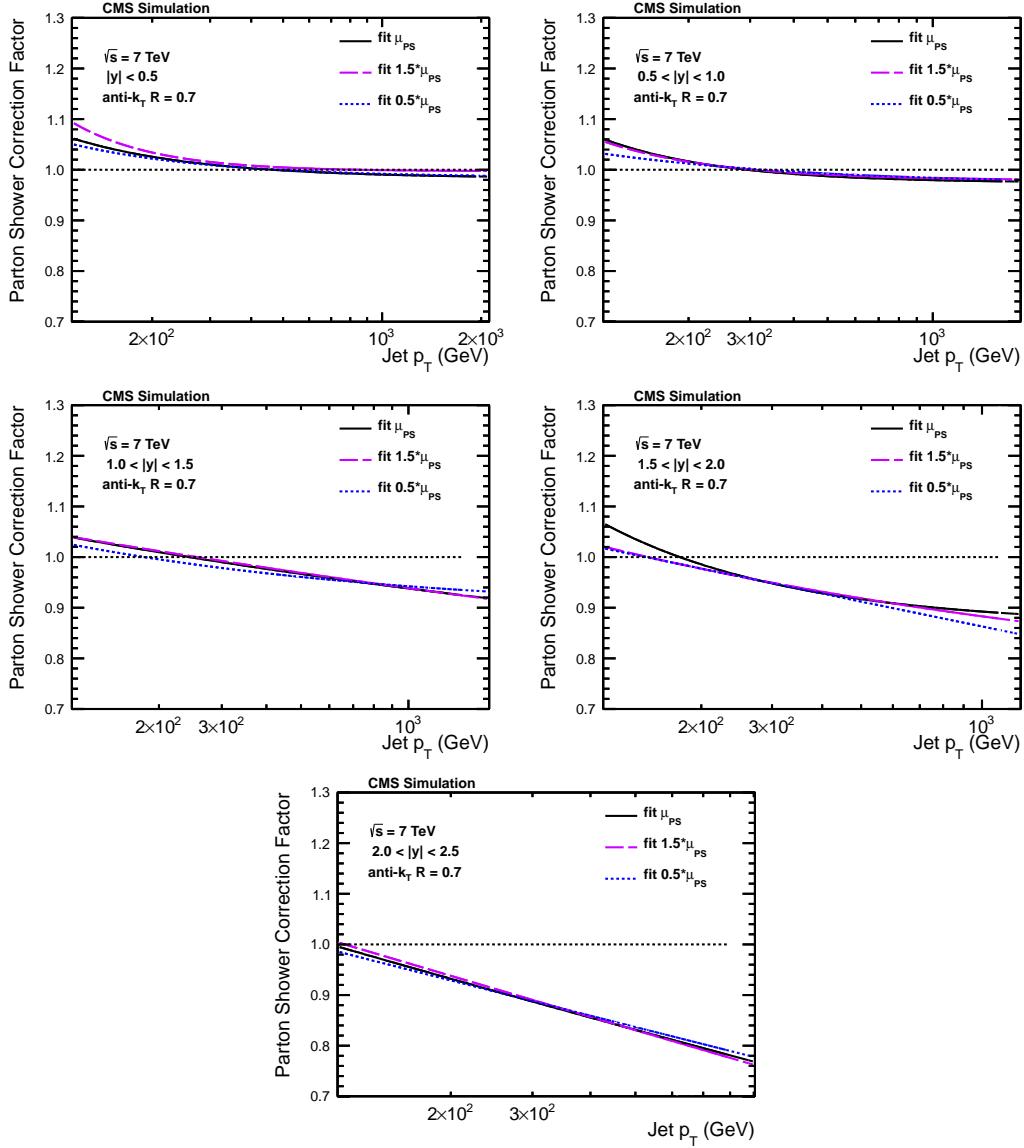


Figure 1.7: Parton shower corrections as derived from POWHEG +PYTHIA6 for different upper scale limits of the parton shower evolution in PYTHIA6 tune Z2. The curves parametrise the correction factors in the five rapidity regions as a function of the jet p_T [59].

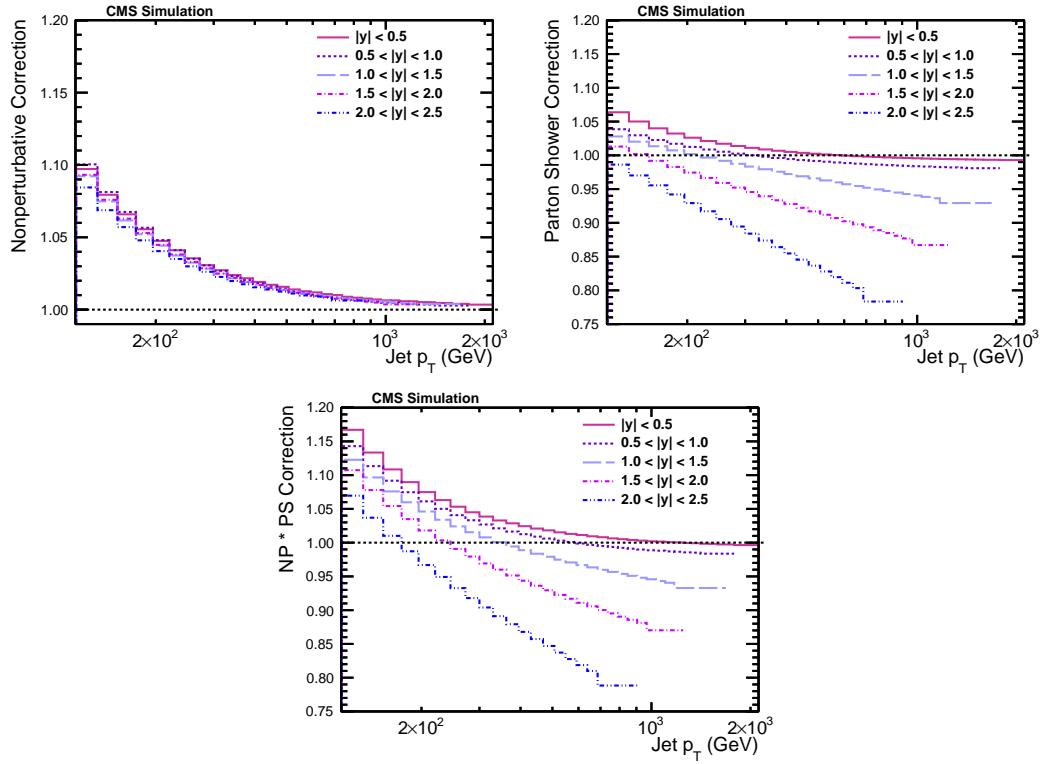


Figure 1.8: Nonperturbative correction (top left) as obtained from the envelope of the predictions of PYTHIA6 tune Z2, HERWIG++ tune 2.3, and POWHEG +PYTHIA6 with the tunes P11 and Z2*, parton shower correction (top right) as obtained from the average of the predictions of POWHEG +PYTHIA6 tune Z2 with scale factor variation, and total correction (bottom) defined as the product of the nonperturbative and parton shower correction [59].

The impact of parton emissions to jet transverse energy spectra has a sizeable effect over the full rapidity region [57, 58], but gives a non-negligible contribution, especially at forward rapidity.

Finally, in Figure 1.8 I present an overview of the nonperturbative, the parton shower, and the total corrections, as they are used in [59], for all five ranges in $|y|$.

1.4 Drell-Yan Production

The Drell-Yan process was first presented by Drell and Yan [3] describing quark anti-quark annihilation into a lepton pair l^+l^- with invariant mass $M^2 = (p_{l^+} + p_{l^-})^2 \gg 1$ GeV². In hadron-hadron collisions, the quark and anti-quark are constituents of the two incoming hadrons and can create an off-shell virtual boson (Z or γ^*), which then directly decays into two leptons. In the following, the Drell-Yan cross section is calculated first in the parton model and including perturbative corrections, following the calculations in [8, 75].

1.4.1 Cross Section

The cross section in the parton model follows eq. (1.22). At lowest-order in pQCD, the sub-process $\hat{\sigma}_{q\bar{q} \rightarrow l^+l^-}$ can be obtained from the $e^+e^- \rightarrow l^+l^-$ cross section, including a factor $1/N_c = 1/3$ to take into account the colour state of the initial-state quarks

$$\hat{\sigma}(q(p_1)\bar{q}(p_2) \rightarrow l^+l^-) = \frac{4\pi\alpha^2}{3\hat{s}} \frac{1}{N_c} Q_q^2. \quad (1.33)$$

The quark electric charge is labelled as Q_q and the centre-of-mass energy of the partonic cross section is given by eq. (1.21). The quantities corresponding to the partonic process are labelled with $\hat{\cdot}$. The incoming $q\bar{q}$ reveal different collision energies $\sqrt{\hat{s}}$, hence different invariant masses of the lepton pair can be produced, corresponding to the production of a virtual photon, $q\bar{q} \rightarrow \gamma^* \rightarrow l^+l^-$, as well as for higher $\sqrt{\hat{s}}$, W and Z bosons are produced. Thus, the differential lepton pair invariant mass M distribution is considered. The sub-process cross section yields

$$\frac{d\hat{\sigma}}{dM^2} = \frac{4\pi\alpha^2}{3M^2} \frac{1}{N_c} Q_q^2 \delta(\hat{s} - M^2). \quad (1.34)$$

The differential cross section of the Drell-Yan process derived in the parton model yields [8]

$$\begin{aligned} \frac{d\sigma}{dM^2} &= \int_0^1 dx_1 dx_2 \sum_{q=1}^{N_f} \{f_q(x_1)f_{\bar{q}}(x_2) + f_{\bar{q}}(x_1)f_q(x_2)\} \frac{d\hat{\sigma}}{dM^2}(q\bar{q} \rightarrow l^+l^-) \\ &\stackrel{(1.34)}{=} \frac{4\pi\alpha^2}{3M^2} \frac{1}{N_c} \int_0^1 dx_1 dx_2 \delta(x_1 x_2 s - M^2) \left[\sum_{q=1}^{N_f} Q_q^2 \{f_q(x_1)f_{\bar{q}}(x_2) + f_{\bar{q}}(x_1)f_q(x_2)\} \right] \end{aligned}$$

$$= \frac{4\pi\alpha^2}{3M^2} \frac{1}{N_c} \int_0^1 dx_1 dx_2 \delta(x_1 x_2 s - M^2) P_{q\bar{q}}(x_1, x_2). \quad (1.35)$$

The sum runs over the N_f quark flavours and the term $f_{\bar{q}}(x_1)f_q(x_2)$ specifies the additional contribution from anti-quark quark contributions. The dependence on the PDFs is defined in the joint $q\bar{q}$ probability function $P_{q\bar{q}}(x_1, x_2)$ ² [75]. In the parton model, the PDFs $f_i(x)$ do not depend on the scale of the process, as it is written in eq. (1.22). Thus, the lepton pair cross section can be written as a function of a dimensionless scaling variable $\tau = \frac{M^2}{s}$ by multiplying eq. (1.35) with M^4

$$\begin{aligned} M^4 \frac{d\sigma}{dM^2} &= \frac{4\pi\alpha^2}{3N_c} \tau \int_0^1 dx_1 dx_2 \delta(x_1 x_2 - \tau) P_{q\bar{q}}(x_1, x_2) \\ &= \frac{4\pi\alpha^2}{3N_c} \tau \mathcal{F}(\tau). \end{aligned} \quad (1.36)$$

In general, the differential cross section is a function of the centre-of-mass energy s and the mass of the system M , but eq. (1.36) shows that the cross section depends on the scaling variable $\tau = \frac{M^2}{s}$.

1.4.2 Perturbative QCD Corrections

The NLO partonic cross section was briefly addressed in eq. (1.23). The $\mathcal{O}(\alpha_S)$ QCD corrections to the partonic cross section correspond to loop and real correction, where the latter correlates to additional parton emissions. Adopting the scaling behaviour of the LO parton model cross section in eq. (1.36), the partonic cross section can be written as

$$M^4 \frac{d\hat{\sigma}}{dM^2} = \frac{4\pi\alpha^2}{3N_c} \tau \hat{\mathcal{F}}(\tau) \quad (1.37)$$

with

$$\hat{\mathcal{F}}(\tau) = Q_q^2 \delta(1 - \tau). \quad (1.38)$$

In pQCD, the function $\hat{\mathcal{F}}$ can be expanded in the strong coupling α_S

$$\hat{\mathcal{F}}(\tau) = \hat{\mathcal{F}}_0(\tau) + \frac{\alpha_S}{2\pi} \hat{\mathcal{F}}_1(\tau) + \dots \quad (1.39)$$

The LO and NLO diagrams of the Drell-Yan process are illustrated in Figure 1.9.

The real and loop corrections in the cross section calculation introduce divergences, of different types, which have to be handled in the calculation. There are three different types of singularities:

1. *Ultraviolet (UV)* divergences from loop diagrams, when the energy of the emitted parton tends to infinity.

² $P_{q\bar{q}}(x_1, x_2)$ is not to be mistaken with the splitting function $P_{q\bar{q}}$. The electric charge Q_q is included in the definition of $P_{q\bar{q}}(x_1, x_2)$.

2. *Infra-red (IR)* divergences from real and virtual contributions due to soft-gluon emissions. These singularities occur due to massless particles (when the energy tends to zero) and are therefore also called mass-singularities.
3. *Collinear* divergences, which are induced by parton splittings of the initial-state partons.

The UV divergences are treated with renormalisation, where different types of regularisation schemes are used. The dimensional regularisation scheme and the massive gluon scheme are two approaches discussed in [75]. In the dimensional regularisation, the integration variables in space time dimension are changed from 4 to $4 - 2\epsilon$ and a renormalisation scale μ_R is introduced. The massive gluon scheme regularises the real and virtual corrections by defining a mass to the gluon. The IR divergences can be handled when combining the real and virtual contributions in the calculation. If the emitted parton becomes soft, the diagram reveals a singularity, which can be cancelled by a soft singularity, provided by the divergence of one of the virtual diagrams. Therefore, the IR singularities can be regulated by adding the contributions from real and virtual diagrams. Additionally, the real diagrams also introduce collinear divergences, i.e. the branching angle of the emitted parton tends to zero. These soft and collinear divergences have to be handled in the perturbative calculation.

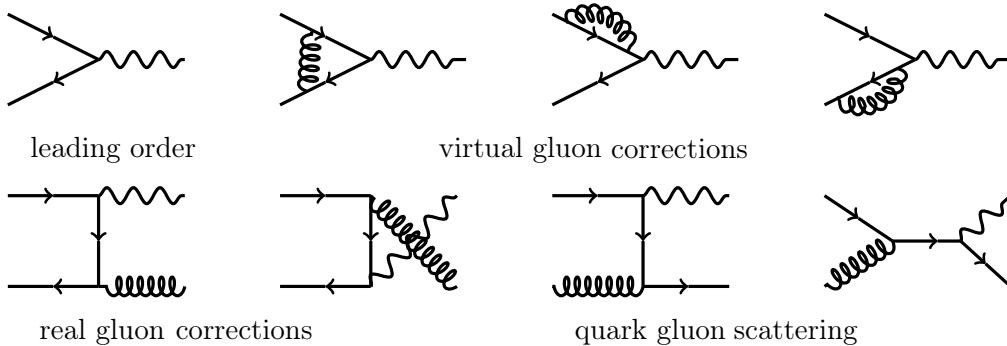


Figure 1.9: LO and NLO feynman diagrams for the Drell-Yan process. The solid lines correspond to quarks or anti-quarks, the curly lines represent gluons and the wavy lines are Z^0/γ^* .

The partonic cross section contains contributions from virtual and real gluon corrections, $\hat{\mathcal{F}}_1^{q\bar{q},V}$ and $\hat{\mathcal{F}}_1^{q\bar{q},R}$, and corrections due to quark gluon scattering $\hat{\mathcal{F}}_1^{qg}$. The results at leading order in ϵ in the dimensional regularisation scheme yield [8]

$$\begin{aligned} \hat{\mathcal{F}}_1^{q\bar{q}} &= \hat{\mathcal{F}}_1^{q\bar{q},V} + \hat{\mathcal{F}}_1^{q\bar{q},R} \\ &= Q_q^2 \frac{\alpha_S(\mu^2)}{2\pi} \left[2 \left(-\frac{1}{\epsilon} - \ln(4\pi) + \gamma_E + \ln \frac{M^2}{\mu^2} \right) P_{q\bar{q}}^{(0)}(\tau) + D_q(\tau) \right], \end{aligned} \quad (1.40)$$

$$\hat{\mathcal{F}}_1^{qg} = Q_q^2 \frac{\alpha_S(\mu^2)}{2\pi} \left[\left(-\frac{1}{\epsilon} - \ln(4\pi) + \gamma_E + \ln \frac{M^2}{\mu^2} \right) P_{qg}^{(0)}(\tau) + D_g(\tau) \right]. \quad (1.41)$$

The constant $\gamma_E = 0.5772$ is the Euler constant, and the functions D_q and D_g are

defined as [8]

$$D_q(z) = \frac{4}{3} \left[4(1+z^2) \left(\frac{\ln(1-z)}{1-z} \right)_+ - 2 \frac{1+z^2}{1-z} \ln z + \delta(1-z) \left(\frac{2\pi^2}{3} - 8 \right) \right], \quad (1.42)$$

$$D_g(z) = \frac{1}{2} \left[(z^2 + (1-z)^2) \ln \frac{(1-z)^2}{z} + \frac{1}{2} + 3z - \frac{7}{2}z^2 \right]. \quad (1.43)$$

The coefficients of the $\frac{1}{\epsilon}$ divergences are the LO splitting functions defined in eq. (1.25). From eq. (1.40) one can see that only collinear divergences remain when real and virtual diagrams are added. The soft singularities at $z = 1$ are regularised by the plus prescription, which is defined via the equation

$$\int_0^1 dx \frac{f(x)}{(1-x)_+} = \int_0^1 dx \frac{f(x) - f(1)}{1-x} \quad (1.44)$$

for any sufficiently smooth function f at the end point with the properties

$$\frac{1}{(1-x)_+} = \frac{1}{1-x} \quad \text{for } 0 \leq x < 1, \quad (1.45)$$

and

$$\int_0^1 dx [f(x)]_+ = 0. \quad (1.46)$$

The collinear singularities are absorbed in the PDFs when the partonic cross section is convoluted with parton densities. The PDFs are redefined

$$f_i(x, \mu_F) = f_i(x) c_i^{\text{divergent}}(x, \alpha_S, \mu_F). \quad (1.47)$$

The *renormalised* PDFs are the measurable quantities and depend on the so-called factorisation scale μ_F . In order to verify these calculations, the divergent part of the cross section has to be included in the nonperturbative part. This observation is known as the factorisation theorem, which factorises the cross section into perturbative and nonperturbative parts.

1.4.3 Transverse Momentum Distribution

The Drell-Yan transverse momentum distribution is crucial to test perturbative as well as nonperturbative QCD. In this analysis, Drell-Yan production is investigated in the decay channel of two opposite charged muons. Therefore, the Drell-Yan lepton pair is referred to as the dimuon system. The Drell-Yan lepton pair vectorial transverse momentum $\vec{p}_T^{\mu\mu}$ is defined as the momenta of the two muons

$$\vec{p}_T^{\mu\mu} = |\vec{p}_T^{\mu_1} + \vec{p}_T^{\mu_2}|, \quad (1.48)$$

which can be translated into the p_T of the Z^0/γ^* . Before the two partons initiate in the scatter, their contribution to the transverse motion can be neglected. At LO this

yields, due to momentum conservation in the transverse plane, that the two outgoing leptons are back-to-back and the dimuon system has $p_T^{\mu\mu} = 0$. However, in general, additional partons are emitted from the initial-state partons, which correspond to higher-order corrections in perturbation theory. At NLO, an additional gluon can be emitted from one of the two quarks, which is then measured as a jet in the detector. In this way, the Drell-Yan dimuon pair gains $p_T^{\mu\mu} > 0$ and is balanced by the radiated gluon $p_T^{\mu\mu} \propto p_T^{\text{gluon}}$. A schematic illustration of the contributions in the transverse plane is presented in Figure 1.10.

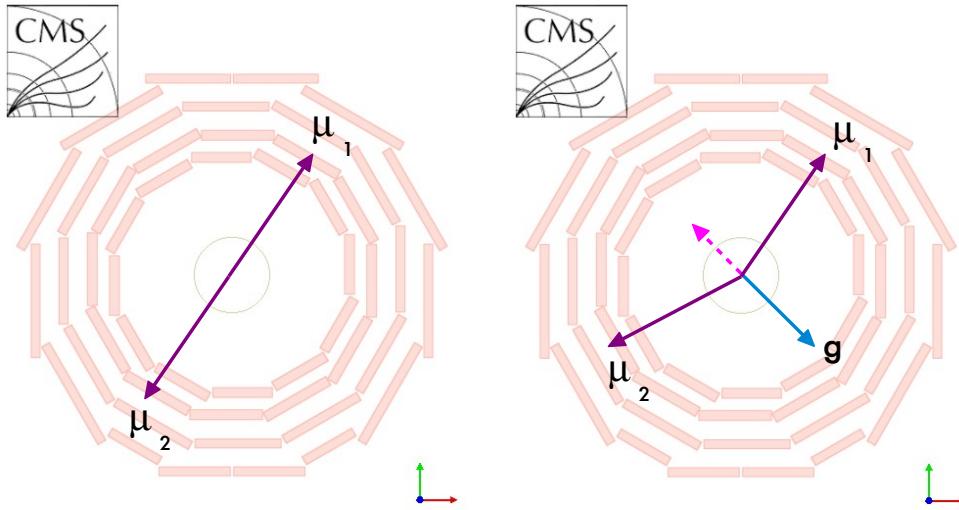


Figure 1.10: Drell-Yan production decaying into two muons, showing the contributions in the transverse plane of the detector at LO (left) and NLO (right). At LO, the dimuon system has zero transverse momentum, resulting in a back-to-back composition of the two muons (represented by the two arrows). At NLO one additional gluon (shown here in blue) is emitted providing non-zero p_T to the dimuon system. The gluon is measured as a jet in the detector. The purple dashed arrow indicates the contribution of the dimuon system, which is not directly measured with the detector, but by the two individual muons.

The perturbative calculation of the Drell-Yan $p_T^{\mu\mu}$ distribution, or the differential cross section as a function of $p_T^{\mu\mu}$, comprises different theoretical calculations. In the following I concentrate on the Drell-Yan process producing a virtual photon. The production of the massive Z boson can be included in a similar way. At small $p_T^{\mu\mu}$, the distribution is dominated by multiple soft-gluon emissions, which have to be treated in an all-order resummation. At high $p_T^{\mu\mu}$, the function follows perturbative QCD at fixed-order, where $2 \rightarrow 2$ processes, like $q\bar{q} \rightarrow \gamma^*g$ or $qg \rightarrow \gamma^*q$, are relevant. The partonic cross section for a $2 \rightarrow 2$ process can be written as

$$\frac{d\hat{\sigma}}{d\hat{t}} = \frac{1}{16\pi} \frac{1}{\hat{s} + Q^2} \frac{1}{\hat{s}} |\mathcal{M}|^2, \quad (1.49)$$

including the virtual mass of the γ^* as Q^2 . The squared matrix-element $|\mathcal{M}|^2$ represents the probability amplitude of the scattering process. The Mandelstam variables represent the different scattering diagrams for the $2 \rightarrow 2$ process and are defined via the incoming and outgoing momentum variables. In the case of the annihilation sub-process $q\bar{q} \rightarrow \gamma^*g$, the variables yield ³

$$\hat{s} = (p_q + p_{\bar{q}})^2, \quad (1.50)$$

$$\hat{t} = (q_\gamma - p_q)^2, \quad (1.51)$$

$$\hat{u} = (p_g - p_q)^2, \quad (1.52)$$

with the squared virtual photon mass $q_\gamma = M^2$. The squared matrix-elements at parton level can be written as [75]

$$|\mathcal{M}^{q\bar{q} \rightarrow \gamma^*g}|^2 = 16\pi\alpha_S\alpha \frac{8}{9} \left[\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}} + \frac{2M^2\hat{s}}{\hat{u}\hat{t}} \right], \quad (1.53)$$

$$|\mathcal{M}^{qg \rightarrow \gamma^*q}|^2 = 16\pi\alpha_S\alpha \frac{1}{3} \left[-\frac{\hat{t}}{\hat{s}} - \frac{\hat{s}}{\hat{t}} + \frac{2M^2\hat{u}}{\hat{s}\hat{t}} \right]. \quad (1.54)$$

Using the partonic variables $z = M^2/\hat{s}$, $\hat{s} + \hat{t} + \hat{u} = M^2$, and the relation

$$\frac{1}{\hat{t}\hat{u}} = -\frac{1}{\hat{s}(1-z)} \left(\frac{1}{\hat{t}} + \frac{1}{\hat{u}} \right), \quad (1.55)$$

the matrix-element for the annihilation process yields

$$\begin{aligned} |\mathcal{M}^{q\bar{q} \rightarrow \gamma^*g}|^2 &= 16\pi\alpha_S\alpha \frac{8}{9} \left[\frac{\hat{u}}{\hat{t}} + \frac{\hat{t}}{\hat{u}} + \frac{2M^2\hat{s}}{\hat{u}\hat{t}} \right] \\ &= 16\pi\alpha_S\alpha \frac{8}{9} \frac{1}{\hat{t}\hat{u}} [\hat{u}^2 + \hat{t}^2 + 2M^2\hat{s}] \\ &= 16\pi\alpha_S\alpha \frac{8}{9} \frac{1}{\hat{t}\hat{u}} [(\hat{u} + \hat{t})^2 - (M^2 - \hat{s})^2 - 2\hat{t}\hat{u} + \hat{s}^2 + M^4] \\ &= 16\pi\alpha_S\alpha \frac{8}{9} \left[(\hat{s}^2 + M^4) \left(-\frac{1}{\hat{s}(1-z)} \right) \left(\frac{1}{\hat{t}} + \frac{1}{\hat{u}} \right) - 2 \right] \\ &= 16\pi\alpha_S\alpha \frac{8}{9} \left[(\hat{s}^2 + \hat{s}^2 z^2) \left(-\frac{1}{\hat{s}(1-z)} \right) \left(\frac{1}{\hat{t}} + \frac{1}{\hat{u}} \right) - 2 \right] \\ &= 16\pi\alpha_S\alpha \frac{8}{9} \left[\left(\frac{1+z^2}{1-z} \right) \left(-\frac{\hat{s}}{\hat{t}} - \frac{\hat{s}}{\hat{u}} \right) - 2 \right] \\ &= 16\pi\alpha_S\alpha \frac{2}{3} \left[P_{qq}(z) \left(-\frac{\hat{s}}{\hat{t}} - \frac{\hat{s}}{\hat{u}} \right) - \frac{8}{3} \right]. \end{aligned} \quad (1.56)$$

The matrix-element depends on the splitting function $P_{qq}(z)$. A similar calculation can be done for the scattering amplitude $qg \rightarrow \gamma^*q$, where the matrix-element depends on the splitting function $P_{qg}(z)$. The poles in the $2 \rightarrow 2$ matrix-elements at $\hat{t} = 0$ and $\hat{u} = 0$ introduce a divergence at $p_T^{\mu\mu} = 0$ in the perturbative calculation. The leading contribution at small p_T is caused by the soft and collinear gluon emissions from $q\bar{q} \rightarrow \gamma^*g$. The p_T of the emitted gluon can be written as $p_T^2 = \frac{\hat{u}\hat{t}}{\hat{s}}$, which

³The analogous definition yields for the Compton sub-process $qg \rightarrow \gamma^*q$ [75].

leads to a divergence in $1/p_T^2$ represented by the last term in eq. (1.53) [21]. In order to obtain the perturbative cross section one has to convolute the PDFs with the matrix-elements and integrate over the defined phase space. The integration over the gluon rapidity yields a logarithmic divergence $\ln \frac{M^2}{p_T^2}$. After some calculation the differential cross section in p_T including real gluon correction can be expressed as [75]

$$\frac{d\sigma^R}{dp_T^2} \propto \alpha_S \frac{\ln \frac{M^2}{p_T^2}}{p_T^2}. \quad (1.57)$$

The virtual corrections contribute only at small $p_T = 0$ in the form of a δ -function [8]

$$\frac{d\sigma^V}{dp_T^2} \propto \delta(p_T^2). \quad (1.58)$$

In order to regularise the divergence at small p_T a cut-off $p_{T\min}$ can be introduced in the integration, which leads to a finite result proportional to $\alpha_S \ln^2 \left(\frac{M^2}{p_{T\min}^2} \right)$. Thus, the perturbative expansion of the differential cross section can be interpreted as an expansion in $\alpha_S^n \ln^{2n-1} \left(\frac{M^2}{p_T^2} \right)$, which means for each additional gluon emission the cross section yields an additional power in α_S and additional logarithms [21].

However, the dimuon $p_T^{\mu\mu}$ distribution comprises contributions from perturbative as well as nonperturbative contributions at small $p_T^{\mu\mu}$. The intrinsic transverse momentum k_t of the partons inside the colliding protons can not be neglected at small $p_T^{\mu\mu}$ providing a crucial impact on the cross section calculation. A purely perturbative expansion of the cross section is only sufficient if $p_T \gg \langle k_t \rangle$. On the other hand, if $p_T \ll M$, the emissions of multiple soft gluons becomes important and the leading contributions in the perturbative series can not be neglected [8].

The differential cross section representing the leading contributions of multiple soft-gluon emissions obeys the perturbative expansion in α_S [8]

$$\frac{1}{\sigma} \frac{d\sigma}{dp_T^2} \sim \frac{1}{p_T^2} \left(A_1 \alpha_S \ln \frac{M^2}{p_T^2} + A_2 \alpha_S^2 \ln^3 \frac{M^2}{p_T^2} + \dots \right), \quad (1.59)$$

where A_i are calculable coefficients. The perturbative expansion in α_S is only reliable if the scales are of the same order, but higher-order terms are non-negligible when $p_T \ll M$, which means

$$\alpha_S \ln^2 \frac{M^2}{p_T^2} \gtrsim 1. \quad (1.60)$$

In this case, a fixed-order perturbative expansion is not appropriate to describe the small p_T behaviour. A solution to describe the region of soft-gluon emissions at low p_T is given by the resummation approach. The contributions from the leading logarithms in eq. (1.59) can be resummed to all orders in perturbation theory and yield [8, 76]

$$\begin{aligned} \frac{1}{\sigma} \frac{d\sigma}{dp_T^2} &\sim \frac{d}{dp_T^2} \exp \left(-\frac{\alpha_S}{2\pi} C_F \ln^2 \left(\frac{M^2}{p_T^2} \right) \right) \\ &= \frac{\alpha_S}{\pi} C_F \frac{\ln \frac{M^2}{p_T^2}}{p_T^2} \exp \left(-\frac{\alpha_S}{2\pi} C_F \ln^2 \frac{M^2}{p_T^2} \right) \end{aligned} \quad (1.61)$$

and vanishes at $p_T = 0$. The colour constant C_F is $\frac{4}{3}$. The large logarithmic contributions can be handled by resummation to all orders, which leads to the *Sudakov suppression* in form of an exponential function in eq. (1.61).

1.4.4 Resummation

In general, the inclusive perturbative partonic cross section includes contributions, which correspond to soft-parton emissions. These contributions are enhanced when $z \rightarrow 1$, where $z = \frac{M^2}{\hat{s}}$. The soft-gluon resummation is also called threshold resummation because when $z \rightarrow 1$, the partonic energy only produces the final state. For illustration a quark line emitting n gluons is considered, as pictured in Figure 1.11. The emitted gluon i carries a momentum fraction $1 - z_i$ of the initial quark. Thus, the energy of the quark decreases continuously by the radiation of gluons. When the emitted gluon becomes soft, an additional term in the inclusive cross section occurs, which is enhanced at $z \rightarrow 1$. Integrated over the whole gluon space a series of terms is obtained [77] as

$$\frac{\ln^k(1-z)}{1-z} \quad 0 \leq k \leq 2n-1. \quad (1.62)$$

These leading-log contributions introduce singularities at $z \rightarrow 1$ and have to be handled by regulating the divergences by the plus prescription

$$\alpha_S^n \left[\frac{\ln^k(1-z)}{1-z} \right]_+ \quad 0 \leq k \leq 2n-1. \quad (1.63)$$

But, if the momentum fraction z_i of the emitted gluon is large enough that

$$\alpha_S \ln^2(1-z) \gtrsim 1 \quad (1.64)$$

the logarithmic contributions are all of the same order and any fixed-order calculation would not give a reasonable result. Therefore, the partonic cross section has to be resummed to all orders in perturbation theory. In terms of resummation, the dominant contributions in the perturbative expansion are singled out and resummed by the use of an evolution equation [21].

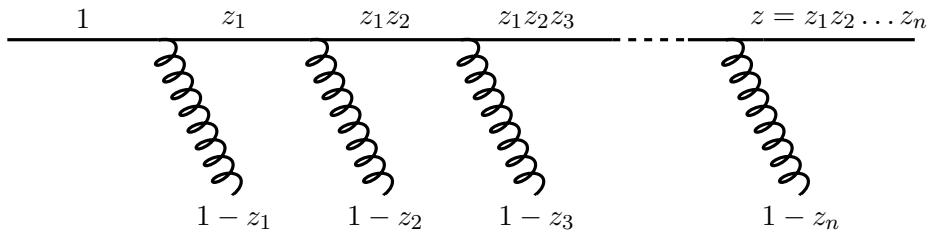


Figure 1.11: A quark line emitting n gluons. The momentum of the quark decreases by emitting gluons with a momentum fraction z_i of the quark.

In the case of the transverse momentum distribution, the large logarithms appear as (cf. eq. (1.59))

$$\alpha_S^n \ln^{2n-1} \frac{M^2}{p_T^2}, \quad (1.65)$$

and the resummed differential cross section can be expressed with the help of the Sudakov suppression in eq. (1.61). However, that the differential cross section in eq. (1.61) vanishes at $p_T = 0$ does not imply that the multi-gluon emissions are soft. The approach of the exponential suppression function assumes that only soft gluons are produced when the p_T of the Z^0/γ^* bosons is close to zero, but actually their vector transverse momentum sum should be small. Thus, in a more complete calculation the multiple gluon emissions with small vector p_T sum describe the rise of the cross section at low p_T [21]. A full calculation of the p_T resummation is provided by the Collins Soper Sterman (CSS) formalism [78, 79]. In the CSS formalism contributions from all orders in α_S are resummed providing a finite result at small p_T . In a more general way, the soft-gluon resummation can be investigated in the TMD factorisation [80, 81]. In this approach the p_T dependence of the PDFs of the initial partons is taken into account in the calculation.

Another approach to study the effect of the soft-gluon effects is the parton shower implementation in MC simulations. In comparison to the analytical calculations of all-order resummation, the parton shower is implemented numerically in MC generators. The parton shower algorithm is based on describing the soft-gluon emissions by using evolution equations. They take into account the leading-log singularities and restrict the phase space to an ordered parton cascade. The initial-state parton cascade is described by the DGLAP evolution, eq. (1.24), and the solution can be written with the help of a *Sudakov form factor* [21]

$$\Delta(t) = \exp \left(- \int_{t_0}^t \frac{dt'}{t'} \int \frac{dz}{z} \frac{\alpha_S}{2\pi} P(z) \frac{f(x/z, t')}{f(x, t)} \right). \quad (1.66)$$

The Sudakov form factor can be interpreted as the probability for a parton to evolve from the hard scale t to a softer scale t_0 without emitting a parton harder than a certain threshold. $P(z)$ is given by the splitting function for the corresponding branching and the PDF ratio gives a weight to the branching according to the PDFs at the different scales. In the model of parton shower the soft-gluon emissions are generated successively along the evolution variable t according to the Sudakov form factor, with kinematic values like the momentum fraction z and azimuthal angle ϕ [21]. The Sudakov form factor resums all soft-gluon emissions, but the implementation of large energy and wide angle emissions is not described.

However, it is difficult to obtain information on the soft-gluon emission by investigating the low p_T region. In practice, nonperturbative effects contribute to this region and yield a composition of nonperturbative as well as perturbative calculations.

In order to open up the phase space for soft-gluon resummation, I analyse the Drell-Yan process including additional jets in the production. When requiring additional jets above a certain p_T threshold, in association with the Drell-Yan dimuon pair, the leading-log contributions can be interpreted as multi-jet emissions. Thus, perturbative all-order resummation can be probed by well-defined multi-jets.

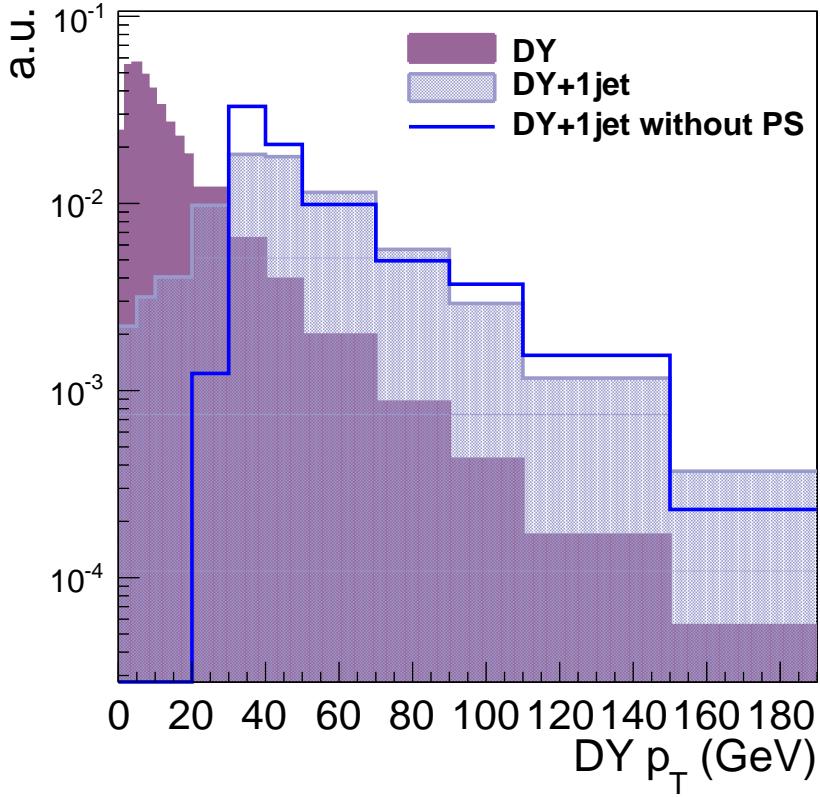


Figure 1.12: Effect of the parton shower (PS) in the Drell-Yan (DY) p_T distribution simulated with PYTHIA6.

In Figure 1.12 I compare the inclusive Drell-Yan p_T distribution to the p_T distribution of a Drell-Yan lepton pair with at least one jet above a certain p_T threshold. It is observed that the maximum of the distribution is shifted towards higher values in p_T , allowing to study the low p_T region in more detail. The distributions are produced by PYTHIA6 for the hard matrix-element and the parton shower. When switching off the parton shower in the generation, I observe that the cross section is dropping to zero just below the p_T peak. Thus, the region dominated by the soft-gluon resummation is described by the parton shower in the MC generator. At the bottom line, the numerical parton shower calculation and the analytical resummation are complementary in terms of soft-gluon dynamics.

1.4.5 Jets in Drell-Yan Events

The NLO and higher-order calculations of the Drell-Yan cross section provide a framework to compare to the measurement. The additional emitted gluons and quarks from the initial-state radiation or MPI are, however, not directly measurable

in the detector. The coloured partons in the final state form colour-neutral particles and combine into hadrons. A bunch of such collimated hadrons is then called a jet. Thus, the QCD emissions, treated by higher-order corrections, in the perturbative calculation correspond to jets in the final state. At higher centre-of-mass energies, the QCD corrections become more and more important, resulting in a large number of jets in addition to the Drell-Yan lepton pair. However, it is an exerteive and long calculation to include higher-order α_S contributions into the perturbative series. Thus, a full order perturbative calculation of final states would be too complicated. A successful approach to treat the higher-order emissions in the perturbative calculations is the parton shower model, including the emissions in the evolution equations in LLA. In this way, the parton shower exhibits an approximation with simplified dynamics in order to obtain a complete description of the multi-jet events.

In the case of MADGRAPH the hard matrix-element is calculated at LO. The fixed-order perturbative calculation includes at most four final-state partons in addition to the Drell-Yan lepton pair. This yields a maximum number of four jets from the matrix-element calculation. However, in order to obtain an optimal description of the final-state dynamics the soft-parton emissions have to be added to the matrix-element calculation. Thus, additional jets can be produced by the parton shower algorithm implemented in PYTHIA6.

2 The Compact Muon Solenoid at the LHC

The Compact Muon Solenoid (CMS) is one of the main particle detectors at the LHC. In this chapter the main components of the detector, which are relevant for the measurement of the final objects (muons and jets), are explained. First, a brief introduction to the LHC is given. Second, the sub-detectors of CMS are presented from the inner to the outer parts. Furthermore, the trigger system of CMS is presented. The trigger paths, which are used in this analysis and the corresponding trigger efficiencies are studied. At the end, the detector simulation is discussed. The information in this chapter is taken from [82–84] if not otherwise mentioned. For a more detailed explanation, I refer the reader to the latter references.

2.1 The LHC Accelerator

The Large Hadron Collider (LHC) [85] is the world’s largest and highest-energy particle accelerator. This giant machine is located at CERN, the well-known scientific research centre, near Geneva. The scientists working at the LHC try to provide answers to the fundamental questions in nature. With the very high energy provided by the LHC it is possible to study cross sections of benchmark processes at an energy scale never investigated before. The purpose of the LHC experiments is to test the theory of the SM, but also to find new physics.

The LHC was designed to accelerate protons as well as heavy ions. The storage ring has a circumference of 27 km and is placed underground. Two high-energy beams are accelerated in opposite direction in two separate beam pipes. Each beam contains bunches of hadrons. At the design luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ there are 2808 proton bunches with $1.15 \cdot 10^{11}$ protons inside each bunch, which are separated by a distance of 25 ns. After the two long shutdowns of the LHC, in which the accelerator is upgraded, it reaches its design proton-proton centre-of-mass energy of $\sqrt{s} = 14 \text{ TeV}$. Running with its design parameters, the LHC is able to produce on average 20 collisions per bunch crossing and 800 million collisions per second.

Superconducting magnets provide a strong magnetic field of 8 Tesla to keep the particles along the circuit. There are 1232 dipole magnets to bend the beam and 392 quadrupole magnets to focus the beam. In order to ensure the best performance with an excellent efficiency of the superconducting magnets a cooling, using liquid helium, to -271.3°C is required.

The proton beams are made to collide at four positions along the ring, where the main experiments are located. The particle detectors ATLAS, CMS, LHCb and ALICE are built to study different physics topics. ATLAS and CMS are so-called multi-purpose detectors, and are devoted to search for new particles and physics

theories beyond the SM. LHCb is designed to study B physics and to address the question why there is more matter than anti-matter in the universe, by performing a precise measurement of CP-violation. ALICE is an experiment to investigate the quark-gluon plasma, produced in lead-lead collisions, and to draw conclusions on the structure of the universe in the first 10^{-10} seconds.

2.1.1 Luminosity

The meaning of luminosity in collider physics is to quantify the performance and the intensity of the collisions. The luminosity \mathcal{L} [86] defines the possibility to produce a certain number of events N of a process with cross section σ in a certain time interval dt

$$\frac{dN}{dt} = \mathcal{L} \cdot \sigma. \quad (2.1)$$

In the case of two colliding beams and assuming Gaussian profiles in all dimensions with variance $\sigma_{x,y}$, the luminosity yields

$$\mathcal{L} = \frac{N_1 N_2 f N_b}{4\pi \sigma_x \sigma_y}, \quad (2.2)$$

with the number of particles N_i per bunch i , and the number of bunches per beam N_b . The revolution frequency is labelled as f . Thus, the luminosity is totally dependent on experimental collider parameters and characterises the performance of the accelerator. In general, the accelerator performance is arranged, such to optimise the luminosity, yielding a larger amount of data to be analysed. The continuous increase of the luminosity delivered by LHC and recorded by CMS in 2011 is shown in Figure 2.1. The time integrated luminosity yields

$$\mathcal{L}_{\text{int}} = \int dt \mathcal{L}, \quad (2.3)$$

and is needed to calculate the cross section of the measured process.

The luminosity can be measured in two ways: online and offline. The online measurement uses empty hits in the towers of the hadronic forward calorimeter, and the offline method is based on the production rate of the primary vertices. With these two measurements the average number of events is obtained. However, to obtain the absolute luminosity an absolute normalisation is needed to account for the size of the beam. The absolute calibration scale is determined from van der Meer scans [88], in which the transverse separation of the two beams is scanned relative to each other. This method yields a determination of the size of the beams at their interaction point.

The uncertainty on the luminosity measurement can lead to a relevant systematic uncertainty on the cross section measurement. Thus, standard well-measured processes are used to cancel the dependence on the luminosity uncertainty. In general, normalised distributions are determined, using as a normalisation factor the cross section of a standard candle. In this analysis the final cross sections are normalised using the $Z \rightarrow \mu\mu$ cross section defined in the Z mass range of 60 – 120 GeV.

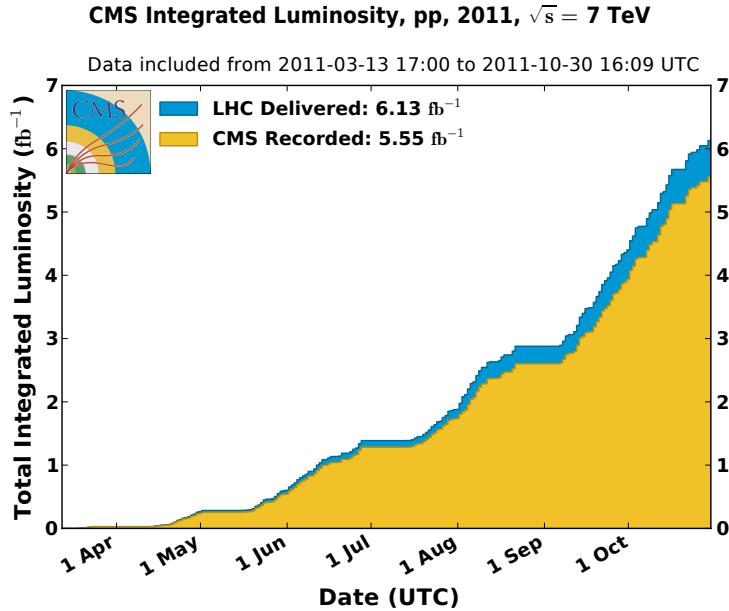


Figure 2.1: Total integrated luminosity in 2011 [87].

2.1.2 Data Taking Period 2011

Since late 2009 the LHC is running and delivers great insight into particle physics. In 2010 the first proton-proton collisions with centre-of-mass energy of 7 TeV took place. A large variety of physics benchmark processes are summarised in Ref. [83]. In 2010 the commissioning and the validation of the machine's performance were

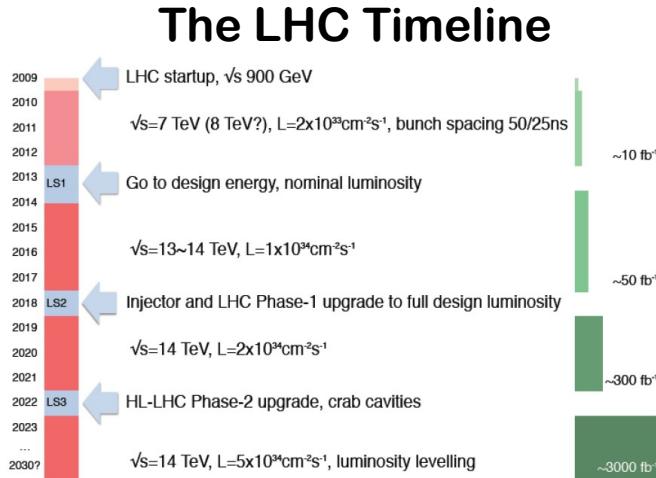


Figure 2.2: LHC timeline [89]. The development of the LHC integrated luminosity and collision energies of the past and the future are presented.

studied. The intensity of protons was increased in steps, each step followed by a few days of system performance checks.

The CMS data taking is performed when beams are stable and is divided into so-called runs, which characterise periods with constant detector conditions. Each run is subdivided into lumi sections, which are defined by the time a proton needs for 2^{18} LHC orbits [90], which corresponds to 23.31 seconds. The 2010 data taking ended with beams of 368 bunches of $\approx 2.1 \times 10^{11}$ protons per bunch, and a peak luminosity of $2.1 \times 10^{32} \text{ cm}^{-2}\text{s}^{-1}$ resulting in a total integrated luminosity of 0.04 fb^{-1} [91]. In Figure 2.2 the LHC timeline shows the chronological energy development of the LHC, including the achieved luminosity but also the expectation for the future. At the moment, the LHC resides in the long shutdown 1 (LS1) preparing for the design centre-of-mass energy of 14 TeV. During 2011 data taking the beam energy resumed at 3.5 TeV, but the number of bunches was increased by running with 50 ns. A total integrated luminosity of 6 fb^{-1} was delivered, out of which CMS recorded 5.6 fb^{-1} (cf. Figure 2.1).

The data events analysed in this thesis were produced with a centre-of-mass energy of 7 TeV and have been recorded in 2011. The corresponding integrated luminosity was 4.9 fb^{-1} . Due to the advanced and well-understood performance of the CMS detector, first years of data taking gave an insight in previously discovered physics processes. The large amount of data recorded with CMS at $\sqrt{s} = 7 \text{ TeV}$ with $\mathcal{L} = 4.9 \text{ fb}^{-1}$ enables scientists to measure known processes with unprecedented precision. In Figure 2.3 I present the measured dimuon invariant mass spectrum, which reveals all known resonances, spanning a range over three orders of magnitude.

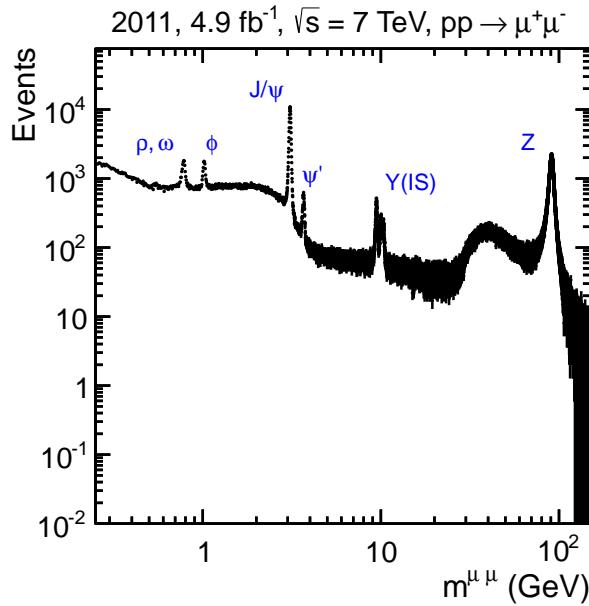


Figure 2.3: 2011 data recorded with CMS at a centre-of-mass energy of 7 TeV in proton-proton collisions. The events are triggered by double and single muon trigger. The dimuon invariant mass spectrum reveals the resonances of mesons and the Z boson.

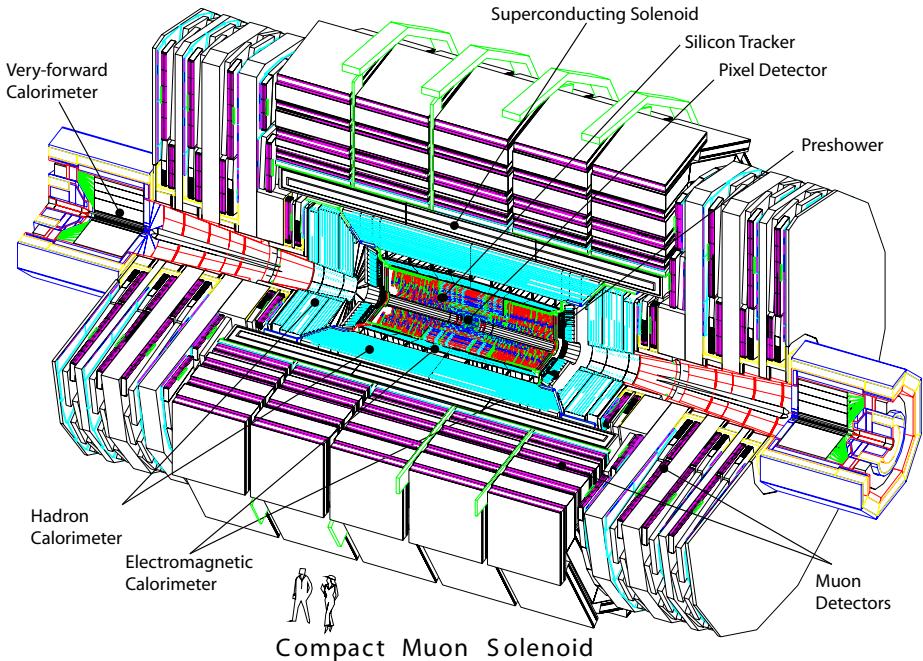


Figure 2.4: The CMS detector and its subcomponents, relevant for the particle identification. The several layers are build around the beam pipe resulting in a cylindrical form of the detector [82].

2.2 The CMS Detector

CMS is one of the main experiments at the LHC with the purpose to study particle physics at the TeV scale, at which an insight to low cross section processes is provided. At such energies the physics processes yield a dominant contribution from QCD background events. One of the main challenges of CMS is, thus, to provide an excellent particle identification and momentum resolution in order to perform a precise measurement of benchmark cross sections. The reconstruction of lepton signatures and jet energy resolution is from great importance in this study. A further challenge is the identification of the particle's origin, the real interaction point of the collisions, i.e. the primary vertex, and the decay vertex. Thus, CMS is designed to have a good time resolution and a high-granularity, with many detector readout channels and an excellent synchronisation among them.

The construction of the detector is arranged in several layers. Each layer has a different function in order to gain information of the produced particles. Combining the findings of each part results in particle identification. The geometric specifications of the detector and its subcomponents are listed in Table 2.1. The subcomponents are arranged around the beam-pipe resulting in a cylindrical form of the detector. CMS comprises a barrel and two endcaps at the front and back of the detector. In Figure 2.4 a schematic overview of the detector and the several layers is presented.

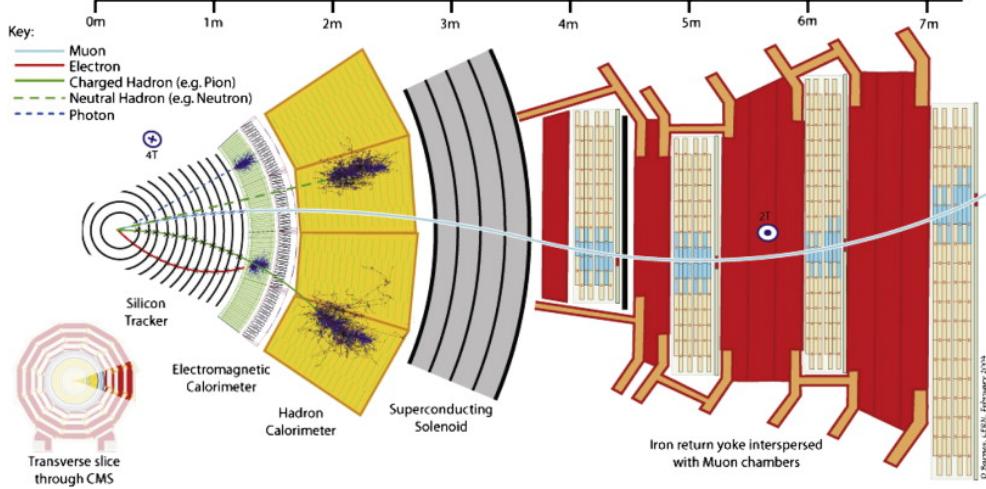


Figure 2.5: A transverse slice through CMS. The trajectories of different particles leave a signature in the parts of the detector providing additional information [92].

	Weight (in t)	Length (in m)	Radius (in m)	
CMS	12500	21.6	7.5	
		Coverage in $ \eta $	Radius/Distance (in m)	Purpose
Solenoidal Magnet	4 layers	-	2.95 - 3.25	bending of charged particle trajectories
Tracker	pixel strip	< 2.4 < 2.4	0.05 - 0.11 0.20 - 1.16	measurement of charge and momentum
ECAL	barrel endcap	< 1.479 1.479 – 3.0	1.24 - 1.86 3.2	energy measurement of electrons and photons
HCAL	barrel endcap forward	< 1.4 1.3 – 3.0 3.0 – 5.0	1.77 - 2.95 3.9 11.2	energy measurement of hadrons
Muon System	barrel endcap	< 1.2 0.9 – 2.4	3.8 - 7.38 5.0	identification and measurement of muons

Table 2.1: Geometric information about CMS and its components. The individual detector parts are listed with their purposes. The coverage in pseudorapidity is presented as well as the inner and outer radii $r_i - r_o$ of the barrel parts and the distance z to the interaction point from the endcap.

From the inner to the outer components the barrel is composed of a tracking system followed by electromagnetic and hadronic calorimeters. The magnetic coil is built as a solenoid around the tracker and calorimeter. The outer part of the barrel and the endcaps compose the muon system embedded in an iron yoke. In Figure 2.5 a slice of CMS in the transverse plane to the beam axis is shown. The trajectories of different particles, and how they yield a signature in the detector components are presented.

The Cartesian coordinate system of CMS is defined as a right-handed coordinate system with the origin defined at the collision point: the x axis points to the centre of the LHC ring, the y axis points vertically upwards and the z axis in the direction of the anti-clockwise beam. The corresponding cylindrical coordinates are defined by the azimuthal angle ϕ , measured in the xy plane starting from x , and the polar angle θ , measured in the rz plane starting from z .

The pseudorapidity η is defined via the polar angle θ as

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

The plane transverse to the beam axis is referred to as the transverse plane and variables measured in this plane are labelled by subscript T. Hence, the transverse momentum is defined as $p_T = \sqrt{p_x^2 + p_y^2}$.

The different subcomponents of CMS, which are relevant for the detection of muons and jets are discussed in the following.

2.2.1 Magnet

Charged particles emerging from the interaction point are affected by the strong magnetic field of the superconducting coil. The magnetic field of the solenoid bends the particle according to its charge in a circular path. The coil is surrounded by the muon system embedded in an iron yoke, which returns the magnetic flux. Inside the coil the tracking system is situated to measure the trajectories of the charged particles. The measurement of the curvature results in the determination of the momentum of the particle. In order to achieve the required momentum resolution for high momentum muons, which is of $\Delta p/p \approx 10\%$ at $p = 1$ TeV, the magnet system has to adapt the design of the detector. Thus, the magnet has the form of a solenoid composed of four-layers windings, which induce an axial magnetic field of 3.8 T.

2.2.2 Inner Tracking System

The inner tracking system has the purpose of revealing the trajectories of charged particles, by measuring their tracks. From the track information the efficient and precise reconstruction of charge, position and momentum of the particle can be performed. Moreover, the track information is used to reconstruct the primary and

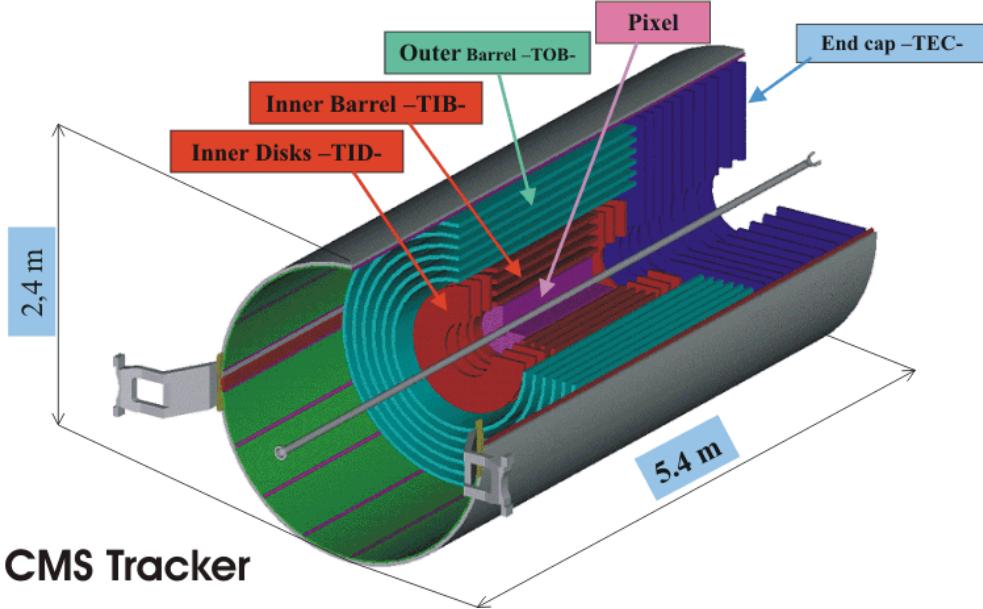


Figure 2.6: The CMS tracking system [92]. The tracker is composed of an inner pixel and an outer silicon strip detector. The combination of six TOB, four TIB and on each endcaps three TID and nine TEC in the silicon trip detector ensures a track covering up to $|\eta| < 2.4$.

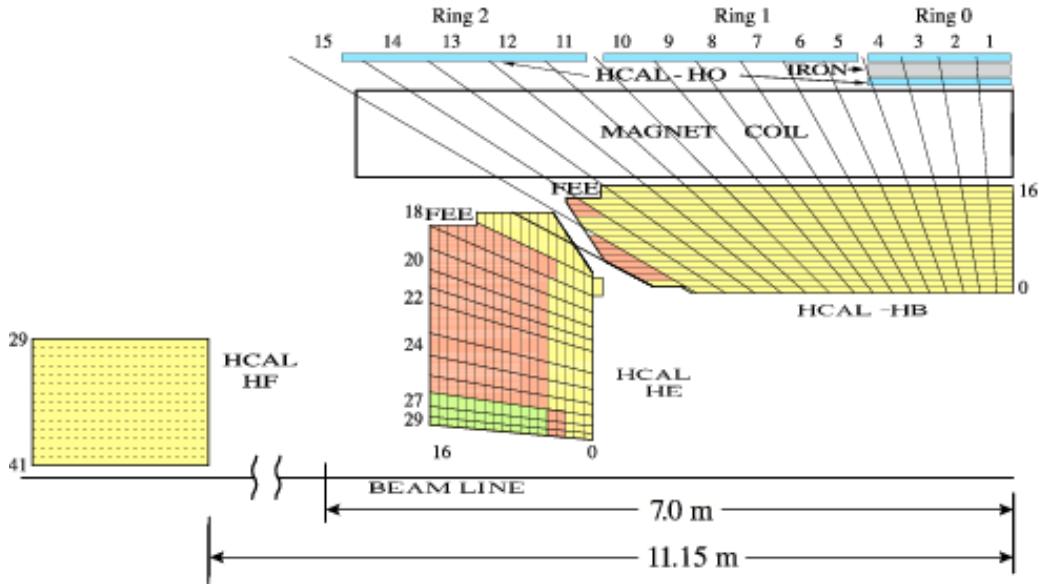


Figure 2.7: A quarter slice of the CMS hadronic calorimeter (HCAL) [93]. The different components, hadronic barrel (HB), hadronic endcap (HE), hadronic outer (HO), and hadronic forward (HF), are arranged around the solenoid.

secondary vertices of particles as well as the impact parameter. The primary vertex is the real interaction point¹, the position where the particles collide. The information on the primary vertex is the basis of the reconstruction of the beam spot, the point, at which the proton bunches overlap. The secondary vertex information is needed to identify heavy quarks and τ leptons. The impact parameter is defined as the transverse distance from the reconstructed track to the interaction point and is needed to discriminate cosmic muons.

The inner tracking system comprises a silicon pixel and a silicon strip detector. The structure of the tracking system is shown in Figure 2.6. Closest to the interaction point the particle flux is the highest and pixel detectors are used. In the intermediate and outermost distance to the collision the track detection is performed by silicon micro strip detectors.

The tracker has the form of a barrel and two endcaps. In the forward region two pixel and nine micro strip layers are arranged in each of the two endcaps. In order to measure an optimal vertex resolution, the pixel layout uses an almost squared shape of $100 \times 150 \mu\text{m}^2$. In the barrel, three layers with 760 pixel modules are arranged as half-ladders. The two endcap disks are assembled in a turbine-like geometry with 20° -rotated blades comprising 672 pixel modules. The silicon strip detector comprises ten layers in the barrel and three inner and nine outer disks in each of the two endcaps, providing a covered range of $|\eta| < 2.4$. The Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB) provide a measurement in both $r - \phi$ and $r - z$ coordinates. The inner silicon strips have size of $10 \text{ cm} \times 80 \mu\text{m}$ with an occupancy of 2–3%, and the outer silicon strips have a size of $25 \text{ cm} \times 180 \mu\text{m}$ with an occupancy of about 1%. The inner tracker is shorter than the outer tracker in order to avoid excessively shallow track crossing angles. In the transition region between barrel and endcaps three additional inner disks, Tracker Inner Disk (TID), are situated. The TID and the Tracker End Caps (TEC) are arranged as rings in the endcaps. In total the entire silicon strip detector consists of 15400 modules with an active silicon area of 200 m^2 .

2.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) measures the energy deposit of electrons and photons passing through the detector. The calorimeter is made of lead tungsten (PbWO_4) crystals, which yield a fast response and fine granularity. The electrons and photons interact with the material and initiate electromagnetic showers by bremsstrahlung, photo-effect, Compton effect or pair production. The lead tungsten crystals are tilted by 3° to optimise the readout of the electromagnetic showers. The barrel consists of 61200 crystals covering a range in pseudorapidity up to $|\eta| < 1.479$. The electric signal readout is performed by silicon avalanche photodiodes. In the region of the endcaps 7324 crystals are arranged and measure the range of $1.479 < |\eta| < 3.0$. The signal is amplified and read out by vacuum photo-triodes.

¹One has to distinguish the real interaction point from the nominal interaction point. The latter is defined as the design position of the collision.

2.2.4 Hadronic Calorimeter

The hadronic calorimeter (HCAL) measures the energy deposit of charged and neutral hadrons. It comprises several subcomponents: Hadronic Barrel (HB), Hadronic Endcap (HE), Hadronic Outer (HO) and Hadronic Forward (HF). The design of the HCAL is such, that some parts of the detector are embedded in the magnetic coil, which surrounds the HCAL. A longitudinal slice of the HCAL is shown in Figure 2.7. The HB and HE are made of layers of brass absorbers alternated with plastic scintillators. The brass plates absorb the hadrons through their repeated scatterings with the nuclei, and inducing an hadronic shower. The active material is made of scintillator material. The emitted scintillator light is measured with photo-detectors. The HB covers a range of $|\eta| < 1.4$ and HE extends up to $1.3 < |\eta| < 3.0$. The readout modules are arranged in towers with size $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for HB and HE respectively. Due to space constraints within the solenoid the thickness of HB is limited. In order to catch the leakage from HB, additional scintillators in HO are positioned outside the magnetic coil. HO has the structure of five rings, similar to the iron yoke and the muon system. HB, HE, and HO are constructed in azimuthal wedges of $\Delta\phi = 20^\circ$. In the forward direction the radiation hard component HF is located, which is made of iron absorbers and quartz fibres. The HF provides a cover of $3.0 < |\eta| < 5.0$. The HF detector components play an important role when requiring jets in the forward region. Furthermore, the detection of hadrons over a wide range in η is necessary for measuring the imbalance in the transverse plane, which is an indirect measurement of neutrinos and other particles that do not interact with the detector material.

2.2.5 Muon System

The muon system consists of muon chambers, also called stations, with the purpose of identifying and measuring minimum-ionising muons. The choice of the gaseous tracking detector depends on the geometry. In the central region ($|\eta| < 1.2$), in the barrel, the muon rate and the residual magnetic field is low. Here, drift tube (DT) chambers are employed as tracking detectors. In the muon barrel (MB) four stations of cylindric detectors altered with the iron yoke are assembled, MB1 - MB4, and five wheels are arranged along the beam. In the endcap regions ($0.9 < |\eta| < 2.4$), the muon rate and the magnetic field is high and cathode strip chambers (CSC) are installed. The design of the CSCs provide precise space and time information as well as a fast response, which is important for a high magnetic field and high particle flux. Additionally, resistive plate chambers (RPC) are located in the barrel and endcaps, with the purpose of an additional positioning measurement, but they are also used for timing and triggering. In the endcaps the CSCs and the RPCs consist of four disks transverse to the beam and two (three) concentric rings in the innermost (outermost) stations. The structure of the muon chambers is shown in Figure 2.8.

The resolution of the muon momentum p measured with the muon system depends on multiple scatterings in the iron yoke and is nearly constant in p . The inner tracking

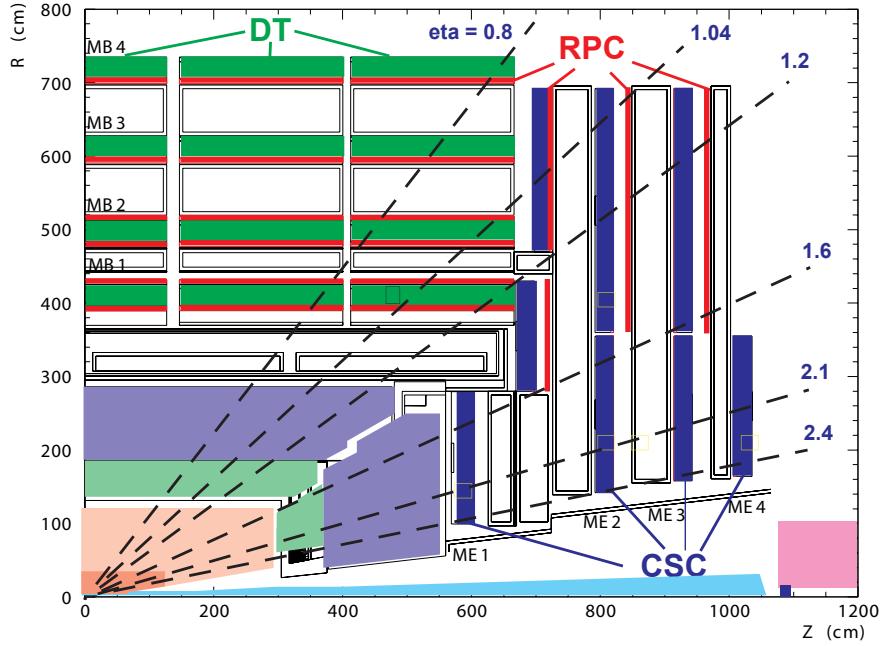


Figure 2.8: The muon system [82]. In the barrel the drift tube chambers (DT) are embedded in the iron yoke. In the endcap regions cathode strip chambers (CSC) are used for muon identification. For an optimal positioning measurement resistive plate chambers (RPC) are installed in the barrel and endcaps.

system already provides a good resolution of the muon momentum for muons with $p < 100$ GeV. For higher momentum muons the resolution gets worse, but this can be compensated by increasing the length of the measured trajectory. Therefore, by combining the tracker and the muon system reconstruction, the resolution of the muon momentum can be optimised. The momentum resolution as a function of p for the three different track reconstructions in the tracker, muon system and the combination of both is shown in Figure 2.9. The resolution is shown for two different pseudorapidity bins. The resolution is better in the central region, than in the forward region, due to the larger amount of material in the endcaps, which can lead to more multiple scatterings. The transverse momentum resolution of muons with $20 < p_T < 100$ GeV is $1.3 - 2\%$ in the barrel and 6% in the endcaps. For muons with $p_T < 1$ TeV the resolution is better than 10% in the barrel [94]. The good momentum resolution, in combination with the high muon selection efficiency and low fake rate ensures an excellent muon identification.

2.2.6 Trigger

The high collision rate of 40 MHz yields a large amount of data, which is not possible to record. Thus, the event rate has to be optimised during data taking, by a pre-

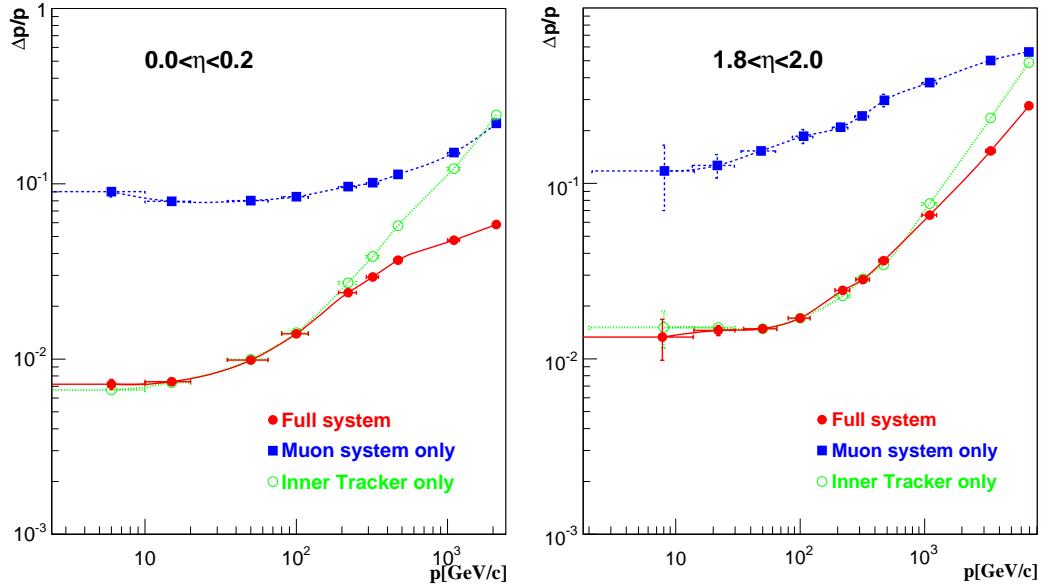


Figure 2.9: The muon momentum resolution in two different pseudorapidity regions [82]. The resolutions obtained from tracker information, muon system information, and global fits combining both are compared.

selection of events. The trigger system of CMS enables to select only the relevant, or interesting, events that are necessary for a data analysis. The trigger has the design to control the data taking. The trigger system is arranged in two steps. The Level-1 trigger (L1T) downsizes the event rate to about 10 - 100 kHz, and the events are passed to the High-Level trigger (HLT), which reduces the rate to 100 Hz. The trigger selection depends on the instantaneous luminosity and can be adjusted during runtime. Thus, triggers in runs with a high luminosity have to be prescaled, to reduce the output rate. This means the selected events are reduced to a fraction given by the prescale factor. The trigger menu consists of different trigger paths, each path corresponds to certain selection criteria.

In this analysis muon triggers are employed. The criteria of an event selection depend on the muon transverse momentum and the muon isolation. In the following the trigger system is explained in the concept of selecting muon candidates.

The trigger system uses an online trigger menu, which allows the event selection to be performed during data taking. L1T is based on detector electronics and the event information is passed through a readout network to the online event filter system, which executes the software for HLT on a processor farm. In the first step, L1T, a minimised event reconstruction is applied from calorimeter information and muon hits. Whether the event is selected or discarded is decided in $3.2 \mu\text{s}$ when the information is stored in pipeline memories. L1T has a specific architecture, which is shown in Figure 2.10. Local trigger objects are generated with local input from energy deposits in the calorimeter. Additionally, tracks are reconstructed locally in DTs and CSCs, and RPCs are used to construct the muon candidates. The local information is combined in the global muon trigger, which can also include information on isolation and signatures of minimum-ionising particles. At the end

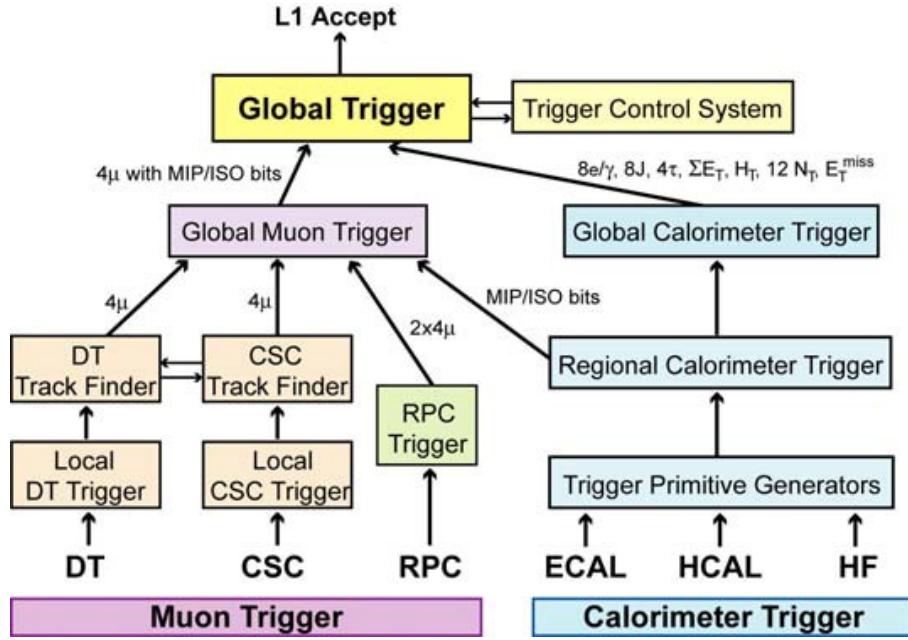


Figure 2.10: Architecture of Level-1 trigger [84].

it is decided whether the event is recorded or not. The events accepted by L1T are transferred to HLT.

The HLT trigger menu is divided in trigger paths, depending on physics objects with certain kinematic criteria. The trigger paths used in this thesis depend on the muon p_T . Thus, a fast reconstruction of physics objects is performed using all detector components in order to accept the event according to the trigger path criteria. The HLT muon reconstruction is divided into several parts [95]. First, the Level-1 (L1) reconstructed segments obtained from L1T are used to construct the *standalone muons* (L2), by performing track fits, using hits from the muon chambers. The transverse momentum of the standalone muon is checked to verify the selection requirement and the isolation is checked with respect to the calorimeter hits. In the next step (L3), for each standalone muon a tracker track is reconstructed. A fit is performed to find the optimal tracker track matching the standalone muon. This collection of objects is referred to as *global muons*. The tracker isolation is checked and if the p_T requirement is fulfilled the event is accepted and a full event reconstruction is applied. The muon reconstruction is explained in section 3.3.1.

Trigger Efficiency

In this analysis unprescaled HLT muon triggers with lowest available p_T threshold are used. In the first half of 2011 data taking a single muon trigger was used and in the second half a double muon trigger was employed. The different trigger paths, run ranges, and the integrated luminosity are listed in Table 2.2. The High-Level trigger

paths used in this analysis are a single isolated muon selection with p_T threshold of 17 GeV (HLT_IsoMu17), and a double muon trigger with p_T thresholds of 13 and 8 GeV (HLT_Mu13_Mu8) for the leading and subleading muon respectively.

The isolation requirement ensures that the muon emerges from an electroweak process. The isolation variable is defined such that no additional activity is found around the muon in a cone of $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. A detailed study on the muon isolation criterion is discussed in section 3.3.2.

The trigger event selection has an efficiency depending on the muon p_T and η . This means, the misidentification of a muon with p_T and η introduces an error in the triggering (selecting or discarding the event). The efficiency should show an increase around the p_T threshold of the muon. The trigger efficiencies, also called trigger-turn-on curves, are estimated using the tag-and-probe method applied to muons originating from Z resonances. The tag-and-probe method is based on clean dimuon signal events, like $Z \rightarrow \mu\mu$. A tight selection on one of the muons is applied, referring to as the tag muon, which ensures the muon merging from a resonance decay yielding a high purity. A looser selection is applied on the other muon referred to as probe muon. The efficiency is defined as the fraction of probe muons passing the selection criteria under study.

Trigger	Run Range	Luminosity (pb^{-1})
HLT_IsoMu17	160404-170248	1170
HLT_Mu13_Mu8	170249-180252	3723

Table 2.2: High-Level trigger paths used in this analysis. A single muon trigger with p_T threshold of 17 GeV and isolation selection, and a double muon trigger with p_T thresholds of 13 and 8 GeV is used.

