Karlsruhe Institute of Technology

IEKP-KA/2012-24

Measurement of the three-jet mass cross-section at $\sqrt{s} = 7$TeV

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT) genehmigte

DISSERTATION

von

Dipl.-Phys. Fred Markus Stober
aus Karlsruhe

Referent: Prof. Dr. Günter Quast
Institut für Experimentelle Kernphysik

Korreferent: Prof. Dr. Wim de Boer
Institut für Experimentelle Kernphysik

Tag der Mündlichen Prüfung: 19. Oktober 2012

Introduction

Since the dawn of time, mankind strives to understand the basic concepts and properties of the world. Data is gathered in experiments while theories are formulated to explain the measurements and to give predictions for future experiments. In the last decades, this interplay between experiment and theory has lead to the discovery of four fundamental forces. The force most know for it's impact on the daily life is the gravitational force, which is well described by the theory of general relativity. However in the context of particle physics, the other three forces are dominant. The strong, electromagnetic and weak force, which rule the processes in particle physics, are contained within the framework of the Standard Model. The Standard Model allows to explain and calculate an incredible number of observations and measurements.

However the Standard Model also leaves many questions unanswered: Is it possible to further unify the fundamental forces into one unified force at higher energies? Are there phenomena in particle physics which cannot be explained by the Standard Model or even contradict its predictions? Are the predictions given by the Standard Model valid up to the highest energies or is there a more fundamental theory?

This study hopes to shed light on at least part of that last question: Are the predictions given by the theory of the strong interaction valid up to the highest energies?

In order to answer all these raised questions, experimental particle physics pushes the boundaries of the available theories with increasingly sophisticated instruments. To accomplish this, it is necessary to coordinate many independent and geographically separate teams of physicists, the design and mass production of high-tech equipment and the utilisation of vast computing resources for data analysis. One of the latest such experiments is the Large Hadron Collider (LHC), which delivers proton-proton collisions at centre-of-mass energies on the TeV scale.

These collisions mainly create quarks and gluons, which undergo a process called hadronization, resulting in a collimated stream of colour-neutral

objects for each created parton. To reconstruct the parton properties from the detector measurement of such a stream, the concept of a jet is introduced. A jet is an object, which tries to collect all particles originating from the same initial parton. There are several different jet algorithms available, which can perform this clustering.

Due to the nature of the collisions, a large number of events with multiple jets, ruled by the laws of quantum chromodynamics (QCD) which describes the strong force, is available. One of the fundamental tasks in the area of experimental QCD is the study of multi-jet observables and the comparison between the measurement and the best available theory prediction.

This study performs such a comparison for the three-jet mass at next-to-leading order theory precision and a centre-of-mass energy of 7 TeV. The three-jet mass is defined as the invariant mass of the three jets with the largest transverse momentum passing a set of acceptance and analysis cuts. The three-jet mass cross section $d^2\sigma/dm_3dy_{\max}$ is measured double differentially in the three-jet mass m_3 and the maximum rapidity of the three-jet system $y_{\max}$. The binning of the three-jet mass is resolution driven, while the rapidity binning used to partition the phase-space into disjoint regions is loosely based on the CMS detector geometry.

Over the course of seven chapters, the necessary theoretical and experimental background is introduced and the results for the measurement of the three-jet mass cross-section is presented.

The first chapter of this thesis gives an introduction to the Standard Model of particle physics with an emphasis on the physics related to quantum chromodynamics and the strong coupling constant. Different shower and hadronization models employed by the used Monte-Carlo programs are also described.

In the second chapter, the Large Hadron Collider is presented. The LHC experiment providing the data for this study, the Compact Muon Solenoid (CMS), is described in greater detail. In particular, the CMS sub-detectors concerned with the energy measurements of jets are emphasised.

Chapter three deals with the computing infrastructure built for the LHC and used to process the datasets for this analysis. The chapter also gives a short overview of the analysis environment developed by the CMS collaboration and several programs used to give theory predictions. Other software, written to support this analysis, is introduced as well. This includes grid-control which was used to utilise the provided computing resources and the data processing framework Kappa which forms the foundation of the analysis code.

The topic of chapter four is the reconstruction of jet from detector data and Monte-Carlo simulation. A short introduction to the particle-flow re-

construction applied by CMS is given. This is followed by the description of the jet clustering algorithms used in the analysis. The chapter closes with an explanation of the jet energy corrections, which are necessary to correct the measured and simulated reconstructed jets for detector effects.

The analysis part of the thesis starts in chapter five with the next-to-leading order (NLO) theory prediction for the three-jet mass cross-section. The different cut scenarios used in this study are also introduced. K-factors describing the difference between the leading order and next-to-leading order prediction are discussed for two different factorisation and renormalisation scale choices. NLO predictions for the three-jet mass cross-section are presented for different parton density functions (PDF). This is followed by a discussion about the scale, PDF and α_s uncertainties of the prediction. The presentation of the non-perturbative corrections and a new method to derive them concludes this chapter.

Chapter six details the analysis of the data measured by CMS. The chapter starts with an introduction of the used datasets and follows up with the event selection. This is followed by a discussion of the observed differences between the measured, uncorrected three-jet mass distribution in data and the distribution taken from fully simulated events. As a next step, the three-jet mass resolution is measured to determine an optimal binning for the three-jet mass distribution. Several unfolding algorithms are studied in order to facilitate future comparisons of the measured three-jet mass distribution with the results from other experiments and theoretical calculations. The chapter closes with a detailed study of the measurement uncertainties.

Chapter seven concludes this thesis with a summary and outlook. In the outlook, the theory predictions and measurements from the previous two chapters are used to derive information about the running of the strong coupling constant at high energy scales. The chapter closes with an overview showing how the results of this study extend the established set of measurements at very high scales.

Contents

Chapter 1

The Standard Model of Particle Physics

1.1 Introduction

One of the primary goals of science is trying to understand the natural principles which rule the world around us. In Ancient Greece, this question was treated as a predominantly philosophical problem and was little influenced by empiricism. During the Middle Ages scholars in the Arab world increasingly emphasised experimentation as well as quantification to answer physical problems. Since the 17th century, the systematic observation, measurement and experiment, and the formulation, testing, and modification of hypotheses as manifest in the scientific method played an important part in the development of the natural sciences. This incremental systematisation of science has lead to a wide spectrum of scientific research involving huge experiments and worldwide collaborations, which try to answer the fundamental questions about the universe around us.

In particular, scientists on the quest to understand the origin and composition of matter made great strides. At the start of the 19th century, increasingly sophisticated experiments began to discover more and more particles when looking at atomic and subatomic reactions. In the 1970s the work on this issue culminated in a simple and elegant theory, which describes most observed particles as composite objects build up from a small number of fundamental particles. This theory of the fundamental constituents of matter and the interactions among them is called the Standard Model of Particle Physics. All observed interactions can be described in terms of the four fundamental forces: gravitation, weak, electromagnetic and strong interaction. However there are large differences between the range and strength of

these forces. While gravitation and the electromagnetic force have an infinite range, the weak and strong interactions are only relevant over microscopic distances. The Standard Model manages to describe the three forces of the strong, electromagnetic and weak interaction using a common mathematical language.

Although gravitation is a force with a large influence on macroscopic objects, it is the weakest of all interactions when looking at the force between microscopic objects, at currently accessible distances which are relevant for the description of particle interactions in collider experiments. The introduction of gravity in the Standard Model is also hindered by the fact that there is currently no theory, which describes gravitation in terms of a quantum theory.

The Standard Model has become successful because it can provide answers to many questions within particle physics to a very high precision with comparably few input parameters, demonstrating the predictive power of the theory. This success has lead to the construction of a wide range of different experiments to test the theory in as many aspects as possible, in order to find deviations which might give hints towards an even more unified theory. From the theory side, improvements to the tools and mathematical techniques for these predictions has lead to algorithms and simulation software, which are able to give high precision results. The relevant tools for this study are described in chapter 3. In the area of collider experiments, the LHC accelerator and its experiments like CMS have opened up a new energy frontier in the last few years and are described in chapter 2. The following introduction to the Standard Model is based on [1, 2, 3, 4, 5] and [6], which contain more details about the mathematical structure underlying the theory.

Unit Conventions In particle physics it is convenient to use the so-called Heaviside–Lorentz units instead of SI units. This means, among other conventions, that rather than $\hbar = 1.0546 \times 10^{-34}\,\mathrm{J\,s}$ and $c = 299\,792\,458\,\mathrm{m\,s^{-1}}$ the following definition is used:

$$\hbar = 1 \text{ and } c = 1.$$

With this convention all relevant units can be expressed in terms of electronvolt eV, with

$$1\,\mathrm{eV} = 1.6022 \times 10^{-19}\,\mathrm{J},$$

for example particle masses (eV), momenta (eV) and other quantities like distances (eV^{-1}). A unit commonly used to express cross sections in particle physics is the barn, which is defined as

$$1\,\mathrm{b} = 10^{-28}\,\mathrm{m}^2$$

1.1.1 Mathematical Background

The Standard Model of particle physics is based on the mathematical principles of quantum field theory. Quantum field theory in turn is founded on the principles of local symmetries, gauge invariance and quantum mechanics.

The fundamental particles within the Standard Model are described as excited states of space-time coordinate dependent quantum fields $\psi_i(x)$. Lagrangian field theory describes the dynamics of these fields, using the techniques from Lagrangian mechanics on these fields $\psi_i(x)$.

Lagrangian Field Theory

In the framework of a Lagrangian field theory, with a given Lagrangian $\mathcal{L}(\psi_i, \partial_\mu \psi_i)$ as a function of the fields ψ_i and their first derivatives $\partial_\mu \psi_i$, the equation of motion can be derived by minimising the action S, which is a functional of ψ_i and $\partial_\mu \psi_i$

$$S = \int d^4x \, \mathcal{L}(\psi_i, \partial_\mu \psi_i)$$

The result of this procedure is the Euler-Lagrange equation:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \psi_i \right)} \right) - \frac{\partial \mathcal{L}}{\partial \psi_i} = 0$$

Here the Lagrangian $\mathcal{L}$ is taken as an axiomatic starting point defined through model building. Starting from very simple Lagrangians, this formalism can be used to derive important equations known from relativistic quantum mechanics:

- The Klein-Gordon equation for a real spin-0 field $\phi(x)$ with mass m:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2} m^2 \phi^2 \Rightarrow \left(\partial^\mu \partial_\mu + m^2 \right) \, \phi(x) = 0$$

- The Dirac equation for a spin-1/2 spinor field $\psi(x)$:

$$\begin{aligned} \mathcal{L} &= \overline{\psi} \left(i \gamma^\mu \partial_\mu - m \right) \psi \\ &\Rightarrow \; i \partial_\mu \bar{\psi} \gamma^\mu + m \bar{\psi} = 0 \text{ and } \left(i \gamma^\mu \partial_\mu - m \right) \, \psi(x) = 0 \end{aligned}$$

  with the Dirac matrices γ^μ and $\bar{\psi} = \psi^\dagger \gamma^0$. In this Lagrangian ψ and $\bar{\psi}$ are treated as two independent variables. Both resulting equations are equivalent due to $\gamma^0 \gamma^\mu = -\gamma^\mu \gamma^0$.

Local Symmetries and Gauge Invariance

One of the pillars underpinning the Standard Model are symmetries. Within the Standard Model these symmetries give rise to conserved quantum numbers and the boson fields of the model according to the principles of local symmetries and gauge invariance. From Noether's theorem it is known that every symmetry induces a conserved quantity. Time invariance of a physical process leads to the conservation of energy and invariance under rotation leads inevitably to conservation of angular momentum. These conservation laws originate from the fact that the solution of the equation of motion does not change under the corresponding symmetry transformations. Therefore, a theory is invariant under a symmetry group G, represented by a unitary operator U, if the fields $\psi(x)$ and $\psi'(x)$, given by

$$\psi(x) \rightarrow \psi'(x) = U\psi(x)$$

follow the same equation of motion.

For every transformation which lets the Lagrangian invariant ($\delta\mathcal{L} = 0$), Noether's theorem states that there exists an associated Noether current $j^\mu(x)$ and a corresponding Noether charge Q

$$\partial_\mu j^\mu(x) = 0 \text{ and } Q := \int d^3x\, j^0.$$

Since this Noether current is a conserved quantity, the Noether charge itself is time invariant.

$$\frac{dQ}{dt} = 0 \text{ if } \delta\mathcal{L} = 0$$

For discrete symmetries these Noether charges are the conserved quantum numbers of the theory. Noethers theorem applies to both, local and global symmetries. The gauge symmetries of the Standard Model are local symmetries, which appears plausible when considering that global symmetries act on different space-time points in the same manner, regardless of how far they are separated or causally connected. Local symmetries on the other hand are implemented by making the parameters of the gauge group G and thus their representations $U = U(x)$ space-time dependent. In the following, the non-Abelian gauge group SU(N) is used because this gauge group is very important for the Standard Model[1]. A quantum field theory based on this non-Abelian group is also called a Yang-Mills theory. It was first formulated by Yang and Mills in the 1950s in an attempt to describe the strong isospin[7]

[1] $SU(N)$ is the set of all unitary $N \times N$ matrices with determinant 1.

and today forms the basis for the electroweak unification and quantum chromodynamics. The group SU(N) has $N^2 - 1$ independent generators T^a and every element of this Lie-group can be parametrised in the form:

$$U(x) = \exp\left(i\theta_a(x)T^a\right) \in \mathrm{SU(N)}$$

with space-time dependent parameters $\theta_a(x)$ and summing over same indices. The generators T^a of the group SU(N) form the Lie-Algebra $\mathfrak{su}$(N):

$$[T^a, T^b] = if^{abc}T^c \tag{1.1}$$

with the structure constants f^{abc}. The knowledge of these structure constants in turn is sufficient to determine the underlying group.

Any equation of motion for a Lagrangian containing normal derivatives ∂_μ, like the Lagrangian for a free spin-1/2 particle

$$\mathcal{L}_{\text{free}} = \bar{\psi}\left(i\gamma_\mu\partial^\mu - m\right)\psi$$

is not invariant under local gauge transformations $U(x)$. The derivatives give rise to terms of the form $\partial_\mu\theta(x)$ in the transformed Lagrangian, which breaks gauge invariance. However, this problem can be solved using a method known as minimal substitution. With this method, the derivative ∂_μ is replaced by the covariant derivative D_μ to preserve the invariance. The covariant derivative introduces a new vector field, the so called gauge field A^a_μ:

$$\partial_\mu \to D_\mu := \partial_\mu - ig\,A^a_\mu\,T^a$$

This gives rise to a new term in the Lagrangian, which describes the interaction between the gauge field A^a_μ and the field ψ:

$$\mathcal{L}_{\text{free}} \to \bar{\psi}\left(i\gamma_\mu D^\mu - m\right)\psi = \bar{\psi}\left(i\gamma_\mu\partial_\mu - m\right)\psi + \underline{g_s\,\bar{\psi}\gamma_\mu A^a_\mu T^a\psi}$$

To fully describe this gauge field A_μ, an appropriate kinematic term $\mathcal{L}_{\text{kin}}$ needs to be added to the Lagrangian. After looking for possible forms of this term, it is clear that only terms with the field strength tensor $F^a_{\mu\nu}$ as given below are allowed. Trying to introduce mass terms of the form $m^2A^a_\mu A^{a,\mu}$ results in breaking the gauge invariance, therefore the gauge field A^a_μ has to be massless.

The field strength tensor for a SU(N) gauge theory with generators T^a and structure constants f^{abc} is defined as:

$$F^a_{\mu\nu} := \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\,f^{abc}A^b_\mu A^c_\nu$$

	Generation			Isospin	Hyper-charge	el. Charge	Colour
	1.	2.	3.	T_3	Y	$Q = T_3 + Y$	
Quarks	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$+\frac{1}{6}$	$\begin{pmatrix} +\frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}$	r,g,b
	u_R	c_R	t_R	0	$+\frac{2}{3}$	$+\frac{2}{3}$	r,g,b
	d_R	s_R	b_R	0	$-\frac{1}{3}$	$-\frac{1}{3}$	r,g,b
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	$\begin{pmatrix} +\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$	$-\frac{1}{2}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	-
	e_R	μ_R	τ_R	0	-1	-1	-

Table 1.1: Overview of the fundamental particles of the Standard Model.

The additional term with this field strength tensor is called the gauge part of the Lagrangian:

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{4} F^a_{\mu\nu}(x) F^{a,\mu\nu}(x)$$

This process of restoring the gauge invariance of the Lagrangian and choosing the vector field A^a_μ is called the gauging of the theory.

1.1.2 Fundamental Particles and Interactions

In the Standard Model all matter is made of quarks and leptons. They are point-like, structureless spin-1/2 fermions. All Standard Model interactions between these particles can be described by local gauge theories, in which the forces are mediated by fundamental spin-1 gauge bosons.

Aiming towards the understanding of quarks and leptons and their interactions in such a framework, one has to discover the underlying conservation laws of the observed forces. After a long time of studies, three symmetries belonging to two areas of the theories have been identified. The two basic components of the Standard Model are: the spontaneously broken $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ electroweak theory, and the unbroken $\mathrm{SU}(3)_C$ colour gauge theory, known as quantum chromodynamics (QCD) which will be described in more detail in section 1.2.

Unfortunately there is no deeper understanding of why exactly the gauge group of the Standard Model is $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$. The Standard

Force	Range [m]	Force carrier	Boson mass [GeV]	J^P
Strong	10^{-15}	8 gluons (g)	0	1^-
Electromagnetic	∞	photon (γ)	0	1^-
Weak	10^{-13}	$W^\pm$ and Z	80.42 and 91.19	1
Gravitational	∞	graviton (?)	?	2 (?)

Table 1.2: Overview of the fundamental forces. There is no consistent quantum theory of gravity, however it is possible to make some general considerations about the properties like the spin of its hypothetical force carrier.

Model gauge group gives rise to the following gauge boson content:

$$\begin{array}{ccccc} \mathrm{SU}(3)_C & \times & \mathrm{SU}(2)_L & \times & \mathrm{U}(1)_Y \\ \downarrow & & \downarrow & & \downarrow \\ G^\alpha_\mu & & W^a_\mu & & B_\mu \\ {\scriptstyle \alpha=1\ldots8} & & {\scriptstyle a=1\ldots3} & & \end{array}$$

There are three spin-1 fields $W^a_\mu(x)$ that are associated with the $\mathrm{SU}(2)_L$ gauge group and one field $B_\mu(x)$ associated with the group $\mathrm{U}(1)_Y$. The subscript L indicates that only left-handed fermions carry this quantum number. The subscript Y represents the quantum number associated with the group $\mathrm{U}(1)_Y$. The four spin-1 gauge bosons $W^a_\mu(x), B(x)$ are related to the experimentally observed bosons $W^\pm, Z$ and γ that mediate the weak and electromagnetic interaction via

$$W^\pm_\mu = \frac{1}{\sqrt{2}}\left(W^1_\mu \mp iW^2_\mu\right)$$
$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\vartheta_W & \sin\vartheta_W \\ -\sin\vartheta_W & \cos\vartheta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}$$

where the photon field is named A_μ. Due to the process of spontaneous symmetry breaking, the symmetry $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ is broken down to the gauge symmetry $\mathrm{U}(1)_{\mathrm{em}}$. The corresponding gauge boson for this symmetry is the photon and the corresponding quantum number is the ordinary electric charge. The same mechanism that is responsible for breaking the symmetry allows the physical bosons $W^\pm$ and Z to acquire mass. This process is caused by a massive spin-0 particle, the Higgs boson. A new particle, which could be the Higgs boson, was discovered by CMS and ATLAS[8] in June 2012 with a mass of 125.3 GeV[9]. Further details on the Higgs mechanism are available in [4, 10, 11].

The eight particles associated with the gauge group $\mathrm{SU}(3)_C$ are called gluons $G^\alpha_\mu(x)$. Gluons are thought to be massless spin-1 particles. The

subscript C denotes the charge of the particles, colour. Any particle that transforms according to this gauge group is said to be coloured or to carry colour. The interaction corresponding to this group is the strong interaction.

There are three different colour charges r, g, b. This stands in contrast to the commonly known electric charge which only has one charge.

An important hint that the Standard Model is not yet a complete description of the subatomic world is the observation of neutrino oscillations, which necessitates the neutrinos to have a mass. However it is not possible to include neutrino mass term into the Standard Model Lagrangian without spoiling the renormalisability[2] of the theory or adding right handed neutrinos.