The trigger efficiencies for the single muon trigger are shown in Figure 2.11 for four different pseudorapidity regions as a function of the probe muon p_T . The trigger efficiencies for the double muon trigger have to be estimated separately for the different p_T thresholds (legs) of the two muons. The results for the harder p_T leg of 13 GeV is presented in Figure 2.12 for four different pseudorapidity ranges as a function of the probe muon p_T . The trigger efficiency plots show a dependence on the muon p_T in the form of general turn-on curves, where the increase of the efficiency is around the p_T threshold. The dependence in pseudorapidity is very small.

In order to account for the trigger inefficiency each MC event is assigned a weight for the isolated muon trigger

$$\omega_{\text{trig}}(p_T, \eta) = 1 - (1 - \epsilon^+(p_T, \eta)) \cdot (1 - \epsilon^-(p_T, \eta)), \quad (2.5)$$

where ϵ^+ (ϵ^-) is the p_T and η dependent single muon trigger efficiency for the two muons, with positive and negative charge respectively. The double muon trigger has asymmetric p_T thresholds for the leading and subleading muon. Therefore the efficiency is different for the two p_T legs of the trigger. The weight for the double

muon trigger is defined as [96]

$$\omega_{\text{trig}}(p_T, \eta) = \epsilon^{\text{high},+}(p_T, \eta) \epsilon^{\text{low},-}(p_T, \eta) \\ + \epsilon^{\text{high},-}(p_T, \eta) \epsilon^{\text{low},+}(p_T, \eta) \\ - \epsilon^{\text{high},+}(p_T, \eta) \epsilon^{\text{high},-}(p_T, \eta), \quad (2.6)$$

where ϵ^{high} is the p_T and η dependent efficiency of the higher p_T leg and ϵ^{low} the efficiency of the leg with the lower p_T threshold.

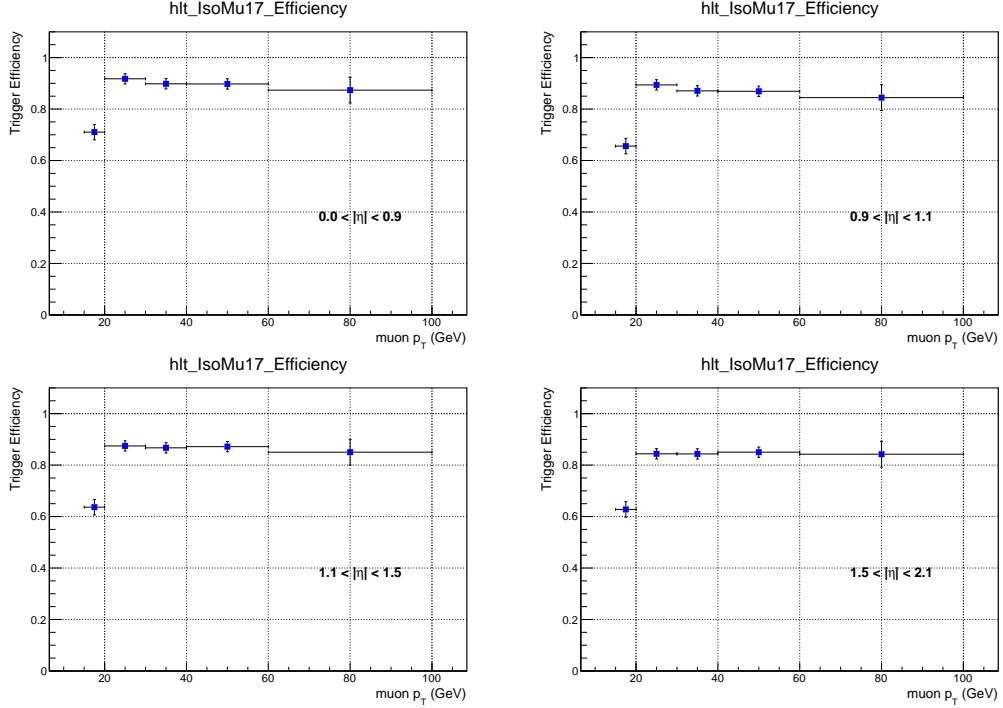


Figure 2.11: Trigger turn-on curves of HLT_IsoMu17 in four pseudorapidity regions as a function of the muon p_T . The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events [97].

In this analysis the dimuon distributions are also measured for high dimuon invariant masses (> 1 TeV). At high dimuon invariant mass mostly high p_T muons contribute. When high p_T muons pass through the iron of the return yoke, multiple scatterings and additional radiation influences the curvature of the muon. These effects can introduce electromagnetic showers, which imply additional hits in the muon chambers. Thus, the reconstructed track fits are biased and the muon momentum calculation can be affected. The standard approach to estimate the effect on the trigger efficiency of high p_T muons is to extrapolate the values from the Z peak to higher masses. This study introduces a systematic uncertainty of the order of 2–3% in the extrapolation to high p_T . In this analysis the systematic uncertainty of the total muon efficiency factor², which is around 2% for muons with $p_T < 100$ GeV, is

²The total muon efficiency is the product of the trigger efficiency, isolation, and identification efficiency. The efficiencies for the muon identification and isolation are discussed in section 3.3.4

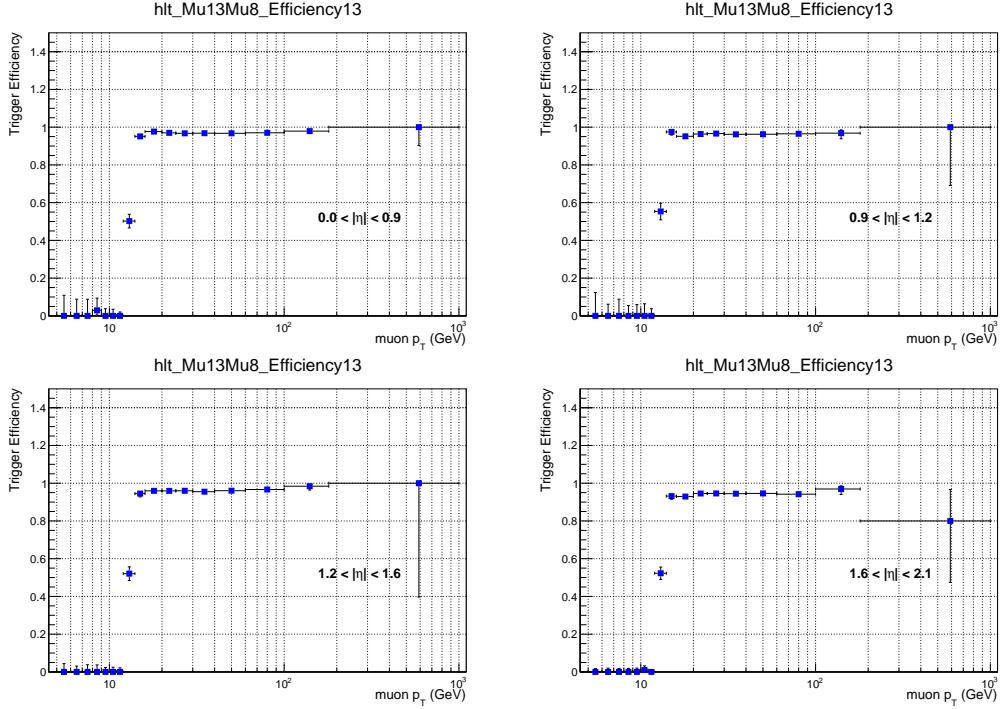


Figure 2.12: Trigger turn-on curves for the trigger leg of 13 GeV of HLT_Mu13_Mu8 in four pseudorapidity regions as a function of the muon p_T . The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events [97]. The turn-on curves for the lower leg of 8 GeV can be found in appendix A.

increased to 5% for high p_T muons [98].

2.3 Detector Simulation with GEANT4

The measured data obtained from CMS are based on interactions of particles with the material of the detector components. Thus, physics observables determined from measured data include a bias due to detector effects. In order to compare reconstructed collision events to MC predictions, the MC simulation has to be generated on the so-called *detector level*. The MC generators presented in section 1.3.1 only model physics processes on parton or generator level. Thus, the interaction of particles with the detector material has to be simulated separately. The events from the SMC event generators are passed to a detector simulation, based on GEANT4 [99, 100], which also includes a full event reconstruction. The software of GEANT4 is interfaced with the CMS software and includes a full detector geometry and performance. Additionally, the architecture of the CMS trigger system is included, providing full reconstructed events in the same form as real collision data.

3 Event Reconstruction and Object Identification

In order to perform an optimal measurement of muons and jets, the CMS detector is constructed to detect particles with high efficiency and momentum resolution, but also an excellent offline reconstruction of the final objects is needed. This chapter is devoted to the reconstruction algorithms and selection criteria of muons and jets measured with CMS. First, the primary vertex reconstruction is presented followed by the muon reconstruction algorithm used in this analysis. The muon identification selection in order to obtain a good quality muon is described, including the corresponding selection efficiency. Further, the jet reconstruction algorithm and identification are explained.

3.1 Primary Vertex Reconstruction

The vertex reconstruction defines the precise determination of the exact collision point in the scattering process. Due to the increasing instantaneous luminosity at the LHC the probability to have more than one proton-proton collision in the same bunch crossing is large. Thus, it is necessary to reconstruct not only the primary vertex position, but also the vertex multiplicity, and to gain a precise assignment of the tracks to the vertices. The offline primary vertex reconstruction is performed with the Deterministic Annealing (DA) clustering method [101]. First, the tracks are arranged in *clusters* according to the z coordinate of the track point closest to the beam line. Second, a three dimensional vertex fit is performed for each track in each cluster. The vertex fit uses the full track information from the tracking system.

To suppress the contamination from $\gamma\gamma \rightarrow \mu\mu$ events a minimum requirement on the number of degrees of freedom¹ on the primary vertex fit is set to four.

The outcome of the vertex reconstruction allows multiple possibilities of the primary vertex. Thus, all possible vertices are stored.

The requirements in this analysis on the primary vertex are:

- z position within 24 cm of the nominal detector centre
- radial position within 2 cm from the beam spot

¹The number of degrees of freedom are defined as $n_{\text{ndof}} = 2 \sum_{i=1}^{N_{\text{tracks}}} w_i - 3$ with the track weight w_i of the i -th track. The weights are defined between $0 \leq w_i \leq 1$ and the value is assigned, taking into account the compatibility with the common vertex. This means, if the track is associated with the common vertex, the weight is set to one [102].

From the set of selected primary vertices the hard interaction vertex is selected as the vertex with the maximum sum of the transverse momenta of the tracks, associated with this vertex.

In order to reproduce the number of reconstructed primary vertices in simulation, a weight factor is applied to the simulation based on the instantaneous luminosity. The "official" pileup reweighting procedure for 2011 MC datasets [103] is explained in the following.

3.1.1 Pileup Reweighting

In general, the MC samples are generated such that the conditions (e.g. number of primary vertices, luminosity, pileup interactions, etc.) match the expected conditions for a certain data taking period. However, due to the high luminosity in 2011 runs several proton-proton interactions happen simultaneously (pileup), which makes it more difficult to reproduce the same number of reconstructed primary vertices in MC generation as observed in data. In Figure 3.1(a) the number of primary vertices in data is compared to MADGRAPH MC sample. The MC simulation does not reproduce the data distribution, although the effect of pileup is included in the MC simulation. This is because, the pileup distribution depends on the beam condition, which is not known at the time the MC sample is generated. In order to reproduce the observed distributions in data by the generated MC distributions, a weight has to be assigned to the MC events. The distribution of the reconstructed primary vertices is sensitive to the implementation of the primary vertex reconstruction and the effects from the UE. Thus, the reweighting of the MC samples is not performed as a function of the number of primary vertices.

Here the MC events are reweighted by the number of pileup interactions from data $n_{\text{PUI}}^{(\text{data})}$ over the number of pileup interactions from the simulation $n_{\text{PUI}}^{(\text{MC})}$ as a function of the *MC truth* information²

$$\omega_{\text{PU}}(n_{\text{PUI}}^{(\text{truth})}) = \frac{n_{\text{PUI}}^{(\text{data})}}{n_{\text{PUI}}^{(\text{MC})}}. \quad (3.1)$$

The MC pileup distribution is available from the MC generation. However, the pileup distribution in data has to be estimated taking into account the luminosity information [103]. The instantaneous luminosity, $\mathcal{L}_{\text{bx},\text{ls}}$, per bunch crossing (bx) per lumi section (ls) is proportional to the pileup rate. Multiplying $\mathcal{L}_{\text{bx},\text{ls}}$ by the inelastic proton-proton cross section and dividing by the circulation rate f , yields the number of expected pileup events per bunch crossing and lumi section

$$N^{\text{PU}} = \frac{\mathcal{L}_{\text{bx},\text{ls}} \cdot \sigma_{\text{inel.}}}{f}. \quad (3.2)$$

The number of pileup events for each bunch crossing can be described by a Poisson distribution with mean value defined as the average number of interactions [104]. The final pileup distribution in data is obtained by generating a Poisson distribution

²MC truth variables corresponds to variables defined on generator level and are independent of the detector.

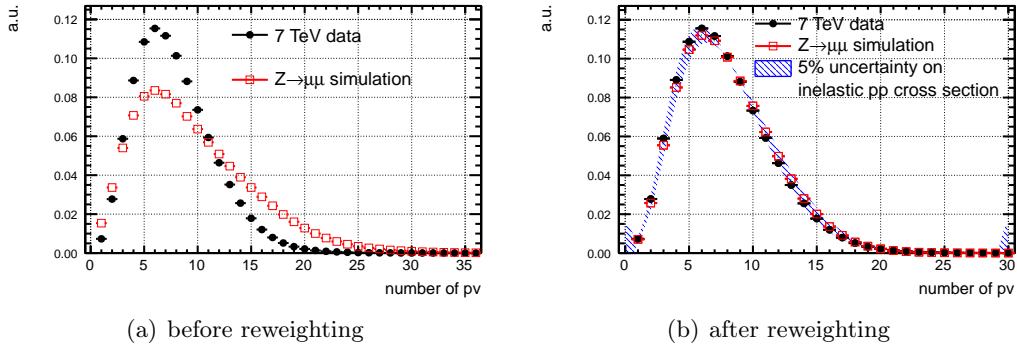


Figure 3.1: Number of reconstructed primary vertices (pv) in data and simulation. Figure (a) compares data to the unreweighted MC sample. Figure (b) presents the pileup reweighting of the MC sample, taking into account 5% uncertainty on the inelastic proton-proton (pp) cross section in the estimation of the data pileup distribution.

from the expected number of pileup events for each lumi section and bunch crossing, weighted by its integrated luminosity.

The pileup reweighted MC sample provides a sufficient description of the number of reconstructed primary vertices in data. The result after applying the weight is shown in Figure 3.1(b). Furthermore, a 5% uncertainty [105] on the inelastic proton-proton cross section in eq. (3.2) is applied. The distribution, including the systematic uncertainty, is presented with an uncertainty band and the MC predictions are located within this envelope.

An alternative approach to reweight the MC prediction is described in appendix B.

3.2 Particle-Flow Reconstruction

In CMS various reconstruction algorithms are based on different detector components. Combining the information obtained from all the subcomponents allows a precise measurement of the particle kinematics. In order to discriminate the final-state particles from the dominant SM background processes, it is essential to use an accurate particle reconstruction and identification.

The concept of *Particle-Flow (PF) Reconstruction* [106] is a method to produce physics objects, i.e. electrons, photons, muons, charged and neutral hadrons, including all physics properties. Information from all sub-detectors are used to construct the particle candidates with a precise four-momentum and high resolution. The list of reconstructed physics objects is called *PF Candidates*. The PF Candidates can be used as an input for jet, missing transverse energy, and τ reconstruction.

The PF Reconstruction is based on charged particle tracks, calorimeter clusters, and information from the muon system. Combining the information from all sub-components of the detector, the measurement can be performed with high efficiency,

while keeping the fake rate low.

An iterative tracker algorithm is used to transform reconstructed tracker hits into tracks with small fake rate. The iterative method is initiated by track seeds (tracker hits) selected with tight criteria. The seeding criteria are loosened in each iteration and hits assigned to tracks are removed. The relaxed requirements improve the track efficiency and the fake rate is kept low.

Additionally to the track reconstruction, a cluster algorithm is performed. The reconstructed calorimeter hits are converted to *PF Clusters*. Topology clusters are build from seeds (energy in calorimeter cells) by adding sequentially neighbouring cells above a certain energy threshold. The PF Cluster energy and position is determined taking into account the energy and position of each individual cell.

The different elements, charged particle tracks, calorimeter clusters, and muon tracks, are combined and the PF Tracks are matched to the PF Clusters forming the *PF Blocks*. A link algorithm to built the PF Blocks is developed to connect pairs of elements in the event, but also to avoid double counting from different detector components. Typically, the PF Blocks contain one, two or three elements.

Finally, the *PF Algorithm* creates the PF Candidates based on the information of the PF Blocks. The PF algorithm is performed for each PF Block. From the collection of tracks, the tracks associated with muons are excluded, building the muon collection and the remaining tracks from the *PF Track Collection*. More information on the muon reconstruction follows in section 3.3.1.

To reconstruct charged hadrons, the tracks must have a relative uncertainty on the transverse momentum smaller than the relative calorimeter energy resolution. This selection reduces the track fake rate. The tracks are connected to a number of ECAL and HCAL clusters. Several tracks can be linked to one calorimeter cluster. To distinguish between neutral and charged hadrons the sum of the track momenta is compared to the calibrated calorimetric energy. The reconstruction of forward particle production uses only calorimeter information measured with the HF calorimeter between $3.0 < |\eta| < 5.0$, due to the absence of the tracking system in the forward region.

The PF Candidates are the basis for the muon identification and jet types used in this analysis. A more detailed description of the algorithms and reconstruction performance is given in [106].

3.3 Muons in CMS

As the name Compact Muon Solenoid implies, muons are one of the main particles to detect with CMS. Muons can penetrate large detector regions without being absorbed by interaction with the detector material. They cross the iron return yoke and are not stopped by the calorimeters. They only dispose a small amount of energy in the calorimeter, that is why the muons are embraced by the term *minimum ionising particle (MIP)*. The muons leave a characteristic signal in the muon chambers, which can be used to discriminate to other particle decays.

This analysis is based on low and high energetic muons covering a wide range in invariant mass of the dimuon system. Thus, this section is devoted to the recon-

struction and identification of the muons. Further, the isolation definition and the efficiency are presented.

3.3.1 Muon Reconstruction

In section 2.2.6 the different muon objects, standalone-, tracker-, and global muon, were defined. The two main reconstruction algorithms, *tracker muon reconstruction* and *global muon reconstruction*, use the information from the inner tracker and the muon system. The details on the reconstruction algorithms are reviewed in [94].

The global muon reconstruction follows an outside-in principle. The starting point are standalone muon tracks, for which matching tracker tracks are found. A fit is performed including the standalone track and the tracker track by extrapolating from the inner muon station to the outer tracker. The best-matching tracker track is selected based on the information on the χ^2 of the fit.

The tracker muon reconstruction follows an inside-out approach. Possible muon candidates are selected from tracker tracks with $p_T > 0.5$ GeV and total momentum of $p > 2.5$ GeV. An extrapolation to the muon system is performed taking into account the impact on the magnetic field and effects on the interaction with the detector material. To each extrapolated track a matching muon segment, either in the sub-detectors CSC or DT, is selected if certain geometrical conditions are fulfilled and the tracker muon is defined.

Due to the high tracker and standalone track reconstruction efficiency around 99% of the muon candidates are reconstructed either as global or tracker muons or as both. The muon momentum resolution, as presented in Figure 2.9, shows that the combination of both algorithms provides an unambiguous muon reconstruction.

3.3.2 Muon Identification

This analysis of the differential Drell-Yan measurement is based on global and tracker muons. The event is selected if the Drell-Yan process decays into two muons, which can be reconstructed by the tracker and global muon reconstruction algorithm. In order to perform a precise measurement based on muon kinematics, the selection criteria are optimised to minimise the muon fake rate while keeping the selection efficiency high. The basic muon identification algorithm used in CMS is based on the PF Algorithm. The details on the selection criteria are summarised in Table 3.1.

Most criteria suppress muons from decays in flight, cosmic muons, and hadronic punch-through. The latter describes processes, when a fraction of a hadronic shower leaks into the muon system and is misidentified as a muon. Further, the requirements on the track reconstruction in the tracker and muon chambers, and on the global fit, ensure a good estimate on the muon trajectory by reducing the mismatching of tracker tracks and standalone tracks. Additionally, an accurate p_T measurement can be performed.

3.3.3 Particle-Flow Isolation

In an electroweak process the final-state muons are well separated from additional activity of other particles; they are isolated. A muon emerging from a QCD process would be accompanied by a large number of low momentum jets (multi-jet events), thus, leaving tracks and energy deposits in the tracker and calorimeter. The isolation condition of a muon is necessary to discriminate between the different physics production processes of the muons. The muons under study are based on isolated muons. Therefore, a selection on the energy surrounding the muon in a certain cone in $\eta - \phi$ space, defined by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, is required.

Selection Criterion	Comment
The muon track must have more than one hit in the pixel detector.	The rate of muons from decays in flight is suppressed.
The global-muon-track fit is performed using at least one hit in the muon system.	The rate of punch-through hadrons and muons from decays in flight are suppressed.
The muon track must have more than eleven tracker hist (in total pixel and strip).	Minimum number of tracker points is needed to ensure an optimal p_T measurement.
The muon candidate is selected from at least two hits in the muon stations (in different layers).	The rate of punch-through hadrons and accidental track-to-segment matching are suppressed.
The normalised χ^2/ndof of the global-muon-track fit has to be less than ten.	The rate of punch-through hadrons and muons from decays in flight are suppressed.
The impact parameter in the transverse plane has to fulfil $ d_0 < 0.04$ cm w.r.t. primary vertex.	The rate of cosmic muons and muons from decays in flight are suppressed.
The longitudinal impact parameter has to fulfil $ d_z < 0.2$ cm w.r.t. primary vertex.	The rate of cosmic muons and muons from decays in flight are suppressed.

Table 3.1: Baseline muon selection criteria for 2011 data defined by CMS.

There are different approaches to define the optimal isolation variable depending on the tracker and calorimeter output. In this analysis the *relative combined PF based*

isolation variable

$$\text{Iso}_{\mu}^{\text{PF}} = \frac{\sum (p_{\text{T}}^{\text{charged}} + p_{\text{T}}^{\gamma} + p_{\text{T}}^{\text{neutral}})}{p_{\text{T}}^{\mu}} \quad (3.3)$$

is used. The sum runs over the transverse momenta of all charged particles emerging from the hard interaction vertex. Only particles within a cone of $\Delta R < 0.4$ around the muon momentum are selected. In order to reduce the amount of energy coming from additional collisions, the sum has to be corrected for activity coming from pileup events. In the case of charged particles only particles, which correspond to the primary vertex are selected. In the case of neutral particles the track information is not available and the above requirement can not be arranged. Thus, a mean energy density ρ in a cone around the neutral particles is subtracted ($\Delta\rho$ correction). The

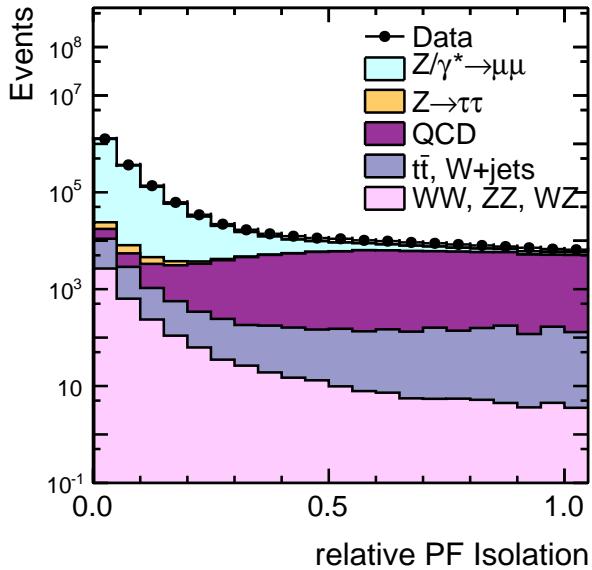


Figure 3.2: Relative PF Isolation variable. The data is compared to simulation of signal and several background contributions.

relative $\Delta\rho$ corrected PF muon isolation is shown in Figure 3.2. The background contribution from QCD (multi-jet) processes dominates for high values of the isolation variable. At lower values the $Z^0/\gamma^* \rightarrow \mu\mu$ process is dominant. Thus, the isolation selection for the dimuon system is required as

$$\text{Iso}_{\mu}^{\text{PF}} < 0.10 \quad \text{for} \quad p_{\text{T}}^{\mu} > 20 \text{ GeV}, \quad (3.4)$$

$$\text{Iso}_{\mu}^{\text{PF}} < 0.15 \quad \text{for} \quad p_{\text{T}}^{\mu} < 20 \text{ GeV}. \quad (3.5)$$

In this analysis the muons are selected according to the relative PF based selection criteria, presented above, and fulfil the relative combined PF isolation requirement in eq. (3.4) and (3.5). The looser criterion for muons with $p_{\text{T}} < 20$ GeV is needed to increase the selection efficiency on the softer muons. This is a crucial request for measurements of low invariant dimuon masses.

3.3.4 Muon Efficiency

The efficiency on the muon identification selection is estimated using the tag-and-probe method. The details on this method were presented in section 2.2.6. The identification efficiency is estimated from data and MC prediction using $Z \rightarrow \mu\mu$ events. The resulting values as a function of the muon p_T and η are presented in Table 3.2. Differences between the MC and data efficiency are of the order of a few percent and within the statistical uncertainties. Correction factor are defined as the ratio of data over MC efficiencies, subsequently referred to as scale factors. The last column in Table 3.2 shows the scale factors $\rho_{ID} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}$. The scale factors are applied as an event weight to the MC events,

$$\omega_{ID}(p_T, \eta) = \rho_{ID}. \quad (3.6)$$

p_T (GeV) & η bin	MC Efficiency	Data Efficiency	Scale Factor
10.0 - 15.0 $ \eta < 1.5$	0.686 ± 0.004	0.680 ± 0.012	0.991 ± 0.018
10.0 - 15.0 $1.5 < \eta < 2.1$	0.652 ± 0.004	0.676 ± 0.014	1.036 ± 0.022
15.0 - 20.0 $ \eta < 1.5$	0.751 ± 0.002	0.748 ± 0.007	0.995 ± 0.010
15.0 - 20.0 $1.5 < \eta < 2.1$	0.718 ± 0.003	0.720 ± 0.010	1.002 ± 0.014
> 20.0 $ \eta < 1.5$	0.930 ± 0.001	0.922 ± 0.001	0.992 ± 0.001
> 20.0 $1.5 < \eta < 2.1$	0.895 ± 0.001	0.890 ± 0.001	0.994 ± 0.001

Table 3.2: Muon identification efficiency scale factors obtained from data and MC simulation in bins of muon p_T and η [107]. Statistical uncertainties are included and propagated to the scale factors.

3.4 Jets in CMS

In high energy collisions it is most likely that the Drell-Yan muon pair is accompanied by additional hard-parton emissions. As discussed in section 1.4.5 the hard partons can not be detected directly, but in form of jets. The partons form colour-neutral states by fragmentation and hadronisation. These hadrons can be grouped in a collimated flow of particles, with defined direction and a defined extension in

$\eta - \phi$ plane.

In order to gain information on the initial parton, which induced the jet, a precise reconstruction and identification is important. Measuring jets with highest accuracy and optimal resolution allows to probe hard partons emerging from the hard scatter.

This section is devoted to the precise jet reconstruction and identification. Additionally, jet energy corrections are explained and jet energy resolution smearing is presented.

3.4.1 Jet Reconstruction

A jet can be described as a bunch of hadrons and other particles grouped in, for example, $\eta - \phi$ space. The jet is produced by hadronisation of quarks and gluons. In Figure 3.3 a sketch of a jet is illustrated. The figure shows a quark induced jet. The initial produced quark emits other partons, and after fragmentation and hadronisation many hadrons (π, K, \dots) are created, which are collimated in the same direction of the original parton. In the detector the energy of the hadrons can be

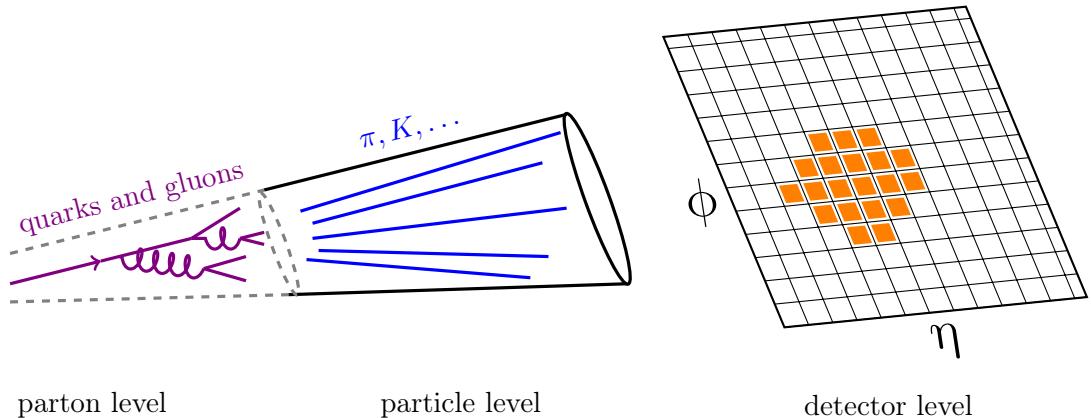


Figure 3.3: A sketch of a jet, presented at different levels. The lines and curly lines represent the partons emerging from the hard scattering (parton level). After fragmentation of partons they hadronise to hadrons and travel in the direction of the initial parton. The final-state particles (this means after fragmentation) can be used as input for jet reconstruction (particle level). The charged and neutral hadrons are grouped together in space in $\eta - \phi$. The hadrons leave energy in the tracker and calorimeter cells, which can be reconstructed to define the jet (detector level).

measured in the calorimeter and the tracks can be observed in the tracker system. Thus, different subsystems can be used to reconstruct the jets. The physics objects, which can be measured are defined by jet algorithms. They are defined by their four-momenta, which correspond to the sum of the four-momenta of all input objects. The jet algorithm should be easily applicable in experimental and theoretical calculations, in order to enable a direct comparison between the measured and calculated cross section. Thus, as input for jet algorithms detector signals (detector level) and well

as final-state particles (particle level) can be used.

Jet Algorithm

Jet algorithms describe how to combine the particles or detector signals in a jet and how the clustering is controlled by a distance parameter R .

There are two main categories of jet algorithms: cone [108] and recombination algorithms [109, 110]. The iterative cone algorithm is used by CMS for online trigger in the HLT, due to the short and calculable execution time. It uses calorimeter towers as seeds and an iterative method maximises the energy flow within a cone of radius R around the cone axis. The sequential recombination algorithm is a cluster algorithm. Within its framework particles close to each other in $\eta - \phi$ space are combined demanding certain distance requirements.

The k_t algorithm [110] combines two objects i, j depending on their distance d_{ij} to each other and to the beam d_{iB} . The resulting composition of particles is called a jet. The relevant distances for the k_t algorithm are

$$d_{ij} = \min \left(k_{t_i}^{2p}, k_{t_j}^{2p} \right) \frac{\Delta R_{ij}^2}{R^2}, \quad (3.7)$$

$$d_{iB} = k_{t_i}^{2p}, \quad (3.8)$$

where k_t is the transverse momentum of the particle and $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ defines the angular distance between the two objects. The iterative algorithm proceeds as follows; first the distance of particle i to the closest particle j d_{ij} and to the beam d_{iB} is calculated. If $d_{ij} < d_{iB}$ the two particles are combined by adding the four-momenta of the two particles. If $d_{ij} > d_{iB}$ the particle i is considered as a jet and excluded from the list of possible jet constituents. These steps are repeated until all objects are clustered.

The characteristic parameters p and R define different types of the k_t algorithm, where R provides a weight to d_{iB} and p gives the power to the momentum scale. The k_t algorithm uses $p = 1$, the Cambridge-Aachen algorithm implements $p = 0$ and the anti- k_t algorithm runs with $p = -1$.

In this analysis the anti- k_t algorithm [111] is employed to define the jets. This algorithm implies important features of reconstruction algorithms because it is fast, the reconstructed jet has the most conic form compared to the other algorithms, and infra-red and collinear (IRC) safety is ensured. The latter point ensures that the jet algorithm is insensitive of adding soft or collinear parton splittings.

Jet Types

The jet algorithms are based on input objects. There are different objects, which can be used as input for the jet algorithm, defining different jet types. The input objects can be defined by measured objects or from MC simulation.

In the case of MC simulation the clustering method can be performed on generated

stable particles. The collection of reconstructed jets based on the MC truth information is independent of the detector and is called *GenJets* (defined on generator level).

In the case of measured objects³ there exist four different types depending on the combination of reconstructed objects. *CaloJets* are defined using energy deposits in the calorimeter towers only. The second approach are *TrackJets*, which use track information only. The tracks can be matched to CaloJets to improve the energy resolution. The combination of CaloJets and tracks is called jet-plus-track jets, *JPTJet*.

In this analysis *PFJets* are employed. The PFJets are based on the PF definition, see section 3.2. The anti- k_t algorithm is run on the collection of PF Candidates and the four-momenta of each PF Candidate is added. The clustering is performed for the individual PF Candidates.

3.4.2 Jet Identification

In this work the jet type is PFJets. In order to select only real hadronic jets they must be separated from the two muons in the final state by $\Delta R > 0.5$.

In order to reduce the amount of jets coming from pileup a jet identification is included. The jet identifier uses vertex information and jet shape information to ensure the jet is a good jet [112]. Additionally, certain requirements are added to reduce the fraction of fake jets coming from noise in the calorimeter. Furthermore, the identification criteria are arranged to select only hadronic jets with high efficiency. The identification criteria are defined in CMS by the *Loose PFJet Identification* [113]:

- $|\eta^{\text{jet}}| > 2.4$: The neutral hadron energy fraction should be < 0.99 , the photon energy fraction should be < 0.99 and the number of constituents should be larger than 1.
- $|\eta^{\text{jet}}| < 2.4$: The neutral hadron energy fraction should be < 0.99 , the photon energy fraction should be < 0.99 and the number of constituents should be larger than 1. Additionally, the number of charged constituents and the charged energy fraction have to be larger than 0 and the electron energy fraction < 0.99 .

3.4.3 Jet Energy Scale and Jet Energy Resolution

The determination of the original parton kinematics requires a precise determination of the jet energy scale (JES) and jet energy resolution (JER). The measured jet parameters deviate from the real parameters of the particle level jet, due to detector effects. The systematic differences can be compensated by the calibration of the detector response. The JES is the calibration point of the jet energy and defined as the average value of the jet response. The latter is defined as the ratio of the

³These jet types can also be reconstructed from MC generation including detector simulation.

measured jet p_T on detector level over the generated jet p_T [114]

$$\mathcal{R} = \frac{p_T^{\text{detector level}}}{p_T^{\text{particle level}}} \quad (3.9)$$

A precise knowledge of the JES of hadronic jets enables to calibrate the measured jet energy by applying a jet energy correction (JEC) depending on η^{jet} and p_T^{jet} . Thus, after calibration the JES should be 1, $\langle \mathcal{R} \rangle = 1$. Furthermore, a precise knowledge of the JER is needed. The JER corresponds to the typical deviation of the measured energy from its expected value and can be interpreted as the width of the response distribution. A detailed introduction to the JEC and JER for the different kinds of jet types is given in [115].

I calculated the jet transverse momentum response for simulated events using

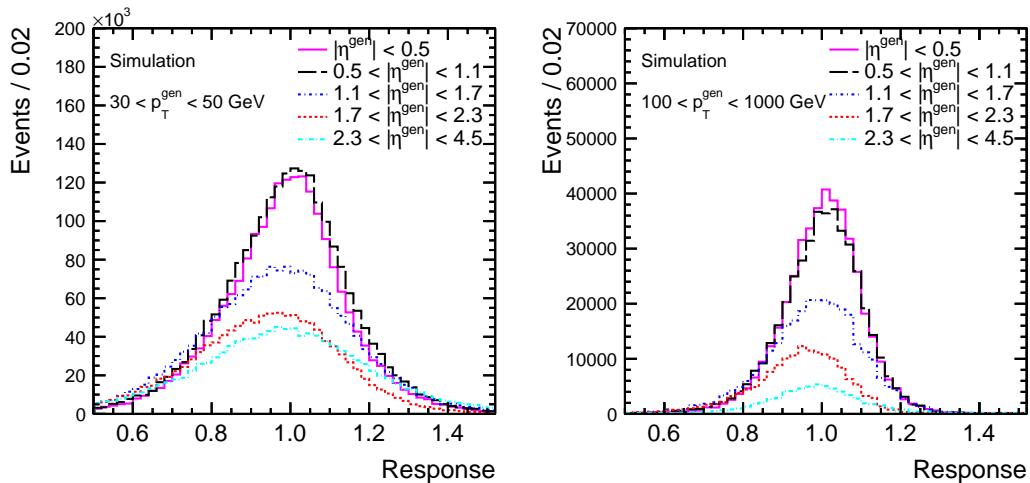


Figure 3.4: MC truth response for high and low p_T jets compared in different η bins. The response after JEC and JER smearing is shown.

MADGRAPH +PYTHIA6, shown in Figure 3.4. On generator level the particle level jet p_T refers to the MC truth information and is defined as the GenJet p_T . The Drell-Yan events are selected in the Z invariant mass range of $60 < m^{\mu\mu} < 120$ GeV. The response is plotted for two different generated jet p_T of $30 < p_T^{\text{gen}} < 50$ GeV and $100 < p_T^{\text{gen}} < 1000$ GeV in different bins of jet η^{gen} . The results are shown after the jet energy calibration and the average response is around a value of ~ 1 . The uncertainty on the JEC increases with η , due to the missing tracking information in the forward region. A dependence of the JER on η and p_T is observed. For higher jet p_T the width of the response is smaller than for lower jet p_T and with increasing jet pseudorapidity the response smears out.

Jet Energy Correction

The JEC can be factorised into several components, which imply different effects of the jet reconstruction: First the reconstructed jets, *RecoJets*, have to be matched

to the GenJets. A global calibration factor is defined as the ratio of the energy of RecoJets over GenJets as a function of p_T^{jet} and η^{jet} . This calibration factor is applied to data and detector level simulation and removes the energy contribution in a jet coming from pileup events. However, differences between data and simulation give rise to additional corrections in the defined phase space. The aim of the correction in the second step (L2Relative) is to flatten the η dependence of the jet response. The forward produced jets are corrected to have the same response as centrally produced jets. For this the p_T balance of dijets is used. In the third step (L3Absolute) the jet response in p_T is flattened. The absolute correction is used to scale the p_T response of the jets. The correction makes use of the p_T balance in $Z + \text{jet}$ and $\gamma + \text{jet}$ events. At the end the residual corrections (L2L3Residual) are applied to data only, which take into account remaining small differences between data and MC simulation and fix the relative energy scale [115].

The jet p_T resolution was studied in dijet and $\gamma + \text{jet}$ events in [116] and it was observed that the measured resolution in simulation is better than in data. In order to improve the agreement of data and simulation, the simulated jets have to be smeared to describe the data. The smearing procedure is based on a scaling of the reconstructed and corrected jet p_T dependent on the p_T difference between the matched RecoJet and GenJet:

$$p_T \rightarrow p_T^{\text{gen}} + c(p_T - p_T^{\text{gen}}). \quad (3.10)$$

The factor c is the score resolution scaling factor defined as the ratio of data over MC resolution. The scaling factors are given in [117] and are estimated from dijet events. The JER smearing only works for well matched RecoJets to GenJets. This means, for each reconstructed jet j^{reco} the closest generated jet j^{gen} in $\eta - \phi$ space is selected. The two jets are matched if the distance $\Delta R = \sqrt{\Delta\eta^2(j^{\text{reco}}, j^{\text{gen}}) + \Delta\phi^2(j^{\text{reco}}, j^{\text{gen}})}$ is smaller than a certain value. In this analysis the matching value is chosen to be $\Delta R < 0.4$. Figure 3.4 shows the response after JER smearing and JEC. With an ideal Gaussian shape of the response distribution the JER corresponds to the variance σ of the function. The comparison of the root mean square (RMS) values of the response distributions for low and high p_T jets shows good agreement. Furthermore, the ratio of the RMS of including JER and not including JER smearing reproduces the correction factor c .

4 Drell-Yan + Jets Analysis

In this chapter the details on the Drell-Yan + jets measurement are presented. Preliminary results on this analysis are published within CMS in Ref. [118].

The Drell-Yan cross section measurement is performed differentially in the dimuon transverse momentum $p_T^{\mu\mu}$. As explained in section 1.4.3, the $p_T^{\mu\mu}$ distribution provides a crucial test of pQCD. The different regions of the dimuon $p_T^{\mu\mu}$ distribution are calculated by several theoretical calculations. In the region of $p_T^{\mu\mu} > Q$ (Q is the scale of the hard process) the cross section follows a fixed-order calculation, but in the region of $p_T^{\mu\mu} \ll Q$ the truncated perturbative expansion is not reliable and large-logs have to be resummed to all orders in pQCD. In the intermediate region nonperturbative effects contribute due to the intrinsic motion of the partons in the colliding hadrons. Thus, the sensitivity of different calculations and models depend on a physical scale Q , which can be interpreted as the invariant mass of the Drell-Yan dimuon system $m^{\mu\mu}$. The cross section measurement is performed differentially in mass $m^{\mu\mu}$, to use it as scale of the process. This is the first time in CMS that the $p_T^{\mu\mu}$ distribution is measured differentially in the dilepton invariant mass. Moreover, the inclusive Drell-Yan production in the mass range of the Higgs boson ($m^{DY} \sim 126$ GeV) can be used to probe soft-gluon and multi-jet resummation from a quark and gluon induced process, by comparing the Drell-Yan and Higgs cross section. Details of how the comparisons are relevant for the future high luminosity runs at the LHC are described in [119] (see appendix G).

The effect of resummation is of great interest for this analysis. As motivated in section 1.4.4, the phase space for soft-gluon resummation can be enlarged when requiring additional jets in association with the Drell-Yan boson production. For the first time in CMS the jets are selected in the full η region of the detector. Comparing inclusive Drell-Yan production to Drell-Yan and at least one jet above a certain p_T threshold, an increase of the region where resummation is important is observed. In this way it is possible to study the effect of resummation at low p_T by measuring multi-jet emissions. In this analysis three event topologies are compared:

1) Inclusive DY

Inclusive Drell-Yan production with no further requirement on the jet selection

2) DY + 1 jet

Inclusive Drell-Yan production with at least one jet above a p_T threshold of 30 GeV

3) DY + 2 jets

Inclusive Drell-Yan production with at least two jets above a p_T threshold of 30 GeV

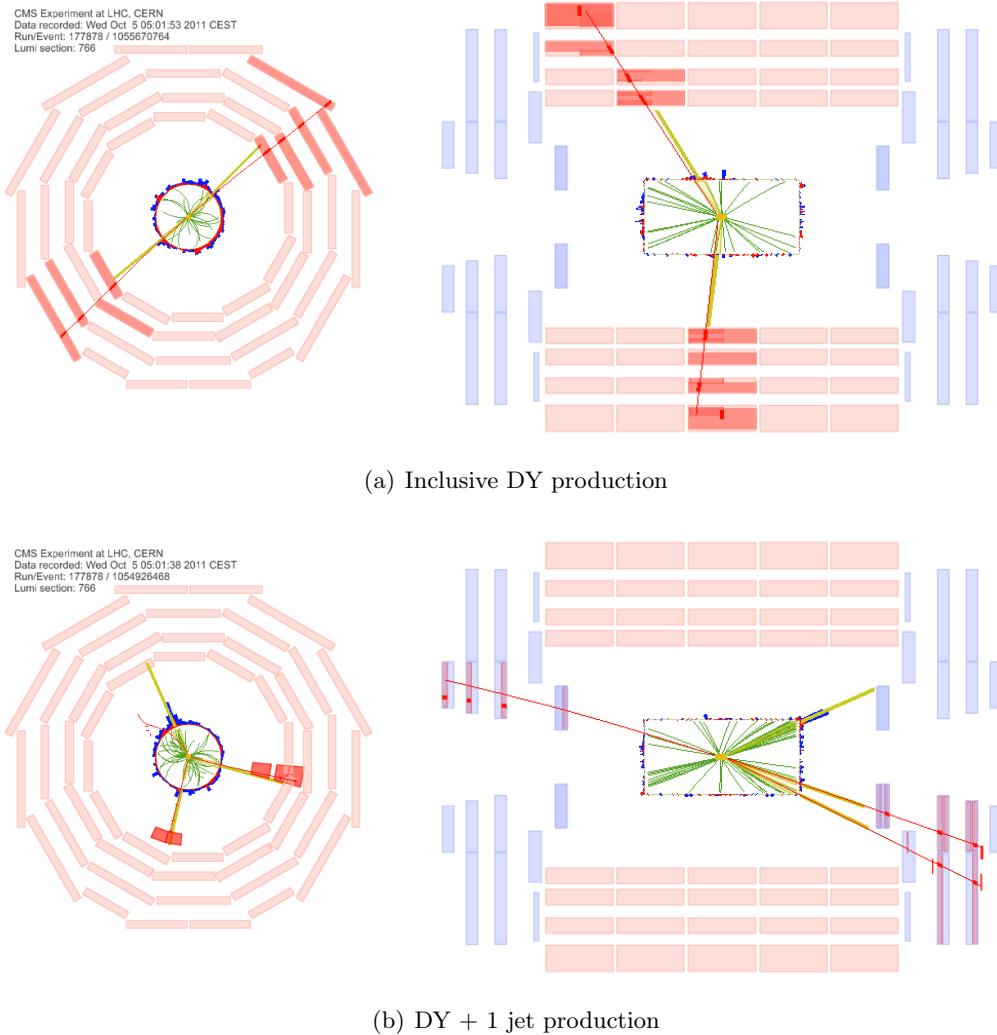


Figure 4.1: Event display of inclusive DY and DY + 1 jet production in the transverse (left) and longitudinal (right) plane of the detector. The red lines represent muon candidates and yellow blocks represent PF jets. Red and blue rectangles represent energy deposits in the towers of ECAL and HCAL respectively. The solid green lines represent reconstructed particle tracks.

The event topologies 1) and 2) are presented in an event display of the CMS detector in Figure 4.1(a) and 4.1(b) respectively. The visualisation of the dimuon signal is presented in the transverse (left) and longitudinal (right) plane of the detector. The red lines represent the muon candidates leaving signals in the different subcomponents of the detector. The yellow lines represent the PF reconstructed jets. In Figure 4.1(b) one can see nicely the balancing of the jet and the Drell-Yan dimuon system in the transverse plane, as it was demonstrated in section 1.4.3. In case of inclusive DY production, the muons are produced in the central region, and are nearly back-to-back. In the case of the DY + 1 jet production, the muons and jets

are detected in the forward region by the endcaps (illustrated in blue). The third muon in the event display has only a small p_T curvature and is not selected.

Further observables to study the multi-jet emissions in the Drell-Yan + jet topologies are provided by angular distributions. In this analysis a forward Drell-Yan pair with $|\eta| > 2.5$ and jets in a pseudorapidity range up to $|\eta| < 4.5$ are selected. A sketch of the process is illustrated in Figure 4.2. This is the first time the Drell-Yan boson is selected in association with jets in such a wide range of pseudorapidity in an CMS analysis. The cross section of the rapidity separation of the Drell-Yan and the leading jet $|\Delta y(\mu\mu, j_1)|$ is sensitive to multi-gluon emission [120]. At large rapidity separation the fixed-order calculation is expected to fail and multi-jet emissions are important. Due to the fact that the Drell-Yan is selected to be forward, the observable $|\Delta y(\mu\mu, j_1)|$ is sensitive to small- x resummation. Additionally, the jet multiplicity above a certain p_T threshold as a function of $|\Delta y(\mu\mu, j_1)|$ is investigated to study the role of multi-gluon emissions at large rapidity.

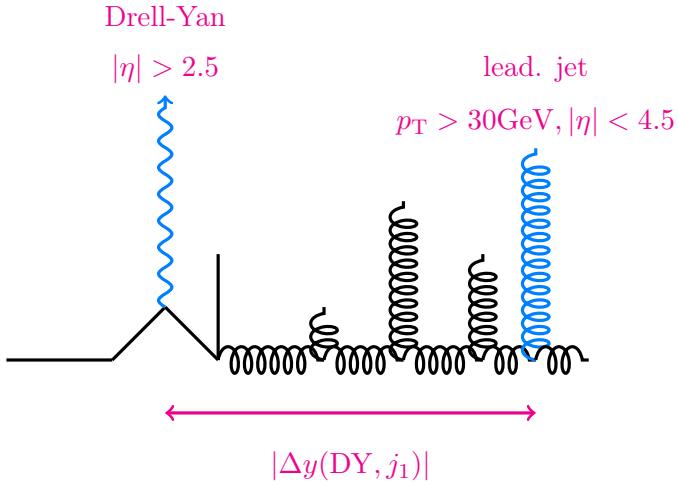


Figure 4.2: Illustration of rapidity separation of the forward Drell-Yan production (represented by the Z^0/γ^*) and leading jet (represented by the gluon emission in blue).

Concluding, the Drell-Yan lepton pair production provides a crucial test of the SM, by measuring the differential cross section, which is sensitive to different theory calculations in pQCD.

Although Drell-Yan production provides a clean and well-understood signal in the Z invariant mass range, in the low and high invariant mass ranges contributions from different background processes mimic the signal and contaminate the measurement. Thus, a clean and well-understood signal is desired in the full invariant mass range under study. By applying selection requirements on the kinematic variables of the Drell-Yan lepton pair and jets, the rate of signal-to-background events can be optimised. The background contributions have to be modelled precisely in simulation or

can be estimated from data events.

In this chapter the data and simulated samples of signal and background processes are presented. Second, the selection requirements on the Drell-Yan dimuons and jets are introduced. Control distributions, which present the detector level data in comparison to the MC signal plus background processes, are shown¹. The unfolding method, to correct the data events for detector effects, is presented. Finally, the systematic uncertainties on the measurement are summarised.

4.1 Data and Simulated Samples

The Drell-Yan measurement is based on proton-proton collision data recorded in 2011 with the CMS detector at $\sqrt{s} = 7$ TeV. The data events are collected with single and double muon triggers during 2011 runs, as listed in Table 2.2. The total number of recorded events correspond to an integrated luminosity of 4.9 fb^{-1} .

In order to obtain good quality data, the online and offline data taking has to be monitored (data quality monitoring (DQM)). The different data streams are qualified and certified in stages of different lumi sections. During online and offline data taking the detector components are controlled and the full event reconstruction is checked. Good quality data correspond to the requirements that the beams are stable, and the detector components and the magnetic field worked properly during the runs.

Process	Generator	σ (pb)	Number of generated events
$Z + \text{jets}$ ($m_{\text{ll}} > 50$ GeV)	MADGRAPH	3048	36,179,628
$Z + \text{jets}$ ($m_{\text{ll}} < 50$ GeV)	MADGRAPH	9530	31,480,628
$Z \rightarrow \tau\tau$	MADGRAPH	3048	36,179,628
$t\bar{t}$	MADGRAPH	165	17,889,708
$W + \text{jets}$	MADGRAPH	31314	81,345,384
QCD ($p_{\text{T}}^{\mu} > 15$ GeV)	PYTHIA6	84679	25,080,240
WW	PYTHIA6	43	4,225,916
WZ	PYTHIA6	18.2	4,265,243
ZZ	PYTHIA6	5.9	4,191,045

Table 4.1: Signal and background MC samples.

The determination of background events from processes that result in two muons as well as the determination of systematic uncertainties is performed using MC event samples. The MC samples are produced by the MADGRAPH (version 5) and PYTHIA6 (version 6.4.33) event generators. These samples are passed through the full CMS detector simulation based on GEANT4, which includes trigger simulations, full chain

¹ Adding signal and background contributions refer to a stacked histogram.

of the CMS event reconstruction, and pileup simulation.

The signal and background samples used for MC simulation are listed in Table 4.1. To compare the observed data with the simulation of the experiment, the simulated events are normalised to obtain the same distribution of the mean number of proton-proton collisions per bunch crossing. For this reason a normalisation factor is assigned to the MC events according to the number of expected events in data

$$\omega_{\text{norm}} = \frac{\sigma \mathcal{L}}{N_{\text{processed}}}, \quad (4.1)$$

where σ is the cross section of each process as listed in Table 4.1, taken from Ref. [121], \mathcal{L} is the integrated data luminosity, and $N_{\text{processed}}$ is the number of generated MC events.

The Drell-Yan signal samples are generated by MADGRAPH and are split into a sample of dileptons with invariant mass $10 < m^{\text{DY}} < 50$ GeV and another with $m^{\text{DY}} > 50$ GeV. Electroweak production of the Drell-Yan lepton pair is not included in the MADGRAPH MC predictions.

Different background processes contribute to the final state of two leptons and jets and can be misidentified as signal events. The precise modelling of the background processes is necessary in order not to bias the signal. The estimation and validation of the different background processes is presented in section 4.4. The contributions from $Z \rightarrow \tau\tau$, $t\bar{t}$ and $W+\text{jets}$ production are generated by MADGRAPH. All samples are interfaced with PYTHIA6 to simulate the parton shower, hadronisation and UE. In all cases the tune Z2 is used. The background contributions from dibosons (WW , WZ , ZZ) and QCD multi-jets are produced with PYTHIA6 Z2.

4.2 Event Selection

The Drell-Yan process in the muon channel provides a characteristic signature of two isolated muons with opposite charge. The identification and isolation of the muons was explained in section 3.3. The muons must correspond to the same primary vertex (cf. section 3.1) to ensure they are from the same collision. The measurement is performed for a pair of muons with $p_T^{\text{lead}\mu} > 20$ GeV and $p_T^{\text{sublead}\mu} > 10$ GeV. The p_T criteria are well above the trigger thresholds to minimise the bias from the trigger efficiency. The muons are selected within $|\eta^\mu| < 2.1$, which corresponds to the fiducial region of the trigger acceptance to optimise the trigger efficiency.

To validate the performance of the detector simulation the MC simulation is compared to measured events in so-called control distributions. In Figure 4.3 control distributions of the kinematic variables p_T , η , and ϕ of the positive muon μ^+ are presented. The data points are compared to signal plus background MC events and correspond to the number of events on detector level. The distributions represent the muon kinematics after the kinematic selection is applied. A good agreement between data and MC simulation is observed.

The dimuon system is based on the two selected muons by choosing the two leading

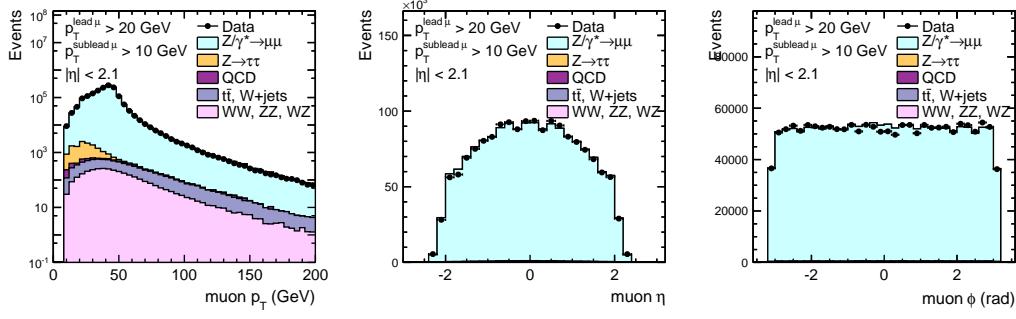


Figure 4.3: Muon kinematic variables after pre-selection. Data is compared to signal and background MC predictions.

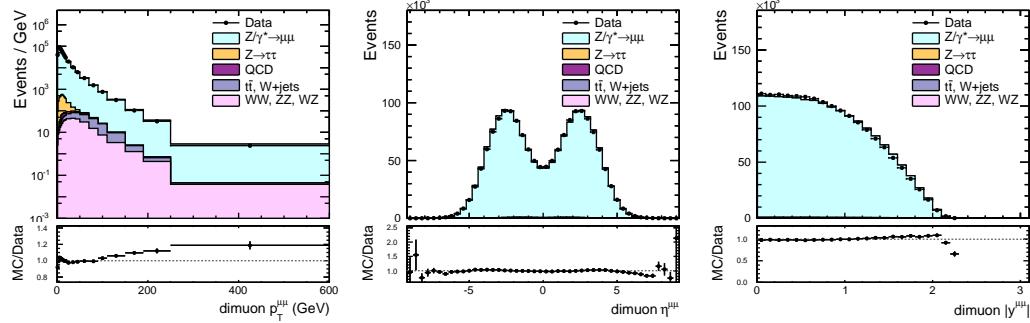


Figure 4.4: Dimuon kinematic variables. The number of events are normalised to the binwidth.

muons. The dimuon variables are defined as the vectorial sums of the two individual muon kinematic variables. The Drell-Yan cross section measurement is performed covering a wide range of the dimuon mass $30 < m^{\mu\mu} < 1500$ GeV. In Figure 4.4 the kinematic variables p_T , η , and y of the dimuon are presented. The plots show the number of events on detector level normalised to binwidth. In addition MC/data ratios are presented. All in all, a good agreement between data and MC predictions is observed. Although the two individual muons are selected with $|\eta| < 2.1$, the dimuon system can reach much larger values in pseudorapidity. In the dimuon $p_T^{\mu\mu}$ distribution the MC predictions overestimate the data for high $p_T^{\mu\mu}$ ($p_T^{\mu\mu} > 100$ GeV) of 10-20%. This behaviour is also seen in other CMS analyses [122].

The jets are selected from PF Jets with $p_T^{\text{jet}} > 30$ GeV and $|\eta^{\text{jet}}| < 4.5$ (cf. section 3.4). The control distributions of the leading and subleading jet are presented. In Figure 4.5 the kinematic variables p_T , η , and y are shown. The leading and sub-leading jet have both a hard p_T tail. The leading jet is mainly produced centrally, whereas the subleading jet has a wider η distribution.

In the following the combination of identification, isolation (as presented in section 3.3 and 3.4), and kinematic criteria of the muons, dimuons and jets are referred to as pre-selection. The event selection is applied on data and MC events.

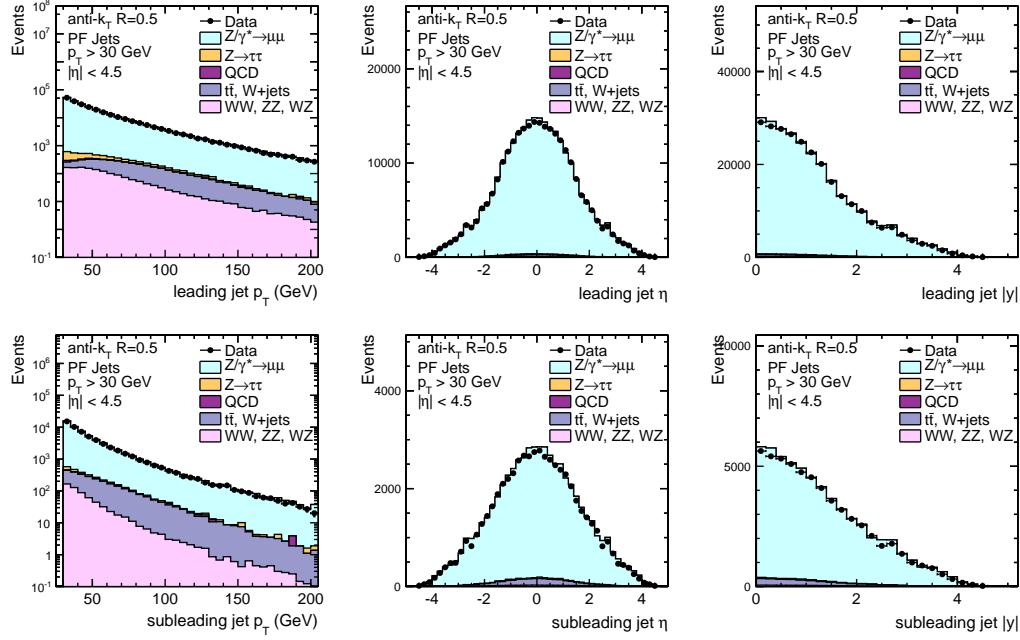


Figure 4.5: Leading (top) and subleading (bottom) jet variables after the pre-selection. Data is compared to signal and background MC predictions.

4.3 Monte Carlo Corrections

There exist several corrections to simulation in order to improve the agreement with data. The corrections are applied as an event weight ω_{MC} to the MC event and are combined in a product of individual weights, taking into account different effects. In addition to the absolute normalisation ω_{norm} (cf. eq. (4.1)), the MC corrections take into account the effect of pileup (cf. eq. (3.1)), trigger efficiencies (cf. eq. (2.5) and (2.6)), and efficiency correction factors (cf. eq. (3.6)) due to muon identification. The total correction can thus be defined as the product

$$\omega_{\text{MC}} = \omega_{\text{norm}} \cdot \omega_{\text{PU}} \cdot \omega_{\text{trigger}} \cdot \omega_{\text{ID}}. \quad (4.2)$$

4.3.1 QED Final-State Radiation

The final-state leptons of the Drell-Yan process are able to emit photons, which is referred to as QED FSR. The radiated photons are emitted nearly collinear to the final-state muon and can affect the invariant mass of the dimuon when the emitted photons have sizeable energy. The MC truth final-state muons can thus be defined with respect to the QED FSR. The final-state muons before QED FSR refer to born muons, while muons after QED FSR are defined as bare muons. The definition is illustrated in Figure 4.6(a). The effect of the QED FSR on the dimuon invariant mass is shown in Figure 4.6(b) by comparing $m^{\mu\mu}$ before and after QED FSR. It is

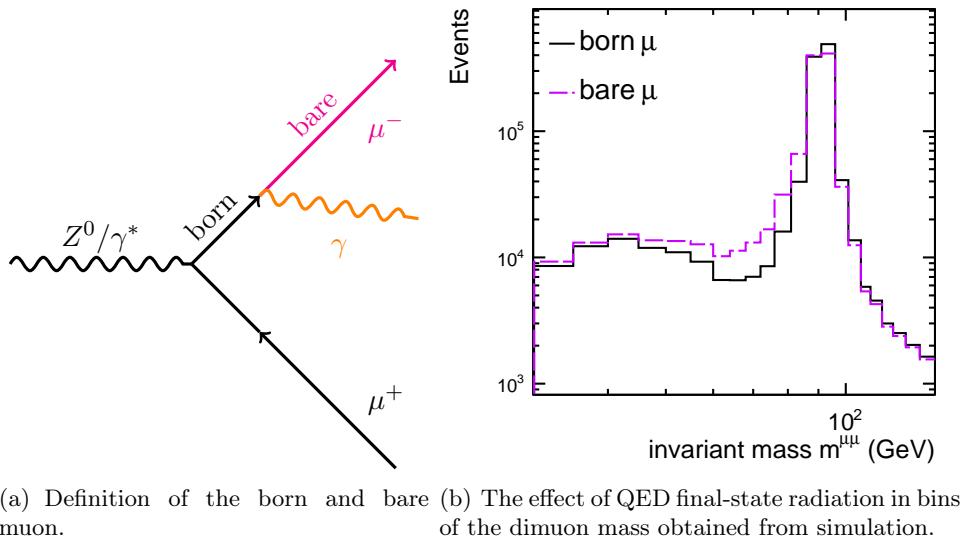


Figure 4.6: An illustration of the definition of the born and bare muons in the final state is shown in (a) and the number of events as a function of the dimuon invariant mass for bare and born muons is presented in (b).

observed that the effect is most significant just below the Z resonance. The other invariant mass regions are not significantly affected. In this analysis the final-state muons refer to muons after QED FSR, which are bare muons.

4.3.2 Muon Momentum Scale Correction

The muon momentum measurement depends on the alignment of the tracker. After tracker alignment a residual misalignment remains, which affects the reconstruction of the muon momentum. In order to recover the bias in data and MC simulation, the muon momentum scale has to be corrected. A detailed explanation on how the muon momentum scale correction is extracted is given in [123].

The muon momentum scale correction is derived as a function of the muon charge, η , and ϕ . Without the muon scale correction the average Z mass reveals a non-physical dependence on the muon η and ϕ . The comparison of data and simulation is shown in Figure 4.7(a). Including the muon scale correction improves the curvature of the muon with respect to the magnetic field and deformations in the transverse plane and it takes into account alignment effects. The average Z mass as a function of the ϕ observable of the muon after the muon scale correction is shown in Figure 4.7(b). The non-physical behaviour before the correction shows a dependence on the muon ϕ . After the correction data and simulation show a flat behaviour of the bosons mass in ϕ .

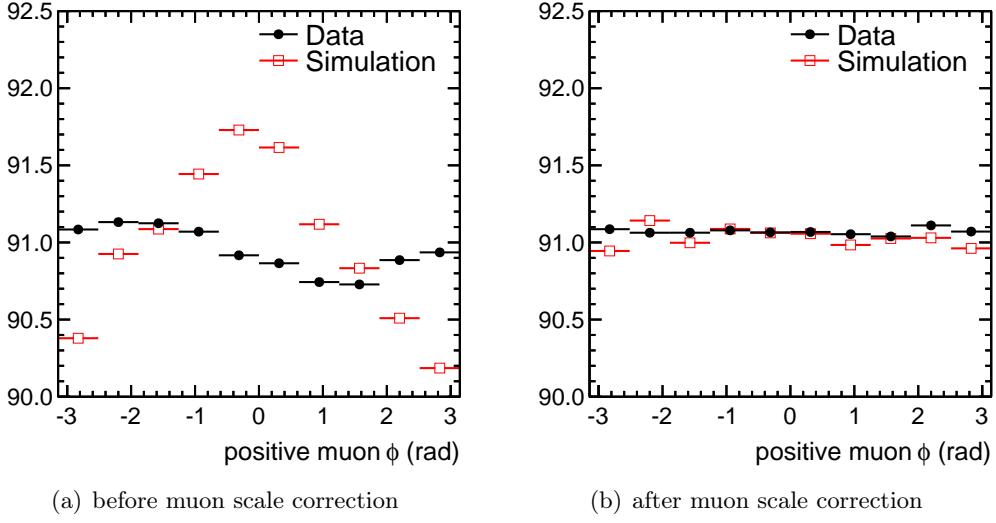


Figure 4.7: Invariant mass of the dimuon as a function of the positive muon ϕ . Before the muon scale correction (a) the data and simulation predictions show a non-physical dependence on the muon variable. In (b) a flat dependence after the correction is observed.

4.4 Background Estimation

Different processes, which have the same signature in the final state, can be misidentified as Drell-Yan dimuon pairs and are called background processes. With well-chosen selection requirements a large contamination from background processes can be reduced. However the remaining background contributions have to be modelled precisely. The background events are subtracted from data events in order to be able to compare to signal MC predictions. Thus, an accurate simulation (estimated from data or MC simulation) of the background processes is needed in order not to bias the signal.

The Drell-Yan cross section measurement is performed in bins of the dimuon invariant mass. The contribution from different background processes depends on the invariant mass of the dimuon, as shown in Figure 4.8 (left). In the Z peak region (60-120 GeV) the signal is nearly background free and the dominant contribution is Z boson exchange. In the low mass region (30-60 GeV) the main background contributions are due to multi-jet QCD events and the Z decaying into τ pairs. Beyond the Z mass (120-1500 GeV) the dominant source of additional dimuon production are $t\bar{t}$ and electroweak processes.

The different event topologies depend on the number of jets in the fiducial region of $p_T^{\text{jet}} > 30 \text{ GeV}$ and $|\eta^{\text{jet}}| < 4.5$. In Figure 4.8 (right) the jet multiplicity is presented. It is observed that the background contribution from top quark pair events becomes significantly large, when $N_{\text{jets}} \geq 2$. Therefore, it is of great importance to reduce the top quark pair background contribution as much as possible, while keeping the

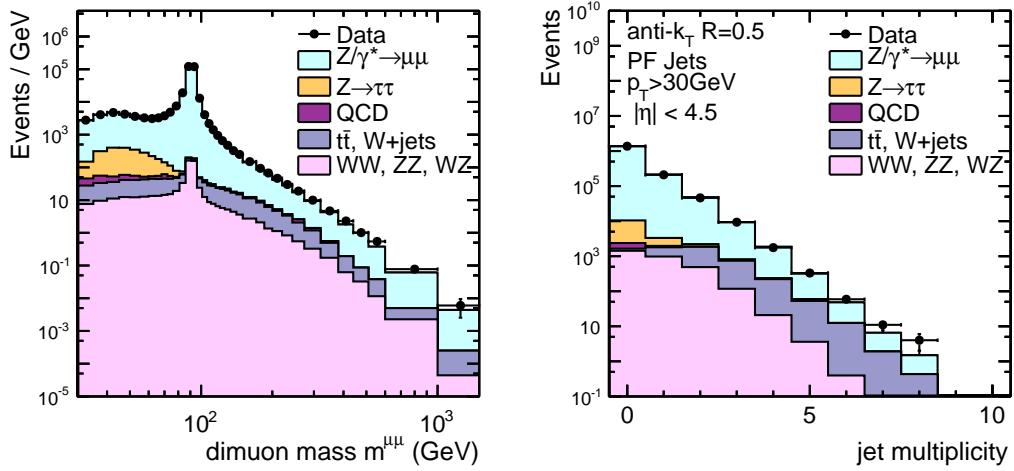


Figure 4.8: Dimuon invariant mass (left) and jet multiplicity (right). The different background processes are compared to signal MC events and data.

signal yields high.

In Figure 4.9 - 4.11 the dimuon $p_T^{\mu\mu}$ distribution in five different invariant mass ranges is shown for the inclusive DY, DY + 1 jet, and DY + 2 jets production respectively. To which extend the different background processes play a role in the different regions of phase space (low and high invariant mass, low and high p_T , inclusive production to DY + 2 jets) is presented in these figures.

In the following the simulation and normalisation of the background processes are described.

4.4.1 QCD Background Estimation from Data Events

Hard QCD processes in proton-proton collisions produce multi-jet events in the final state. The jets are mainly produced by semi-leptonic decays of charm and bottom quarks, and are in general accompanied by non-isolated leptons in the final state. The isolation requirement, described in section 3.3.2, reduces most of the QCD background contribution. The remaining fraction of QCD events, e.g. coming from misreconstructed jets identified as isolated leptons, is estimated using a method based on data events. A technique using data events is performed to calculate the number of QCD events.

The muon net charge is uniformly distributed in QCD events, thus the number of opposite-sign (OS) and same-sign (SS) dimuon events are the same in the QCD sample. To estimate the background contribution the data are separated according to the charge of the two muons in one opposite-sign ($\mu^+\mu^-$) and one same-sign ($\mu^+\mu^+$ or $\mu^-\mu^-$) sample. In Figure 4.12(a) and 4.12(b) the invariant mass distribution, before the isolation criterion is applied on the OS and SS dimuon data sample, are shown, respectively. It is observed that the SS dimuon sample is dominated by QCD processes. The SS sample can be used to estimate the QCD events of OS dimuons.

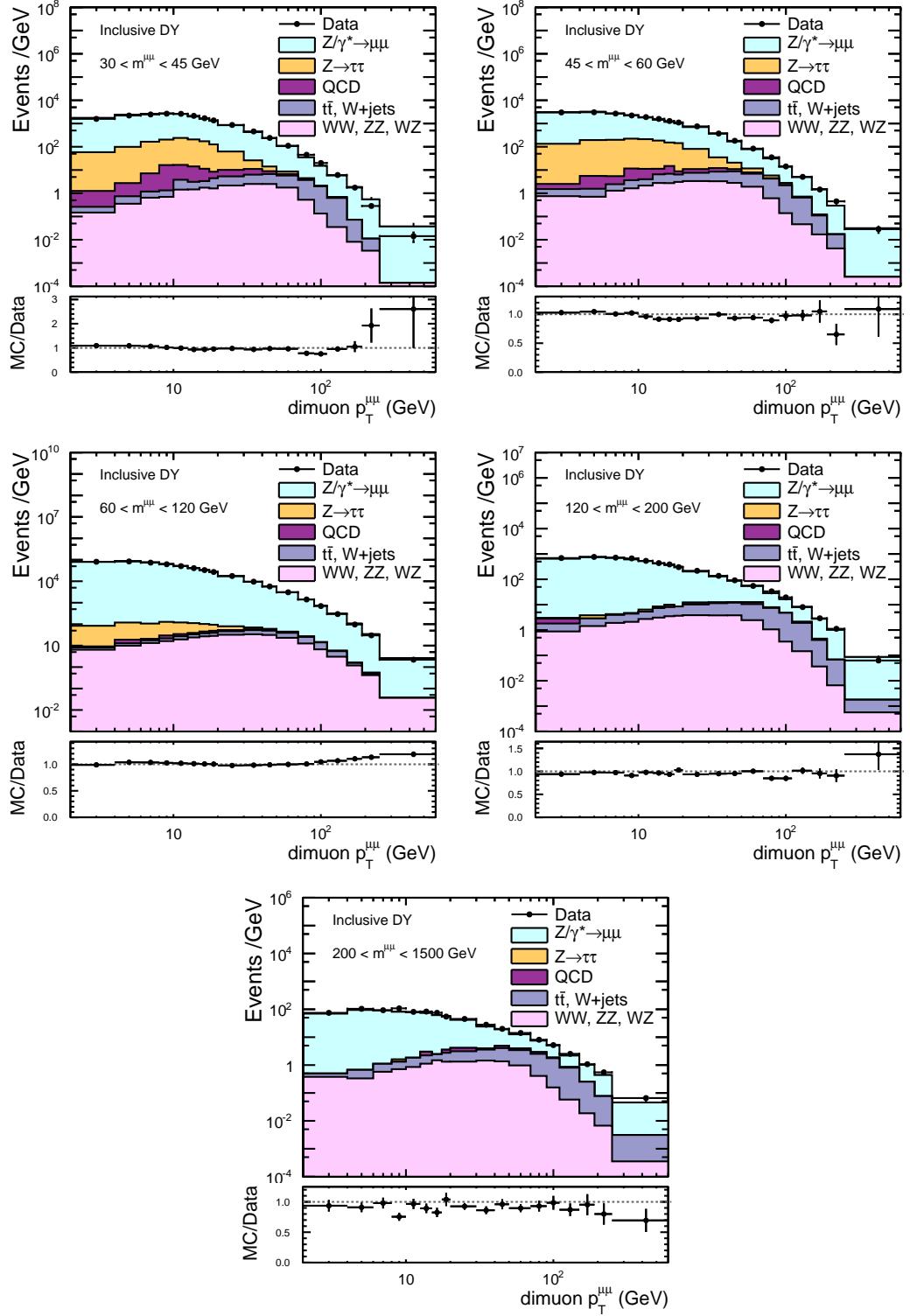


Figure 4.9: Dimuon $p_T^{\mu\mu}$ distribution (inclusive DY) in five different dimuon invariant mass ranges. The number of events are normalised to the binwidth.

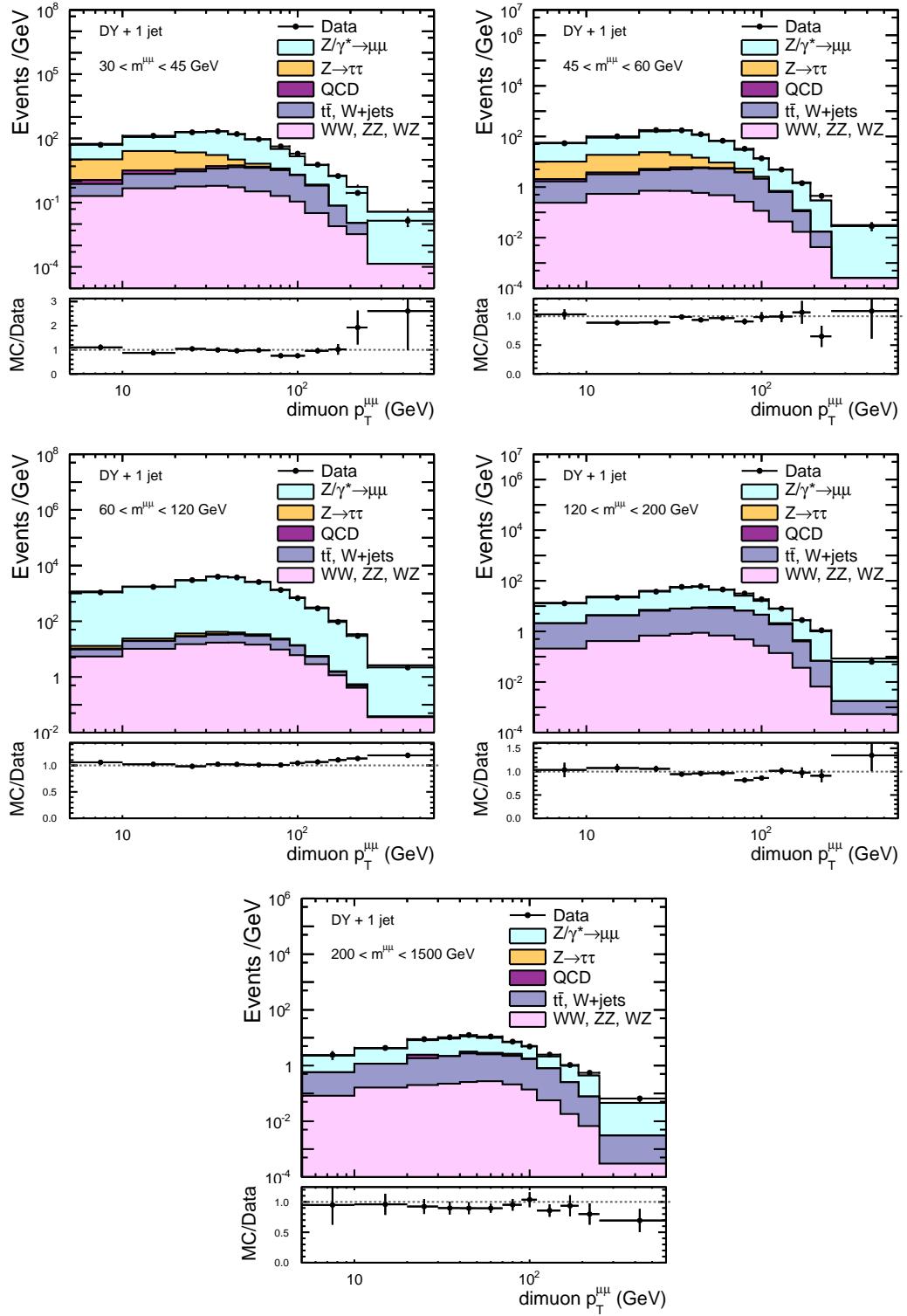


Figure 4.10: Dimuon $p_T^{\mu\mu}$ distribution (DY + 1 jet) in five different dimuon invariant mass ranges. The number of events are normalised to the binwidth.