Free Parameters of the Standard Model Just like the Standard Model is not able to explain the origin of its underlying gauge group, there are some fundamental parameters in the model whose values it can not predict. These free parameters have to be measured in experiments. By testing the self-consistency of the Standard Model with independent precision measurements of observables related to a common free parameter of the model, it might be possible to uncover tensions and perhaps find hints pointing towards a theory beyond the Standard Model. The fundamental, free parameters are:

- (1) The Weinberg-angle ϑ_W.
- (3) The three gauge couplings g_1, g_2,g_3 of the electroweak and strong interaction.
- (9) The masses of the six quarks and three leptons (or the equivalent Yukawa couplings to the Higgs field).
- (1) The mass of the Higgs boson.
- (4) The matrix elements of the CKM-Matrix[12] that describes the mixing between the down-type quark eigenstates for the strong and weak interaction.

These parameters are adding up to the 18 free parameters of the Standard Model. Extensions of the Standard Model which incorporate the observation of Neutrino oscillations require at least three additional parameters for the Neutrino masses and four parameters describing their mixing.

[2]Renormalisability is an important property of the theory and further described in 1.2.2

1.2 Quantum Chromodynamics

1.2.1 Introduction

Historically, Quantum chromodynamics was introduced to solve a problem with the observation of the Δ^{++} particle. Naively, its wave function has the form $\Psi_\Delta = \psi_{\text{flavour}}\psi_{\text{orbital}}\psi_{\text{spin}}$. Since the particle is composed of three up-quarks, the first part is symmetric under the exchange of two quarks. Also, since the spin of this particle is 3/2, all three quark spins are pointing spatially into the same direction, so the spin wave-function is also symmetric. Finally, the interchange of the quarks in the orbital part yields a factor $(-1)^L$, which is $+1$ because it is assumed that $L = 0$ for the ground state of this particle. This means that the particle is a spin 3/2 fermion with a symmetric wave-function. However the spin-statistics theorem states that a fermion must necessarily be totally antisymmetric with respect to the interchange of the quantum numbers of its constituents.

Although it looked like the Δ^{++} particle violates the spin-statistics theorem, physicists were reluctant to abandon the theorem since it can be derived from a small set of fundamental axioms whose validity was without question. Thus, the existence of another quantum number was postulated to circumvent this fundamental problem and to save the theorem.

The conservation of this new quantum number is associated with the symmetry group SU(3) and the local gauge symmetry gives rise to the gauge bosons of the strong interaction - the gluons. These gluons carry both, colour and anti-colour, and from group theory it is known that they are grouped into an octet and a singlet. The gluon octet can be represented by

$$r\bar{b}, r\bar{g}, b\bar{g}, b\bar{r}, g\bar{r}, g\bar{b}, \frac{1}{\sqrt{2}}(r\bar{r} - b\bar{b}), \frac{1}{\sqrt{6}}(r\bar{r} + b\bar{b} - 2g\bar{g}).$$

while the singlet can be written as:

$$\frac{1}{\sqrt{3}}(r\bar{r} + b\bar{b} + g\bar{g})$$

Mathematically, the generators T^a of the SU(3) group are described by the Gell-Mann matrices λ^a with $T^a = \lambda^a/2$. As shown in the preceding sections, the Lagrangian density which is invariant under local $SU(3)$ transformations can be constructed as:

$$\begin{aligned} \mathcal{L}_{\text{QCD}} &= \bar{\psi}\left(i\gamma_\mu\partial_\mu - m\right)\psi - \frac{1}{4}F^a_{\mu\nu}F^{a,\mu\nu} + g_s\,\bar{\psi}\gamma_\mu G^a_\mu T^a\psi \\ \text{with } F^a_{\mu\nu} &= \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s\, f^{abc}G^b_\mu G^c_\nu \end{aligned}$$

The Lagrangian of the theory is split into three parts. The first part is the free quark Lagrangian. The second part describes the kinematics of the gluon field and the third part the interaction vertex between the quark ψ and the gluon field G^a_μ. The coupling between them is given by the strong gauge coupling g_s. It is common to speak about the strong coupling constant $\alpha_s = g_s^2/4\pi$ instead of g_s.

Especially interesting is the term $F^a_{\mu\nu}F^{a,\mu\nu}$ describing the kinematics of the gluon field. Due to its non-Abelian group, the structure constants f^{abc} are non-vanishing as described in equation 1.1. This gives rise to interactions of the form $g_s\, G^\mu G^\nu \partial_\mu G_\nu$, the triple gluon vertex, and $g_s^2\, G^\mu G^\nu G_\mu G_\nu$, the quadruple gluon vertex.

Together with the quark interaction $g_s\, \bar{\psi}\gamma_\mu G^a_\mu T^a \psi$, these terms represent the three fundamental QCD interaction vertices. From the corresponding Feynman rules, it is possible to write down matrix elements for the transition between any given initial and final states using the prescriptions of Quantum field theory.

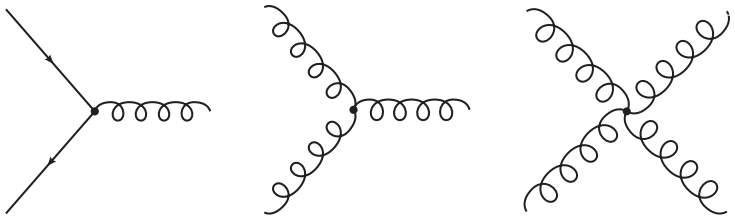


Figure 1.1: The three fundamental QCD interaction vertices.

1.2.2 The Running QCD Coupling α_s

The coupling strength α_s, the analogue to the fine-structure constant α in QED, is the principal free parameter in the theory of the strong interaction. The behaviour of α_s at different energy scales is of great importance to understand the interaction between quarks and gluons.

Asymptotic Freedom and Confinement

Experiments show that quarks and gluons behave like free particles for short time intervals ($t < 10^{-24}\,\mathrm{s}$), small distances ($\ll 1\,\mathrm{fm}$) or high energies ($q^2 > 1\,\mathrm{GeV}^2$). This implies that the strength of the strong coupling α_s decreases

for large momentum transfers q^2. It is similar to the electromagnetic screening effect, but on small instead of large scales. The effect, which is due to the gluon self-coupling, is called asymptotic freedom.

At large distances however ($\sim 1\,\mathrm{fm}$) or small values of q^2 ($q^2 < 1\,\mathrm{GeV}$) no free coloured objects are observed. A probable explanation is that for sufficiently large distances ($> 1\,\mathrm{fm}$) the colour field between them has accumulated a sufficient amount of energy to create new quark-antiquark pairs from the vacuum. In turn, these quark-antiquark pairs fragment into colourless hadrons. This observation called confinement implies that the coupling strength α_s, becomes large and eventually diverges in the regime of large distances ($\sim 1\,\mathrm{fm}$) or small momentum transfers ($q^2 < 1\,\mathrm{GeV}^2$). This also means that the perturbative treatment of partons (i.e. quarks and gluons) in this region is no longer possible and that the strong force has a limited range although the gluons are massless.

From the QCD gauge group it is possible to derive a differential equation which describes the energy dependence of α_s. The initial value for the equation is a free parameter which must be derived from experimental data. Determining α_s at a specific energy scale q is therefore a fundamental task in QCD. It is also possible to perform self-consistency tests of the theory by probing several different energy scales and comparing the predictions of the running of α_s.

Renormalisation

The following introduction to renormalisation theory follows the path laid out in [13]. In a quantum field theory like QCD, a dimensionless physical observable X can be expressed by a perturbation series in the coupling parameter α_s. Calculating X in standard QCD implies operating with bare charges and masses of point particles, which are free of self-interactions, however this means one has to deal with additional emissions and ultraviolet (UV) divergences from the integration of loop momenta. They manifest themselves in the form of gluon or quark self-energy and vertex corrections. These divergences are removed by a procedure called renormalisation. There are different schemes to do this (like the $\overline{MS}$ scheme) but they have in common that they introduce a second energy scale μ and that X and α_s become functions of μ. Here μ represents the point at which the removal of the UV divergences is performed. A natural choice of the renormalisation scale is the energy scale of the process [14]. The observable X can be expressed as a function of α_s and a single energy scale q when neglecting all masses in the problem. Since X is dimensionless, one can assume that:

$$X = X(q^2/\mu^2, \alpha_{s\mu}), \qquad \alpha_{s\mu} = \alpha_s(\mu^2)$$

As the choice of μ is arbitrary, the observable has to be independent of μ^2 for constant α_s. This can be mathematically expressed by:

$$\mu^2 \frac{d}{d\mu^2} X(q^2/\mu^2, \alpha_{s\mu}) \stackrel{!}{=} 0 = \left(\mu^2 \frac{\partial}{\partial \mu^2} + \mu^2 \frac{\partial \alpha_{s\mu}}{\partial \mu^2} \frac{\partial}{\partial \alpha_{s\mu}} \right) X(q^2/\mu^2, \alpha_{s\mu}) \quad (1.2)$$

It follows that any explicit dependence of X on μ^2 arising via q^2/μ^2 must cancel the one associated with α_s. Equation 1.2 is called the renormalisation group equation (RGE) for X. An example for such an observable X is the matrix element of a perturbative QCD process

Energy Dependence of α_s

The next step is to introduce the β function:

$$\beta(\alpha_{s\mu}) = \mu^2 \frac{\partial \alpha_{s\mu}}{\partial \mu^2} \quad (1.3)$$

To see how this leads to a description of α_s at different energy scales, the variable $t = \ln(q^2/\mu^2)$ is introduced. With this variable, equation 1.2 becomes:

$$\left(-\frac{\partial}{\partial t} + \beta(\alpha_{s\mu}) \frac{\partial}{\partial \alpha_{s\mu}} \right) X(e^t, \alpha_{s\mu})$$

By implicitly defining a new function - the running coupling $\alpha_s(q^2)$, this first order differential equation can be solved:

$$t = \int_{\alpha_{s\mu}}^{\alpha_s(q^2)} \frac{d\alpha}{\beta(\alpha)} \quad (1.4)$$

By differentiating 1.2 with respect to t at fixed $\alpha_{s\mu}$ and with respect to $\alpha_{s\mu}$ at fixed t, one obtains:

$$1 = \frac{1}{\beta(\alpha_s(q^2))} \frac{\partial \alpha_s(q^2)}{\partial t} \text{ and } 0 = \frac{\partial \alpha_s(q^2)}{\partial \alpha_{s\mu}} \frac{1}{\beta(\alpha_s(q^2))} - \frac{1}{\beta(\alpha_{s\mu})}$$

$$\left(-\frac{\partial}{\partial t} + \beta(\alpha_{s\mu}) \frac{\partial}{\partial \alpha_{s\mu}} \right) \alpha_s(q^2) = 0$$

From which follows that $X(1, \alpha_s(q^2))$ is a solution to equation 1.2. This remarkable result shows that the complete energy scale dependence of X can be absorbed into the running of the coupling constant $\alpha_s(q^2)$ [15]. However this result is only valid under the assumption above that q^2 is so large that one can neglect all masses. From this result also follows that if it is possible

to calculate X at the scale $q^2 = \mu^2$ at some definite order of perturbation theory, one can predict the scale dependency of X by solving 1.4.

Expanding the β function perturbatively in terms of α_s yields:

$$\beta(\alpha_s(q^2)) = -\beta_0\alpha_s^2(q^2) - \beta_1\alpha_s^3(q^2) - \beta_2\alpha_s^4(q^2) + \mathcal{O}(\alpha_s^5) \tag{1.5}$$

From QCD loop calculations the coefficients β_i can be calculated[16]:

$$\begin{aligned}
\beta_0 &= \frac{33 - 2n_f}{12\pi} \\
\beta_1 &= \frac{153 - 19n_f}{24\pi^2} \\
\beta_2 &= \frac{77139 - 15099n_f + 325n_f^2}{3456\pi^3}
\end{aligned}$$

where n_f is the number of quarks with a mass less than the energy scale q. Solving equation 1.3 with ansatz 1.5 in leading order, gives the one-loop solution for the energy dependence of $\alpha_s(q^2)$:

$$\alpha_s(q^2) = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2)\beta_0 \ln\left(\frac{q^2}{\mu^2}\right)} \tag{1.6}$$

It is common to evaluate equation 1.6 at $\mu = M_Z$ to compare between different measurements. According to the Particle Data Group[16], the current best estimate of the strong coupling is:

$$\alpha_s(M_Z) = 0.1176 \pm 0.002$$

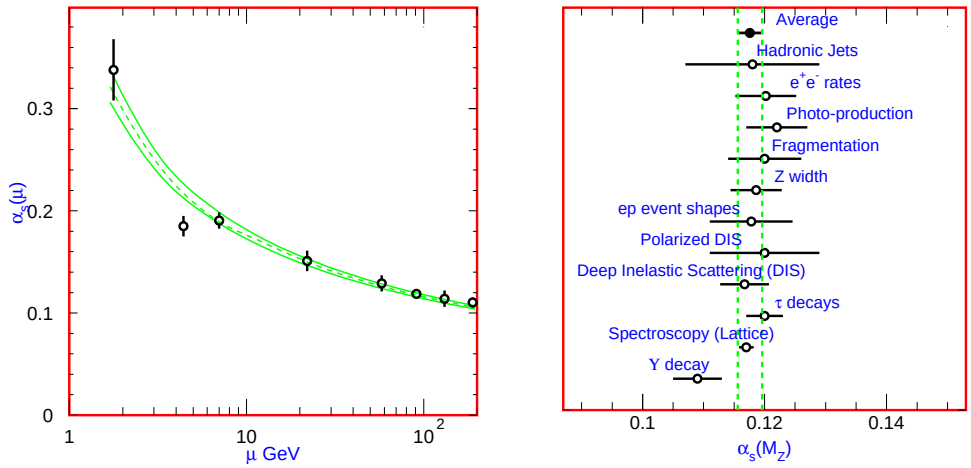
Already at one loop level, the solution 1.6 shows the behaviour of α_s which was measured in experiments. In figure 1.2 and equation 1.6 the aforementioned confinement and asymptotic freedom can be observed.

Further details are available in [17] and [18].

1.3 Bound States in QCD

Coloured objects like quarks and gluons are confined together with other coloured objects into colourless objects. These objects are known as hadrons. Currently, two kinds of such bound states of quarks have been observed. The fermionic baryons qqq and the bosonic mesons $q\bar{q}$ as well as the corresponding anti-particles. From lattice gauge[3] QCD computations there are

[3] A lattice gauge theory is a gauge theory on a discretised space-time.



(a) Summary of the values of $\alpha_s(\mu)$ at the values of μ where they are measured. The lines show the central values and the $\pm 1\sigma$ limits. The figure clearly demonstrates the decrease in $\alpha_s(\mu)$ with increasing μ.

(b) Summary of the value of $\alpha_s(M_Z)$ from various processes. The values shown indicate the process and the measured value of α_s extrapolated to $\mu = M_Z$.

Figure 1.2: Energy dependence of the QCD coupling $\alpha_s = g_s^2/4\pi$ (Source [16])

predictions[16] of more exotic compound particles like tetra-quarks $q\bar{q}q\bar{q}$, penta-quarks $qqqq\bar{q}$ or glue-balls which consist only of gluons, but there is currently no experimental evidence for their existence.

The large number of catalogued hadrons is due to the large number of combinations which are possible by having six different flavours and their anti-partners. A large number of excited particles are also assigned their own name because their energies due to spin and orbital momentum can reach or even surpass the order of magnitude of their ground state energy.

These bound states are subjected to conservation and quantisation laws: Baryon number conservation holds true (counting anti-baryons as -1), the charge of hadrons is an integer multiple of the elementary charge and the colour charge vanishes. While hadrons can be described by their valence quarks, which give rise to the hadron quantum numbers, the mass of bound states is much larger than the sum of their bare masses. However it is possible to calculate many masses and lifetimes using the already mentioned lattice gauge QCD computations.

There are only two known stable bound states with very long lifetimes, the proton and the neutron. There are lower limits on the proton lifetime of $2{\cdot}10^{29}$ years, and the reason for this long lifetime is a topic under study. However the neutron will decay when not bound with other hadrons in a nucleus after

a free mean lifetime of just 881 s. Due to their stable nature even when not bound in atoms, protons are ideal candidates to study hadronic interactions.

To study the quite complicated inner dynamics of these bound states, point-like particles are used in deep-inelastic scattering experiments to probe the proton structure. To resolve the proton structure, energies of the order of at least 1 GeV are needed.

These probes uncovered that baryons consist of three valence quarks bound together by a gluon field. The gluon field between the valence quarks carries a large amount of the momentum of the hadron and causes by virtual pair production and annihilation the appearance of sea-quarks, which can be heavy quarks (like s, c, b) even in light hadrons like the proton. The exact composition of protons and neutrons is given by parton distribution functions (PDFs). Here the quantity $f_i(x, q^2)dx$ gives the probability that a parton of type i (quark, antiquark or gluon) carries a momentum fraction between x and $x + dx$ of the total hadron momentum for a probe at scale q^2. An example for a PDF set is shown in figure 1.3. For the theoretical description of these distributions, it is important to recall that the partons are in a collinear configuration with very small relative momentum. Hence the interaction between them is very strong and cannot be calculated perturbatively. However, it is possible to formulate a model, which allows to apply the renormalisation group equation to the problem.

It is necessary to introduce an arbitrary scale μ, which is analogous to the renormalisation scale μ_R and factorises the problem into two terms, one governing the perturbative, short-distance physics and one term describing the non-perturbative, long-distance interactions. The factorisation scale is usually chosen to be the same as the renormalisation scale. The terms which absorb the incalculable effects in the model are cancelled when applying the renormalisation group equation and one is left with the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP)[19] equation, which governs the evolution of the parton distribution functions between different scales. So although it is not possible to calculate the distribution function $f_i\left(x, \mu^2\right)$ at any particular value of μ^2, as soon as it is measured at a certain scale, the distribution at another scale can be predicted. The DGLAP equations for quarks and gluons are:

$$\begin{aligned}
\mu^2 \frac{\partial q_i\left(x, \mu^2\right)}{\partial \mu^2} &= \frac{\alpha_s\left(\mu_R^2\right)}{2\pi} \int_x^1 \frac{dy}{y} \left[P_{q_i \to q_i g}(x/y) \cdot q_i\left(y, \mu^2\right) + P_{g \to qq}(x/y) \cdot g\left(y, \mu^2\right) \right] \quad (1.7) \\
\mu^2 \frac{\partial g\left(x, \mu^2\right)}{\partial \mu^2} &= \frac{\alpha_s\left(\mu_R^2\right)}{2\pi} \int_x^1 \frac{dy}{y} \left[\sum\nolimits_i P_{q_i \to q_i g}(x/y) \cdot q_i\left(y, \mu^2\right) + P_{g \to gg}(x/y) \cdot g\left(y, \mu^2\right) \right]
\end{aligned}$$

with the Altarelli-Parisi splitting functions $P_{a \to bc}(z)$. The splitting function

has the interpretation of describing the probability of finding the branching $a \to bc$, where z is the energy fraction taken by one of the daughter particles.

The DGLAP equation thereby shows how the distribution for a parton with momentum fraction x, which absorbs a probe, is given by the integral over y of the corresponding distribution for a quark with momentum fraction y, which radiated away via a gluon a fraction x/y of its momentum with probability" $\alpha_s(\mu^2)/2\pi P_{f_j \to f_i g}(x/y)$. The probability is high for large momentum fractions, which expresses how high-momentum quarks lose momentum by radiating gluons. This predicts that the distributions $f_i(x, \mu^2)$ get smaller at large x as μ^2 increases and larger at small x, while keeping the integral of the distribution over x constant. In addition, the radiated gluons produce more quark-pairs at small x. Therefore the nucleon can be imagined as having more and more constituents, all contributing to its total momentum, as the structure is probed on smaller distances, corresponding to larger scales μ.

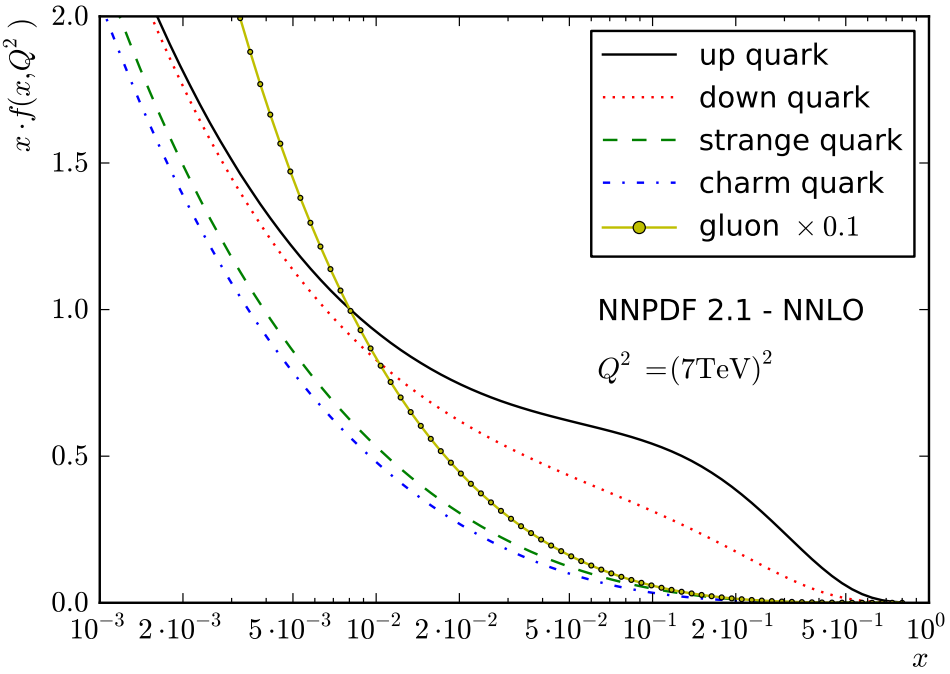


Figure 1.3: NNPDF2.1 NNLO parton distribution function for different quark flavours and gluons at $Q^2 = (7\,\text{TeV})^2$

1.3.1 Hadron-Hadron interactions

The mathematical description of the interaction between complex objects like hadrons are possible because of the previously discussed ability to decouple the long ranged hadron interactions from the short ranged parton interactions. This decoupling implies that the PDFs, which use experimental input from other experiments, like electron-proton colliders, can also be used in the description of proton-proton interactions. Using the PDFs $f_i\left(H_j\right)$ for the partons p_i of the involved hadrons H_j, it is possible to rewrite the cross section for the interaction of two hadrons H_1 and H_2 into some final state in the following way:

$$d\sigma\left(H_1H_2 \to \dots\right) = \int_0^1 dx_1 dx_2 \sum_{i,j} f_i\left(H_1, x_1, \mu_F^2\right) \cdot f_j\left(H_2, x_2, \mu_F^2\right) \cdot d\sigma\left(p_i p_j \to \dots\right)$$

This allows to determine the cross section $d\sigma\left(H_1H_2 \to \dots\right)$ for the hadron interaction from the cross sections $d\sigma\left(p_ip_j \to \dots\right)$ of all the partons p_i, p_j composing the hadrons.

1.4 Partons and Hadrons

Using the Standard Model, it is possible to describe the hard interaction between two colliding fundamental particles. Using PDFs it is also possible to describe complex hadron-hadron collisions. The results of such computations for QCD processes are a small number of partons. Yet in experiments a large number of colour neutral hadrons is observed. The Monte Carlo simulation which describes the transition from partons to hadrons can be factorised into two parts: the parton shower and the hadronization model.

1.4.1 Parton Shower

In electrodynamics accelerated electric charges radiate charge-less photons whereas in QCD accelerated colour charges radiate gluons, which carry colour and therefore also radiate themselves.

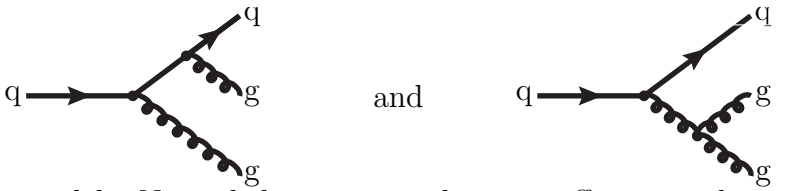
Two general approaches exist for the modelling of these perturbative corrections. A detailed treatment would consist in calculating the Feynman diagrams order by order. This takes into account the exact kinematics of the problem and the full interference and helicity structure. The problem is that such computations become increasingly difficult in higher orders and computationally prohibitive as the number of diagrams grows factorially.

The second option, used in Monte Carlo event generators, is to model the effect using the parton-shower approach. It is an approximative perturbative treatment of QCD dynamics and is applicable above a given transferred momentum q^2, with q^2 of the order of $1\,\mathrm{GeV}^2$. The following description is based on the lectures presented in [20].

The parton evolution is characterised by mostly soft and collinear emissions since such branchings are logarithmically enhanced. Solving QCD parton evolution equations, specifically the DGLAP QCD evolution equations in the leading-log approximation with a fixed cut-off regularisation leads to a model where the emissions are described by Markov processes with branchings into $a \to bc$, specifically for QCD: $q \to qg$, $g \to gg$ and $g \to q\bar{q}$.

Therefore, tree-level matrix elements for an n parton state can be approximated by a product of splitting functions $P_{a\to bc}(z)$. This product corresponds to a sequence of one-parton emissions from the initial state (e.g. a 2 parton state). z is the energy fraction taken by one of the daughter particles. The succession of these emissions can be ordered according to some resolution scale ρ. Common choices for this scale are in the Monte Carlo event generator Pythia the virtuality q^2 or the transverse momentum p_T. An angular ordered parton shower algorithm is used for example in the Monte Carlo program Herwig. From the splitting functions the Sudakov form factor $\Delta(\rho_n, \rho_{n+1})$ can be derived which describes the probability for having no emissions between two resolution scales ρ_n and ρ_{n+1}. This Sudakov form factor allows to compute the probability $P(\rho_n)$ for the n-th splitting at a given resolution scale ρ_n. With this information it is possible to simulate the parton shower with Monte Carlo methods.

Using the splitting probabilities means that coherence effects are not directly taken into account. Therefore, there is no interference between the two diagrams



in this model. Nevertheless, most coherence effects can be taken into account with the aforementioned angular ordering. It is possible to approximate the interference by assuming that the gluon is only emitted in an angular cone given by the previous emission[21]. Another way to include the coherence effects is to use a colour dipole description were the radiation is from colour-connected partons[22]. The Monte-Carlo program Sherpa applies the CKKW[23] method to provide a consistent matching between multi-particle matrix elements and parton shower.

The parton shower can be applied to both the incoming and outgoing partons of a given hard process, resulting in the initial- and final-state shower. The initial-state shower is less well understood theoretically, but can be modelled with a "backwards evolution" approach in which the parton shower that preceded the hard scattering is subsequently reconstructed.

1.4.2 Hadronization models

Quarks and gluons produced in hard scattering processes and the parton shower cannot exist as free particles due to QCD confinement. What can be measured in a detector is the result of a transformation process from coloured partons into a stream of colourless hadrons.

The process of building baryons or mesons from partons by pair production mechanisms or radiating off gluons is a non-perturbative phenomenon and therefore only phenomenological models exist to describe this process so far. Best known are the Lund String Fragmentation model and the Cluster Fragmentation model.

Lund String Fragmentation Model[4]

The Lund string model (described in [24] and summarised here) is based on the linear part of the QCD diquark potential[25] at long distances.

$$V(r) = -\frac{4}{3}\frac{\alpha_s(r)\,\hbar\,c}{r} + \kappa \cdot r$$

The colour field of a $q\bar{q}$ pair which is moving apart, increases linearly in strength with the string constant $\kappa \approx 1\,\mathrm{GeV/\,fm}$, transferring energy from the quarks to the gluon field. At long distances, the self interactions of the gluons collapse the field lines, creating a string-like field configuration. Gluons from the partonic final state are incorporated as kinks in the strings, carrying energy and momentum. The gluon self interaction is strong enough that the transverse size of the string is negligible. This motivates the description of the string as a massless, one-dimensional, relativistic object with only longitudinal excitations. These longitudinal excitations allow the transfer of energy between the two quarks and the field.

If enough energy is accumulated in the string to allow pair production, the colour field breaks up into another $q'\bar{q}'$ pair, resulting in the two colour-singlets $q\bar{q}'$ and $q'\bar{q}$. In order to generate the quark pair during break-up, the Lund model invokes the idea of quantum mechanical tunnelling. Hence,

[4]The Lund string model was developed at the University of Lund in Sweden.

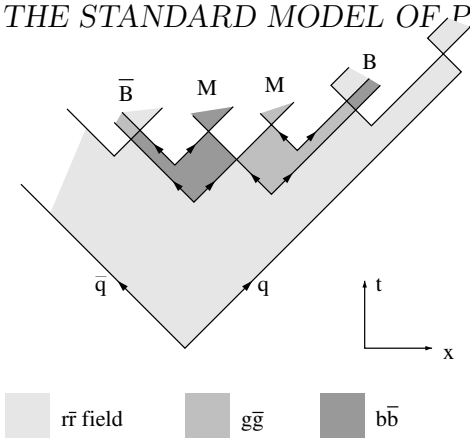


Figure 1.4: An example how two baryons and mesons are created in the popcorn model by independent pair production processes [26].

heavy quarks like charm, bottom and top quarks are not expected to be produced in the soft fragmentation.

Since the string is assumed to have no transverse excitations, the p_T acquired by the created quark is compensated by the corresponding anti-quark. If the invariant mass of either one of the remaining strings between $q\bar{q}'$ or $q'\bar{q}$ is large enough, further breaks may occur. This break-up process is assumed to proceed until only on-mass-shell hadrons remain. When the quarks from adjacent string breaks are combined to form the final state hadrons, it is necessary to invoke an algorithm to choose which hadrons are formed that way. Notably their spin and flavour composition must be assigned based on tuned models.

Meson production is easily understood in the Lund String model as a quark-antiquark pair connected by an unbroken string while baryon production is not so straightforward. There are two common models to describe baryon production in the context of the Lund String model[26]. In the diquark approach, a string can also be broken up into an antidiquark and a diquark which become constituents of an anti-baryon respectively a baryon.

The contending popcorn model is in better agreement with experimental results. In this model, as shown in figure 1.4, a diquark is produced in a stepwise manner and not as a single unit.

Once the Lund String model is fitted to e^+e^--collision data, there is little freedom to tune its parameters to other types of collisions. The model is physically well motivated and gives, after tuning, a good description of the observed data.

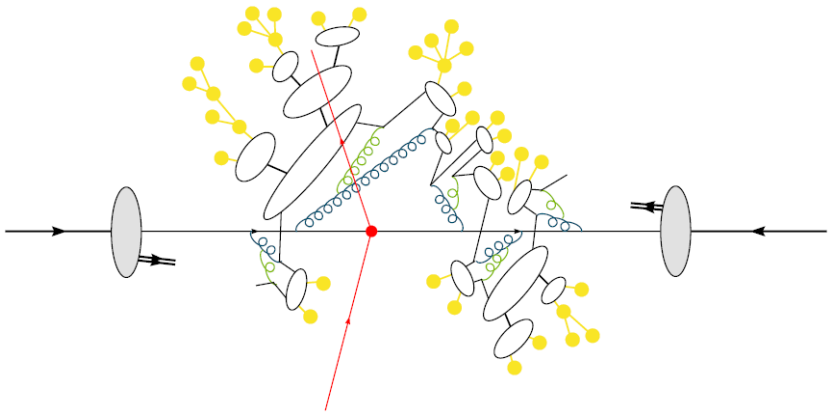


Figure 1.5: Sketch of the simulation of a pp collision using the cluster hadronisation model. The hard signal process produced a pair of leptons (red). After initial and final state parton showers (green), Gluons are split into quark pairs. Then colourless clusters (white) are formed. These fission into lighter clusters or into pairs of hadrons (yellow), which may decay further. (Source [29])

Cluster Hadronisation Model

The Cluster Hadronisation Model (described in [27] and summarised here) is based on the idea of preconfinement. Preconfinement is the property that quarks and gluons produced in the parton shower become organised in clusters of colour singlets.

The first step of the Cluster Hadronization Model is to take the partonic final state and split the gluons non-perturbatively into quark-antiquark pairs. In the Herwig++ implementation, the gluon decays isotropically into the possible quark flavours to separate the two colours carried by the gluon. In Sherpa, the decay of these gluons is performed in a dipole-inspired framework, where the colour partner accounts for the recoil of the massless gluon[28]. After the gluons are decayed, the event contains only colour connected quarks, diquarks and their antiparticles.

The colour singlets formed by the colour connected parton pairs are formed into low-mass clusters with the momentum given by the sum of the momenta of the constituents. These colour singlet clusters have a mass spectrum peaked at low masses ($m \sim 1\,\mathrm{GeV}$) and can be regarded as highly excited hadron resonances. They are subsequently split into hadrons according to phase space weight and spin. However, a small fraction of clusters is too heavy ($m > 4\,\mathrm{GeV}$) for this to be a reasonable approach. Therefore, these heavy clusters are anisotropically split into lighter clusters before they

decay in a process called cluster fission. Although only 15% of all clusters have to be split, the threshold for cluster fission is of great importance since 50% of all observed baryons are descendants of such clusters.

To finally decay a cluster of a given flavour $q_1\bar{q}_2$ into hadrons, a quark-antiquark $q\bar{q}$ or diquark-antidiquark $Q\bar{Q}$ pair is created and a pair of hadrons with flavours $q_1\bar{q}$ / $q_1\bar{Q}$ and $q\bar{q}_2$ / $Q\bar{q}_2$ is formed. The selection of the produced hadrons is based on flavour, spin and available phase space.

Chapter 2

The CMS Detector at the Large Hadron Collider

In particle physics, hadron colliders have a long tradition as discovery machines, which are used to explore physics at the highest energies. In the early 1930s, cyclotrons with energies of around 1 MeV and circumferences of around 30 cm were used to study fundamental nuclear interactions.

Over many decades, technological progress enabled the construction of increasingly sophisticated, larger and more powerful colliders. In 1983, the 7 km long Super Proton Synchrotron, located at CERN[1] in Geneva, recorded data which proved the existence of the W and Z bosons [30][31] by colliding protons at centre-of-mass energies of around 450 MeV. The Tevatron at Fermilab in Batavia/Illinois measured in 1995 for the first time the mass of the top quark [32]. Today, the Large Hadron Collider [33] (LHC) and the connected experiments at CERN represent the pinnacle in the field of high energy particle collider and detector design. A major success of the LHC was the discovery of a new boson with a mass of 125 GeV in July 2012.

The physics program at the LHC stretches across the whole field of High Energy Physics: Measurements of Standard Model parameters at high energy scales as well as searches for the Higgs boson, Supersymmetry, extra dimensions or other new phenomena. For this purpose, the LHC provides four interaction points where proton-proton or collisions with lead ions take place. At each of these four points, a complex detector identifies the particles coming from the collisions and measures their kinematic properties.

The heavy ion experiment ALICE [34] is dedicated to the investigation of strongly interacting hadronic matter as it is produced in collisions of lead

[1] European Organisation for Nuclear Research established by the "Conseil Européen pour la Recherche Nucléaire"

ions. At high energy densities, hints for the formation of a new phase of matter, the quark-gluon plasma, is observed in these interactions[35]. The LHCb [36] collaboration performs precision measurements on CP violation and rare decays in order to reveal possible indications for new physics and to further constrain the Standard Model. CMS [37] and ATLAS [38] are two general purpose experiments with complementary detector technologies. Located in their vicinity and sharing the interaction point are TOTEM [39] and LHCf [40], which are dedicated to cross section measurements and forward physics.

2.1 The Large Hadron Collider

2.1.1 Introduction

The Large Hadron Collider is a superconducting hadron accelerator and collider. It is installed in the 26.7 km long tunnel which previously housed the Large Electron Positron (LEP) collider at CERN. The tunnel lies between 45 m and 170 m below the surface on a slightly inclined plane across the French-Swiss border near Geneva.

Between 1989 and 2000, LEP and its detectors allowed in particular the measurement of the W and Z boson properties with unprecedented accuracy[41] by colliding electron with positrons with centre-of-mass energies of up to 209 GeV. The limiting factor for LEP was the energy loss due to synchrotron radiation which increases proportionally to the fourth power of the inverse mass of the particles. Using particles with a higher mass this synchrotron radiation problem can be circumvented. Therefore by using hadron beams, the LHC will be able to reach much higher energies than LEP while utilising the same tunnel.

The LHC project proposal was approved in December 1994 and achieved its first beam in September 2008. After a serious problem with the connection between the superconducting magnets[42], the first collisions were recorded in November 2009. Since February 2010, the LHC runs at a centre of mass energy of 7000 GeV. The energy was increased to 8000 GeV for the 2012 data-taking. After this data-taking period, some necessary upgrades will be performed to meet the design goals of providing proton and lead ion collisions with a centre-of-mass energy of the nuclei of up to 14 TeV and 5.52 TeV respectively.

2.1.2 LHC Accelerator Chain

The hadron beams are accelerated by several existing facilities before being injected into the LHC. An overview of the pre-acceleration chain is shown in figure 2.1. For a proton run the linear collider LINAC 2 accelerates the particles to $E_p = 50\,\text{MeV}$ and transfers them into a chain of synchrotrons. After the Proton Synchrotron Booster ($E_p = 1.4\,\text{GeV}$), the Proton Synchrotron ($E_p = 25\,\text{GeV}$) and the Super Proton Synchrotron ($E_p = 450\,\text{GeV}$) the protons are injected into the Large Hadron Collider. To fill all bunches of the beam takes about 260 s for each beam direction. Once the filling procedure is finished, the protons are accelerated for about 15 min to their final energy of currently $E_p = 4000\,\text{GeV}$[43].

To perform this acceleration, the LHC uses eight radio-frequency cavities per beam, each delivering 2 MV. Particle which are exactly synchronised with the frequency of the cavities are called synchronous particles. Other particles in the accelerator oscillate longitudinally around the synchronous particles. This means that the particles in the beams are not uniformly spread around the circumference, but rather tightly packed around the synchronous particles in so-called bunches. The LHC can contain up to 2808 such bunches, consisting of roughly 10^{11} protons and separated by a bunch spacing of 25 ns

These particles bunches are kept on a circular path along the ring by 1232 dipole magnets, designed to deliver a magnetic field strength of 8.33 T. In order to focus the beams and counteract other interactions, 858 quadrupole magnets and a large number of sextupoles, octupoles and decapoles are used. With complex beam optics, the two particle beams are further focused and made to intersect each other at four interaction points.

2.1.3 Luminosity and Machine Parameters

The cross section σ for a physical process is proportional to the number of reactions per second, which occur at an interaction point of a particle collider:

$$\frac{\partial N}{\partial t} = \mathcal{L}\,\sigma$$

The factor of proportionality is called the machine luminosity $\mathcal{L}$ and can be calculated with the specific machine and beam parameters. Assuming a Gaussian beam distribution, the luminosity can be written as:

$$\mathcal{L} = \frac{N_b^2\, n_b\, f_{\text{rev}}\, \gamma\, F}{4\,\pi\,\varepsilon_n \beta^*}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency and γ the relativistic gamma factor.

		Achieved	Design	Unit
Energy per nucleon	E	3.5	7	TeV
Dipole field	B		8.33	T
Instantaneous luminosity	$\mathcal{L}$	3.55×10^{33}	10^{34}	$\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$
Bunch separation	t_b	25	25	ns
Number of bunches	n_b	1380	2808	
Number of particles per bunch	N_b	1.5×10^{11}	1.15×10^{11}	
Norm. transverse beam emittance	ε_n	2	3.75	μm
β-value at interaction point	β^*	0.55	0.55	m
Revolution frequency	f_{rev}	11.2	11.2	kHz
Gamma factor	γ	3730	7460	
Luminosity lifetime	τ_L	17	15	h
Bunch crossing frequency	f_b	40	40	MHz
Number of collisions per crossing	n_c	≈ 25	≈ 20	

Table 2.1: Comparison of achieved and design parameters of the LHC machine [33, 44, 45, 46, 47].

ε_n is the normalised transverse beam emittance - a measure for the phase space area associated with the two transverse degrees of freedom. The beam properties at the interaction point are described by the betatron function β^* and the geometric luminosity reduction factor F.

During a physics run, the luminosity of the LHC decreases mainly because of the particle loss during collisions, but also due to deteriorating beam properties. This decay is characterised by the luminosity lifetime τ_L. Integrating the luminosity over one run with length T_{run} yields

$$L_{\mathrm{int}} = \mathcal{L}_0 \tau_L \left[1 - \exp\left(-\frac{T_{\mathrm{run}}}{\tau_L}\right)\right]$$

After a short start-up phase with 450 GeV and 1.38 TeV beam energy, the LHC delivered $47\,\mathrm{pb}^{-1}$ of integrated luminosity with 3.5 TeV beams to CMS in 2010. With the experience gained during this run, it was possible to collect $5.7\,\mathrm{fb}^{-1}$ proton-proton collision data in the following year 2011[48]. Using reasonable run times one obtains an integrated luminosity L_{int} of about $80\,\mathrm{fb}^{-1}$ to $120\,\mathrm{fb}^{-1}$ per year for the CMS experiment when running with design parameters.

The luminosity is continuously measured by each experiment[49]. CMS uses two methods to determine the luminosity. The first is based on the occupancy of calorimeter cells of the forward hadronic calorimeter (HF method).

The second method is based on counting the number of reconstructed vertices in events with low activity. Both approaches allow a relative luminosity measurement. The absolute luminosity normalisation can be performed using either a comparison with simulated events or by performing a so-called van-der-Meer scan[50]. Since the former method has large systematic uncertainties, CMS is using the van-der-Meer scans performed in separate fills for the official luminosity measurement. It is based on the measurement of the relative interaction rates as a function of the transverse beam separation, which is varied during each such scan. This allows to determine the effective beam size and the maximum achievable collision rate.