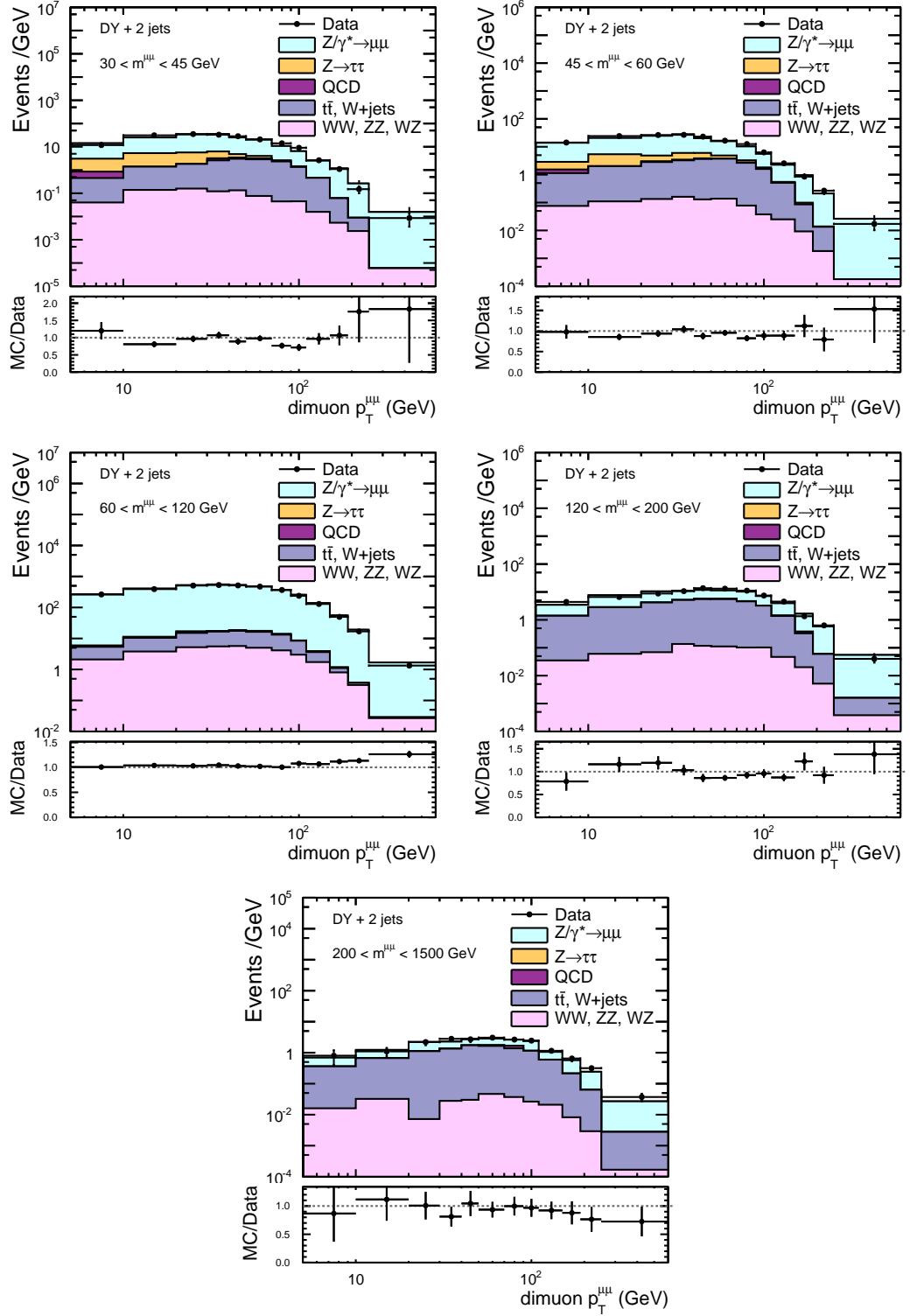


Figure 4.11: Dimuon $p_T^{\mu\mu}$ distribution ($DY + 2 \text{ jets}$) in five different dimuon invariant mass ranges. The number of events are normalised to the binwidth.

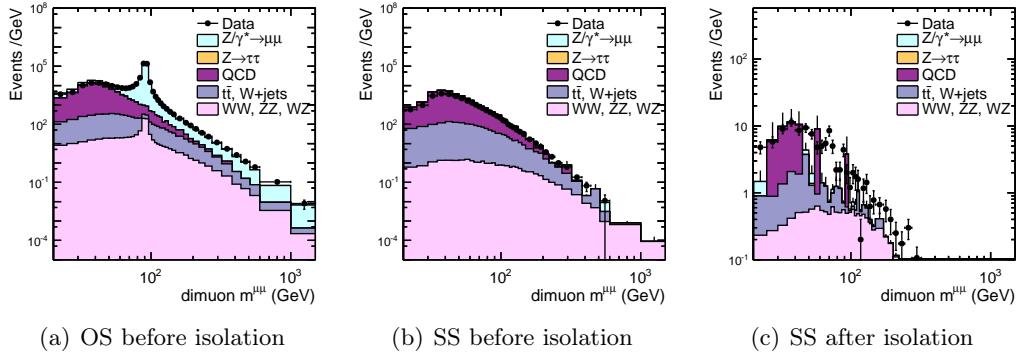


Figure 4.12: Invariant mass distribution for the opposite-sign (OS) and same-sign (SS) data samples. Figures (a) and (b) show the contribution before the isolation requirement. Figure (c) shows the remaining events in the same signed data sample after isolation requirement on the dimuon system.

The number of OS QCD events $N_{\text{QCD}}^{\text{OS}}$ are then estimated from the number of the SS data events $N_{\text{data}}^{\text{SS}}$, with isolated muons, by multiplying $N_{\text{data}}^{\text{SS}}$ with a correction factor to account for differences between the SS and OS sample

$$N_{\text{QCD}}^{\text{OS}} = N_{\text{data}}^{\text{SS}}(\text{Iso}_{\mu_{1(2)}} < 0.1(0.15)) \cdot R_{\text{QCD}}. \quad (4.3)$$

The correction R_{QCD} can not be calculated precisely in the signal-region, because the isolation requirement removes most of the QCD events, as shown in Figure 4.12(c). Thus, R_{QCD} is obtained in a region, which is defined as the QCD dominant region with non-isolated muons, by using an anti-isolation of $\text{Iso}_{\mu_{1,2}} > 0.5$. The correction factor can be calculated by the ratio of OS over SS data events, with anti-isolation

$$R_{\text{QCD}} = \frac{N_{\text{data}}^{\text{OS}}(\text{Iso}_{\mu_{1,2}} > 0.5)}{N_{\text{data}}^{\text{SS}}(\text{Iso}_{\mu_{1,2}} > 0.5)}. \quad (4.4)$$

The number of events in the SS and OS sample are studied for different anti-isolation requirements and the ratio R_{QCD} is calculated. The results are presented in Table 4.2. Furthermore, it is verified that the shape of the SS and OS sample, after anti-isolation selection on the two muons, is similar in data and MC simulation, as shown in Figure 4.13.

The number of OS events is approximately twice the number of SS events, therefore a correction factor of ~ 2 is applied to the number of SS data events. A cross-check with a MC QCD-enriched sample estimated by PYTHIA6 is performed. The value of the factor R_{QCD} is presented in Table 4.2 and is in agreement with data. An uncertainty on the ratio is obtained by taking the maximal difference between the nominal value, with $R_{\text{QCD}}(\text{Iso}_{\mu_{1,2}} > 0.5)$, and the ratios obtained by varying the isolation requirement between 0.2 and 0.4. The correction factor is estimated to be 1.98 ± 0.14 .

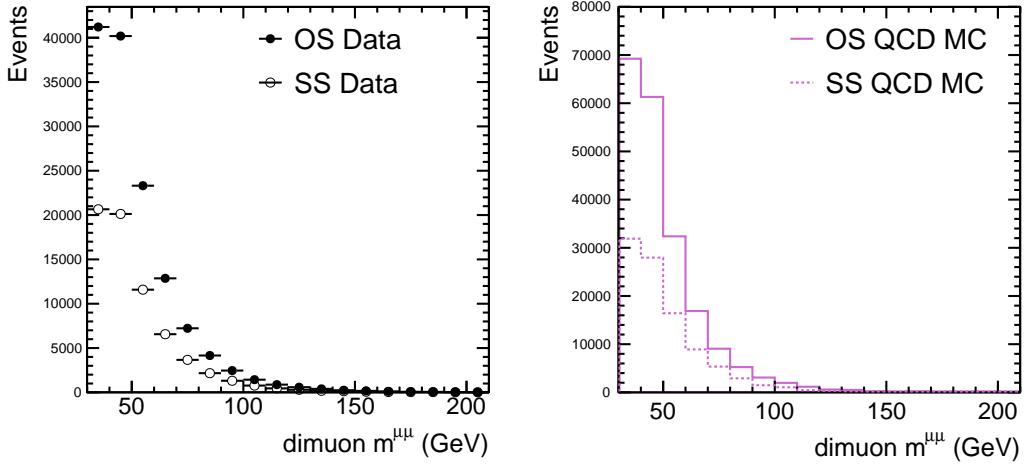


Figure 4.13: Opposite- and same-sign number of dimuon events as a function of the dimuon invariant mass in data (left) and MC events (right), with anti-isolation.

R_{QCD}	Iso > 0.2	Iso > 0.3	Iso > 0.4	Iso > 0.5
Data	2.12	2.07	2.02	1.98
MC	2.02	1.99	1.95	2.04

Table 4.2: Ratio of number of events in same-sign and opposite-sign dimuon samples with anti-isolation requirement in data and MC events.

4.4.2 Top Quark Pair Production

The top quark can decay into a bottom quark by radiation of a W boson, which can decay into a lepton and neutrino. The neutrino is not observed in the detector and can only be identified via missing transverse energy ($E_{\text{T}}^{\text{miss}}$). The $E_{\text{T}}^{\text{miss}}$ variable is defined as the energy imbalance in the transverse plane of the detector. This imbalance can occur due to particles, which do not leave energy in the calorimeter and pass the detector unobserved. The missing energy in top quark decays corresponds mainly to undetected neutrinos. The missing transverse energy is shown in Figure 4.14. The $t\bar{t}$ process is dominant in the region of $E_{\text{T}}^{\text{miss}} > 80$ GeV.

The top quark pair production is estimated from a simulated sample generated by MADGRAPH. The normalisation of the background contribution is taken from the CMS measurement of top quark pair production cross section [124] and a systematic uncertainty of 8% is assigned to the normalisation. Additionally the $t\bar{t}$ background process is controlled in the region with, $E_{\text{T}}^{\text{miss}} > 80$ GeV, where the $t\bar{t}$ contribution is dominant. A correction factor to take into account differences between data and MC events is derived by comparing the yields in the $t\bar{t}$ dominant region with

$E_T^{\text{miss}} > 80$ GeV, and is found to be 0.95 [97], which means the MC simulation is describing the data well.

A study of the top quark pair kinematics modelling the region of $E_T^{\text{miss}} > 80$ GeV is provided in appendix C. To reduce background contamination of $t\bar{t}$, a E_T^{miss} requirement of $E_T^{\text{miss}} < 80$ GeV is applied.

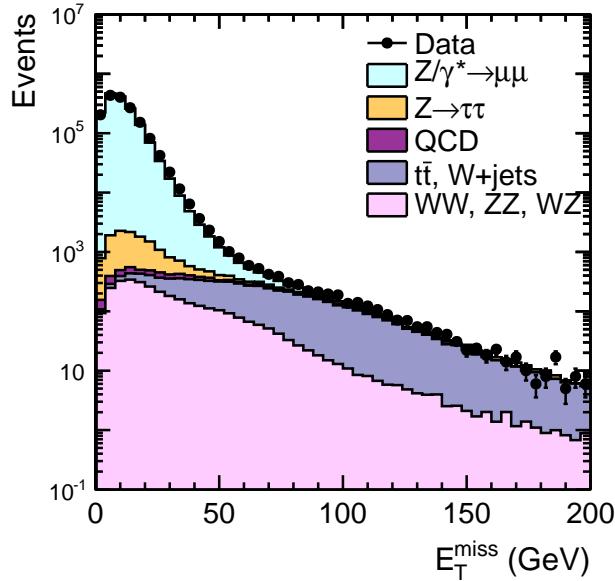


Figure 4.14: Distribution of missing transverse energy before applying the E_T^{miss} requirement. The $t\bar{t}$ background contribution is controlled in the region with $E_T^{\text{miss}} > 80$ GeV, where the top quark pair events dominate the sample.

4.4.3 Diboson and $W+$ Jets Production

Contributions from diboson and single W production are estimated using a simulated sample. The yields of diboson production are scaled to the cross sections predicted by theory at NLO [125, 126]. The contribution from single W production is normalised to the inclusive cross section as measured with CMS [127].

4.4.4 $Z \rightarrow \tau\tau$ Background Process

Another background process comes from Z decays into τ leptons, which subsequently decay into muons. The shape of $Z \rightarrow \tau\tau$ background process is estimated by MADGRAPH and normalised to the Z cross section measured with CMS [127].

4.4.5 Summary of Background Estimation

The relevant background and signal MC event yields are summarised in Table 4.3 and in Figure 4.15. The MC event yields are scaled to data luminosity and all efficiency corrections are taken into account. The background from QCD multi-jet events is calculated from data. The number of events are presented after different selection steps. It is observed that the signal-to-background ratio improves after the first selection criteria. Especially the isolation criterion reduces significantly the multi-jet background contribution and the E_T^{miss} requirement reduces mainly the $t\bar{t}$ background contribution. The last three columns of Table 4.3 show the event yields after the requirement of a forward Drell-Yan production ($|\eta| > 2.5$) in association with at least one or two jets with $p_T^{\text{jet}} > 30$ GeV. In Figure 4.15 the MC yields are compared to data events.

Selection	$Z^0/\gamma^* \rightarrow \mu\mu$	$Z \rightarrow \tau\tau$	QCD	$t\bar{t}, W + \text{jets}$	diboson
Muon η, p_T	1934183.5	12511.9	268080.0	23139.4	4138.4
Isolation	1660631.6	10049.0	1406.0	5268.5	3337.5
Mass	1626177.4	9942.2	1038.0	4838.3	3222.8
E_T^{miss}	1625992.4	9849.3	992.0	3277.3	3033.9
DY + 1 jet	262561.2	1730.5	262.0	3066.4	1607.6
DY + 2 jets	55744.2	375.0	90.0	2257.1	631.8
fwd DY	795275.4	3158.6	136.0	195.1	398.4
fwd DY + 1 jet	29567.3	182.8	16.0	182.3	148.9
fwd DY + 2 jets	6245.4	39.0	6.0	126.9	52.1

Table 4.3: Comparison of the background process event yields for different selection steps. The MC yield is scaled to the data luminosity and selection efficiencies are taken into account.

4.5 Particle Level Correction

The measurement is influenced by detector effects and the physics observables can not be detected directly. The detector components employ different methods to measure particles. The measurement depends on the area of the detector and the detector components. The particle also may produce secondary particles when interacting with the detector material. All these detector effects influence the measurement and affect the physical quantities under study.

There exist different sources of detector effects: first, the measurement is affected by the acceptance and the efficiency of the detector components. Second, migra-

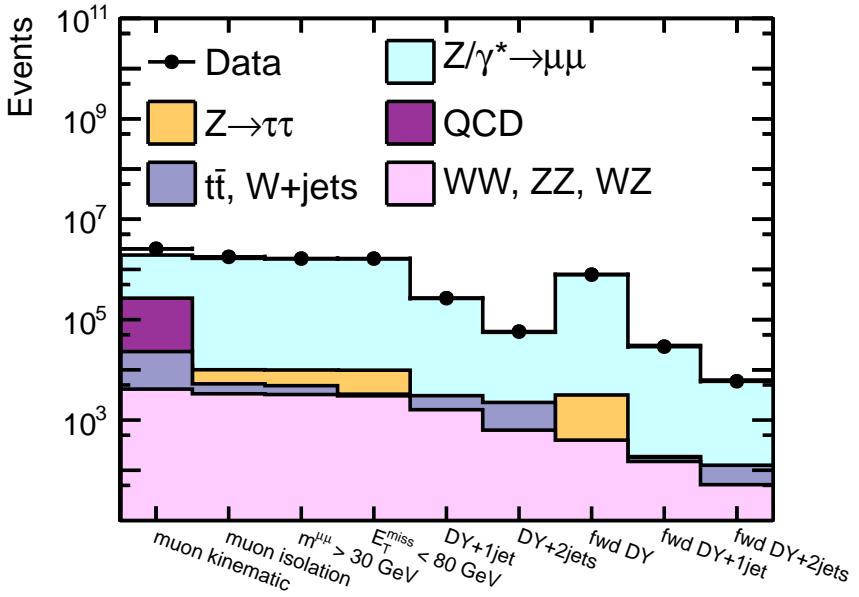


Figure 4.15: Data and MC yields after different selection steps. The number of MC events are scaled to data luminosity and selection efficiencies are taken into account.

tions of the measured observable between detector and particle level, and the limited resolution of the detector influence the measurement. In a differential cross section measurement the events are counted in a defined region of phase space, called bin. Due to the finite resolution, the physics observables can only be determined with finite precision, which becomes noticeable in migrations between bins. In order to perform an detector-independent measurement the bias due to the detector impact has to be reduced and remaining migration effects have to be recovered.

The measured quantities can be corrected for detector effects by taking the individual effects into account by an efficiency correction. However, a more advanced and practical method is to unfold the measured quantities to reproduce the actual true physics quantities. The relation between the measured and true physics quantities can be estimated from MC simulation. The simulation of detector effects is performed using the GEANT4 software (see section 2.3). Thus, in the unfolding method the true physics quantities can be estimated by inversely applying the detector response to the measured values.

The control distributions in the previous section illustrate that MADGRAPH MC predictions model well the distributions observed in data. The good agreement between data and MC events assures the MC simulation models well the detector response and the MADGRAPH sample can be used to unfold the observed number of data events from the detector effects.

The unfolding has to be derived for each variable and corresponds to the correction of detector inefficiencies, but also accounts for resolution and bin migrations of the observable under study. A correction to the total phase space, to account for

detector acceptance, is not applied. The migrations in dimuon $p_T^{\mu\mu}$ and $|\Delta y(\mu\mu, j_1)|$ are studied in the following sections with $Z^0/\gamma^* \rightarrow \mu\mu$ simulation by MADGRAPH.

The final results presented in this analysis refer to the fiducial region of $|\eta^\mu| < 2.1$ and $p_T > 20$ (10) GeV for the leading (subleading) muon. The impact of the pre-selection requirements on the dimuon $p_T^{\mu\mu}$ distribution is presented in Figure 4.16. The dimuon p_T distribution in the fiducial region is compared to the distribution obtained in the full phase space. Three invariant mass ranges are compared from 30 to 120 GeV. The effect of the acceptance is large at low $p_T^{\mu\mu}$ in the lowest invariant mass range 30 – 45 GeV, but it becomes smaller in the high $p_T^{\mu\mu}$ region. With increasing invariant mass the effect is negligible and for invariant masses above 60 GeV no impact is observed.

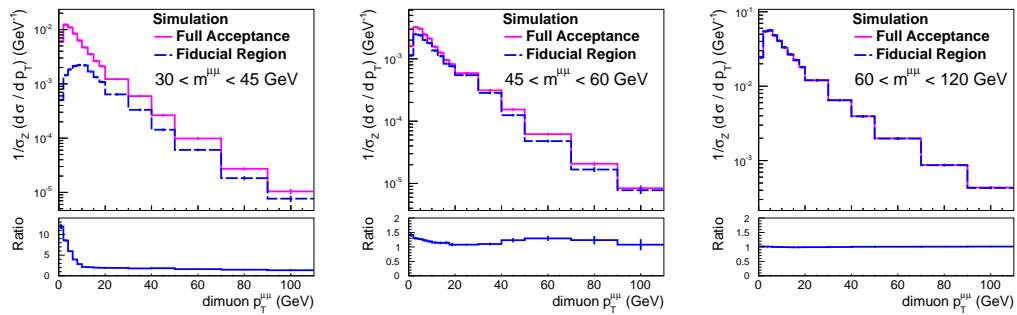


Figure 4.16: Dimuon $p_T^{\mu\mu}$ distribution comparing the fiducial region ($|\eta^\mu| < 2.1$ and $p_T > 20$ (10) GeV) and the full acceptance. The effect of an acceptance correction is presented for three invariant mass ranges.

4.5.1 Migration Study

This section is devoted to study the different migration effects, which can occur when measuring observables and mapping them to the corresponding MC truth variable. In the following I label the reconstructed observables with "reco" and the generated (from MC truth) with "gen" (which correspond to variables defined on particle level).

To study the effect of migrations the generator level and detector level events have to be matched by matching the jets in $\eta - \phi$ space with $\Delta R(\text{jet}^{\text{reco}}, \text{jet}^{\text{gen}}) < 0.4$. The matched jets are defined as "good" jets [112], when they have a matched generator jet above $p_T > 8$ GeV.

There are several relevant sources of migrations affecting the dimuon $p_T^{\mu\mu}$ and the $|\Delta y(\mu\mu, j_1)|$ distributions:

M1 Signal migrations between generator and detector level bins

M2 Migrations into the phase space from background

M3 Migrations outside the phase space

Migrations in p_T or $|\Delta y|$ due to limitations in the binning (M1) of the reconstructed and generated variables can occur. It can happen that e.g. a dimuon on generator level appears in the p_T bin i and on detector level the reconstructed p_T appears in bin k ($k \neq i$) and vice versa. Variables to study the bin migrations are purity and stability of a bin i . Purity is defined as

$$P_i = \frac{N_{\text{matched}}(E_{\text{reco}} \in i \wedge E_{\text{gen}} \in i)}{N_{\text{matched}}(E_{\text{reco}} \in i)}, \quad (4.5)$$

with the number of events on generator and detector level, E_{gen} and E_{reco} , respectively. Stability is defined as

$$S_i = \frac{N_{\text{matched}}(E_{\text{reco}} \in i \wedge E_{\text{gen}} \in i)}{N_{\text{matched}}(E_{\text{gen}} \in i)}, \quad (4.6)$$

referring to the number of generated events $E_{\text{gen}} \in i$, which remain also on detector level $E_{\text{reco}} \in i$. Purity and stability quantify to which extend the measured variable is sensitive to migration effects [128].

Further migrations can come from sources as noted in M2 and M3. The defined jet p_T threshold of 30 GeV can introduce migrations inside and outside the fiducial region. For example, a generated jet has $p_T^{\text{gen}} = 35$ GeV, but due to detector effects the jet p_T^{reco} is reconstructed with < 30 GeV. Thus, the event corresponds on generator level to the event selection, but is not selected on detector level. Concluding, due to migrations around the jet p_T threshold some fraction of events at detector level do not have a corresponding jet on generator level and vice versa. To which extend the migrations inside and outside the phase space influence the observables can be interpreted by background and acceptance defined as

$$B_i = 1 - \frac{N_{\text{matched}}(E_{\text{reco}} \in i)}{N_{\text{all}}(E_{\text{reco}} \in i)}, \quad A_i = \frac{N_{\text{matched}}(E_{\text{gen}} \in i)}{N_{\text{all}}(E_{\text{gen}} \in i)}. \quad (4.7)$$

The background variable B_i of bin i illustrates the relevance from migration into the phase space and acceptance A_i of bin i shows how many events on generator level can be identified with an event after the reconstruction. N_{matched} refers to the number of events with matching conditions, while N_{all} takes all events into account. Small migration effects are revealed by small background contributions and high acceptance values.

In the following the migration effects due to detector resolution and contributions from migrations inside and outside the phase space are discussed separately.

Detector Resolution

A reconstructed $p_T^{\mu\mu}$ can appear in a bin different from its truth $p_T^{\mu\mu}$ due to limits in the detector resolution. Thus, the bin size is chosen to be larger than the resolution to limit migration effects to neighbouring bins. Still migration effects remain and bias the measured distribution. In the case of small migration effects the purity is large and the detector resolution effects can be taken into account by an efficiency correction. However, for a low bin-purity large migration effects between neighbouring bins can occur and a more advance unfolding technique is needed.

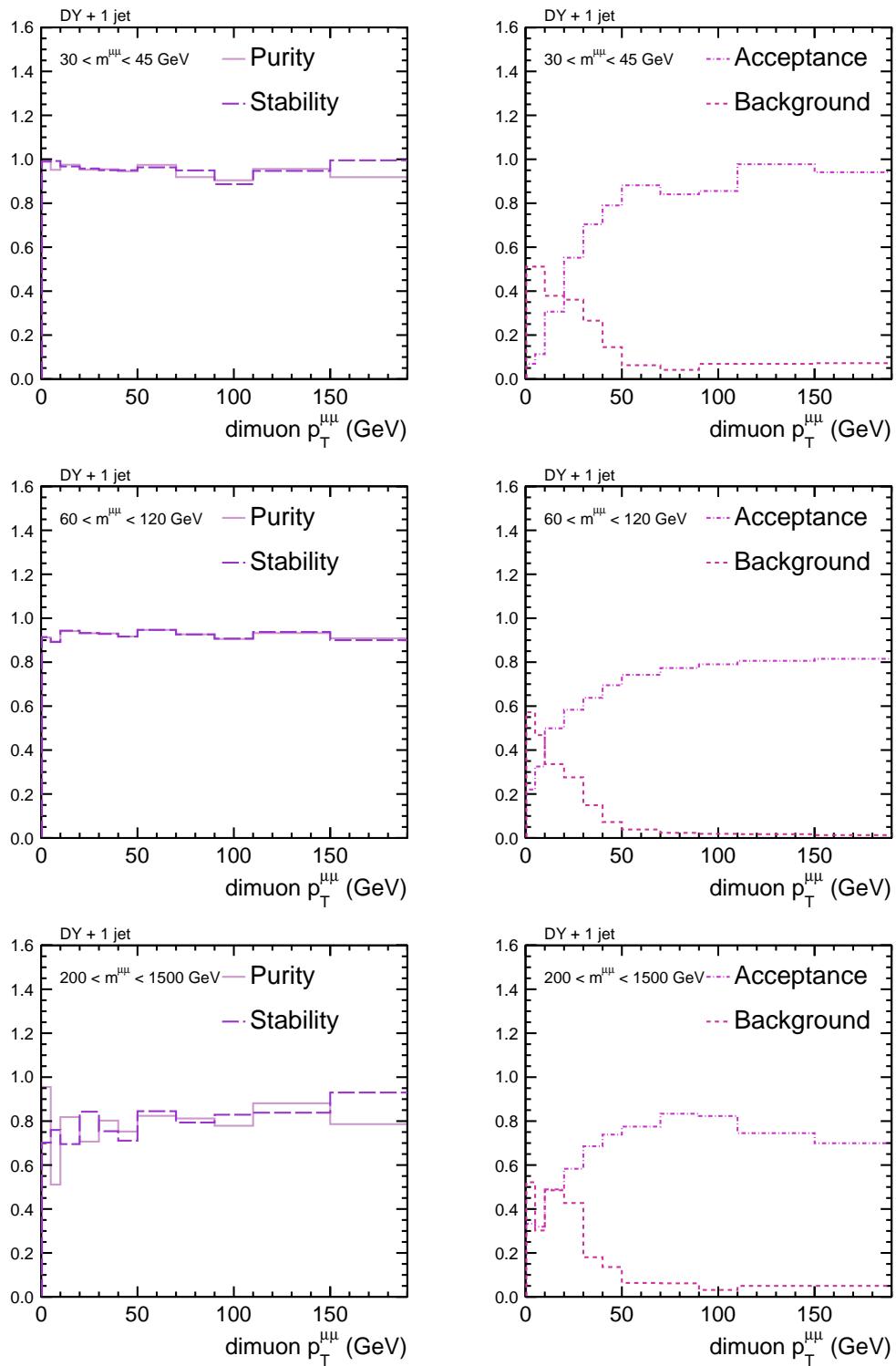


Figure 4.17: Purity, stability, background and acceptance for $DY + 1 \text{ jet}$ transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass ($30 - 45$, $60 - 120$, $120 - 1500 \text{ GeV}$).

The purity of the dimuon transverse momentum distribution for DY + 1 jet production is shown in Figure 4.17. For small $m^{\mu\mu}$ the purity is around $> 95\%$ and for larger $m^{\mu\mu}$ the bin-purity decreases to $\sim 70 - 90\%$. This behaviour is similar also for the other invariant mass ranges, and for the inclusive DY and DY + 2 jets production (cf. appendix D). All in all, the purity is $> 60\%$ for all different production processes. Additionally, the bin-stability is calculated. It is observed that the stability of the dimuon $p_T^{\mu\mu}$ is similar to the purity. Thus, the dimuon $p_T^{\mu\mu}$ distribution reveals small bin migrations.

In the case of the differential cross section in $|\Delta y(\mu\mu, j_1)|$ the purity and stability is typically $> 80\%$ for the different mass ranges. In Figure 4.18 the purity and stability as a function of $|\Delta y(\mu\mu, j_1)|$ are shown in the case of DY + 1 jet production. At small $|\Delta y(\mu\mu, j_1)|$ the bin-purity and stability are close to one, while with increasing rapidity separation the migrations increase with purity and stability to $\sim 80\%$.

Pileup Jets

The migrations inside and outside the phase space can be studied with the acceptance and background variables (eq. (4.7)). The variables are presented in Figure 4.17 and 4.18 for DY + 1 jet as a function of $p_T^{\mu\mu}$ and $|\Delta y(\mu\mu, j_1)|$. In the case of the dimuon $p_T^{\mu\mu}$ it is observed that the contribution from background is $< 10\%$ for $p_T^{\mu\mu} > 50$ GeV and $\sim 20 - 50\%$ for $p_T^{\mu\mu} < 50$ GeV. Small migrations are illustrated for high $p_T^{\mu\mu}$ in large values of A , while at low $p_T^{\mu\mu}$ the migrations outside the phase space increase. In the case of the $|\Delta y(\mu\mu, j_1)|$ observable, migrations from background are $> 60\%$ for low invariant mass and $|\Delta y| > 4$. With increasing invariant mass the migration effects reduce. The acceptance at small $m^{\mu\mu}$ is $< 30\%$, which means the migrations outside the phase space are not negligible and have to be accounted for in the unfolding procedure. With increasing invariant mass the acceptance increases to 60% .

The reasons for the migrations can come from two different sources. First, jets can come from pileup interactions and appear in the phase space after reconstruction, second, migrations around the p_T threshold introduce a bias.

Migrations into the phase space due to the jet p_T migrations and pileup are treated in independent ways. To decouple the two effects four different event classes are compared:

- all reconstructed jets with $p_T^{\text{reco}} > 30$ GeV
- matched jets with $p_T^{\text{reco}} > 30$ GeV and $p_T^{\text{gen}} > 30$ GeV
- matched jets with $p_T^{\text{reco}} > 30$ GeV and $p_T^{\text{gen}} < 30$ GeV
- no matched jet on generator level to the reconstructed jet with $p_T^{\text{reco}} > 30$ GeV

The first class corresponds to the inclusive class, where no further requirement on the generator jet is applied. The second class refers to reconstructed jets, with a matched jet on generator level. The third class correspond to the fraction of events,

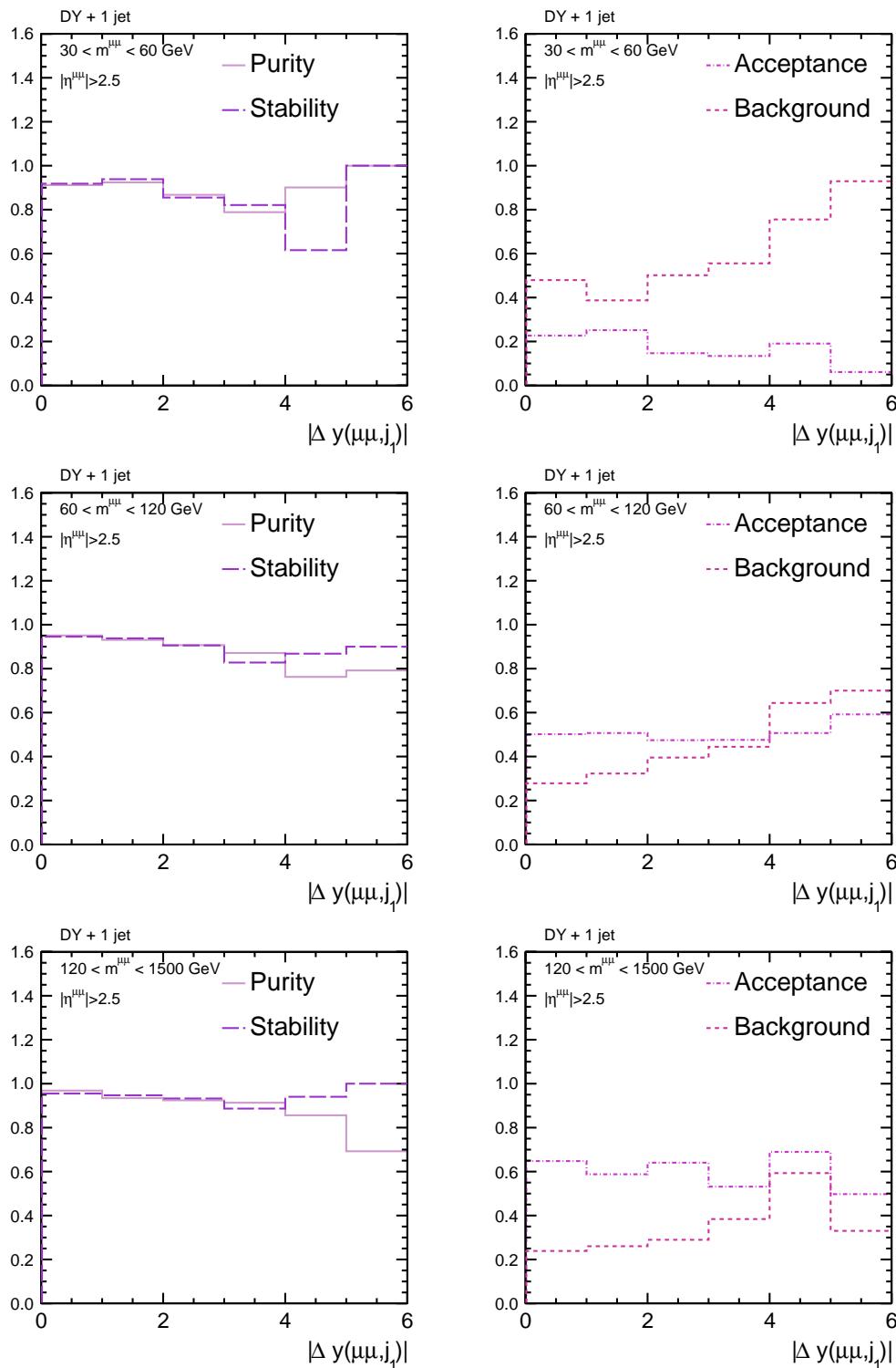


Figure 4.18: Purity, stability, acceptance and background for the DY + 1 jet production as a function of $|\Delta y(\mu\mu, j_1)|$ in three bins of the invariant mass ($30 - 60$, $60 - 120$, $120 - 1500$ GeV).

which have an impact of migrations around the jet p_T threshold. Finally, the fourth class refers to the fraction of events, which are affected from pileup jets.

In Figure 4.19 the jet transverse momentum, pseudorapidity and multiplicity is plotted for the four classes. The ratio is taken to the inclusive class where all reconstructed jets are selected above threshold. Small migrations in p_T are visible in the region of $p_T < 50$ GeV, and pileup jets show an impact of $< 10\%$. The effect from pileup jets is negligible in the central region but in the forward region $|\eta| > 2.5$ it increases up to $20 - 30\%$. This behaviour can be explained, due to the reason that, the tracking information is only existent in the central region. In the forward region the pileup jet rejection can not be applied. The jet multiplicity plot presents that up to four jets can be identified as pileup jets in the event. However, only a small fraction of events are due to migrations, with $\sim 15\%$ and $\sim 5\%$ for p_T^{jet} migrations and pileup jets respectively for $N_{\text{jets}} = 1$. For higher jet multiplicities the fraction is negligible.

The different classes are further studied in the cases of dimuon transverse momentum and $|\Delta y|$ of the Drell-Yan and the leading jet in different invariant mass ranges in Figure 4.20 - 4.23.

The dimuon $p_T^{\mu\mu}$ for DY + 1 jet and DY + 2 jets in the invariant mass ranges of 30 – 45 GeV, 60 – 120 GeV, and 120 – 200 GeV are shown in Figure 4.20 and 4.21, respectively. In the first two dimuon $p_T^{\mu\mu}$ bins (0 – 10 GeV) the contribution from pileup jets is the largest. In the low invariant mass ranges (30 – 45 GeV and 45 – 60 GeV) and at low $p_T^{\mu\mu}$ the contribution from pileup jets is around 30 – 40%. In all the other bins in $p_T^{\mu\mu}$ and $m^{\mu\mu}$ the effect is below 20%. The migration effect in and out of the fiducial region is of the order of 30% at low dimuon $p_T^{\mu\mu}$. For $p_T^{\mu\mu} > 40$ GeV the migration effects are $< 20\%$. The migration effects and fraction of pileup jets reduce with going from DY + 1 jet to DY + 2 jets production.

The comparison of the different effects for the $|\Delta y|$ distribution in the case of DY + 1 jet and DY + 2 jets are shown in Figure 4.22 and 4.23 respectively. The effect from pileup jets in $|\Delta y|$ becomes important for $|\Delta y| > 4$. In this region the contribution is $\sim 40\%$ and increases with $m^{\mu\mu}$. To reduce the impact of pileup jets the cross section measurement in $|\Delta y|$ is performed in six bins up to $|\Delta y| = 6$. The migrations around the p_T threshold show no dependence on $|\Delta y(\mu\mu, j_1)|$. The effect is $\sim 20\%$ and reduces to 15% for $200 < m^{\mu\mu} < 1500$ GeV. In DY + 2 jets production the migrations decrease to $< 10\%$.

The jets coming from pileup interactions and the effects due to jet p_T migrations into the fiducial region are taken into account in the unfolding method.

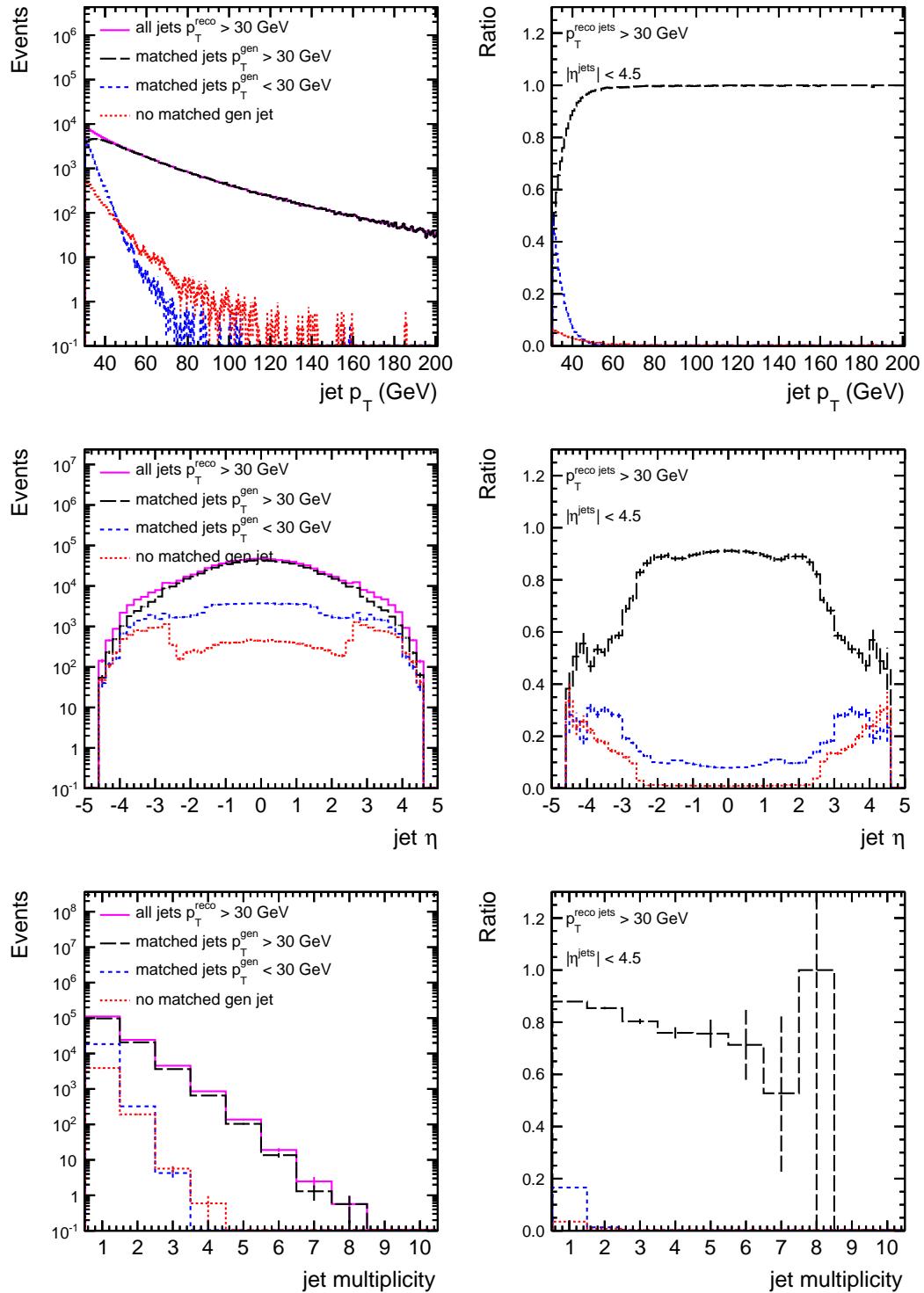


Figure 4.19: Jet control distributions comparing different event classes to investigate the migration effects due to the jet p_T and pileup jets. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.

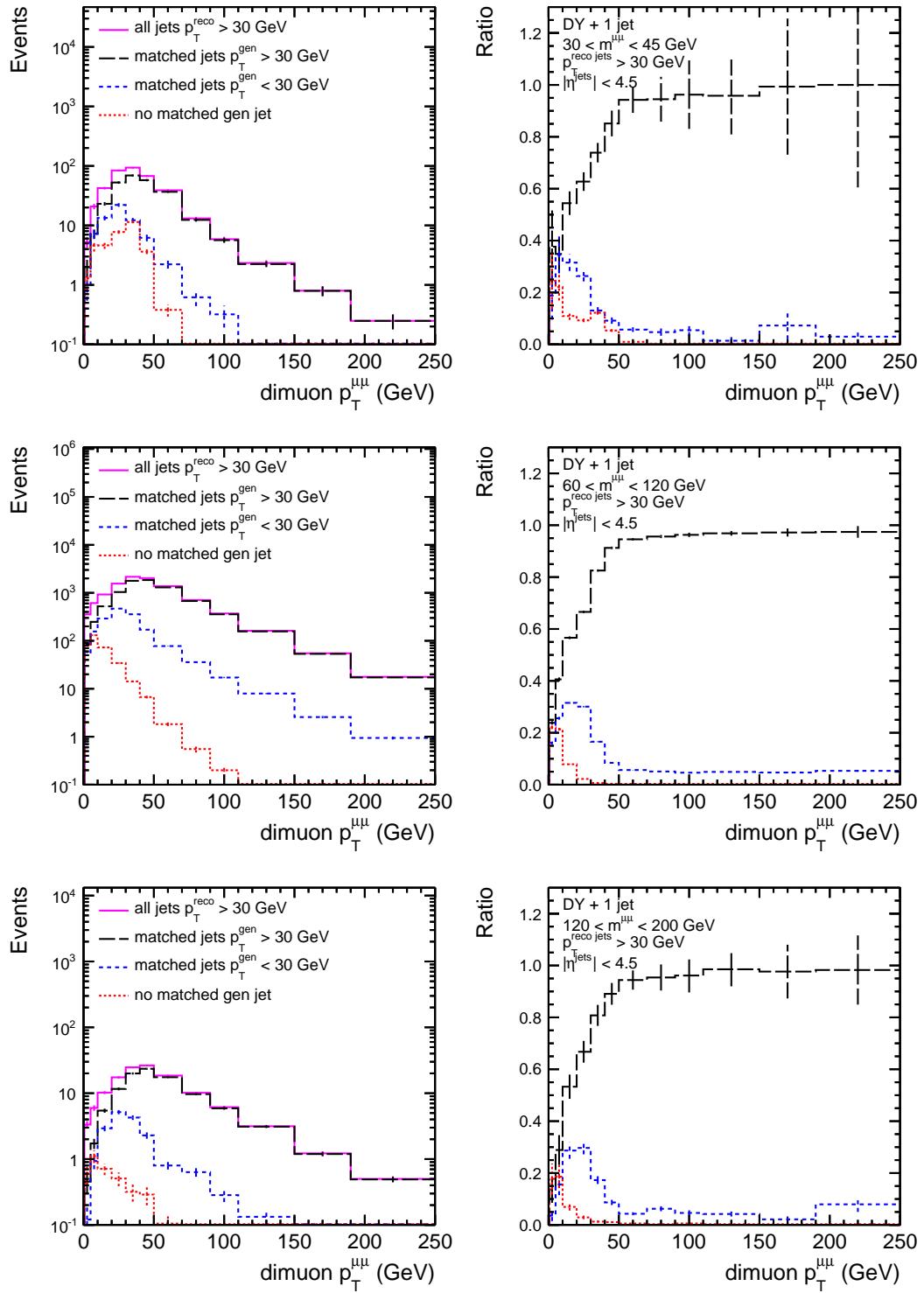


Figure 4.20: Dimuon transverse momentum for the DY + 1 jet case in three different invariant mass ranges. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.

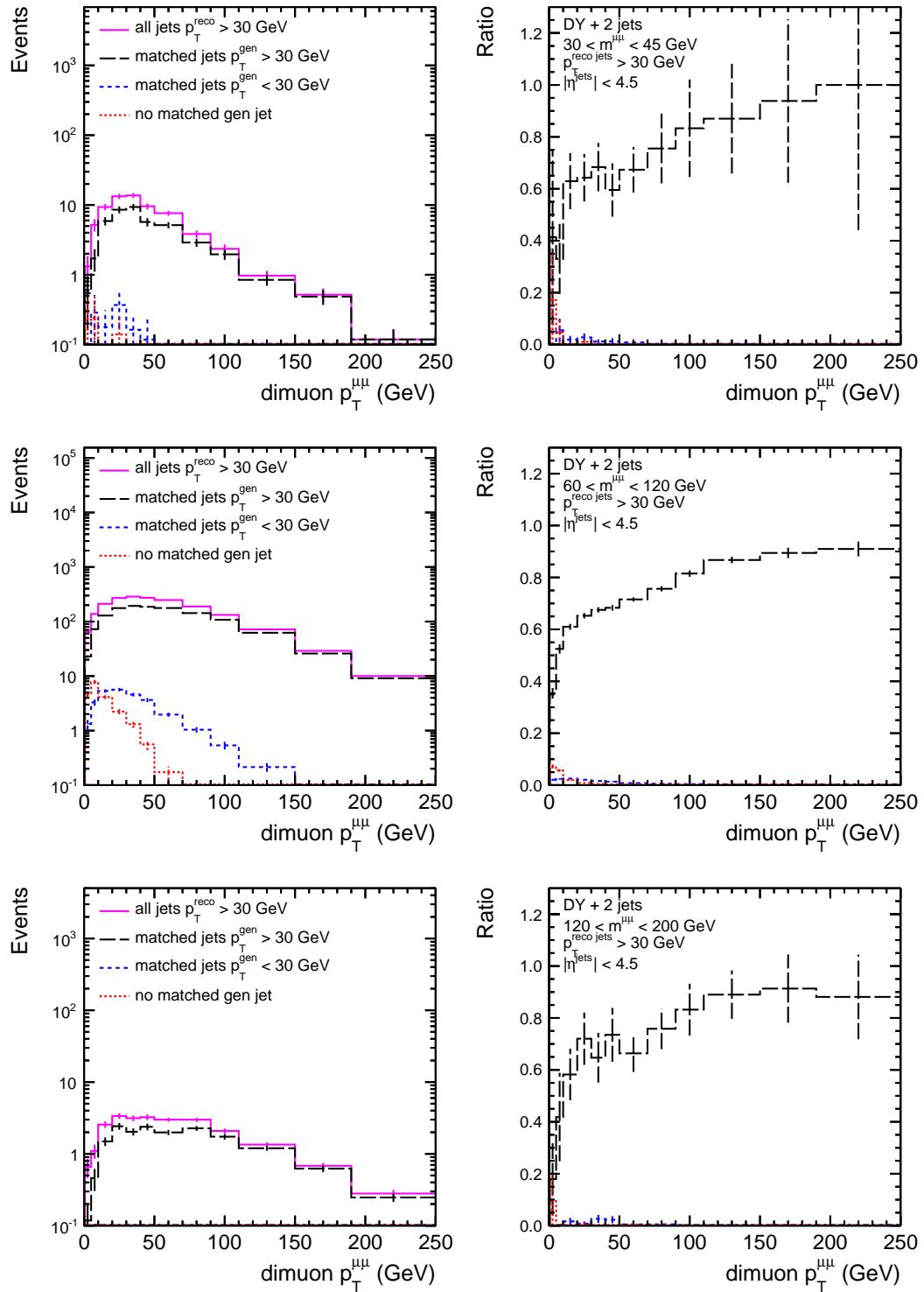


Figure 4.21: Dimuon transverse momentum for the DY + 2 jets case in three different invariant mass ranges. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.

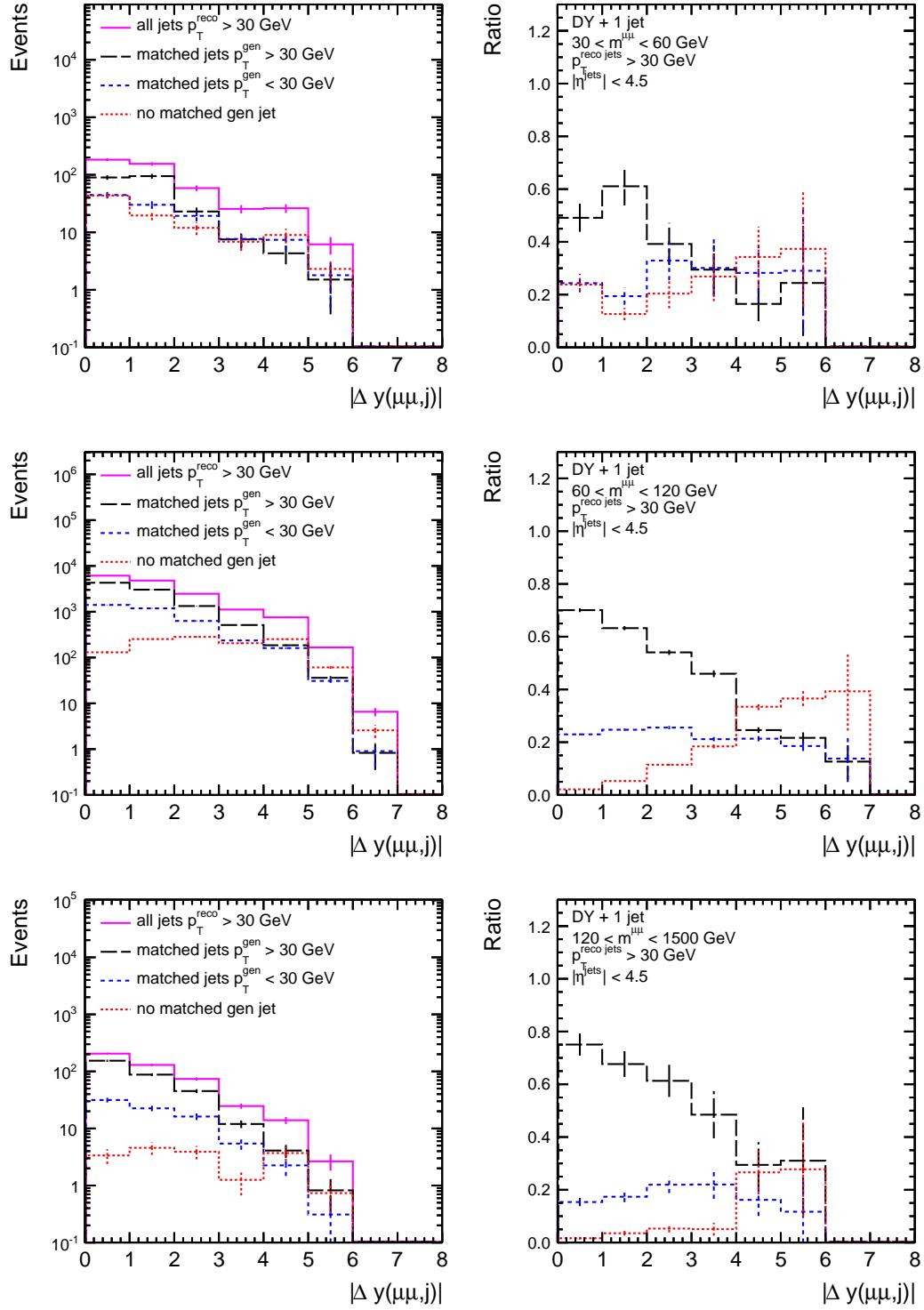


Figure 4.22: Absolute rapidity separation between the leading jet and the forward Drell-Yan is shown for the different event topologies for DY + 1 jet production in three different invariant mass ranges. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.

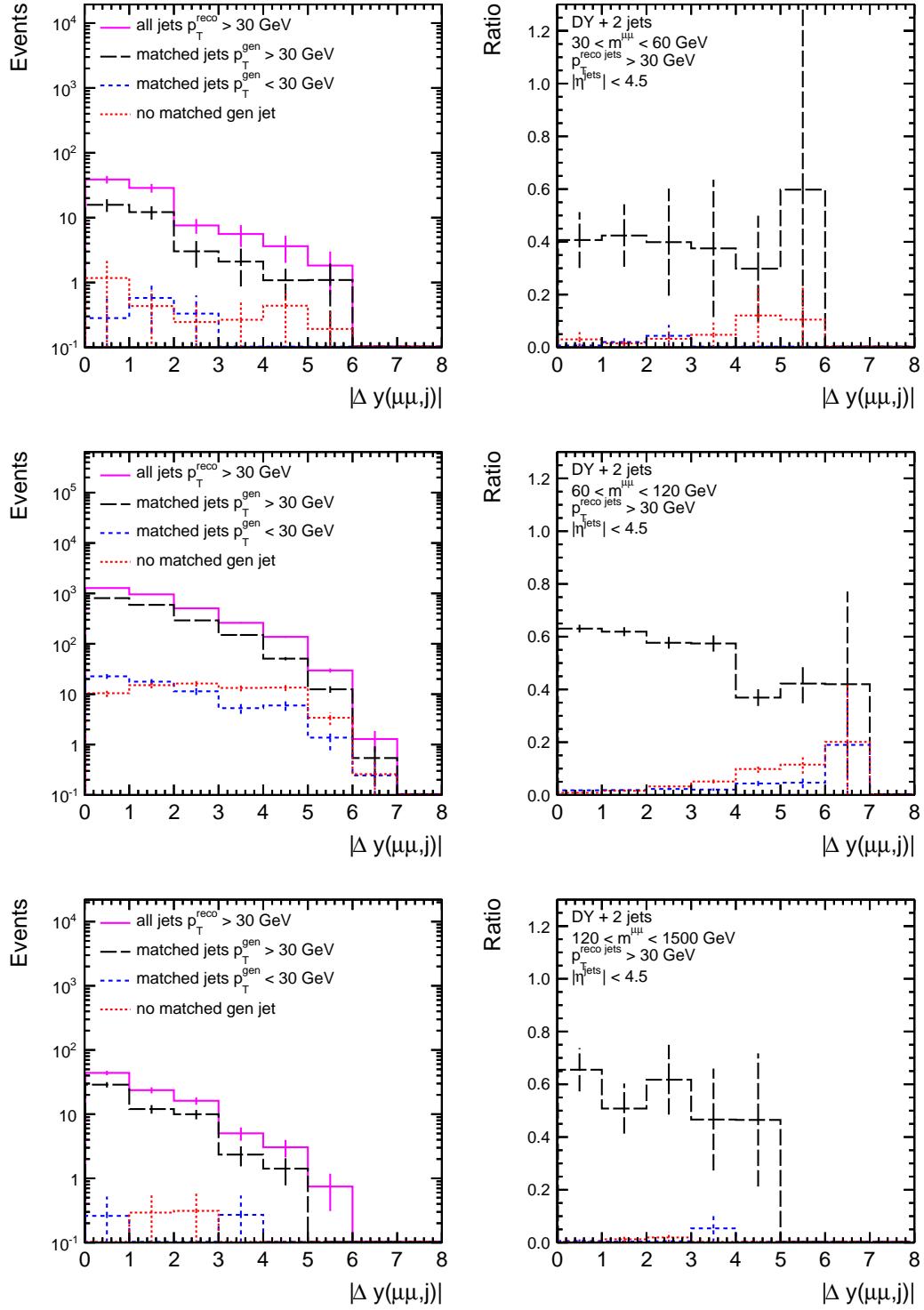


Figure 4.23: Absolute rapidity separation between the leading jet and the forward Drell-Yan is shown for the different event topologies for DY + 2 jets production in three different invariant mass ranges. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.

4.5.2 Unfolding

Unfolding is defined as a technique to decouple the physical distribution from detector effects to obtain the best estimation of the true distribution. The unfolding method used in this analysis is based on the Bayes' theorem and was developed by D'Agostini [129].

In general the detector response can be represented by a two-dimensional matrix (response matrix) $(C_{i,j})_{\substack{i=1,\dots,n \\ j=1,\dots,m}}$. In the following the original values of the physics quantities are referred to as "true". The number of true events x^{true} in bin j can be connected to the measured events y^{det} in bin i by [128]

$$y_i^{\text{det}} = \sum_{j=1}^m C_{ij} x_j^{\text{true}}, \quad i = 1, \dots, n. \quad (4.8)$$

The response matrix includes amongst other the migrations of type M1, which can be represented as a sub-matrix C^{mig} of the response matrix C , as illustrated in Figure 4.24. In the following the full detector response matrix is labelled with response matrix and the sub-matrix including bin migrations from detector to particle level and vice versa is called migration matrix. The diagonal elements of the migration matrix provide the probability to reconstruct a measured observable correctly within its true bin.

Furthermore, migrations of type M2 and M3 are included in the response matrix to account for migrations into and outside the phase space. Background events (as defined in eq. (4.7)) can be handled by adding additional bins to the matrix. These additional contributions are defined as follows: "miss" events are defined as signal contribution, which do not appear in any detector level bin, and "fake" events reveal the contribution, when the detector level contribution has no corresponding signal on generator level. An illustration of the 2d migration matrix including the fake and miss contribution is shown in Figure 4.24. The x and y axes represent the generator and detector level bins. The overflow and underflow bins are filled with the fake and miss events, while the 2d migration matrix C^{mig} reveals the bin migrations.

The normalised migration matrices of the dimuon $p_T^{\mu\mu}$ for the three event topologies of two invariant mass ranges (30 – 45 GeV and 120 – 200 GeV) are shown in Figure 4.25. The diagonal elements represent the fraction of correctly matched events on generator and detector level. The off-diagonal elements illustrate the amount of migrations. The off-diagonal elements show migrations less than 10%. In addition to the high bin-purity and stability, the migration matrix indicates small migration effects in the transverse momentum variable. The corresponding response matrices for the other invariant mass ranges are shown in appendix D. The normalised response matrices for $|\Delta y(\mu\mu, j_1)|$ for two invariant mass ranges (30 – 60 GeV and 120 – 1500 GeV) for DY + 1 jet and DY + 2 jets production is presented in Figure 4.26. It is observed that migrations appear over a wide range in $|\Delta y|$. Although the fraction of migrated events is < 30% in the DY + 1 jet production the migration in $|\Delta y|$ can occur over the whole phase space. In the case of DY + 2 jets production the fraction of migrated events increases to $\sim 50\%$ in the low invariant mass range. In order to take the migrations into account an advanced unfolding technique

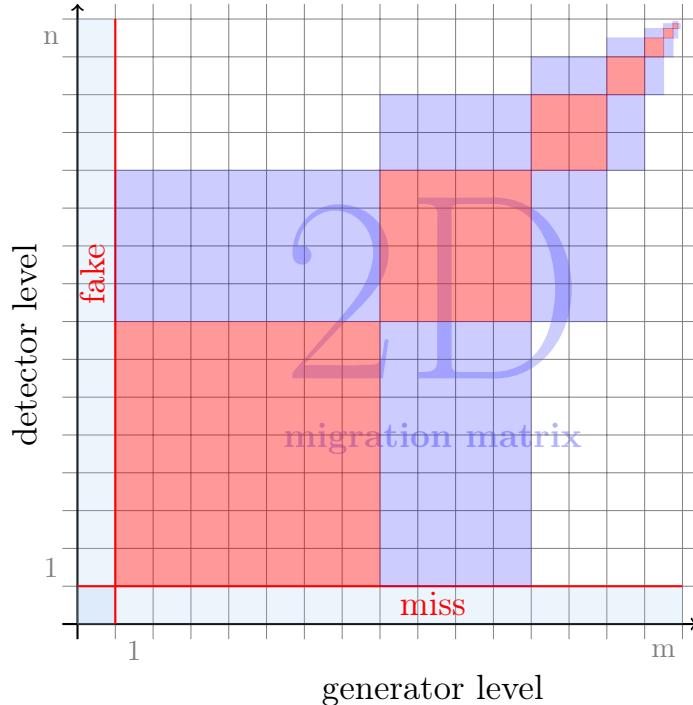


Figure 4.24: Illustration of the unfolding response matrix. Contributions from fake and miss events are considered in the unfolding procedure. The effect of bin migrations is shown in the 2d matrix.

is employed. The theoretical basic information and the mathematical and physical implementation is described in the following.

Bayes Unfolding

In general, to solve eq. (4.8) the response matrix has to be inverted, which is mathematically possible if $n = m$ and $\det C \neq 0$. However, complications in the inversion of the unfolding matrix can appear, if the matrix is singular or if large fluctuations complicate the calculation.

An easy calculation of a simple correction factor c_k to account for detector effects can be obtained with the bin-by-bin method. The correction factor is derived from the ratio

$$c_k = \frac{\sum_{i=1}^n C_{ik}}{\sum_{j=1}^m C_{kj}}. \quad (4.9)$$

by taking the ratio of the sum of generator bin values over the sum of detector bin values. The factor can than be applied to the measured value in bin k . However, this method can strongly depend on the MC generator, which is used to fill the response matrix. Further, this simple bin-by-bin method is not reliable if contributions from background are large and it only yields an appropriate unfolded result if migrations are small [128].

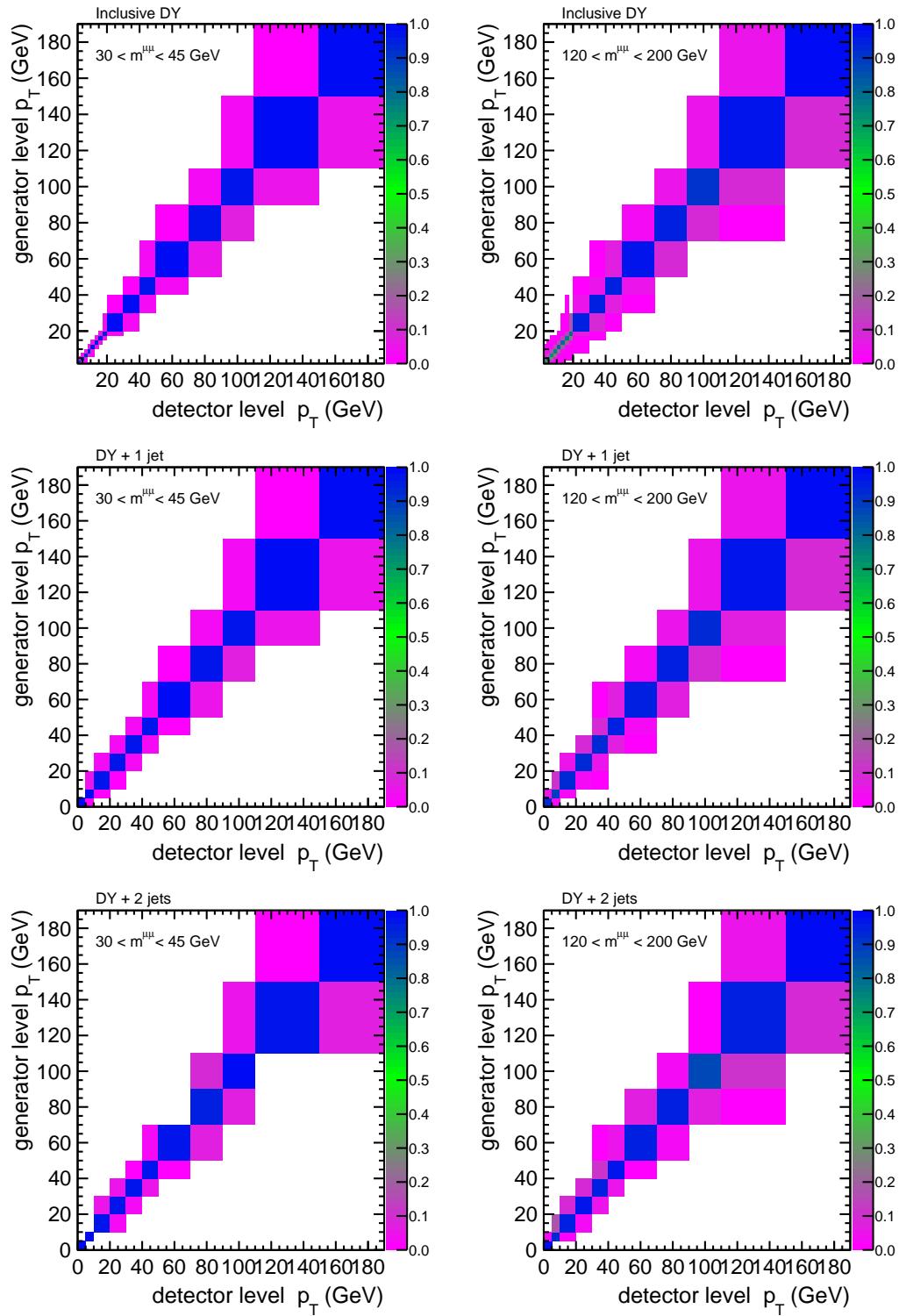


Figure 4.25: The normalised migration matrix of the dimuon $p_T^{\mu\mu}$ distribution for the low ($30 - 45$ GeV) and high ($120 - 200$ GeV) invariant mass range. The inclusive DY, DY + 1 jet, and DY + 2 jets are compared.

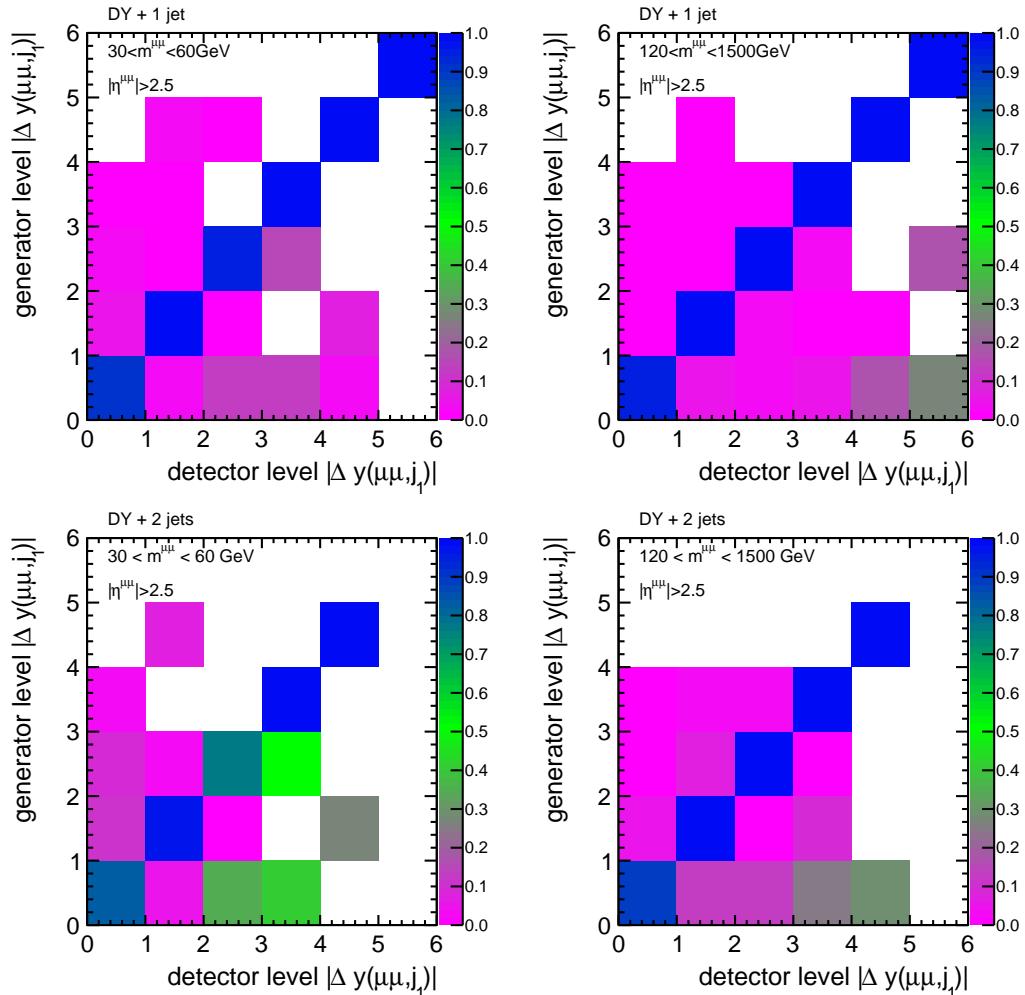


Figure 4.26: The normalised migration matrix of $|\Delta y(\mu\mu, j_1)|$ for the low ($30 - 60 \text{ GeV}$) and high ($120 - 1500 \text{ GeV}$) invariant mass range. The $DY + 1 \text{ jet}$, and $DY + 2 \text{ jets}$ are compared.

In this analysis an iterative approach, proposed by D'Agostini in [129], based on the Bayes' theorem, is used. In the following I will give a brief explanation of the method proposed in [129]. The notations and formulas are taken from [129, 130].

The Bayes theorem is defined in terms of causes C_i and effects E_j . The effect E_j can be caused by several causes C_i , but without the knowledge of the exact cause. In our measurement the cause corresponds to the true value T_i in bin i and the effect refers to the reconstructed value R_j in the detector level bin j . The aim of the unfolding is to estimate the number of expected data events in the true bin $n(T_i^{\text{data}})$, which is given by

$$n(T_i^{\text{data}}) = \frac{1}{\epsilon_i} \sum_{j=1}^{n_R} n(R_j^{\text{data}}) \cdot P(T_i^{\text{MC}}|R_j^{\text{MC}}), \quad (4.10)$$

with the number of events in the detector level bin j , $n(R_j^{\text{data}})$. The probability that the true value T_i results in the reconstructed value R_j yields

$$P(T_i^{\text{MC}}|R_j^{\text{MC}}) = \frac{P(R_j^{\text{MC}}|T_i^{\text{MC}}) \cdot P_0(T_i^{\text{MC}})}{\sum_{l=1}^{n_T} P(R_l^{\text{MC}}|T_l^{\text{MC}}) \cdot P_0(T_l^{\text{MC}})}, \quad (4.11)$$

and is estimated from MC simulation. The sum in eq. (4.10) runs over the number of detector level bins, n_R , and the sum in eq. (4.11) runs over the number of true bins, n_T . $P(R_j^{\text{MC}}|T_i^{\text{MC}})$ is the probability that the reconstructed value R_j^{MC} is caused by T_i^{MC} and $P_0(T_i^{\text{MC}})$ is an initial probability, also called prior. The efficiency ϵ_i is defined as

$$\epsilon_i = \sum_{j=1}^{n_R} P(R_j^{\text{MC}}|T_i^{\text{MC}}), \quad 0 \leq \epsilon_i \leq 1. \quad (4.12)$$

and corresponds to the efficiency that the true value T_i can come from any detector level bin. Eq. (4.10) can be written in terms of the response matrix, using eq. (4.11) and (4.12), as

$$n(T_i^{\text{data}}) = \sum_{j=1}^{n_R} M_{ij}^{\text{MC}} \cdot n(R_j^{\text{data}}) \quad (4.13)$$

with

$$M_{ij}^{\text{MC}} = \frac{P(R_j^{\text{MC}}|T_i^{\text{MC}}) \cdot P_0(T_i^{\text{MC}})}{\sum_{l=1}^{n_R} P(R_l^{\text{MC}}|T_i^{\text{MC}}) \sum_{l=1}^{n_T} P(R_l^{\text{MC}}|T_l^{\text{MC}}) \cdot P_0(T_l^{\text{MC}})}. \quad (4.14)$$

The number of data events in the true bin $n(T_i^{\text{data}})$ can be calculated by multiplying the response matrix M_{ij}^{MC} with the number of events in the detector level bin $n(R_j^{\text{data}})$. To relate eq. (4.8) with eq. (4.13) the matrix M corresponds to the inverse of the matrix C .

The probabilities $P(R_j^{\text{MC}}|T_i^{\text{MC}})$ can be estimated from migration, efficiency and resolution using the response matrix estimated from MC generator predictions. Thus, the iterative method only requires the unfolding matrix (4.14) estimated from MC simulation. The result of the method is independent of the initial probability $P_0(T_i^{\text{MC}})$, which is used as a starting point of the iteration to calculate the matrix (4.14). The distribution of $P_0(T_i^{\text{MC}})$ is arbitrary (e.g. flat or estimated from MC simulation) and is replaced each iteration by an updated distribution [131]. The iterative procedure follows the steps

1. Starting point is the initial probability $P_0(T^{\text{MC}})$, estimated e.g. from MC predictions, and calculate the initial number of expected events from the total number of observed events N_{obs} by $n_0(T^{\text{data}}) = P_0(T^{\text{MC}}) \cdot N_{\text{obs}}$.
2. Calculate $n_1(T^{\text{data}})$ with eq. (4.13) and (4.14) using the information from the response matrix. The probabilities $P(R^{\text{MC}}|T^{\text{MC}})$ are obtained from the MC generator and $P_0(T^{\text{MC}})$ is taken as prior.
3. Calculate the updated prior distribution $P_1(T)$ with $P_1(T) = \frac{n_1(T^{\text{data}})}{N_{\text{true}}}$, where N_{true} is the total number of events in the true bins. Calculate $n_2(T^{\text{data}})$ with eq. (4.13), but replace $P_0(T^{\text{MC}})$ with $P_1(T)$ in eq. (4.14).
4. Compare $n_2(T^{\text{data}})$ with $n_1(T^{\text{data}})$ by performing a χ^2 fit. If the value of χ^2 is small enough the algorithm can be stopped. Otherwise continue with the program from point 3.

The D'Agostini iterative procedure does not depend on the prior distribution and thus, only uses the response matrix simulated with the MC generator. Employing the iterative method the output converges to the true data distribution.

The implementation of the Bayes unfolding is performed within the RooUnfold package [132]. A regularisation parameter, to control the iterative unfolding, is defined as the number of iterations. In this analysis the same number of true and detector level bins are used ($n_T = n_R$). The signal data events in the final distributions refer to unfolded results using the MADGRAPH MC sample to model the response matrix.

Consistency Test

Tests can be performed in order to check if the unfolding procedure using MADGRAPH provides a reasonable result. The consistency test is a MC based cross check and the method and results are presented in the following. I use the $Z^0/\gamma^* \rightarrow \mu\mu$ MADGRAPH MC predictions to model the measured as well as the MC truth predictions. In the consistency check it is validated if the results after unfolding can reproduce the MC truth predictions.

First the $Z^0/\gamma^* \rightarrow \mu\mu$ MADGRAPH MC sample is used to calculate the response matrix. In addition the MADGRAPH predictions are used to estimate the "measured distribution". The Bayes unfolding is performed and the unfolded result is compared to the values from MADGRAPH on generator level. The results are shown in Figure 4.27. The MC truth distribution is compared to the measured, and unfolded events. Overall the unfolded events can reproduce the MC truth events within statistical uncertainties. The same result is obtained for the other invariant mass ranges and for the DY + 1 jet and DY + 2 jets topologies. From this one can conclude that the unfolding procedure is correctly used and the results are stable.

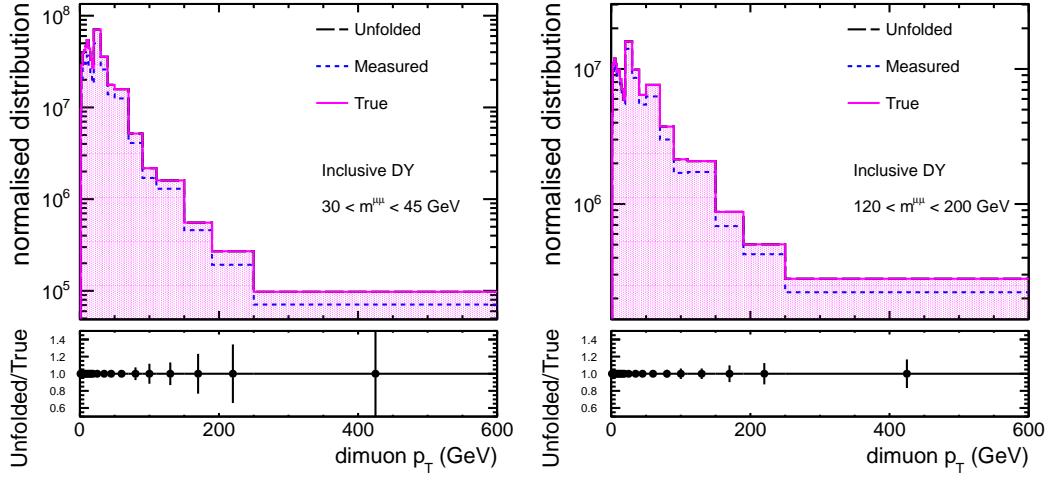


Figure 4.27: Comparing measured, MC truth, and unfolded dimuon $p_T^{\mu\mu}$ using MADGRAPH for the response matrix and for the measured distribution.

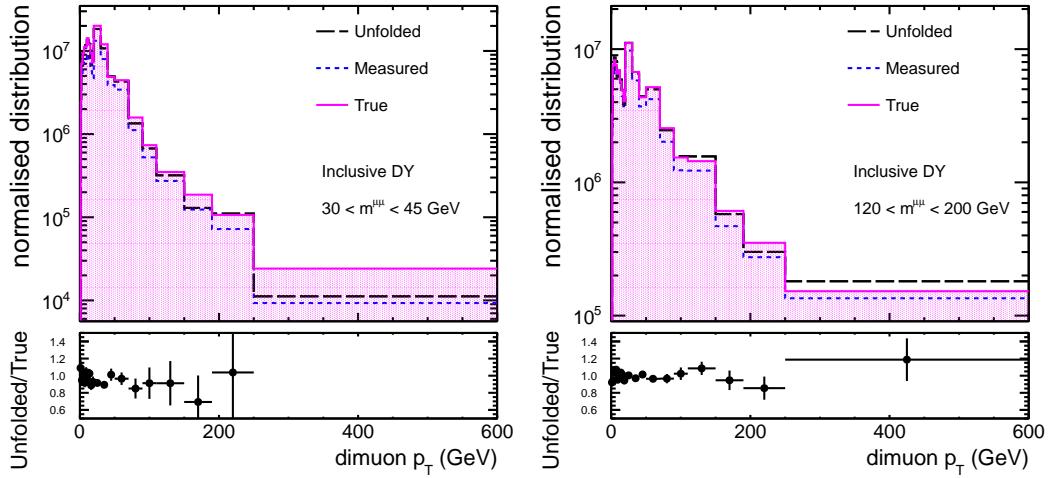


Figure 4.28: Comparing measured, MC truth, and unfolded dimuon $p_T^{\mu\mu}$ using two statistically independent samples of MADGRAPH.

In the next step the MC sample is split into two statistically independent samples. One sub-sample is used to estimate the response matrix and the other sub-sample is used as "measured distribution". The results from Bayes unfolding are shown in Figure 4.28. At low p_T in the low invariant mass range (30 – 45 GeV) the unfolded result overestimate the MC truth values. In the high invariant mass range (200 – 1500 GeV) the unfolded values agree with the MC truth predictions within statistical uncertainties. Some fluctuations are seen due to limits in statistics when splitting the MADGRAPH sample. Similar results are obtained for the $|\Delta y(\mu\mu, j_1)|$ distribution. For the unfolding of the signal data events the total MADGRAPH sample is used.

The consistency tests provide evidence of allowing to use MADGRAPH MC generator predictions for unfolding the final distributions. Furthermore, the test provides a verification of the accuracy of the Bayes unfolding method.