Phase	Energy E_p	Instantaneous luminosity $\mathcal{L}$
2010	3.5 TeV	up to $2 \times 10^{30}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$
2011	3.5 TeV	up to $3.55 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$
2012	4 TeV	more than $6.25 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$
Design	7 TeV	$1.05 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$

Table 2.2: Development of the LHC beam energy and instantaneous luminosity for pp collisions over the years (Source: [48, 44]).

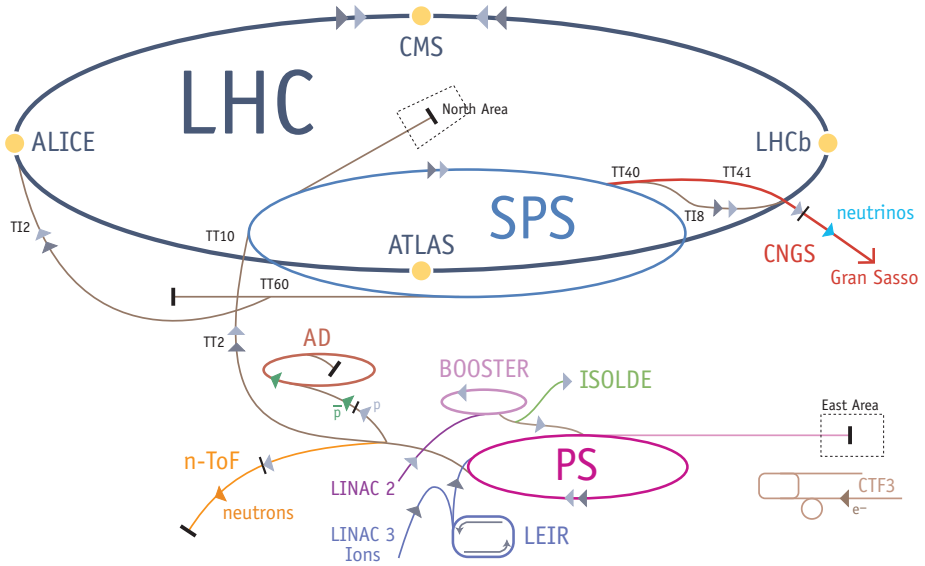


Figure 2.1: Overview of the accelerator chain. The hadron beams are accelerated by several successive facilities to the LHC injection energy of 450 GeV before being accelerated in the LHC to 7 TeV[43].

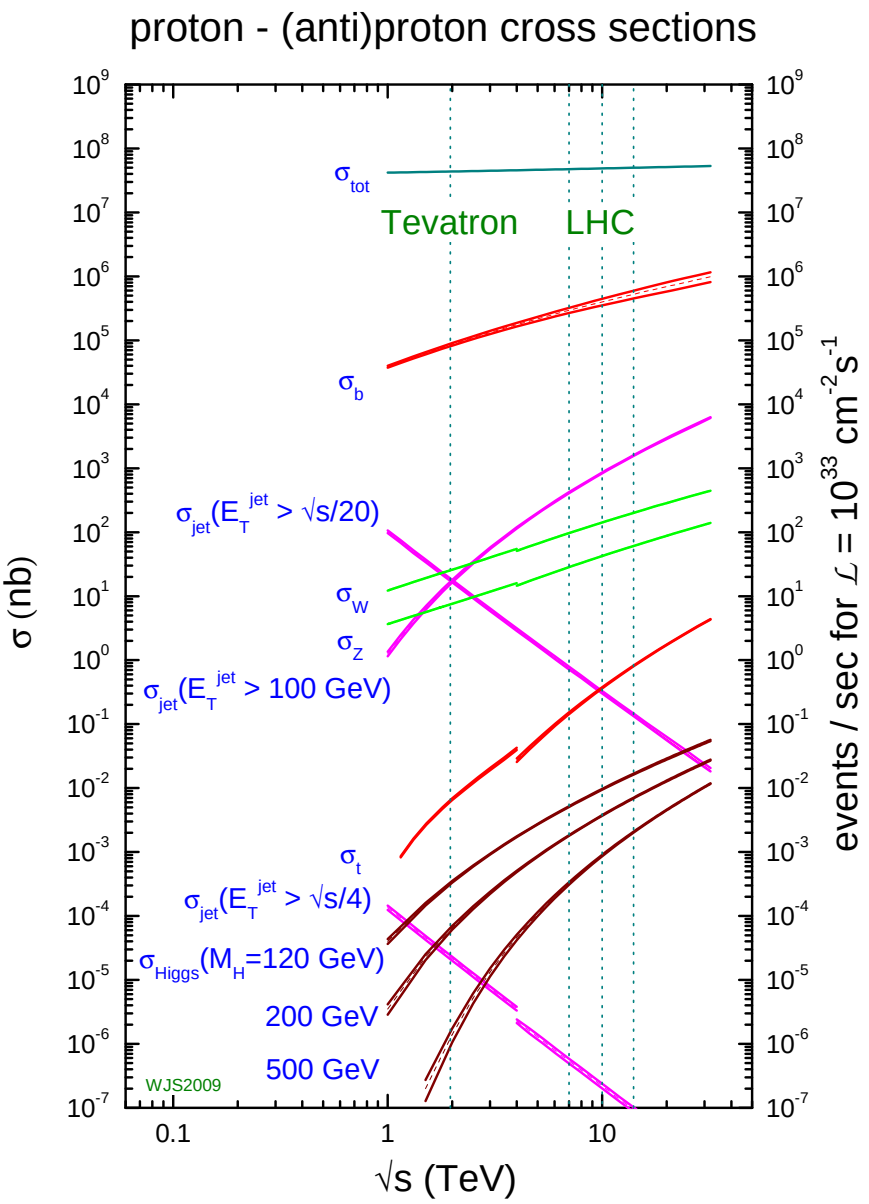
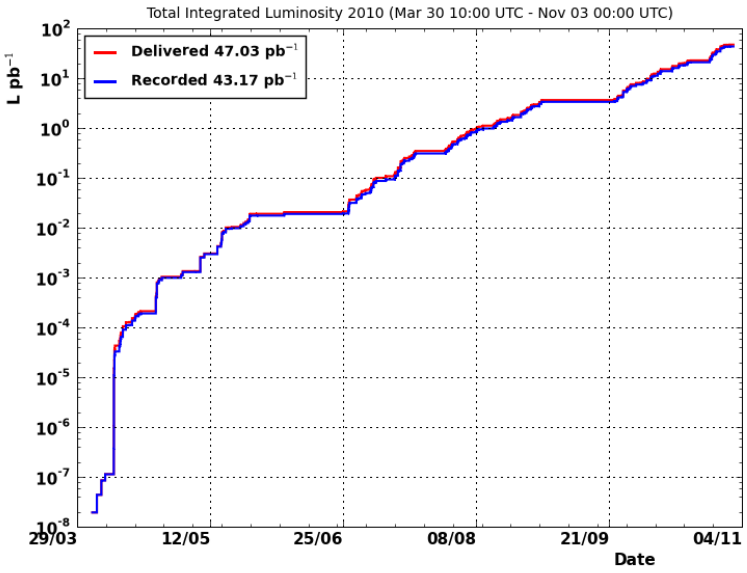
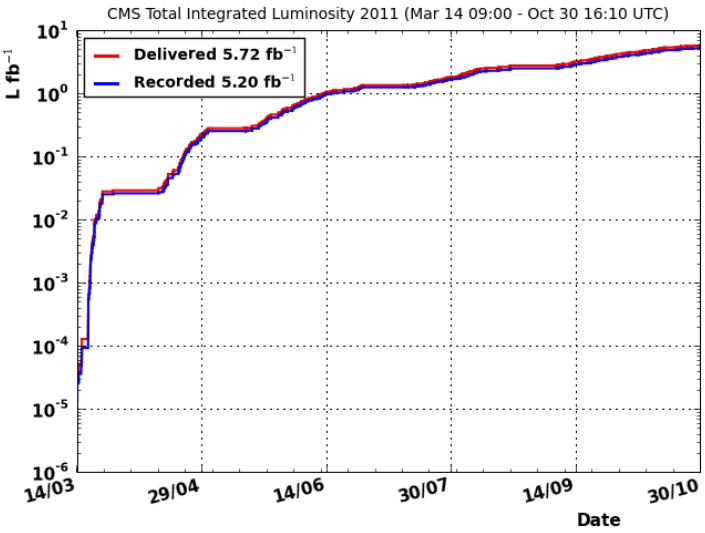


Figure 2.2: Overview of the cross-sections for different physics processes. At the centre-of-mass energies of the LHC, jet production is the dominating process. (Source: [16])



(a) 2010



(b) 2011

Figure 2.3: Integrated luminosity delivered by the LHC at P5 and recorded by CMS in 2010 and 2011. Flat regions correspond to technical stops of the accelerator (Source: [51]).

2.2 The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) experiment is located at Point 5 of the LHC near Cessy in France, opposite to the CERN main campus at Meyrin. CMS is a general purpose detector built to run at the highest luminosity at the LHC. The design of CMS is driven by having a high-performance muon system, an excellent electromagnetic calorimeter, a high quality central tracking and a hermetic hadron calorimeter. The main volume of the CMS detector is a multi-layered cylinder, 21.6 m long and 14.6 m in diameter, weighing more than $12,500$ t. The innermost layer is a silicon-based particle tracker, surrounded by electromagnetic and hadronic calorimeters for measuring particle energies. They are fit inside a central superconducting solenoid magnet. The outermost layer of the detector consists of the muon chambers and a small part of the hadronic calorimeter. In contrast to some other experiments at the LHC, the compact nature of the detector allowed the construction at ground level. The subsequent sections summarise the detailed description of the detector given in [37].

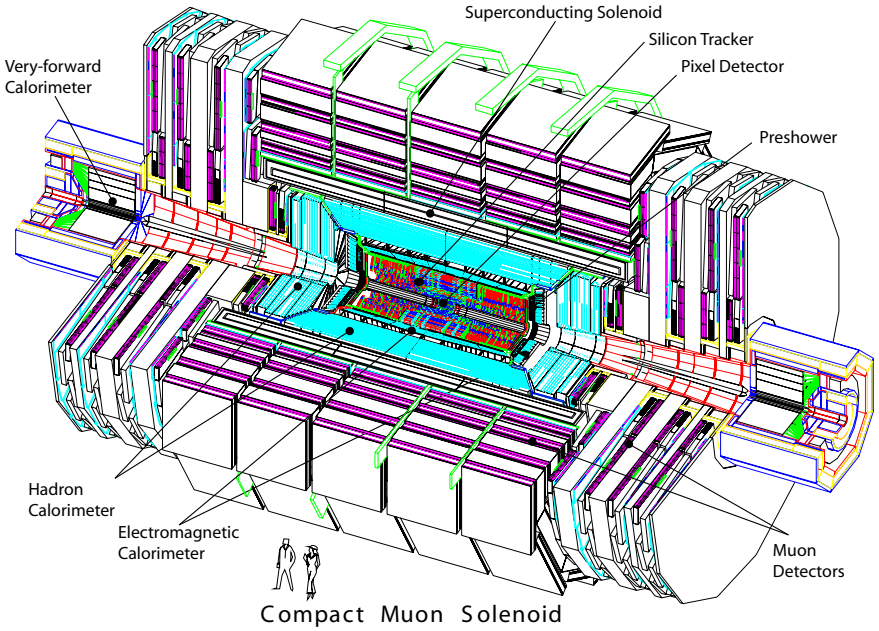


Figure 2.4: Schematic view of the CMS detector and its components[37]

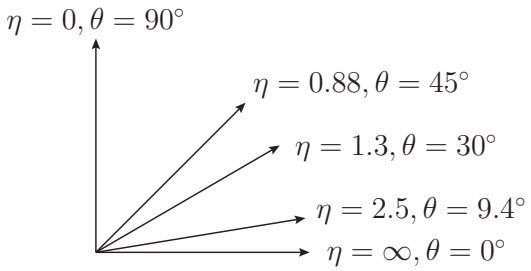


Figure 2.5: Relationship between the polar angle and pseudo-rapidity.

2.2.1 Conventions and Definitions

The coordinate system adopted by CMS has the origin centred at the nominal collision point inside the experiment. The x-axis points radially inward towards the centre of the LHC, the y-axis is pointing vertically upward and the z-axis points along the beam direction towards the Jura mountains, building a right-handed coordinate system. The azimuthal angle ϕ is measured from the x-axis in the x-y plane and the radial coordinate in this plane r gives the radial distance from the beam pipe. The polar angle θ is measured from the z axis with the +z-direction corresponding to $\theta = 0$ and the -z-direction corresponding to $\theta = \pi$.

Instead of the polar angle, the rapidity y or pseudo-rapidity η are commonly used. The differential particle flux in these two variables is approximately constant at hadron colliders[52], making them a good choice as phase-space partition variables. The rapidity has the useful property that it is invariant under longitudinal boosts in the z-direction and is defined as

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

Using the pseudo-rapidity η, which is solely defined by spatial coordinates, it is possible to avoid an energy dependency:

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

For massless particles rapidity and pseudo-rapidity are equivalent. Information about energy and momentum is often given by its transversal components E_T and p_T which are related to the Cartesian coordinates via:

$$p_T = \sqrt{p_x^2 + p_y^2}, \qquad E_T = E \frac{p_T}{\sqrt{p_T^2 + p_z^2}}$$

The angular distance between two objects coming from the origin is described by

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

2.2.2 Inner Tracking System

The inner tracking system (figure 2.6) of CMS is designed to provide a precise measurement of charged particle trajectories emerging from LHC collisions. It surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m, covering a pseudo-rapidity range of up to $|\eta| < 2.5$. The tracking system consists of two main subsystems, the pixel detector with three barrel layers at radii of 4.4 cm, 7.3 cm and 10.2 cm and a silicon strip tracker with 10 barrel detection layers extending outwards up to a radius of 1.1 m. Each subsystem is completed by end-caps which consist of 2 disks in the pixel detector and 3 plus 9 disks in the strip tracker on each side of the barrel.

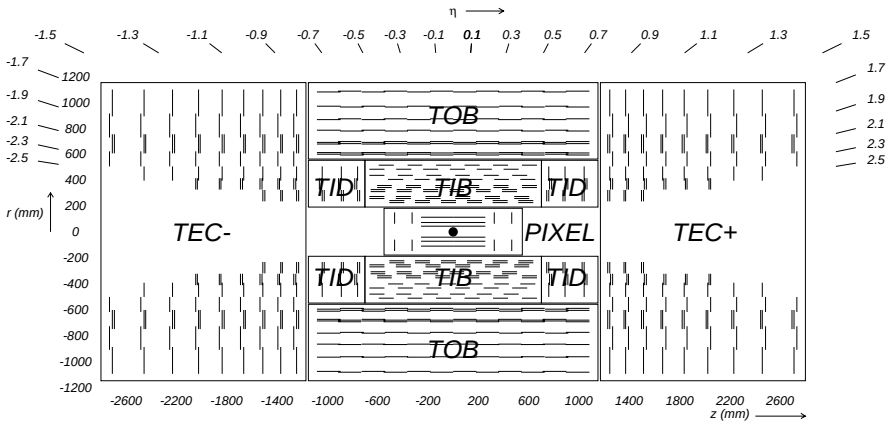


Figure 2.6: Lateral cut through the Inner Tracking System (Source: [37])

With about 200 m^2 of active silicon area the CMS tracker is the largest silicon tracker ever built. In addition to the issues of producing such a large detector, the operating conditions for the tracking system are very challenging. During each LHC bunch crossing at design luminosity about 1000 particles are hitting the tracker on average. To resolve so many particles the innermost pixel detector has a small cell size of $100 \times 150 \mu m^2$ in r-ϕ and z leading to an occupancy in the order of 10^{-4} per pixel and bunch crossing. The high spatial resolution achieved using these approximately 66 million pixels allows

the precise measurement of secondary vertices and impact parameters which is important for the efficient identification of long lived particles.

Further away from the interaction point, the reduced particle flux allows the usage of cost-saving strip detectors with a typical cell size of $10\,\text{cm} \times 80\mu\text{m}$ in the intermediary range and cell sizes of $25\,\text{cm} \times 180\,\text{mm}$ in the outer region of the tracker. To get position information along the long side of the strips some layers use double-sided modules with a stereo angle of 100 mrad.

Measurement of particle tracks

To measure the tracks of charged particles, multiple layers of silicon detectors are depleted of free charges by applying a high voltage to them. Any charged particle travelling through depleted silicon leaves an ionisation trail causing a small current. This current can be measured and is registered as a hit in the detector cell. With sophisticated pattern recognition algorithms it is possible to reconstruct a set of probable paths for the particle through the tracker. The very homogeneous 3.8 T strong magnetic field inside of the CMS solenoid causes even high p_T particles to leave tracks with a small curvature. On the other hand, low p_T particles can even be forced into helical tracks with a small-radius. Therefore it is reliably possible to measure the particle momentum and charge from the curvature of the tracks.

2.2.3 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) of CMS is a hermetic calorimeter made of lead tungstate ($PbWO_4$) crystals as scintillating material. This material was chosen because it is very dense ($8.28\,\text{g}\,\text{cm}^{-3}$) and has a short radiation length ($X_0 = 0.89\,\text{cm}$). These properties allow for a compact calorimeter design where electrons, positrons and photons deposit their energy within a very small depth. The Molière radius, the radius of the cylinder containing 90% of the energy deposition, for this material is rather small (2.19 cm). This allows to measure the energy with a fine granularity. $PbWO_4$ is also radiation hard and fast scintillating (80% of the light is emitted within $t_b = 25\,\text{ns}$), making it an ideal scintillator material for the conditions at the LHC. The crystals emit violet scintillation light which is measured by photo-detectors. The relatively low light yield (50 Photons/ MeV) requires the use of photo-detectors with intrinsic gain. The electromagnetic calorimeter consists of three subsystems, the barrel, the end-cap and the pre-shower detector as shown in figure 2.7.

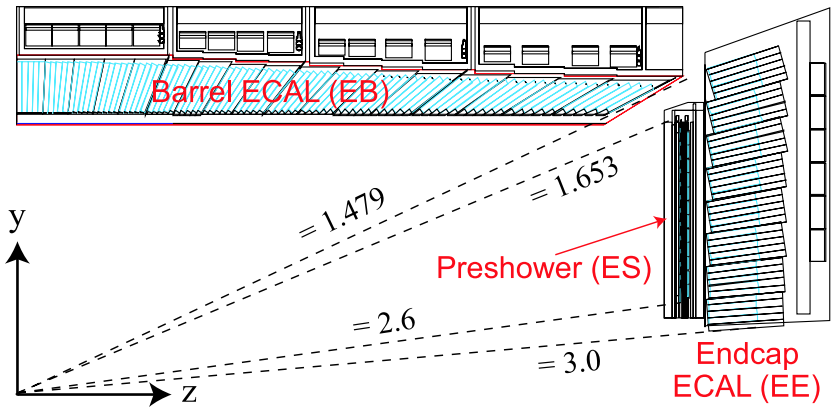


Figure 2.7: Cross section through the Electromagnetic Calorimeter (Source: [37])

Electromagnetic Calorimeter Barrel The barrel (EB) covers the pseudo-rapidity range $|\eta| < 1.479$ with 61200 crystals. They are positioned and shaped in such a way that the ECAL has a homogeneous crystal distribution in η. The crystal cross section corresponds to approximately 0.0174×0.0174 in the $\eta - \phi$ plane and they have a length of 230 mm corresponding to $25.8\, X_0$. Crystals for each half-barrel are grouped into 18 super-modules each covering 20° in ϕ. Each super-module comprises four modules with 500 crystals in the first module and 400 crystals in the remaining three. The crystals are grouped in arrays of 2×5 crystals which together with the integrated avalanche photo-diodes form a sub-module.

Electromagnetic Calorimeter End-caps The end-caps (EE) cover the rapidity range $1.479 < |\eta| < 3.0$ and are each composed of 7324 crystals. The crystals in the end-cap are arranged in an x-y-grid, resulting in an η dependent crystal cross section ranging from 0.0174×0.0174 to 0.05×0.05 in the $\eta - \phi$ plane. For the end-cap radiation resistant vacuum photo-triodes are used as photo-detectors.

Electromagnetic Pre-shower Detector Most of the end-cap surface is covered by the pre-shower detector (ES). It consists of two orthogonal planes of silicon strip sensors interleaved with lead. Its principal aim is to improve the neutral pion and photon discrimination in the end-caps within $1.653 < |\eta| < 2.6$. It also improves the position determination of electrons, positrons and photons with high granularity.

Details of electromagnetic energy deposition

When a photon, electron or positron enters the dense calorimeter material, an electromagnetic shower develops due to bremsstrahlung and electron-positron pair production. Via Compton scattering and the photo-electric effect, the incident particle as well as the shower particles, deposit their energy in the calorimeter material. The deposited energy causes the excitation of atomic states, which return to the ground state by photon emission. The number of emitted photons is proportional to the fraction of the energy that has been deposited in the crystal and this scintillation light can be measured by photo detectors. To reconstruct the whole shower different algorithms are available to find clusters of crystals belonging to the shower. The incident particle energy is estimated as the sum of the energies measured by each of the crystals belonging to the cluster.

The energy resolution of the ECAL has been determined with electron test beams with energies ranging from 20 GeV to 250 GeV[53]. The resolution can be parametrised with the NSC resolution formula:

$$\left(\frac{\sigma_E}{E}\right)^2 = \frac{N^2}{E^2} + \frac{S^2}{E} + C^2 \tag{2.1}$$

with noise (N), stochastic (S) and constant terms. The study fitted the measured resolution and found for the ECAL the following values: $N = 0.0415\,\mathrm{GeV}, S = 0.028\sqrt{\mathrm{GeV}}$ and $C = 0.003$.

2.2.4 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) plays an important role in the identification and measurement of particles by measuring the energy and direction of hadron jets and together with the ECAL the missing transverse energy flow in an event. The size of missing E_T in an event can be inferred by using all measured transverse energy components, since energy and momentum conservation dictates that the sum of these components must vanish. The residual part of this balancing is the missing E_T. Large values of missing E_T are a crucial signature in many searches for new physics at the LHC. For good missing E_T resolution, a hermetic calorimetry coverage is required.

The HCAL consists of four parts: the barrel, end-cap, forward and outer calorimeter, which is shown schematically in figure 2.8.

Hadron Barrel Calorimeter The Hadron Barrel Calorimeter (HB) is the innermost HCAL subsystem and covers the pseudo-rapidity range of $|\eta| < 1.305$. It is divided into 36 wedges made of 14 flat brass absorber plates

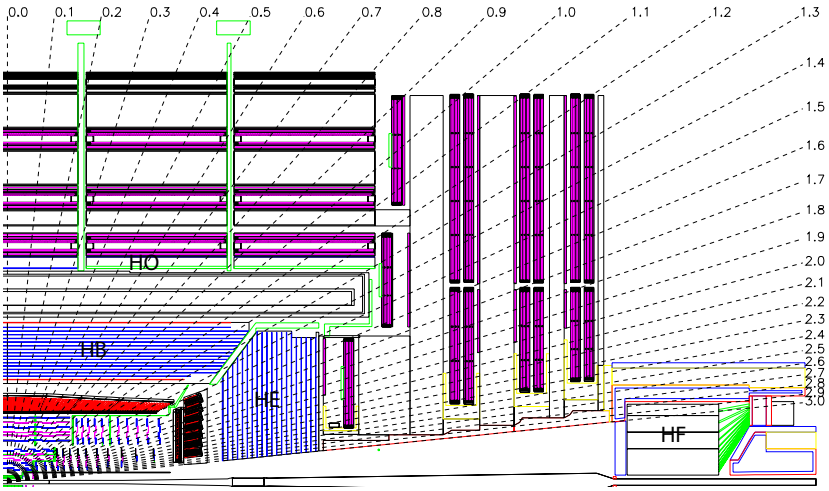


Figure 2.8: Schematic overview of the detector in the r-z plane. The four subsystems of the Hadronic Calorimeter, the Hadron Barrel (HB), the Hadron End-cap (HE), Hadron Outer (HO) and Hadron Forward Calorimeter (HF) are highlighted. The dashed lines indicate the pseudo-rapidity. (Source: [37])

(nuclear interaction length $\lambda = 16.4\,\text{cm}$), each segmented into four ϕ sectors. The front and back plates of each wedge are made of steel for structural reasons. To accomplish a high hermeticity, the plates are put together in a staggered way, leaving no dead material for the whole radial extent of the wedges. Together the absorbers have a thickness ranging from $5.39\,\lambda$ at $\eta = 0$ to $10.3\,\lambda$ at $|\eta| = 1.3$. The wedges themselves are aligned parallel to the beam axis and are forming the two half barrels (HB+ and HB-) of the HB. Between the absorber plates lies the plastic scintillator, segmented into 16 η sectors. This geometry allows for a resolution of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The 16th HB tower located at $|\eta| = 1.305 - 1.392$ of the HB overlaps with the end-cap and is not counted into the coverage.

Hadron End-cap Calorimeter The HCAL end-caps (HE) cover a large pseudo-rapidity range of $1.305 < |\eta| < 3.0$. A big engineering challenge was the requirement to build the heavy structure using non-magnetic materials, since the HE is placed at the end of the solenoid and must avoid distorting the magnetic field. The general structure of the HB closely follows the material and design choices of the HB. The cartridge brass absorber material is interleaved with plastic scintillators which are connected to photo-detectors by wavelength shifting fibres. The segmentation of the scintillators leads to a granularity of 0.087×0.087 in the $\eta - \phi$ plane for $|\eta| < 1.6$ and 0.17×0.17

for $|\eta| \geq 1.6$.

Hadron Outer Calorimeter The HO is placed outside of the magnet and utilises the solenoid coil as additional absorber material. Located in the barrel region the HO covers pseudo-rapidities of $|\eta| < 1.262$. The tiles of the HO are roughly aligned to the HB with a granularity of 0.087×0.087 in η and ϕ. Especially in the central pseudo-rapidity region the HO enables to recover shower leakage caused by the short interaction length. However high noise levels currently prevent the use of this sub-detector in the general event reconstruction.

Hadron Forward Calorimeter The HF calorimeter is a hadronic calorimeter composed of iron absorber with quartz fibres as active medium. The Cerenkov light generated by the relativistic components of the shower is detected by an array of photo-multipliers. By measuring Cerenkov light, the HF is mostly sensitive to the electromagnetic component of the shower. The construction of the calorimeter, which covers the pseudo-rapidity interval $2.8 < |\eta| < 5.2$ was mainly driven by the requirements of a high radiation environment. The main physics goal of the HF calorimeter is to tag high energy jets generated by vector-boson fusion events.

The Interaction of Hadrons with Matter

On entering matter, hadrons deposit their energy by causing a hadronic shower in the material. The hadronic showering process is dominated by a succession of inelastic hadron interactions, causing a large number of secondary particles. These secondary particles are mostly pions and nucleons and are produced with large transverse momentum. Consequently, hadronic showers spread more laterally than electromagnetic showers. The longitudinal development of the hadronic multiplication process is measured using the nuclear interaction length λ.

There is also an electromagnetic component present in hadronic showers due to the relatively frequent production of π^0. The energy measurement of a hadronic shower is based on the same principle as for the electromagnetic shower. The shower develops until a certain energy threshold is reached after which the energy is deposited by ionisation and low-energy hadronic activity.

The resolution for the HCAL is given in terms of formula 2.1 and was determined for the HCAL subsystems separately[54]. The HB, HO and HE resolution is described by $N = 0, S = 0.847\sqrt{\text{GeV}}$ and $C = 0.074$. For the HF resolution, the measurement yielded $N = 0, S = 1.98\sqrt{\text{GeV}}$ and $C = 0.09$.

2.2.5 Superconducting Solenoid

The superconducting magnet for CMS is designed to produce a 4 T field in a free bore of 6 m diameter and 12.5 m length with a stored energy of 2.6 GJ at full current. The high magnetic flux is necessary for a high momentum resolution in the tracker and muon system and is returned through a 10,000 t yoke. The distinctive feature of the 220 t magnet is the 4-layer winding made from an aluminium-stabilised reinforced NbTi conductor. The solenoid is operated at a temperature of 4 K, below the temperature were the NbTi conductor becomes superconducting. The magnet will be operated at 3.8 T for the foreseeable future.

2.2.6 Muon System

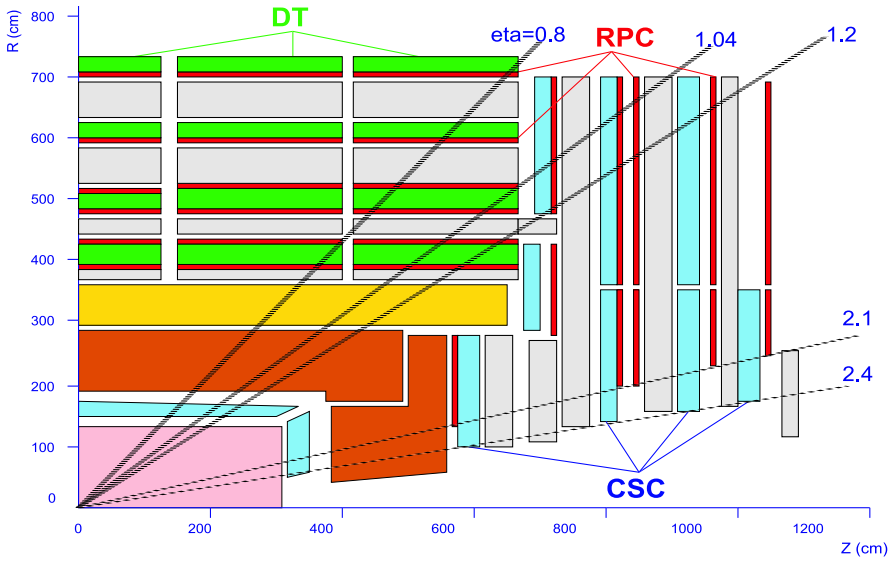


Figure 2.9: This schematic view of the detector shows the muon chambers, with drift tubes (DT), resistive plate chambers (RPC) and cathode strip chambers (CSC). The iron return yoke components are shown in grey, the superconducting solenoid is shown in yellow (Source: [37])

As is implied by the experiment's middle name, the detection of muons is of central importance to CMS. The muon detector chambers cover the pseudo-rapidity range of $|\eta| < 2.4$ and lie in the outermost region of the detector (see figure 2.9). Due to their minimal interaction with matter, muons deposit only a small amount of energy while traversing the many layers of detector material and the solenoid between the interaction point

and the muon system. All other detectable particles from the interaction point are stopped in one of the inner detector systems.

The muon system uses three different technologies for muon identification: drift tubes (DT) in the barrel region ($|\eta| < 1.2$), cathode strip chambers (CSC) in the end-cap region ($1.2 < |\eta| < 2.4$) and resistive plate chambers (RPC) in both the barrel and the end-caps. The DT and CSC detectors are used to obtain a precise measurement of the position of the muons and thus the momentum of the muons, whereas the RPCs are dedicated to provide fast information for the Level-1 trigger. The muon identification and momentum information collected by the muon system can be combined with tracker information to yield well reconstructed muons with excellent momentum resolution.

2.2.7 Forward Detectors

Centauro And Strange Object Research

The Centauro[2] And Strange Object Research (also abbreviated as CASTOR) detector is a quartz-tungsten sampling calorimeter similar to the HF, designed for the very forward region. It is installed 14.4 m from the interaction point covering $5.2 < |\eta| < 6.6$.

CASTOR is a Cerenkov light based calorimeter like the HF but uses tungsten as absorber material instead of iron. CASTOR is useful to the study of the very forward, baryon-rich region in heavy ion collisions.

Zero Degree Calorimeter

The combined zero degree calorimeter (ZDC) is designed as a combination of sampling quartz-tungsten electromagnetic and hadronic calorimeters. Two identical ZDC will be located in the tunnel between the beam pipes 140 m away from each side of CMS. They will detect very forward photons and neutrons with $|\eta| > 8.3$. ZDC information can be used for a variety of physics measurements as well as improving the collision centrality determination in heavy-ion collisions.

[2]Centauro events are highly asymmetric events observed in cosmic-ray detectors[55][56].

2.3 Data Acquisition and Trigger

Under normal conditions, the LHC runs with a beam crossing frequency of 25 ns, resulting in 40 million bunch crossings per second. Since there are usually more than 20 collisions per bunch crossing, depending on the beam parameters, the Data Acquisition (DAQ) system has to read in a large amount of data within the available, very small time window. Since it is not possible to read out and accumulate this large amount of data from all sub-detectors and store it permanently with current hardware, only a small subset of all collisions can be preserved. Therefore it is necessary to decide in a very fast and efficient way whether collision data should be discarded or kept for processing. CMS solves this problem by using two trigger levels to reduce the data by a factor of 10^5, the Level 1 trigger (L1) and the High Level Trigger (HLT). However a positive trigger decision by the L1 or HLT trigger does not necessarily mean that the event is forwarded to the HLT or storage system. Many trigger paths apply a so called trigger prescale. This means that only a predefined fraction of events with a positive trigger decision are passed on to the next step in the processing chain. This data reduction through prescales is necessary to record a representative sample of events for trigger paths with loose requirements without over-straining the storage systems. A thorough description of the CMS DAQ system can be found in [57, 58, 37].

2.3.1 Level 1 Trigger

The L1 trigger is based on custom-designed hardware, which can work with a trigger rate of around 100 kHz. This rate is mostly limited by the speed of the detector readout electronics. Wherever possible, the dedicated hardware is programmable in order to enable adaptations to the event selection. The L1 trigger combines the detector data in local, regional and global components in order of increasing regional extend, to extract the final L1 trigger decision. This allows for example to trigger on high energy deposition in a single calorimeter tower, muon tracks from regional hit patterns or large transverse energy. The structure of the L1 trigger is shown in 2.10.

2.3.2 High Level Trigger and Data Acquisition

All collision data, which pass the L1 trigger are harvested by the DAQ system for processing by the HLT filter farm located near the detector at Point 5. The DAQ system uses a high bandwidth network to collate the data from the more than 600 data sources of the detector. This event builder network is

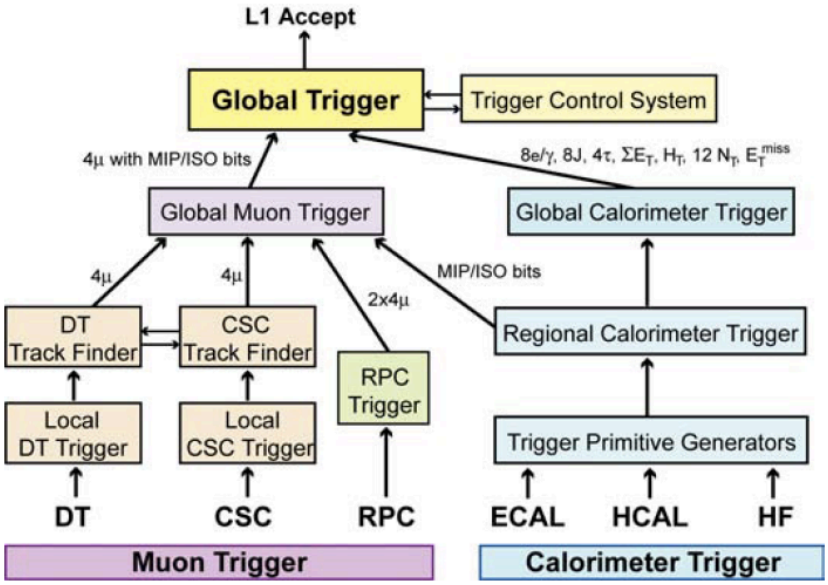


Figure 2.10: Overview of the L1 Trigger system. The system consists of two main parts, the Muon and Calorimeter Trigger. The Muon Trigger combines the information from the DT, CSC and RPC sub-detectors. Calorimeter Trigger decisions are based on ECAL and HCAL data. Both Trigger systems have several data collection stages and combine their results in the Global Trigger to form the L1 trigger decision. (Source: [37, 57])

split into 8 independent parts called slices, which assemble the detector data fragments into full events. The event data is then processed by a computing farm consisting of $\approx$ 8000 CPU cores, which runs a light-weight form of the CMS reconstruction software on the incoming data. The reconstructed data are then used to determine the HLT trigger decision. An event with a positive trigger decision is forwarded to the Storage Manager system, which writes the event data into files and transfers them to the T0 located on the main CERN campus for further processing. A schematic overview of the system is given in figure 2.11.

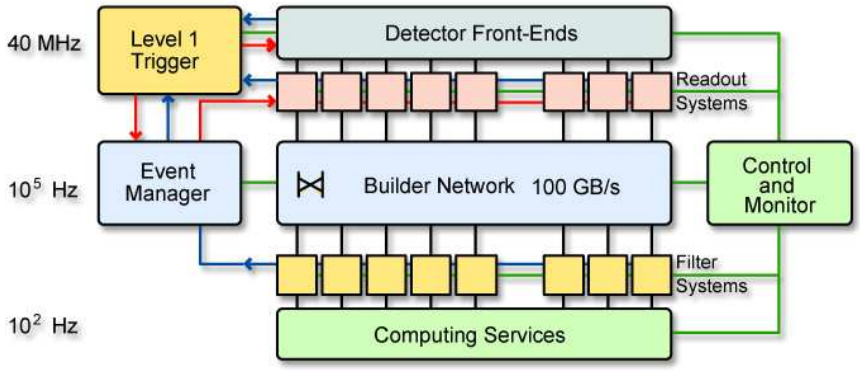


Figure 2.11: Overview of the HLT and DAQ system. The L1 trigger and detector devices (frontends) are working with trigger / readout rates of 40 MHz. The L1 trigger reduces this to the HLT trigger rate, at which the sequential data from the frontends is combined to form event data packages. Finally, the HLT filter systems reduce the data rate further to allow event processing and storage by the computing systems. (Source: [37, 57])

Chapter 3

Software and Computing

The LHC experiments are faced with many challenges in the computing area. In order to perform a physics analysis, very complex software is needed to generate Monte-Carlo events, simulate the detector response and study observables. The theory community takes care of the first part and offers a wide variety of programs to calculate cross-sections (e.g. NLOJet++/fastNLO) and generate Monte-Carlo events (e.g. Pythia, Herwig++ or Sherpa). The LHC collaborations give their members access to software packages (CMSSW) which interface these generators to their detector simulation and allow to run over the recorded data. They also provide a general framework to implement physics analyses. However for sophisticated analyses, it is necessary to extend the existing framework or even build an own framework (e.g. Kappa) which allows to process data at a faster rate and in more elaborate ways than foreseen by the framework provided by the collaboration.

To run these complex software packages, a powerful computing infrastructure is necessary. At the same time, an analysis needs access to the datasets collected by the detector, simulated data from event generators and the processing power to run on. However the data rates of the experiments and their detailed detector simulations lead to dataset sizes not seen in previous collider experiments. In addition, the precise theory calculations and the statistical analysis tools involved in an analysis can use a very large amount of computing power.

3.1 Worldwide LHC Computing Grid

In order to overcome these challenges, the LHC experiments decided to build a common distributed computing and storage infrastructure. In analogy to

the electricity grid, the Worldwide LHC Computing Grid (WLCG) provides scientists with an easy interface to access the available data of their collaboration and the computing power to analyse them. The WLCG was designed as a heterogeneous distributed system with a simple interface hiding the peculiarities of the different computing sites from the end user. Using the interface, it doesn't matter to the user whether his analysis runs on a small cluster in their basement or a large computer farm in another country. The software which provides the common interface is called grid middle-ware. While there are several middle-ware implementations available, the the LHC experiments decided to use the gLite[59] software for the majority of their systems.

The storage infrastructure on the grid provides easy access with high availability and redundant resources to speed up access to datasets which are in high demand. To ensure long term reproducibility of results, backup facilities in the form of large tape storage systems are also incorporated.

The computing centres are able to schedule the processing task according to certain priority lists. This allows to run certain high priority jobs concurrently with a large number of other end user jobs. These high priority jobs are usually managed centrally for the whole collaboration and consist for example of data reprocessing jobs using new detector alignment and calibration information and better adjusted physical models. Other high priority jobs are performing Monte-Carlo events generation and detector simulation. The sites participating in the WLCG are scattered around the whole world and mirror the geographic distribution of high energy physics centres.

3.1.1 The Tiered Architecture

The WLCG does not use an entirely decentralised architecture, since the main source of data is an experiment situated at a single point. To reflect the natural flow of information from the experiment to the user, the computing infrastructure has a hierarchical structure with four tiers. At the top of the hierarchy resides the computing centre at CERN - called the Tier-0, which directly receives the data from the experiments. Both raw detector data and reconstructed event data are stored locally on tape as backup. This data is simultaneously forwarded via high bandwidth connections to the globally distributed Tier-1 centres. The Tier-1 centres provide a large amount of the storage and computing resources to run more comprehensive reconstruction steps, general reprocessing, skimming jobs[1] and other tasks involving the raw detector data. Connected to one or more Tier-1 sites are

[1] Details of these jobs are explained in section 3.3.