4.6 Cross Section Definition

The differential cross section is measured as a function of the variable $X = p_T^{\mu\mu}$, $|\Delta y(\mu\mu, j_1)|$ by counting the number of events in the defined region of phase space using a limited binning in X_i , $i = 1, \dots, N_{\text{bins}}$. To obtain the cross section value in bin i the corrected number of signal events has to be divided by the total luminosity of $\mathcal{L} = 4.9 \text{ fb}^{-1}$ and the binwidth ΔX_i . The differential cross section in bin i is then obtained by the formula

$$\frac{d\sigma}{dX_i} = \frac{N_i^{\text{sig,unfolded}}}{\Delta X_i \cdot \mathcal{L}}, \quad (4.15)$$

where $N_i^{\text{sig,unfolded}}$ corresponds to the number of signal events after unfolding in bin i

$$N_i^{\text{sig,unfolded}} = \sum_{j=1}^{N_{\text{bins}}} C_{ij}^{-1} \times (N_j^{\text{obs}} - N_j^{\text{bkg}}). \quad (4.16)$$

The number of signal events is obtained by subtracting the number of background events, N_i^{bkg} , which are estimated from data and MC simulation, from the number of observed events, N_i^{obs} . The matrix multiplication in eq. (4.16) represents the unfolding method to correct for bin migrations and detector resolution and efficiency.

The cross section calculation, as presented in eq. (4.15), is performed in five bins of the dimuon invariant mass $m^{\mu\mu}$. The final results are presented as normalised distributions, by normalising the differential cross section by the total integrated cross section in the Z invariant mass range ($60 < m^{\mu\mu} < 120 \text{ GeV}$)

$$\sigma_Z = \sum_{j=1}^{N_{\text{bins}}} \frac{d\sigma}{dX_j} \cdot \Delta X_j, \quad (4.17)$$

with respect to the three topologies, inclusive DY, DY + 1 jet, and DY + 2 jets. The dependence on the luminosity cancels by taking the ratio. The normalised cross sections are

$$\left(\frac{1}{\sigma_Z} \right) \frac{d\sigma}{dp_T^{\mu\mu}} \quad \text{and} \quad \left(\frac{1}{\sigma_Z} \right) \frac{d\sigma}{d|\Delta y(\mu\mu, j_1)|}. \quad (4.18)$$

4.7 Systematic Uncertainties

An experimental measurement is always specified with a measurement uncertainty. The data points are given with a statistical and systematic uncertainty. The latter characterises the precision of the measurement or measured value. Reasons for systematic uncertainties on measurements are given by several causes, e.g. influences from the measurement devices, or environmental conditions. Some uncertainties can be reduced by calibration of the detector components, but not all uncertainties can be eliminated. In general the impact on the measurement composition is considered in the data analysis.

In this section the main sources for systematic uncertainties on data are described. The MC generator MADGRAPH is used to estimate the uncertainties. The final results in section 5.1 and 5.2 show the cross section ratio of eq. (4.18). The ratio to σ_Z reduces the systematic uncertainty, due to the cancellation of the uncertainty on the luminosity.

Furthermore, a theory uncertainty on the choice of scale parameters (μ_F and μ_R) is estimated by POWHEG ($Z + 1$ jet) MiNLO [133] with the PDF set of HERA-PDF1.5NLO [134] and the tune P0 [74].

4.7.1 Experimental Uncertainties

Jet Energy Scale and Jet Energy Resolution

The jet energy correction with respect to the JES and JER are known with a limited precision. Thus, the jet energy correction can influence the fraction of selected events in the fiducial region with $p_T^{\text{jet}} > 30$ GeV and $|\eta^{\text{jet}}| < 4.5$. The correction of the jet energy leads into a systematic uncertainty on the cross section. An uncertainty between 3 – 5% on the jet energy scale is assigned depending on the transverse momentum and the pseudorapidity of the jet [116]. The score resolution scaling factor (cf. eq. (3.10)), which was defined as the ratio of data over MC resolution, was estimated with a precision of 5 – 15%, depending on the jet η [117]. The total uncertainty on the JEC and JER is obtained by varying the JEC and the scaling factor within the limits of uncertainty for the MC generator predictions. This directly translates into an uncertainty on the measured cross section by using the scaled MADGRAPH predictions for the unfolding matrix. The JEC and JER uncertainty is the dominant uncertainty source in the distributions including jets.

Model Uncertainty

To correct for detector effects and bin migrations the MADGRAPH +PYTHIA6 predictions are used to fill the response matrix and model the migration effects. MADGRAPH to data comparisons provide already a good agreement on detector level and thus, MADGRAPH can be used to model the response matrix. However, the theory model

does not necessarily describe the same migrations observed in data. Therefore, an uncertainty on the input model, which is used to fill the response matrix, is needed.

The migration effects depend on the shape of the distribution. A flat distribution reveals small migrations, while a steeply falling distribution predicts larger migrations. The theory, to be used to correct for detector effects, should already describe the data on detector level, in order to predict the shape and thus the migrations in an optimal way.

To estimate an uncertainty on the model, a parametrisation of the data points is employed. The parametrisation gives an estimate on how the shape of the distribution can vary within the statistical uncertainty of data. The parameters are given with an uncertainty taking into account the statistical uncertainty on the data. By changing the fit parameters within their uncertainty, one can estimate to which extend the shape of the distribution can vary but still being consistent with data.

The parametrisation function is chosen in order to describe the data on detector level. In case of the dimuon $p_T^{\mu\mu}$ distribution, the low $p_T^{\mu\mu}$ region is parametrised by a Gaussian function,

$$g(p_T^{\mu\mu}) = c \cdot e^{-\frac{1}{2} \left(\frac{p_T^{\mu\mu} - m}{s} \right)^2}, \quad (4.19)$$

and the $p_T^{\mu\mu}$ tail is parametrised by an exponential function,

$$e(p_T^{\mu\mu}) = e^{p_0 + p_1 \cdot p_T^{\mu\mu}}. \quad (4.20)$$

The parameters c, m, s (which represent a constant, mean, and sigma respectively) and p_0 and p_1 are obtained by the fitting procedure.

In case of the $|\Delta y(\mu\mu, j_1)|$ distribution, the data is parametrised by a linear function

$$l(|\Delta y|) = p_0 \cdot |\Delta y| + p_1. \quad (4.21)$$

The uncertainty on the fit parameters are used to reweight the MADGRAPH +PYTHIA6 predictions, according an event weight defined as $\omega_{\text{model}} = \frac{p^\pm}{p^{\text{nominal}}}$. p^\pm indicates the parametrisation with varying the fit parameters up and down their uncertainties and p^{nominal} refers to the parametrisation with nominal fit parameters. The nominal predictions from MADGRAPH and the model uncertainty obtained from the reweighted results show good agreement to data, as presented in Figures 4.29 and 4.30. The data values are compared to the MADGRAPH +PYTHIA6 predictions with model uncertainties. To propagate the model uncertainty on the cross section result, the results of the unfolding are compared using a response matrix estimated by varying the MADGRAPH prediction up and down its model uncertainty.

Pileup Reweighting

The pileup distribution in data is estimated from the luminosity information and the inelastic proton-proton cross section. The inelastic proton-proton cross section is estimated with a precision of 5% by CMS [105]. The uncertainty on the inclusive cross section influences the pileup modelling in the data distribution and is taken into account in the analysis. The uncertainty is obtained by varying the inelastic cross section by $\pm 5\%$ around the nominal value.

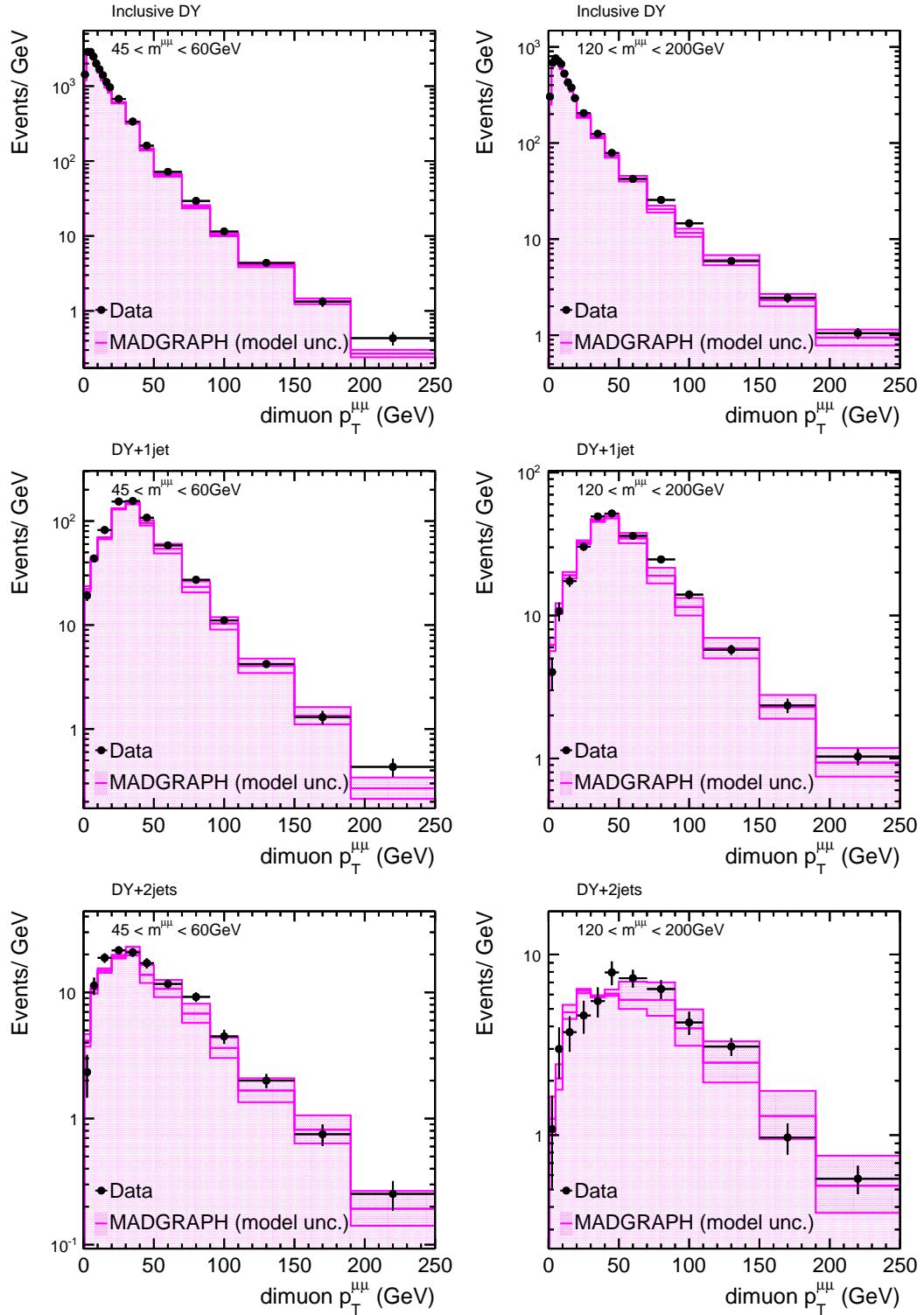


Figure 4.29: The $p_T^{\mu\mu}$ distribution for low and high invariant masses in the inclusive DY, DY + 1 jet, and DY + 2 jets topology. Data is compared to MADGRAPH + PYTHIA6 predictions plus a model uncertainty derived from the uncertainties on the data parametrisation.

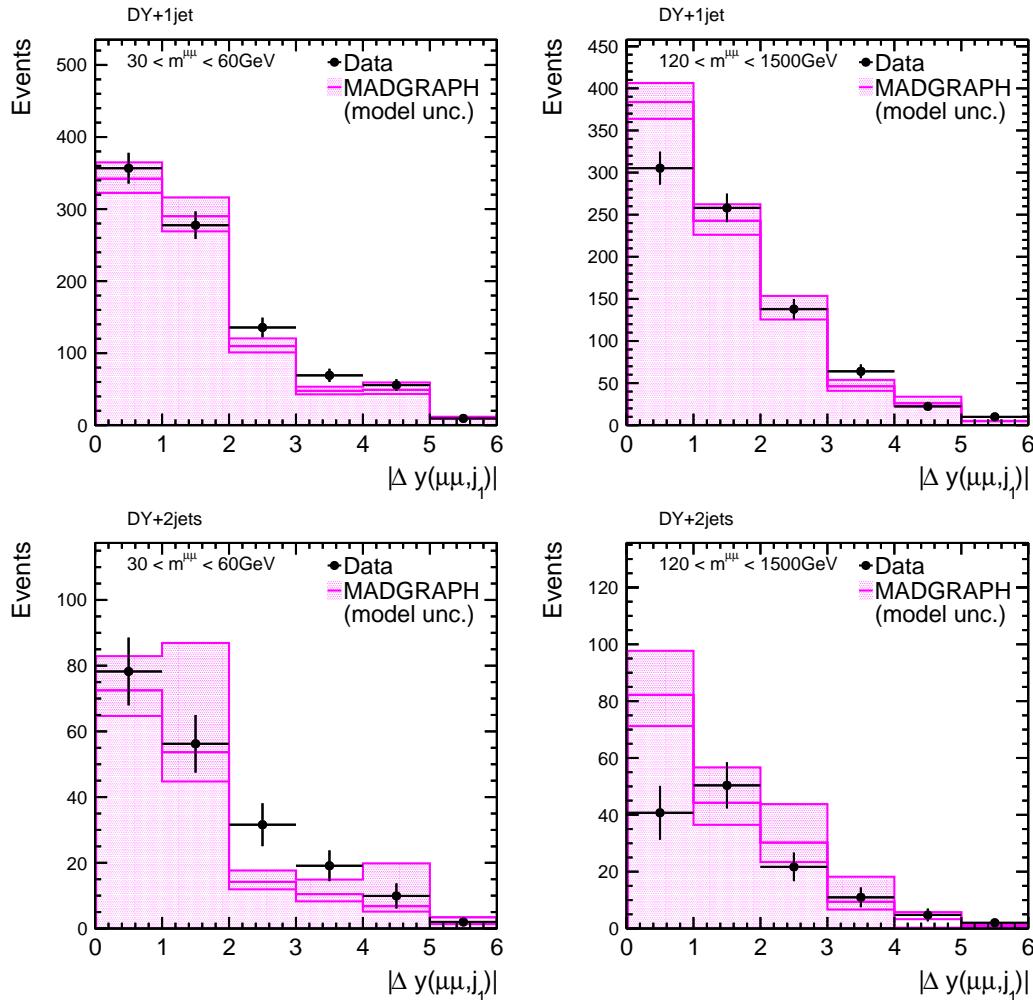


Figure 4.30: The $|\Delta y(\mu\mu, j_1)|$ distribution for low and high invariant masses in the DY + 1 jet and DY + 2 jets topology. Data is compared to MADGRAPH + PYTHIA6 predictions plus a model uncertainty derived from the uncertainties on the data parametrisation.

Efficiency Correction

The muon isolation and identification efficiency as well as the trigger efficiency are estimated using the tag-and-probe method. The statistical and systematic uncertainties on the correction factors are propagated to the cross section. The uncertainty on the correction factor is estimated to be 2% [97] for muons with $p_T < 100$ GeV. The uncertainty for muons with $p_T > 100$ GeV, is estimated by the extrapolation of correction factors from low to high muon p_T in using the total correction factor calculated at low p_T , adding a systematic uncertainty of the order of 3% for the extrapolation to high p_T [98]. The efficiency correction factors are changed within their uncertainty to estimate an uncertainty on the cross section.

Background Estimation

The final results show the background-subtracted data yields. Therefore, the uncertainty due to the estimation of the background process is assigned. This uncertainty includes the statistical uncertainty in the number of events in the background processes as well as an uncertainty to take into account the shape and normalisation of the background contribution.

The uncertainty due to the limited number of simulated events is assigned by varying the number of background events by $\pm\sqrt{N_i}$, where N_i is the number of MC events in bin i .

The uncertainty on the normalisation and background estimation depends on the background process. An uncertainty on the QCD background estimation is assigned to the ratio calculation of opposite-sign over same-sign data events by varying the anti-isolation requirement. The uncertainty on the ratio is estimated to be 8%. The $t\bar{t}$ background normalisation is taken from CMS measurement and an uncertainty of 8% is assigned [124]. Additionally, an uncertainty on the $t\bar{t}$ correction factor, to take into account differences between data and simulation, is assigned to be 10% [97]. The yields of diboson production are scaled to NLO theory predictions [125, 126] and an uncertainty of 30% is assigned on the normalisation. The single W background is estimated by PYTHIA and normalised to the inclusive cross section as measured by CMS [127]. The uncertainty is driven by limited MC statistics of the selected W + jets events. The uncertainty in $Z \rightarrow \tau\tau$ normalisation, estimated from the CMS measurement, is 4% [127].

The total systematic uncertainty due to the estimation of different background processes and limitations in statistics is obtained by varying the number of background events within the systematic uncertainty of the shape and normalisation, and by $\pm\sqrt{N_i}$, in the event bin i .

4.7.2 Summary of Systematic Uncertainties

The typical numbers of the relative systematic uncertainties due to JEC and JER, model dependence, pileup reweighting, efficiency correction, and background estimation, are listed in Table 4.4, 4.5, 4.6, and 4.7. The values correspond to inclusive DY, DY + 1 jet, and DY + 2 jets production. The different sources of systematic effects are considered to be uncorrelated and the individual uncertainties are added in quadrature in the final cross section.

$m^{\mu\mu}$ bin (GeV)	JEC & JER (%)	Model (%)	Pileup Rew. (%)	Eff. (%)	Bkg. Est. (%)
30 - 45	-	0.1	0.5	0.5	0.5
45 - 60	-	0.1	0.3	1	0.5
60 - 120	-	0.1	0.1	0.1	0.1
120 - 200	-	0.5	0.5	0.5	0.5
200 - 1500	-	2	1	2	2

Table 4.4: Summary of typical systematic uncertainties of the inclusive DY transverse momentum in different bins of the dimuon mass. The total systematic uncertainties are given by the quadratic sum of each individual sources.

$m^{\mu\mu}$ bin (GeV)	JEC & JER (%)	Model (%)	Pileup Rew. (%)	Eff. (%)	Bkg. Est. (%)
30 - 45	6	1	1	1	1
45 - 60	6	1	1	1	1
60 - 120	6	0.1	0.1	0.1	0.1
120 - 200	7	1	1	2	2
200 - 1500	7	2	1	4	3

Table 4.5: Summary of typical systematic uncertainties of the DY + 1 jet transverse momentum in different bins of the dimuon mass. The total systematic uncertainties are given by the quadratic sum of each individual sources.

$m^{\mu\mu}$ bin (GeV)	JEC & JER (%)	Model (%)	Pileup Rew. (%)	Eff. (%)	Bkg. Est. (%)
30 - 45	8	2	2	1	3
45 - 60	6	2	1	2	2
60 - 120	4	0.5	0.1	0.5	0.1
120 - 200	9	3	1	4	6
200 - 1500	13	7	4	8	8

Table 4.6: Summary of typical systematic uncertainties of the DY + 2 jets transverse momentum in different bins of the dimuon mass. The total systematic uncertainties are given by the quadratic sum of each individual sources.

$m^{\mu\mu}$ bin (GeV)	JEC & JER (%)	Model (%)	Pileup Rew. (%)	Eff. (%)	Bkg. Est. (%)
30 - 60	20 12	4 10	2 2	1 1	1 1
60 - 120	5 2	0.5 0.5	1 1	0.1 0.1	0.1 0.5
120 - 1500	5 7	1 5	1 1	2 2	1 2

Table 4.7: Summary of typical systematic uncertainties of $|\Delta y(\mu\mu, j_1)|$ distribution in DY + 1 jet and DY + 2 jets (first and second column respectively) events in different bins of the dimuon mass. The total systematic uncertainties are given by the quadratic sum of each individual sources.

4.7.3 Theory Uncertainties

The choice of the factorisation and renormalisation scale influences the theory prediction. The theory uncertainties are determined with the MC generator POWHEG. The uncertainties on the scale choices are obtained by reweighting the events in the POWHEG MiNLO production of $Z + 1$ jet [133] at NLO. The weight corresponds to different choices of scales. This method provides an easy calculation of the uncertainties without generating different MC samples for each scale variation. The central value of the factorisation and renormalisation scales are varied by a factor of two above and below. The results of the POWHEG prediction for the dimuon $p_T^{\mu\mu}$ and $|\Delta y(\mu\mu, j_1)|$ are shown in Figures 4.31 and 4.32. The generation of the Drell-Yan events with POWHEG is done in the Z invariant mass range of 60 – 120 GeV. The envelope of all predictions is shown as a band and the central value is shown as a histogram.

Small uncertainties due to the renormalisation and factorisation scale variation are observed at low $p_T^{\mu\mu}$. The uncertainties increase with increasing $p_T^{\mu\mu}$ and are largest for inclusive DY production around 25%. In the cases of DY + 1 jet and DY + 2 jets production the uncertainty decrease and is around 10% at high $p_T^{\mu\mu}$. In the

$|\Delta y(\mu\mu, j_1)|$ distribution the scale variation of μ_R and μ_F have a small effect of 10% on the differential cross section at small rapidity differences. Changing μ_R and μ_F a maximal uncertainty of $\sim 30\%$ is observed at large rapidity differences .

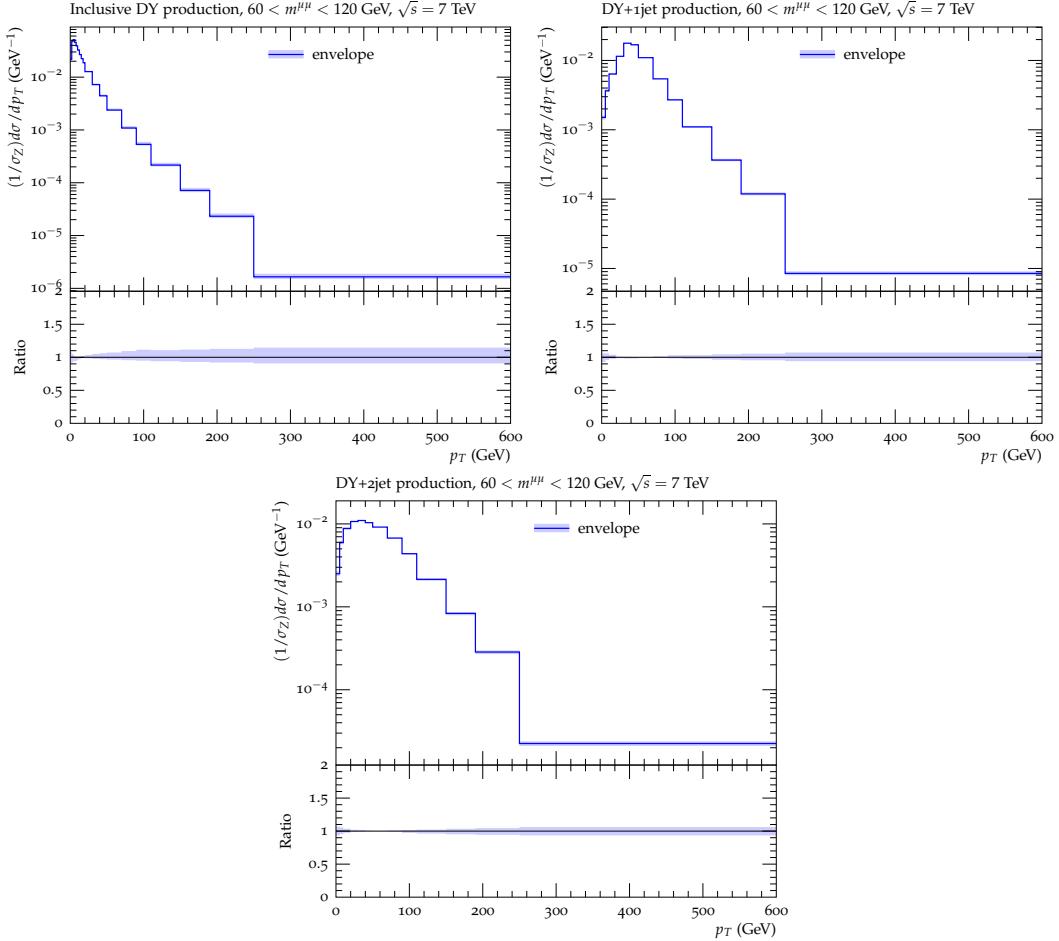


Figure 4.31: The renormalisation and factorisation scale variation by a factor of two above and below the central value. The envelope of all predictions is shown as a band. The dimuon $p_T^{\mu\mu}$ distribution is shown for the inclusive DY, DY + 1 jet and DY + 2 jets production in the Z invariant mass range.

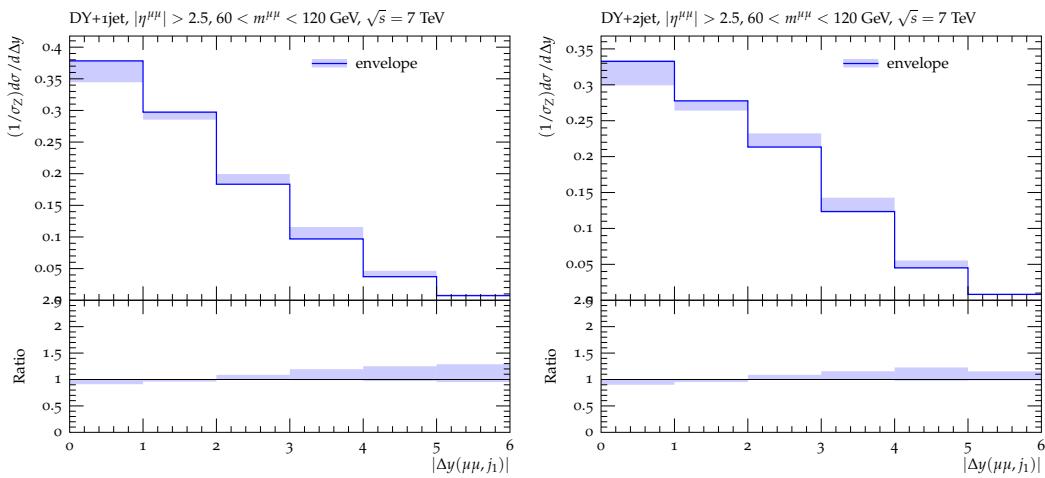


Figure 4.32: The renormalisation and factorisation scale variation by a factor of two above and below the central value. The envelope of all predictions is shown as a band. The $|\Delta y(\mu\mu, j_1)|$ distribution is shown for $DY + 1$ jet and $DY + 2$ jets production in Z invariant mass range.

5 Drell-Yan + Jets Results

The normalised differential Drell-Yan cross section as a function of the dimuon transverse momentum, $p_T^{\mu\mu}$, and the rapidity separation between the leading jet and the dimuon pair, $|\Delta y(\mu\mu, j_1)|$, is presented. The cross section is measured in ranges of the dimuon invariant mass covering a range of $30 < m^{\mu\mu} < 1500$ GeV. Three different event topologies are compared: inclusive DY, DY + 1 jet, DY + 2 jets production. The background-subtracted data yields are corrected to stable-particle level. The systematic uncertainties are presented as gray bands and the statistical uncertainties are shown as errorbars.

5.1 Drell-Yan Transverse Momentum Distribution

The Drell-Yan p_T distribution is a meaningful observable to study different physics effects in different regions of pQCD. This chapter presents the results on the cross section as a function of $p_T^{\mu\mu}$, as well as investigations of phenomenological aspects of the $p_T^{\mu\mu}$ distribution.

I present the results as a function of the invariant mass $m^{\mu\mu}$, to investigate the effect on the choice of the scale of the hard process. Furthermore, I compare the measurement to various theory predictions, estimated by MC event generators. Moreover, the theory of resummation is investigated by studying the effect of parton shower algorithms implemented in the MC simulation. In the event selection of Drell-Yan and associated jets the cross section is compared to the exclusive Drell-Yan plus jet production, in order to investigate the impact of additional jets, in terms of multi-jet resummation, at low $p_T^{\mu\mu}$.

Transverse Momentum Distribution as a Function of the Dimuon Invariant Mass

The normalised distribution as a function of the dimuon $p_T^{\mu\mu}$ is presented in Figures 5.1, 5.2, and 5.3, for inclusive DY, DY + 1 jet, and DY + 2 jets production respectively. The data are compared to theory predictions estimated by MADGRAPH +PYTHIA6. Additionally, the ratios of MC predictions over data are presented. The $p_T^{\mu\mu}$ distribution is shown in five invariant mass ranges: 30 – 45, 45 – 60, 60 – 120, 120 – 200, and 200 – 1500 GeV.

The distributions are shown in logarithmic scale in $p_T^{\mu\mu}$ and on the y-axis. The

cross section values extend over a range of five orders of magnitude for the low invariant mass range, $30 - 45$ GeV, decreasing to two orders of magnitude for the high invariant mass range, $200 - 1500$ GeV.

The $p_T^{\mu\mu}$ distribution for inclusive DY production (Figure 5.1) has a maximum around $p_T^{\mu\mu} \sim 5$ GeV. However, in the low invariant mass range, the distribution is affected by the muon p_T selection of $p_T > 20$ (10) GeV for leading (subleading) muon. In the mass range of $30 - 45$ GeV, the maximum value is around 10 GeV. The effect of the muon p_T selection on the dimuon $p_T^{\mu\mu}$ distribution is studied and the impact on low invariant masses of $m^{\mu\mu} < 60$ GeV was presented in Figure 4.16. The position of the maximum of ~ 10 GeV in the invariant mass range of $30 - 45$ GeV is due to the muon p_T selection and not due to the invariant mass selection.

The rise of the $p_T^{\mu\mu}$ distribution in the range of $0 < p_T^{\mu\mu} < 5$ GeV is described by resummation. The predictions by MADGRAPH +PYTHIA 6 provide a relatively good agreement to data. In the first $p_T^{\mu\mu}$ bin from $0 - 2$ GeV MADGRAPH underestimates the data by 10%.

The choice of the dimuon mass does not affect the low $p_T^{\mu\mu}$ region. However, the tail of the distribution changes with increasing invariant mass. The $p_T^{\mu\mu}$ tail is reasonably well described within the uncertainties by MADGRAPH +PYTHIA 6. Some discrepancies in the high $p_T^{\mu\mu}$ bins are observed, which are also remarked in other CMS analysis, e.g. [122]. The slope of the tail is steeply falling for $30 < m^{\mu\mu} < 45$ GeV and becomes flatter with higher $m^{\mu\mu}$, due to changing the scale of the hard process by changing $m^{\mu\mu}$. In processes with high $m^{\mu\mu}$, higher values in $p_T^{\mu\mu}$ can be observed. The agreement of data and theory in the high $p_T^{\mu\mu}$ bin improves with increasing invariant mass.

It is observed that, in the differential cross section as a function of $p_T^{\mu\mu}$ in the inclusive DY production, the scale of the process does not affect the resummation calculation at low $p_T^{\mu\mu}$, but results in a change of the slope of the $p_T^{\mu\mu}$ tail.

In the case of DY + 1 jet production (Figure 5.2) the maximum of the distribution is shifted to higher values of $p_T^{\mu\mu} \sim 35$ GeV, compared to inclusive Drell-Yan production. The shift of the maximum to higher values was motivated in section 1.4.4. In the event selection jets with $p_T^{\text{jet}} > 30$ GeV are required in addition to the Drell-Yan pair. This condition leads to the production of Drell-Yan lepton pairs with $p_T^{\mu\mu} \sim p_T^{\text{jet}}$. Due to the balancing of the Drell-Yan dimuon and the jet in the transverse plane, the Drell-Yan pair is recoiled by the jet. Thus, the requirement of associated jets above a p_T threshold of 30 GeV produces mainly dimuon pairs slightly above the p_T^{jet} threshold. Moreover, the Drell-Yan pair can be balanced by soft gluons, which lead to a contribution at low $p_T^{\mu\mu}$. For DY + 1 jet production the low $p_T^{\mu\mu}$ region is enlarged and soft-gluon resummation effects can be investigated in a wider range.

MADGRAPH overestimates data for $p_T^{\mu\mu} < 10$ GeV. A difference of 60% is observed for $30 < m^{\mu\mu} < 45$ and $45 < m^{\mu\mu} < 60$ GeV, and of 20% for $60 < m^{\mu\mu} < 120$ GeV. The differences are still within the systematic uncertainties. The comparison reveals that MADGRAPH predicts more low p_T Drell-Yan pairs than observed in data. The agreement improves with increasing $p_T^{\mu\mu}$. In the region where fixed-order calculations contribute dominantly, the MC prediction agrees reasonably well with data. At very high $p_T^{\mu\mu}$ MC results overestimate data, which was observed already in the inclusive DY distributions. With increasing invariant mass the decline in the tail flattens as

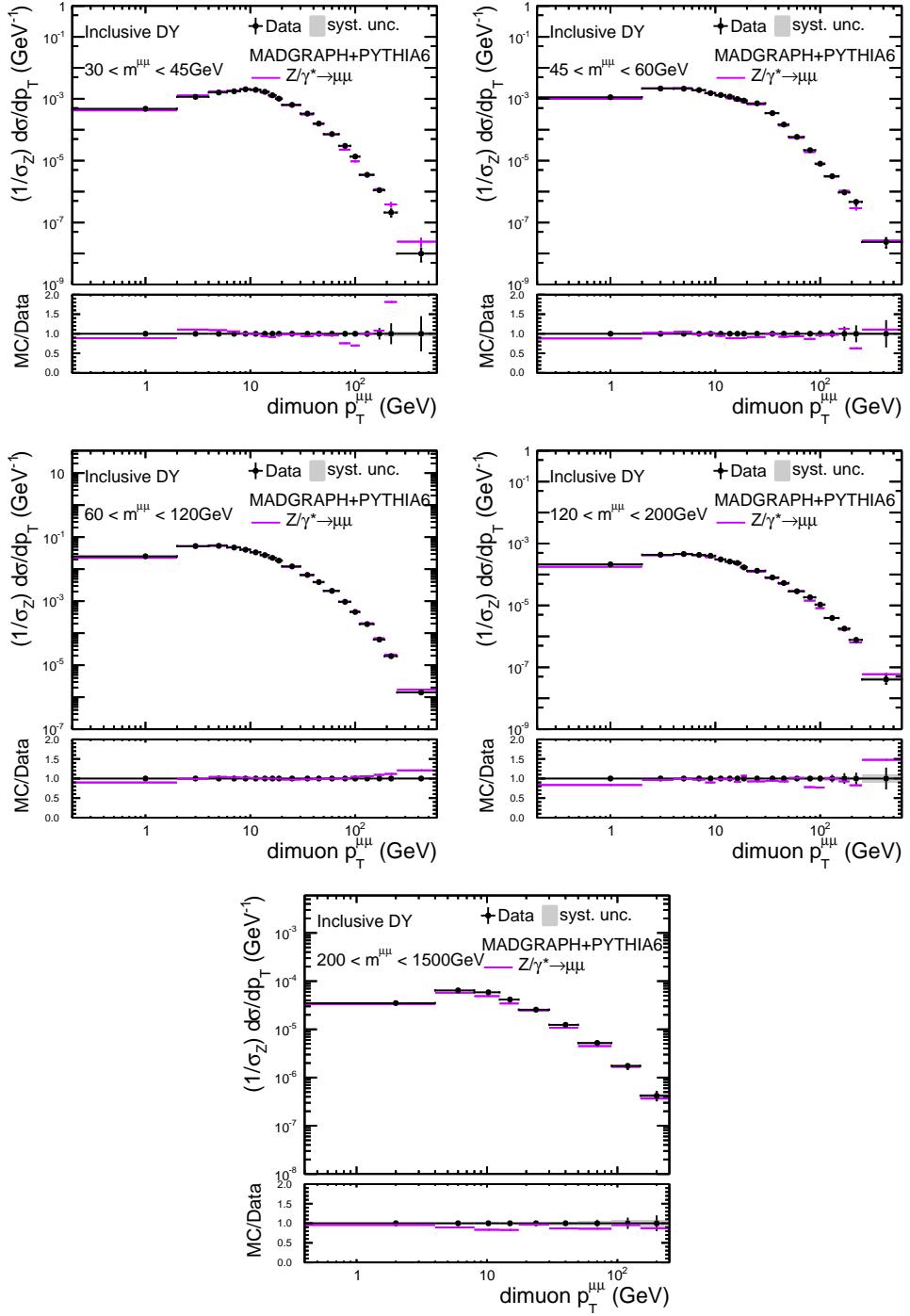


Figure 5.1: The normalised differential cross section for inclusive Drell-Yan production in five invariant mass ranges $m^{\mu\mu}$ as a function of $p_T^{\mu\mu}$. The results for the inclusive Drell-Yan production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

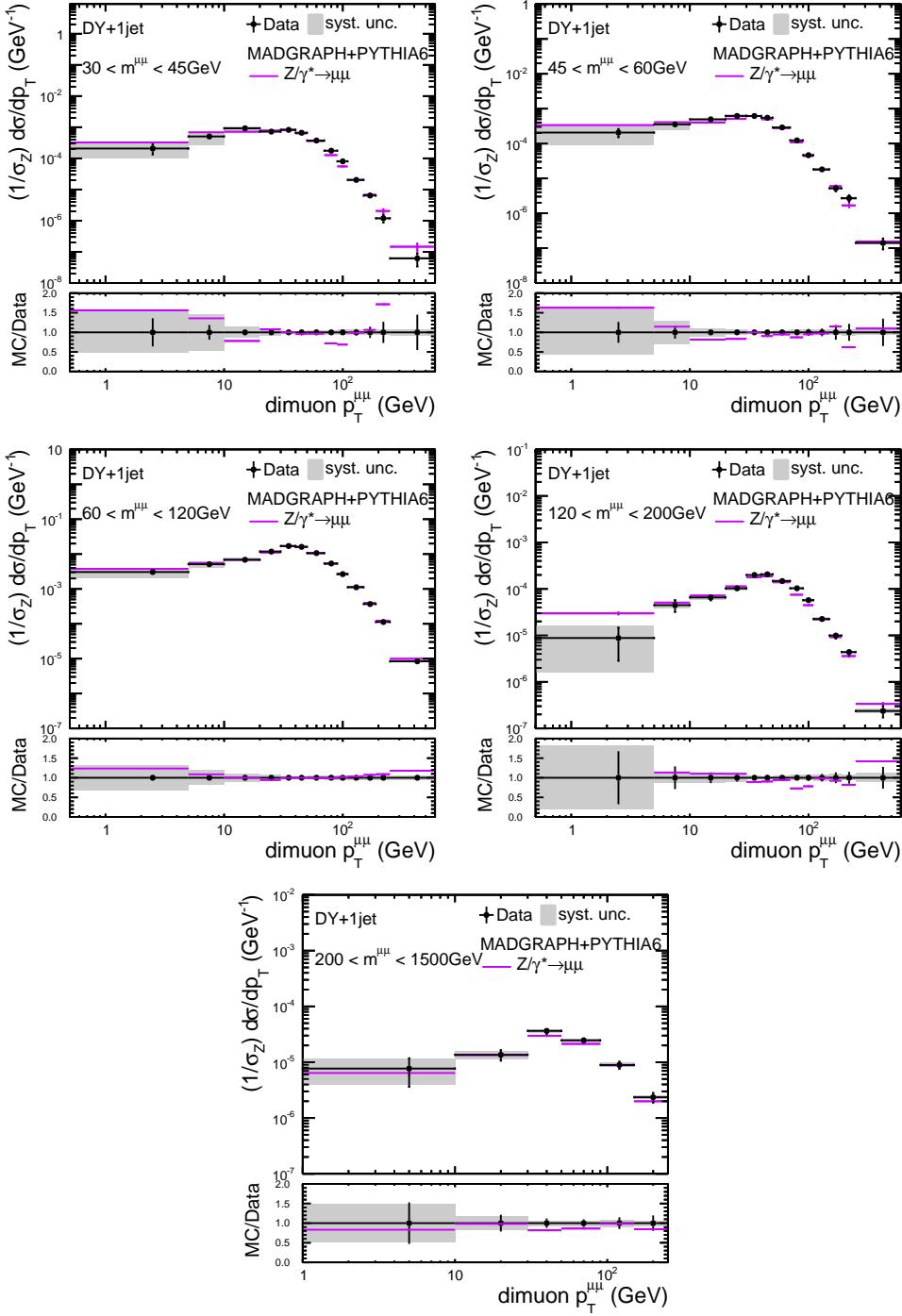


Figure 5.2: The normalised differential cross section in five invariant mass ranges $m^{\mu\mu}$ as a function of $p_T^{\mu\mu}$. The results for the DY + 1 jet production are compared to the predictions of MADGRAPH + PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

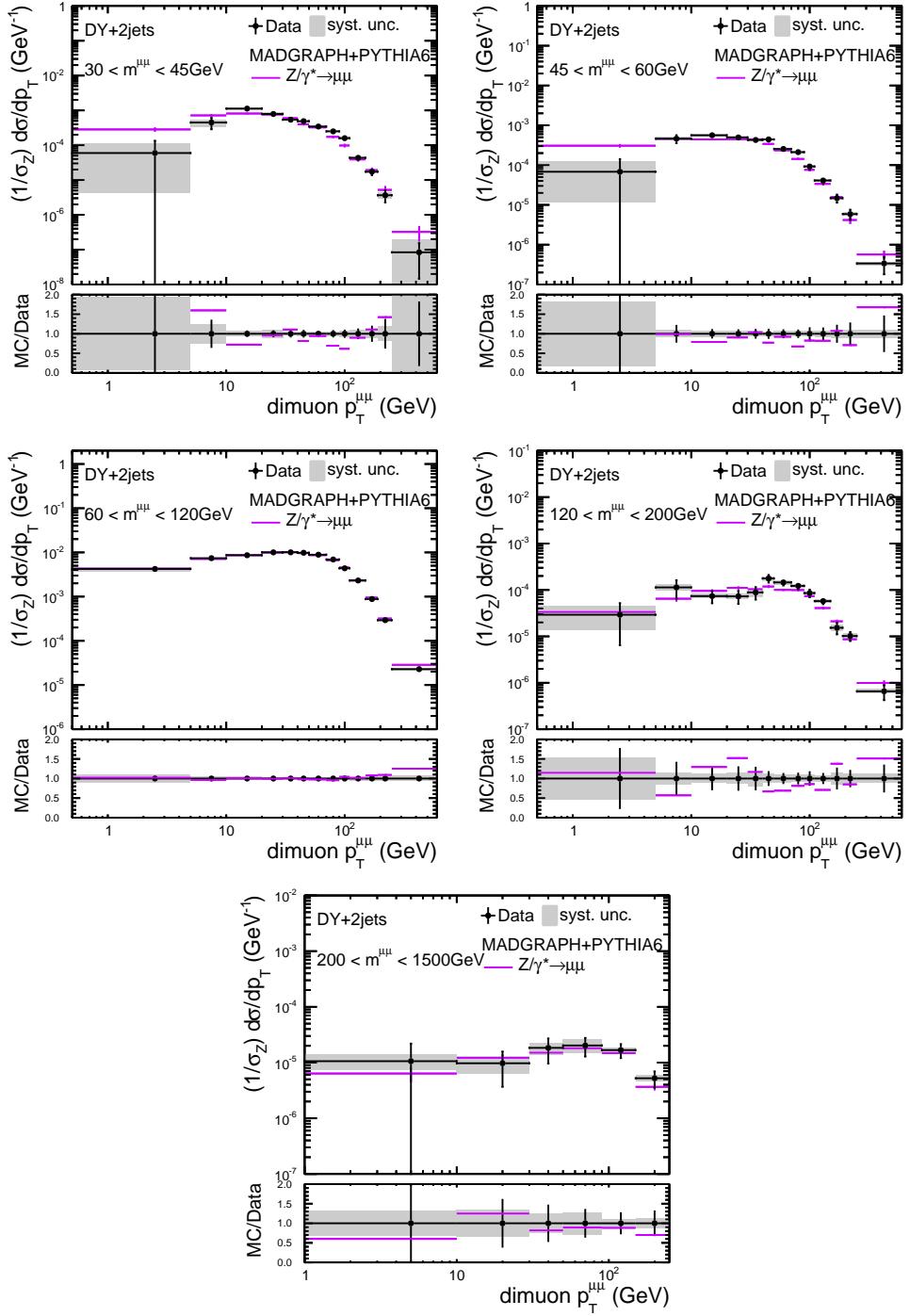


Figure 5.3: The normalised differential cross section in five invariant mass ranges $m^{\mu\mu}$ as a function of $p_T^{\mu\mu}$. The results for the DY + 2 jets production are compared to the predictions of MADGRAPH + PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

larger values of $p_T^{\mu\mu}$ can be reached. The agreement of data and MC predictions improve in the high $p_T^{\mu\mu}$ bins for high $m^{\mu\mu}$. The region of resummation is not influenced by the choice of the dimuon mass.

The dimuon $p_T^{\mu\mu}$ distribution for DY + 2 jets production is presented in Figure 5.3. The maximum of the normalised cross section is around ~ 35 GeV, the same value as for DY + 1 jet production. In the DY + 2 jets event topology, in most of the Drell-Yan events the dimuon is recoiled by the leading jet. However, some events are observed, in which the two jets balance each other and the Drell-Yan pair is produced with low $p_T^{\mu\mu} < 35$ GeV and is balanced by softer partons. This effect can be observed in the low $p_T^{\mu\mu}$ region, where the rise of the cross section is much flatter in the DY + 2 jets topology compared to DY + 1 jet production.

MADGRAPH shows a relatively good agreement to data. In the low $p_T^{\mu\mu}$ bin and the low $m^{\mu\mu}$ range MADGRAPH overestimates the data, but also large uncertainties are observed due to JEC. With increasing invariant mass the agreement improves up to an optimal description of data in the Z resonance range of $60 < m^{\mu\mu} < 120$ GeV.

All in all MADGRAPH +PYTHIA6 is able to describe the normalised differential cross section as a function of $p_T^{\mu\mu}$. Although some discrepancies are observed in some regions of phase space, the agreement of data and simulation is relatively good. The matrix-element calculation plus parton shower can describe the rise at small $p_T^{\mu\mu}$ as well as the fixed-order perturbative tail. In order to investigate resummation in more detail, the cross section results are compared to various matrix-element plus parton shower calculations.

Comparison to Fixed-Order Perturbative Theory Calculations plus Parton Shower

The application of soft-gluon resummation in SMC event generators is implemented via parton shower algorithms. How well the parton showers describe the mathematical principles of resummation is one topic in this thesis. It is of interest to understand the resummation in terms of soft-gluon emissions via parton showers (soft-gluon resummation) or summing contributions from hard emissions, like multi-jet events (multi-jet resummation). To draw conclusions on the resummation method in the dimuon $p_T^{\mu\mu}$ distribution, the data is compared to different MC generator predictions, which correspond to different orders in α_S in the perturbative calculation.

In the previous paragraph the cross section results were compared to MADGRAPH +PYTHIA6. The matrix-element calculation of MADGRAPH includes the calculation of the hard process for $2 \rightarrow Z^0/\gamma^* + 4$, which means the production of Z^0/γ^* is accompanied by at most four additional partons. The parton shower and UE is simulated with PYTHIA6.

A list of generated processes at different orders in α_S is presented in Table 5.1. POWHEG ($Z + 2$ jets) MiNLO [135] using the PDF set of HERAPDF1.5NLO [134] provides the calculation for $Z^0/\gamma^* + 2$ partons at NLO. The first parton emission is calculated by POWHEG and the parton shower algorithm is implemented using PYTHIA6 (tune Z2*).

The SMC event generator PYTHIA6 features the option of generating the Drell-Yan process at LO and at first-order in α_S . The matrix-element calculation of the $2 \rightarrow 1$ or $2 \rightarrow 2$ process is determined at LO and the parton shower algorithm is generated at NLL. The tune Z2* is employed. This set of MC predictions yields fixed-order perturbative calculations of the hard matrix-element at different orders in α_S .

Process	$\mathcal{O}(\alpha_S)$	Generator
$Z^0/\gamma^* + 4$ partons	LO	MADGRAPH
$Z^0/\gamma^* + 2$ partons	NLO	POWHEG MiNLO
$Z^0/\gamma^* + 1$ parton	LO	PYTHIA
Z^0/γ^*	LO	PYTHIA

Table 5.1: Different higher-order processes of the Drell-Yan production generated by various MC generators.

In the previous paragraph, it was observed that, the region of resummation is not significantly affected by the scale of the hard process ($m^{\mu\mu}$). Therefore, the following comparisons are performed in the Z resonance invariant mass range from $60 < m^{\mu\mu} < 120$ GeV.

The results are presented in Figure 5.4 for the inclusive DY, DY + 1 jet, and DY + 2 jets production in logarithmic scale in $p_T^{\mu\mu}$. With this option the rise at low $p_T^{\mu\mu}$ is much more visible and the different topologies can be compared more easily. The shift of the maximum, from inclusive DY production to Drell-Yan plus associated jets, is presented in this figure. The increase of the cross section at low $p_T^{\mu\mu}$ for DY + 1 jet and DY + 2 jets production is nicely visible.

In the case of inclusive DY production the different MC predictions provide a good agreement with the measured cross section. In the low $p_T^{\mu\mu}$ region as well as in the perturbative fixed-order region, the MC simulations agree to data.

In the case of DY + 1 jet production the comparison to PYTHIA6 estimated at $\mathcal{O}(\alpha_S)$ is included. Two regions of the $p_T^{\mu\mu}$ distribution can be discussed separately. In the low $p_T^{\mu\mu}$ region ($p_T^{\mu\mu} < 35$ GeV) and in the high $p_T^{\mu\mu}$ tail ($p_T^{\mu\mu} > 35$ GeV) a very different level of agreement of data and MC predictions can be observed. For $p_T^{\mu\mu} > 35$ GeV all MC predictions provide a good agreement to data. The fixed-order perturbative calculations of different orders are all able to describe the tail of the distribution. However, at low $p_T^{\mu\mu} < 35$ GeV, larger discrepancies between the four calculations are observed. MADGRAPH +PYTHIA6, provides a good agreement to data. The MC values overestimate the data around $\sim 20\%$ but are still within the systematic uncertainties. POWHEG comparisons to data reveal a maximal difference of 50%, which decreases with larger values of $p_T^{\mu\mu}$. The two predictions from PYTHIA yield a poor agreement to data at low $p_T^{\mu\mu}$. The inclusive Z^0/γ^* production underestimates data by maximal 50% and the $2 \rightarrow 2$ process gives a too low cross section and is not able to describe the rise of the distribution.

In the case of DY + 2 jets production the comparison to MADGRAPH +PYTHIA6,

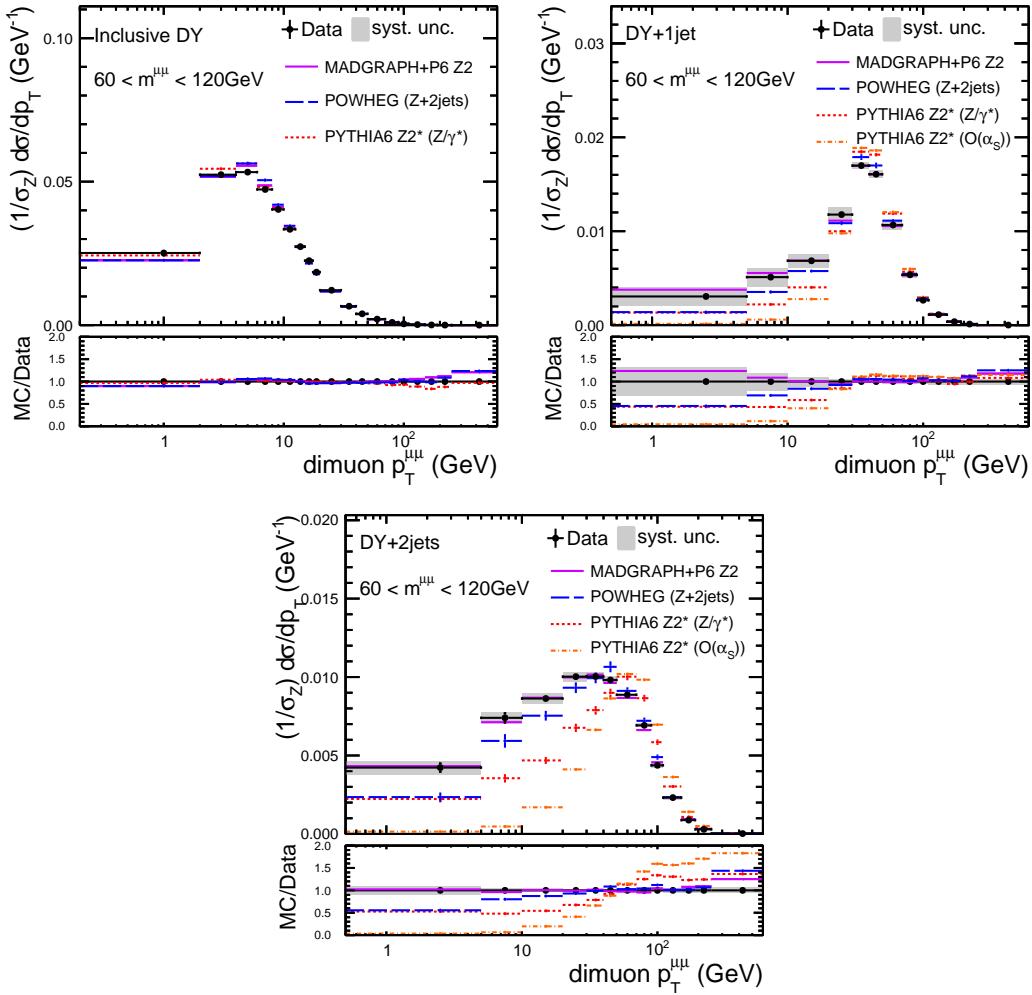


Figure 5.4: The normalised differential cross section in the Z resonance mass range as a function of $p_T^{\mu\mu}$. The results for the inclusive DY, DY + 1 jet, and DY + 2 jets productions are compared to the predictions of MADGRAPH +PYTHIA6, POWHEG +PYTHIA6, and PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

shows the best agreement to data. The cross section results from MADGRAPH yield a good agreement with the data over the whole region in $p_T^{\mu\mu}$. The second best agreement is obtained by POWHEG +PYTHIA6. The hard $p_T^{\mu\mu}$ tail can be described nicely, while in the low $p_T^{\mu\mu}$ region a lack of low $p_T^{\mu\mu}$ dimuons is observed. In the case of the SMC generator PYTHIA6, the predictions can not describe the shape of the $p_T^{\mu\mu}$ distribution for DY + 2 jets production. The cross section values at low $p_T^{\mu\mu}$ underestimate the data, and the $p_T^{\mu\mu}$ tail can not be described neither by LO nor by the first-order calculation.

The low $p_T^{\mu\mu}$ region is sensitive to resummation as motivated in section 1.4.4. In Figure 1.12 it was observed that the p_T distribution of LO DY + 1 jet production

drops to zero, when excluding parton shower. Moreover, the DY + 1 jet production at LO plus parton shower is not sufficient to describe the data. Higher-order matrix-element calculations are needed to provide a reasonable basis for the cross section calculation. The theory predictions provided by matched fixed-order calculations plus parton shower algorithms from MADGRAPH plus PYTHIA6 provide a better agreement to data, than lower-order calculations.

Concluding, the rise at low $p_T^{\mu\mu}$ can be described by the parton showers for inclusive DY production. In this topology the parton shower approach in the MC simulation is sufficient to estimate the soft-gluon resummation. However, in Drell-Yan plus jet topologies a higher-order matrix-element calculation is needed in combination with the parton shower algorithm to be able to describe the rise at low $p_T^{\mu\mu}$. Thus, the low $p_T^{\mu\mu}$ region requires contributions of hard emissions, calculated at fixed-order pQCD, plus soft-gluon resummation, modelled by the parton shower.

In the following the impact of the parton shower is studied in more detail in the Drell-Yan plus jets topologies.

Effect of Parton Shower on the Dimuon Transverse Momentum Distribution

The parton shower in the MC simulation is an explicit simulation of soft-gluon resummation. In the Drell-Yan lepton pair production only the initial-state parton emissions affect the transverse momentum distribution directly and influence the rise of the cross section at small $p_T^{\mu\mu}$. In DY + 1 jet production at LO the low $p_T^{\mu\mu}$ Drell-Yan pair can only be recoiled by the soft-gluon emissions. In case of DY + 2 jets production at LO and NLO, the low $p_T^{\mu\mu}$ dimuons can also be recoiled by jets from the higher-order corrections in the matrix-element calculation. In the following the impact of parton shower is investigated using POWHEG ($Z + 2$ jets) MiNLO, which generates DY + 2 jets at NLO. The initial-state parton shower is explicitly excluded in the simulation and the difference to the nominal setting is compared at low $p_T^{\mu\mu}$.

The normalised cross section as a function of the dimuon $p_T^{\mu\mu}$ is compared to MC simulation by MADGRAPH +PYTHIA6 and POWHEG +PYTHIA6 with and without parton showers in Figure 5.6. It is observed that the effect of soft-parton emissions is negligible in the high $p_T^{\mu\mu}$ tail, where fixed-order calculations dominate. However, at small $p_T^{\mu\mu}$ a small impact is observed when excluding parton showers in the MC simulation. The difference between the two contributions, including and excluding initial-state parton emissions, is $\sim 10\%$.

Only a small amount of Drell-Yan pairs are really recoiled by soft-gluon emissions. In order to obtain small Drell-Yan p_T events, the process can be interpreted as radiation of Z^0/γ^* boson by one of the hard partons. The feynman diagrams for Drell-Yan plus jet production at different orders in α_S is illustrated in Figure 5.5. The LO DY + 1 jet diagram shows how the Drell-Yan pair is emitted by the hard parton. In this case additional emission recoil the Drell-Yan plus parton system. A process like this is significantly sensitive to soft-gluon emission, and the patron shower is required to balance the low p_T Drell-Yan pair. The LO and NLO DY +

2 jets diagrams illustrate the Z^0/γ^* radiation by one of the hard partons and the system can be recoiled by other emissions. Thus, the low p_T Drell-Yan events in POWHEG ($Z + 2$ jets) are sensitive to additional hard emission from the matrix-element calculation. Comparing the predictions from MADGRAPH and POWHEG, one can observe that with increasing number of additional partons the agreement to data improves.

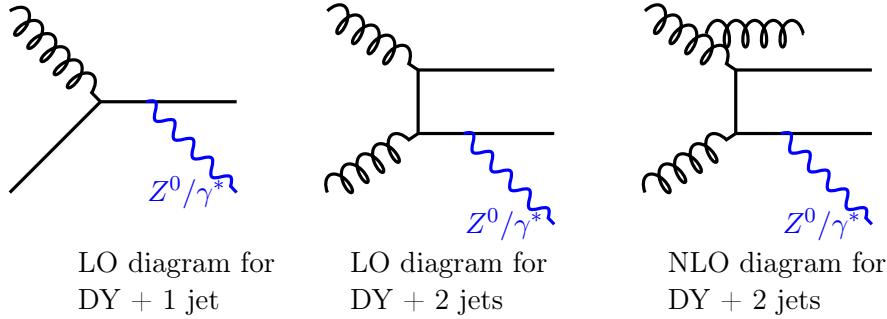


Figure 5.5: Radiation of Z^0/γ^* at different orders in α_S .

It is verified that the spectrum at low $p_T^{\mu\mu}$ in Drell-Yan plus jet events is based mainly on the hard-parton emission, calculated at fixed-order in perturbation theory. Thus, the low $p_T^{\mu\mu}$ is sensitive by hard-parton emissions plus resummation in form of parton shower.

In order to investigate the production of low $p_T^{\mu\mu}$ events by Z^0/γ^* radiation by one of the hard partons, the inclusive DY + 1 jet events are compared to the exclusive production of Drell-Yan plus exactly one jet with $p_T^{\text{jet}} > 30$ GeV.

Exclusive Drell-Yan plus Jet Production

It was observed that the effect of soft-gluon resummation in terms of parton showers is small for DY + 2 jets production at NLO. By comparing exclusive and inclusive Drell-Yan plus jet production one can investigate the impact of additional jets emerging from the hard-parton emission. In the case of low $p_T^{\mu\mu}$ production, by Z^0/γ^* radiation by one of the hard partons, the impact of the hard emissions in form of multi-jet events can be investigated.

In Figure 5.7 the normalised cross section for inclusive DY + 1 jet and Drell-Yan plus exactly one jet (DY + excl 1 jet) above $p_T^{\text{jet}} > 30$ GeV is compared. The theory prediction from MADGRAPH +PYTHIA 6 is added in the comparison. It is observed that the cross section in the exclusive Drell-Yan plus jet production yields slightly smaller values at low $p_T^{\mu\mu}$. However, the agreement is within the uncertainties.

One can conclude that additional hard emissions are needed to describe the cross section at low $p_T^{\mu\mu}$, but do not result in a jet above the p_T^{jet} threshold of 30 GeV.

All in all, the studies on the normalised cross section measurement as a function

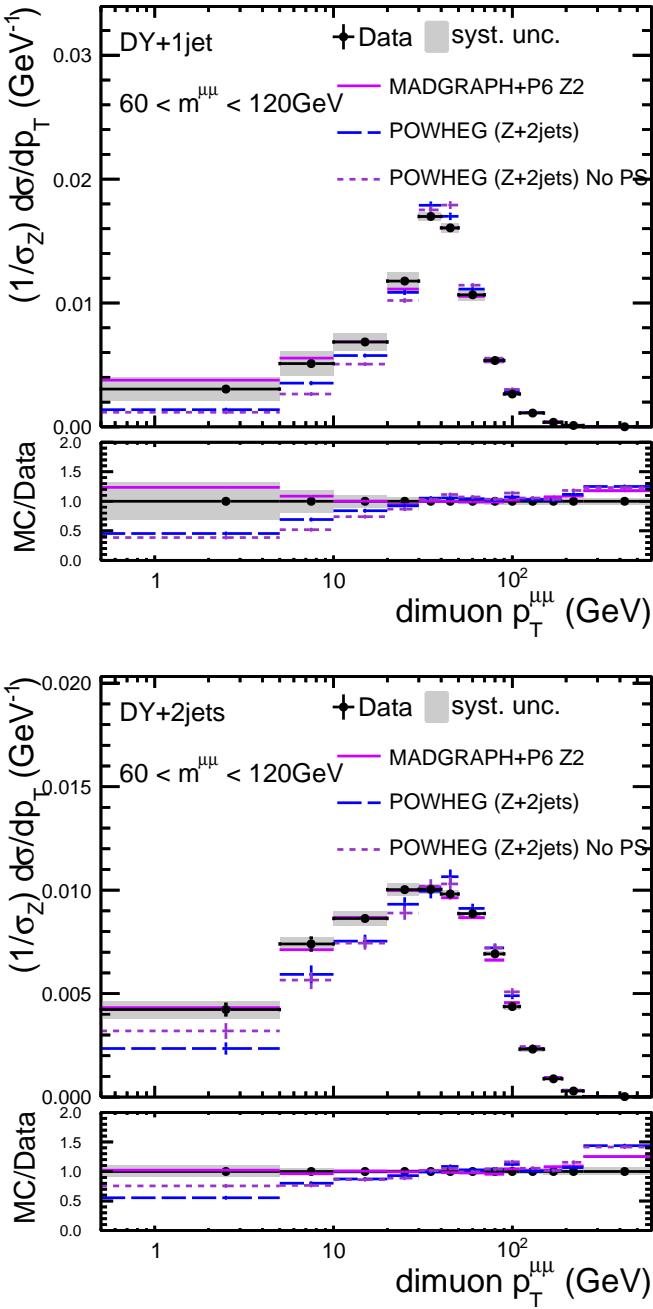


Figure 5.6: The normalised differential cross section in the Z resonance mass range as a function of $p_T^{\mu\mu}$. The results for the DY + 1 jet and DY + 2 jets productions are compared to the predictions of MADGRAPH +PYTHIA6 and POWHEG +PYTHIA6 with and without initial-state parton emissions (parton shower = PS). The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

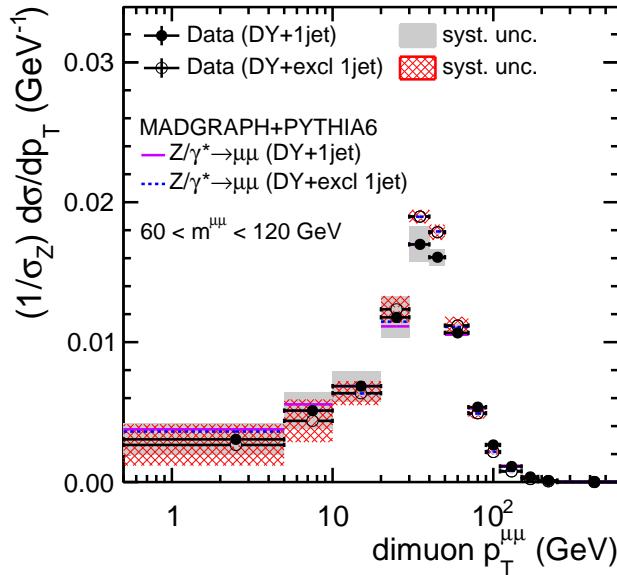


Figure 5.7: The normalised differential cross section in the Z resonance mass range as a function of $p_T^{\mu\mu}$. The results for inclusive and exclusive DY + 1 jet production are compared to the predictions of MADGRAPH + PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

of the dimuon $p_T^{\mu\mu}$ reveal a deep insight into phenomenological investigations of resummation and multi-jet events. It is observed that the differential cross section shows an impact on the dimuon invariant mass range in the slope of the hard $p_T^{\mu\mu}$ tail. The low $p_T^{\mu\mu}$ region is not affected by the change of $m^{\mu\mu}$. The normalised cross section is compared to various MC simulations, which refer to different higher-order calculations of the hard scattering process plus parton shower simulation. The best agreement to data is observed for MADGRAPH + PYTHIA 6. MADGRAPH generates the hard process with at most four final-state partons in the calculation. The parton shower model of PYTHIA6 describes nicely the rise of the cross section at low $p_T^{\mu\mu}$ for inclusive DY production. However, for Drell-Yan plus jet events the LO calculation plus parton showers are not sufficient to describe resummation. A more appropriate result can be obtained by higher-order calculations of the hard scattering plus parton showers, e.g. provided by POWHEG + PYTHIA or MADGRAPH + PYTHIA. The effect of the parton shower is investigated in more detail for DY + 2 jets production at NLO. The contribution from initial-state soft-gluon emissions is small in DY + 1 jet and DY + 2 jets topologies. It can be concluded that the low $p_T^{\mu\mu}$ region in Drell-Yan plus jets production is sensitive to additional hard-parton radiation. It is observed that the agreement improved when including higher-order calculations. Summarising, Drell-Yan cross section with associated jets can probe hard-parton emissions at low $p_T^{\mu\mu}$ in terms of higher-order matrix-element calculations. The normalised cross section as a function of the dimuon $p_T^{\mu\mu}$ is sensitive to a combination of hard-parton emissions as well as soft-gluon resummation.

5.2 Rapidity Separation between Drell-Yan and the leading Jet

The normalised differential Drell-Yan plus jet cross section as a function of the absolute rapidity separation between the dimuon pair and the leading jet is presented. The cross section is measured in three ranges of the dimuon invariant mass from $30 - 60$, $60 - 120$, and $120 - 1500$ GeV. The event topologies of DY + 1 jet and DY + 2 jets are compared. The background-subtracted data yields are corrected to stable-particle level. The systematic uncertainties are presented as gray bands and the statistical uncertainties are shown as errorbars.

In this analysis the dimuon system is selected to be forward in pseudorapidity with $|\eta| > 2.5$. The Drell-Yan boson production with associated jets ($p_T^{\text{jet}} > 30$ GeV) at large rapidities is sensitive to small- x resummation. At large rapidities the fixed-order calculations are expected to fail, due to the fact that, in the perturbative calculations large logarithms in terms of $\ln(1/x)$ appear. At small- x the resummation of the leading-logs is performed via the BFKL or CCFM evolution. Thus, the distribution of the rapidity difference of the forward Drell-Yan and the leading jet is a meaningful observable to study multi-gluon emissions at large rapidities.

In this section, I present the normalised differential cross section as a function of $|\Delta y(\mu\mu, j_1)|$ and the average jet multiplicity as a function of the rapidity separation. The results are compared to various MC predictions, calculated at fixed-order plus parton showers.

In the region of Drell-Yan + jet production at large rapidities, very large as well as very low values of x can be reached. In the case of forward Drell-Yan production, the quark (or anti-quark) inducing the Drell-Yan process yields a large momentum fraction x of the incoming proton. Therefore, any initial-state parton shower is produced at even larger rapidity. However, the parton emissions induced from the opposite side of the evolution ladder, fill the region of phase space investigated in this work. The relation between the observables of rapidity y and momentum fraction x is connected to the invariant mass M (cf. eq. (1.28) and Figure 1.5.) by

$$x_{1,2} = \frac{M}{\sqrt{s}} e^{\pm y}. \quad (5.1)$$

Thus, by fixing two observables, the remaining can be investigated. In the case of forward Drell-Yan production, one can reach small (large) values in x by requiring low (high) invariant masses. Typical values in x for low dimuon masses of 30 GeV are less than 10^{-4} , when requiring the Drell-Yan pair to be forward.

Differential Cross Sections as a Function of $|\Delta y(\mu\mu, j_1)|$ at Low and High Invariant Masses

The normalised differential cross section in three invariant mass regions is presented in Figure 5.8 for DY + 1 jet production and in Figure 5.9 for DY + 2 jets production. The measured cross section decreases as a function of $|\Delta y(\mu\mu, j_1)|$. Comparing DY + 1 jet and DY + 2 jets production no significant change in the shape of the

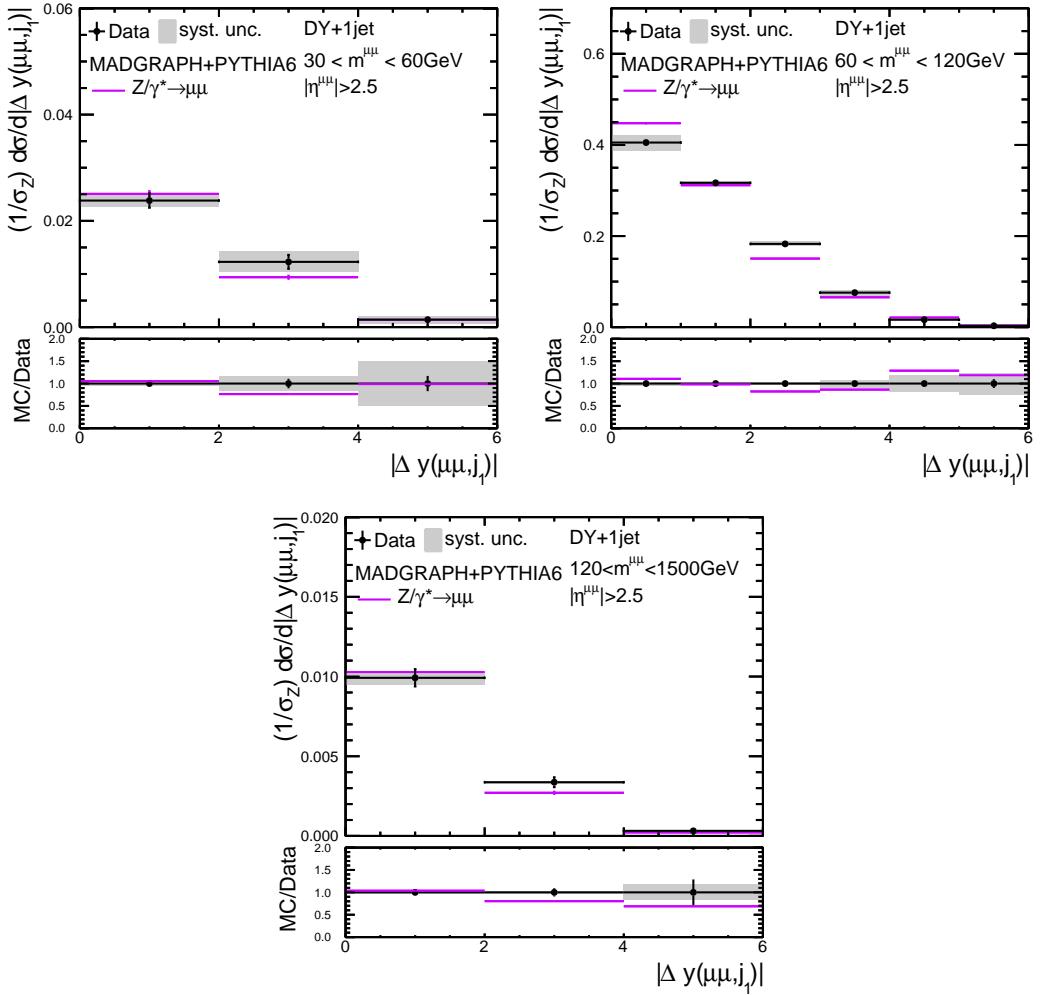


Figure 5.8: The normalised differential cross section for DY + 1 jet production in three invariant mass ranges $m^{\mu\mu}$ as a function of $|\Delta y(\mu\mu, j_1)|$. The results for the DY + 1 jet production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

distributions is observed. Slightly smaller values are predicted for DY + 2 jets production. The cross section values are measured in three bins of $|\Delta y|$ in the mass ranges of 30 – 60 and 120 – 1500 GeV. In the Z resonance range the normalised cross section is measured in six bins. The measurement is compared to MADGRAPH +PYTHIA6. The predictions of MADGRAPH +PYTHIA6 are shifted slightly towards smaller rapidity differences and the Drell-Yan is predicted to be closer to the jet than observed in data.

Additionally, the measured cross section is compared to different MC predictions. The simulation of POWHEG ($Z + 2$ jet) MiNLO is included in the cross section comparison as well as the LO calculations of PYTHIA6. In the Z invariant mass region

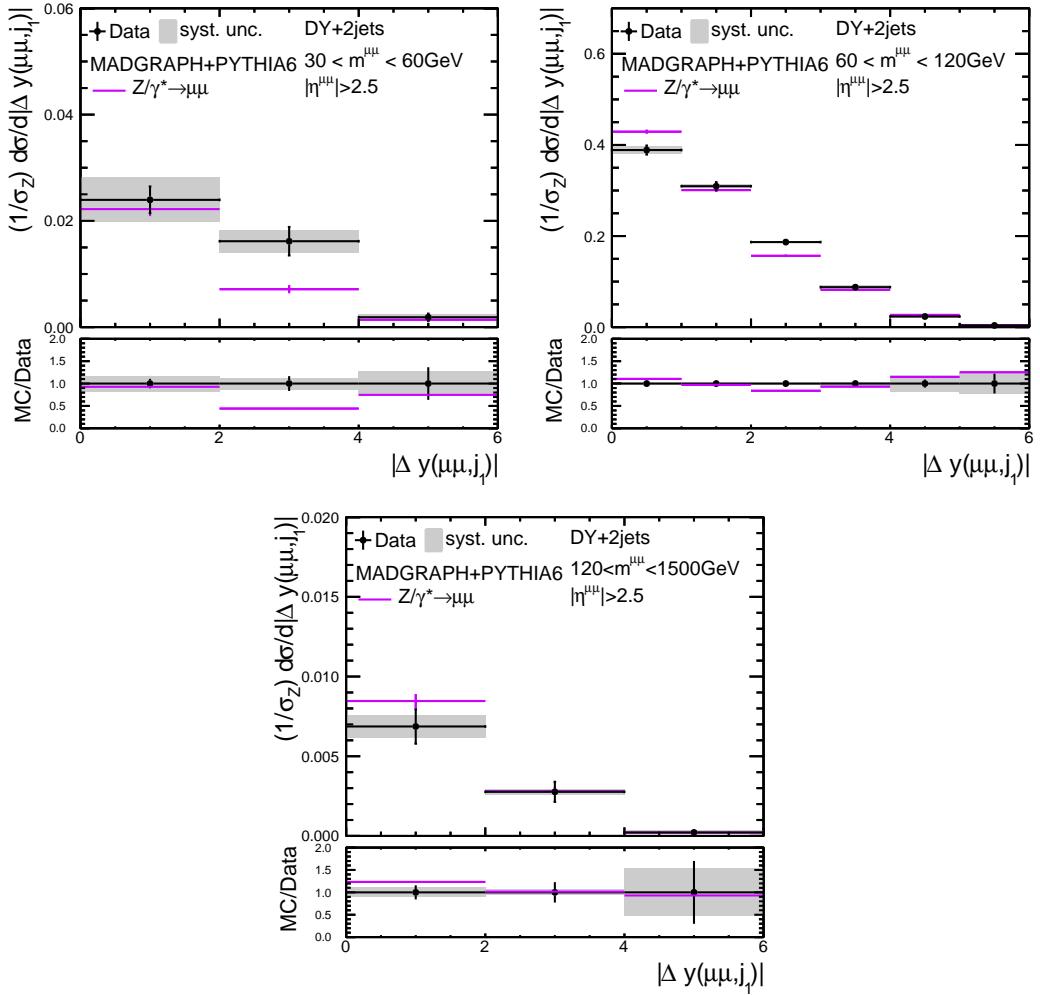


Figure 5.9: The normalised differential cross section for DY + 2 jets production in three invariant mass ranges $m^{\mu\mu}$ as a function of $|\Delta y(\mu\mu, j_1)|$. The results for the DY + 2 jets production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

predictions from CASCADE are shown. The event simulation is generated with an off-shell matrix-element including a TMD density function JH-2013-set2 [56]. The evolution follows the CCFM parton branching. In Figures 5.10 and 5.11 the results are presented for DY + 1 jet and DY + 2 jets production respectively.

In the low invariant mass region the MC predictions are not able to reproduce the measured cross section as a function of $|\Delta y|$. The MC calculations predict a too low cross section. The predictions of MADGRAPH +PYTHIA6 are within the statistical and systematic uncertainties of data. The predictions of POWHEG ($Z + 2\text{jets}$) at NLO underestimate the data at low masses, but PYTHIA6 is even further off.

In the Z resonance region POWHEG +PYTHIA6 reveals similar results than obtained from MADGRAPH +PYTHIA6. Moreover, the LO and first order calculations

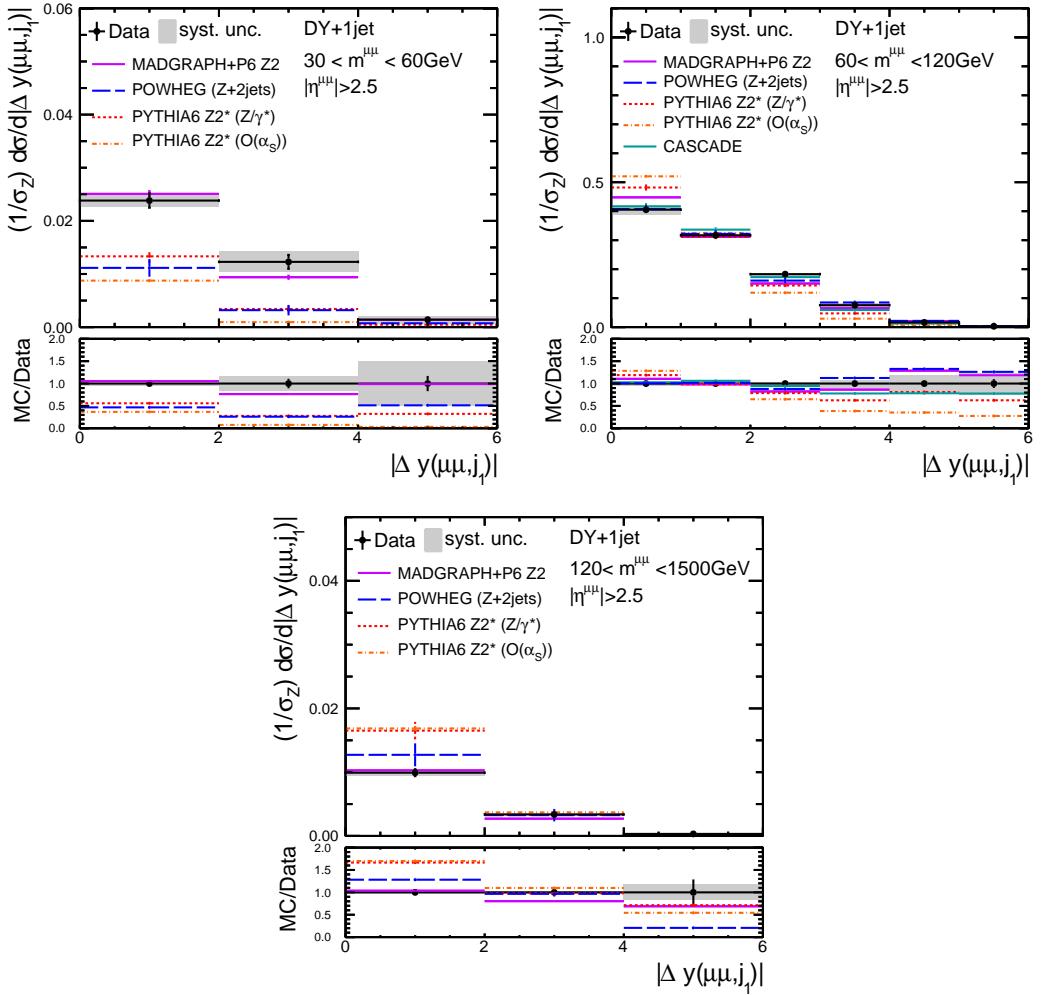


Figure 5.10: The normalised differential cross section in three invariant mass ranges as a function of $|\Delta y(\mu\mu, j_1)|$. The results for the DY + 1 jet production are compared to the predictions of MADGRAPH +PYTHIA6, POWHEG +PYTHIA6, CASCADE, and PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

of PYTHIA6 fail to describe the large $|\Delta y|$ region. In the bin of $1 < |\Delta y(\mu\mu, j_1)| < 2$, the cross section value is nicely described by all MC predictions.

The predictions of CASCADE are included in the comparison. CASCADE shows a good agreement to data for $|\Delta y| < 3$. At large rapidity differences CASCADE underestimates the data around 10 – 30%.

In the high invariant mass region for DY + 1 jet production, the MC over data ratio shows the trend towards smaller $|\Delta y|$. All MC predictions overestimate the cross section at low $|\Delta y|$ and underestimate the data at large $|\Delta y|$. In the case of DY + 2 jets production only MADGRAPH is able to reproduce the data. The other MC simulations predict too high values for $120 < m^{\mu\mu} < 1500$ GeV.

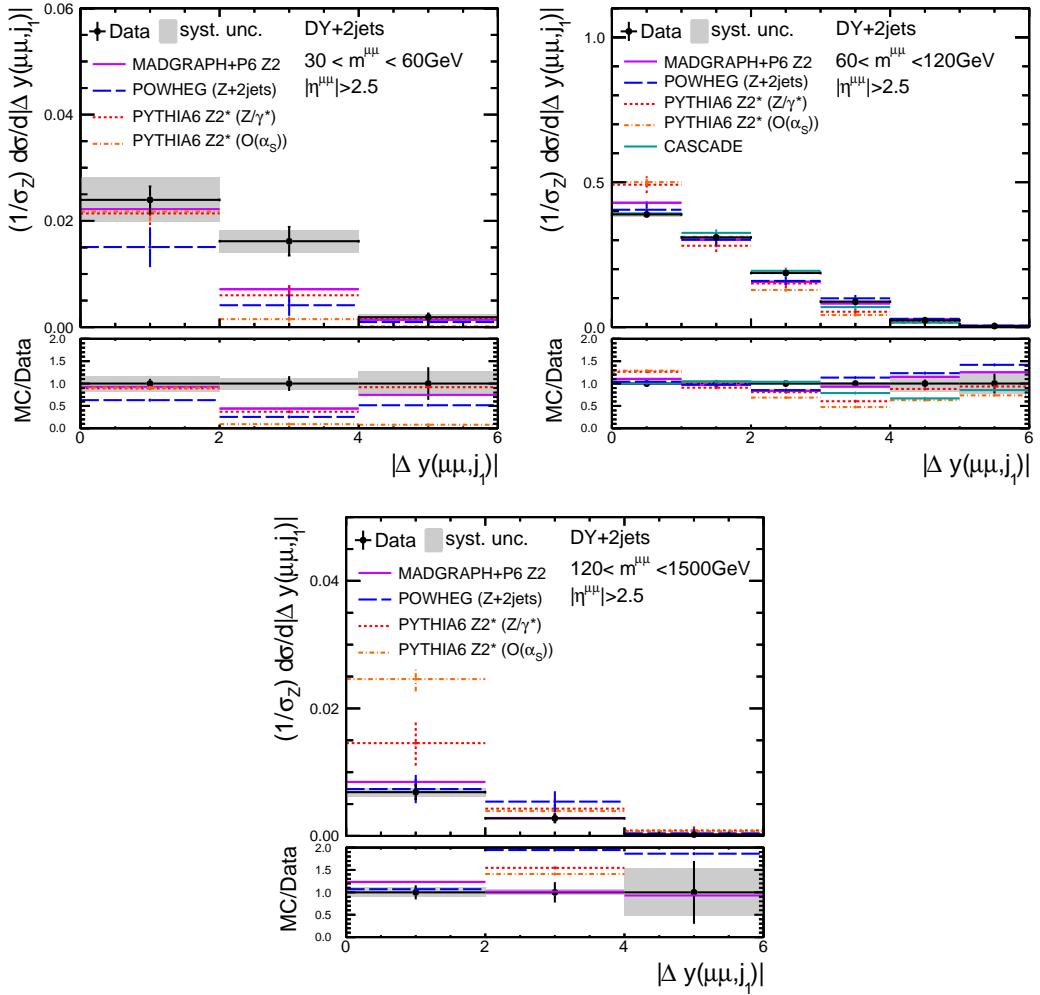


Figure 5.11: The normalised differential cross section in three invariant mass ranges as a function of $|\Delta y(\mu\mu, j_1)|$. The results for the DY + 2 jets production are compared to the predictions of MADGRAPH +PYTHIA6, POWHEG +PYTHIA6, CASCADE, and PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.

Average Jet Multiplicity as a Function of $|\Delta y(\mu\mu, j_1)|$

In the previously defined region of phase space, with Drell-Yan and associated jet production at large rapidities, the measured cross section as a function of $|\Delta y(\mu\mu, j_1)|$ is sensitive to small- x resummation. Higher-order corrections in the forward Drell-Yan plus jet production can be investigated by the average number of jets. The average jet multiplicity is studied as a function of $|\Delta y(\mu\mu, j_1)|$, where jets are observed with $p_T^{\text{jet}} > 30$ GeV and $|\eta^{\text{jet}}| < 4.5$. The measured jet multiplicity is compared to MC simulation accounting for different theoretical calculations at LO and NLO plus collinear resummation. In Figure 5.12 the distribution is presented in the Z reso-

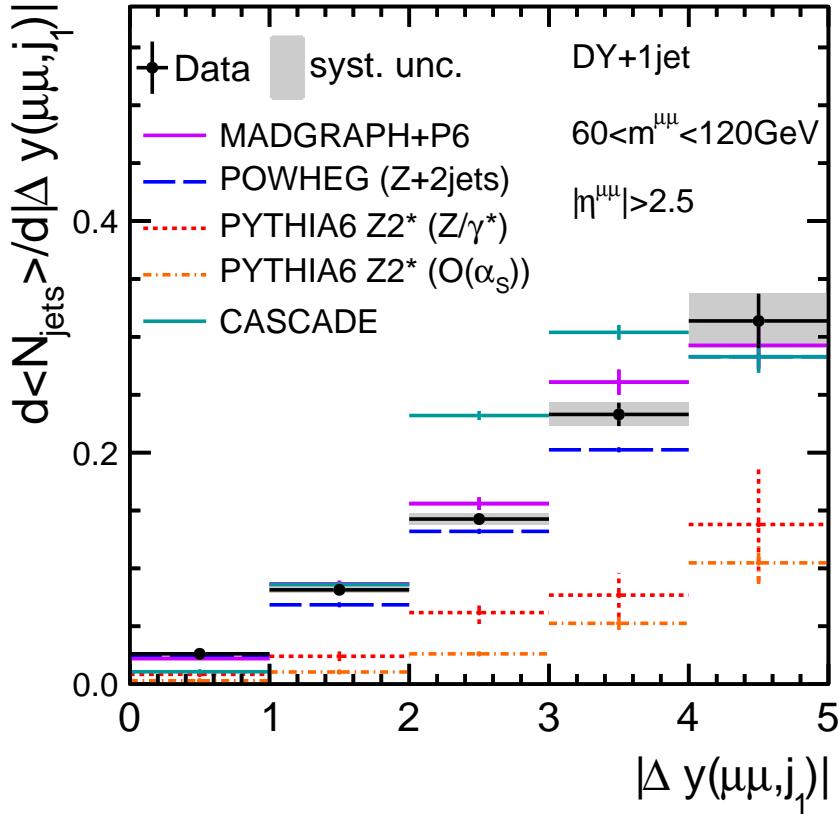


Figure 5.12: Average jet multiplicity as a function of the absolute rapidity separation between the forward Drell-Yan pair and the leading jet. The jets are measured in the fiducial region of $p_T^{\text{jet}} > 30 \text{ GeV}$ and $|\eta^{\text{jet}}| < 4.5$. Data are compared to various MC simulations.

nance invariant mass region. An increase of the average number of jets is observed with $|\Delta y|$. The jet multiplicity increases up to 30% at $|\Delta y| = 5$. PYTHIA underestimates the rise of the average jet multiplicity. The higher-order calculations provided by MADGRAPH and POWHEG yield a reasonable good agreement to data. CASCADE predicts a steeper rise of the jet multiplicity and predicts more jets. In the last $|\Delta y|$ bin the MC predictions of MADGRAPH, POWHEG (the POWHEG prediction is below the CASCADE line), and CASCADE underestimate the data, but the agreement is still within the uncertainties.

The collinear resummation provided by PYTHIA6 for LO production is not sufficient to describe data and higher-order calculations are needed to describe the average jet multiplicity. The resummation of the gluon cascade predicts more jets than observed.

6 Summary

Studies of differential Drell-Yan cross sections provide deep insights into QCD. Since, the Drell-Yan process is sensitive to QCD initial-state radiation, the parton emissions can be investigated in detail. Higher-order corrections are important in the perturbative calculation of the Drell-Yan + jets cross section. The higher-order corrections are included as QCD emissions and correspond to soft-gluon radiations or hard-parton emissions, where the latter can be observed as jets in the final state.

At low Drell-Yan transverse momentum the soft-gluon emissions introduce a divergence in the fixed-order calculation and have to be resummed to all orders in perturbation theory, in order to describe the observed cross section. The p_T^{DY} distribution is sensitive to soft-gluon resummation at $p_T^{\text{DY}} \ll m^{\text{DY}}$. In an event selection with Drell-Yan and associated jets with $p_T^{\text{jet}} > 30$ GeV, the maximum of the distribution is shifted to larger values and the region of resummation increases. In this topology the soft-gluon emissions can be studied in a much wider range.

In this work, the Drell-Yan dimuon pairs in association with jets at a centre-of-mass energy of 7 TeV were measured. The results of the differential cross section measurement were presented as a function of the dimuon $p_T^{\mu\mu}$. The process was studied in different regions of the dimuon invariant mass $m^{\mu\mu}$ from 30 to 1500 GeV, to use the mass as a choice of the scale of the process. Three different event topologies, inclusive DY, DY + 1 jet, and DY + 2 jets, were investigated. The differential cross section was normalised to the total integrated cross section in the Z invariant mass region (60 – 120 GeV) for each topology. The results correspond to the fiducial region defined by the leading and subleading muon with $p_T > 20$ and 10 GeV respectively, and $|\eta| < 2.1$. The jets were selected with $p_T^{\text{jet}} > 30$ GeV in $|\eta^{\text{jet}}| < 4.5$.

The normalised $p_T^{\mu\mu}$ distribution was compared to various theory estimations from fixed-order plus collinear resummation, generated by MC event generators. The matrix-element calculation was obtained at different orders in α_S in the various MC generators. The implementation of the soft-gluon resummation was simulated by parton shower algorithms. MADGRAPH generated the Drell-Yan production at LO with at most four additional parton emissions in the matrix-element. The MC generator POWHEG calculated DY + 2 jets at NLO, which included the first hard emission. Both predictions were interfaced with PYTHIA6 to model the parton shower and the UE. Drell-Yan cross section predictions provided by PYTHIA6 at LO and first-order in α_S were also included in the comparison.

MADGRAPH +PYTHIA6 provided the best agreement to data over the whole range in $p_T^{\mu\mu}$ and for all three event topologies. With increasing invariant mass $m^{\mu\mu}$ the tail of the $p_T^{\mu\mu}$ distribution got steeper and higher values in $p_T^{\mu\mu}$ were reached. The change of the mass did not affect the low $p_T^{\mu\mu}$ region. Small discrepancies for small invariant masses between data and MC simulation were observed. MADGRAPH +PYTHIA6 is able to describe the rise at small $p_T^{\mu\mu}$ as well as the high $p_T^{\mu\mu}$ tail. The $Z + 2$

jets predictions at NLO from POWHEG provided a good agreement in the inclusive Drell-Yan $p_T^{\mu\mu}$ distribution. However, for DY + 1 jet and DY + 2 jets production, the calculation underestimated the data in the first $p_T^{\mu\mu}$ bins. The high $p_T^{\mu\mu}$ tail is well described, but in the low $p_T^{\mu\mu}$ region some discrepancies were observed. PYTHIA 6 was able to describe the inclusive Drell-Yan p_T distribution. For DY + 1 jet and DY + 2 jets production the LO prediction from PYTHIA6 were not able to reproduce the data.

From this one can conclude that the Drell-Yan + jets cross section calculation requires a computation including higher-order emissions in the matrix-element calculation. With increasing number of additional partons in the matrix-element calculation, the agreement to data improved. Thus, a combination of hard-parton emissions plus soft-gluon resummation is needed to describe the resummation of the Drell-Yan plus jets $p_T^{\mu\mu}$ distribution.

The impact of the parton shower was studied in more detail in the case of the DY + 2 jets production at NLO provided by POWHEG. It was observed that the effect of parton shower is very small at low $p_T^{\mu\mu}$, around $\sim 10\%$. In case of DY + 1 jet production at LO, the parton shower is required to recoil the p_T of the Drell-Yan system. However, in Drell-Yan + jets at higher orders, the Z^0/γ^* can be radiated by one of the hard partons and both recoil the other hard-parton emissions. In this process, the effect of parton shower does not influence the dimuon $p_T^{\mu\mu}$ significantly. Comparing inclusive and exclusive Drell-Yan plus jets production only small differences in the normalised cross sections were observed, which were still within the systematic uncertainties.

Summarising, the low $p_T^{\mu\mu}$ region in Drell-Yan plus jet topologies is sensitive to additional hard-parton emissions, calculated in the matrix-element computation. The combination of one to three additional partons and the soft-parton resummation in terms of parton showers are needed to describe the low $p_T^{\mu\mu}$ resummation in Drell-Yan plus jets cross sections.

In an event selection with forward Drell-Yan production ($|\eta^{\mu\mu}| > 2.5$) the DY + jet cross section is sensitive to small- x resummation. At large rapidities the multi-gluon emissions can be investigated in detail. The differential cross section as a function of the rapidity separation between the leading jet and the Drell-Yan dimuon pair was presented. The normalised cross section results were compared to various MC generators in three invariant mass regions. The predictions from higher-order calculations plus soft-gluon resummation provided by MADGRAPH +PYTHIA6 and POWHEG +PYTHIA6 were presented. Furthermore, the distributions obtained from CASCADE were included in the comparison in the Z resonance invariant mass region. In addition the average number of jets as a function of $|\Delta y(\mu\mu, j_1)|$ was shown.

MADGRAPH +PYTHIA6 provided the best agreement to data. The estimation from POWHEG provided similar results in the Z invariant mass range. However, at low and high dimuon invariant masses the predictions underestimated and overestimated the data, respectively. The predictions from PYTHIA6 were even worse and were not able to describe the normalised cross section. At large rapidity separation the LO calculation failed and higher-order calculations provided a better agreement to data. The predictions obtained from CASCADE showed a good agreement to data at low $|\Delta y(\mu\mu, j_1)|$. At large rapidity separation CASCADE underestimated the data.

The average jet multiplicity is nicely reproduced by MADGRAPH +PYTHIA6 and

POWHEG +PYTHIA6. CASCADE predicted even higher jet multiplicities. The rise of the data increased up to 30% for rapidity differences of five. The lowest order calculation of PYTHIA could not reproduce the rise of the distribution.

Summarising, probing the Drell-Yan plus jets processes at low $p_T^{\mu\mu}$ and large rapidity differences, provides important information on soft-gluon resummation as well as multi-jet events. Higher-order calculations in combination with collinear resummation is required to reproduce the normalised distributions. With increasing number of partons included in the hard matrix-element calculation the agreement to data improves. Thus, the Drell-Yan plus jets topologies are sensitive to additional hard-parton emissions as well as multi-gluon emissions, which can be simulated by parton shower algorithms.

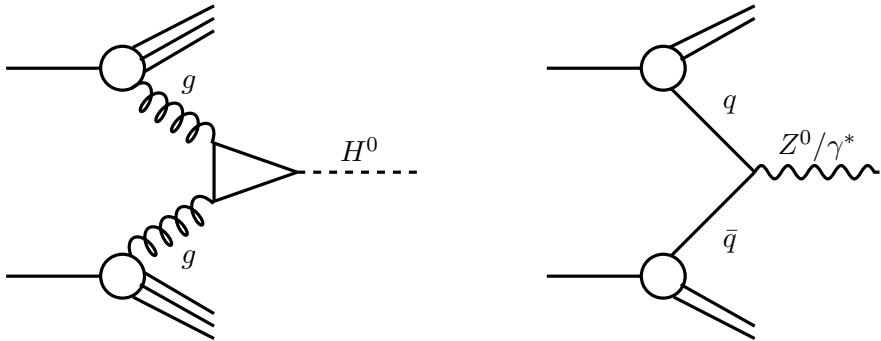


Figure 6.1: Illustration of Higgs and Drell-Yan production, comparing a gluon and quark induced process.

With this work at hand, one now is in the position to extend the study by measuring also Higgs boson production. By comparing Drell-Yan and Higgs production in a mass range around the Higgs mass, a gluon and quark induced process can be compared by measuring colour singlet final states. The Higgs production via the gluon gluon fusion is the dominant production process. The Higgs process provides good mass resolution via the four-lepton final state ($H \rightarrow ZZ \rightarrow 4l$) and was measured by CMS in [136]. The data was collected with an integrated luminosity of $\mathcal{L} = 24.8 \text{ fb}^{-1}$. Around 24 events are selected in a mass range around $m^H = 126 \text{ GeV}$. This provides a production cross section of around 1 fb for the Higgs boson in the four-lepton final state.

The production of the neutral Higgs boson via a top quark loop and the Drell-Yan process is shown in Figure 6.1. At low boson $p_T \ll m^H$ the contributions from quark and gluon Sudakov form factors can be investigated and information on the gluon radiation pattern from initial states can be extracted. At large boson $p_T > m^H$ the multi-jet emissions dominate and recoil against the Higgs or Drell-Yan pair. The multi-jet topology of the gluon and quark induced process can be investigated.

A precise measurement of the differential Higgs cross section (e.g. as a function of the boson p_T) can be used in means to study QCD cross section also with high pileup conditions. The aspect of pileup is a crucial issue to consider in the upcoming high luminosity phase with the LHC. By measuring the ratios of differential cross section for Higgs and Drell-Yan production pileup effects largely cancel out.

The feasibility study of measuring systematic differences between Higgs and Drell-

Yan differential distributions is presented in [119, 137]. This work addresses a broad range of physics issues, which are needed for a detailed understanding of the Higgs process.

Appendix A

Trigger Efficiency

This analysis uses double and single muon High-Level triggers HLT_Mu13_Mu8 and HLT_IsoMu17. The trigger efficiencies for the two triggers in four different pseudorapidity ranges of the muon as a function of the probe muon p_T are shown in the following. The efficiency calculation takes into account the 2011 run periods A and B.

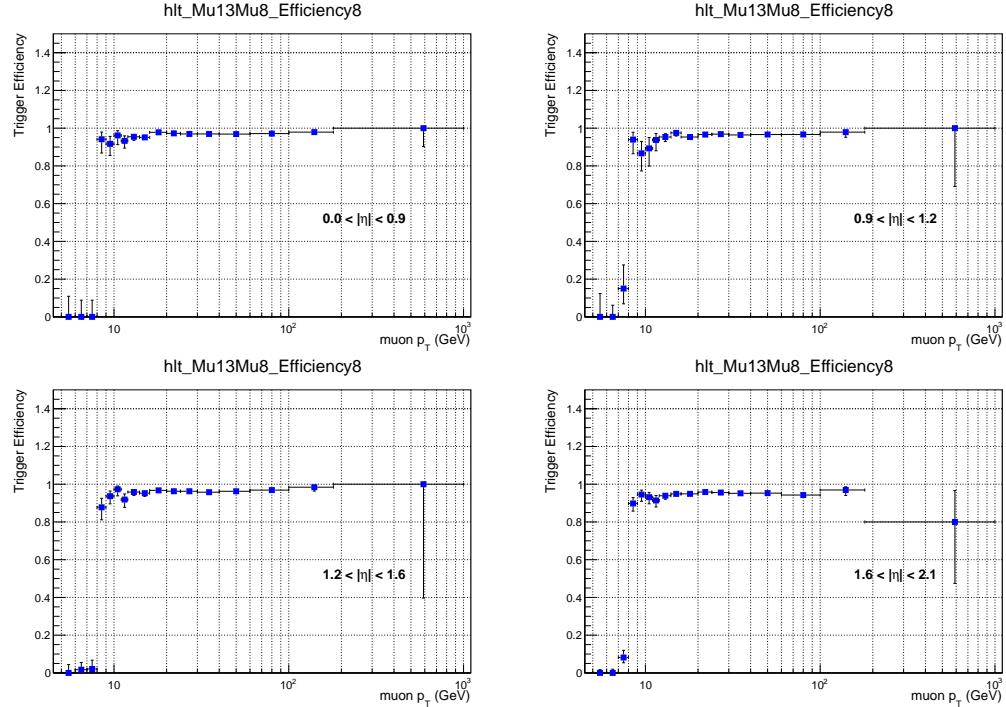


Figure A.1: Trigger turn-on curves for the trigger leg of 8 GeV of HLT_Mu13_Mu8 in four pseudorapidity regions as a function of the muon p_T [97]. The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events.

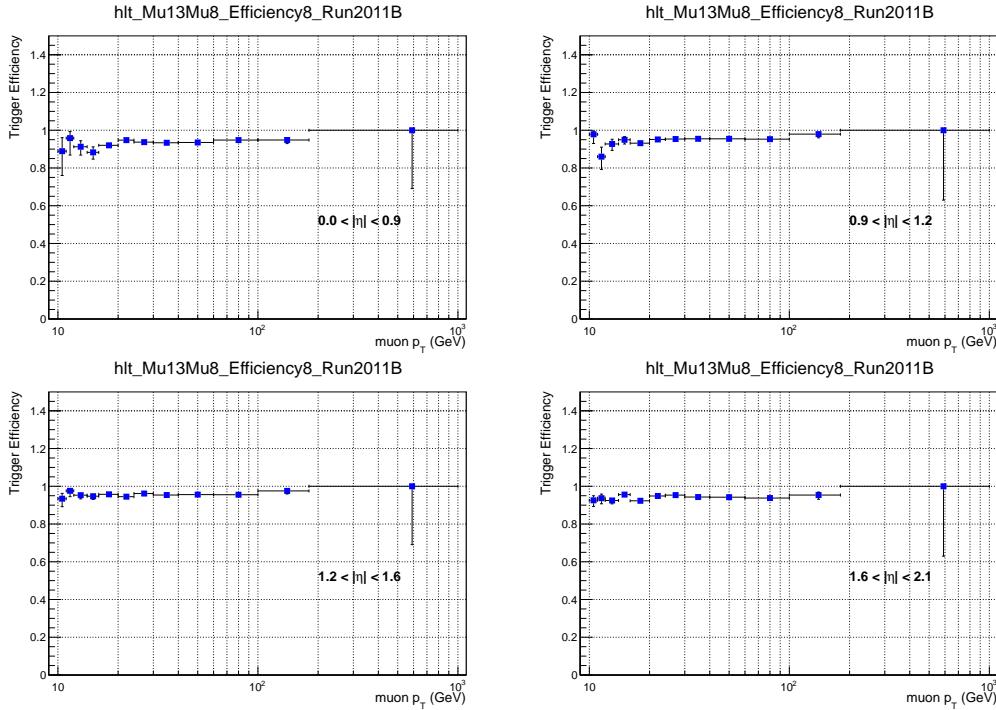


Figure A.2: Trigger turn-on curves for runs in 2011B for the trigger leg of 8 GeV of HLT_Mu13_Mu8 in four pseudorapidity regions as a function of the muon p_T [97]. The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events.

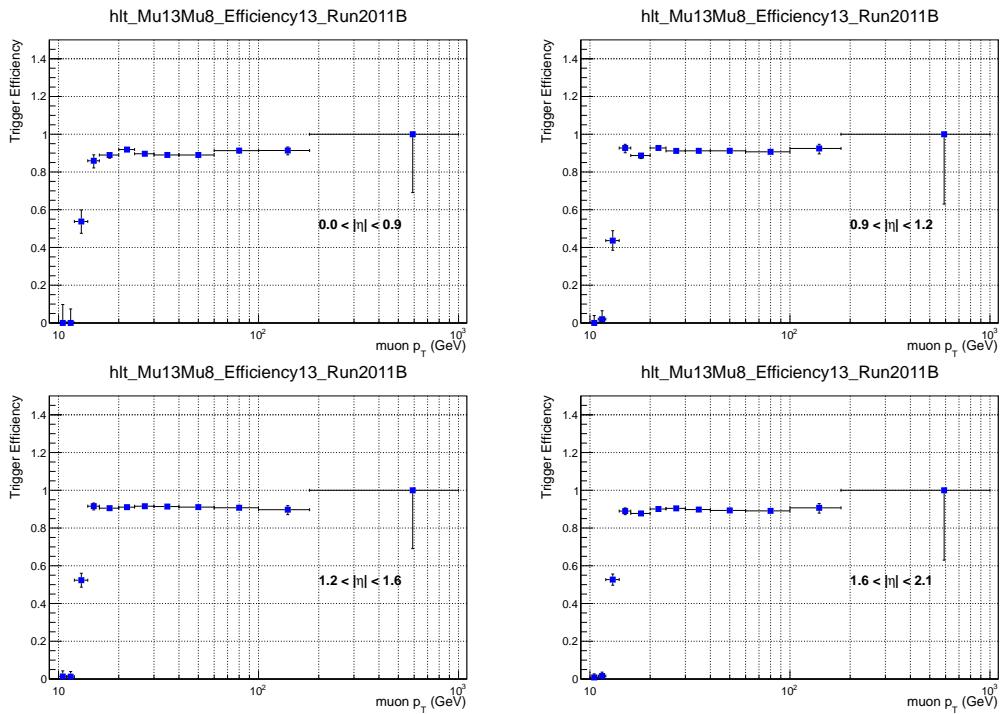


Figure A.3: Trigger turn-on curves for runs in 2011B for the trigger leg of 13 GeV of HLT_Mu13_Mu8 in four pseudorapidity regions as a function of the muon p_T [97]. The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events.

Appendix B

Primary Vertex Reweighting

An alternative method to reproduce the number of primary vertices in data is presented. This method is an iterative procedure, which makes use of the correlation of the number of primary vertices and the number of pileup interactions produced by MC generator (here MADGRAPH + PYTHIA6 is used).

To produce the same number of primary vertices as observed in data, a weight is applied to the simulated events, according to the difference in the number of primary vertices between data and MC simulation. The weight is defined as the ratio

$$\omega = \frac{nPV^{\text{data}} (\langle nPV \rangle)}{nPV^{\text{MC}} (\langle nPV \rangle)}, \quad (\text{B.1})$$

where nPV^{data} and nPV^{MC} indicate the number of primary vertices in data and MC simulation. $\langle nPV \rangle$ is defined as the number of primary vertices averaged over the number of pileup interactions (PUI). In Figure B.1 the relation of the average

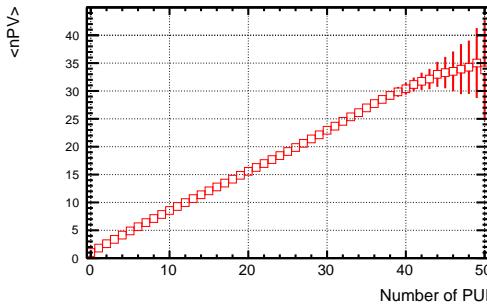


Figure B.1: Average number of reconstructed primary vertices $\langle nPV \rangle$ as a function of the number of pileup interactions (PUI) in simulation.

number of reconstructed vertices and the number of pileup interactions is shown, estimated by MADGRAPH.

The weight in eq. (B.1) is applied on MC events as a function of the average number of primary vertices depending on the number of pileup interactions in the simulation. The weight is applied iteratively in order to obtain a good agreement between data and MC predictions in the number of reconstructed primary vertices. After eight iterations a satisfying agreement between data and MC predictions is observed. The normalised distribution of the number of primary vertices after reweighting is shown

in Figure B.2.

The reweighting procedure used in this thesis is explained in section 3.1.1. The

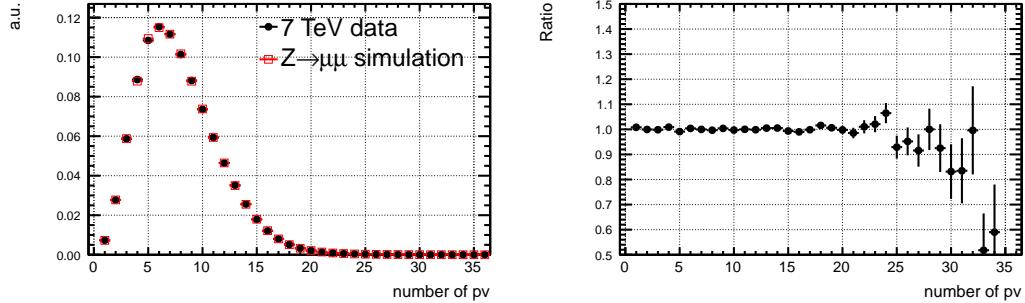


Figure B.2: Number of primary vertices (pv) in data and simulation. The figure on the left shows the distribution after reweighting and the figure on the right shows the ratio of data over simulation. A good agreement between data and MC predictions is observed after the reweighing.

difference between the results, using the two reweighting procedures, is small and within the statistical uncertainties.

Appendix C

Study of Top Quark Background Estimation

The top quark decays via the weak interaction by emitting a W boson and producing a bottom quark. The W boson can decay subsequently into a lepton and neutrino, where the latter can be identified via missing transverse energy.

The top quark process is dominant in the region with $E_T^{\text{miss}} > 80 \text{ GeV}$, to which I refer to as side-band region, in the following. In Figure C.1 the dimuon invariant mass is shown in the region of $E_T^{\text{miss}} > 80 \text{ GeV}$. The top quark background process is dominant over the whole phase space, but largest at high masses with $m^{\mu\mu} > 200 \text{ GeV}$. The top quark background process is simulated with `MADGRAPH` and normalised using the measurement from CMS [124]. In the following the predictions of the $t\bar{t}$ background are validated. It is investigated how well the MC predictions model the $t\bar{t}$ production in comparison to data. I present the results in the side-band region and for high invariant masses of the dimuon with $200 < m^{\mu\mu} < 1500 \text{ GeV}$, due to the fact that it is the dominant region for $t\bar{t}$ production.

In Figure C.2 the dimuon transverse momentum distribution is shown for inclusive

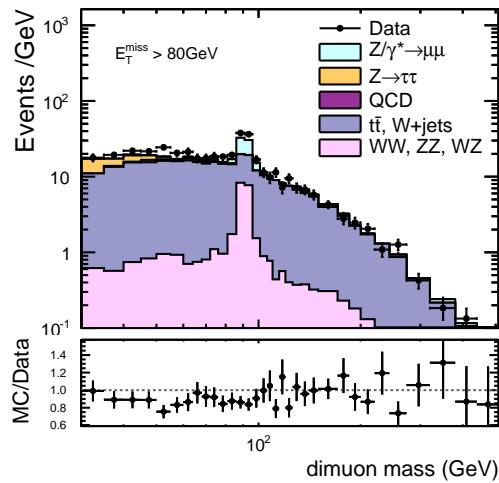


Figure C.1: Distribution of dimuon invariant mass with inverse missing transverse energy requirement. The $t\bar{t}$ background contribution is controlled in the side-band region, $E_T^{\text{miss}} > 80 \text{ GeV}$, where the top quark pair events dominate the sample.

DY, DY + 1 jet, and DY + 2 jets production. The top quark pair contribution

is the main background process in the side-band region. The agreement of data and MC predictions are within statistical uncertainties. As shown previously, in Figure 4.14, the distribution of missing transverse energy is well described by the MC predictions. MADGRAPH MC predictions model well the top quark pair background in the side-band region and this conclusion can also be assigned in the control-region of $E_T^{\text{miss}} < 80 \text{ GeV}$, due to the fact that, the E_T^{miss} variable is well reproduced by MC predictions.

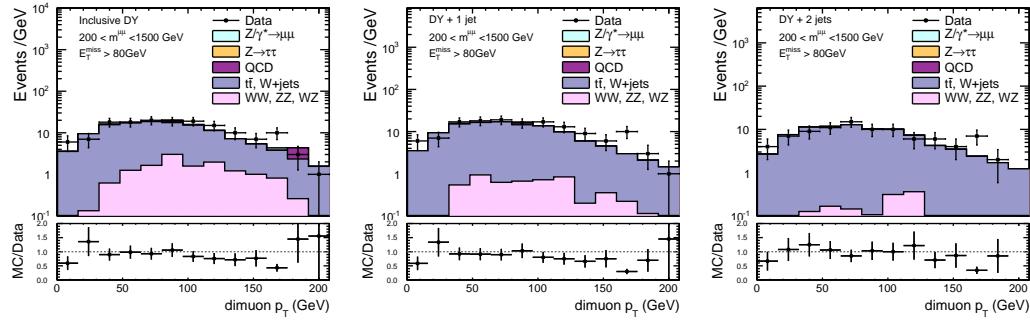


Figure C.2: Distribution of dimuon transverse momentum with inverse missing transverse energy requirement. The $t\bar{t}$ background contribution is controlled in the side-band region, $E_T^{\text{miss}} > 80 \text{ GeV}$, where the top quark pair events dominate the sample. A good agreement of data and MC simulation is shown.

Study of Top Quark Background in association with b-Jets

The top quark decays into a bottom quark before it is able to hadronise. The bottom quark decay via the electroweak force into lighter quarks is highly suppressed by the CKM matrix. Therefore the b quark hadronises into a B-meson before it is able to decay. The jets from bottom quark hadronisation can be identified as b-jets. To identify jets emerging from b-hadrons, several identification algorithms [138] are developed, depending on a discriminator value to distinguish jets, from lighter flavours as e.g. u, d, or s, and jets from heavy flavours, as b and c quarks. The b-jet identification algorithms require an identification as well as an misidentification efficiency. The b-jets are defined by the Combined Secondary Vertex algorithm (CSV) [139] using the medium working point (0.679). The medium working point is defined in such way that the contamination from non b-jets is less than 1%. The efficiency for b-jets and misidentification probability is computed in [139] in $t\bar{t}$ events. The CSV algorithm provides an efficiency of $\sim 70\%$ depending on the p_T and η .

The b-jets are selected in the region with $p_T > 30 \text{ GeV}$ and $|\eta| < 4.5$. In the following a study on the b-jet multiplicity is shown. A requirement to reduce the top quark pair background would include events with zero b-jets. However, the signal Drell-Yan events are influenced by requiring a selection with zero b-jet multiplicity. Therefore, it is a crucial decision to take if to apply a b-jets = 0 selection in order

to reduce the $t\bar{t}$ background, while keeping the number of signal events high. In

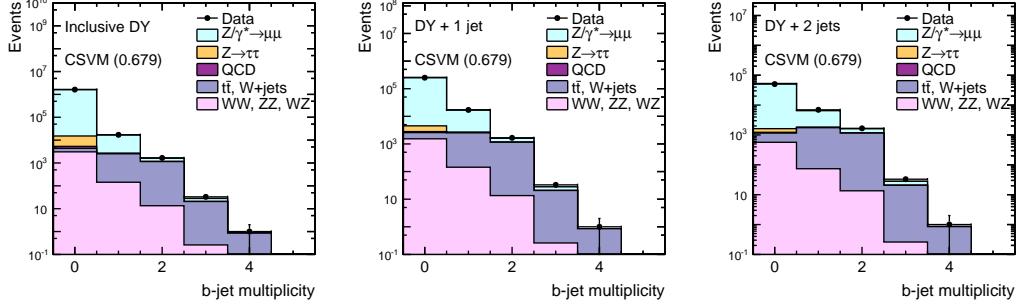


Figure C.3: Number of b-jets for inclusive DY, DY + 1 jet and DY + 2 jets events.

Figure C.3 the b-jet multiplicity is shown for inclusive DY, DY + 1 jet, and DY + 2 jets production. In the case of DY + 2 jets production the background contribution from top quark pair production is dominant.

The efficiency for applying a zero b-jet selection in the case of inclusive DY is 99% for Drell-Yan signal events and 25% for $t\bar{t}$ background events. The $t\bar{t}$ background contribution is therefore reduced significantly, while the signal events are kept with high efficiency. However, for DY + 1 jet and DY + 2 jets production the rejected number of signal events for b-jet multiplicity = 0 is not negligible. In the case of DY + 1 jet production the efficiency for signal Drell-Yan production yields 94% and 22% for $t\bar{t}$ production. Selecting DY + 2 jets the efficiency on the Drell-Yan sample reduces to 90% and the selection efficiency for $t\bar{t}$ production yields 17%.

In Figure C.4 the dimuon transverse momentum distribution is shown for DY + 2 jets in the high invariant mass range of 200 – 1500 GeV. The three cases of inclusive b-jets, b-jet multiplicity = 0, and number of b-jets > 0 is shown. It is observed that $t\bar{t}$ events are rejected mainly by applying zero b-jets selection, but this is also accompanied by a loss of Drell-Yan signal events. Therefore, I do not apply a b-jet requirement in the selection in order not to bias our Drell-Yan signal events.

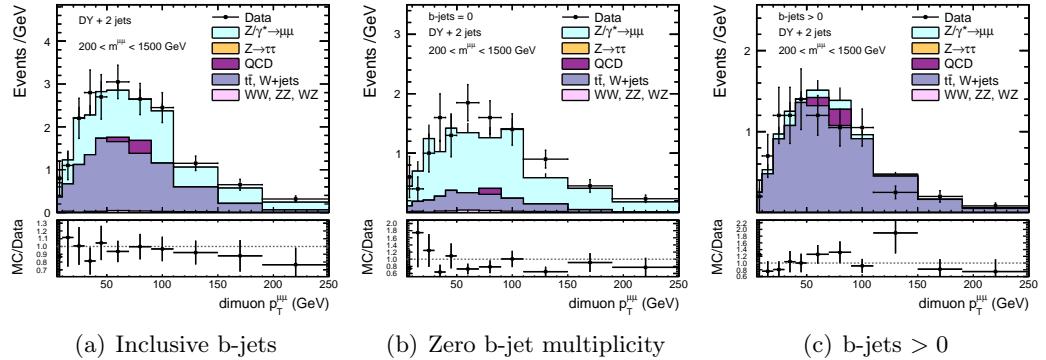


Figure C.4: Dimuon $p_T^{\mu\mu}$ for DY + 2 jets in the invariant mass rage of 200 – 1500 GeV for three different selections in the number of b-jets.

Appendix D

Purity, Stability, Acceptance and Background

The quantities purity and stability (cf. eg. (4.5) and (4.6)) quantify to which extend the measured variable is sensitive to migrations effects [128]. The migrations inside and outside the phase space are illustrated by background and acceptance, eq. (4.7). Here, the quantities are presented for the remaining mass ranges, which were not shown in section 4.5. Furthermore, the migrations matrices are presented.

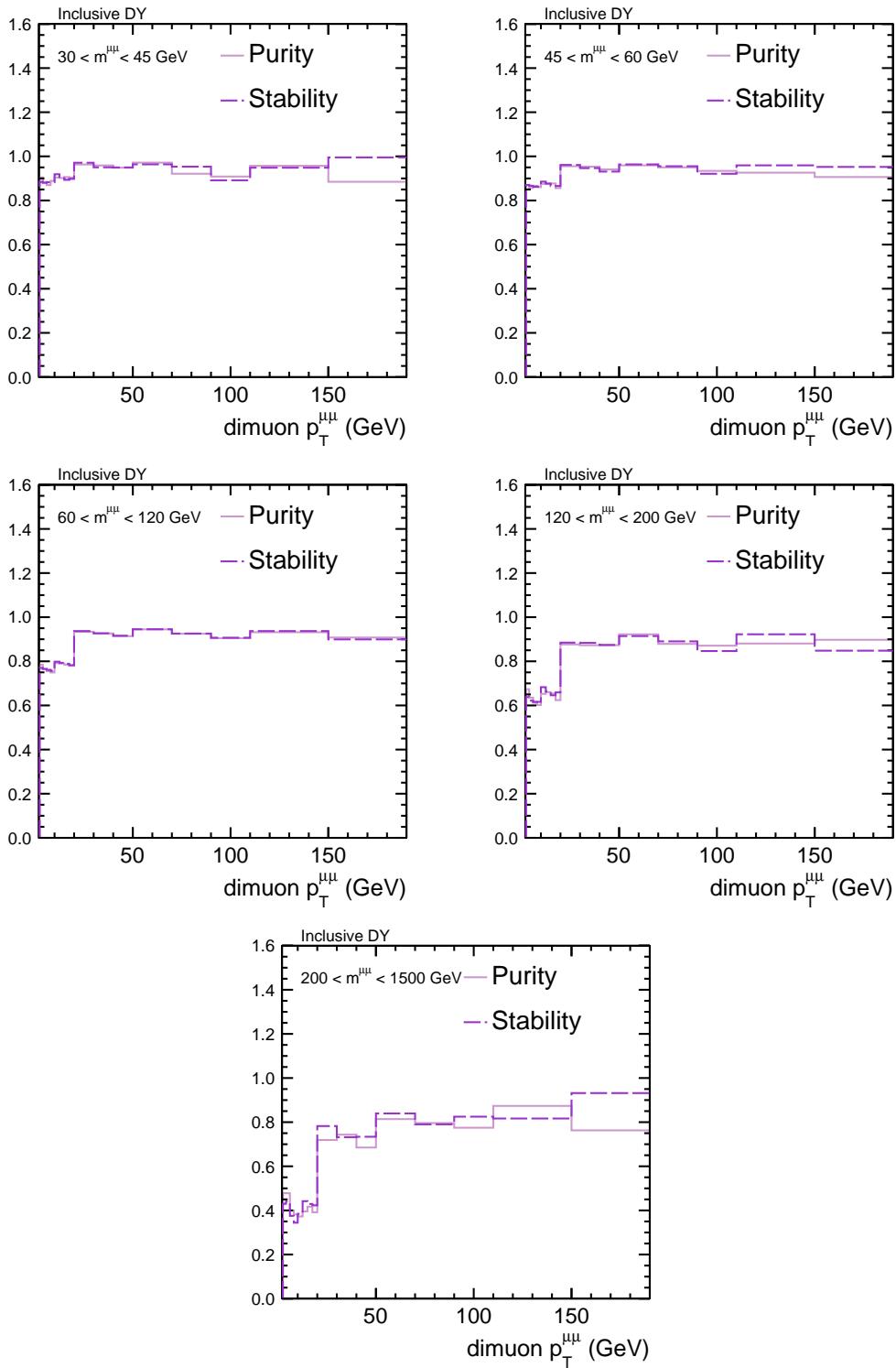


Figure D.1: Purity, stability for inclusive DY transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.

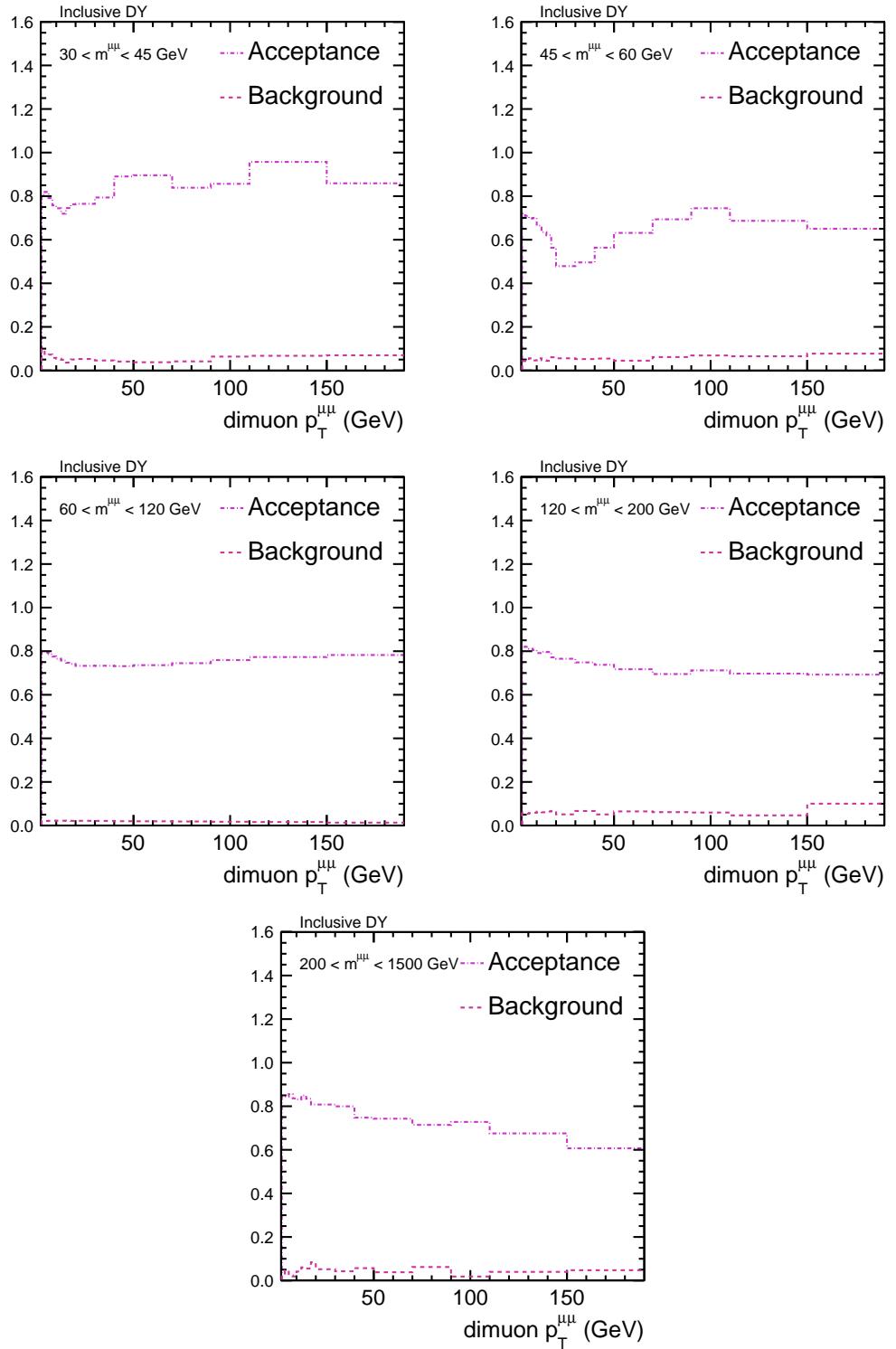


Figure D.2: Background and acceptance for inclusive DY transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.

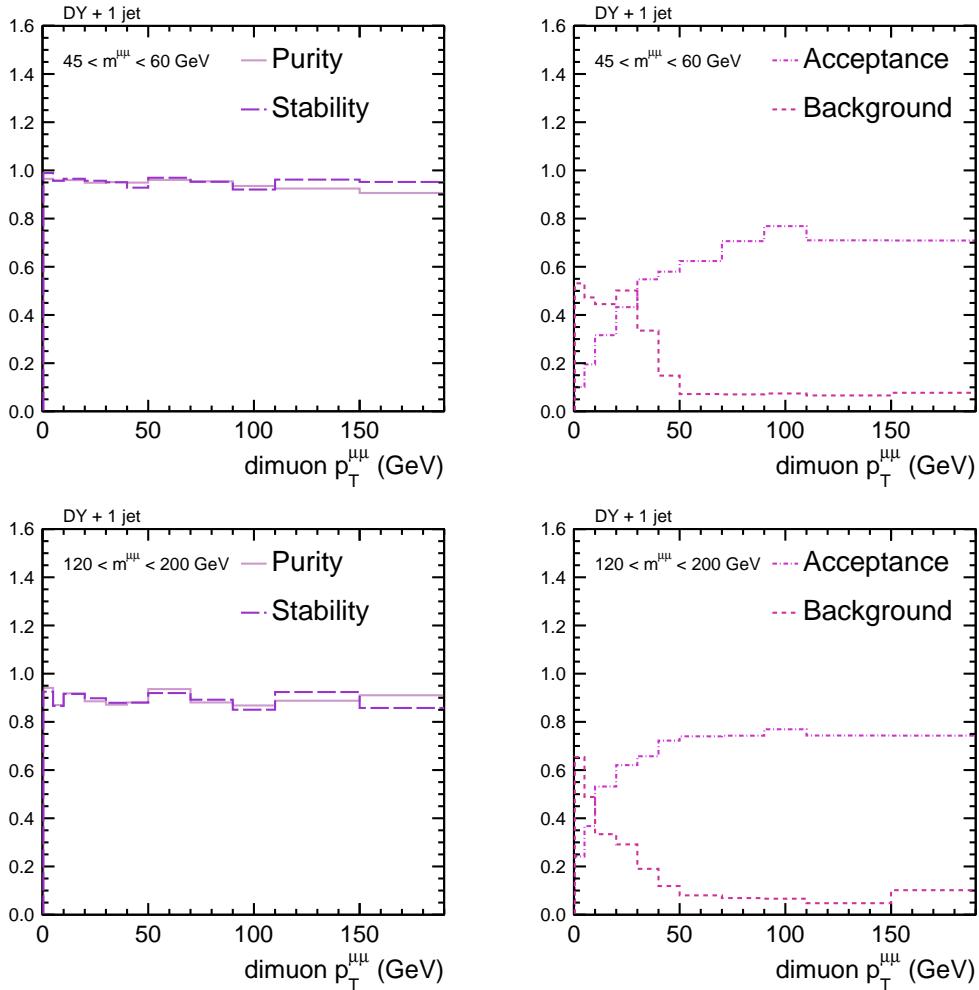


Figure D.3: Purity, stability, background and acceptance for DY + 1 jet transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.

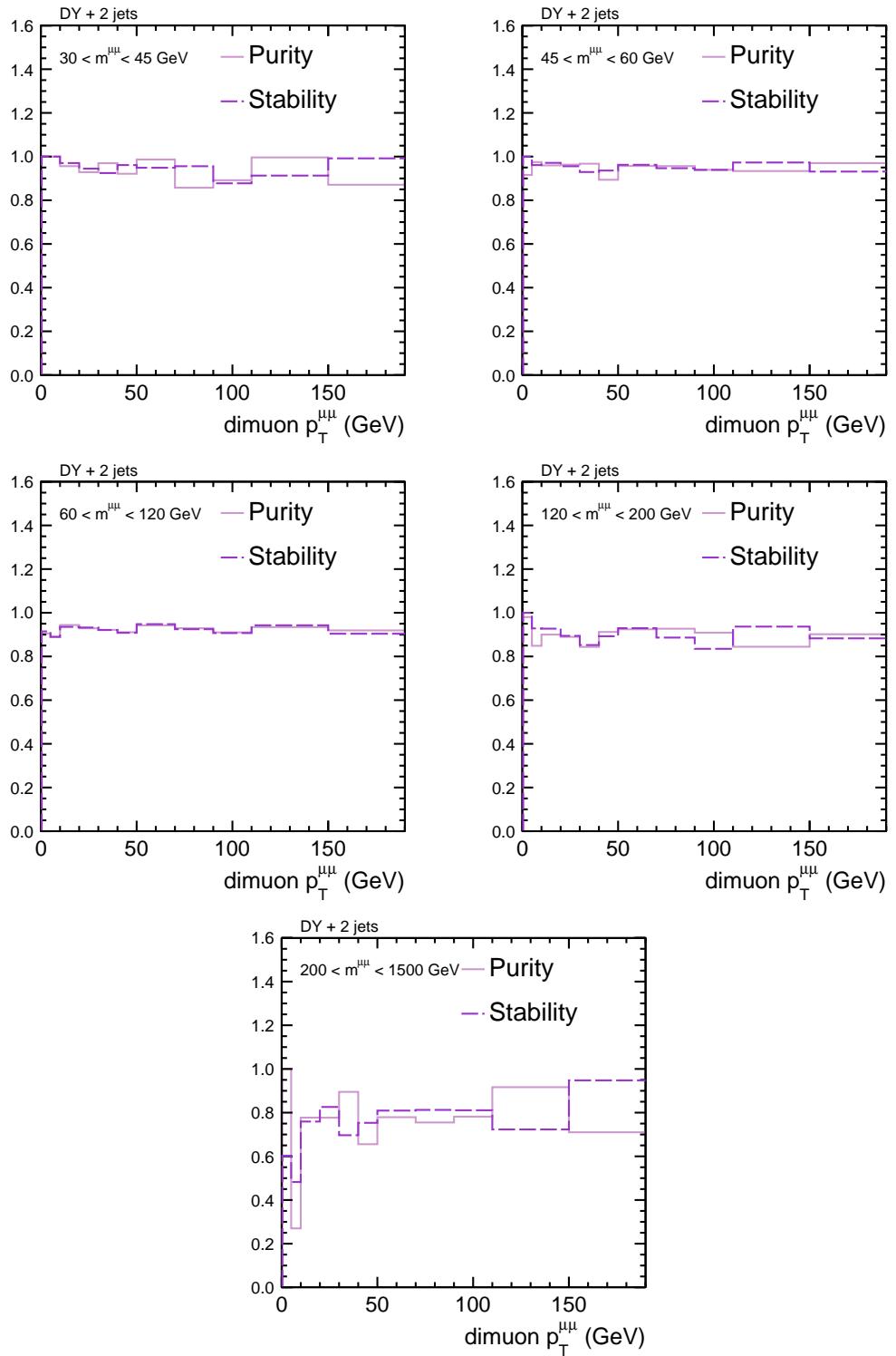


Figure D.4: Purity, stability for DY + 2 jets transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.

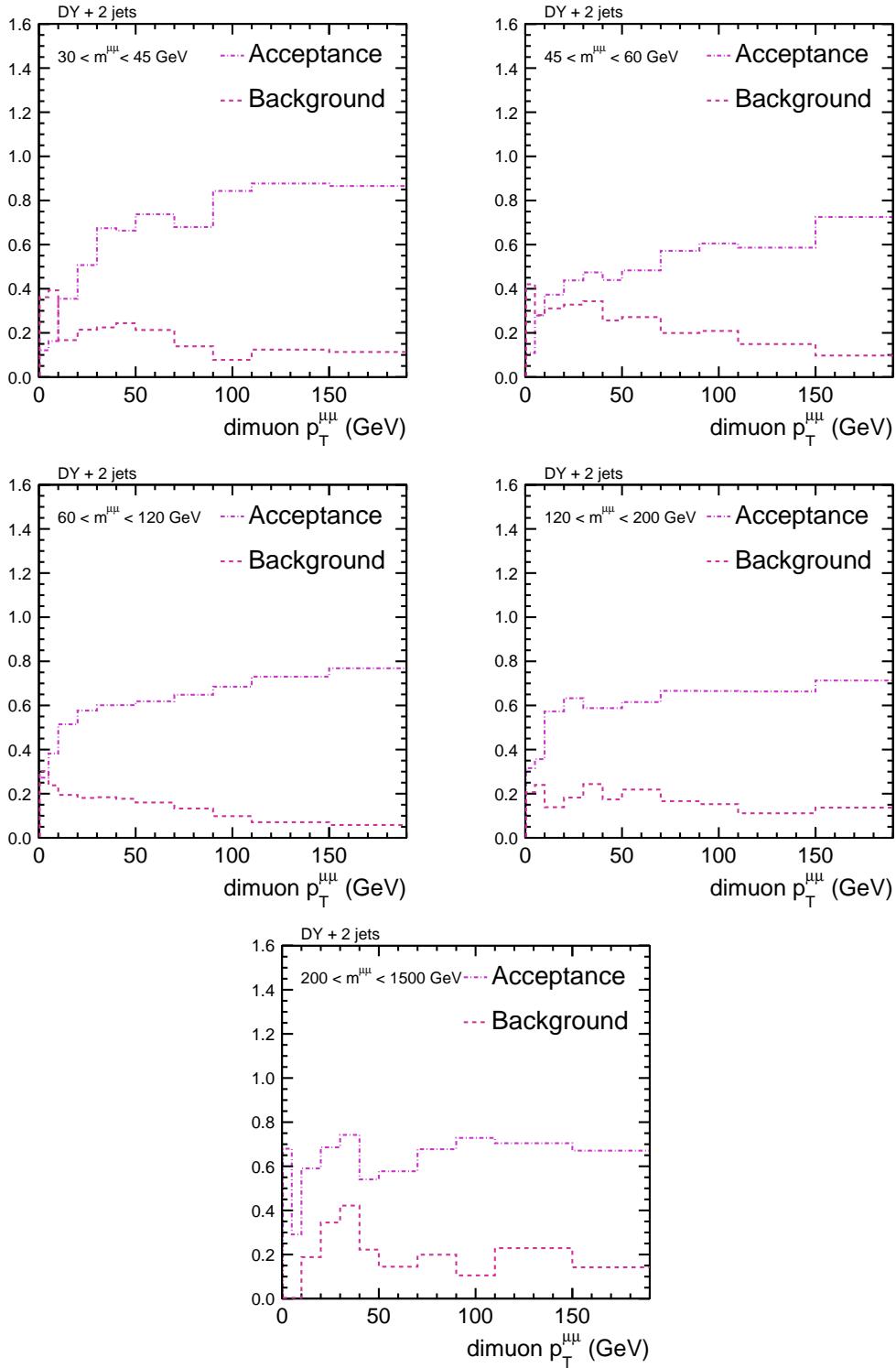


Figure D.5: Background and acceptance for $DY + 2 \text{ jets}$ transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.

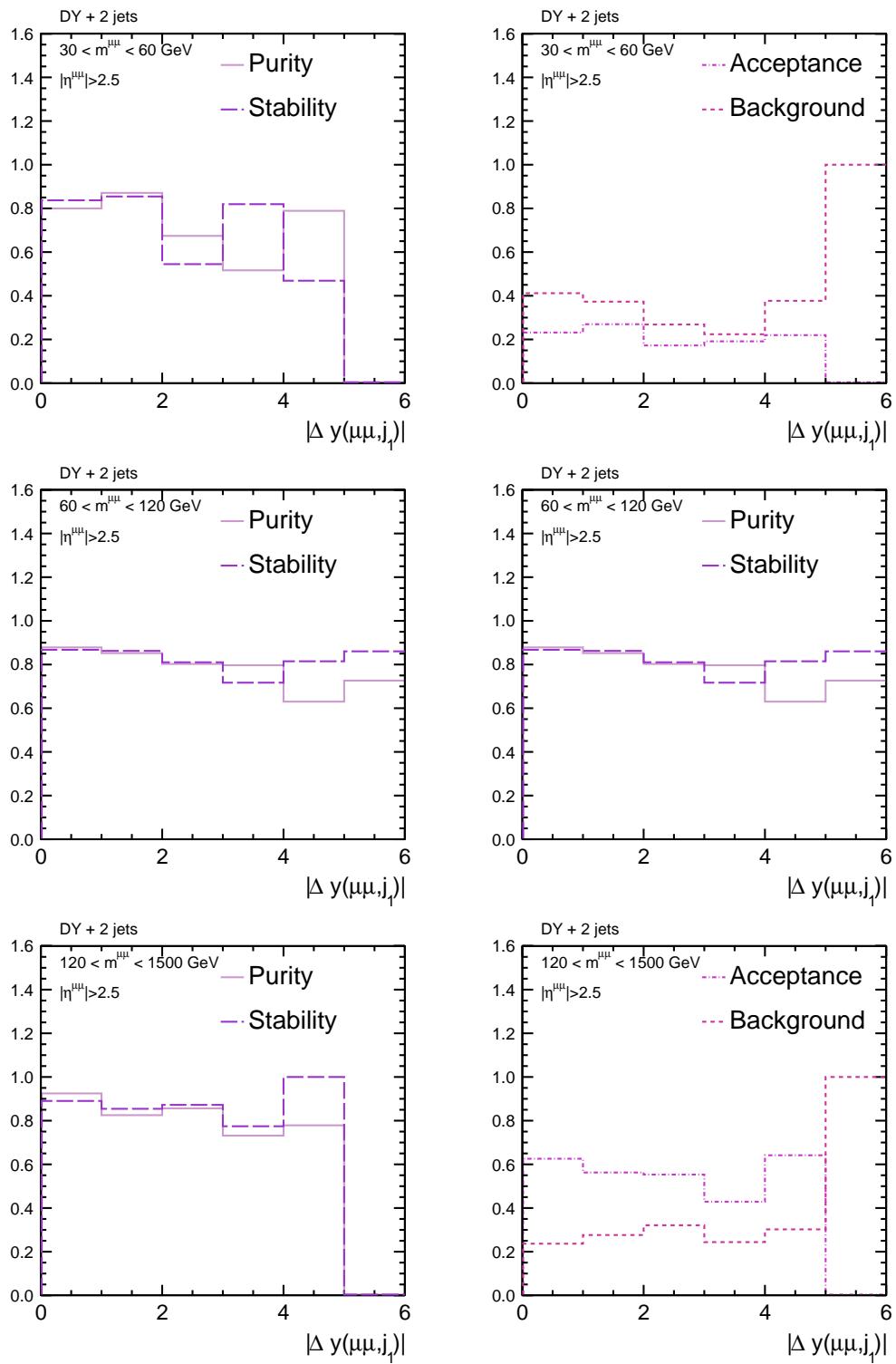


Figure D.6: Purity, stability, background and acceptance as a function of $|\Delta y(\mu\mu, j_1)|$ in $DY + 2 \text{ jets}$ production.

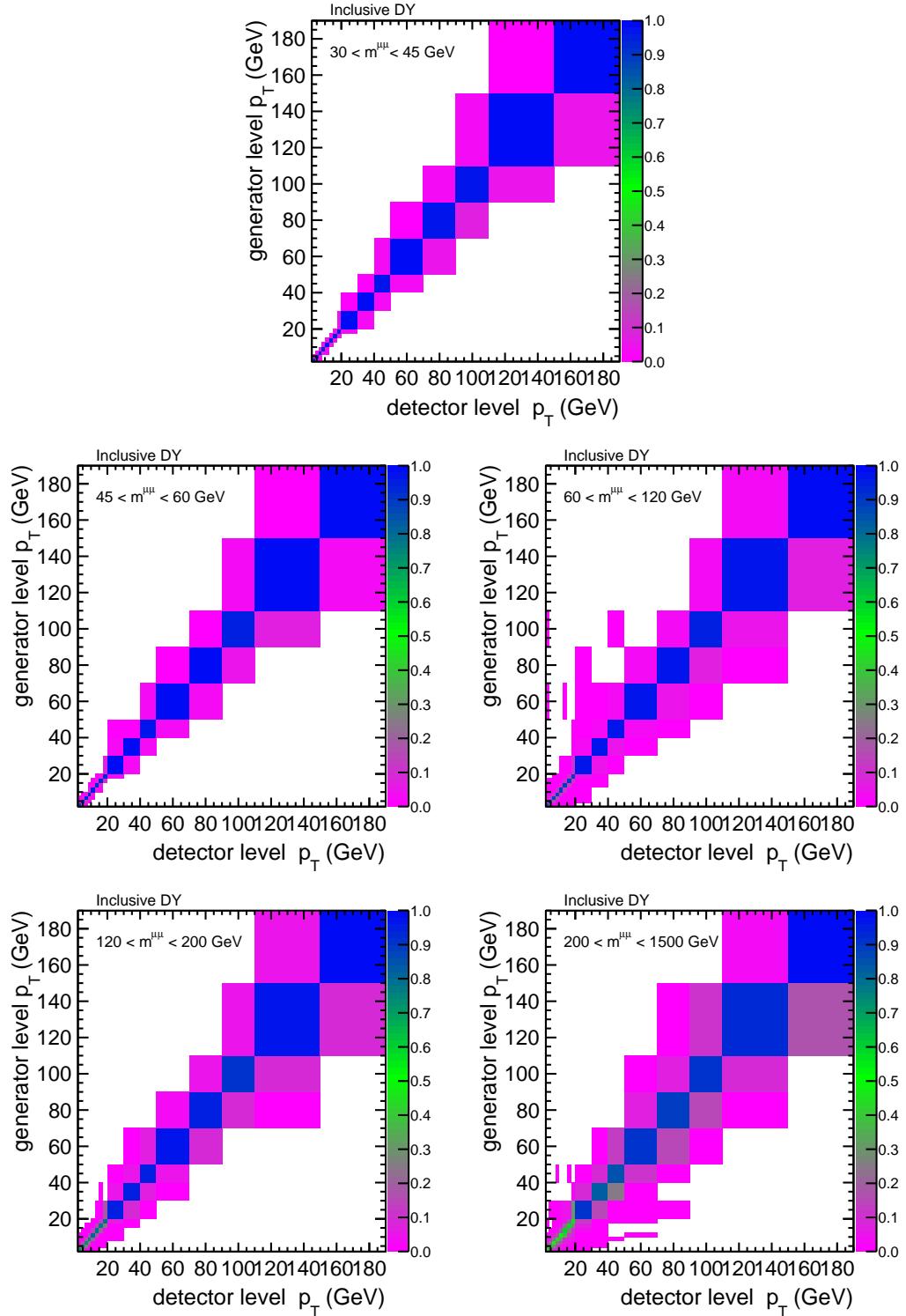


Figure D.7: Normalised migration matrix in matched events for the inclusive DY transverse momentum distribution in bins of the dimuon mass.

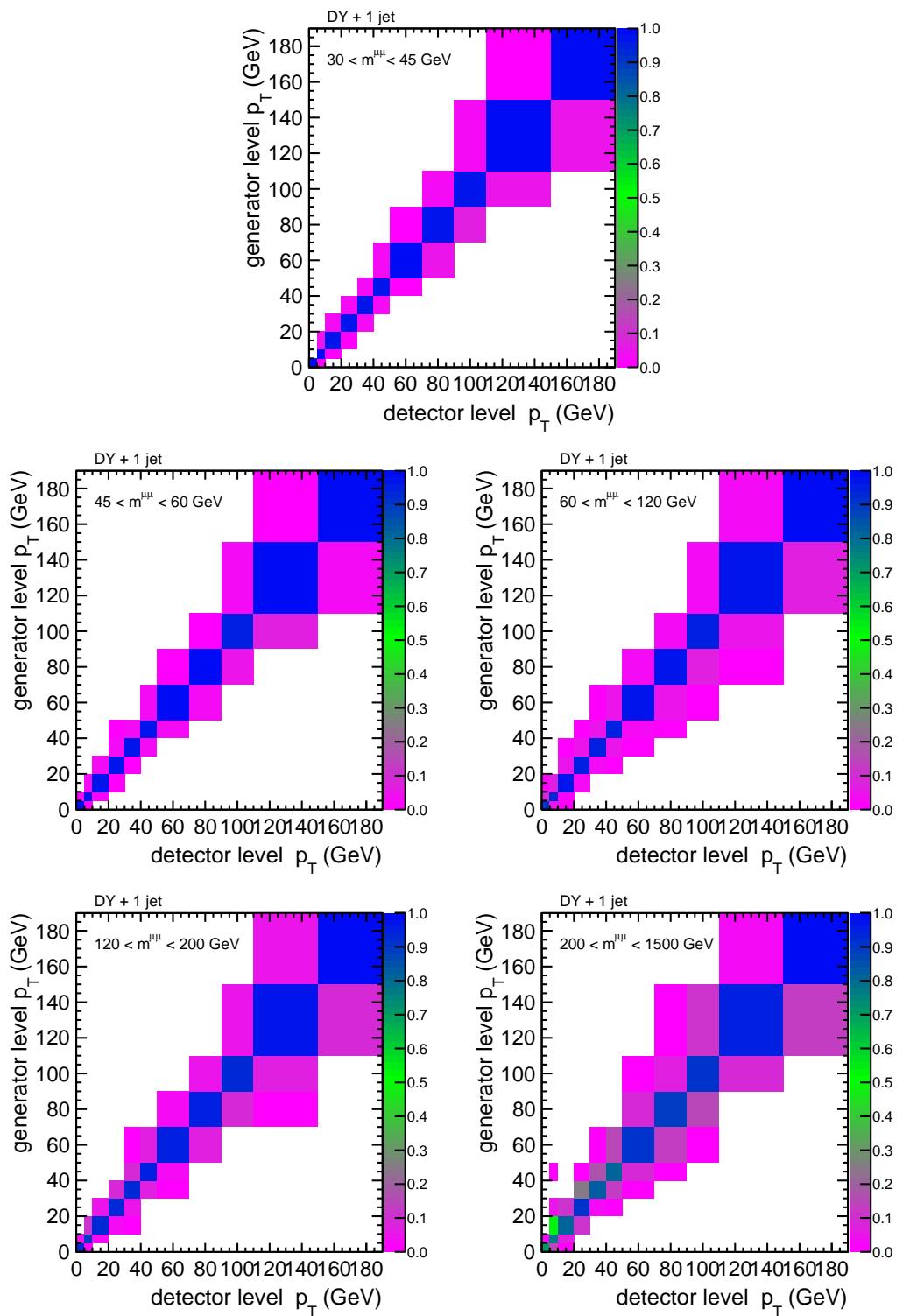


Figure D.8: Normalised migration matrix in matched events for DY + 1 jet transverse momentum distribution in bins of the dimuon mass.

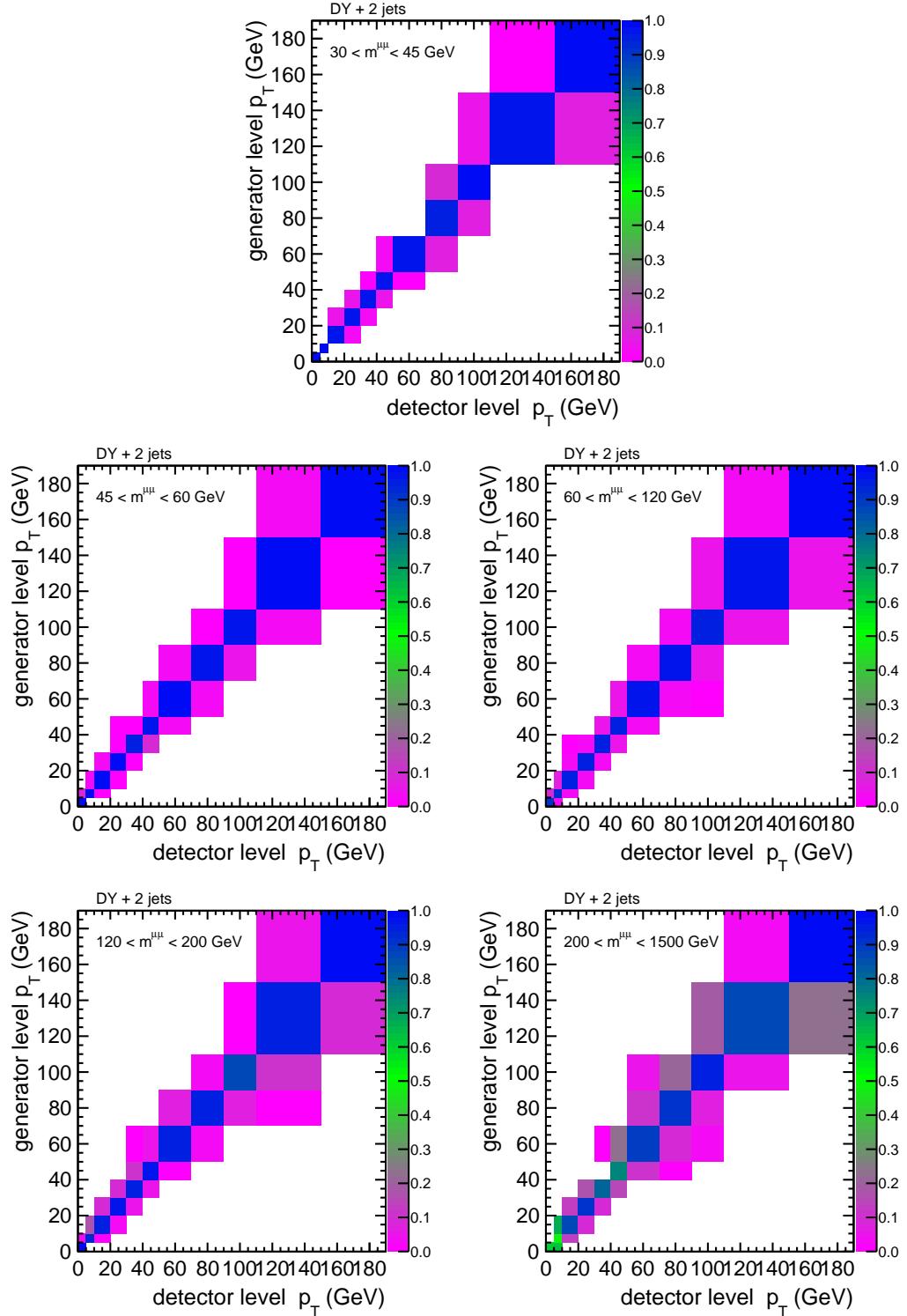


Figure D.9: Normalised migration matrix in matched events for DY + 2 jets transverse momentum distribution in bins of the dimuon mass.

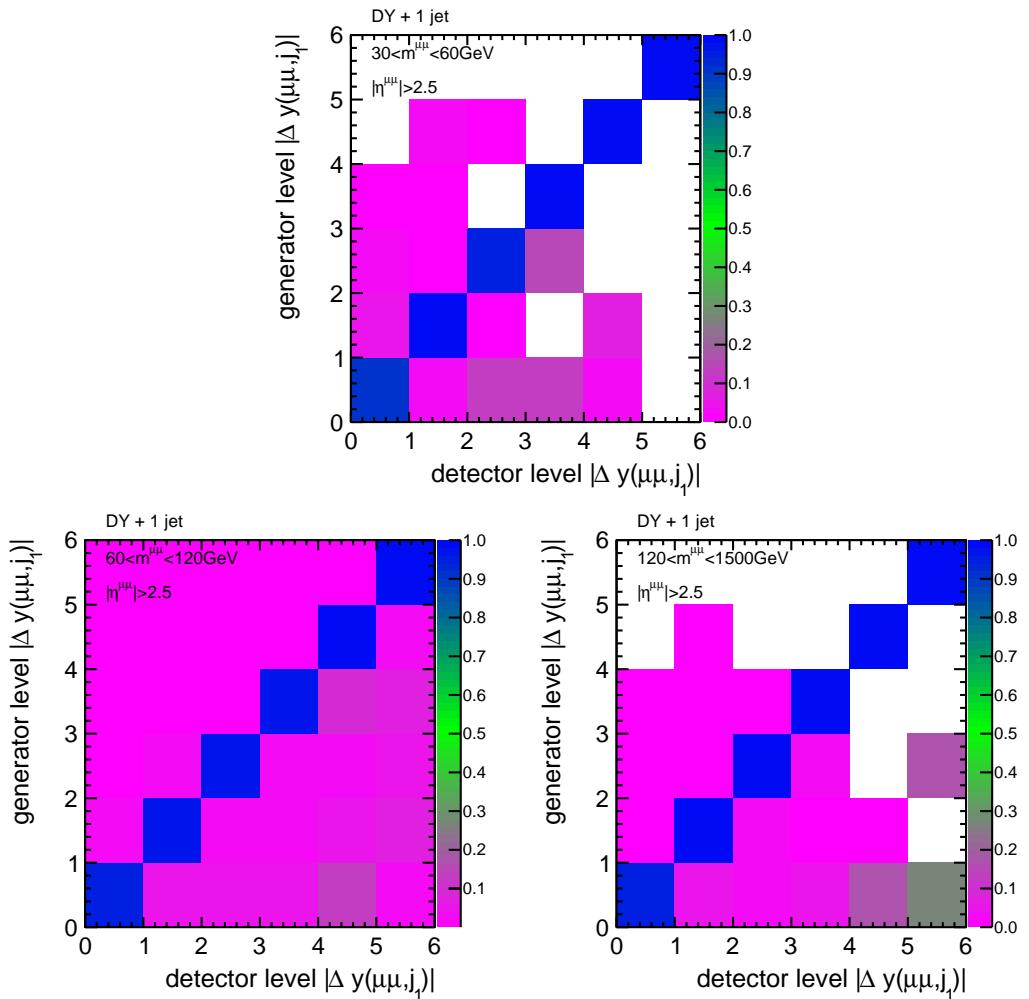


Figure D.10: Normalised migration matrix in for DY + 1 jet $|\Delta y(\mu\mu, j_1)|$ distribution in bins of the dimuon mass.

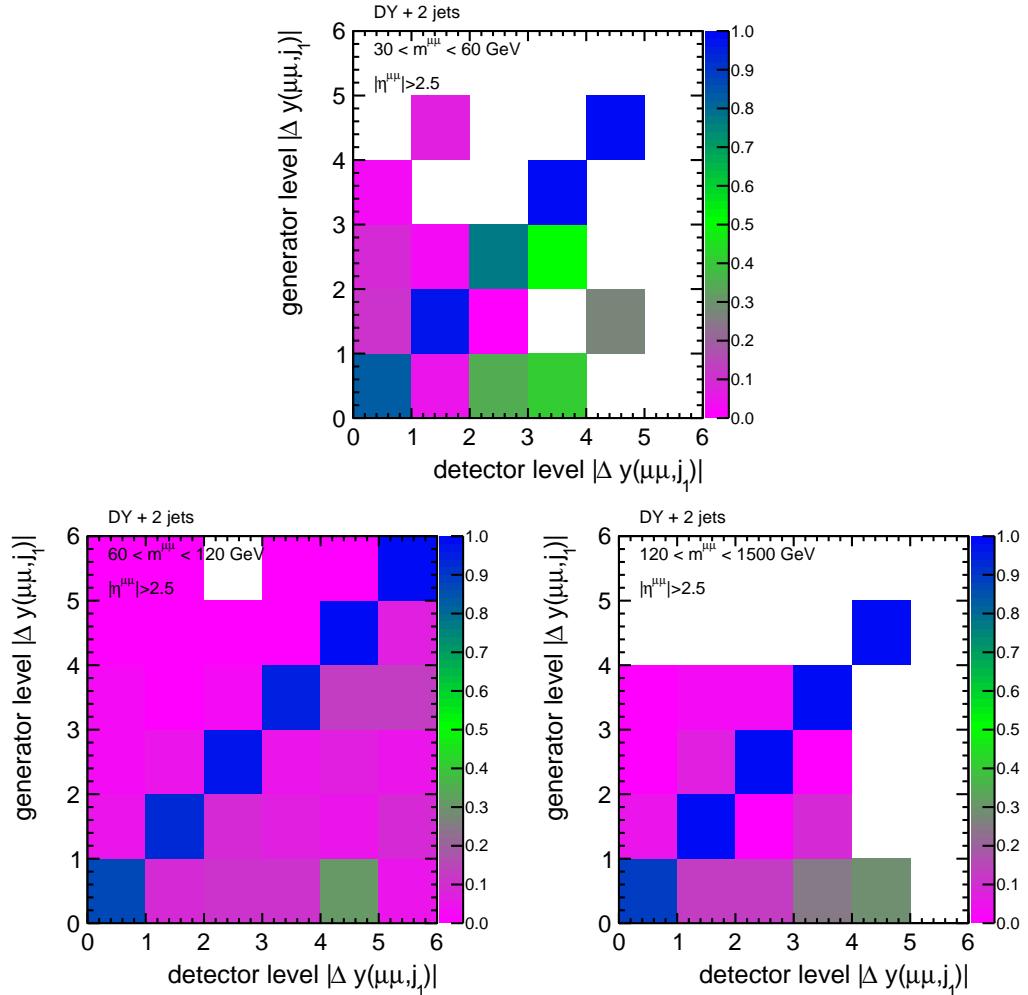


Figure D.11: Normalised migration matrix in for DY + 2 jets $|\Delta y(\mu\mu, j_1)|$ distribution in bins of the dimuon mass.