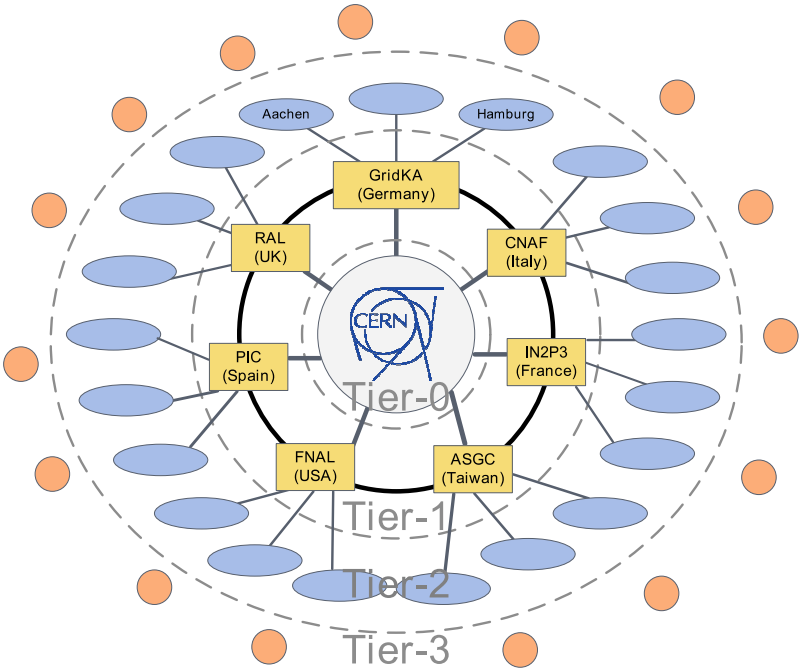


Figure 3.1: Tiered structure of the grid

the Tier-2 sites. While the Tier-1 sites provide the bulk of the available storage, the Tier-2 sites supply the main part of the computing resources used for example in Monte Carlo productions and user analyses. Tier-3 sites are small independent clusters participating in the WLCG and accessing the data from the other sites without defined commitments to the experiment like the Tier-1 and Tier-2 sites. As optional components in the computing model, they provide mainly resources to their own users.

3.1.2 User authentication

User authentication on the grid is based on certificates. In order to use grid services, it is necessary for the user to register his certificate at one or more Virtual Organisations (VO) and to authenticate at a Virtual Organisation Membership Service (VOMS) server. The first step in the authentication process is to create a so-called “proxy”, a certificate signed by the user and used for identification purposes on the grid. This proxy is sent to the VOMS server, which signs the proxy in case the user is authorised to use the resources of the VO. This proxy is then sent together with each user job to authenticate the user at grid services. Based on user information contained in this proxy,

like VO or group membership, user requests are authorised and prioritised at the grid resources. For security reasons the proxy has a limited lifetime and an expired proxy will abort grid jobs and deny access to storage systems.

3.1.3 Job work-flow

To process data on the grid, a user needs an account on a user interface (UI). The UI provides the necessary job submission and retrieval programs (popular choices are detailed in [60, 61]) and the tools to get a proxy from the VOMS server. With this proxy the user is then able to submit jobs to the Workload Management Service (WMS) server. A job consists of a file describing the job requirements, input and output files including the script or program to run. The general job setup and the requirements are specified using the Job Description Language (JDL). All these files are contained in the so-called input sandbox, which is send to the WMS during the job submission.

The WMS server queries other grid services to find the best suited Computing Element (CE) for the user job based on the requirements in the JDL file. The WMS server considers various metrics like the distance to any requested datasets, queue length and the sites and other factors in this matching process. Once the optimal computing element fulfilling the requirements is found, the WMS queues the job at a CE.

Computing Elements are the grid-widely visible portals to local batch queues at a site. The computing elements are organised as a collection of Worker Nodes (WNs), the building blocks of the batch system. These WNs themselves are the computing resources on which the user jobs are running. A batch system handles the distribution and monitoring of jobs at at CE. It schedules the job to run at a certain worker node, where the job has access to the local storage systems and other Storage Elements (SE) via the proxy credentials.

Such a storage element provides uniform access to the accessible storage space of a site. The SE may control large disk arrays, tape storage systems and the like. After the termination of the job, the output is transferred in a so-called output sandbox to the WMS server. There the job output remains until the user fetches the output sandbox.

3.2 Standalone Simulation and Analysis Software

This section gives an overview over the software tools used by the analysis. The cross-section programs and event generators which are used for the

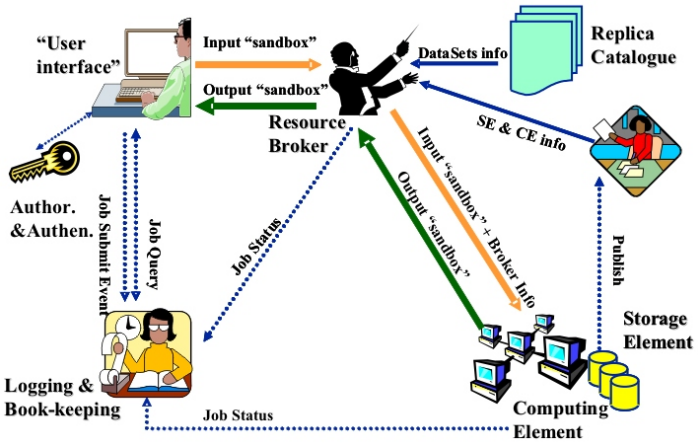


Figure 3.2: Detailed job work-flow on the grid. The resource broker is the central element in the work-flow.

theory prediction and Monte-Carlo studies are introduced. Other important programs related to the analysis work-flow, starting with the job and data processing and ending with the visualisation of results are also mentioned.

3.2.1 Pythia

Pythia [62] is a general purpose event generator which can be used to simulate high-energetic collisions of electrons, positrons, protons and anti-protons. It combines many theoretical and empirical models to describe parton remnants, initial- and final-state radiation, fragmentation, and multiple interactions. Pythia uses the Lund string hadronization model[24] and offers a multitude of tunable parameters for every aspect of the event generation. The development of this generator started in 1978 and due to its long history, it is well tested and widely accepted in the particle physics community and used by many physics groups in CMS. This analysis uses the most widely used version, Pythia 6, which is written in Fortran 77. However there is also a new implementation in C++ available in the form of Pythia 8.

3.2.2 Herwig++

Herwig++ [27] is a multi-purpose event generator which uses the cluster hadronisation model, in contrast to Pythia's Lund string model. There is a

sophisticated multiple parton interaction model [63] and a simple parametrisation[2] for the underlying event available. It also features an improved angular ordered parton evolution compared with the predecessor Herwig 6, which used a model in which radiation with angles of less than m/E, where m is the mass and E the energy of the heavy particle was forbidden. This resulted in a dead-cone around the particle without any radiation. In Herwig++ this approach is replaced by a smooth suppression of radiation in the direction of the particle, resulting in a better description of the process. Herwig++ also includes many improvements to the simulation of meson and tau decays. In case the corresponding matrix elements are available, it uses them to accurately decay particles. It also has the capability to simulate spin correlations through-out the event simulation. Herwig++ is the successor of the very successful Herwig 6 (Hadron Emission Reactions With Interfering Gluons) Monte Carlo event generator [21]. It is based on the ThePEG framework [64], which provides a common platform for using and building event generators in C++.

3.2.3 Sherpa

The Monte Carlo event generator Sherpa allows the Simulation of High-Energy Reactions of Particles (Sherpa). The major difference between Sherpa and Herwig++ or Pythia is the ability to use a built-in matrix-element generator, which automatically calculates and integrates tree-level amplitudes for processes with user-defined input and output states. In order to merge any multi-jet final states with the results from a parton shower algorithm, Sherpa employs the CKKW[23] matching technique[65]. The parton shower algorithm is based on the cluster fragmentation model like Herwig++ while the multiple parton interaction model is based on the ideas implemented in Pythia.

3.2.4 NLOJet++ / fastNLO

NLOJet++ is a C++ program for calculating jet cross section. It uses a slightly modified Catani-Seymour dipole subtraction method[66], which is able to calculate jet cross sections in arbitrary scattering processes to next-to-leading accuracy in perturbative QCD. At the moment, NLOJet++ can calculate one-, two- and three-jet observables at next-to-leading order and offers a general framework to implement specific scenarios, like the three-jet

[2]Based on data gathered by the UA5 collaboration.

mass cross-section. In order to compare the theory prediction of this program with data, additional non-perturbative corrections are necessary.

It should be noted, that the program is unable to simulate individual events suited for detector simulations. The partonic states simulated by NLOJet++ can carry very large positive and negative weights and therefore only the ensemble mean can be interpreted as a physical result. In order to ensure the cancellation of these large weights, a very large number events has to be simulated. This makes the calculations done by NLOJet++ very computing intensive.

However for PDF fits or in systematic studies the same cross-section calculation needs to be repeated many times for different parameters. In order to make these computations more efficient, the fastNLO package can be used.

fastNLO uses NLOJet++ to perform the initial perturbative calculation. In this calculation, the fractional proton momentum x dependence of the PDFs and the scale dependence of α_s and the PDFs are approximated using interpolation functions between fixed support points. This yields tables of coefficients for the x and scale support points, which can be used to quickly evaluate the cross section for different PDFs, values of $\alpha_s(M_Z)$ and scale choices.

3.2.5 ROOT

ROOT [67] is a powerful object-oriented data analysis framework written in C++. It offers efficient access to the data from the experiments, exhaustive analysis tools and versatile visualisation options for data using object oriented interfaces. The user can use these interfaces with compiled and interpreted C++. There are also language bindings to high level languages such as Python and Ruby available. The PyROOT interface between Python and ROOT was extensively used throughout this work to transfer histograms from ROOT to Python for visualisation and statistical analysis.

ROOT offers its own tree-based file format for storage, which enables writing arbitrary data structures to manifold storage backends. For data analysis ROOT has integrated several libraries like MINUIT2, TMVA and RooFit to let the user do curve fitting, minimisation, multivariate data analyses, neural network studies and much more. To display the results of the analysis, ROOT provides comprehensive histogramming and drawing routines in 2D and 3D.

3.2.6 Python, NumPy and matplotlib

The scripting language Python[68] is used in conjunction with the scientific computing package NumPy[69] for a large part of the statistical analysis in this study. It provides powerful array objects, on which sophisticated statistical functions can be applied. There is also a large suite of linear algebra tools available in the package. For the data visualisation, the matplotlib package was used. Matplotlib[70] gives a very convenient high-level interface to produce publication quality figures and was used throughout this work.

3.2.7 grid-control

The grid submission tool grid-control[61, 71] was used for all data processing steps in this analysis. The author is for several years now the main developer of this program, which was also used in a large number of other works. It makes access to both local and grid computing resources very transparent and contains powerful tools to manage complex job work-flows. The support of parametrised jobs in particular was heavily used during the evaluation of the uncertainties in this study.

3.3 CMSSW - The CMS Software Framework

CMSSW is the software framework of the CMS collaboration [72, 73]. It offers a common framework for event generation, simulation, data analysis and online data taking. It consists of many packages and modules and is based on the Framework Core (FWCore) which organises the usage of modules and in turn utilises many external libraries like ROOT or FastJet. The analysis work-flow is based on a bus paradigm where modules are arranged along paths which produce or read different types of products, e.g. a jet collection. An analysis is setup using python configuration files where the user can add or remove modules from the work-flow or modify the predefined module settings. This modular approach allows to switch easily between different reconstruction methods and detector simulations. CMSSW also keeps track of the data describing the conditions and calibrations of the reconstruction algorithm of each data object.

3.3.1 Event Data Model

CMSSW is based on the so-called Event Data Model (EDM) which uses the event as the basic processing unit. In this concept, the event is a general container for all data that was taken during a measured or simulated physics

event as well as all objects derived from the data. Such an event content can include objects representing the raw detector readout, reconstructed objects, detector simulation products, and analysis objects relating to a single beam crossing or simulation thereof.

When an event is processed by a module, that module can both get data from the event and put data back into the event. When data is put into the event, the provenance information about the module that created the data will be stored with the data into the event.

Event Setup

Besides the event data, there is additional information necessary for the event processing. In CMSSW, information of this kind is accessed via the so called event setup. The meta-data of the event, data about the detector environment and status, is not tied to a single event but is valid for a specified period of time. This is because the running conditions and the performance of the detector can change in time. For example defects like hot cells in the calorimeter, which report continuously large energies, are marked as such in the event setup. From Data Quality Monitoring (DQM) the evolution of these changes is known. This is also the reason for the fact that many parts of this information are stored in external databases instead of the data files themselves, in order to be able to update certain information more regularly. This allows to retroactively change the data quality status in case previously unnoticed problems are uncovered.

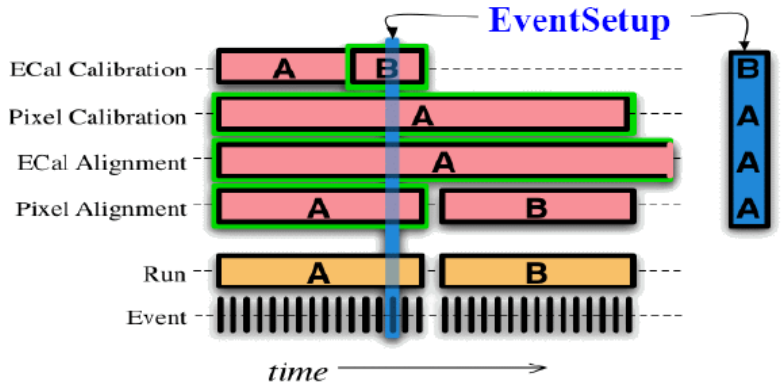


Figure 3.3: Illustration of the CMS event setup. Each collision event has a specific set of calibration and alignment measurements or other run conditions associated with it.

In order to handle these time-dependant changes of the event setup, the concept of an interval of validity (IOV) is introduced. Such an IOV can

span hours and therefore, a specific instance of event setup meta-data can be valid for many millions of events (see figure 3.3). The auxiliary information in such a record will include calibration and alignment measurements, geometry descriptions and other run conditions.

3.3.2 Modular Analysis Work-flow

The purpose of a module in CMSSW is to allow the independent development and verification of distinct elements of triggering, simulation, reconstruction, and analysis. Each module encapsulates a clearly defined event-processing functionality. The modules may not communicate directly with each other, which allows them to be independently tested and reused. Communication between modules is only possible through the event and in the direction of the set-up work-flow. The events are processed in this work-flow by passing them along the configured path of modules. An example of such a work-flow is shown in figure 3.4. The available modules can be classified into different categories.

A source module is at the start of the work-flow and provides the events which are to be processed. These events can be from a file on storage, an external event generator or the CMS data acquisition system.

Producer modules can read data from the event and generate new event objects. All reconstruction algorithms are implemented in the form of such a producer. These modules are also often linked against an external library (Geant 4 [74], FastJet [75], ...) which holds a substantial amount of functionality of the producer. A prominent example for such a module is the FastJetProducer, which takes some input objects from the event, applies one of the jet clustering algorithms from FastJet and puts the resulting jets collection into the event.

A filter module on the other hand evaluates the event according to some user defined criteria and signals whether or not to proceed with processing this event. An example for such a filter is the GoodVertexFilter, which takes a vertex collection as input and aborts the processing of the event when certain quality criteria are not fulfilled.

The third class of modules is formed by the Analyzer modules, which are used to perform user processing steps like generating histograms. While these modules can access products from the event and study them, they are not allowed to put products into the event or influence subsequent scheduled steps.

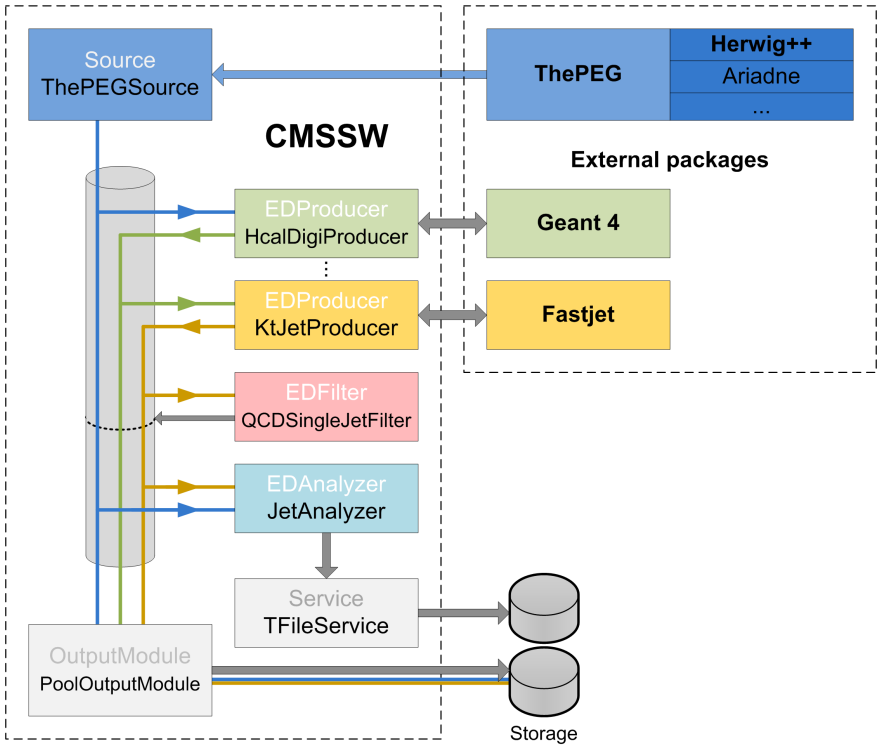


Figure 3.4: Short overview of a typical CMSSW job work-flow. In the first step Monte Carlo events are generated using the ThePEGInterface. Afterwards a simulation is run and jets are reconstructed. The jets are then filtered and analysed in a user module. At the end the results are written to disk.

Monte Carlo Event Generation

Monte Carlo event generators in CMSSW provide the framework with fully hadronised final states. The simulation usually starts with a leading order calculation of matrix elements. On the result, a parton shower algorithm is applied to provide an approximative perturbative treatment of higher order QCD dynamics. The partonic final state is then combined with a hadronization model to give final state particles like hadrons, leptons and photons. The generators provide their output in the form of an HepMC[76] event record, which consists of a detailed description of the simulated interaction. CMSSW offers interfaces to many generators, ranging from the general purpose Monte Carlo event generator Pythia and Herwig++, the multi-jet generators Sherpa, MadGraph and Alpgen, to particular generators such as the black hole event generator Charybdis[77]. The configuration of these generators can be done directly from within the CMSSW configuration file.

Detector Simulation

The hadronic final state from a Monte Carlo event generator can be passed to the detector simulation. There are two detector simulations available in CMSSW. Both take as input the HepMC event record produced by a Monte Carlo generator. The full simulation is using Geant 4 [74] to provide an accurate simulation of the detector response using a detailed detector geometry. It simulates in detail the interaction of each particle and its secondary particles with the detector material in each detector component. These very complex computations take several minutes for a typical event.

To reduce the massive amount of processing power necessary for Monte Carlo studies, there is also a fast simulation available. The fast simulation uses Pythia to simulate the particle decays together with various simplifications and parametrisations in the interaction of particles with the detector. As a result the simulation of a single event can take approximately one second. However the fast simulation has the drawback that it has to be tuned to the Geant based detector simulation to give reasonable results.

Both detector simulations provide a collection of hits in the different detector subsystems as output. These hits can be used as input for the reconstruction algorithms, just like hits taken from the data acquisition systems.

Event Reconstruction

To allow the analysis of an event with the measured or simulated detector output, it is necessary to run reconstruction algorithms on the data. For these further analysis steps there are many different types of physical objects to be reconstructed. The hits in the silicon strip tracker have to be matched to give a particle track and the calorimeter towers have to be clustered into jets. From the hits in the tracker and muon chambers the muon momentum and energy have to be derived. Following the CMSSW design, the different models for the reconstruction of these objects are each contained in a module which allows the user to choose the most appropriate algorithm for the analysis. By default, the reconstruction of the jet objects is performed with the anti-k$_T$ algorithm for two different jet sizes $R = 0.5$ and $R = 0.7$ and is using the jet clustering implementation from the FastJet [75] package.

3.3.3 Kappa

Kappa is the acronym for KArlsruhe Package for Physics Analyses. It is a powerful skimming framework implemented within CMSSW in the form of an analyser plug-in and consists of many different modules which work

similarly to the producer plug-ins known from CMSSW. The modules extract information from within the CMS software, perform processing steps on it and then stores the restructured data using the ROOT object I/O system.

The Kappa framework gives the modules the ability to select input objects from CMSSW using simple regular expressions. Since it is a common task to process many input objects in a similar fashion, this is a very convenient tool not provided by the normal CMS software. A set of base classes allow to easily develop Kappa modules which iterate over collections of input objects or perform basic selection steps like applying a cut on the transverse momentum. A normal Kappa module is therefore very compact and usually just consists of the selection of output variables and some processing steps to calculate derived variables. The framework employs several techniques which hide much of the complexity connected to working with the ROOT I/O system. Since the modules only store the information absolutely necessary for the analysis, the space occupied by a Kappa dataset is more than one order of magnitude smaller than the space needed by the unprocessed CMSSW dataset.