Appendix E

Underlying Event Tunes and Double Parton Scattering

CMS Physics Analysis Summary

Contact: cms-pag-conveners-generators@cern.ch

2014/04/29

Underlying Event Tunes and Double Parton Scattering

The CMS Collaboration

Abstract

Using the “Rivet” and “Professor” framework, we construct a new PYTHIA 6 tune using the CTEQ6L1 PDF and two new PYTHIA 8 UE tunes (one using CTEQ6L1 and one using the HERAPDF1.5LO). By simultaneously fitting CDF data from $\bar{p}p$ collisions at 300 GeV, 900 GeV, and 1.96 TeV together with CMS data for $p\bar{p}$ collisions at 7 TeV, we test the Underlying Event (UE) models and constrain their parameters, allowing for more precise predictions at 13 TeV and 14 TeV. The consistency of these new tunes with measurements of double-parton scattering (DPS) is also investigated.

1 Introduction

A QCD Monte Carlo model simulation of a hadron-hadron collision, in which a “hard” 2-to-2 parton scattering has occurred, has several components. The “hard scattering” component of the event consists of particles that result from the hadronization of the two outgoing partons (i.e. the primary outgoing two “jets”) plus the particles that arise from initial and final state radiation (ISR & FSR). The “underlying event” (UE) consists of the beam-beam remnants (BBR) and particles that arise from multiple parton interactions (MPI). The BBR are what is left over after a parton is knocked out of each of the initial two beam hadrons. MPI are additional “soft” or “semi-hard” 2-to-2 parton-parton scatterings that occur within the same hadron-hadron collision.

The perturbative $2 \rightarrow 2$ parton-parton differential cross section diverges like $1/\hat{p}_T^4$, where \hat{p}_T is the transverse momentum of the outgoing parton in the parton-parton center-of-mass frame. QCD Monte Carlo models, such as PYTHIA [1–4], regulate this divergence by including a smooth cut-off p_{T_0} as follows: $1/\hat{p}_T^4 \rightarrow 1/(\hat{p}_T^2 + p_{T_0}^2)^2$. This approaches the perturbative result for large scales and is finite as $\hat{p}_T \rightarrow 0$. The primary hard scattering processes and the MPI are regulated in the same way with the one parameter p_{T_0} . However, this cut-off is expected to have a dependence on the center-of-mass energy of the hadron-hadron collision, E_{cm} . PYTHIA parameterizes this energy dependence as follows:

$$p_{T_0}(E_{cm}) = p_{T_0^{REF}} \times (E_{cm}/E_0)^\epsilon, \quad (1)$$

where E_0 is the chosen reference energy and the parameter ϵ determines the energy dependence. At a given center-of-mass energy the amount of MPI depends on the cut-off p_{T_0} , the parton distribution functions (PDF), and the overlap of the matter distributions of the two colliding hadrons. Smaller values of p_{T_0} result in more MPI due to a larger MPI cross-section. Table 1 shows the parameters in PYTHIA 6 [1] and PYTHIA 8 [5] that, together with the selected PDF, determine the energy dependence of MPI. The QCD Monte Carlo generators have parameters which may be adjusted to control the behavior of their event modeling. A specified set of these parameters, that has been adjusted to better fit some aspects of the data, is referred to as a “tune” [6–8].

Table 1: Shows the parameters in PYTHIA 6 and PYTHIA 8 that, together with the selected PDF, determine the energy dependence of MPI. For historic reasons, the reference energy, E_0 , is chosen to be 1.8 TeV.

Parameter	PYTHIA 6	PYTHIA 8
MPI Cut-off, $p_{T_0^{REF}}$, at $E = E_0$	PARP(82)	MultipleInteractions:pT0Ref
Center-of-Mass Reference energy, E_0	PARP(89)	MultipleInteractions:ecmRef
MPI Energy Extrapolation Parameter, ϵ	PARP(90)	MultipleInteractions:ecmPow

Minimum bias (MB) is a generic term which refers to events that are selected with a “loose” trigger that accepts a large fraction of the overall inelastic cross section. To study the UE, we use MB data, however, MB and UE are not the same object. The majority of MB collisions are “soft”, while the UE is studied in events in which a hard-scattering has occurred. One uses the structure of the hard hadron-hadron collision to experimentally study the UE [9]. On an event-by-event basis, a “leading object” is used to define regions of η - ϕ space, where η is the pseudo-rapidity and ϕ is the azimuthal scattering angle. The “transverse” region is perpendicular to the hard-scattering and is very sensitive to the modeling of the UE.

Most of the time, MPI are much “softer” than the primary “hard” scattering; however, occasionally two “hard” $2 \rightarrow 2$ parton scatterings can occur within the same hadron-hadron collision. This is referred to as double parton scattering (DPS) and is typically described in terms of an effective cross section parameter, σ_{eff} , defined as follows:

$$\sigma_{AB} = \frac{\sigma_A \sigma_B}{\sigma_{\text{eff}}}, \quad (2)$$

where σ_A and σ_B are the inclusive cross sections for individual hard scatterings of type A and B, respectively, and σ_{AB} is the cross section for producing both scatterings in the same hadron-hadron collision. If A and B are indistinguishable, as in 4-jet production, a statistical factor of 1/2 must be inserted. Note that this equation holds, only in case a factorization of the two hard scatters is assumed. Furthermore, one expects σ_{eff} to be independent of A and B and within the experimental uncertainties, there is no indication that σ_{eff} is energy dependent [10]. Measurements of σ_{eff} have been recently released in $p\bar{p}$ collision at 7 TeV, by the ATLAS [11] and the CMS [12] experiments in a W+dijet final state. They show that the value of σ_{eff} is around 15-20 mb at 7 TeV. Further measurements have been also performed by older experiments (UA2 [13], AFS [14], and CDF [15]), mainly in a channel with four jets in the final state, at different energies. So far, no evidence of an energy dependence of σ_{eff} has been observed, due to large experimental measurements.

In Section 2, we study the center-of-mass energy dependence of the components of the UE using recent CDF data for $\bar{p}p$ collisions at 300 GeV, 900 GeV, and 1.96 TeV [16] together with CMS data for $p\bar{p}$ collisions at 7 TeV [17]. The 300 GeV and 900 GeV data are a result of the “Tevatron Energy Scan” which was performed just before the Tevatron was shut down. Using the “Rivet” and “Professor” framework [18, 19] we construct a new PYTHIA 6 UE tune with the CTEQ6L1 PDF [20] and two new PYTHIA 8 UE tunes (one using CTEQ6L1 [20] and one using HERAPDF1.5LO [21] PDF sets). The choice of the PYTHIA 6 generator stems from the fact that it has been the Monte Carlo reference since the early data in CMS, and we want to check which level of agreement can be reached by a new tune in the description of UE data at different collision energies; the PYTHIA 8 event generator, instead, is recommended by the authors for future predictions, and it is interesting to compare the performance of new tunes, with predictions obtained with PYTHIA 6. Since the tunes, currently used in CMS, for PYTHIA 6 and PYTHIA 8, respectively Z2*lep and 4C, are interfaced with the CTEQ6L1 PDF set, we perform new tunes by using the same PDF set with the two event generators. Moreover, we perform a new different tune with PYTHIA 8, when using the HERAPDF1.5LO PDF set, since this tune can be used in combination with Next-to-Leading-Order (NLO) calculations and the corresponding NLO PDF set, which is determined from exactly the same data. This is a first attempt of performing a tune, which uses a PDF set with both a LO and a NLO fit. There are other PDF sets in the market with the same feature and they might also be used for a tune extraction in the future. Having fixed the parameters of a given MPI model, one can make an unambiguous prediction of σ_{eff} in Equation 2. In Section 3 we investigate if the values of σ_{eff} determined from fitting the UE in a hard-scattering process are consistent with the values determined from tunes of DPS measurements. Section 4 is reserved for summary and conclusions.

2 New CMS UE Tunes

Previous UE studies in CMS have used the leading charged particle as the “leading object” [22, 23], and a Z-boson [24] as the “leading object”. The CDF and CMS data we use here select the highest transverse momentum charged particle in the event, $p_T \text{ max}$, as the “leading ob-

ject" and includes charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < \eta_{\text{cut}}$, with $\eta_{\text{cut}} = 0.8$. The "toward" region is defined by $\Delta\phi < 60^\circ$ and $|\eta| < \eta_{\text{cut}}$, while the "away" region is $\Delta\phi > 120^\circ$ and $|\eta| < \eta_{\text{cut}}$. In constructing the "transverse" density one adds together the two "transverse" regions: "transverse-1" ($60^\circ < -\Delta\phi < 120^\circ$, $|\eta| < \eta_{\text{cut}}$) and "transverse-2" ($60^\circ < \Delta\phi < 120^\circ$, $|\eta| < \eta_{\text{cut}}$) and divides by the area of η - ϕ space, $\Delta\eta\Delta\phi = 2\eta_{\text{cut}} \times 2\pi/3$. However, the "transverse" region can be separated into the "transMAX" and "transMIN" regions. As shown in Figure 1, on an event-by-event basis, one defines "transMAX" ("transMIN") to be the maximum (minimum) number of charged particles or the scalar p_T sum of charged particles in the two "transverse" regions: "transverse-1" and "transverse-2". Again densities are formed by dividing by the area in η - ϕ space, where the "transMAX" and "transMIN" regions each have an area of $\Delta\eta\Delta\phi = 2\eta_{\text{cut}} \times 2\pi/6$. Hence, the "transverse" density (also referred to as "transAVE") is the average of the "transMAX" and the "transMIN" densities. For events with large initial or final-state radiation the "transMAX" region often contains the third jet, while both the "transMAX" and "transMIN" regions receive contributions from the MPI and BBR components. Thus, the "transMIN" region is very sensitive to the MPI and BBR components of the UE, while "transDIF" ("transMAX" minus the "transMIN") is very sensitive to initial-state radiation (ISR) and the final-state radiation (FSR) [25].

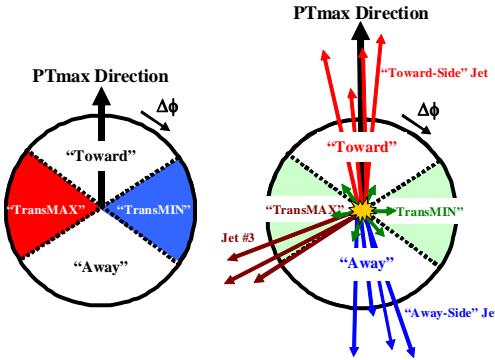


Figure 1: (left) Illustration of correlations in azimuthal angle $\Delta\phi$ relative to the direction of the leading charged particle in the event, p_T max. The relative angle $\Delta\phi = \phi - \phi_{\text{MAX}}$, where ϕ_{MAX} is the azimuthal angle of p_T max and ϕ is the azimuthal angle of a charged particle. On an event by event basis, we define "transMAX" ("transMIN") to be the maximum (minimum) of the two "transverse" regions transverse-1 and transverse-2 shown in Fig. 1. The "transverse" region (i.e. "transAVE") is the average of the "transMAX" and the "transMIN" regions. (right) illustration of the topology of a hadron-hadron collision in which a "hard" parton-parton collision has occurred. The "toward" region contains the "toward-side" jet, while the "away" region, on the average, contains the "away-side" jet. For events with large initial or final-state radiation the "transMAX" region contains the third jet, while both the "transMAX" and "transMIN" regions receive contributions from the MPI and beam-beam remnants. Thus, the "transMIN" region is very sensitive to the MPI and BBR, while the "transMAX" minus the "transMIN" (i.e. "transDIF") is very sensitive to initial and final-state radiation (ISR & FSR).

2.1 PYTHIA 6 UE Tune

The PYTHIA 6 Tune Z2*lep [23] is a previous CMS UE tune which was constructed by fitting the CMS charged-particle jet UE data at 900 GeV and 7 TeV [22]. Furthermore, it uses an improved set of fragmentation parameters. Only data on the "transAVE" charged particle and p_T sum density were used, since data on "transMAX", "transMIN", and "transDIF" were not available at that time. Starting with the Tune Z2*lep parameters, we construct a

new PYTHIA 6 tune by varying the five parameters shown in Table 2 in an attempt to fit the “transMAX” and “transMIN” charged particle p_T sum densities at four center-of-mass energies: 300 GeV, 900 GeV, 1.96 TeV, and 7 TeV. We use the same PDF as Tune Z2*lep, CTEQ6L1 [20]. In addition to varying the MPI energy dependence parameters (Table 1), we also vary the color-reconnection suppression, PARP(77), the color-reconnection strength, PARP(78), and the core matter fraction, PARP(83). It is not necessary to include the “transAVE” and “transDIF” densities in the fit, since they can be constructed from “transMAX” and “transMIN”. Using the “Rivet” and “Professor” framework we arrive at the best fit values of the five parameters shown in Table 2. The new tunes are called after CMS TUNE PYTHIA6 SET 1: CUETP6S1.

Table 2: Shows the tuning range, Tune Z2*lep values, and the best fit values for the CUETP6S1 tune. We use the same PDF as Tune Z2*lep, CTEQ6L1.

PYTHIA 6 Parameter	Tuning Range	Tune Z2*lep	CUETP6S1
PARP(82) - MPI Cut-off (GeV)	1.6 - 2.2	1.921	1.9096
PARP(90) - MPI Energy Extrapolation	0.18 - 0.28	0.227	0.2479
PARP(77) - CR Suppression	0.25 - 1.15	1.016	0.6646
PARP(78) - CR Strength	0.2 - 0.8	0.538	0.5454
PARP(83) C- Matter fraction in core	0.1 - 1.0	0.356	0.8217
Reduced χ^2	-	-	0.915

Figures 2-5 compare the CDF data 300 GeV, 900 GeV, and 1.96 TeV and the CMS data at 7 TeV on the charged particle and p_T sum densities in the “transMIN” and “transMAX” regions as defined by the leading charged particle, p_T max, as a function of p_T max with PYTHIA 6 Tune Z2*, PYTHIA 6 Tune Z2*lep and with the new CMS PYTHIA 6 tune, CUETP6S1. This is the first tune to use the “transMAX” and “transMIN” data. Previously we only had data on “transAVE”. Tune Z2* and Tune Z2*lep fit the data remarkably well considering they were constructed with “transAVE” data at just two energies; 900 GeV and 7 TeV. However, the new CMS PYTHIA 6 tune, CUETP6S1, does an even better job describing the data.

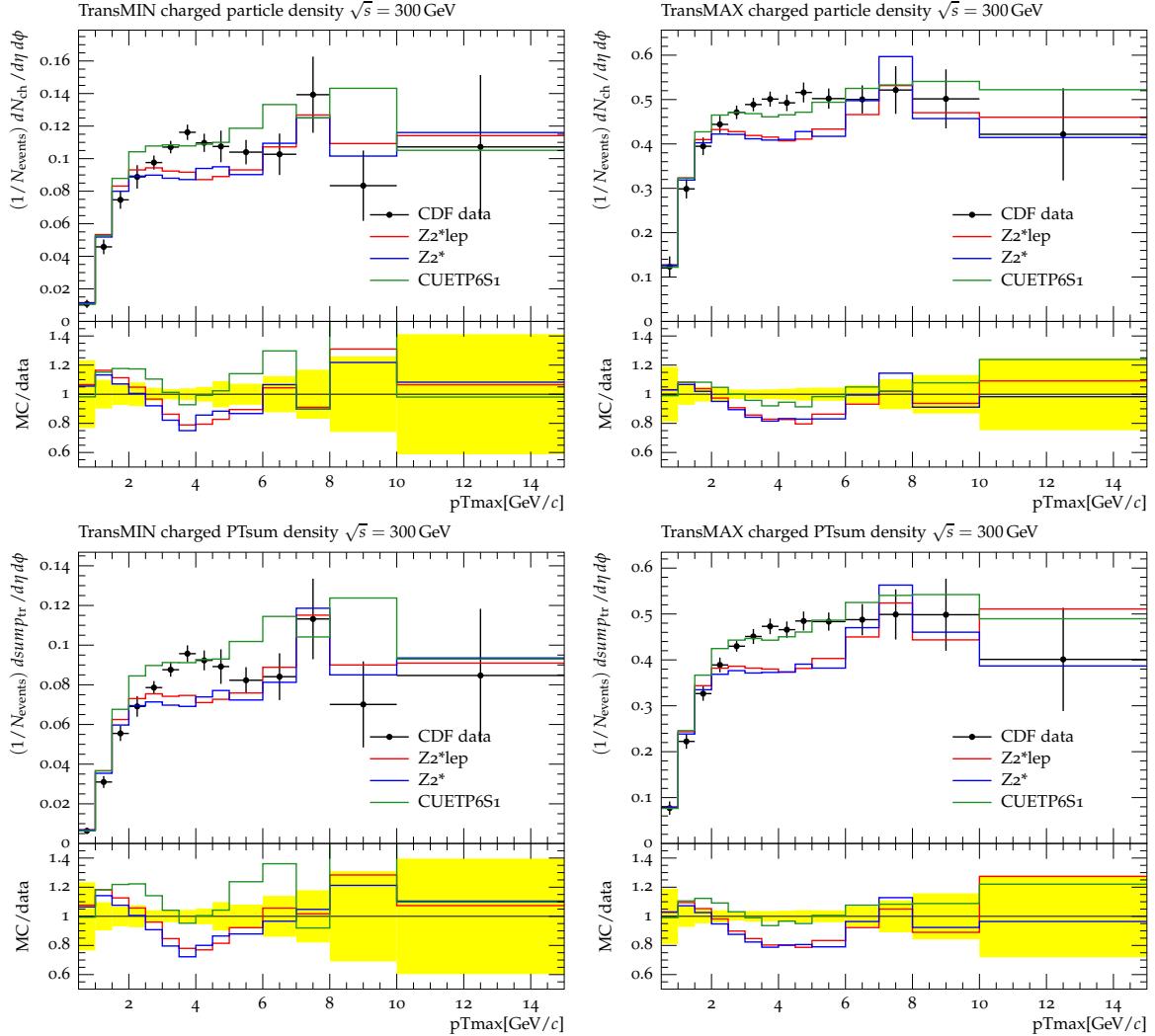


Figure 2: CDF data for $\bar{p}p$ collisions at 300 GeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transMIN” (left column) and “transMAX” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 6 tune Z_2^* , Tune $Z_2^*\text{lep}$ and the new CMS PYTHIA 6 tune.

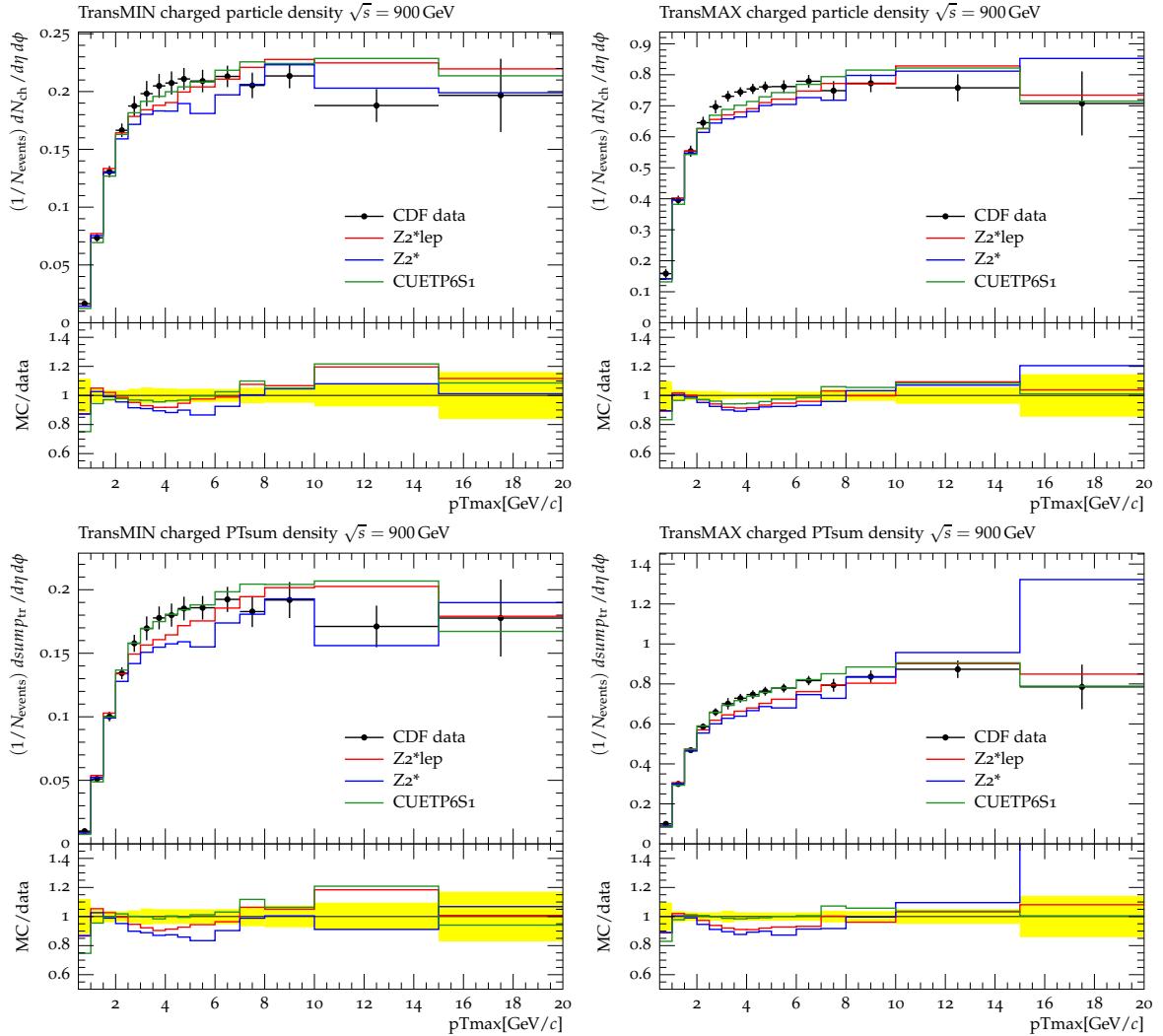


Figure 3: CDF data for $\bar{p}p$ collisions at 900 GeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transMIN” (left column) and “transMAX” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 6 tune $Z2^*$, Tune $Z2^*\text{lep}$ and the new CMS PYTHIA 6 tune.

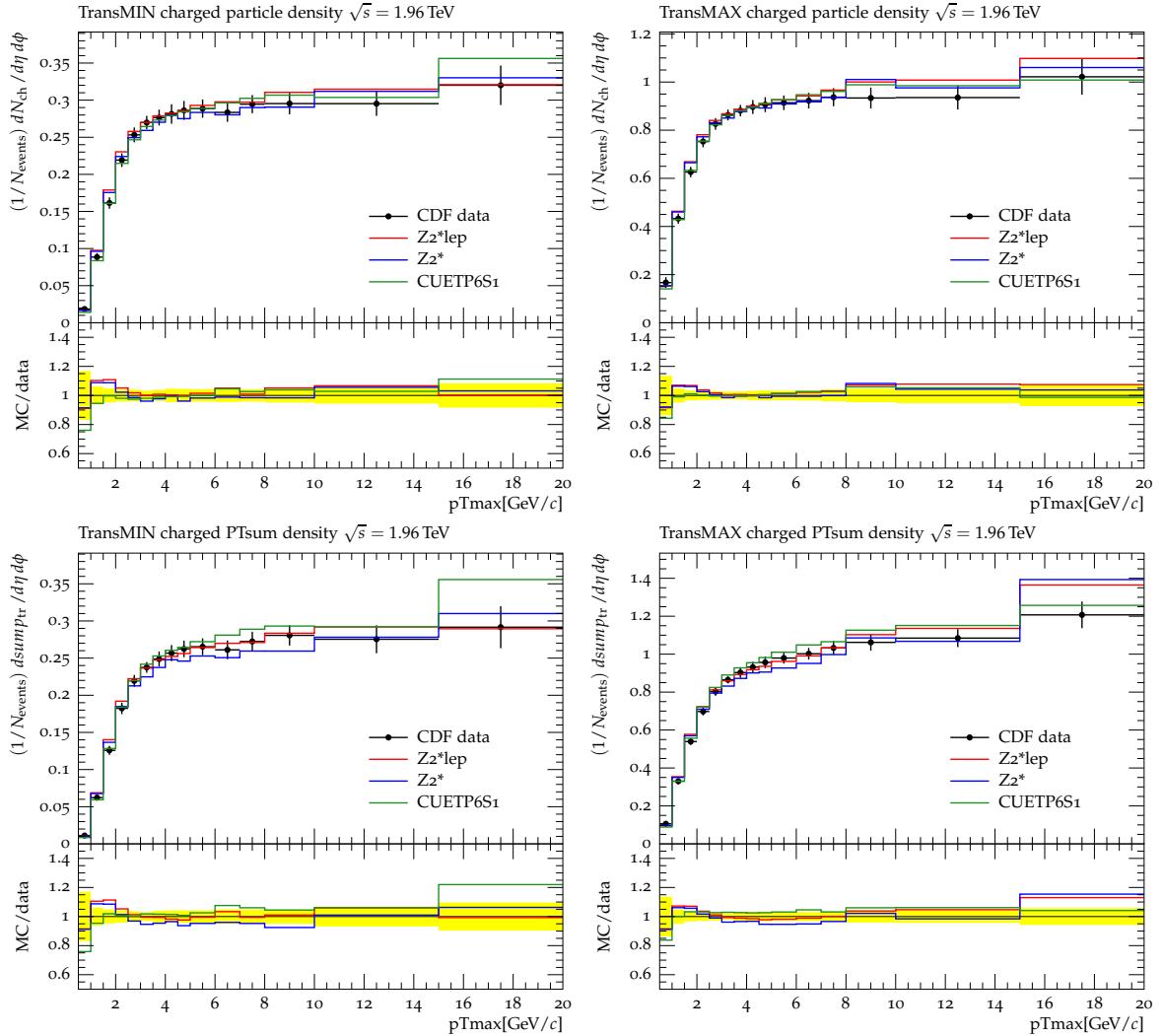


Figure 4: CDF data for $\bar{p}p$ collisions at 1.96 TeV: charged particle density (top row) and p_{T} sum density (bottom row) for charged particles with $p_{\text{T}} > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transMIN” (left column) and “transMAX” (right column) regions as defined by the leading charged particle, as a function of p_{T} max. The data are compared with PYTHIA 6 tune Z2*, Tune Z2*lep and the new CMS PYTHIA 6 tune.

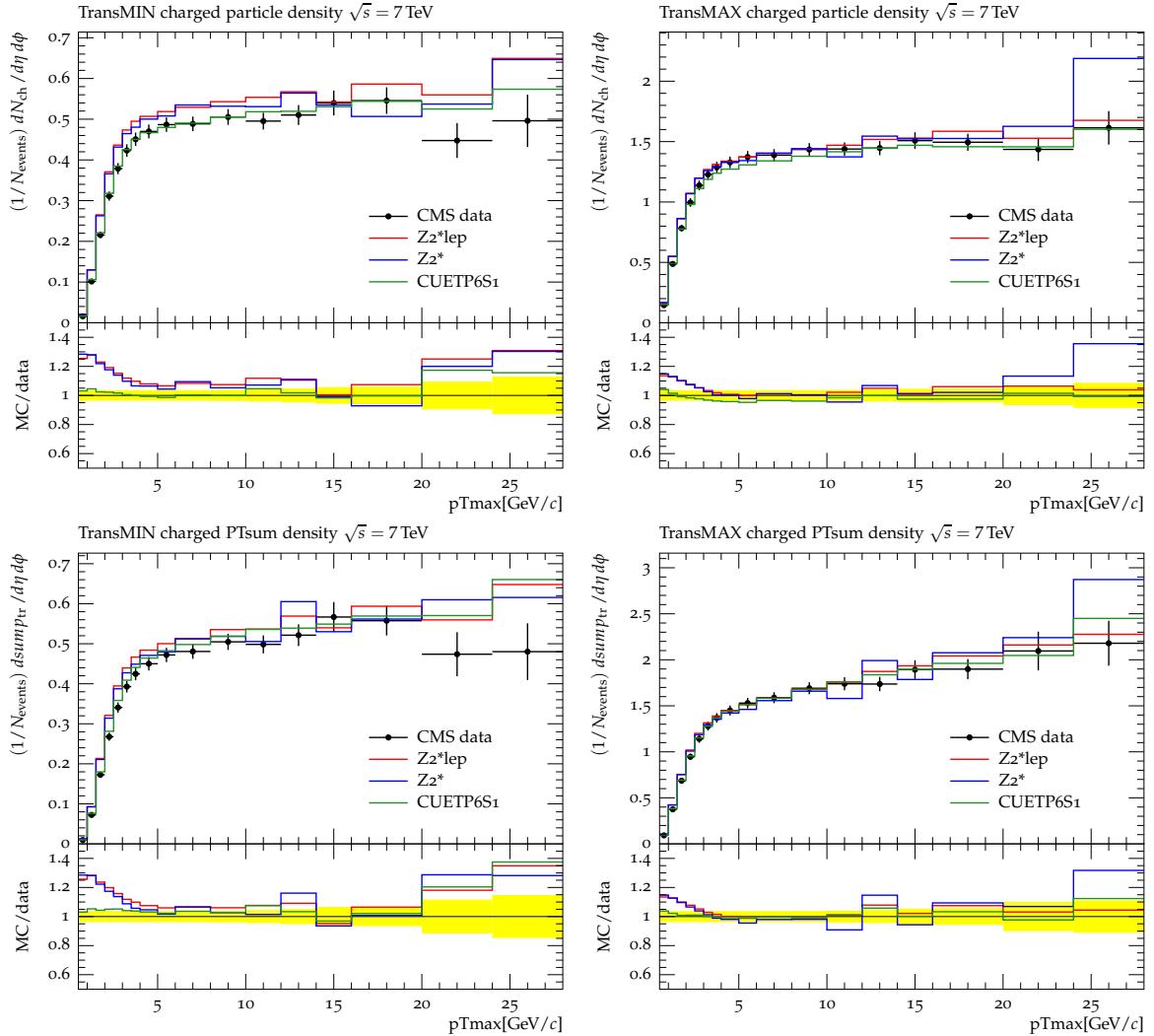


Figure 5: CMS data for pp collisions at 7 TeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transMIN” (left column) and “transMAX” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 6 tune Z2*, Tune Z2*lep and the new CMS PYTHIA 6 tune.

2.2 PYTHIA 8 UE Tunes

Starting with the parameters of Corke and Sjöstrand's PYTHIA 8 Tune 4C [26], we construct two new tunes, one using the CTEQ6L1 PDF [20] and one using the HERAPDF1.5LO [21]. The tunes are constructed by varying the four parameters shown in Table 3 in an attempt to fit the "transMAX" and "transMIN" charged particle and p_T sum densities at three center-of-mass energies: 900 GeV, 1.96 TeV, and 7 TeV. In addition to varying the MPI energy dependence parameters (Table 1), we also vary BBR reconnect range, `reconnectRange`. As is done in Tune 4C, we use an exponential matter overlap function, by setting the switch "bProfile" to 3 in the PYTHIA 8 run card, and vary the exponential shape, `expPow`. Here we exclude the 300 GeV data, because by starting with the parameters of Tune 4C and varying the four parameters in Table 3, we are unable to get a good fit to these data. The parameters obtained for the new two tunes are shown in Table 3. Note that the amount of colour reconnection changes between the two new tunes; it is higher for the HERAPDF1.5LO and lower for the CTEQ6L1 PDF set. This feature depends on the parton dynamics at small longitudinal momenta, which is different for the two PDF sets. The new tunes are called after CMS UE TUNE PYTHIA8 SET1: CUETP8S1.

Table 3: Shows the tuning range, Tune 4C values, and the best fit values for the PYTHIA 8 CTEQ6L1 and HERAPDF1.5LO tunes, respectively called CUETP8S1-CTEQ6L1 and CUETP8S1-HERAPDF1.5LO.

PYTHIA 8 Parameter	Tuning Range	Tune 4C	CUETP8S1- CTEQ6L1	CUETP8S1- HERAPDF1.5LO
MultipleInteractions:pT0Ref (GeV)	1.0 - 3.0	2.085	3.1006	2.0001
MultipleInteractions:ecmPow	0.0 - 0.4	0.19	0.2106	0.2499
MultipleInteractions:expPow	0.4 - 10.0	2.0	1.6089	1.6905
BeamRemnants:reconnectRange	0.0 - 9.0	1.5	3.3126	6.0964
Reduced χ^2	-	-	0.952	1.13

Figures 6-8 compare the CDF data 900 GeV, and 1.96 TeV and the CMS data at 7 TeV on the charged particle and p_T sum densities in the "transMIN" and "transMAX" regions as defined by the leading charged particle, as a function of p_T max with PYTHIA 8 Tune 4C and with the two new CMS PYTHIA 8 tunes. Both the new CMS PYTHIA 8 tunes are an improvement over Tune 4C and describe the data fairly well. The PYTHIA 8 tunes do not describe very well the data at 300 GeV, while the new PYTHIA 6 tune describes data very well. This is because the PYTHIA 6 tune uses a double-Gaussian matter distribution, while the PYTHIA 8 tunes use a single exponential matter overlap function. PYTHIA 8 tunes with double-Gaussian matter distributions might improve the comparisons at 300 GeV.

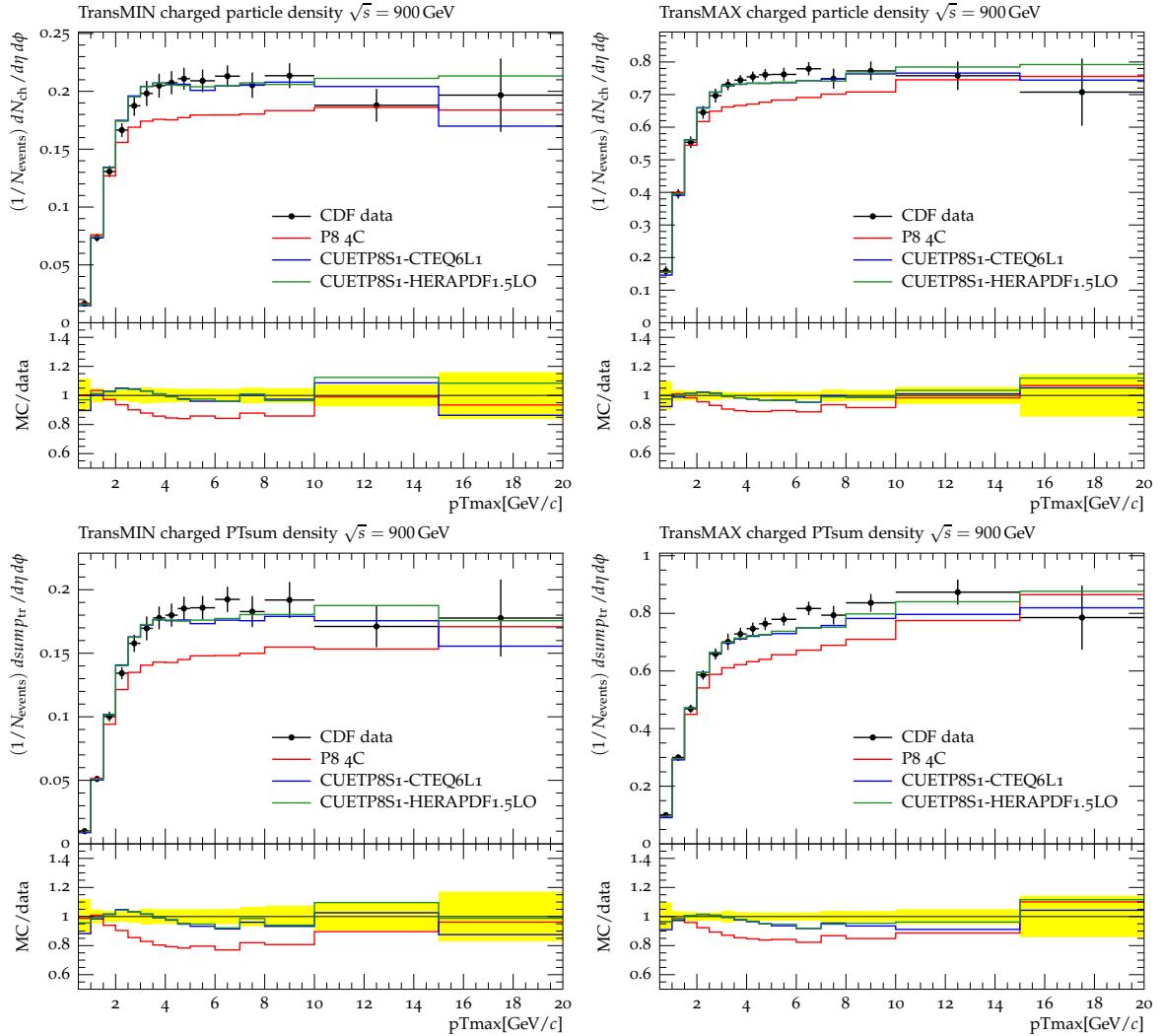


Figure 6: CDF data for $\bar{p}p$ collisions at 900 GeV: charged particle density (top row) and p_{T} sum density (bottom row) for charged particles with $p_{\text{T}} > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transMIN” (left column) and “transMAX” (right column) regions as defined by the leading charged particle, as a function of p_{T} max. The data are compared with PYTHIA 8 Tune 4C and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO.

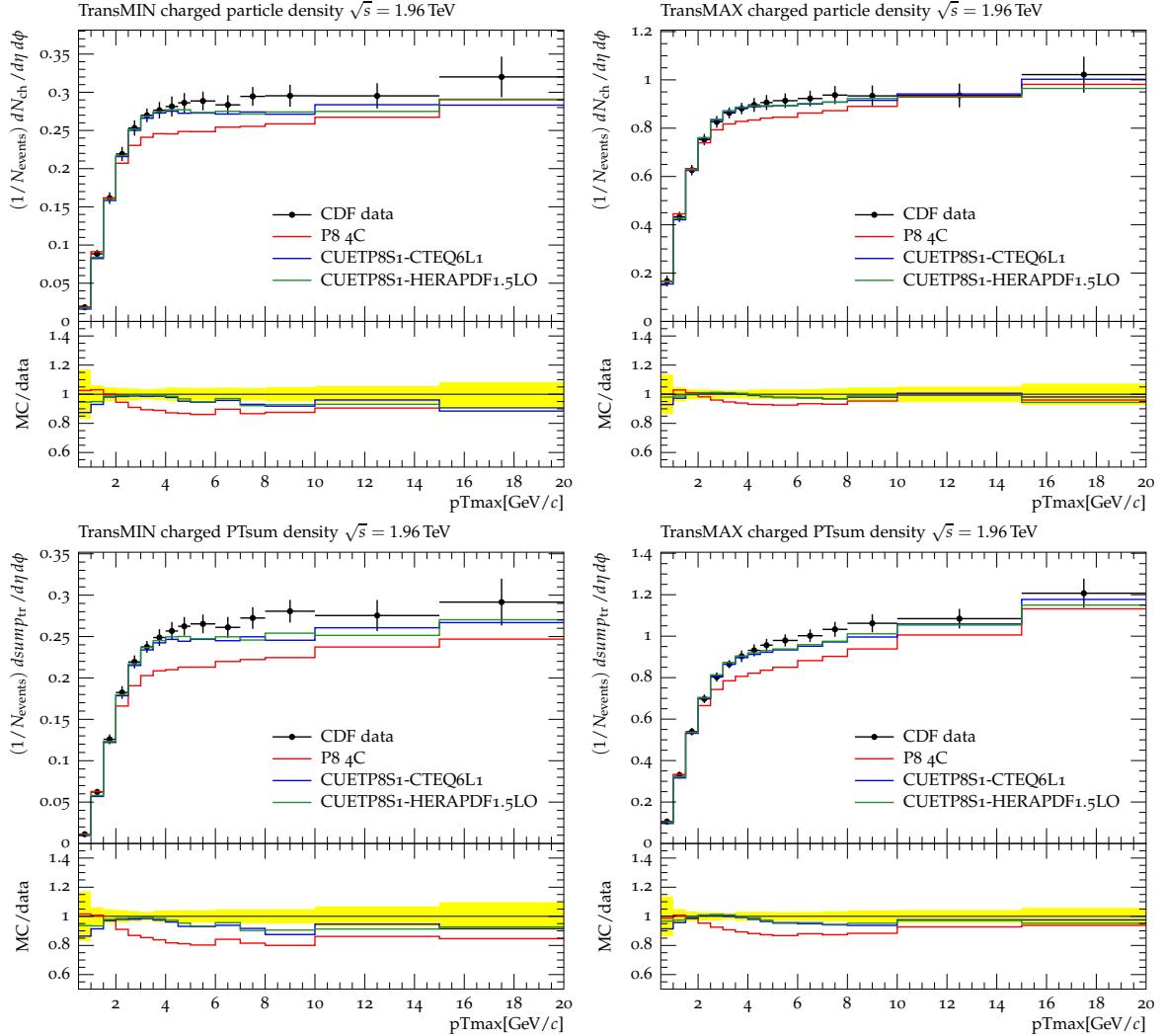


Figure 7: CDF data for $\bar{p}p$ collisions at 1.96 TeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transMIN” (left column) and “transMAX” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 8 Tune 4C and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO.

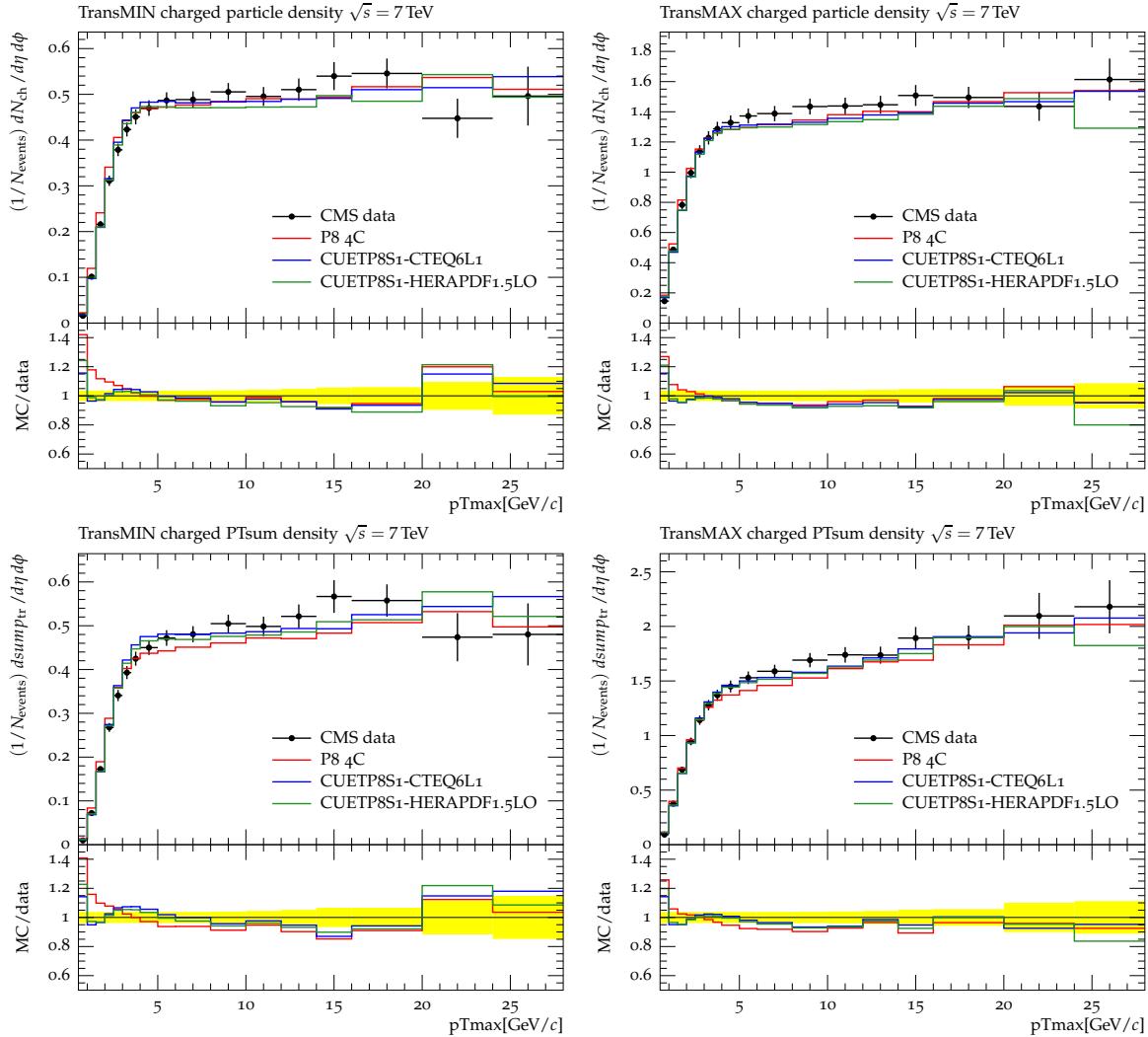


Figure 8: CMS data for pp collisions at 7 TeV: charged particle density (top row) and p_{T} sum density (bottom row) for charged particles with $p_{\text{T}} > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transMIN” (left column) and “transMAX” (right column) regions as defined by the leading charged particle, as a function of p_{T} max. The data are compared with PYTHIA 8 Tune 4C and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO.

2.3 Validation of the new CMS UE tunes

Predictions obtained with the new tunes are compared to other measurements which are not used in the tuning procedure. This helps to give a more general picture of the performance of the energy-dependent tune. The following observables are investigated:

- Forward energy flow measured in pp collisions at 7 TeV by the CMS experiment [27]; it includes the measurement of the energy flow in MB events and in hard events, where a dijet system with $p_T > 20$ GeV is required;
- Forward charged particle multiplicity as a function of pseudorapidity, $dN/d\eta$, measured in pp collisions at 7 TeV by the TOTEM experiment [28];
- Central charged particle multiplicity as a function of pseudorapidity, $dN/d\eta$, measured in pp collisions at 7 TeV by the ALICE experiment [29];
- Traditional UE measurement, measured in pp collisions at 7 TeV by the ATLAS experiment [30]; charged particle multiplicity and p_T sum are considered in the transverse, toward and away regions are considered, as a function of the leading charged particle p_T .

Results are shown for the old tunes, PYTHIA 6 Z2* and PYTHIA 8 4C tunes, and for the new UE tunes, described in this paper. Figure 9 shows the forward energy flow in two types of events, MB and dijet ones, and the charged particle multiplicity as a function of η , $dN/d\eta$, in the central and forward regions. A good agreement is achieved by the predictions of the new tunes for most of the compared measurements. The energy flow in dijet events and $dN/d\eta$ in the central region are well described by the new tunes throughout the whole spectrum. The energy flow in MB events is also well described, except in the very forward region, where a slightly lower value is predicted by the new tunes. This effect is also seen in the forward $dN/d\eta$, where a lower multiplicity is predicted by all tunes. However, the description provided by the new tunes is very good and for some observables, like $dN/d\eta$ in the central region, it is better than the one achieved by the old tune predictions. A further improvement might be fulfilled by tuning observables in the forward region.

Figure 10 shows the charged particle multiplicity and the p_T sum in the transverse, toward and away regions, as a function of the leading track p_T . The description of these measurements is very good for the new tunes. The plateau region, at $p_T > 4$ GeV is reproduced by all tunes in all regions. The low p_T region ($p_T < 4$ GeV) shows slight differences between the predictions of less than 10%. However, the agreement of predictions provided by the new tunes with the measurement, is notable.

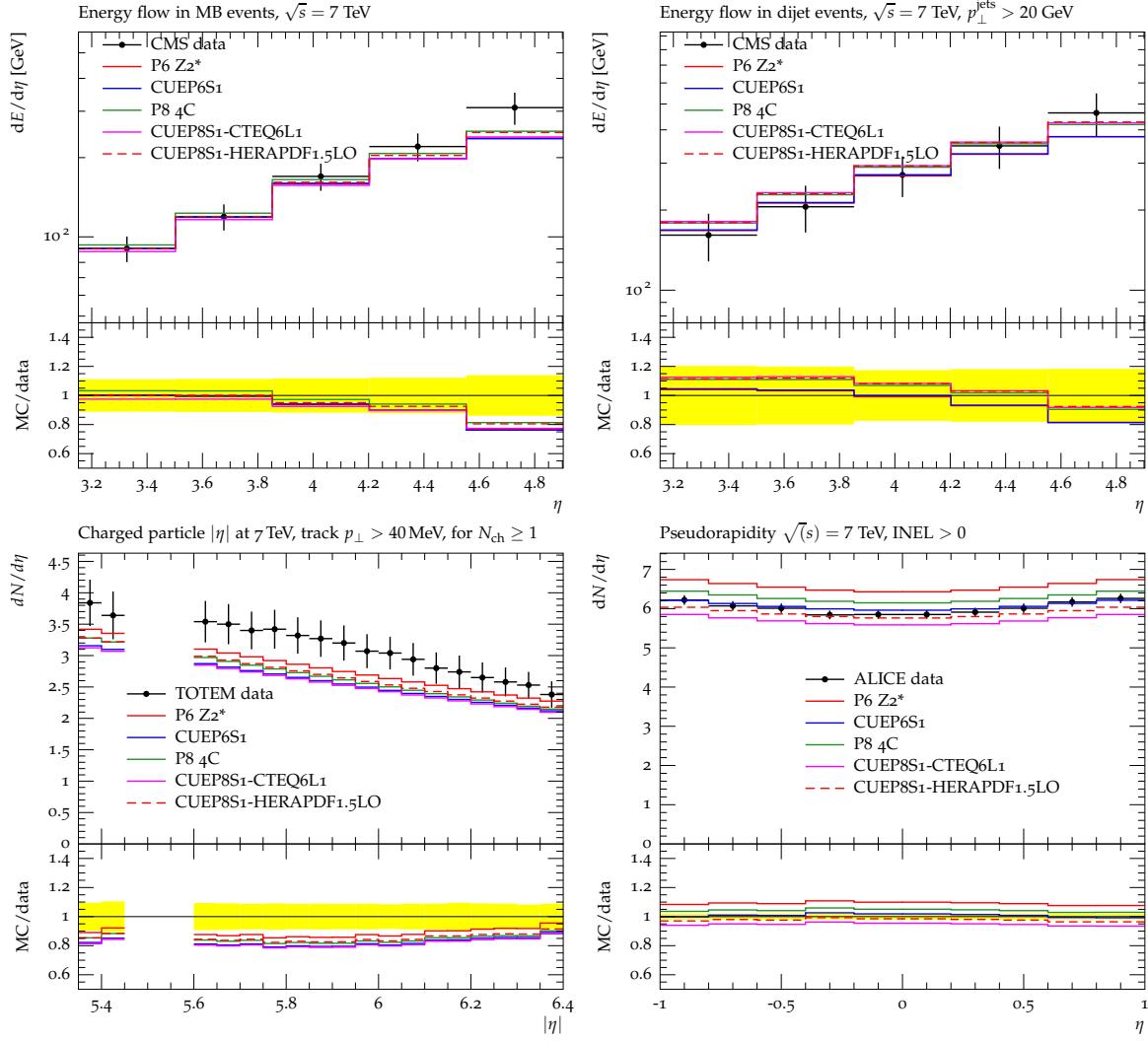


Figure 9: CMS data on the forward energy flow in MB (top left) and dijet (top right) events, ALICE data and TOTEM data on charged particle pseudorapidity in, respectively, the central and forward region. The data are compared with PYTHIA 6 tune Z₂* , PYTHIA 8 Tune 4C, the new PYTHIA 6 tune, and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO. Also shows the ratio of the tunes with the data.

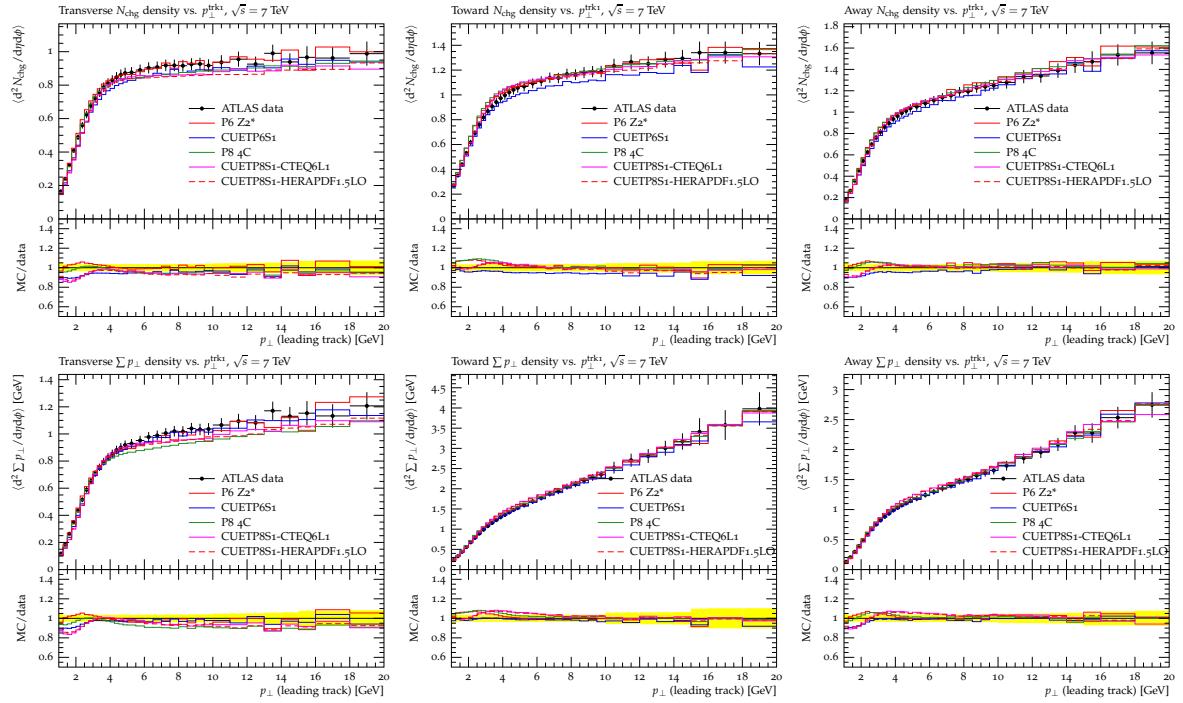


Figure 10: ATLAS data on the charged particle multiplicity N_{ch} (top) and p_T sum (bottom) measured in the transverse (left), toward (center) and away (right) regions. The data are compared with PYTHIA 6 tune $Z2^*$, PYTHIA 8 Tune 4C, the new PYTHIA 6 tune, and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO. Also shows the ratio of the tunes with the data.

2.4 Predictions for pp collisions at 13 TeV

Figure 11 shows the predictions for pp collisions at 13 TeV for PYTHIA 8 Tune 4C, PYTHIA 6 tune Z2*, the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO, and the new CMS PYTHIA 6 tune for the charged particle density and the charged p_T sum density in the “transMIN” and “transMAX” regions as defined by the leading charged particle, as a function of p_T max. The goal of this study is to make reliable predictions at 13 TeV and 14 TeV. The old tunes predict a higher contribution (20–30%) for leading $p_T < 5$ GeV, in both the TransMIN and TransMAX regions with respect to the predictions of the new tunes, which are very close to each other. At higher leading particle p_T , in the plateaux region, all tunes predict the same amount of charged particle multiplicity and p_T sum. The fact that the three new tunes predict similar results at 13 TeV gives confidence that we are able to describe the energy dependence fairly well.

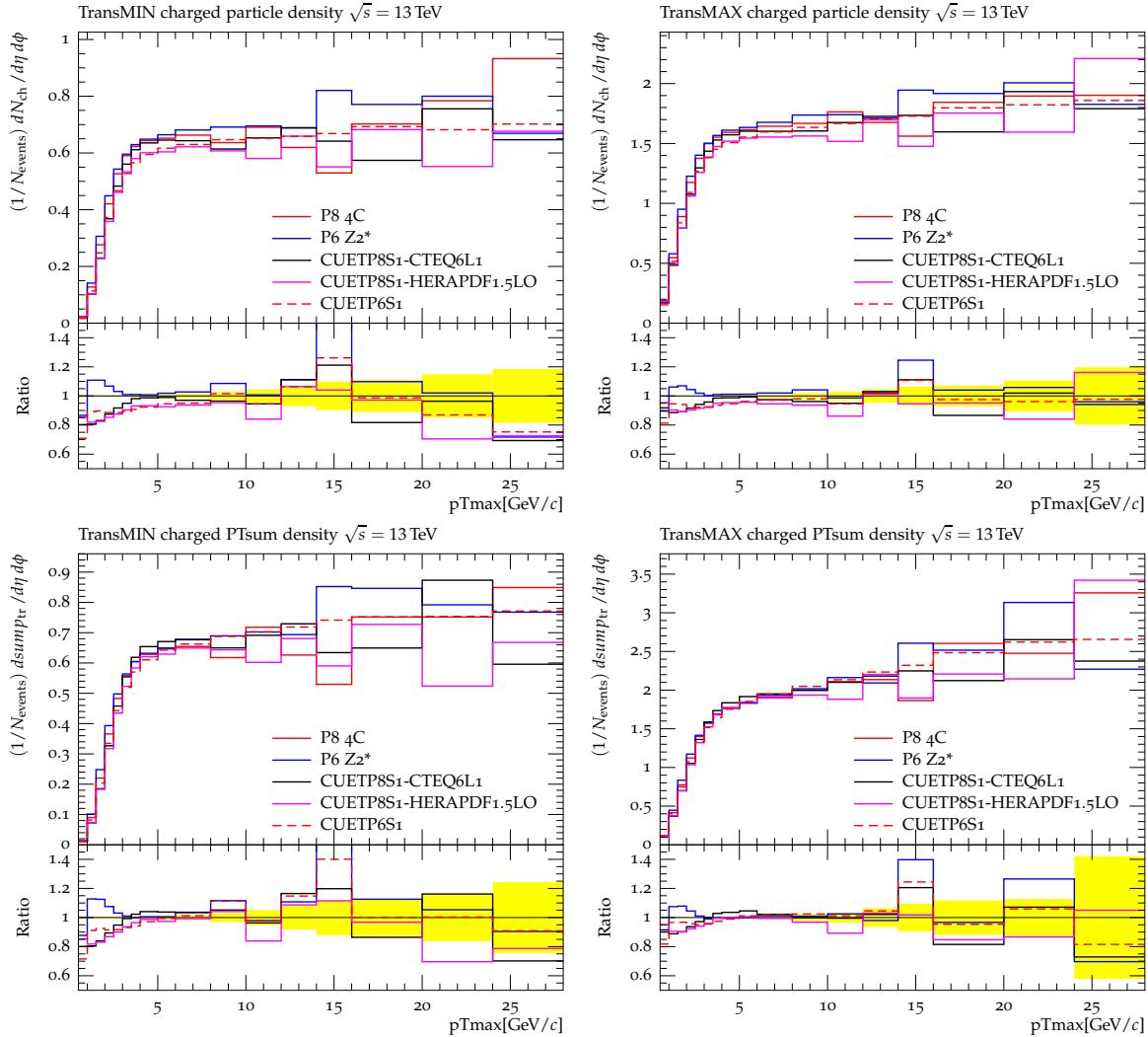


Figure 11: Predictions for pp collisions at 13 TeV for PYTHIA 8 Tune 4C, PYTHIA 6 tune $Z2^*$, the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO, and the the new CMS PYTHIA 6 tune: charged particle density (left column) and p_T sum density (right column) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transMIN” (top row), and the “transMAX” (bottom row) regions as defined by the leading charged particle, as a function of p_T max. Also shown are the ratios of the new CMS tunes to Tune 4C.

3 Tuning Double Parton Scattering Observables

Having determined the parameters of a given MPI model, one can make an unambiguous prediction of σ_{eff} in Equation 2. For example, PYTHIA 8 Tune 4C predicts a σ_{eff} of around 30.3 mb. In PYTHIA 8, σ_{eff} depends primarily on the matter overlap function, which for bProfile = 3 is determined by the exponential shape parameter, expPow, and the MPI cross section determined by pT0Ref. It also depends weakly on the BBR reconnect range, reconnectRange. These are the same parameters we varied in determining the new PYTHIA 8 UE tunes in Table 3.

The value of σ_{eff} is then predicted from the MPI parameters. We obtain the value of σ_{eff} through the PYTHIA 8 output, when two hard scatterings are forced to occur during the same pp collision, and the specific MPI parameters are set for the UE simulation. The DPS model, implemented in PYTHIA 8, assumes the independence of the hard scatterings, and follows the master formula 2. The value of σ_{eff} is calculated by PYTHIA 8, from the value of the non-diffractive cross section, fixed and equal to 50.91 mb, and an enhancement/depletion factor, which expresses the dependence of DPS events on the collision impact parameter. In fact, intuitively, more central collisions are likely to have a higher probability of a second hard scattering than peripheral ones. The enhancement/depletion factor depends only on the UE parameters, namely on the parameters which set the matter overlap function of the two protons and the MPI regulators. Note that the extraction of σ_{eff} in this way, is possible with PYTHIA 8, only when two hard scatterings are set in the generation, in order to initiate the calculation of the enhancement/depletion factor. Values of σ_{eff} , determined from fitting DPS observables, are then compared with those determined from fits to the UE observables.

In order to estimate the uncertainty of the values of σ_{eff} obtained with this method, the eigentunes provided by Professor are taken into account; they represent orthogonal variations of the parameters in the parameter space, such that the obtained results are still within the experimental uncertainties. The experimental uncertainties include both statistical and systematic ones, added in quadrature. The eigentunes are defined by values of the parameters which determine a variation of the χ^2 of the fit by one unit, with respect to the best tune. The uncertainties, associated to each variation of the tuned parameters are then summed in quadrature. Up- and down- variations in the parameter space are considered separately. Thus, asymmetric uncertainties are given for the values of σ_{eff} .

By following this procedure, the value of σ_{eff} predicted by the new two UE tunes, CUETP8S1-CTEQ6L1 and CUETP8S1-HERAPDF1.5LO, has been determined with the associated uncertainties. Table 4 shows the value of σ_{eff} for each of the two tunes, together with the value predicted by Tune 4C and the one measured by the CMS Collaboration in the $W+dijet$ channel [12]. A σ_{eff} value around 28-30 mb is obtained by all three tunes, while CMS has measured a lower value, around 21 mb. The CMS σ_{eff} measurement has been performed by defining the DPS contribution from $W+0$ jets and dijet events, generated separately, and mixing them under the same assumptions expressed by Equation 2. The discrepancy between the value of σ_{eff} measured by CMS and the one predicted by the tunes, might be an indication of a slight tension between UE tunes and DPS-based measurements, namely of troubles to describe the “softer” and the “harder” MPI components within the same framework.

A complementary approach might be to perform tunes by considering observables which are sensitive to DPS events. In the following, we investigate if the values of σ_{eff} determined by tunes based on DPS-sensitive observables, are consistent with the predictions of the UE tunes. We use the “Rivet” and “Professor” framework and vary the PYTHIA 8 UE parameters in an attempt to fit DPS observables measured by CMS in $W+dijet$ production [12] and 4-jet production [31]. The DPS observables measure correlations between the outgoing objects in hadron-hadron

Table 4: Values of σ_{eff} obtained for each PYTHIA 8 tune. The values of the old Tune 4C, and the new UE tunes, CUETP8S1-CTEQ6L1 and CUETP8S1-HERAPDF1.5LO, are compared. The uncertainties are obtained from the Professor eigentunes and express the value of σ_{eff} corresponding to a variation of the χ^2 of the fit by one unit, with respect to the best tune. The σ_{eff} value measured by CMS in the W+dijet channel is also quoted.

PYTHIA 8Tune	PYTHIA 8 σ_{eff} value (mb)
Tune 4C	30.3
CUETP8S1-CTEQ6L1	$27.8^{+1.2}_{-1.3}$
CUETP8S1-HERAPDF1.5LO	$29.1^{+2.3}_{-2.0}$
CMS W+dijet measurement [12]	$20.7 \pm 0.8 \text{ (stat)} \pm 6.6 \text{ (syst)}$

collision. Two correlation observables that are sensitive to DPS are ΔS and $\Delta^{\text{rel}} p_T$ defined as follows:

$$\Delta S = \arccos \left(\frac{\vec{p}_T(\text{object}\#1) \cdot \vec{p}_T(\text{object}\#2)}{|\vec{p}_T(\text{object}\#1)| \times |\vec{p}_T(\text{object}\#2)|} \right) \quad (3)$$

$$\Delta^{\text{rel}} p_T = \frac{|\vec{p}_T^{\text{jet}\#1} + \vec{p}_T^{\text{jet}\#2}|}{|\vec{p}_T^{\text{jet}\#1}| + |\vec{p}_T^{\text{jet}\#2}|} \quad (4)$$

where for W+dijet production object#1 is the W-boson and object#2 is the dijet. For 4-jet production object#1 is the hard-jet pair and object#2 is the soft-jet pair. For $\Delta^{\text{rel}} p_T$ in W+dijet production jet#1 and jet#2 are the two dijets, while in 4-jet production jet#1 and jet#2 are the softer two jets.

3.1 DPS in W+Dijet Production

Starting from the parameters of PYTHIA 8 Tune 4C, we construct two W+dijet DPS observables tunes. A partial tune which varies only the exponential shape parameter, `expPow`, and a full tune which varies the four parameters in Table 5. Here we interface the matrix element generated by MADGRAPH [32] with PYTHIA 8 and then tune to the normalized distributions of the correlation observables. In order to describe accurately the observables in W+dijet production, it is necessary to use the higher order matrix elements provided by MADGRAPH. Figure 12 shows the CMS data on the correlation observables ΔS and $\Delta^{\text{rel}} p_T$ measured in the W+dijet production compared with MADGRAPH interfaced with PYTHIA 8 Tune 4C, Tune 4C with no MPI, the new PYTHIA 8 partial tune and the new full tune. Table 5 gives the best fit parameters and the resulting predicted σ_{eff} values. The uncertainties quoted for the σ_{eff} values are the ones provided by the Professor machinery and they do not include any uncertainty on model dependence. Even with a smaller uncertainty, these values of σ_{eff} are compatible with the value measured by CMS.

3.2 DPS in 4-Jet Production

Starting from the parameters of PYTHIA 8 Tune 4C, we construct two 4-jet DPS observables tunes. A partial tune which varies only the exponential shape parameter, `expPow`, and a full tune which varies the four parameters in Table 6. Here we can get a good fit to the four-jet data without including the higher order matrix elements. However, we also get a good fit if

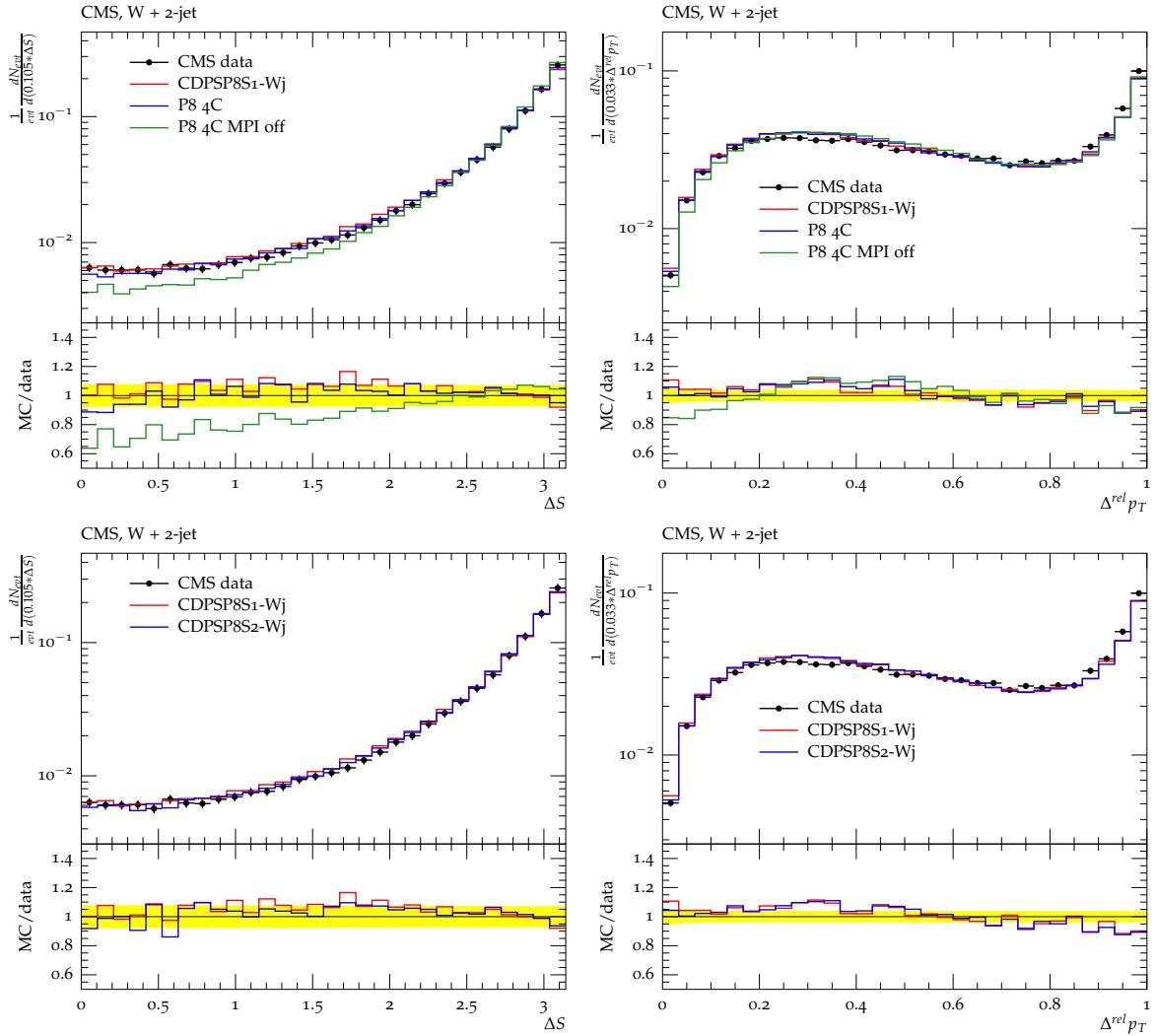


Figure 12: CMS data on the normalized distributions of the correlation observables ΔS (left column) and $\Delta^{\text{rel}} p_T$ (right column) measured in the W+dijet channel compared with MADGRAPH (MG) interfaced with PYTHIA 8 Tune 4C, Tune 4C with no MPI, and the new PYTHIA 8 partial tune (overlap only) (top row) and compared with MG interfaced with the new PYTHIA 8 partial tune (overlap only) and the new full tune (bottom row). Also shows the ratio of the tunes with the data.

Table 5: Best fit values of the parameters used to fit the DPS observables in W+dijet production. Starting from the parameters of PYTHIA 8 Tune 4C using the CTEQ6L1 PDF, the CDPSTP8S1-Wj tune varies only `expPow`, while the CDPSTP8S2-Wj tune varies all four parameters. Also shown are the predicted values σ_{eff} resulting from the tunes. The uncertainties obtained from the eigentunes given by Professor are also indicated for σ_{eff} . The parameters, accompanied by a * as superscript, are the parameters which were not tuned but left unchanged with respect to Tune 4C.

PYTHIA 8 Parameter	Tuning Range	Tune 4C	CDPSTP8S1-Wj	CDPSTP8S2-Wj
MultipleInteractions:pT0Ref (GeV)	1.0 - 3.0	2.085	2.085*	2.501
MultipleInteractions:ecmPow	0.0 - 0.4	0.19	0.19*	0.1791
MultipleInteractions:expPow	0.4 - 10.0	2.0	1.523371	1.1197
BeamRemnants:reconnectRange	0.0 - 9.0	1.5	1.5*	2.5861
Predicted σ_{eff} (mb)		30.3	$25.9^{+2.4}_{-2.9}$	$25.8^{+8.2}_{-4.2}$

we use a higher-order matrix element generated with MADGRAPH. Figure 13 shows the CMS data on the correlation observables ΔS and $\Delta^{\text{rel}} p_T$ measured in 4-jet production compared with PYTHIA 8 Tune 4C, Tune 4C with no MPI, the new PYTHIA 8 partial tune and the new full tune. The new full tune is also shown when interfaced with the matrix element generated with MADGRAPH. Table 6 gives the best fit parameters and the resulting predicted σ_{eff} values. The predicted values correspond to the first determination of σ_{eff} in 4-jet production for pp collisions at 7 TeV. Note that the uncertainties, quoted for σ_{eff} , represent how much the σ_{eff} values can change in the tuning such that the obtained predictions are still compatible within the experimental uncertainties. They do not include any model dependence because this would assume the use of a generator different from PYTHIA 8, which lie outside the purpose of this work. Hence, we do not compare previous values of σ_{eff} , measured in the four-jet channel [13–15], with the ones measured here.

Table 6: Best fit values of the parameters used to fit the DPS observables in 4-jet production. Starting from the parameters of PYTHIA 8 Tune 4C using the CTEQ6L1 PDF, the CDPSTP8S1-4j tune varies only `expPow`, while the CDPSTP8S2-4j tune varies all four parameters. Also shown are the predicted values σ_{eff} resulting from the tunes; the uncertainties obtained from the eigentunes given by Professor are also indicated for σ_{eff} . The parameters, accompanied by a * as superscript, are the parameters which were not tuned but left unchanged with respect to Tune 4C.

PYTHIA 8 Parameter	Tuning Range	Tune 4C	CDPSTP8S1-4j	CDPSTP8S2-4j
MultipleInteractions:pT0Ref (GeV)	1.0 - 3.0	2.085	2.085*	2.1254
MultipleInteractions:ecmPow	0.0 - 0.4	0.19	0.19*	0.3450
MultipleInteractions:expPow	0.4 - 10.0	2.0	1.160	0.6921
BeamRemnants:reconnectRange	0.0 - 9.0	1.5	1.5*	6.5256
Predicted σ_{eff} (mb)		30.3	$21.3^{+1.2}_{-1.6}$	$19.0^{+4.7}_{-3.0}$

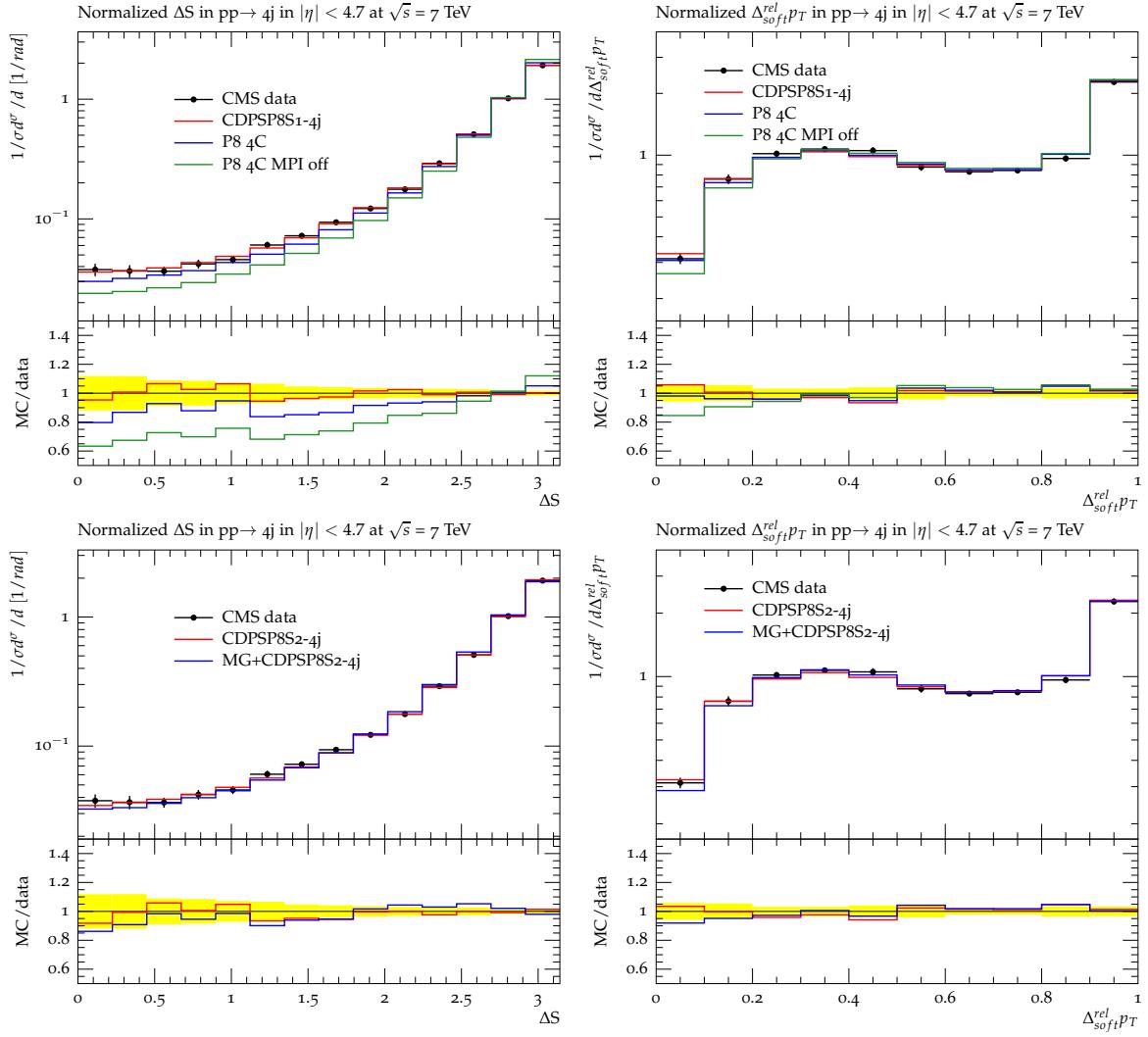


Figure 13: CMS data on the normalized distributions of the correlation observables ΔS (left) and $\Delta_{soft}^{rel} p_T$ (right) measured in 4-jet production compared with PYTHIA 8 Tune 4C, Tune 4C with no MPI, and the new PYTHIA 8 partial tune (top row) and compared with the new PYTHIA 8 tune and with MG interfaced with the new full tune (bottom row). Also shows the ratio of the tunes with the data.

3.3 Comparison of the measured values

The values of σ_{eff} predicted by the new UE and the new DPS tunes show some differences; they are shown in Table 7. While the DPS tunes prefer lower values of σ_{eff} around 20 mb, the same quantity measured in the UE tunes is around 28 mb, closer to the value predicted by Tune 4C. In order to evaluate the compatibility between the UE and DPS tunes, predictions obtained with the CDPSTP8S2-4j tune are compared to UE measurements; charged particle multiplicity and p_{T} sum are considered in the transverse, toward and away regions as a function of leading track p_{T} . Results of these comparisons are shown in Figure 14. Predictions are shown with an error band, corresponding to the total uncertainty obtained from the Professor eigentunes. While the predictions from the nominal tune undershoot the measurements in all regions, the data points are compatible for $p_{\text{T}} > 4$ GeV within the envelopes of the tune uncertainties. For lower p_{T} values, the description is not optimal, with deviations of up to 30–40%. The UE tunes, instead, offer a very good description of the whole spectrum (Figure 10). This difference, together with the discrepancies observed for the values of σ_{eff} for the different tunes, might be an indication of some tension in the simultaneous description of softer and harder MPI, within the same framework.

Table 7: Values of σ_{eff} obtained for each PYTHIA 8 tune. The values of the old Tune 4C, the new UE tunes, CUETP8S1-CTEQ6L1 and CUETP8S1-HERAPDF1.5LO, and the new DPS tunes, CUETP8S1-4j and CDPSTP8S2-4j, are compared. The uncertainties are obtained from the Professor eigentunes and express the value of σ_{eff} to a variation of the χ^2 of the fit by one unit, with respect to the best tune.

PYTHIA 8Tune	PYTHIA 8 σ_{eff} value (mb)
Tune 4C	30.3
CUETP8S1-CTEQ6L1	$27.8^{+1.2}_{-1.3}$
CUETP8S1-HERAPDF1.5LO	$29.1^{+2.3}_{-2.0}$
CDPSTP8S1-4j	$21.3^{+1.2}_{-1.6}$
CDPSTP8S2-4j	$19.0^{+4.7}_{-3.0}$

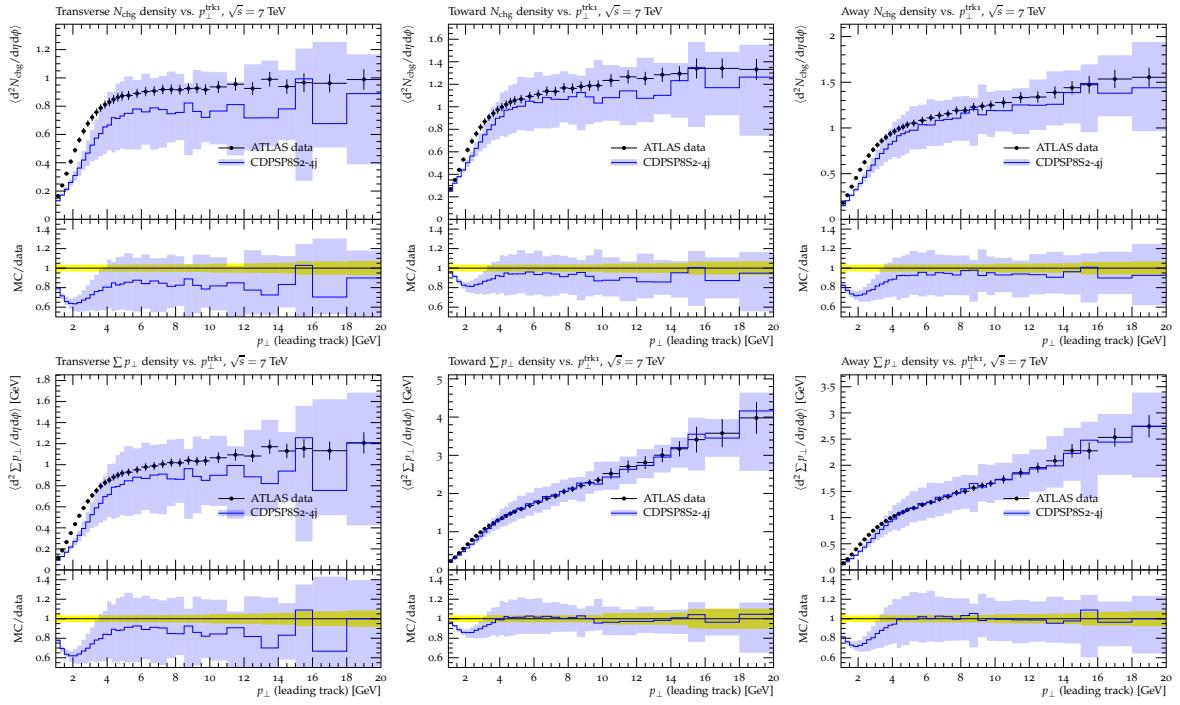


Figure 14: ATLAS data on the charged particle multiplicity N_{ch} (top) and p_T sum (bottom) measured in the transverse (left), toward (center) and away (right) regions compared with CDPSTS2-4j. Also shows the ratio of the tunes with the data and the uncertainties of the predictions based on the Professor eigentunes.

4 Summary and Conclusions

Using the ‘‘Rivet’’ and ‘‘Professor’’ framework, we have constructed a new PYTHIA 6 CTEQ6L1 tune and two new PYTHIA 8 UE tunes (one using CTEQ6L1 and one using HERAPDF1.5LO). By simultaneously fitting more than two center-of-mass energies, one tests the validity of the model of the energy dependence of MPI in Equation 1. We see for the first time that this formalism is describing the energy dependence of the UE very well. Also by fitting several energies one can better constrain the parameters, allowing for more precise predictions at 13 TeV and 14 TeV. The ‘‘transMIN’’ density (sensitive to MPI & BBR) increases much faster with center-of-mass energy than does the ‘‘transDIF’’ density (sensitive to ISR & FSR). The MPI increases like a power of the center-of-mass energy, while the ISR & FSR increase logarithmically. This is the first time we have seen the different energy dependences of these two components. Previously, we only had information on the energy dependence of the ‘‘transAVE’’ density and the new CMS tunes describe quite well the energy dependence of the various components of the UE. Although the data were not used in the fit, the new tunes also fit the CMS charged-particle jet UE data [22] and the ATLAS p_T max UE data [30] very well at 900 GeV and 7 TeV.

Starting with PYTHIA 8 Tune 4C, we have constructed tunes using CMS data on correlation observables in W+dijet [12] and 4-jet [31] production that are supposed to be sensitive to DPS events. The parameter obtained with the new DPS-based tunes are then used to extract a measurement of σ_{eff} . The values of σ_{eff} determined from W+dijet production are compatible with the previous CMS measurements using a different simulation of the DPS contribution. The values determined from 4-jet production correspond to the first measurements of σ_{eff} in this channel in pp collisions at 7 TeV. The uncertainties, quoted for σ_{eff} , represent how much the σ_{eff} values can change in the tuning such that the obtained predictions are still compatible within

the experimental errors and they do not include any model dependence.

The UE tunes attempt to describe “soft” or “semi-hard” MPI scatterings that accompany a hard scattering process, while DPS tunes attempt to describe two hard scatterings within the same hadron-hadron collision. They are clearly related, since DPS occurs when one of the MPI becomes hard. More work needs to be done, but for the first time an attempt to fit and describe “softer” and “harder” MPI within the same framework is tried. We see that the UE and DPS tunes seem to be consistent with each other only in some regions of the UE spectrum, while differences are measured between UE and DPS tunes, in the predicted values of σ_{eff} .

References

- [1] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 physics and manual”, *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
- [2] T. Sjöstrand, L. Lonnblad, and S. Mrenna, “PYTHIA 6.2: Physics and manual”, arXiv:hep-ph/0108264.
- [3] T. Sjöstrand and M. van Zijl, “A Multiple Interaction Model for the Event Structure in Hadron Collisions”, *Phys.Rev.* **D36** (1987) 2019, doi:10.1103/PhysRevD.36.2019.
- [4] M. Bengtsson, T. Sjöstrand, and M. van Zijl, “Initial State Radiation Effects on W and Jet Production”, *Z.Phys.* **C32** (1986) 67, doi:10.1007/BF01441353.
- [5] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1”, *Comput.Phys.Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.
- [6] TeV4LHC QCD Working Group Collaboration, “Tevatron-for-LHC Report of the QCD Working Group”, arXiv:hep-ph/0610012.
- [7] CDF Collaboration, “Studying the ‘underlying event’ at CDF and the LHC”, (2009). FERMILAB-CONF-09-792-E.
- [8] P. Z. Skands, “The Perugia Tunes”, arXiv:0905.3418.
- [9] R. Field, “The underlying event in hadronic collisions”, *Ann.Rev.Nucl.Part.Sci.* **62** (2012) 453–483, doi:10.1146/annurev-nucl-102711-095030.
- [10] M. H. Seymour and A. Siodmok, “Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data”, *JHEP* **1310** (2013) 113, doi:10.1007/JHEP10(2013)113, arXiv:1307.5015.
- [11] ATLAS Collaboration, “Measurement of hard double-parton interactions in $W(\rightarrow l\nu)+2$ jet events at $\sqrt{s}=7$ TeV with the ATLAS detector”, *New J.Phys.* **15** (2013) 033038, doi:10.1088/1367-2630/15/3/033038, arXiv:1301.6872.
- [12] CMS Collaboration, “Study of double parton scattering using W + 2-jet events in proton-proton collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **1403** (2014) 032, doi:10.1007/JHEP03(2014)032, arXiv:1312.5729.
- [13] UA2 Collaboration, “A Study of multi - jet events at the CERN anti-p p collider and a search for double parton scattering”, *Phys.Lett.* **B268** (1991) 145–154, doi:10.1016/0370-2693(91)90937-L.

- [14] AFS Collaboration, “Double parton scattering in pp collisions at $\sqrt{s} = 63$ GeV”, *Z. Phys.* **C34** (1987).
- [15] CDF Collaboration, “Double parton scattering in anti p-p collisions at $\sqrt{s} = 1.8$ TeV”, *Z. Phys.* **D56** (1997).
- [16] R.Field, “The Tevatron Energy Scan: Findings & Surprises”, (September 17,2013). published in the proceedings of the International Symposium on Multiparticle Dynamics.
- [17] CMS Collaboration, “Measurement of the Underlying Event Activity at the LHC at 7 TeV and Comparison with 0.9 TeV”, (2012). CMS-PAS-FSQ-12-020.
- [18] A. Buckley et al., “Rivet user manual”, *Comput.Phys.Commun.* **184** (2013) 2803–2819, doi:10.1016/j.cpc.2013.05.021, arXiv:1003.0694.
- [19] A. Buckley et al., “Systematic event generator tuning for the LHC”, *Eur.Phys.J.* **C65** (2010) 331–357, doi:10.1140/epjc/s10052-009-1196-7, arXiv:0907.2973.
- [20] J. Pumplin et al., “New generation of parton distributions with uncertainties from global QCD analysis”, *JHEP* **0207** (2002) 012, doi:10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195.
- [21] H1 and ZEUS Collaborations, “HERAPDF1.5LO PDF Set with Experimental Uncertainties”, (October,2013). H1prelim-13-141, ZEUS-prel-13-003.
- [22] CMS Collaboration, “Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV”, *JHEP* **1109** (2011) 109, doi:10.1007/JHEP09(2011)109, arXiv:1107.0330.
- [23] CMS Collaboration, “Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s} = 0.9, 2.76$, and 7 TeV”, *JHEP* **1304** (2013) 072, doi:10.1007/JHEP04(2013)072, arXiv:1302.2394.
- [24] CMS Collaboration, “Measurement of the underlying event in the Drell-Yan process in proton-proton collisions at $\sqrt{s} = 7$ TeV”, *Eur.Phys.J.* **C72** (2012) 2080, doi:10.1140/epjc/s10052-012-2080-4, arXiv:1204.1411.
- [25] J. Pumplin, “Hard underlying event correction to inclusive jet cross-sections”, *Phys.Rev.* **D57** (1998) 5787–5792, doi:10.1103/PhysRevD.57.5787, arXiv:hep-ph/9708464.
- [26] R. Corke and T. Sjöstrand, “Interleaved parton showers and tuning prospects”, *JHEP* **03** (2011) 032, doi:10.1007/JHEP03(2011)032, arXiv:1011.1759.
- [27] CMS Collaboration, “Measurement of energy flow at large pseudorapidities in pp collisions at $\sqrt{s} = 0.9$ and 7 TeV”, *JHEP* **1111** (2011) 148, doi:10.1007/JHEP11(2011)148, 10.1007/JHEP02(2012)055, arXiv:1110.0211.
- [28] G. Antchev et al., “First measurement of the total proton-proton cross section at the LHC energy of $\sqrt{s} = 7$ TeV”, *Europhys.Lett.* **96** (2011) 21002, doi:10.1209/0295-5075/96/21002, arXiv:1110.1395.
- [29] ALICE Collaboration, “Charged-particle multiplicity measurement in proton-proton collisions at $\sqrt{s} = 7$ TeV with ALICE at LHC”, *Eur.Phys.J.* **C68** (2010) 345–354, doi:10.1140/epjc/s10052-010-1350-2, arXiv:1004.3514.

-
- [30] ATLAS Collaboration, “Measurement of underlying event characteristics using charged particles in pp collisions at $\sqrt{s} = 900\text{GeV}$ and 7 TeV with the ATLAS detector”, *Phys.Rev.* **D83** (2011) 112001, doi:10.1103/PhysRevD.83.112001, arXiv:1012.0791.
- [31] CMS Collaboration, “Measurement of four-jet production in proton-proton collisions at $\sqrt{s}=7\text{ TeV}$ ”, arXiv:1312.6440. CMS-FSQ-12-013.
- [32] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **1106** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.

A Further comparisons of UE data and predictions of old and new tunes

In this appendix, comparisons between predictions of the new and old tunes and data in the “transDIF” and “transAVE” regions are shown. These are the data which have not been used for the tuning. Figures 15-18 compare the CDF data at 300 GeV, 900 GeV and 1.96 TeV and the CMS data at 7 TeV on the charged particle and p_T sum densities in the “transDIF” and “transAVE” regions as defined by the leading charged particle, as a function of p_T max with PYTHIA 6 tune Z2*, Tune Z2*lep and with the new CMS PYTHIA 6 tune. Figures 19-21 compare the CDF data at 900 GeV and 1.96 TeV and the CMS data at 7 TeV on the charged particle and p_T sum densities in the “transDIF” and “transAVE” regions as defined by the leading charged particle, as a function of p_T max with PYTHIA 8 Tune 4C and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO. Same conclusions can be drawn for the predictions of the new PYTHIA 6 and PYTHIA 8 tunes, as observed for the comparisons with data measured in the “transMIN” and “transMAX” regions. All of the new tunes are able to improve the description of the data over the whole phase space and at all energies, with respect to the level of agreement achieved by predictions of the old tunes.

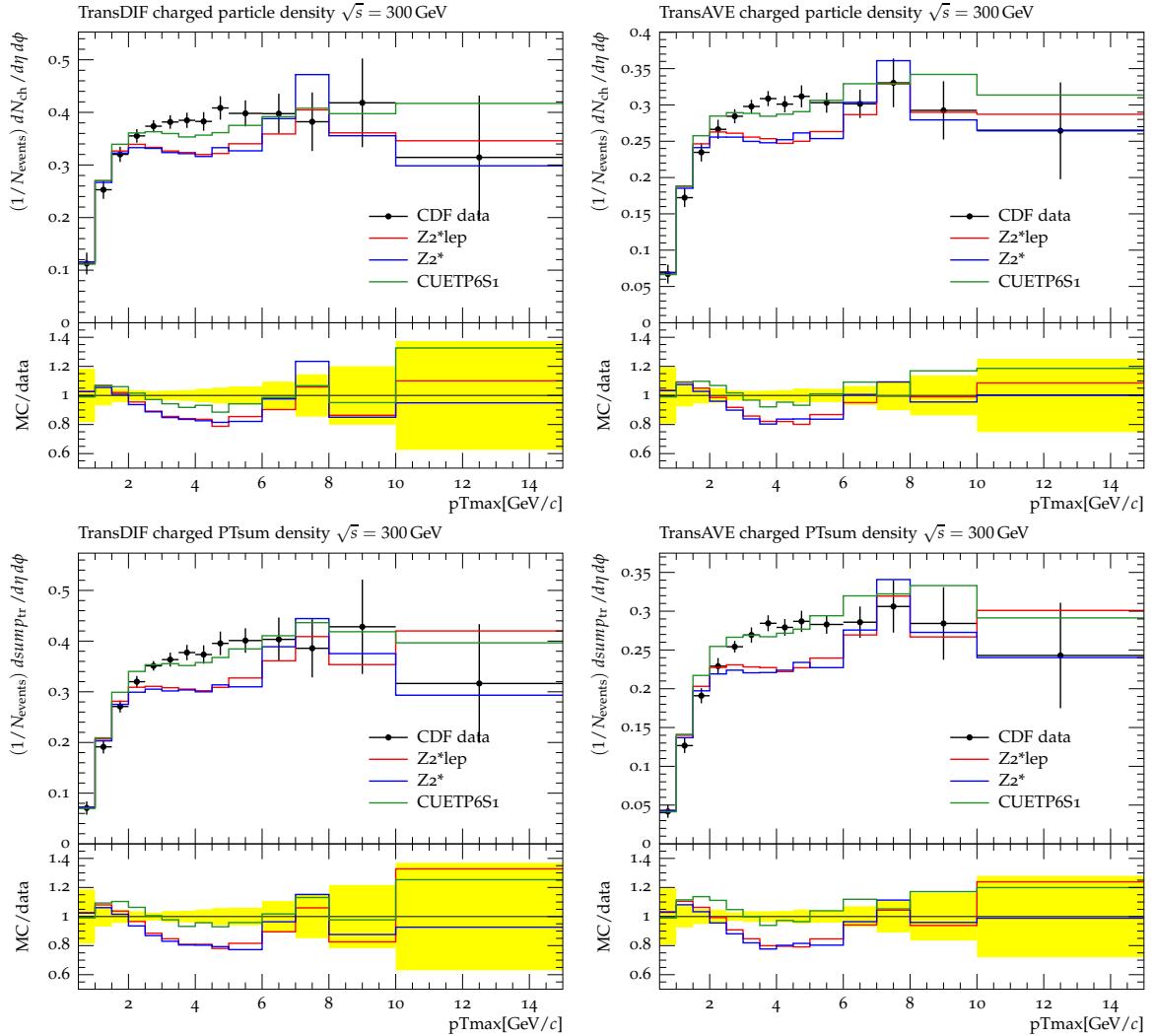


Figure 15: CDF data for $\bar{p}p$ collisions at 300 GeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transDIF” (left column) and “transAVE” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 6 tune Z2*, Tune Z2*lep and the new CMS PYTHIA 6 tune.

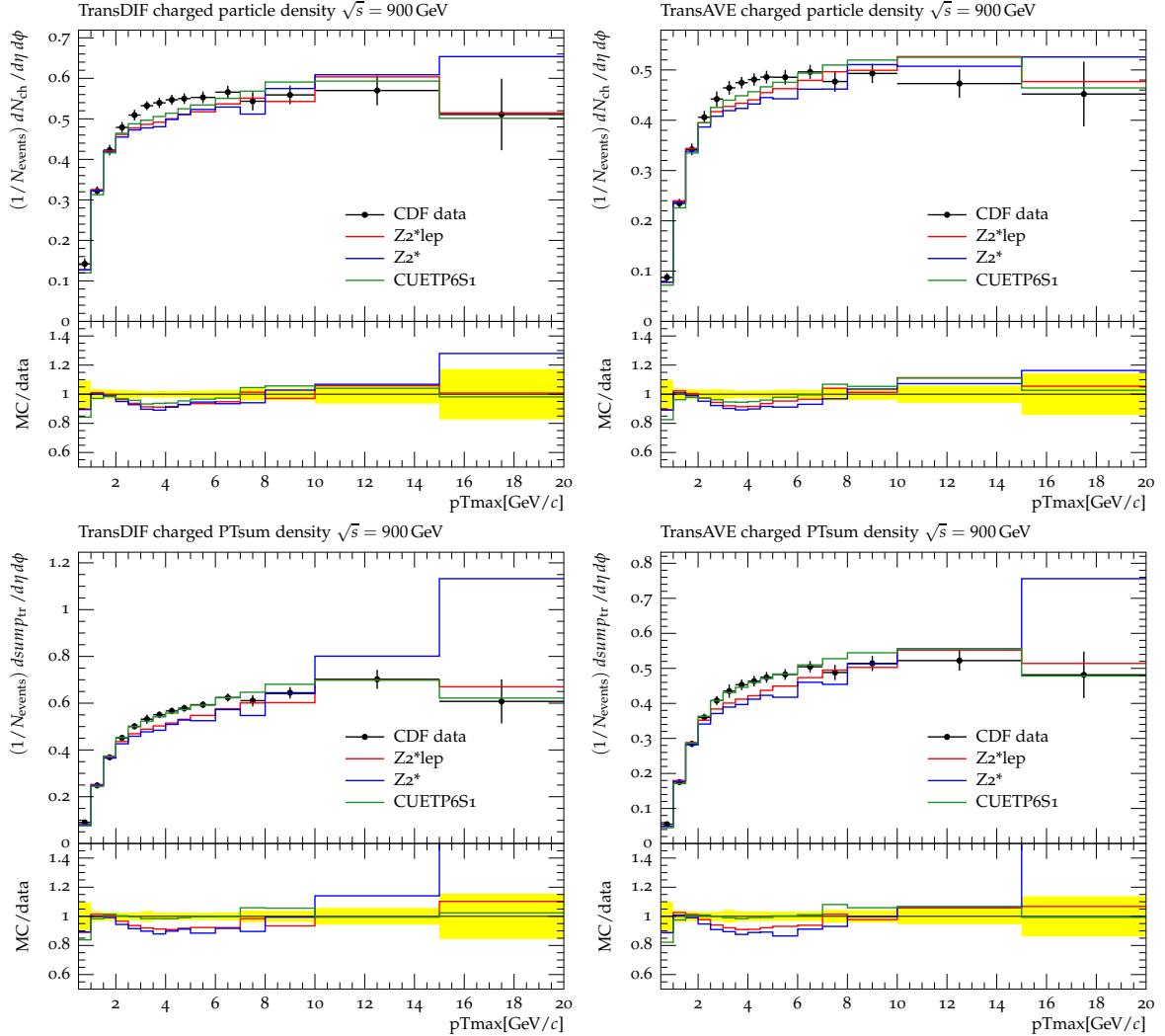


Figure 16: CDF data for $\bar{p}p$ collisions at 900 GeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transDIF” (left column) and “transAVE” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 6 tune Z2*, Tune Z2*lep and the new CMS PYTHIA 6 tune.

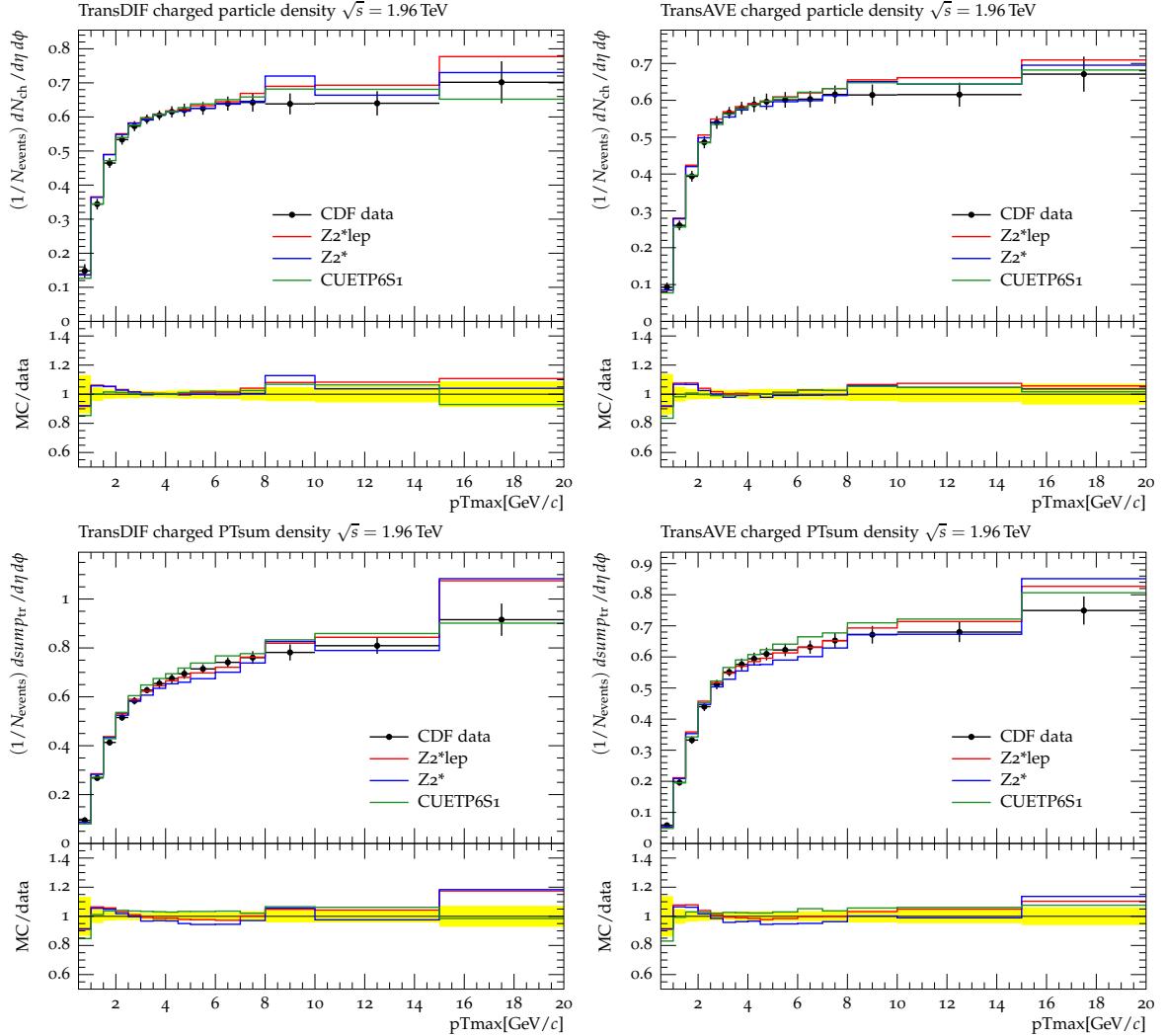


Figure 17: CDF data for $\bar{p}p$ collisions at 1.96 TeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transDIF” (left column) and “transAVE” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 6 tune Z2*, Tune Z2*lep and the new CMS PYTHIA 6 tune.

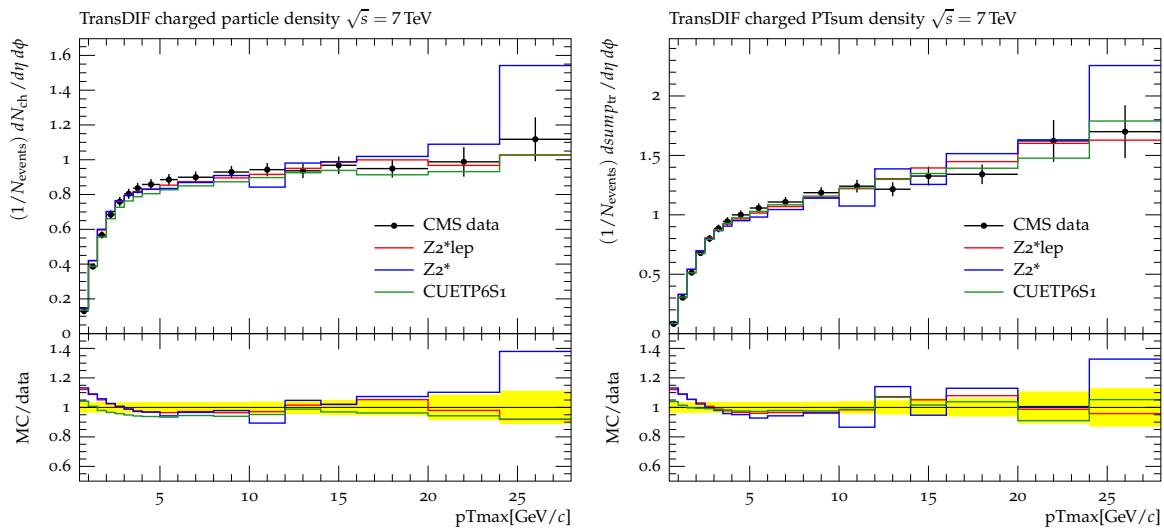


Figure 18: CMS data for pp collisions at 7 TeV: charged particle density (left) and p_T sum density (right) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transDIF” region as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 6 tune $Z2^*$, Tune $Z2^*\text{lep}$ and the new CMS PYTHIA 6 tune.

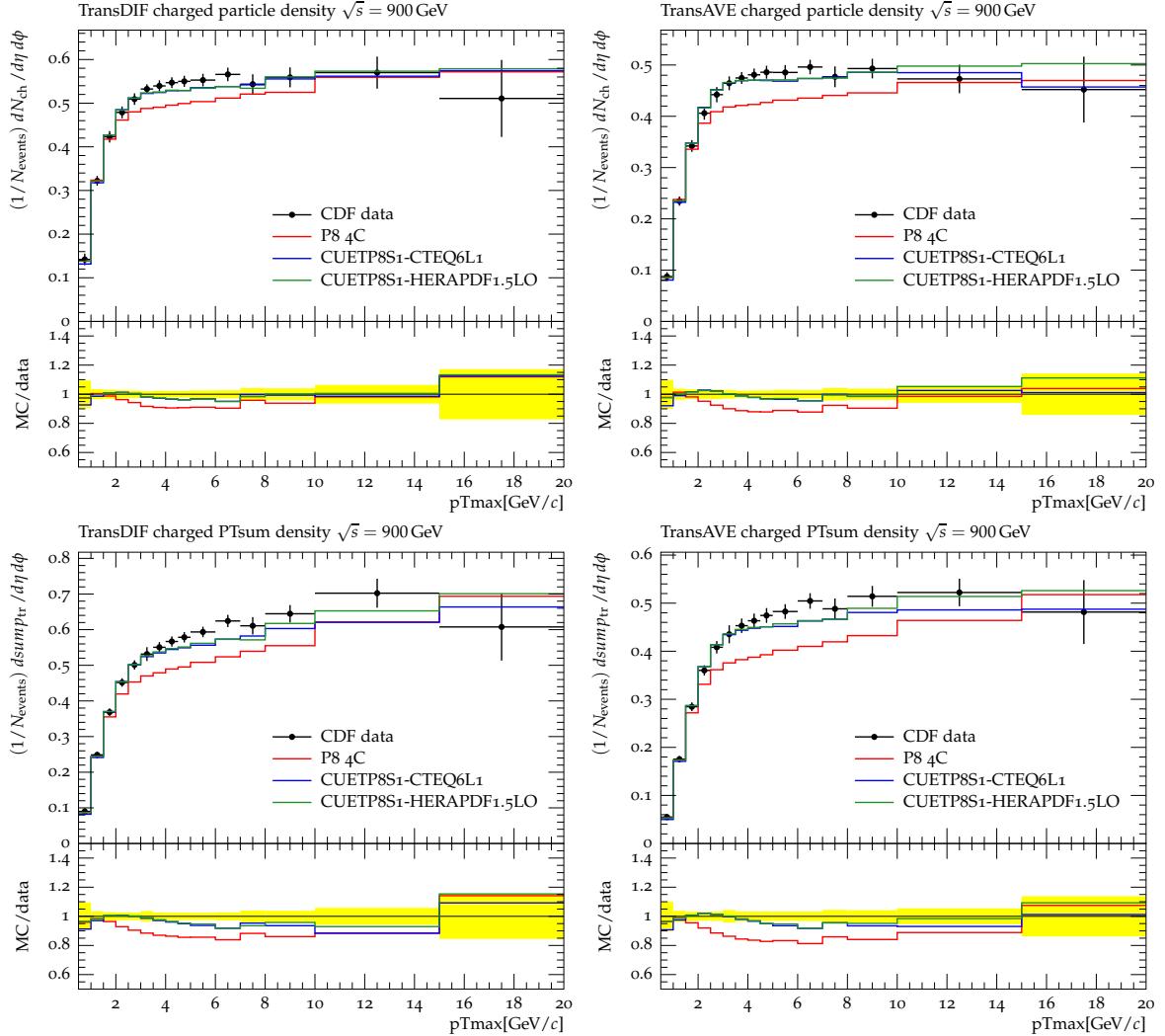


Figure 19: CDF data for $\bar{p}p$ collisions at 900 GeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transDIF” (left column) and “transAVE” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 8 Tune 4C and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO.

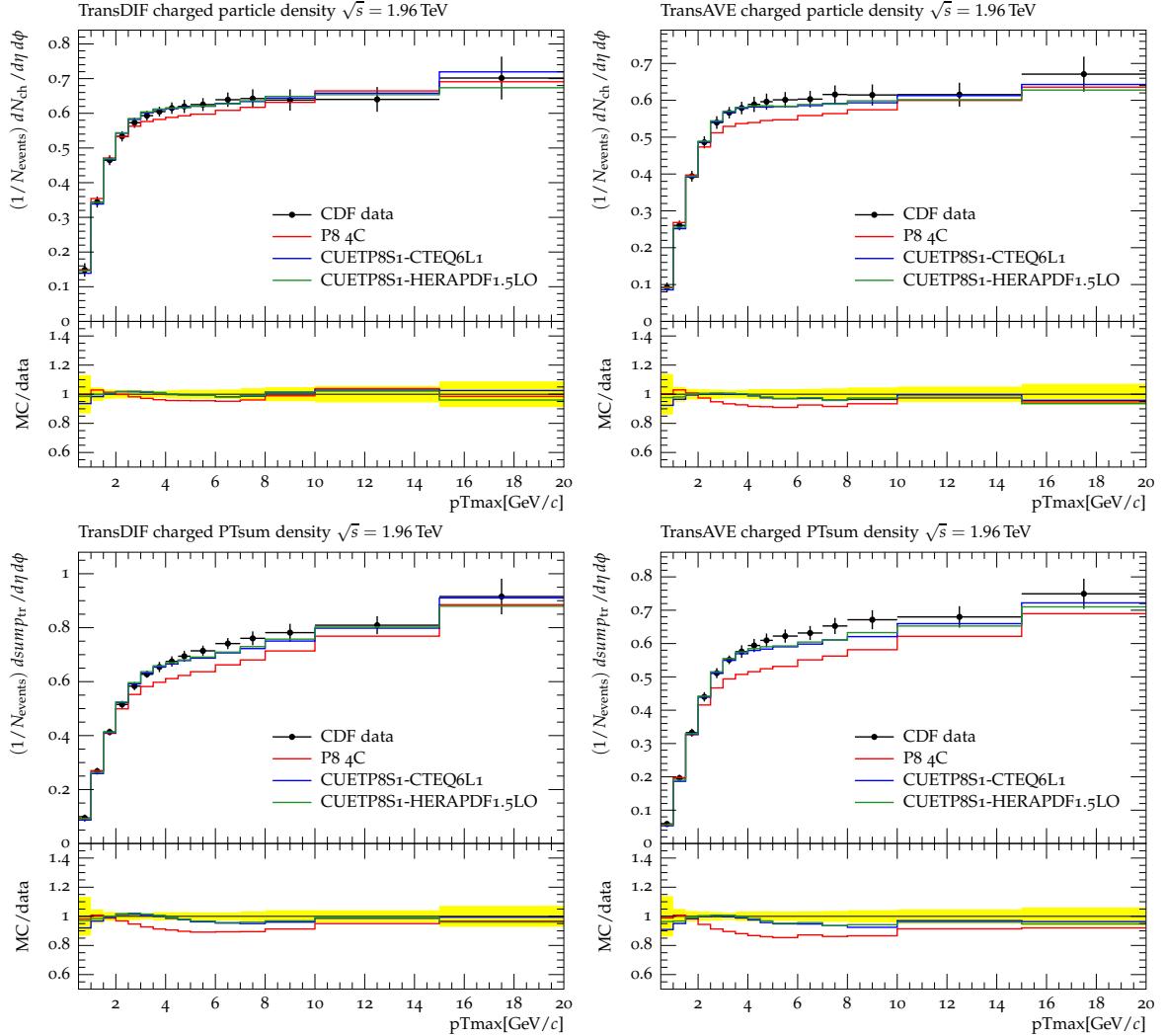


Figure 20: CDF data for $\bar{p}p$ collisions at 1.96 TeV: charged particle density (top row) and p_T sum density (bottom row) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transDIF” (left column) and “transAVE” (right column) regions as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 8 Tune 4C and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO.

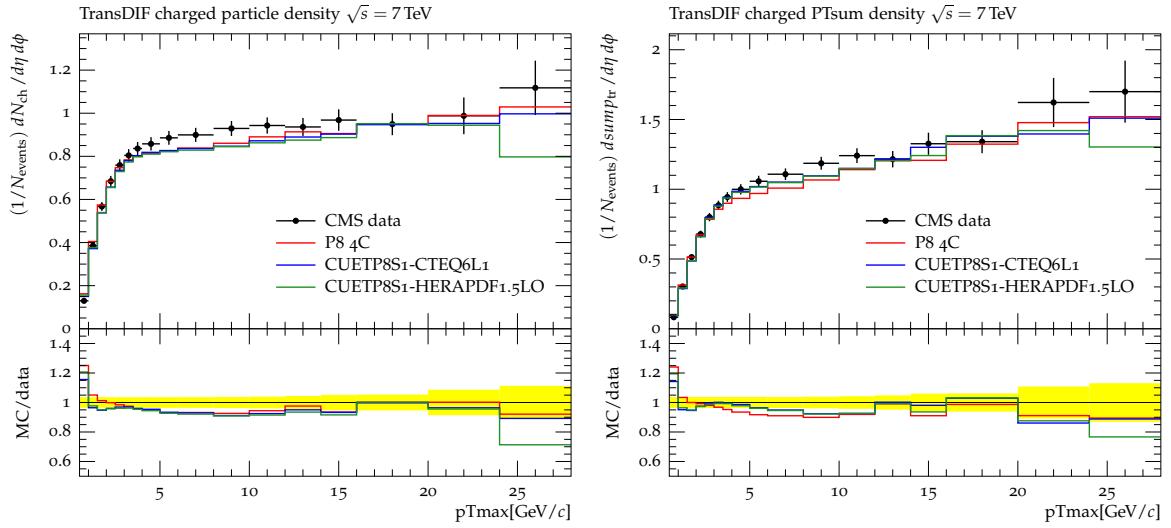


Figure 21: CMS data for pp collisions at 7 TeV: charged particle density (left) and p_T sum density (right) for charged particles with $p_T > 0.5 \text{ GeV}/c$ and $|\eta| < 0.8$ in the “transDIF” region as defined by the leading charged particle, as a function of p_T max. The data are compared with PYTHIA 8 Tune 4C and the two new CMS PYTHIA 8 tunes using CTEQ6L1 and the HERAPDF1.5LO.

Appendix F

Longitudinal momentum shifts, showering, and nonperturbative corrections in matched next-to-leading-order shower event generators

Phys. Rev. D87 (2013) 094009

Longitudinal momentum shifts, showering, and nonperturbative corrections in matched next-to-leading-order shower event generators

S. Dooling,¹ P. Gunnellini,¹ F. Hautmann,^{2,3} and H. Jung^{1,4}

¹*Deutsches Elektronen Synchrotron, D-22603 Hamburg, Germany*

²*Theoretical Physics Department, University of Oxford, Oxford OX1 3NP, United Kingdom*

³*Physics & Astronomy, University of Sussex, Brighton BN1 9QH, United Kingdom*

⁴*Elementaire Deeltjes Fysica, Universiteit Antwerpen, B 2020 Antwerpen, Belgium*

(Received 27 December 2012; published 10 May 2013)

Comparisons of experimental data with theoretical predictions for collider processes containing hadronic jets rely on shower Monte Carlo event generators to include corrections to perturbative calculations from hadronization, parton showering, and multiple parton collisions. We examine current treatments of these corrections and propose alternative methods to take into account nonperturbative effects and parton showering in the context of next-to-leading-order event generators. We point out sizable parton-showering corrections to jet transverse energy spectra at high rapidity and discuss kinematic shifts in longitudinal momentum distributions from initial state showering in the case both of jet production and of heavy mass production at the Large Hadron Collider.

DOI: [10.1103/PhysRevD.87.094009](https://doi.org/10.1103/PhysRevD.87.094009)

PACS numbers: 12.38.-t

I. INTRODUCTION

Phenomenological analyses of collider processes involving the production of hadronic jets rely on event simulation by parton shower Monte Carlo generators [1,2]. The subject of this paper concerns two different, common uses of shower Monte Carlo generators: one in which they are combined with hard scattering matrix elements via a matching scheme, e.g., at the next-to-leading order (NLO) [3,4] in perturbative QCD, and another in which they are used to obtain corrections to perturbative calculations due to hadronization, showering, and multiple parton interactions (see e.g., [5,6]), with such correction factors then being applied to determine realistic predictions, which can be compared with experimental data.

We begin in Sec. II by considering methods to evaluate the nonperturbative (NP) corrections to jet cross sections using shower event generators. We also estimate the corrections that arise from the initial state and final state parton showers and observe that they are sizeable (beyond NLO) in jet transverse energy spectra over the full range of rapidity. We propose a decomposition of the corrections to be applied to fixed NLO calculations, consisting of a truly NP contribution supplemented with a contribution coming from all order resummation via parton showers.

Next, in Sec. III we investigate kinematic aspects of parton showers associated with combining the approximation of collinear, on-shell partons with energy-momentum conservation. The main effect is an event-by-event shift in longitudinal momentum distributions whose size depends on the observable and on the phase space region, and increases with increasing rapidities. We illustrate this by numerical Monte Carlo results in different phase space regions for four specific examples of jet, heavy-quark, electroweak gauge boson, and Higgs boson production.

First results on kinematic shifts have been presented in [7].

The approach of this work may be helpful to analyze corrections to finite-order perturbative calculations for jet observables from parton-showering and nonperturbative dynamics. These encompass both final state fragmentation effects and initial state contributions associated with collinearity approximations. Dynamical high-energy effects on jet final states, distinct from the ones discussed in this paper, have been emphasized in [8–10] due to noncollinear contributions to parton branching processes. We note that both these results and the results in this paper stress the phenomenological relevance of more complete descriptions of QCD parton cascades in terms of transverse momentum dependent parton fragmentation and parton density functions [11–14]. Concluding comments on the results of this work are given in Sec. IV.

II. MONTE CARLO NONPERTURBATIVE CORRECTION FACTORS

In this section we consider methods to evaluate NP and parton shower correction factors. To be definite, we refer to the case of inclusive production of single jets at the LHC [15]. In order to compare theory with experimental data corrected to stable particle level, Refs. [5,6] supplement NLO perturbative calculations with NP corrections estimated from Monte Carlo event generators. Using leading-order Monte Carlo (LO-MC) generators [1,2], the correction factors K_0 are schematically obtained by [5,6]

$$K_0^{\text{NP}} = N_{\text{LO-MC}}^{(\text{ps+mpi+had})} / N_{\text{LO-MC}}^{(\text{ps})}, \quad (1)$$

where (ps + mpi + had) and (ps) mean, respectively, a simulation including parton showers, multiparton interactions, and hadronization, and a simulation including

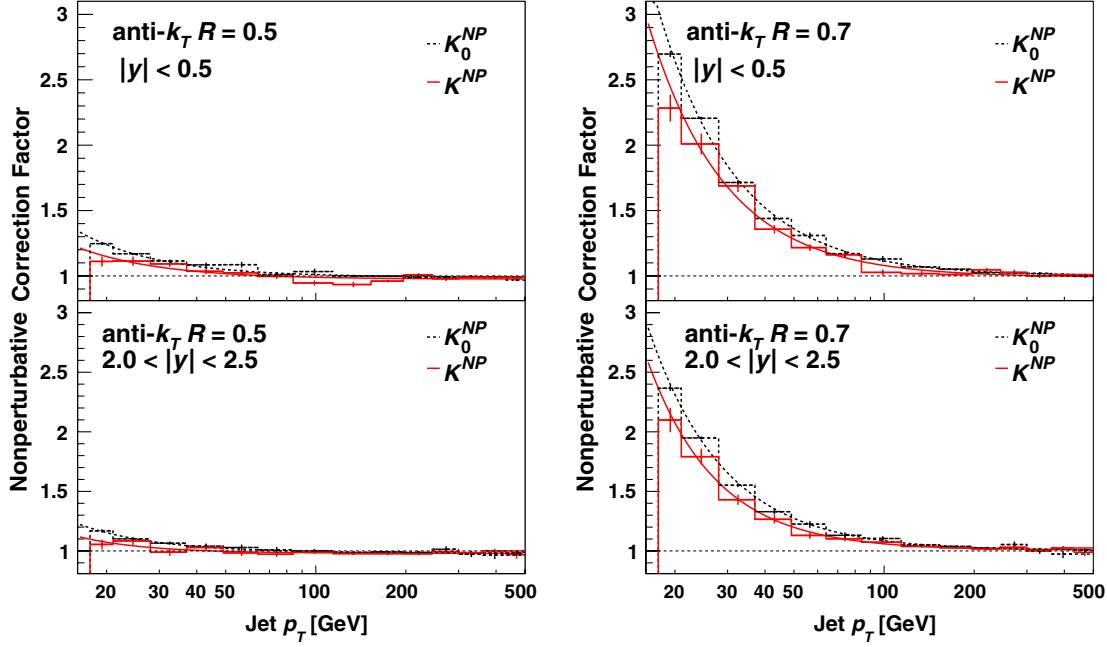


FIG. 1 (color online). The NP correction factors to jet transverse momentum distributions obtained from Eqs. (1) and (2), using PYTHIA and POWHEG respectively, for $|y| < 0.5$ and $2 < |y| < 2.5$. Left: $R = 0.5$. Right: $R = 0.7$.

only parton showers in addition to the LO hard process. Having only LO + PS event generators available, this is the most obvious way to estimate NP corrections to be applied to NLO parton-level calculations. However, when these corrections are combined with NLO parton-level results, a potential inconsistency arises because the radiative correction from the first gluon emission is treated at

different levels of accuracy in the two parts of the calculation.

We here suggest that an alternative method that avoids this is to use NLO Monte Carlo (NLO-MC) generators to determine the correction. In this case one can consistently assign correction factors to be applied to NLO calculations. Moreover, this method allows one to study separately

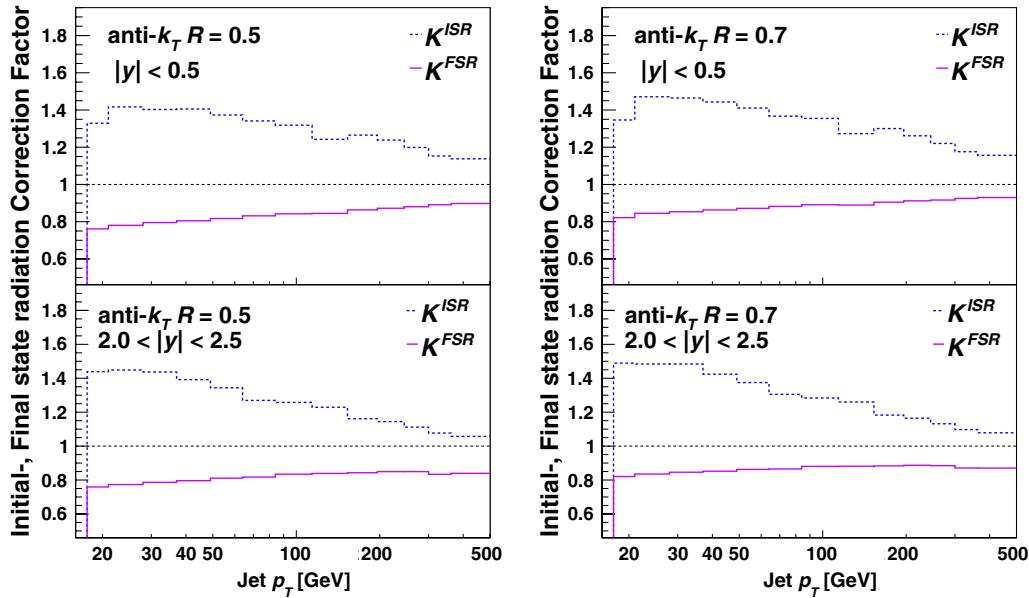


FIG. 2 (color online). The initial and final state parton shower correction factor to jet transverse momentum distributions, obtained from Eq. (3) using POWHEG for $|y| < 0.5$ and $2 < |y| < 2.5$. Left: $R = 0.5$. Right: $R = 0.7$.

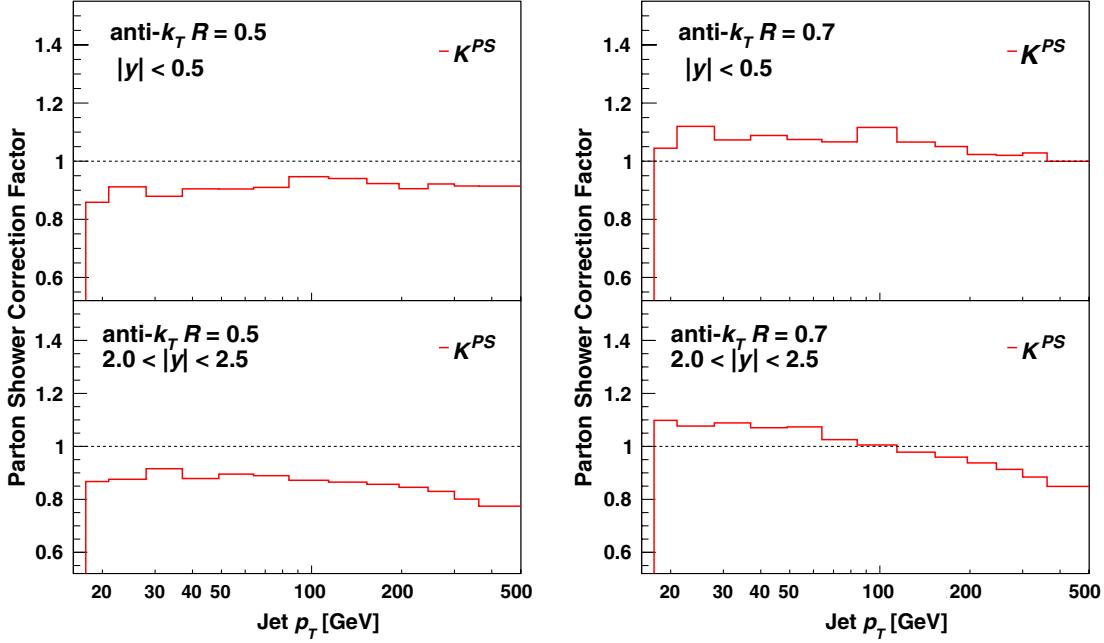


FIG. 3 (color online). The parton shower correction factor to jet transverse momentum distributions, obtained from Eq. (3) using POWHEG for $|y| < 0.5$ and $2 < |y| < 2.5$. Left: $R = 0.5$. Right: $R = 0.7$.

correction factors to the fixed-order calculation due to parton-showering effects. To this end, we introduce the correction factors K^{NP} and K^{PS} as

$$K^{\text{NP}} = N_{\text{NLO-MC}}^{(\text{ps+mpi+had})} / N_{\text{NLO-MC}}^{(\text{ps})}, \quad (2)$$

$$K^{\text{PS}} = N_{\text{NLO-MC}}^{(\text{ps})} / N_{\text{NLO-MC}}^{(0)}, \quad (3)$$

where the denominator in Eq. (3) is defined by switching off all components beyond NLO in the Monte Carlo simulation. The difference between the correction factors in Eqs. (1) and (2) comes primarily from the way in which the multiple parton interaction (MPI) contribution is matched to the NLO calculation. MPI processes have typical transverse momentum scales smaller than the scale of the hard process, which may be defined as the average transverse momentum of the hard partons. This, however, is different in LO and NLO calculations, giving rise to non-negligible numerical differences, which we will show below. The correction factor in Eq. (3), on the other hand, is new. It singles out contributions due to parton showering. This correction factor has not been considered in earlier analyses. We show below its numerical significance. We anticipate that taking properly into account these showering corrections can be relevant in fits for parton distribution functions (pdfs) using inclusive jet data.

In Fig. 1 we compute results for the NP correction factors in Eqs. (1) and (2) to jet transverse momentum distributions. We define jets using the anti- k_T algorithm [16] with jet size $R = 0.5$ and $R = 0.7$. We plot the results versus the jet transverse momentum p_T for different

regions in the jet rapidity y . We show K^{NP} as obtained using the NLO event generator POWHEG [17] and compare it to the result obtained at leading order from PYTHIA [2] (tune Z2 [18] and CTEQ6L1 pdfs [19]). The curves in Fig. 1 illustrate the differences coming from the definition of the hard process.

In Figs. 2 and 3 we compute the corrections from parton shower K^{PS} as obtained from Eq. (3) as a function of the jet p_T for different values of R and different rapidities y . Figure 2 shows the contributions coming from initial state and final state parton showers separately. We note that the initial and final state showers are so interconnected that the combined effect is nontrivial and cannot be obtained by simply adding the two results. In general the effect from parton shower is largest at large $|y|$, where the initial state

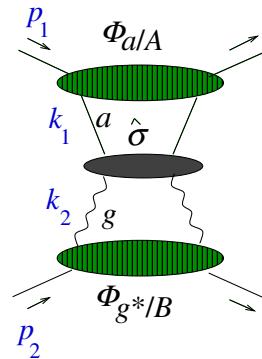


FIG. 4 (color online). Factorized structure of the jet cross section at high rapidity.

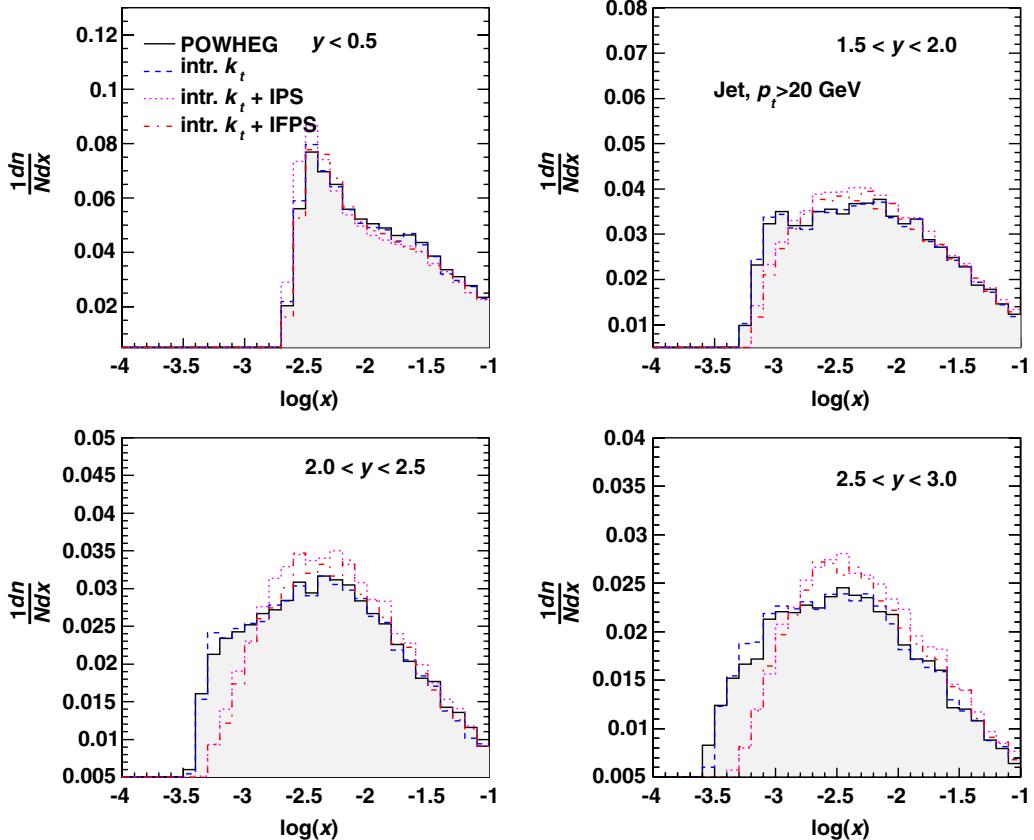


FIG. 5 (color online). Distributions in the parton longitudinal momentum fraction x before (POWHEG) and after parton showering (POWHEG + PS), for inclusive jet production at different rapidities for jets with $p_T > 18$ GeV obtained by the anti- k_t jet algorithm [16] with $R = 0.5$. Shown is the effect of intrinsic k_t , initial (IPS) and initial+final state (IFPS) parton shower.

parton shower is mainly contributing at low p_T , while the final state parton shower is contributing significantly over the whole p_T range. In particular, note in Fig. 3 that, while at central rapidity the combined shower correction is rather flat in p_T , at higher rapidity this is no longer flat and for large p_T it may even dip below the correction from the purely final state shower reported in Fig. 2. This suggests that migration effects become relevant not only in p_T but also in y .

While the NP corrections studied in Fig. 1 become vanishingly small at sufficiently large p_T , the showering correction in Figs. 2 and 3 gives finite effects also for large p_T . Since, as shown by our results, the size of this effect does depend on the value of rapidity y , this will influence the shape of jet distributions and the comparisons of theory predictions with experimental data. In particular, if the showering correction factor is not consistently taken into account, besides the NP corrections, this may affect the determination of parton distribution functions from data sets including jets.

Note that in [5,6] NP correction factors K_0 are applied to the NLO calculation [20], and the data comparison shows that the NLO calculation agrees with data at central rapidities, while increasing deviations are seen with increasing

rapidity at large transverse momentum p_T [5]. A second comparison is performed in [5] with NLO-matched POWHEG calculations [17], showing large differences in the high rapidity region between results obtained by interfacing POWHEG with different shower models [1,2] and different model tunes [18,21].¹ Motivated by this observation, in the next section we consider more closely the kinematics of the initial state parton shower at high rapidity.

III. INITIAL STATE SHOWERING AND KINEMATIC SHIFTS

Let us recall the physical picture [10] of jet production at high rapidity (Fig. 4) based on QCD high-energy factorization [23]. Take the incoming momenta p_1 and p_2 in Fig. 4 in the plus and minus lightcone directions, defined, for any four-vector v^μ , as $v^\pm = (v^0 \pm v^3)/\sqrt{2}$. Let us parametrize the exchanged momenta k_1 and k_2 in terms of purely transverse four-vectors $k_{\perp 1}$ and $k_{\perp 2}$ and longitudinal (light cone) momentum fractions x_i (collinear) and \bar{x}_i

¹Further discussion of parton showering effects on high-rapidity jets may be found in [22].

(anticollinear) as $k_1 = x_1 p_1 + k_{\perp 1} + \bar{x}_1 p_2$ and $k_2 = x_2 p_2 + k_{\perp 2} + \bar{x}_2 p_1$. To single-logarithmic accuracy in the jet rapidity and the jet transverse momentum, we may approximate k_1 and k_2 using strong ordering in the longitudinal momenta and get [10]

$$k_1 \approx x_1 p_1, \quad k_2 \approx x_2 p_2 + k_{\perp 2}, \quad x_1 \gg x_2. \quad (4)$$

The physical picture corresponding to the factorization [10,23] consists of the scattering of a highly off-shell, low- x parton off a nearly on-shell, high- x parton. The calculations [10,22] embody this picture through the longitudinal and transverse momentum dependences of both perturbative and nonperturbative components of the jet cross section, denoted, respectively, by $\hat{\sigma}$ and Φ in Fig. 4. In what follows, however, we will not use the specific content of these calculations, but we will simply use the underlying physical picture as a guidance to examine kinematic effects of collinear approximations.

In the light of this picture, let us consider the NLO-matched shower Monte Carlo calculations, following [7]. In the Monte Carlo event generator first the hard subprocess events with full four-momentum assignments for the external lines are generated. In particular, the momenta $k_j^{(0)}$

($j = 1, 2$) of the partons initiating the hard scatter are on shell, and are taken to be fully collinear with the incoming state momenta p_j ,

$$k_j^{(0)} = x_j p_j \quad (j = 1, 2). \quad (5)$$

Next the showering algorithm is applied, and complete final states are generated including additional QCD radiation from the initial state and final state parton cascades. As a result of QCD showering, the momenta k_j are no longer exactly collinear,

$$k_j \neq x_j p_j \quad (j = 1, 2). \quad (6)$$

Their transverse momentum is to be compensated by a change in the kinematics of the hard scattering subprocess. By energy-momentum conservation, however, this implies a reshuffling, event by event, in the longitudinal momentum fractions x_j of the partons scattering off each other in the hard subprocess. The size of the shift in x_j depends on the emitted transverse momenta.

Let us now focus on jets measured in the rapidity range $y < 2.5$ [6] and examine the effect of the kinematical shift in the longitudinal momentum fractions. To this end we

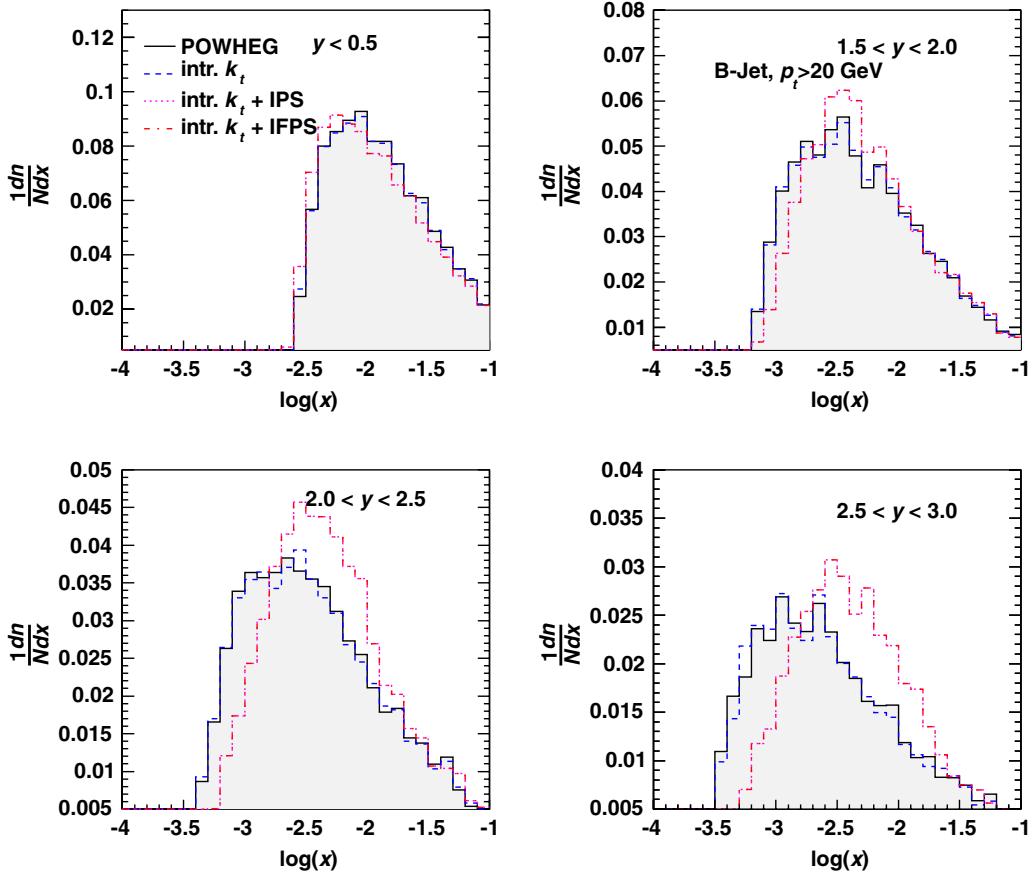


FIG. 6 (color online). Production of b -jets: distribution in the parton longitudinal momentum fraction x , before and after parton showering, for different rapidity regions. Shown is the effect of intrinsic k_t , IPS and IFPS parton shower.

compute the distribution in x_j from POWHEG before parton showering and after parton showering [7]. Figure 5 shows the distribution for one of the x_j partons. We plot the result before showering (POWHEG) and the results of successively including intrinsic k_t , initial state parton shower, and initial + final state parton showers. The results are obtained using the PYTHIA parton shower (tune Z2 [18] and CTEQ6L1 pdfs [19]). This does not include multiple parton interaction and hadronization effects. Using the definition of light cone momentum fractions given at the beginning of this section, the kinematic variable x is computed as $x = (E + p^z)/(2E_{\text{beam}})$, where E and p^z are the energy and z component of momentum of parton j , and E_{beam} is the energy of the hadron beam. The momentum fraction x is first calculated for the partons given by POWHEG before shower and then calculated from the PYTHIA event record after shower.

We see from Fig. 5 that the kinematical reshuffling in the longitudinal momentum fraction is negligible for central rapidities but becomes significant for $y > 1.5$. This effect characterizes the highly asymmetric parton kinematics, which becomes important for the first time at the LHC in significant regions of phase space [10]. Since the perturbative weight for each event is determined by the initial POWHEG simulation, predictions of matched NLO-shower calculations for observables sensitive to this asymmetric

region can be affected significantly by the kinematical shift as shown in Fig. 5. Similarly, since the momentum reshuffling is done after the evaluation of the parton distribution functions, the kinematical shift can affect predictions also through the pdfs. It will be of interest to examine the impact of this phase space region on total cross sections as well.

Let us next consider the case of bottom-flavor jet production [24,25]. The LHC measurements [24,25] are reasonably described by NLO-matched shower generators MC@NLO [26] and POWHEG [27] at central rapidities, and they are below these predictions at large rapidity and large p_T . In Fig. 6 we consider B -jets in different rapidity regions [24] and plot the gluon x distribution from POWHEG before parton showering and after including various components of the parton shower generator, similarly to what is done above for Fig. 5. We use the PYTHIA parton shower (tune Z2 [18], here including hadronization to identify the B -jet). We observe a similar shift in longitudinal momentum with increasing rapidity as in the inclusive jet case.

In Fig. 7 we consider Drell-Yan (DY) production in the mass range $16 < m_{\text{DY}} < 166$ GeV and perform a similar study to what is done above for jets. In this case too we find that the effects of the kinematical reshuffling in x evaluated from POWHEG become non-negligible away from the

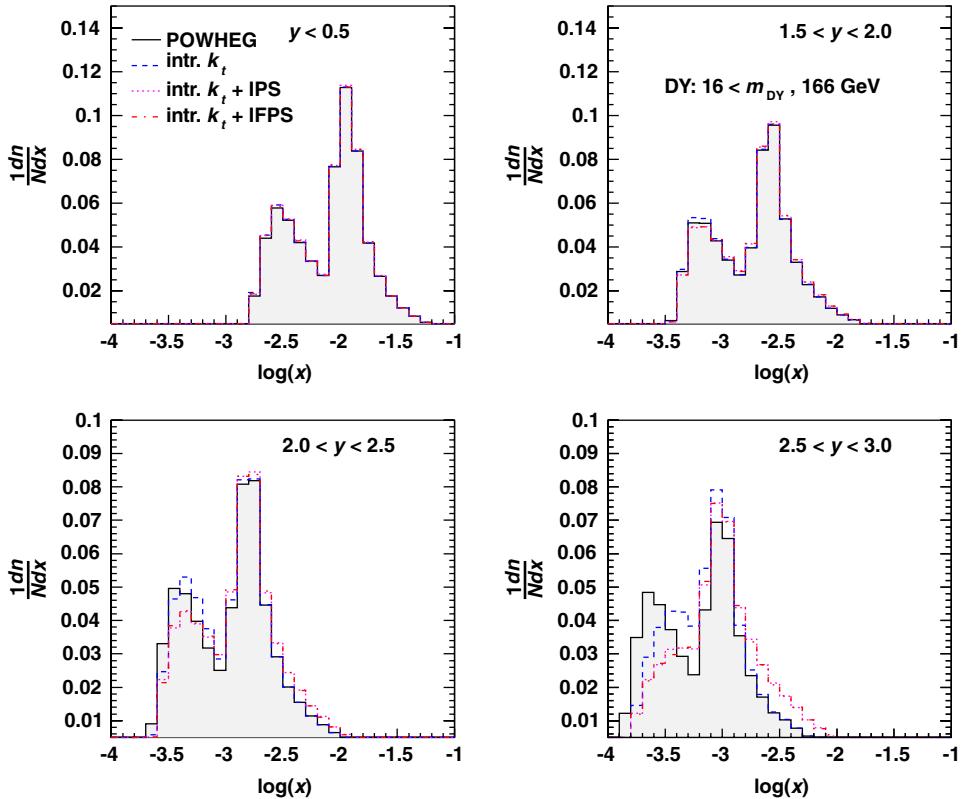


FIG. 7 (color online). Drell-Yan production with $16 < m_{\text{DY}} < 166$ GeV: distribution in the parton longitudinal momentum fraction x before and after showering. Shown is the effect of intrinsic k_t , IPS and IFPS parton shower.

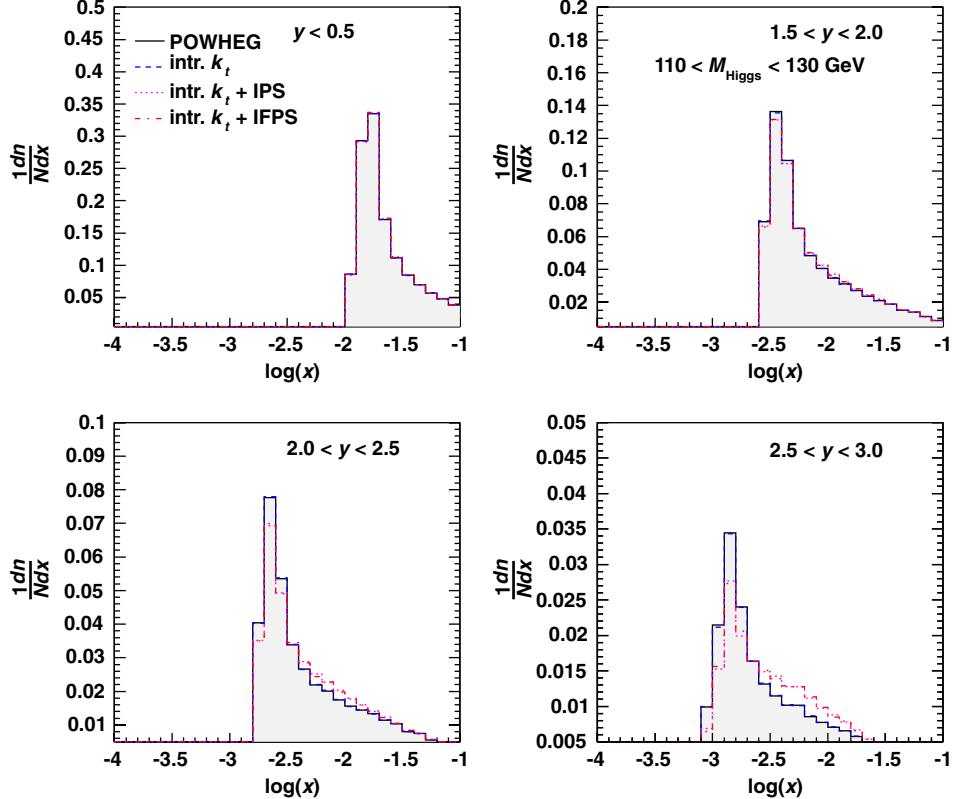


FIG. 8 (color online). Higgs boson production with $110 < m_{\text{Higgs}} < 130$ GeV: distribution in the parton longitudinal momentum fraction x before and after showering. Shown is the effect of intrinsic k_t , IPS and IFPS parton shower.

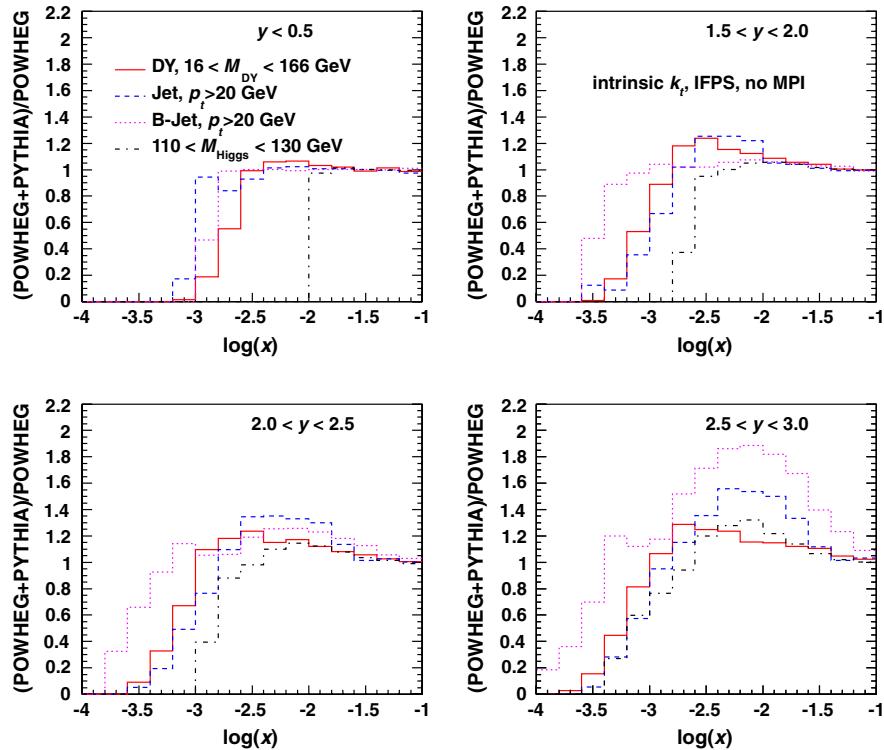


FIG. 9 (color online). Ratio of the cross sections obtained with POWHEG after and before inclusion of initial + final state parton shower and intrinsic k_t for the different processes.

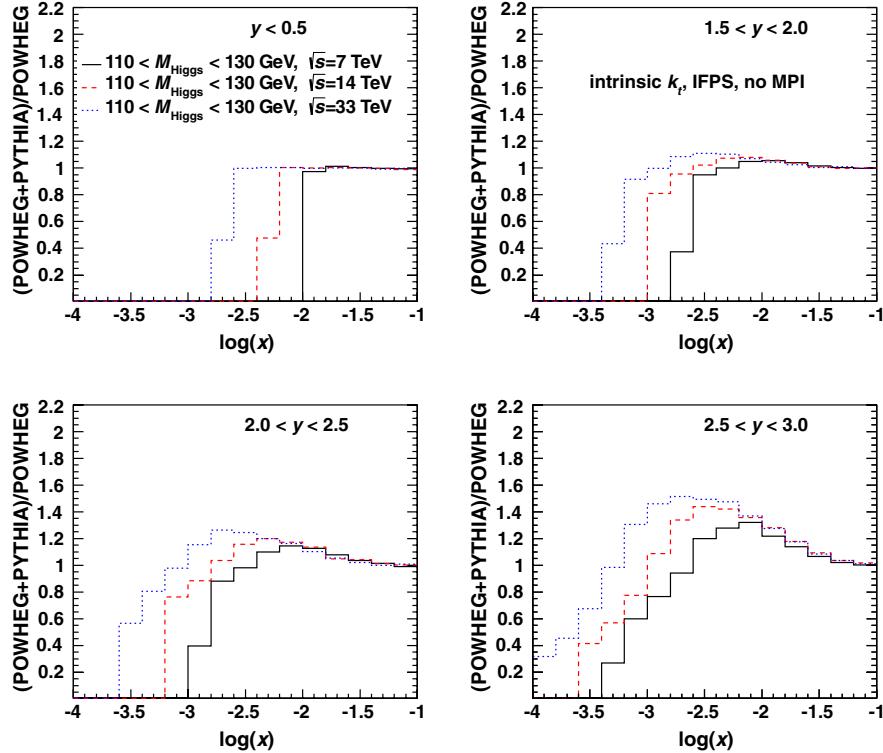


FIG. 10 (color online). Ratio of the cross sections obtained with POWHEG after and before inclusion of initial + final state parton shower and intrinsic k_t for Higgs production at different energies: $\sqrt{s} = 7, 14, 33$ TeV.

central rapidity region. The double peak structure in Fig. 7 comes from the continuum DY production in addition to Z_0 production. It will be of interest to investigate the kinematic reshuffling effect along with the forward Drell-Yan enhancements discussed in [28].

Finally we consider Higgs boson production in Fig. 8 for $110 < m_{\text{Higgs}} < 130$ GeV. We observe a smaller effect at $\sqrt{s} = 7$ GeV than in the previous cases since the x range is limited by the Higgs mass.

Figure 9 summarizes the results in Figs. 5–8 for the ratio of the cross section obtained by POWHEG after inclusion of parton showering to the cross section before parton showering, plotted for different processes. In Fig. 10 we plot this ratio for Higgs boson production at different \sqrt{s} energies of 7, 14, and 33 GeV.

The longitudinal momentum shifts from parton showering computed in this section measure effects from QCD radiation beyond perturbative fixed-order calculations and provide a significant contribution to the correction factors in Sec. II. They affect initial state showers and need to be consistently taken into account in calculations that are used to determine parton density functions. The origin of the kinematical shifts lies with the approximation of collinearity [7] on the partonic states to which the branching algorithms describing showers are applied. Although for explicit calculations we have used a particular NLO-shower matching scheme (POWHEG), the effect is common

to any calculation matching NLO with collinear showers. In calculations using integrated parton density functions the correction factors studied in this paper have to be applied after the evaluation of the cross section (and, as remarked on earlier, this may induce systematic inconsistencies if these corrections are not taken into account properly). On the other hand, this is avoided in approaches using transverse momentum dependent pdfs [11–14,28] from the beginning (transverse momentum dependent pdfs or unintegrated pdfs), as is done for example in the CASCADE event generator [29].

IV. CONCLUSIONS

Theoretical predictions for high-energy collider processes containing hadronic jets require supplementing finite-order perturbative calculations with parton showering and nonperturbative corrections. In this paper we have studied methods to treat parton-showering and nonperturbative corrections in the context of matched NLO-shower event generators.

We have pointed out potential inconsistencies in current approaches that on the one hand apply NP correction factors from leading-order Monte Carlo generators to NLO parton-level predictions and on the other hand fail to include showering corrections. We have proposed methods to address these deficiencies by using consistently

available NLO Monte Carlo tools. We have shown that the differences in the predictions for jet cross sections induced by the modified approach we propose are significant in regions of phase space that are explored with hard probes for the first time at the LHC. In particular, the nonperturbative correction factor K^{NP} introduced in Sec. II gives non-negligible differences at low to intermediate jet p_T , and the showering correction factor K^{PS} of Sec. II gives significant effects over the whole p_T range and is largest at large jet rapidities y .

Because of this y and p_T dependence, taking properly into account NP and showering correction factors changes the shape of jet distributions and affects significantly the comparison of theory predictions with experimental data. The numerical results we have presented show effects as large as 50% in regions of y and p_T phase space relevant to jet measurements at the LHC. The showering correction factor K^{PS} , in particular, can affect the determination of parton distribution functions from fits to experimental data sets comprising inclusive jet measurements.

We have investigated in closer detail the sources of the showering correction from initial state and final state effects. We have observed that the main initial state showering effect comes from kinematical shifts in longitudinal momentum distributions [7] due to combining collinearity approximations with the Monte Carlo implementation of energy-momentum conservation constraints. We have examined the longitudinal shifts for specific processes in Sec. III. This effect is largest for inclusive jets and b -flavor jets at the LHC in the higher rapidity bins. We have

extended the study of longitudinal shifts [7] to the case of Drell-Yan pair production by analyzing the Drell-Yan mass region $16 < m_{\text{DY}} < 166$ GeV and found that the shifts are non-negligible for Drell-Yan production at forward rapidities $y \geq 2$. We have also examined the case of Higgs boson production for $110 < m_{\text{Higgs}} < 130$ GeV and found that the shifts are non-negligible at large rapidities at $\sqrt{s} = 7$ GeV and become more and more important at higher center-of-mass energies.

It will be interesting to study the impact of the effects discussed in this work on phenomenological analyses of LHC final states involving hadronic jets. We expect these effects to also influence determinations of parton distributions. Longitudinal momentum shifts can be avoided in formulations that keep track of noncollinear (i.e., transverse and/or anticolinear) momentum components from the beginning using unintegrated initial state distributions [12,13], also at parton shower level [29,30]. It will be interesting to investigate to what extent this can be exploited to construct approaches in which nonperturbative contributions such as multiple parton interactions, finite transverse momenta, and hadronization are consistently incorporated into parton branching event generators.

ACKNOWLEDGMENTS

We are grateful to Torbjörn Sjöstrand for many discussions concerning hadronization corrections and multiparton interactions in PYTHIA. We are also grateful for many discussions and clarifications on POWHEG to Simone Alioli.