There is also a set of tools and application programming interfaces called KappaTools, which provides comfortable access to the information contained in a Kappa dataset. The foundation of the Kappa framework was basically developed for this analysis, but a large number of Kappa modules were added to allow the use in other studies as well. The first processing step for detector data in this analysis is to process the CMSSW datasets into Kappa datasets. These Kappa datasets are later analysed using the facilities provided by KappaTools.

Chapter 4

Jet Reconstruction

Partons from the hard process of an interaction result in a collimated stream of particles after the parton shower and hadronization. This roughly collinear stream of particles is the experimental signature of quarks and gluons and such signatures are called jets. From the detector point of view, jets manifest themselves as localised clusters of deposited energy accompanied by a large number of tracks pointing towards the deposited energy. All these objects have to be combined into well defined jets to reconstruct information about the initial partons.

For hadronised particles from the Monte Carlo generator these jets are called *particle jets*. The jets reconstructed from energy clusters in the calorimeter are called *calorimeter jets*. When the input particles are reconstructed from particle flow objects, the resulting jets are called *particle-flow jets*.

4.1 Particle flow reconstruction

The particle flow reconstruction [78, 79] used by CMS combines information from all sub-detectors to reconstruct and identify all stable particles in an event. The particle flow algorithm uses five categories to identify particles: photons, electrons, muons, neutral and charged hadrons. The basic elements of the particle flow algorithm are the reconstructed tracks and vertices in the tracker, calorimeter clusters in the ECAL and HCAL and tracks in the muon system.

The tracks used by the particle flow algorithm are taken from the default track reconstruction of CMS, which uses the Combinatorial Track Finder (CTF) algorithm. Likewise, the default vertex reconstruction using the CTF tracks is used to determine the list of primary vertices in the event. Due to pile-up, multiple primary vertices can appear in a single event.

In order to incorporate the hadronic and electromagnetic calorimeter, the particle flow algorithm uses a grid of cells based on the detector granularity in order to identify calorimeter cluster seeds. Seeds with energies above a certain threshold are used in an iterative merging algorithm to form so-called particle flow clusters.

These different elements are linked together into so called particle flow building blocks based on geometry and χ^2 fits. These building blocks are used to reconstruct the final particle flow candidates. Blocks connected with the muon system are analysed to identify particles belonging to the muon category. Similarly, multi-variate methods are employed to reconstruct electrons from blocks with links to the tracks and ECAL clusters. Charged hadrons are identified using the remaining blocks containing calorimeter clusters and links to the tracking system. At this stage only neutral objects without traces in the tracking system are left. Depending on the energy of the reconstructed charged hadrons, ECAL clusters are assigned to photon candidates, while the remaining HCAL clusters are interpreted as deposits from neutral hadrons.

It should be emphasised that in each step of the particle flow reconstruction, great care is taken to avoid any kind of energy double-counting. This is done by removing the particle flow elements of successfully reconstructed particle candidates from any linked blocks and by recalculating the energy of calorimeter clusters. In the end, all particle flow elements are therefore uniquely assigned to a certain particle flow candidate.

The final result of the particle flow algorithm is a set of well identified particles, using the excellent resolution of the tracking system where applicable to determine or improve the momentum measurement.

4.2 Jet Algorithms

There are many different jet algorithms available to cluster a set of input objects together. In the CMS collaboration, the two principal jet reconstruction algorithms are the anti-k_T jet algorithm and the inclusive k_T jet algorithm. Both are reconstructed using the the FastJet package[75, 80, 81]. The two algorithms fulfil two important properties in that they are collinear and infrared safe (see figure 4.1). Collinear safety means that their output is resistant to collinear splitting and merging of input objects. The fact that additional soft input particles do not change the reconstructed jets is called infrared safety.

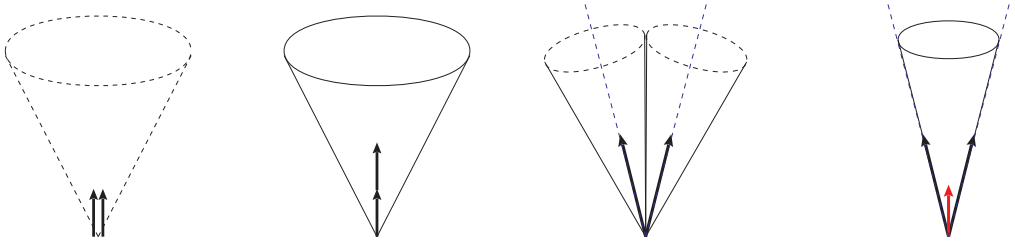
Both algorithms are of sequential recombination type and cluster the objects based on a distance measure in Minkowski or Euclidean[1] space. There

[1]The Cambridge/Aachen jet algorithm computes distances in Euclidean space.

are also cone based algorithms available, which combine the input into cone shaped jets where all constituents lie within a predefined distance from the centre of the cone in Euclidean space.

A jet description is defined by the algorithm, the jet size and the recombination scheme which describes how objects are merged together by the algorithm.

Although several recombination schemes exist, only four vector addition (called E-scheme) is widely used in CMS. However this recombination scheme also has the property that calorimeter objects, which are assumed to be massless, can produce massive jets. Other recombination schemes (e.g. the E_T-scheme) do not necessarily have this property.



(a) Collinear unsafe: A quasi-collinear splitting can lead to different jet configurations (e.g. disappearing jets).

(b) Infrared unstable: The addition of a soft particle (red) can lead to new stable jet configurations.

Figure 4.1: Illustration of collinear and infrared safety violations

4.2.1 Generalised k_T Algorithm

The generalised k_T jet algorithm[82, 83] is a sequential recombination jet algorithm and is infrared and collinear safe. It uses a jet size parameter R and distance measure parameter n and works according to the following algorithm: The clustering procedure starts with a list of input objects, e.g. stable particles or calorimeter towers.

1. For each pair of particles i, j compute the distance measure d_{ij} as well as all beam distances:

$$d_{ij} = \min\left(p_{Ti}^{2n}, p_{Tj}^{2n}\right)\frac{\Delta R_{ij}^2}{R^2}, \qquad d_{iB} = k_{Ti}^{2n}$$
$$\text{where } \Delta R_{ij}^2 = (\Delta y_{ij})^2 + (\Delta\phi_{ij})^2$$

   with the transverse momentum p_{Ti} of particle i, the rapidity difference Δy and the azimuthal angle $\Delta\phi$ between particle i and j. d_{iB} is called the beam distance.

2. The algorithm searches for the smallest distance d_{ij} or d_{iB}. If it is a distance from the category d_{ij}, the two particles i, j are merged using a defined recombination scheme into a new particle k and which replaces the original particles. If the smallest distance is of type d_{iB}, we call particle i a jet and remove it from the list of particles.

3. Repeat from step 1 until all objects are included in jets

The algorithm successively merges objects which have a distance $\Delta R_{ij} < R$. It follows that $\Delta R_{ij} \geq R$ for all final jets i and j. Because the distance is usually calculated in Minkowski space, the projection into Euclidean space has irregular jet shapes in general as shown in figure 4.2. Three important jet algorithms, which belong to the generalised k_T jet algorithm class are:

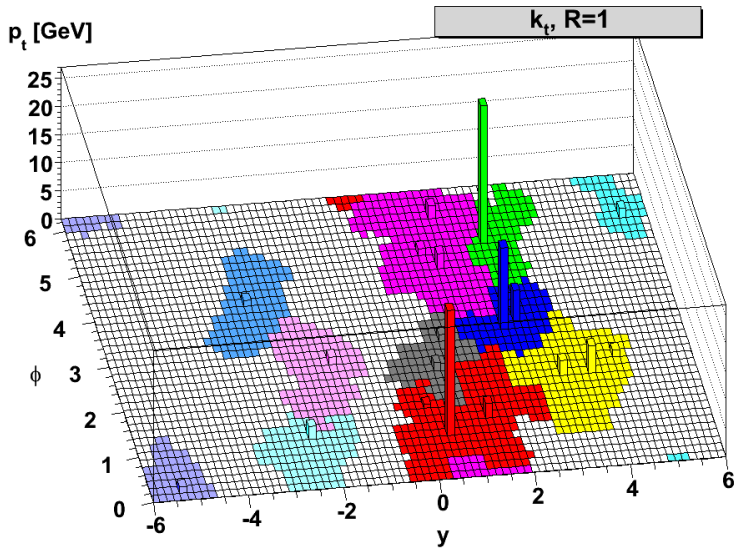


Figure 4.2: Result of the inclusive k_T algorithm. Calorimeter towers on a fixed grid in the $y - \phi$ plane are assigned to jets by the jet algorithm. The colouration shows the area covered by the jet. (Source: [84])

Inclusive k_T algorithm ($n = 1$)

Inclusive k_T algorithm [83] is the archetype of the generalised k_T Algorithm. Descriptively speaking, the inclusive k_T clustering successively undoes the QCD branching in the parton shower to reconstruct the parton.

Cambridge-Aachen algorithm ($n = 0$)

The Cambridge-Aachen algorithm clusters according to the angular distribution of the input objects. The energy of the objects is not considered in the process. Like the k_T algorithm, the resulting jets do not possess a regular shape. However the algorithm has properties which make it useful for the detailed study of jet shapes and sub-jet properties.

Anti-k_T algorithm ($n = -1$)

The Anti-k_T algorithm [85] favours clustering the hard input particles in the event, this is similar to seeded jet algorithms like SiScone[86, 87], in that the jet grows outwards around hard seeds. The form of the distance metric results in cone shaped jets. Due to this similarity to cone-based algorithms and the infrared and collinear safety of the algorithm, this algorithm is the standard algorithm used by the collaboration.

4.2.2 Jet Area

For the correction of the jet energy in the detector[88], the area covered by the jet object in the $y - \phi$ plane is of interest. For a cone-based jet algorithm the nominal jet area is πR^2, however for cluster-based algorithms more sophisticated methods to determine the jet area are necessary. The FastJet package linked to the CMS experiment software uses the Voronoi tessellation method described in [89, 90] to calculate the jet area. The Voronoi tessellation decomposes the $y - \phi$ plane into areas, determined by distances to a set of jets scattered in the plane. The corresponding Voronoi cell (or Voronoi region) of a jet consists of every point in the plane, whose distance to the jet is less than or equal to its distance to any other site. Since there are fast algorithms available to calculate the Voronoi tessellation, the area of a jet can be efficiently calculated.

4.3 Jet energy corrections

Jet algorithms are designed to reconstruct the momentum of an initial parton from both Monte Carlo truth information or detector data. But there are several effects causing the reconstructed jet energy to not correspond to the energy of the parton. For calorimeter jets this is mainly caused by detector effects like electronic noise, pile-up and underlying event energy, non-uniformities and non-linearities in the detector response. Other influences are related to the jet reconstruction itself and also effect particle jets.

They include the fragmentation model, initial- and final-state radiation, the underlying event and pile up.

The official CMS jet energy corrections are using a factorised multi-level approach[91] to correct for these effects. The first three correction levels are mandatory for jets from both data and detector simulation.

1. **Pile-up correction:** The Level 1 correction aims to subtract pile-up from the jet energy. Pile-up refers here to energy deposition from additional pp collisions, occurring close enough in time to the hard scatter to be included in the jet energy. This correction uses the concept of jet area to determine the median energy deposition in the detector per unit area, which is then individually removed from the measured energy of each jet. The input for the correction is therefore the median energy density ρ of the event as determined by the k_T jet algorithm with cone size $R = 0.6$ and the area of the jet calculated with the Voronoi technique described in section 4.2.2. Both variables are provided by the FastJet software.

2. **Relative correction:** Due to the geometry of the detector, the reconstructed energy of jets depends on the pseudo-rapidity. The purpose of the Relative (Level 2) correction is to flatten the jet response as a function of the pseudo-rapidity.

3. **Absolute correction:** The goal of the Level 3 (p_T dependence) correction is to remove jet response variations in the CMS detector as a function of p_T, which is primarily a result from the non-linear calorimeter response.

Since the analysis studies QCD events, only these three correction levels are applied to data and Monte-Carlo. Additional correction levels like the flavour correction, which is intended to correct for the assumption of a QCD flavour composition underlying the determination of the previous corrections steps, are not necessary.

This allows to express the jet energy correction factor k as a function of the transverse momentum, pseudo-rapidity and area A of the uncorrected jet and the median jet energy per area of the event ρ. The corrected jet four-vector p'_i can be calculated from the uncorrected jet four-vector p_i using:

$$p'_i = k(p_{T,i}, \eta_i, A_i, \rho) \cdot p_i$$

The L1,L2 and L3 correction are determined using data-driven methods. In order to better understand the corrections, they are expressed in terms of

one factor which can be derived with Monte-Carlo techniques and a second factor, which describes the differences between the jet response in the detector simulation and the response as measured with data-driven methods. This second correction factor is close to unity and must only be applied to data. It is given in the form of a η and p_T dependent correction and called the L2L3 residual correction.

Chapter 5

Theory prediction for the three-jet mass

The ultimate goal of this analysis is to study a multi-jet observable to gain a better understanding of the theory of quantum chromodynamics at the high scales accessible with the LHC. This is accomplished by performing a detailed comparison between a measurement using the CMS detector and the best available theory prediction. This chapter aims to give an introduction how the theory prediction and the corresponding uncertainties are derived.

As presented in the introduction, the focus is on the measurement of the invariant mass of three-jet events:

$$m_3^2 = (p_1 + p_2 + p_3)^2$$

with the four-vectors $p_1 \dots p_3$ of the three jets with the largest transverse momentum ($p_{T,1} \geq p_{T,2} \geq p_{T,3}$) passing certain selection cuts. The three-jet mass cross section is measured double differentially in the three-jet mass and the maximum rapidity of the three-jet system:

$$\begin{array}{rcl} \dfrac{d^2\sigma}{dm_3 dy_{\max}} & & \\ y_{\max} & = & \max\left(|y_1|, |y_2|, |y_3|\right) \end{array}$$

with the rapidity y_i of the i-th jet. The binning of the three-jet mass m_3 is resolution driven, while the rapidity binning is a simple integer binning, loosely following the CMS detector geometry. Both binnings are identical for the measurement and the theory calculation.

5.0.1 Cut scenarios

The three-jet mass is studied for different three-jet event selection criteria, which are called cut scenarios in the following. All cut scenarios have in common that jets with $|y_i| > 3$ or $p_{T,i} < 50\,\text{GeV}$ are discarded. This allows to use the same acceptance cuts for the theory calculation and the measurement.

The distinguishing feature of the analysis cuts is how the cut on the third jet operates. The first scenario directly applies a lower limit on the transverse momentum of the third jet $p_{T,3} > p_{T,\text{min}}$ and is called the "absolute cut". In this analysis the threshold was chosen to be $p_{T,\text{min}} = 100\,\text{GeV}$. The second cut scenario requires that the ratio $r = p_{T,3}/\langle p_{T,1}, p_{T,2}\rangle$ between the transverse momentum of the third jet and the average transverse momentum of the first and second leading jet is larger than a certain threshold r_{sel}. The relative cut value is set to $r_{\text{sel}} = 0.25$ in this study. This cut is also called the "relative cut". The absolute cut with $p_{T,\text{min}}$ is harder for jets with small transverse momenta, while the relative cut is harder for transverse momenta $p_{T,3}$, as soon as the average $p_{T,12} = \langle p_{T,1}, p_{T,2}\rangle$ of the leading two jets is larger than $p_{T,\text{min}}/r_{\text{sel}}$. For the used values in this analysis the equation resolves to $p_{T,\text{min}}/r_{\text{sel}} = 400\,\text{GeV}$.

The next-to-leading order cross-section predictions for the two cut scenarios is shown in figure 5.1 for the CT10 PDF[92]. The cross section of the absolute cut scenario increases up to a three-jet mass of $\approx 500\,\text{GeV}$ before starting to decline. This increase is due to the large number of three-jet events, which do not fulfil the hard jet selection cuts in the low three-jet mass region. The relative cut scenario with the much softer cut in the low-p_T region doesn't show such a behaviour for the observed low three-jet masses. The fact that the relative cut is harder than the absolute cut above $p_{T,12} > 400\,\text{GeV}$, causes the three-jet mass cross section predictions of the relative cut to drop below the absolute cut around a three-jet mass of $m_3 \approx 1000\,\text{GeV}$.

5.1 Next-to-leading order prediction

The double differential three-jet mass observable can be predicted with the NLOJet++[93] program with next-to-leading order precision. However this calculation is very computationally intensive. In order to accelerate the calculation, NLOJet++ is run together with the FastNLO package[94]. FastNLO reduces the run-time of the evaluation of the three-jet mass cross sections for different PDFs, α_s and scales by storing the matrix element information as a function of the scales and the fractional hadron momenta. Two different choices for the relevant scale of the process were investigated. The first is

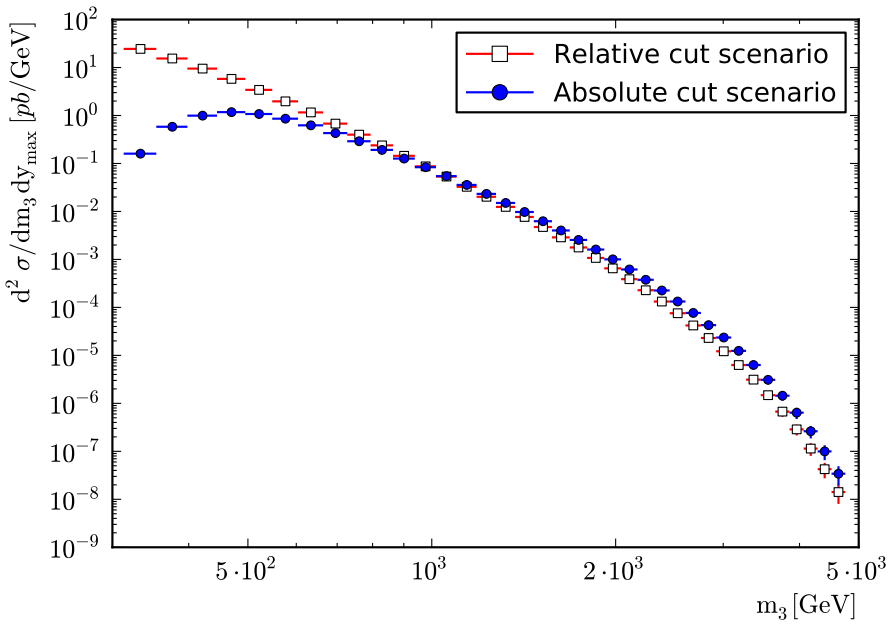


Figure 5.1: Next-to-leading order theory predictions for the three-jet mass calculated using NLOJet++ and fastNLO. The calculation is presented for the two cut scenarios in the innermost rapidity bin using the CT10 - NLO PDF. For the absolute cut scenario, the three-jet mass cross section increases up to a three-jet mass of $\approx 500\,\text{GeV}$ before starting to decline. The increase is caused by the fact, that many events do not fulfil the hard jet selection cuts in the low three-jet mass region. The relative cut scenario has a softer cut in the low-p_T region and therefore a higher three-jet mass cross section. However above $p_{T,12} > 400\,\text{GeV}$, the relative cut is harder and this causes the three-jet mass cross section predictions for the relative cut to drop below the absolute cut around a three-jet mass of $m_3 \approx 1000\,\text{GeV}$.

the average transverse momentum of the leading three jets:

$$p_{T,\text{avg}} = \langle p_{T,1}, p_{T,2}, p_{T,3} \rangle = \frac{p_{T,1} + p_{T,2} + p_{T,3}}{3},$$

while the other choice is half the three-jet mass $m_3/2$ of the event.

The next-to-leading order prediction was calculated for the Anti-k$_T$ 0.5 jet algorithm in the inner two rapidity bins ($|y_{\max}| \leq 1$ and $1 < |y_{\max}| \leq 2$) with both the absolute and relative cut. In addition, theory calculations for Anti-k$_T$ 0.7 jets with the absolute cut scenario are available in the two rapidity bins.