-
- [1] G. Corcella, I. G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M. H. Seymour, and B. R. Webber, *J. High Energy Phys.* **01** (2001) 010; G. Corcella *et al.*, arXiv:hep-ph/0210213.
 - [2] T. Sjöstrand, S. Mrenna, and P. Skands, *J. High Energy Phys.* **05** (2006) 026.
 - [3] P. Nason and B. R. Webber, *Annu. Rev. Nucl. Part. Sci.* **62**, 187 (2012).
 - [4] S. Höche, SLAC Report No. SLAC-PUB-14498, 2011; S. Höche and M. Schönherr, *Phys. Rev. D* **86**, 094042 (2012).
 - [5] G. Aad *et al.* (ATLAS Collaboration), *Phys. Rev. D* **86**, 014022 (2012).
 - [6] S. Chatrchyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **107**, 132001 (2011).
 - [7] F. Hautmann and H. Jung, *Eur. Phys. J. C* **72**, 2254 (2012).
 - [8] G. Marchesini and B. R. Webber, *Nucl. Phys.* **B386**, 215 (1992).
 - [9] F. Hautmann and H. Jung, *J. High Energy Phys.* **10** (2008) 113.
 - [10] M. Deak, F. Hautmann, H. Jung, and K. Kutak, *J. High Energy Phys.* **09** (2009) 121; arXiv:0908.1870; *Eur. Phys. J. C* **72**, 1982 (2012).
 - [11] J. C. Collins, *Foundations of Perturbative QCD* (Cambridge University Press, Cambridge, England, 2011), p. 624.
 - [12] E. Avsar, arXiv:1203.1916; *Int. J. Mod. Phys. Conf. Ser.* **04**, 74 (2011).
 - [13] F. Hautmann, *Acta Phys. Pol. B* **40**, 2139 (2009); *Phys. Lett. B* **655**, 26 (2007); F. Hautmann and H. Jung, *Nucl. Phys. B*, *Proc. Suppl.* **184**, 64 (2008); *AIP Conf. Proc.* **1056**, 79 (2008).
 - [14] P. J. Mulders, *Pramana* **72**, 83 (2009); P. J. Mulders and T. C. Rogers, arXiv:1102.4569.
 - [15] K. Rabbertz, ISMD 2012, Kielce, 2012.
 - [16] M. Cacciari, G. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
 - [17] S. Alioli, K. Hamilton, P. Nason, C. Oleari, and E. Re, *J. High Energy Phys.* **04** (2011) 081.
 - [18] R. D. Field, arXiv:1010.3558.
 - [19] J. Pumplin, D. R. Stump, J. Huston, H.-L. Lai, P. Nadolsky, and W.-K. Tung, *J. High Energy Phys.* **07** (2002) 012.

- [20] Z. Nagy, *Phys. Rev. D* **68**, 094002 (2003).
- [21] P.Z. Skands, *Phys. Rev. D* **82**, 074018 (2010).
- [22] M. Deak *et al.*, arXiv:1206.7090; arXiv:1012.6037.
- [23] S. Catani, M. Ciafaloni, and F. Hautmann, *Phys. Lett. B* **242**, 97 (1990); *Nucl. Phys.* **B366**, 135 (1991); *Phys. Lett. B* **307**, 147 (1993); *Phys. Lett. B* **315**, 157 (1993); *Nucl. Phys.* **B427**, 475 (1994).
- [24] S. Chatrchyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **04** (2012) 084.
- [25] G. Aad *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **71**, 1846 (2011).
- [26] S. Frixione, P. Nason, and B. R. Webber, *J. High Energy Phys.* **08** (2003) 007.
- [27] S. Frixione, P. Nason, and G. Ridolfi, *J. High Energy Phys.* **09** (2007) 126.
- [28] F. Hautmann, M. Hentschinski, and H. Jung, *Nucl. Phys.* **B865**, 54 (2012); arXiv:1205.6358; arXiv:1209.6305.
- [29] H. Jung *et al.*, *Eur. Phys. J. C* **70**, 1237 (2010).
- [30] S. Jadach and M. Skrzypek, *Acta Phys. Pol. B* **40**, 2071 (2009); S. Jadach, M. Jezabek, A. Kusina, W. Placzek, and M. Skrzypek, *Acta Phys. Pol. B* **43**, 2067 (2012).

Appendix G

Higgs as a gluon trigger

Phys. Rev. D88 (2013) 097501

Higgs boson as a gluon triggerP. Cipriano,¹ S. Dooling,¹ A. Grebenyuk,¹ P. Gunnellini,¹ F. Hautmann,^{2,3,4} H. Jung,^{1,5} and P. Katsas¹¹*Deutsches Elektronen Synchrotron, D-22603 Hamburg, Germany*²*Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, United Kingdom*³*Physics Department, University of Oxford, Oxford OX1 3NP, United Kingdom*⁴*Rutherford Appleton Laboratory, Chilton OX11 0QX, United Kingdom*⁵*Elementaire Deeltjes Fysica, Universiteit Antwerpen, B-2020 Antwerp, Belgium*

(Received 30 October 2013; published 26 November 2013)

In the forthcoming high-luminosity phase at the LHC, many of the most interesting measurements for precision QCD studies are hampered by conditions of large pileup, particularly at not very high transverse momenta. We study observables based on measuring ratios of color-singlet currents via Higgs boson and Drell-Yan production, which may be accessed also at large pileup, and used for an experimental program on QCD physics of gluon fusion processes in the LHC high-luminosity runs. We present results of Monte Carlo calculations for a few specific examples.

DOI: [10.1103/PhysRevD.88.097501](https://doi.org/10.1103/PhysRevD.88.097501)

PACS numbers: 14.80.Bn, 12.38.-t

I. INTRODUCTION

The observation of the Higgs boson by the ATLAS and CMS experiments [1] marks the beginning of a revolutionary era in high-energy physics. It affects profoundly the paradigms by which we define the limits of our knowledge on the nature of interactions of elementary particles. This observation gives us confidence in the physical picture of fundamental interactions encoded by the Standard Model (SM) Lagrangian and provides us with guidance in the search for its generalizations.

The electroweak sector of the SM and the nature of electroweak symmetry breaking will be explored in detail in the coming years of operation of the LHC by measuring properties of the observed boson [2]. In this paper, we remark that the observation of the Higgs boson opens up the possibility of a rich experimental program in the strong-interaction sector of the SM as well. In particular, we propose that a program of QCD measurements at high luminosity can be carried out at the LHC by using the Higgs boson as a trigger, focusing on QCD gluonic processes at high mass scales.

Classic collider probes of QCD in e^+e^- annihilation, deep inelastic ep scattering, and Drell-Yan production (DY) all involve color-singlet currents which couple to quarks. With the Higgs, for the first time, LHC experiments will probe QCD by a color-singlet current which, in the heavy top limit, couples to gluons. The physics of gluon fusion processes can be explored from a new perspective compared to experimental investigations over the past three decades. As illustrated below, we propose measuring systematically differences of differential distributions for Higgs and Drell-Yan final states. This comparison allows one to access experimentally distinctive QCD features of gluon fusion physics.

In the next high-intensity phase at the LHC, one faces high pileup conditions leading to large numbers of overlaid events. In these conditions, many of the most interesting

measurements for precision QCD studies, particularly for not very high transverse momenta, become extremely difficult—see e.g. Refs. [2,3]. Here we argue that by studying the differences of Higgs and Drell-Yan for masses around 125 GeV, the effects of pileup largely drop out. This offers the possibility of a program of QCD measurements of great physics interest in the high-luminosity runs of the LHC.

In this paper, we illustrate this by Monte Carlo simulation for three specific examples: the ratio of Higgs vs Drell-Yan p_\perp spectra; the structure of the associated underlying event and charged-particle multiplicities; and the scattering angle in the center-of-mass reference frame. These involve QCD physics both at high transverse momenta and at low transverse momenta, and allow one to study both high- x and low- x physical effects.

We contrast the distinctive features of the Higgs trigger with other LHC short-distance probes such as jets, heavy flavors, and vector boson pairs which either couple perturbatively to color-octet and color-triplet sources on an equal footing, or imply final-state color-charged particles, or both.

We leave to detailed phenomenological investigations the study of the optimal channels to be used to access gluon fusion and suppress Higgs production by vector boson fusion and quark annihilation; of the luminosity requirements for reaching sufficient statistics; and of the different treatment of pileup for different channels.

Very recently, the ATLAS Collaboration has presented its first measurements of Higgs differential cross sections based on the 2012 data set in the diphoton decay channel [4].

II. HIGGS VS DRELL-YAN

Consider first transverse momentum spectra for Higgs bosons and for Drell-Yan (DY) pairs in the invariant mass range $115 \text{ GeV} < M < 135 \text{ GeV}$. Transverse momentum spectra, comparing Higgs and Z bosons, were examined

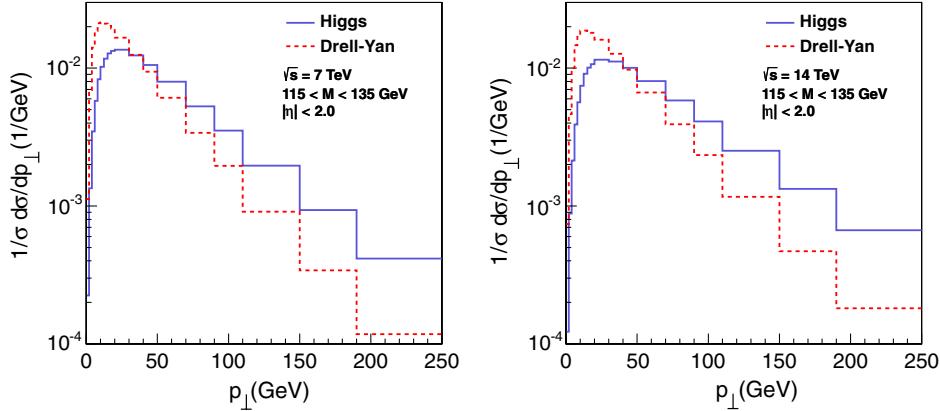


FIG. 1 (color online). Normalized transverse momentum spectra for Higgs bosons and for Drell-Yan pairs.

early on in Ref. [5]. The transverse momentum spectra can be described by QCD factorization in the form

$$d\sigma/dp_\perp = \int H \otimes S \otimes J_1 \otimes J_2, \quad (1)$$

decomposing the cross section into hard (H), soft (S), and collinear-to-initial-states (J_1, J_2) contributions—see e.g. Ref. [6] for analysis of how this decomposition arises. In Fig. 1, we show the result of Monte Carlo simulations for the p_\perp spectra in the central region based on the next-to-leading-order (NLO) POWHEG [7] event generator interfaced with PYTHIA [8] shower, at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 14$ TeV. In Fig. 2, we plot the ratio of the Higgs and DY spectra at invariant mass $115 \text{ GeV} < M < 135 \text{ GeV}$.

The $p_\perp \ll M$ region of the spectrum measures infrared aspects of the cross section in Eq. (1); i.e., (i) the ratio of the gluon vs quark Sudakov form factor [factor S in Eq. (1)], and (ii) the evolution of the collinear-to-initial-states functions [factors J_1, J_2 in Eq. (1)]. In particular, gluon polarization terms $p_\perp^\mu p_\perp^\nu$ in gluon fusion, related to eikonal polarizations at high energy [9], give rise to distinctive radiation patterns from initial-state functions in the Higgs case—see e.g. Ref. [10]. The $p_\perp \geq M$ region measures the ultraviolet function H in Eq. (1) and the features of hard

jets recoiling against the Higgs or DY pair. In particular, the leading-jet contribution to the measured ratio depends on the p_\perp distribution for the spin-1 vs spin-1/2 exchange and on the corresponding color emission probabilities. Further aspects on jet recoil are discussed below in the context of angular distributions.

In the large pileup environment of the high-luminosity LHC runs, one has to deal with the contribution of large numbers of overlaid events. However, this contribution cancels in the comparison of Higgs to DY spectra at fixed invariant mass. Using this comparison, one can go to low p_\perp and access QCD effects in this region experimentally also at high pileup.

Measurements on gluon fusion which can be performed using the Higgs trigger open a new experimental area. They may also be relevant to interpreting data for other, more complex processes, e.g. processes that depend on both quark and gluon channels on an equal footing, or involve color-charged particles in the final state.

One such example is given by top quark production. This is often studied as a process sensitive to gluonic initial states at the LHC. For instance, the top quark p_\perp spectrum [11] receives contributions at low p_\perp from the gluon Sudakov form factor and gluonic initial-state recoil analogous to those discussed above. However, since the final

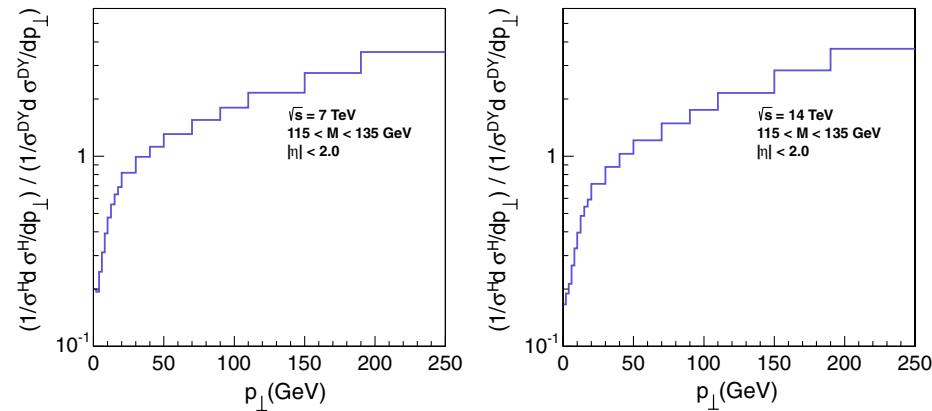


FIG. 2 (color online). Ratio of Higgs to DY spectra vs p_\perp .

state is not a color-singlet current, the analysis of the p_\perp spectrum is made more complex by final soft color emission. The Higgs case serves to single out the initial-state contributions, including gluon polarization effects.

For observables more exclusive than the cross section in Eq. (1), e.g. measuring the associated jets, full QCD factorization formulas are still lacking. For parton shower event generators, inclusive measurements are still useful to control methods [12,13] for merging parton showers and matrix elements. Higgs vs DY studies similar to those considered above can be done, for instance, in boson + jet states, now fixing, in addition to invariant mass, the jet transverse momentum or rapidity.

III. UNDERLYING EVENTS

The structure of underlying events and color flows associated with Higgs boson final states was investigated long ago [14] as a possible method to analyze $gg \rightarrow H$ and $WW \rightarrow H$ production mechanisms. In the case of vector boson final states it was pointed out [15] that the treatment

of parton showers, and in particular of the recoils in the shower, is essential for a proper description of W/Z spectra. This affects the amount of multiparton interactions [16] needed to describe the events [15,17]. Analogous effects may be investigated for gluonic showers [18,19] in the case of events associated with Higgs final states.

We follow the treatment [20] of underlying events in the azimuthal plane, with the directions of the Higgs momentum and the DY-pair momentum, respectively, defining the origin in the azimuthal plane. In Figs. 3 and 4, we show the result of NLO POWHEG + PYTHIA Monte Carlo calculations for charged-particle multiplicities associated with Higgs and DY. (Analogous calculations can be usefully performed for multiplicities of minijets defined e.g. as in Ref. [21].) We plot the average multiplicity vs Higgs and DY p_\perp (Fig. 3) and the multiplicity distribution (Fig. 4) in the transverse region of the azimuthal plane ($60^\circ < |\Delta\phi| < 120^\circ$).

The distributions in the Higgs case are dominated by higher multiplicities from gluon cascades.

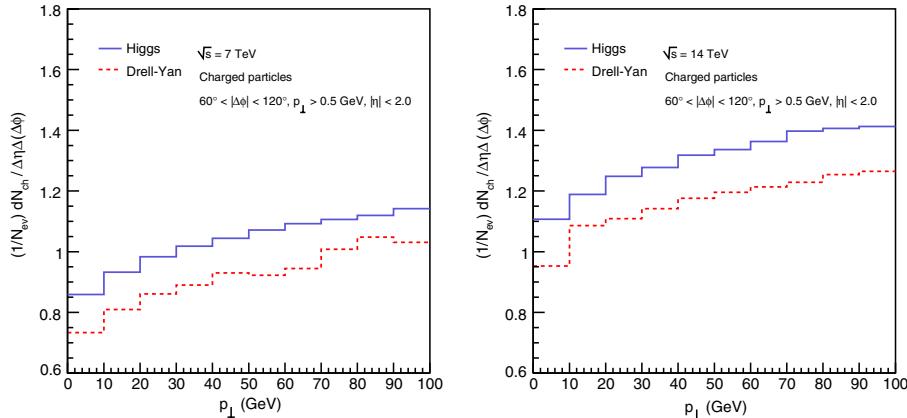


FIG. 3 (color online). Normalized charged-particle average multiplicity in the transverse region of the azimuthal plane vs the Higgs transverse momentum (solid blue line) and the DY transverse momentum (red dashed line).

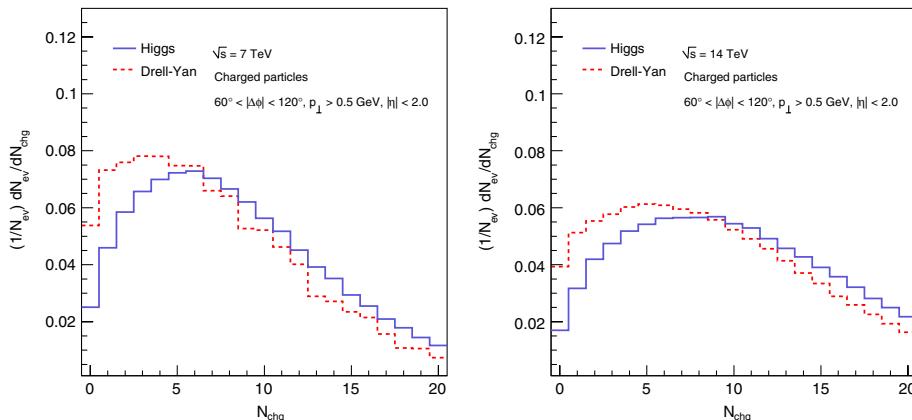


FIG. 4 (color online). Charged-particle multiplicity distribution in the transverse region of the azimuthal plane in the Higgs (solid blue line) and Drell-Yan (red dashed line) cases.

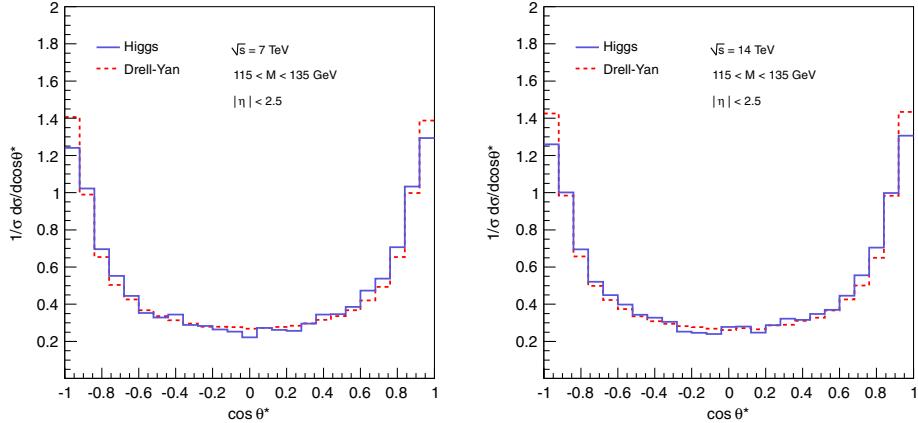


FIG. 5 (color online). Angular distribution in the center-of-mass scattering angle.

Similarly to the case of the previous section, the effects of a large number of overlaid events due to pileup will be reduced if one measures the difference between Higgs and DY underlying event distributions.

IV. ANGULAR DISTRIBUTIONS

Besides soft radiation from underlying events, we consider Higgs vs DY distributions in the case of hard radiation accompanying the heavy bosons, for example boson + jet [22]. For Higgs production, the angular distribution in the scattering angle θ^* of the boson-jet center-of-mass frame is characterized by the scalar coupling to gluons partially canceling the small-angle Coulomb singularity $d\theta^{*2}/\theta^{*4}$ from gluon scattering—see e.g. Ref. [9]. The Drell-Yan θ^* distribution is determined by spin-1/2 exchange. Owing to the cancellation from the scalar coupling to gluons, the angular distributions have the same small-angle asymptotics in the Higgs and DY cases, despite the two processes occurring via spin-1 and spin-1/2 exchange. The $\theta^* \rightarrow 0$ behavior thus tests the Higgs spin at the level of the production cross section.

In Fig. 5, we consider one-jet production associated with Higgs and Z bosons, and show the differential distributions in $\cos\theta^*$, for jet $p_\perp > 20$ GeV and boson-jet invariant mass m such that $200 \text{ GeV} < m < 500 \text{ GeV}$. The rise for increasing $\cos\theta^*$ reflects the mechanism described above. This large $\cos\theta^*$ power counting is the basic reason why the difference between Higgs and DY in the low- p_\perp regions of Figs. 1 and 2 gives a measurement of higher-loop

radiative contributions. Further effects from higher-order color emission may be analyzed via angular correlations in the boson-jet azimuthal plane in the laboratory frame.

In summary, this paper points out that a program of QCD measurements can be carried out in the high-luminosity phase at the LHC, using the Higgs boson as a gluon trigger. By measuring systematically differences between Higgs and Drell-Yan differential distributions for masses around 125 GeV, the effects of pileup largely cancel. Such measurements allow one to access experimentally, for the first time, gluon fusion processes at high mass scales via a color-singlet current. Detailed studies are warranted to investigate quantitatively the reduction of pileup contributions in different channels, the optimal Higgs channels to access gluon fusion by suppressing vector boson fusion and quark annihilation, and the required Higgs statistics. The observables discussed in this paper illustrate that this program spans a broad range of physics issues on strong interactions, from soft gluon dynamics showing up in the ratio of Higgs to DY low- p_\perp spectra, to underlying events and multiple parton interactions associated with gluonic showers, to hard-QCD contributions in large- p_\perp spectra and angular distributions for boson + jet production. These angular distributions in particular may be used to test the spin of the Higgs at the level of production processes.

ACKNOWLEDGMENTS

F. H. gratefully acknowledges the hospitality and support of the Terascale Physics Helmholtz Alliance and DESY.

-
- [1] G. Aad *et al.* (ATLAS Collaboration), *Phys. Lett. B* **716**, 1 (2012); S. Chatrchyan *et al.* (CMS Collaboration), *Phys. Lett. B* **716**, 30 (2012).
 - [2] ATLAS and CMS collaborations, submission to European Strategy for Particle Physics, Krakow, 2012.
 - [3] CMS Collaboration, “QCD at the Extremes”, in European Strategy for Particle Physics, Krakow, 2012.
 - [4] ATLAS Collaboration, Report No. ATLAS-CONF-2013-072.
 - [5] I. Hinchliffe and S. F. Novaes, *Phys. Rev. D* **38**, 3475 (1988).

BRIEF REPORTS

PHYSICAL REVIEW D **88**, 097501 (2013)

- [6] J.C. Collins and F. Hautmann, *Phys. Lett. B* **472**, 129 (2000).
- [7] S. Frixione, P. Nason, and C. Oleari, *J. High Energy Phys.* **11** (2007) 070.
- [8] T. Sjöstrand, S. Mrenna, and P. Skands, *J. High Energy Phys.* **05** (2006) 026.
- [9] F. Hautmann, *Phys. Lett. B* **535**, 159 (2002).
- [10] S. Mantry and F. Petriello, *Phys. Rev. D* **81**, 093007 (2010); **83**, 053007 (2011).
- [11] CMS Collaboration, Report No. CMS-PAS-TOP-12-028; CMS Collaboration, *Eur. Phys. J. C* **73**, 2339 (2013); ATLAS Collabortaion, Report No. ATLAS-CONF-2012-149.
- [12] S. Plätzer, arXiv:1307.0774; *J. High Energy Phys.* **08** (2013) 114.
- [13] L. Lönnblad and S. Prestel, *Prog. Part. Nucl. Phys.* **70**, 1 (2013); arXiv:1211.7228.
- [14] Yu. L. Dokshitzer, V. A. Khoze, and T. Sjöstrand, *Phys. Lett. B* **274**, 116 (1992).
- [15] P. Z. Skands, *Phys. Rev. D* **82**, 074018 (2010).
- [16] T. Sjöstrand and M. van Zijl, *Phys. Rev. D* **36**, 2019 (1987).
- [17] R. D. Field, arXiv:1010.3558.
- [18] L. Lönnblad and M. Sjödahl, *J. High Energy Phys.* **02** (2004) 042; **05** (2005) 038.
- [19] M. Deak *et al.*, arXiv:1006.5401.
- [20] S. Chatrchyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **09** (2011) 109.
- [21] M. Deak, F. Hautmann, H. Jung, and K. Kutak, *Eur. Phys. J. C* **72**, 1982 (2012).
- [22] R. K. Ellis, W.J. Stirling, and B.R. Webber, *QCD and Collider Physics* (Cambridge University Press, Cambridge, 1996).

List of Figures

1.1	Particles of the Standard Model.	5
1.2	Illustration of a proton-proton collision in the parton model. The four-momenta of the protons are defined as $P_{1,2}$, giving the four-momenta of the partons initiating the hard scattering with $p_{1,2}$, depending on the momentum fractions $x_{1,2}$. The interaction of the quark and anti-quark produces a Z^0/γ^* resonance decaying into two leptons l^+l^-	11
1.3	Splitting functions at leading order.	13
1.4	PDFs for quarks, anti-quarks and gluons for energy scale $Q^2 = 10$ GeV 2 (left) and $Q^2 = 10^4$ GeV 2 (right). The distributions are determined from the H1 and ZEUS Collaborations from global fits to data [28].	14
1.5	Kinematic relation between partonic quantities: momentum fraction x , hard scale Q^2 and the final-state quantities: invariant mass M and rapidity y , for different centre-of-mass energies [21].	15
1.6	Nonperturbative corrections as derived in [58] from PYTHIA6 tune Z2 and HERWIG++ with the default tune of version 2.3 in comparison to corrections as derived from POWHEG +PYTHIA6 with the two different underlying event tunes P11 and Z2* [59].	21
1.7	Parton shower corrections as derived from POWHEG +PYTHIA6 for different upper scale limits of the parton shower evolution in PYTHIA6 tune Z2. The curves parametrise the correction factors in the five rapidity regions as a function of the jet p_T [59].	22
1.8	Nonperturbative correction (top left) as obtained from the envelope of the predictions of PYTHIA6 tune Z2, HERWIG++ tune 2.3, and POWHEG +PYTHIA6 with the tunes P11 and Z2*, parton shower correction (top right) as obtained from the average of the predictions of POWHEG +PYTHIA6 tune Z2 with scale factor variation, and total correction (bottom) defined as the product of the nonperturbative and parton shower correction [59].	23
1.9	LO and NLO feynman diagrams for the Drell-Yan process. The solid lines correspond to quarks or anti-quarks, the curly lines represent gluons and the wavy lines are Z^0/γ^*	26

1.10 Drell-Yan production decaying into two muons, showing the contributions in the transverse plane of the detector at LO (left) and NLO (right). At LO, the dimuon system has zero transverse momentum, resulting in a back-to-back composition of the two muons (represented by the two arrows). At NLO one additional gluon (shown here in blue) is emitted providing non-zero p_T to the dimuon system. The gluon is measured as a jet in the detector. The purple dashed arrow indicates the contribution of the dimuon system, which is not directly measured with the detector, but by the two individual muons.	28
1.11 A quark line emitting n gluons. The momentum of the quark decreases by emitting gluons with a momentum fraction z_i of the quark.	31
1.12 Effect of the parton shower (PS) in the Drell-Yan (DY) p_T distribution simulated with PYTHIA6.	33
2.1 Total integrated luminosity in 2011 [87].	37
2.2 LHC timeline [89]. The development of the LHC integrated luminosity and collision energies of the past and the future are presented.	37
2.3 2011 data recorded with CMS at a centre-of-mass energy of 7 TeV in proton-proton collisions. The events are triggered by double and single muon trigger. The dimuon invariant mass spectrum reveals the resonances of mesons and the Z boson.	38
2.4 The CMS detector and its subcomponents, relevant for the particle identification. The several layers are build around the beam pipe resulting in a cylindrical from of the detector [82].	39
2.5 A transverse slice trough CMS. The trajectories of different particles leave a signature in the parts of the detector providing additional information [92].	40
2.6 The CMS tracking system [92]. The tracker is composed of an inner pixel and an outer silicon strip detector. The combination of six TOB, four TIB and on each endcaps three TID and nine TEC in the silicon trip detector ensures a track covering up to $ \eta < 2.4$	42
2.7 A quarter slice of the CMS hadronic calorimeter (HCAL) [93]. The different components, hadronic barrel (HB), hadronic endcap (HE), hadronic outer (HO), and hadronic forward (HF), are arranged around the solenoid.	42
2.8 The muon system [82]. In the barrel the drift tube chambers (DT) are embedded in the iron yoke. In the endcap regions cathode strip chambers (CSC) are used for muon identification. For an optimal positioning measurement resistive plate chambers (RPC) are installed in the barrel and endcaps.	45
2.9 The muon momentum resolution in two different pseudorapidity regions [82]. The resolutions obtained from tracker information, muon system information, and global fits combining both are compared.	46
2.10 Architecture of Level-1 trigger [84].	47
2.11 Trigger turn-on curves of HLT_IsoMu17 in four pseudorapidity regions as a function of the muon p_T . The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events [97].	49

2.12 Trigger turn-on curves for the trigger leg of 13 GeV of HLT_Mu13_Mu8 in four pseudorapidity regions as a function of the muon p_T . The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events [97]. The turn-on curves for the lower leg of 8 GeV can be found in appendix A.	50
3.1 Number of reconstructed primary vertices (pv) in data and simulation. Figure (a) compares data to the unreweighted MC sample. Figure (b) presents the pileup reweighting of the MC sample, taking into account 5% uncertainty on the inelastic proton-proton (pp) cross section in the estimation of the data pileup distribution.	53
3.2 Relative PF Isolation variable. The data is compared to simulation of signal and several background contributions.	57
3.3 A sketch of a jet, presented at different levels. The lines and curly lines represent the partons emerging from the hard scattering (parton level). After fragmentation of partons they hadronise to hadrons and travel in the direction of the initial parton. The final-state particles (this means after fragmentation) can be used as input for jet reconstruction (particle level). The charged and neutral hadrons are grouped together in space in $\eta - \phi$. The hadrons leave energy in the tracker and calorimeter cells, which can be reconstructed to define the jet (detector level).	59
3.4 MC truth response for high and low p_T jets compared in different η bins. The response after JEC and JER smearing is shown.	62
4.1 Event display of inclusive DY and DY + 1 jet production in the transverse (left) and longitudinal (right) plane of the detector. The red lines represent muon candidates and yellow blocks represent PF jets. Red and blue rectangles represent energy deposits in the towers of ECAL and HCAL respectively. The solid green lines represent reconstructed particle tracks.	66
4.2 Illustration of rapidity separation of the forward Drell-Yan production (represented by the Z^0/γ^*) and leading jet (represented by the gluon emission in blue).	67
4.3 Muon kinematic variables after pre-selection. Data is compared to signal and background MC predictions.	70
4.4 Dimuon kinematic variables. The number of events are normalised to the binwidth.	70
4.5 Leading (top) and subleading (bottom) jet variables after the pre-selection. Data is compared to signal and background MC predictions.	71
4.6 An illustration of the definition of the born and bare muons in the final state is shown in (a) and the number of events as a function of the dimuon invariant mass for bare and born muons is presented in (b).	72
4.7 Invariant mass of the dimuon as a function of the positive muon ϕ . Before the muon scale correction (a) the data and simulation predictions show a non-physical dependence on the muon variable. In (b) a flat dependence after the correction is observed.	73
4.8 Dimuon invariant mass (left) and jet multiplicity (right). The different background processes are compared to signal MC events and data. . .	74

4.9	Dimuon $p_T^{\mu\mu}$ distribution (inclusive DY) in five different dimuon invariant mass ranges. The number of events are normalised to the binwidth.	75
4.10	Dimuon $p_T^{\mu\mu}$ distribution (DY + 1 jet) in five different dimuon invariant mass ranges. The number of events are normalised to the binwidth.	76
4.11	Dimuon $p_T^{\mu\mu}$ distribution (DY + 2 jets) in five different dimuon invariant mass ranges. The number of events are normalised to the binwidth.	77
4.12	Invariant mass distribution for the opposite- (OS) and same-sign (SS) data samples. Figures (a) and (b) show the contribution before the isolation requirement. Figure (c) shows the remaining events in the same signed data sample after isolation requirement on the dimuon system.	78
4.13	Opposite- and same-sign number of dimuon events as a function of the dimuon invariant mass in data (left) and MC events (right), with anti-isolation.	79
4.14	Distribution of missing transverse energy before applying the E_T^{miss} requirement. The $t\bar{t}$ background contribution is controlled in the region with $E_T^{\text{miss}} > 80$ GeV, where the top quark pair events dominate the sample.	80
4.15	Data and MC yields after different selection steps. The number of MC events are scaled to data luminosity and selection efficiencies are taken into account.	82
4.16	Dimuon $p_T^{\mu\mu}$ distribution comparing the fiducial region ($ \eta^\mu < 2.1$ and $p_T > 20$ (10) GeV) and the full acceptance. The effect of an acceptance correction is presented for three invariant mass ranges.	83
4.17	Purity, stability, background and acceptance for DY + 1 jet transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass (30 – 45, 60 – 120, 120 – 1500 GeV).	85
4.18	Purity, stability, acceptance and background for the DY + 1 jet production as a function of $ \Delta y(\mu\mu, j_1) $ in three bins of the invariant mass (30 – 60, 60 – 120, 120 – 1500 GeV).	87
4.19	Jet control distributions comparing different event classes to investigate the migration effects due to the jet p_T and pileup jets. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.	89
4.20	Dimuon transverse momentum for the DY + 1 jet case in three different invariant mass ranges. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.	90
4.21	Dimuon transverse momentum for the DY + 2 jets case in three different invariant mass ranges. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.	91
4.22	Absolute rapidity separation between the leading jet and the forward Drell-Yan is shown for the different event topologies for DY + 1 jet production in three different invariant mass ranges. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.	92

4.23	Absolute rapidity separation between the leading jet and the forward Drell-Yan is shown for the different event topologies for DY + 2 jets production in three different invariant mass ranges. The ratio is taken to the inclusive class. The errorbars represent statistical uncertainties.	93
4.24	Illustration of the unfolding response matrix. Contributions from fake and miss events are considered in the unfolding procedure. The effect of bin migrations is shown in the 2d matrix.	95
4.25	The normalised migration matrix of the dimuon $p_T^{\mu\mu}$ distribution for the low ($30 - 45$ GeV) and high ($120 - 200$ GeV) invariant mass range. The inclusive DY, DY + 1 jet, and DY + 2 jets are compared.	96
4.26	The normalised migration matrix of $ \Delta y(\mu\mu, j_1) $ for the low ($30 - 60$ GeV) and high ($120 - 1500$ GeV) invariant mass range. The DY + 1 jet, and DY + 2 jets are compared.	97
4.27	Comparing measured, MC truth, and unfolded dimuon $p_T^{\mu\mu}$ using MADGRAPH for the response matrix and for the measured distribution.	100
4.28	Comparing measured, MC truth, and unfolded dimuon $p_T^{\mu\mu}$ using two statistically independent samples of MADGRAPH.	100
4.29	The $p_T^{\mu\mu}$ distribution for low and high invariant masses in the inclusive DY, DY + 1 jet, and DY + 2 jets topology. Data is compared to MADGRAPH +PYTHIA6 predictions plus a model uncertainty derived from the uncertainties on the data parametrisation.	104
4.30	The $ \Delta y(\mu\mu, j_1) $ distribution for low and high invariant masses in the DY + 1 jet and DY + 2 jets topology. Data is compared to MADGRAPH +PYTHIA6 predictions plus a model uncertainty derived from the uncertainties on the data parametrisation.	105
4.31	The renormalisation and factorisation scale variation by a factor of two above and below the central value. The envelope of all predictions is shown as a band. The dimuon $p_T^{\mu\mu}$ distribution is shown for the inclusive DY, DY + 1 jet and DY + 2 jets production in the Z invariant mass range.	109
4.32	The renormalisation and factorisation scale variation by a factor of two above and below the central value. The envelope of all predictions is shown as a band. The $ \Delta y(\mu\mu, j_1) $ distribution is shown for DY + 1 jet and DY + 2 jets production in Z invariant mass range.	110
5.1	The normalised differential cross section for inclusive Drell-Yan production in five invariant mass ranges $m^{\mu\mu}$ as a function of $p_T^{\mu\mu}$. The results for the inclusive Drell-Yan production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	113
5.2	The normalised differential cross section in five invariant mass ranges $m^{\mu\mu}$ as a function of $p_T^{\mu\mu}$. The results for the DY + 1 jet production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	114

5.3 The normalised differential cross section in five invariant mass ranges $m^{\mu\mu}$ as a function of $p_T^{\mu\mu}$. The results for the DY + 2 jets production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	115
5.4 The normalised differential cross section in the Z resonance mass range as a function of $p_T^{\mu\mu}$. The results for the inclusive DY, DY + 1 jet, and DY + 2 jets productions are compared to the predictions of MADGRAPH +PYTHIA6, POWHEG +PYTHIA6, and PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	118
5.5 Radiation of Z^0/γ^* at different orders in α_S	120
5.6 The normalised differential cross section in the Z resonance mass range as a function of $p_T^{\mu\mu}$. The results for the DY + 1 jet and DY + 2 jets productions are compared to the predictions of MADGRAPH +PYTHIA6 and POWHEG +PYTHIA6 with and without initial-state parton emissions (parton shower = PS). The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	121
5.7 The normalised differential cross section in the Z resonance mass range as a function of $p_T^{\mu\mu}$. The results for inclusive and exclusive DY + 1 jet production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	122
5.8 The normalised differential cross section for DY + 1 jet production in three invariant mass ranges $m^{\mu\mu}$ as a function of $ \Delta y(\mu\mu, j_1) $. The results for the DY + 1 jet production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	124
5.9 The normalised differential cross section for DY + 2 jets production in three invariant mass ranges $m^{\mu\mu}$ as a function of $ \Delta y(\mu\mu, j_1) $. The results for the DY + 2 jets production are compared to the predictions of MADGRAPH +PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	125
5.10 The normalised differential cross section in three invariant mass ranges as a function of $ \Delta y(\mu\mu, j_1) $. The results for the DY + 1 jet production are compared to the predictions of MADGRAPH +PYTHIA6, POWHEG +PYTHIA6, CASCADE, and PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	126
5.11 The normalised differential cross section in three invariant mass ranges as a function of $ \Delta y(\mu\mu, j_1) $. The results for the DY + 2 jets production are compared to the predictions of MADGRAPH +PYTHIA6, POWHEG +PYTHIA6, CASCADE, and PYTHIA6. The errorbars represent statistical uncertainties and the bands represent systematic uncertainties.	127

5.12 Average jet multiplicity as a function of the absolute rapidity separation between the forward Drell-Yan pair and the leading jet. The jets are measured in the fiducial region of $p_T^{\text{jet}} > 30 \text{ GeV}$ and $ \eta^{\text{jet}} < 4.5$. Data are compared to various MC simulations.	128
6.1 Illustration of Higgs and Drell-Yan production, comparing a gluon and quark induced process.	131
A.1 Trigger turn-on curves for the trigger leg of 8 GeV of HLT_Mu13_Mu8 in four pseudorapidity regions as a function of the muon p_T [97]. The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events.	135
A.2 Trigger turn-on curves for runs in 2011B for the trigger leg of 8 GeV of HLT_Mu13_Mu8 in four pseudorapidity regions as a function of the muon p_T [97]. The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events.	136
A.3 Trigger turn-on curves for runs in 2011B for the trigger leg of 13 GeV of HLT_Mu13_Mu8 in four pseudorapidity regions as a function of the muon p_T [97]. The efficiency is estimated using tag-and-probe with $Z \rightarrow \mu\mu$ events.	137
B.1 Average number of reconstructed primary vertices $\langle n_{\text{PV}} \rangle$ as a function of the number of pileup interactions (PUI) in simulation.	139
B.2 Number of primary vertices (pv) in data and simulation. The figure on the left shows the distribution after reweighting and the figure on the right shows the ratio of data over simulation. A good agreement between data and MC predictions is observed after the reweighing.	140
C.1 Distribution of dimuon invariant mass with inverse missing transverse energy requirement. The $t\bar{t}$ background contribution is controlled in the side-band region, $E_T^{\text{miss}} > 80 \text{ GeV}$, where the top quark pair events dominate the sample.	141
C.2 Distribution of dimuon transverse momentum with inverse missing transverse energy requirement. The $t\bar{t}$ background contribution is controlled in the side-band region, $E_T^{\text{miss}} > 80 \text{ GeV}$, where the top quark pair events dominate the sample. A good agreement of data and MC simulation is shown.	142
C.3 Number of b-jets for inclusive DY, DY + 1 jet and DY + 2 jets events.	143
C.4 Dimuon $p_T^{\mu\mu}$ for DY + 2 jets in the invariant mass rage of 200 – 1500 GeV for three different selections in the number of b-jets.	143
D.1 Purity, stability for inclusive DY transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.	146
D.2 Background and acceptance for inclusive DY transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.	147
D.3 Purity, stability, background and acceptance for DY + 1 jet transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.	148
D.4 Purity, stability for DY + 2 jets transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.	149

D.5	Background and acceptance for DY + 2 jets transverse momentum distribution $p_T^{\mu\mu}$ in bins of the dimuon mass.	150
D.6	Purity, stability, background and acceptance as a function of $ \Delta y(\mu\mu, j_1) $ in DY + 2 jets production.	151
D.7	Normalised migration matrix in matched events for the inclusive DY transverse momentum distribution in bins of the dimuon mass.	152
D.8	Normalised migration matrix in matched events for DY + 1 jet transverse momentum distribution in bins of the dimuon mass.	153
D.9	Normalised migration matrix in matched events for DY + 2 jets transverse momentum distribution in bins of the dimuon mass.	154
D.10	Normalised migration matrix in for DY + 1 jet $ \Delta y(\mu\mu, j_1) $ distribution in bins of the dimuon mass.	155
D.11	Normalised migration matrix in for DY + 2 jets $ \Delta y(\mu\mu, j_1) $ distribution in bins of the dimuon mass.	156

List of Tables

1.1	Four fundamental forces in physics.	6
1.2	Properties of the six quark flavours. The given quantum numbers are: electric charge Q, isospin I, charmness C, strangeness S, topness T, bottomness B. The values are taken from the Particle Data Group [7].	9
2.1	Geometric information about CMS and its components. The individual detector parts are listed with their purposes. The coverage in pseudorapidity is presented as well as the inner and outer radii $r_i - r_o$ of the barrel parts and the distance z to the interaction point from the endcap.	40
2.2	High-Level trigger paths used in this analysis. A single muon trigger with p_T threshold of 17 GeV and isolation selection, and a double muon trigger with p_T thresholds of 13 and 8 GeV is used.	48
3.1	Baseline muon selection criteria for 2011 data defined by CMS.	56
3.2	Muon identification efficiency scale factors obtained from data and MC simulation in bins of muon p_T and η [107]. Statistical uncertainties are included and propagated to the scale factors.	58
4.1	Signal and background MC samples.	68
4.2	Ratio of number of events in same-sign and opposite-sign dimuon samples with anti-isolation requirement in data and MC events.	79
4.3	Comparison of the background process event yields for different selection steps. The MC yield is scaled to the data luminosity and selection efficiencies are taken into account.	81
4.4	Summary of typical systematic uncertainties of the inclusive DY transverse momentum in different bins of the dimuon mass. The total systematic uncertainties are given by the quadratic sum of each individual sources.	107
4.5	Summary of typical systematic uncertainties of the DY + 1 jet transverse momentum in different bins of the dimuon mass. The total systematic uncertainties are given by the quadratic sum of each individual sources.	107
4.6	Summary of typical systematic uncertainties of the DY + 2 jets transverse momentum in different bins of the dimuon mass. The total systematic uncertainties are given by the quadratic sum of each individual sources.	108

4.7	Summary of typical systematic uncertainties of $ \Delta y(\mu\mu, j_1) $ distribution in DY + 1 jet and DY + 2 jets (first and second column respectively) events in different bins of the dimuon mass. The total systematic uncertainties are given by the quadratic sum of each individual sources.	108
5.1	Different higher-order processes of the Drell-Yan production generated by various MC generators.	117

Bibliography

- [1] **F. Englert and R. Brout**
Broken Symmetry and the Mass of Gauge Vector Mesons
Phys.Rev.Lett. (1964)13:321–323
- [2] **P. W. Higgs**
Broken Symmetries and the Masses of Gauge Bosons
Phys.Rev.Lett. (1964)13:508–509
- [3] **S. Drell and T.-M. Yan**
Massive Lepton Pair Production in Hadron-Hadron Collisions at High-Energies
Phys.Rev.Lett. (1970)25:316–320
- [4] **O. Phillipsen**
Private lecture notes on Standard Model
- [5] **G. Aad et al. (ATLAS Collaboration)**
Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC
Phys.Lett. (2012)B716:1–29
[hep-ex/1207.7214](#)
- [6] **S. Chatrchyan et al. (CMS Collaboration)**
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC
Phys.Lett. (2012)B716:30–61
[hep-ex/1207.7235](#)
- [7] **J. Beringer et al.**
Particle Data Group
Phys. Rev. D86, 010001 (2012)
<http://pdg.lbl.gov>
- [8] **R. K. Ellis, W. J. Stirling and B. R. Webber**
QCD and Collider Physics
Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge Univ. Press, Cambridge (2003)
- [9] **F. Aaron et al. (H1 Collaboration)**
Measurement of the Azimuthal Correlation between the most Forward Jet and the Scattered Positron in Deep-Inelastic Scattering at HERA
Eur.Phys.J. (2012)C72:1910
[hep-ex/1111.4227](#)
- [10] **C. Adloff et al. (H1 Collaboration)**
Measurement and QCD analysis of jet cross-sections in deep inelastic positron - proton collisions at $s^{**}(1/2)$ of 300-GeV
Eur.Phys.J. (2001)C19:289–311
[hep-ex/0010054](#)
- [11] **A. Aktas et al. (H1 Collaboration)**
Forward jet production in deep inelastic scattering at HERA
Eur.Phys.J. (2006)C46:27–42
[hep-ex/0508055](#)
- [12] **A. Aktas et al. (H1 Collaboration)**
Measurement of inclusive jet production in deep-inelastic scattering at high Q^{**2} and determination of the strong coupling
Phys.Lett. (2007)B653:134–144
[hep-ex/0706.3722](#)
- [13] **S. Chekanov et al. (ZEUS Collaboration)**
Inclusive-jet and dijet cross-sections in deep inelastic scattering at HERA
Nucl.Phys. (2007)B765:1–30
[hep-ex/0608048](#)
- [14] **S. Chekanov et al. (ZEUS Collaboration)**
Forward-jet production in deep inelastic ep scattering at HERA
Eur.Phys.J. (2007)C52:515–530
[hep-ex/0707.3093](#)
- [15] **S. Chekanov et al. (ZEUS Collaboration)**
Forward jet production in deep inelastic ep scattering and low-x parton dynamics at HERA
Phys.Lett. (2006)B632:13–26
[hep-ex/0502029](#)
- [16] **G. Bozzi, S. Catani, G. Ferrera, D. de Florian and M. Grazzini**
Transverse-momentum resummation: A Perturbative study of Z production at the Tevatron
Nucl.Phys. (2009)B815:174–197
[hep-ph/0812.2862](#)
- [17] **A. Lipatov, M. Malyshev and N. Zotov**
Drell-Yan lepton pair production at high energies in the k_T -factorization approach
JHEP (2011)1112:117
[hep-ph/1110.6582](#)
- [18] **V. D. Elvira (D0 and CDF Collaboration)**
QCD, Tevatron Results and LHC Prospects (2008)
[hep-ex/0808.0901](#)
- [19] **R. P. Feynman**
Very high-energy collisions of hadrons
Phys.Rev.Lett. (1969)23:1415–1417
- [20] **J. C. Collins, D. E. Soper and G. F. Sterman**
Factorization of Hard Processes in QCD
Adv.Ser.Direct.High Energy Phys. (1988)5:1–91
[hep-ph/0409313](#)
- [21] **J. M. Campbell, J. Huston and W. Stirling**
Hard Interactions of Quarks and Gluons: A Primer for LHC Physics
Rept.Prog.Phys. (2007)70:89
[hep-ph/0611148](#)
- [22] **Y. L. Dokshitzer**
Calculation of the Structure Functions for Deep Inelastic Scattering and $e^+ e^-$ Annihilation by Perturbation Theory in Quantum Chromodynamics.
Sov.Phys.JETP (1977)46:641–653
- [23] **V. Gribov and L. Lipatov**
Deep inelastic e p scattering in perturbation theory
Sov.J.Nucl.Phys. (1972)15:438–450
- [24] **G. Altarelli and G. Parisi**
Asymptotic Freedom in Parton Language
Nucl.Phys. (1977)B126:298
- [25] **J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky et al.**
New generation of parton distributions with uncertainties from global QCD analysis
JHEP (2002)0207:012
[hep-ph/0201195](#)
- [26] **A. Martin, W. Stirling, R. Thorne and G. Watt**
Parton distributions for the LHC
Eur.Phys.J. (2009)C63:189–285
[hep-ph/0901.0002](#)
- [27] **A. Cooper-Sarkar (ZEUS and H1 Collaboration)**
Extraction of the proton parton density functions using a NLO-QCD fit of the combined H1 and ZEUS inclusive DIS cross sections (2008)page 25
[hep-ph/0808.1854](#)
- [28] https://www.desy.de/h1zeus/combined_results/
- [29] **N. P. Zotov**
Small x physics and hard QCD processes at LHC
PoS (2013)QFTHEP2011:010
[hep-ph/1201.4144](#)

- [30] **D. G. d'Enterria**
Forward Physics at the LHC
Invited Talk at Conference (2007)C07-04-16:1141-1152
[hep-ex/0708.0551](https://arxiv.org/abs/hep-ex/0708.0551)
- [31] **T. Aaltonen et al.** (CDF Collaboration)
Studying the Underlying Event in Drell-Yan and High Transverse Momentum Jet Production at the Tevatron
Phys.Rev. (2010)D82:034001
[hep-ex/1003.3146](https://arxiv.org/abs/hep-ex/1003.3146)
- [32] **A. Knutsson**
Forward Jet Production in ep-Collisions at HERA
Department of Experimental High Energy Physics Lund PhD thesis (2007)
<http://www-h1.desy.de/psfiles/theses/>
- [33] **E. A. Kuraev, L. N. Lipatov and V. S. Fadin**
Multi - Reggeon Processes in the Yang-Mills Theory
Sov.Phys.JETP (1976)44:443-450
- [34] **E. Kuraev, L. Lipatov and V. S. Fadin**
The Pomeranchuk Singularity in Nonabelian Gauge Theories
Sov.Phys.JETP (1977)45:199-204
- [35] **Y. Y. Balitsky and L. N. Lipatov**
The Pomeranchuk singularity in quantum chromodynamics
Sov. J. Nucl. Phys. (1978)28:822-829
- [36] **M. Ciafaloni**
Coherence Effects in Initial Jets at Small q^{**2} / s
Nucl.Phys. (1988)B296:49
- [37] **S. Catani, F. Fiorani and G. Marchesini**
QCD Coherence in Initial State Radiation
Phys.Lett. (1990)B234:339
- [38] **S. Catani, F. Fiorani and G. Marchesini**
Small x Behavior of Initial State Radiation in Perturbative QCD
Nucl.Phys. (1990)B336:18
- [39] **B. Andersson, G. Gustafson, G. Ingelman and T. Sjöstrand**
Parton Fragmentation and String Dynamics
Phys.Rept. (1983)97:31-145
- [40] **B. Andersson, G. Gustafson and B. Soderberg**
A general model for jet fragmentation
Z. Phys. C (1983)20:317-329
- [41] **T. Sjöstrand**
The merging of jets
Phys.Lett. (1984)B142:420-424
- [42] **B. R. Weber**
A QCD model for jet fragmentation including soft gluon interference
Nucl.Phys. (1984)B238:492-528
- [43] **T. Sjöstrand, S. Mrenna and P. Z. Skands**
PYTHIA 6.4 Physics and Manual
JHEP (2006)0605:026
[hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175)
- [44] **R. Field**
Early LHC Underlying Event Data - Findings and Surprises
(2010)
[hep-ph/1010.3558](https://arxiv.org/abs/hep-ph/1010.3558)
- [45] **CMS Collaboration**
Underlying Event Tunes and Double Parton Scattering
CMS-PAS-GEN-14-001
<https://inspirehep.net/record/1293076>
- [46] **R. Corke and T. Sjöstrand**
Interleaved Parton Showers and Tuning Prospects
JHEP (2011)1103:032
[hep-ph/1011.1759](https://arxiv.org/abs/hep-ph/1011.1759)
- [47] **S. Gieseke, D. Grellscheid, K. Hamilton, A. Paepfstatthiou, S. Plätzer et al.**
Herwig++ 2.5 Release Note
(2011)
[hep-ph/1102.1672](https://arxiv.org/abs/hep-ph/1102.1672)
- [48] **S. Gieseke, C. Rohr and A. Siódmiak**
Colour reconnections in Herwig++
Eur.Phys.J. (2012)C72:2225
[hep-ph/1206.0041](https://arxiv.org/abs/hep-ph/1206.0041)
- [49] **F. Maltoni and T. Stelzer**
MadEvent: Automatic event generation with MadGraph
JHEP (2003)0302:027
[hep-ph/0208156](https://arxiv.org/abs/hep-ph/0208156)
- [50] **S. Mrenna and P. Richardson**
Matching matrix elements and parton showers with HERWIG and PYTHIA
JHEP (2004)0405:040
[hep-ph/0312274](https://arxiv.org/abs/hep-ph/0312274)
- [51] **P. Nason**
A New method for combining NLO QCD with shower Monte Carlo algorithms
JHEP (2004)0411:040
[hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146)
- [52] **S. Frixione, P. Nason and C. Oleari**
Matching NLO QCD computations with Parton Shower simulations: the POWHEG method
JHEP (2007)0711:070
[hep-ph/0709.2092](https://arxiv.org/abs/hep-ph/0709.2092)
- [53] **S. Alioli, P. Nason, C. Oleari and E. Re**
NLO vector-boson production matched with shower in POWHEG
JHEP (2008)0807:060
[hep-ph/0805.4802](https://arxiv.org/abs/hep-ph/0805.4802)
- [54] **H. Jung**
The CCFM Monte Carlo generator CASCADE
Comput.Phys.Commun. (2002)143:100-111
[hep-ph/0109102](https://arxiv.org/abs/hep-ph/0109102)
- [55] **H. Jung, S. Baranov, M. Deak, A. Grebenyuk, F. Hautmann et al.**
The CCFM Monte Carlo generator CASCADE version 2.2.03
Eur.Phys.J. (2010)C70:1237-1249
[hep-ph/1008.0152](https://arxiv.org/abs/hep-ph/1008.0152)
- [56] **F. Hautmann and H. Jung**
Transverse momentum dependent gluon density from DIS precision data
Nuclear Physics B (2014)883(0):1 - 19
<http://www.sciencedirect.com/science/article/pii/S0550321314000881>
- [57] **S. Chatrchyan et al.** (CMS Collaboration)
Measurement of the Inclusive Jet Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV
Phys.Rev.Lett. (2011)107:132001
[hep-ex/1106.0208](https://arxiv.org/abs/hep-ex/1106.0208)
- [58] **S. Chatrchyan et al.** (CMS Collaboration)
Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV with the CMS detector
Phys.Rev. (2013)D87(11):112002
[hep-ex/1212.6660](https://arxiv.org/abs/hep-ex/1212.6660)
- [59] **CMS Collaboration**
PDF constraints and extraction of the strong coupling constant from inclusive jet cross section at 7 TeV
CMS-PAS-SMP-12-028
<http://cds.cern.ch/record/1632407>
- [60] **T. Sjöstrand and M. van Zijl**
A Multiple Interaction Model for the Event Structure in Hadron Collisions
Phys.Rev. (1987)D36:2019

- [61] **M. Bahr, S. Gieseke and M. H. Seymour**
Simulation of multiple partonic interactions in Herwig++
JHEP (2008)0807:076
[hep-ph/0803.3633](#)
- [62] **S. Dooling, P. Gunnellini, F. Hautmann and H. Jung**
Longitudinal momentum shifts, showering, and non-perturbative corrections in matched next-to-leading-order shower event generators
Phys.Rev. (2013)D87(9):094009
[hep-ph/1212.6164](#)
- [63] **S. Alioli, P. Nason, C. Oleari and E. Re**
A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX
JHEP (2010)1006:043
[hep-ph/1002.2581](#)
- [64] **S. Alioli, K. Hamilton, P. Nason, C. Oleari and E. Re**
Jet pair production in POWHEG
JHEP (2011)1104:081
[hep-ph/1012.3380](#)
- [65] **F. Hautmann and H. Jung**
Collinearity approximations and kinematic shifts in partonic shower algorithms
Eur.Phys.J. (2012)C72:2254
[hep-ph/1209.6549](#)
- [66] **G. Marchesini and B. Webber**
Final states in heavy quark lepto-production at small x
Nucl.Phys. (1992)B386:215–235
- [67] **F. Hautmann and H. Jung**
Angular correlations in multi-jet final states from k_{\perp} -perpendicular - dependent parton showers
JHEP (2008)0810:113
[hep-ph/0805.1049](#)
- [68] **M. Deak, F. Hautmann, H. Jung and K. Kutak**
Jets in the forward region at the LHC
(2009)
[hep-ph/0908.1870](#)
- [69] **E. Aksar**
TMD factorization and the gluon distribution in high energy QCD
(2012)
[hep-ph/1203.1916](#)
- [70] **E. Aksar**
On the Understanding and Use of ‘Unintegrated’ Parton Distributions in Small- x QCD
Int. J. Mod. Phys. Conf. Ser. (2011)04:74–84
[hep-ph/1108.1181](#)
- [71] **F. Hautmann and H. Jung**
Recent results on unintegrated parton distributions
Nucl.Phys.Proc.Suppl. (2008)184:64–72
[hep-ph/0712.0568](#)
- [72] **F. Hautmann and H. Jung**
Jet correlations from unintegrated parton distributions
AIP Conf. Proc. (2008)1056:79–86
[hep-ph/0808.0873](#)
- [73] **S. Dooling, F. Hautmann and H. Jung**
Hadroproduction of electroweak gauge boson plus jets and TMD parton density functions
Physics Letters B (2014)736(0):293 – 298
[hep-ph/1406.2994](#), <http://www.sciencedirect.com/science/article/pii/S0370269314005371>
- [74] **P. Z. Skands**
Tuning Monte Carlo Generators: The Perugia Tunes
Phys.Rev. (2010)D82:074018
[hep-ph/1005.3457](#)
- [75] **R. Field**
Applications of perturbative QCD
Frontiers in physics. Addison-Wesley, The Advanced Book Program (1989). <http://books.google.de/books?id=rbvAAAAAAJ>
- [76] **S. Ellis, N. Fleishon and W. J. Stirling**
Logarithmic Approximations, Quark Form-factors and Quantum Chromodynamics
Phys.Rev. (1981)D24:1386
- [77] **M. Bonvini**
Resummation of soft and hard gluon radiation in perturbative QCD
Dipartimento di Fisica dell’Università di Genova PhD thesis (2012)
[hep-ph/1212.0480](#)
- [78] **J. C. Collins, D. E. Soper and G. F. Sterman**
Transverse Momentum Distribution in Drell-Yan Pair and W and Z Boson Production
Nucl.Phys. (1985)B250:199
- [79] **C. Balazs and C. Yuan**
Soft gluon effects on lepton pairs at hadron colliders
Phys.Rev. (1997)D56:5558–5583
[hep-ph/9704258](#)
- [80] **J. Collins**
TMD theory, factorization and evolution
International Journal of Modern Physics: Conference Series (2014)25:1460001
<http://www.worldscientific.com/doi/abs/10.1142/S2010194514600015>
- [81] **T. C. Rogers**
TMD Evolution Overview
Presentation at the INT workshop: Studies of 3D Structure of Nucleon (2014)
www.int.washington.edu/talks/WorkShops/int_14_55W/People/Rogers_T/Rogers.pdf
- [82] **G. L. Bayatian et al.** (CMS Collaboration)
CMS Physics: Technical Design Report Volume 1: Detector Performance and Software
Technical Design Report CMS. CERN, Geneva (2006). <https://cds.cern.ch/record/922757>
- [83] **G. L. Bayatian et al.** (CMS Collaboration)
CMS Physics: Technical Design Report Volume 2: Physics Performance
J. Phys. G (2007)34(CERN-LHCC-2006-021. CMS-TDR-8-2):995–1579
<https://cds.cern.ch/record/942733>
- [84] **S. Chatrchyan et al.** (CMS Collaboration)
The CMS experiment at the CERN LHC
JINST (2008)3:S08004
- [85] **O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojevic et al.**
LHC Design Report. 1. The LHC Main Ring CERN-2004-003-V-1
<http://cds.cern.ch/record/782076>
- [86] **W. Herr and B. Muratori**
Concept of luminosity
CERN Accelerator School (2006)Intermediate Course on Accelerator Physics:361–378
- [87] <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>
- [88] **S. van der Meer**
Calibration of the effective beam height in the ISR CERN-ISR-PO-68-31. ISR-PO-68-31 (1968)
<http://cds.cern.ch/record/296752>
- [89] Rolf Heuer in CERN General Meeting January 2013
- [90] **CMS Collaboration**
CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update CMS-PAS-LUM-13-001
<https://cds.cern.ch/record/1598864>
- [91] <http://cerncourier.com/cws/article/cern/54381>
- [92] <http://www.hephy.at/project/cmstracker/gallery/>

- [93] **S. Chatrchyan et al.** (CMS Collaboration) Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data *JINST* (2010)5:T03012 [physics.ins-det/0911.4991](https://arxiv.org/abs/physics/0911.4991)
- [94] **S. Chatrchyan et al.** (CMS Collaboration) Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV *JINST* (2012)7:P10002 [physics.ins-det/1206.4071](https://arxiv.org/abs/physics/1206.4071)
- [95] <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBook>
- [96] https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonHLT#Reference_Efficiencies_for_2011
- [97] **CMS Collaboration** Search for Neutral Higgs Bosons Decaying into Tau Leptons in the Dimuon Channel with CMS in pp Collisions at 7 TeV *CMS-PAS-HIG-12-007, CMS-AN-2012/018* <http://cds.cern.ch/record/1429929>
- [98] **S. G. Lopez** Private Communication
- [99] **S. Agostinelli et al.** (GEANT4) GEANT4: A Simulation toolkit *Nucl.Instrum.Meth.* (2003)A506:250–303
- [100] **J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Dubois et al.** Geant4 developments and applications *IEEE Trans.Nucl.Sci.* (2006)53:270
- [101] **W. Erdmann** Offline Primary Vertex Reconstruction with Deterministic Annealing Clustering *CMS IN -2011/014* (2011) CMS Internal Note
- [102] **CMS Collaboration** Tracking and Primary Vertex Results in First 7 TeV Collisions *CMS-PAS-TRK-10-005* <https://cds.cern.ch/record/1279383/>
- [103] http://home.fnal.gov/~cplager/log/1007/log.html#100729_Idea_for_Pileup_Estimation
- [104] **M. Hildreth** Pileup Issues and Simulation in CMS <https://indico.cern.ch/event/144956/session/1/contribution/17/material/slides/1.pdf>
- [105] **CMS Collaboration** Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV *Physics Letters B* (2013)722(1 - 3):5 – 27 <http://www.sciencedirect.com/science/article/pii/S0370269313002499>
- [106] **CMS Collaboration** Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET *CMS-PAS-PFT-09-001* (2009) <http://cds.cern.ch/record/1194487>
- [107] https://twiki.cern.ch/twiki/bin/viewauth/CMS/HiggsToTauTauWorking2012#Muon_ID_Isolation_EMu_Channel_Va
- [108] **G. P. Salam and G. Soyez** A Practical Seedless Infrared-Safe Cone jet algorithm *JHEP* (2007)0705:086 [hep-ph/0704.0292](https://arxiv.org/abs/hep-ph/0704.0292)
- [109] **S. D. Ellis and D. E. Soper** Successive combination jet algorithm for hadron collisions *Phys.Rev.* (1993)D48:3160–3166 [hep-ph/9305266](https://arxiv.org/abs/hep-ph/9305266)
- [110] **S. Catani, Y. L. Dokshitzer, M. Seymour and B. Webber** Longitudinally invariant K_t clustering algorithms for hadron hadron collisions *Nucl.Phys.* (1993)B406:187–224
- [111] **M. Cacciari, G. P. Salam and G. Soyez** The Anti- $k(t)$ jet clustering algorithm *JHEP* (2008)0804:063 [hep-ph/0802.1189](https://arxiv.org/abs/hep-ph/0802.1189)
- [112] **CMS Collaboration** Pileup Jet Identification *CMS-PAS-JME-13-005* (2013) <http://cds.cern.ch/record/1581583>
- [113] <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID>
- [114] **M. Schröder** Quality of Jet Measurements and Impact on a Search for New Physics at CMS *Department Physik der Universität Hamburg PhD thesis* (2012) <http://www-library.desy.de/diss12.html>
- [115] **CMS Collaboration** Jet Energy Corrections determination at 7 TeV *CMS-PAS-JME-10-010* (2010) <http://cds.cern.ch/record/1308178>
- [116] **S. Chatrchyan et al.** (CMS Collaboration) Determination of jet energy calibration and transverse momentum resolution in CMS *Journal of Instrumentation* (2011)6(11):P11002 <http://stacks.iop.org/1748-0221/6/i=11/a=P11002>
- [117] <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution>
- [118] **CMS Collaboration** Measurement of the double differential Drell-Yan cross section at low and high masses of the di-muon system at $\text{sqrt}s = 7$ TeV (2014)(CMS-PAS-FSQ-13-003) <http://cds.cern.ch/record/1670353>
- [119] **P. Cipriano, S. Dooling, A. Grebenyuk, P. Gunnellini, F. Hautmann et al.** Higgs boson as a gluon trigger *Phys.Rev.* (2013)D88(9):097501 [hep-ph/1308.1655](https://arxiv.org/abs/hep-ph/1308.1655)
- [120] **J. Andersen, V. Del Duca, F. Maltoni and W. J. Stirling** W boson production with associated jets at large rapidities *JHEP* (2001)0105:048 [hep-ph/0105146](https://arxiv.org/abs/hep-ph/0105146)
- [121] **R. Chierici et al.** Standard Model Cross Sections for CMS at 7 TeV <https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSections>
- [122] **CMS Collaboration** Measurement of Z production as a function of pT, Y *CMS-PAS-SMP-13-013* <http://cds.cern.ch/record/1700115>
- [123] **A. Bodek, A. van Dyne, J. Han, W. Sakumoto and A. Strelnikov** Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments *Eur.Phys.J.* (2012)C72:2194 [hep-ex/1208.3710](https://arxiv.org/abs/hep-ex/1208.3710)
- [124] **CMS Collaboration** Combination of top pair production cross section measurements *CMS-PAS-TOP-11-024* (2011) [http://cds.cern.ch/record/1401250/](http://cds.cern.ch/record/1401250)
- [125] **K. Melnikov and F. Petriello** Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_S^3)$ *Phys. Rev. D* (2006)74:114017 [http://link.aps.org/doi/10.1103/PhysRevD.74.114017](https://arxiv.org/abs/hep-ph/0603103)

- [126] **K. Melnikov and F. Petriello**
W Boson Production Cross Section at the Large Hadron Collider with $\mathcal{O}(\alpha_S^2)$ Corrections
Phys. Rev. Lett. (2006)96:231803
<http://link.aps.org/doi/10.1103/PhysRevLett.96.231803>
- [127] **V. Khachatryan et al.** (CMS Collaboration)
Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV
JHEP (2011)1101:080
[hep-ex/1012.2466](https://arxiv.org/abs/hep-ex/1012.2466)
- [128] **S. Schmitt**
TUnfold: an algorithm for correcting migration effects in high energy physics
JINST (2012)7:T10003
[1205.6201](https://arxiv.org/abs/1205.6201)
- [129] **G. D'Agostini**
A Multidimensional unfolding method based on Bayes theorem
Nucl.Instrum.Meth. (1995)A362:487
- [130] **B. Wynne**
Imagiro: an implementation of Bayesian iterative unfolding for high energy physics
(2012)
<https://cds.cern.ch/record/1434247/>
- [131] **T. Adye**
Corrected error calculation for iterative Bayesian unfolding
(2011)
http://hepunx.rl.ac.uk/~adye/software/unfold/bayes_errors.pdf
- [132] **T. Adye**
<http://hepunx.rl.ac.uk/~adye/software/unfold/RooUnfold.html>
- [133] **S. Alioli, P. Nason, C. Oleari and E. Re**
Vector boson plus one jet production in POWHEG
JHEP (2011)1101:095
[hep-ph/1009.5594](https://arxiv.org/abs/hep-ph/1009.5594)
- [134] **H1 and ZEUS Collaborations**
PDF fits including HERA-II high Q^2 data (HERA-PDF1.5)
H1prelim-10-142, ZEUS-prel-10-018
<http://www-h1.desy.de/h1/www/publications/htmlsplit/H1prelim-10-142.long.html>
- [135] **J. M. Campbell, R. K. Ellis, P. Nason and G. Zanderighi**
W and Z bosons in association with two jets using the POWHEG method
JHEP (2013)1308:005
[hep-ph/1303.5447](https://arxiv.org/abs/hep-ph/1303.5447)
- [136] **S. Chatrchyan et al.** (CMS Collaboration)
Measurement of the properties of a Higgs boson in the four-lepton final state
Phys.Rev. (2014)D89:092007
[hep-ex/1312.5353](https://arxiv.org/abs/hep-ex/1312.5353)
- [137] **H. Van Haevermaet, P. Cipriano, S. Dooling, A. Grebenyuk, P. Gunnellini et al.**
Higgs boson as a gluon trigger: the study of QCD in high pile-up environments
(2014)
[hep-ph/1407.2815](https://arxiv.org/abs/hep-ph/1407.2815)
- [138] **S. Chatrchyan et al.** (CMS Collaboration)
Identification of b-quark jets with the CMS experiment
JINST (2013)8:P04013
[hep-ex/1211.4462](https://arxiv.org/abs/hep-ex/1211.4462)
- [139] **S. Bhattacharya** (CMS Collaboration)
Efficiency measurement of b-tagging algorithms developed by the CMS experiment
(2011)
[hep-ex/1110.4569](https://arxiv.org/abs/hep-ex/1110.4569)
- [140] **K. Yamamoto**
Measurement of the Drell-Yan differential cross section $d(\sigma)/d(p_T)$ at high dilepton mass in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector
Iowa State University PhD thesis (2012)
http://lib.dr.iastate.edu/physastro_etd/
- [141] **A.-K. N. Sanchez**
Studies of W and Z Bosons with the CMS Experiment at the LHC
ETH Zurich PhD thesis (2012)
<http://e-collection.library.ethz.ch/view/eth:6180>
- [142] **J. Hauk**
Measurement of Associated Z^0 -Boson and b-Jet Production in Proton-Proton Collisions with the CMS Experiment
Department Physik der Universität Hamburg PhD thesis (2012)
<http://cms.desy.de/e48945/>
- [143] **J. Bjorken and E. A. Paschos**
Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon
Phys.Rev. (1969)185:1975–1982
- [144] **S. Chatrchyan et al.** (CMS Collaboration)
Measurement of the differential and double-differential Drell-Yan cross sections in proton-proton collisions at $\sqrt{s} = 7$ TeV
JHEP (2013)1312:030
[hep-ex/1310.7291](https://arxiv.org/abs/hep-ex/1310.7291)
- [145] **S. Schumann and F. Krauss**
A Parton shower algorithm based on Catani-Seymour dipole factorisation
JHEP (2008)0803:038
[hep-ph/0709.1027](https://arxiv.org/abs/hep-ph/0709.1027)
- [146] **CMS Collaboration**
Performance of Jet Algorithms in CMS
CMS-PAS-JME-07-003
[http://cds.cern.ch/record/1198227](https://cds.cern.ch/record/1198227)
- [147] **J. E. Huth, N. Wainer, K. Meier, N. J. Hadley, F. Aversa, M. Greco, P. Chiappetta, J. P. Guillet, S. Ellis, Z. Kunzst and D. E. Soper**
Toward a standardization of jet definitions
(1990)(FERMILAB-CONF-90-249-E):7 p
- [148] **CMS Collaboration**
Jet Performance in pp Collisions at 7 TeV
CMS-PAS-JME-10-003 (2010)
<https://cds.cern.ch/record/1279362>
- [149] **CMS Collaboration**
Jet Energy Resolution in CMS at $\sqrt{s}=7$ TeV
CMS-PAS-JME-10-014 (2011)
[http://cds.cern.ch/record/1339945](https://cds.cern.ch/record/1339945)
- [150] **F. Hautmann**
High-multiplicity final states and transverse-momentum dependent parton showering at hadron colliders
(2009)
[hep-ph/0909.1240](https://arxiv.org/abs/hep-ph/0909.1240)
- [151] **Y. Dokshitzer, D. D'yakonov and S. Troyan**
On the transverse momentum distribution of massive lepton pairs
Physics Letters B (1978)79(3):269 – 272
<http://www.sciencedirect.com/science/article/pii/037026937890240X>
- [152] **G. Parisi and R. Petronzio**
Small transverse momentum distributions in hard processes
Nuclear Physics B (1979)154(3):427 – 440
<http://www.sciencedirect.com/science/article/pii/0550321379900403>
- [153] **M. Duong-van**
The Transverse Momentum Distribution of the Dileptons in Hadronic Processes
SLAC-PUB-1819 (1976)
[http://cds.cern.ch/record/117337](https://cds.cern.ch/record/117337)

Bibliography

- [154] **M. Grazzini**
Soft gluon resummation vs parton shower simulation
Presentation at the MCWS workshop (2006)
<http://moby.mib.infn.it/~nason/mcws2/Grazzini.pdf>
- [155] **S. Alioli, J. R. Andersen, C. Oleari, E. Re and J. M. Smillie**
Probing higher-order corrections in dijet production at the LHC
Phys. Rev. (2012)D85:114034
[hep-ph/1202.1475](http://arxiv.org/abs/hep-ph/1202.1475)
- [156] **J. R. Andersen, T. Hapola and J. M. Smillie**
W Plus Multiple Jets at the LHC with High Energy Jets
JHEP (2012)1209:047
[hep-ph/1206.6763](http://arxiv.org/abs/hep-ph/1206.6763)
- [157] **Dittmaier, S et al.** (LHC Higgs Cross Section Working Group)
Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables
CERN, Geneva (2011). <http://cds.cern.ch/record/1318996>
- [158] <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt14TeV>
- [159] **ATLAS Collaboration**
Differential cross sections of the Higgs boson measured in the diphoton decay channel using 8 TeV pp collisions
ATLAS-CONF-2013-072, ATLAS-COM-CONF-2013-086 (2013)
<http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2013-072/>

Statement of authorship

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Hamburg, 08 September 2014

Acknowledgements

I would like to express my deepest gratitude to Dr. Hannes Jung, my research supervisor, for his valuable guidance over the last three years. Lieber Hannes, ich danke Dir vielmals für die vielen Dinge die Du mir beigebracht hast. Danke, dass Deine Bürotür immer offen stand und ich immer zu Dir kommen konnte. Du hast mich unterstützt und mich ermutigt mehr zu schaffen. Von Dir habe ich gelernt immer zielstrebig und mit Freude zu arbeiten, aber auch mal nach Hause zu gehen, wenn das Wetter schön ist. Danke, dass Du es mir ermöglicht hast mit vielen verschiedenen Menschen über meine Arbeit zu sprechen und zu diskutieren. Deine Begeisterung, wenn Du mal wieder eine neue Idee hattest, hat mich stets angesteckt und mir Freude an der Arbeit bereitet. Vielen Dank für alles.

My gratitude to my referees Prof. Dr. Peter Schleper, Prof. Dr. Günther Dissertori and Dr. Hannes Jung for agreeing to evaluate my thesis. I would also like to thank Prof. Dr. Peter Schleper, Dr. Hannes Jung, Dr. Alexei Raspereza, Prof. Dr. Dieter Horns, and Prof. Dr. Bernd Kniehl for taking part in my disputation and evaluating my work.

I would also like to thank Alexei, for his invaluable support and his help in clarifying the physics and coding. Alexei, it was a pleasure working with you. Your constant willingness to help, especially your enthusiastic support prior to and during the approvals, is very much appreciated.

Additionally I would like to thank Armin, for his support in the computing and coding elements. It was a pleasure working with you.

I would like to express my up most appreciation to Prof. Dr. Francesco Hautmann, for his productive and intriguing theory discussions. Francesco, your ideas and advice encouraged me go on to explore particle physics further.

To Dr. Jochen Heitger. Lieber Jochen, vielen Dank, dass Du mich in die Welt der Teilchenphysik geführt hast. Du hast mich unterstützt als ich mit meiner Arbeit für die Diplomarbeit anfing und standest mir bis zum Ende bei. Du hast mir beigebracht wissenschaftlich zu denken und zu arbeiten. Vielen Dank!

I am particularly grateful to have shared my office, work, chocolate, and my moods (good and bad), with my colleague and friend, Paolo. Your cheerful manner is contagious and it was a pleasure to have spent the last years with you by my side.

My special thanks are extended to Carmen, who had to suffer answering my questions, over and over again. Danke, Carmen, dass Du so streng mit mir warst, aber mich auch immer wieder aufgebaut hast. Die gemeinsamen DESY Fahrradtouren, Sportabende und Tanzeinlagen in meiner Küche werde ich auf ewig in Erinnerung haben.

I would also like to thank Ivan, for his lively early morning discussions in physics and other topics. It is a pleasure to start the working day with a laugh.

I am particularly grateful to be a member of the Physics & Cookies Group, where

problems as well as happy moments were always shared. I would like to express my gratitude to Albert, Nastja, Panos, Pedro, Mira, Juan, Hans, Benoit, Ola, Patrick and Thomasz.

I would also like to thank Johannes, Jasone, Jan, Matthias and Francesco for physics discussions and fun conversations.

It has been an honour to have been accepted into the DESY CMS working group. Thank you for the friendly atmosphere in the coffee room and the meetings. It was a pleasure knowing you.

To the FSQ and small-x working group and especially to the conveners, Jonathan, David, Gabor, Albert and Grigory. Thank you all for the enthusiastic encouragement and useful critiques on my research. Thank you for encouraging me to give presentations and go to conferences. The assistance provided by you all is greatly appreciated.

To the SMP working group and especially Klaus, Panos and Maxime. Thank you for your support concerning the nonperturbative and parton shower corrections. It was a pleasure working with you all.

Für meine Mädels. Danke, dass Ihr mir immer Mut zugesprochen habt und mit mir die letzten Tage bis zur Abgabe gezählt habt. Ihr seid immer für mich da, wenn ich Euch brauche und habt mit mir schon jede Prüfung überstanden.

Für Christine. Ich danke Dir, dass Du mich immer wieder gepusht hast, ob es nun im Laufen oder bei der Arbeit war. Du hast immer zugehört und mich aufgebaut. Danke, dass Du für mich da bist.

Für Paul, Clärchen und Peter. Vielen Dank für Eure Unterstützung in den letzten Jahren. Ihr habt mir geholfen die Zeit zu überstehen, mehr als Ihr vielleicht denkt. Danke, für die Kerzen, die Ihr für mich angezündet habt, und für die Marmelade, die einfach die Beste ist.

Für Christian. Du warst vom ersten bis zum letzten Satz dieser Arbeit an meiner Seite. Du hast mir die Kraft gegeben jeden Morgen aufzustehen und noch eine Seite zu schreiben, auch wenn es mir manchmal schwer viel. Danke, dass Du für mich da bist. xx

Für meine kleinen Schwestern Kelly und Chelsea. Danke, dass Ihr immer an mich geglaubt habt, mir zugehört habt. Ich bin sehr stolz darauf Eure große Schwester zu sein. Ihr seid wunderbar. xx

Für Mami und Papi. Danke für Eure Unterstützung in allem was ich mache. Ihr seid immer für mich da und habt nie an mir gezweifelt. Papi, danke, dass Du immer glücklich bist und mir Kraft gibst. Deine guten Vorzäte geben mir Mut und von Dir weiß ich, dass man alles schaffen kann, wenn man es nur will. Ich danke Dir für Deine Hilfe in der englischen Sprache, die ich niemals so gut beherrschen werde wie Du die Deutsche! Mami, danke, dass Du immer für mich da bist, wenn es mir schlecht geht. Ohne Deine Zwangspausen hätte ich es nicht geschafft. Du weißt einfach was das Beste für mich ist. Ich liebe euch! xx