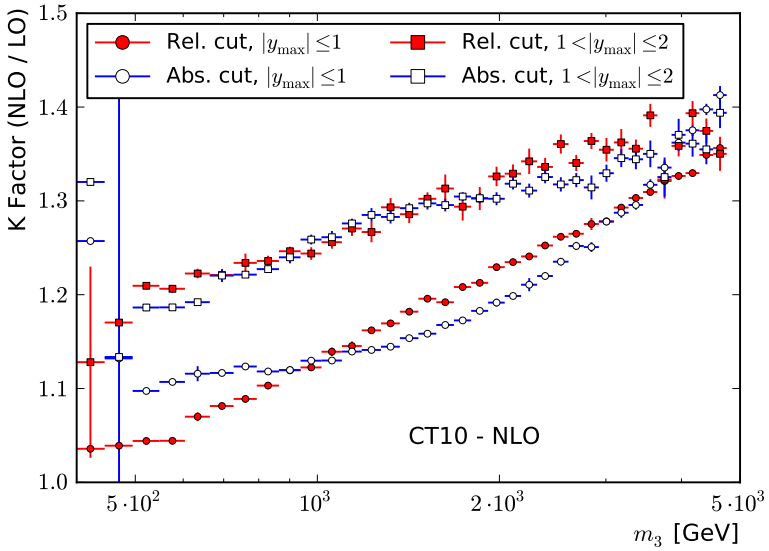
5.1.1 NLO K-factors

One of the basic checks of a perturbative calculation is to take a look at the impact of an additional term in the perturbative series on the result. The difference between the leading order and next-to-leading order prediction can be expressed in the form of a so-called K-factor. There are multiple ways to define NLO K-factors, however for the studies performed in this analysis, the K-factor is defined as the ratio between the double-differential three-jet mass cross-section prediction at next-to-leading order precision $\sigma_{\text{NLO}}/dm_3dy_{\max}$ and the prediction at leading order precision $\sigma_{\text{LO}}/dm_3dy_{\max}$:

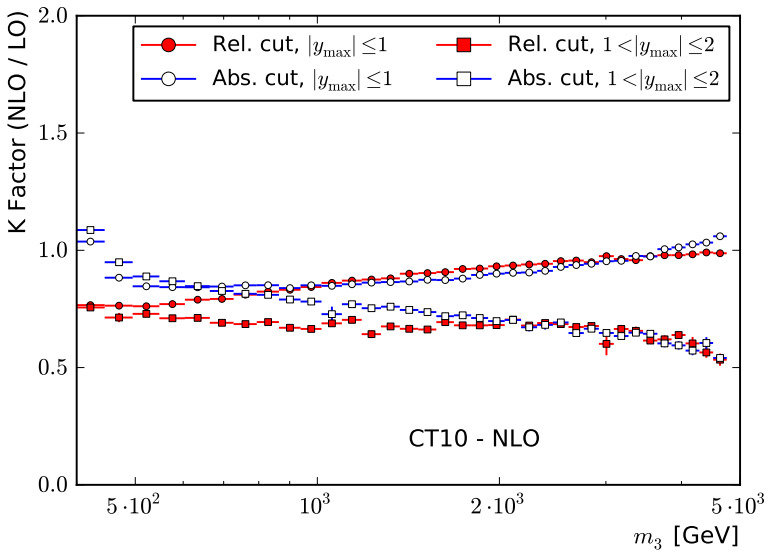
$$k_{\text{NLO}} = \frac{\sigma_{\text{NLO}}/dm_3dy_{\max}}{\sigma_{\text{LO}}/dm_3dy_{\max}}$$

The applied cuts, parton density functions, α_S and scale choices for the factorisation and renormalisation scale μ_F, μ_R for both $\sigma_{\text{NLO}}/dm_3dy_{\max}$ and $\sigma_{\text{LO}}/dm_3dy_{\max}$ are identical in this case. In particular the perturbative order of the used PDF is the same for the two predictions. The error on the K-factor is determined from the propagated statistical uncertainty of the two predictions.

The K-factors for the Anti-k$_T$ 0.5 jet algorithm are presented in figure 5.2. The size of the next-to-leading order corrections, ranging between -50 and $+70$ percent of the leading order calculation, is relatively small. This means that the leading order calculation already gives a good approximation of the three-jet mass cross-section. The figure also shows that the behaviour of the K-factors in general is very similar for the two cut scenarios. Larger differences from this rule are visible for the K-factors in the low mass region, determined with the average p_T as scale. In this region, the absolute cut is harder than the relative cut and has an direct influence on the average p_T scale. Another general feature visible in figure 5.2 is that the K-factor for



(a) $\mu = m_3/2$



(b) $\mu = p_{T,\text{avg}}$

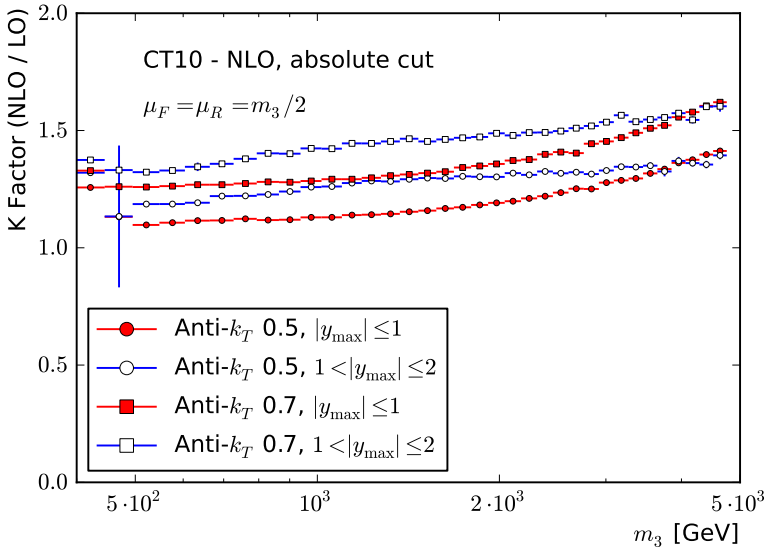
Figure 5.2: These figures presents the K-factors of the three-jet mass prediction with the Anti-k_T 0.5 jet algorithm for the absolute and relative cut in two rapidity bins. On the top panel, the event scale used in the theory calculation is set to $m_3/2$, while the bottom panel uses $\mu = p_{T,\text{avg}}$. The size of the next-to-leading order corrections is relatively small, confirming the good approximation already present at leading order. The presented behaviour is very similar for the two cut scenarios. Differences are observed for K-factors in the low mass region, with the average p_T as scale.

the outer rapidity bin $1 < |y_{\max}| \leq 2$ is farther away from unity than the inner rapidity bin. This means the next-to-leading order corrections become more important for the outer rapidity bins.

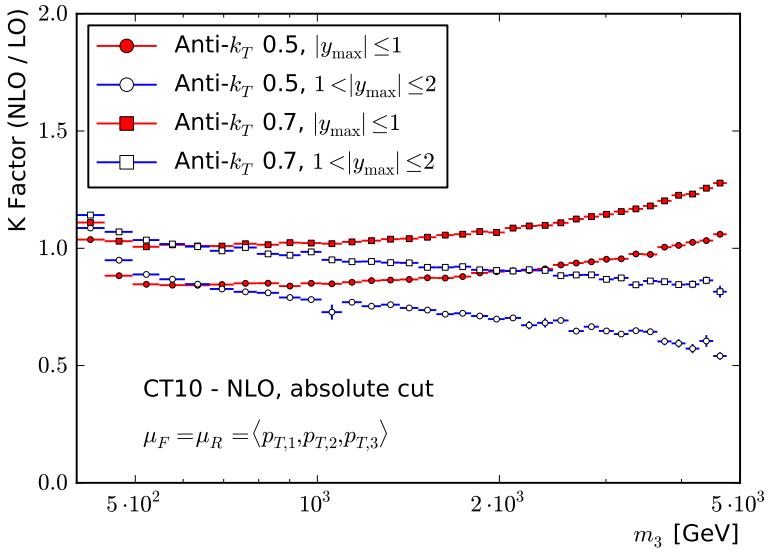
Figure 5.3 shows the K-factors for both the Anti-k$_T$ 0.5 and Anti-k$_T$ 0.7 jet algorithm in the absolute cut scenario. The K-factor for the Anti-k$_T$ 0.7 jets is always larger than the K-factor of the Anti-k$_T$ 0.5 jet algorithm. However they show the same behaviour as a function of the three-jet mass. This is demonstrated in figure 5.4, which shows the ratio of the K-factors for the two jet algorithms. The difference between the two algorithms is smaller for the theory calculation using $m_3/2$ as scale of the event. For this scale choice, the difference is also the same for both the inner and outer rapidity bins. Using $p_{T,\text{avg}}$ as scale, the deviation between the two jet algorithms becomes larger and shows a slope for the outer rapidity bin.

A direct comparison between the influence of the two scales is given in figure 5.5. First, the K-factors with the relative cut scenario are shown for different scale choices in figure 5.5(a). This plot demonstrates, that this choice in the NLO corrections becomes more important in the outer rapidity bins. The figure also shows that for the relative cut scenario, the scale choice directly influences the sign of the next-to-leading order correction. For the $p_{T,\text{avg}}$ scale, the leading order calculation needs to be corrected downwards, while for the $m_3/2$ scale, the LO result is corrected upwards. Figure 5.5(b) presents the ratio between the K-factors determined with the $m_3/2$ scale and the K-factors using the $p_{T,\text{avg}}$ scale. It is clearly visible that for the inner rapidity bin, the scale choice simply translates into an offset in the size of the NLO correction. However for the outer rapidity bin, a slope is clearly visible, which means that the choice of scale becomes more important for the high three-jet mass region.

The theory prediction using the $m_3/2$ scale exhibits K-factors in a small range, features a direct correlation with the observable and shows only small differences between the NLO corrections for different jet sizes and rapidity regions. Therefore, this scale is chosen as the default scale in the following discussions.



(a) $\mu = m_3/2$



(b) $\mu = p_{T,\text{avg}}$

Figure 5.3: These figures presents the K-factors of the three-jet mass prediction with the absolute cut scenario in two rapidity bins for both the Anti-k_T 0.5 and 0.7 jet algorithm. On the top side, the event scale used in the theory calculation is set to $m_3/2$, while the calculation in the bottom plot uses $\mu = p_{T,\text{avg}}$. The K-factor for the Anti-k$_T$ 0.7 jets is consistently larger than the K-factor for the Anti-k$_T$ 0.5 jet algorithm. However both show the same three-jet mass dependence. The difference between the two jet sizes can by minimised by choosing $m_3/2$ as scale of the event.

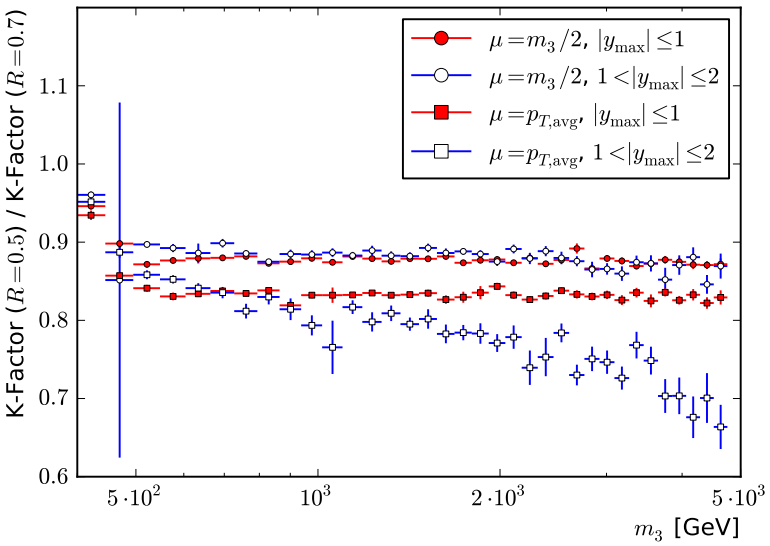
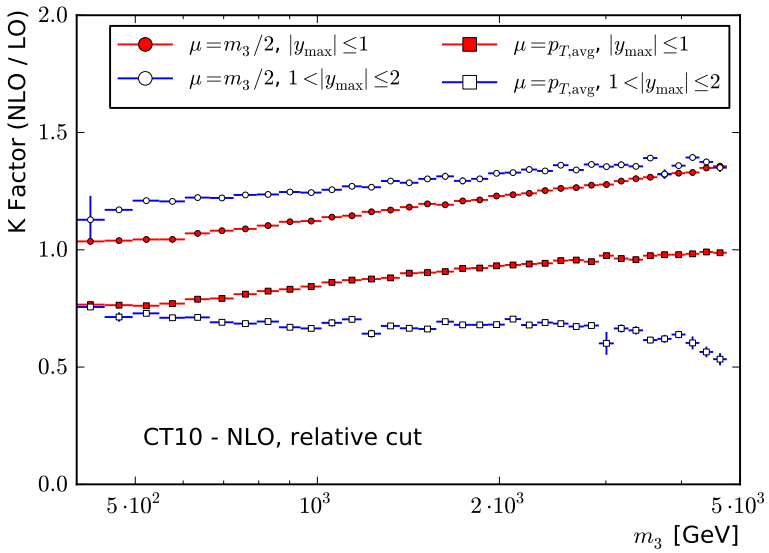
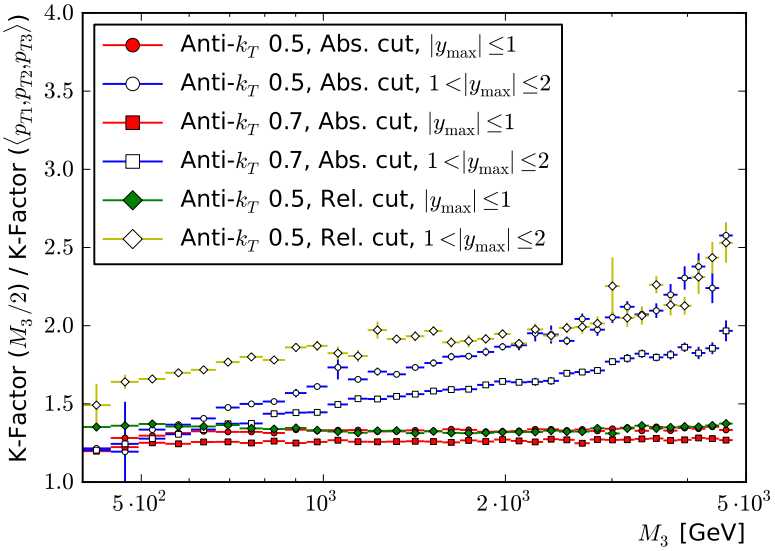


Figure 5.4: Plot of the ratio between the K-factors of the three-jet mass prediction for the Anti-k_T 0.5 and 0.7 jet algorithm. The results are from the calculations using the absolute cut scenario in two different rapidity bins. This demonstrates that the scale choice becomes more important in the outer rapidity bins. This also shows that the influence of the jet size on the NLO corrections is independent of the three-jet mass for a proper scale choice ($m_3/2$).



(a) Plot of the K-factors determined in the relative cut scenario with the Anti-k_T 0.5 jet algorithm. For the $p_{T,\text{avg}}$ scale, the leading order calculation needs to be corrected downwards, while for the $m_3/2$ scale, the LO result is corrected upwards.



(b) This figure shows the ratio between the K-factors determined with $m_3/2$ and $p_{T,\text{avg}}$. In the inner rapidity bins (filled markers), the difference can be described by a simple correction factor. The outer rapidity bins show different shapes for the two scale choices.

Figure 5.5: Overview of the scale dependence of the three-jet mass prediction. These plots demonstrate, how the scale choice directly influences the sign and shape of the next-to-leading order correction.

5.2 Uncertainties

The theory prediction for the three-jet mass measurement is afflicted with five major uncertainty sources. This section describes the statistical error, scale uncertainties, the PDF and α_s uncertainties. The uncertainty on the non-perturbative corrections is discussed in section 5.3.

For each uncertainty source, the complete covariance matrix is available in appendix D. A tabular overview of the theory uncertainties is given in table 5.1. Figure 5.6 shows an overview of the theoretical uncertainties.

$400 - 3000\,\mathrm{GeV}$	Min	Average	Max
Scale uncertainty	$^{+3\%}_{-10\%}$	$^{+6\%}_{-12\%}$	$^{+9\%}_{-14\%}$
PDF uncertainty	$^{+2\%}_{-1\%}$	$^{+3\%}_{-3\%}$	$^{+5\%}_{-5\%}$
α_s uncertainty	$^{+4\%}_{-4\%}$	$^{+4\%}_{-4\%}$	$^{+4\%}_{-4\%}$
NP uncertainty	$^{+2\%}_{-2\%}$	$^{+2\%}_{-2\%}$	$^{+5\%}_{-5\%}$
Statistical uncertainty	$^{<+1\%}_{<-1\%}$	$^{<+1\%}_{<-1\%}$	$^{<+1\%}_{<-1\%}$

Table 5.1: Overview of the theoretical uncertainties for the NNPDF 2.1 - NNLO prediction using the Anti-k_T 0.7 jet algorithm and the absolute cut scenario in the innermost rapidity bin.

5.2.1 Statistical uncertainties

The calculation of the fastNLO tables with NLOJet++, which allow the extraction of the theory prediction, has a very high computational complexity. In order to handle these computing requirements, the task to calculate a given fastNLO table is split into a large number of jobs n_{jobs}. These statistically independent jobs, which are combined to form the final result, can be interpreted as a toy sample of NLO calculations.

To get the final result, the tables produced by fastNLO are combined to form a single table. This table is used to determine the mean value $\bar{m}_{3.i}$ for the three-jet mass prediction. The statistical error on the NLO prediction is given by the uncertainty on the arithmetic mean of the jobs. Therefore, the statistical uncertainty can be calculated with

$$\Delta m_{3,i} = \sqrt{\frac{1}{n_{\mathrm{jobs}}(n_{\mathrm{jobs}}-1)} \sum_{k=1}^{n_{\mathrm{jobs}}} \left(m_{3,i}^{k} - \bar{m}_{3.i}\right)^2}$$

which uses the three-jet mass cross-section prediction $m_{3,i}^{k}$ for the three-jet mass bin i calculated in processing job number k.

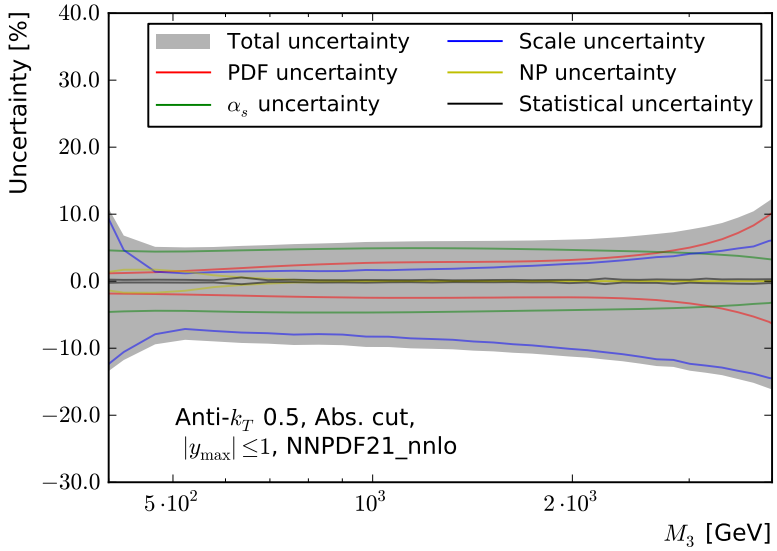


Figure 5.6: Breakdown of the theory uncertainties for the NNPDF 2.1 - NNLO prediction of the three-jet mass using the Anti-k_T 0.5 jet algorithm and the absolute cut scenario in the innermost rapidity bin. The lower uncertainty is dominated by the scale uncertainty, while the upper uncertainty derives from the PDF and α_s uncertainty.

5.2.2 Scale uncertainties

In order to perform the next-to-leading order cross section calculations, a factorisation and renormalisation scale has to be chosen. If it were possible to perform the calculation at all orders, the result would not depended on either scale. However with a truncated perturbative series, the scale dependencies of the final result does not cancel. This means that the observable can react to changes in the size of the scale.

The scale μ should reflect the energy scale of the interaction at which the calculation should be performed.. For this study, two scales were considered - the average transverse momentum of the three leading jets $p_{T,\text{avg}} = \langle p_{T,1}, p_{T,2}, p_{T,3} \rangle$ and half the three-jet mass $m_3/2$. The asymmetric uncertainty $\Delta m_{3,i}^{\text{up/down}}$ on the i-th three-jet mass bin due to the selection of the event scale is simply the envelope of the theory calculations $m_{3,i}^{(k)}$ performed for n different factorisation and renormalisation scales, while taking $(\mu_R, \mu_F) = (\mu, \mu)$ as the central value $m_{3,i}^{(\mu,\mu)}$:

$$\begin{aligned} \Delta m_{3,i}^{\text{up}} &= \max_n \left(m_{3,i}^{(k)} - m_{3,i}^{(\mu,\mu)}, 0 \right) \\ \Delta m_{3,i}^{\text{down}} &= \max_n \left(m_{3,i}^{(\mu,\mu)} - m_{3,i}^{(k)}, 0 \right) \end{aligned}$$

There are two widely used ways to perform the scale variation. The two point variation uses the scales:

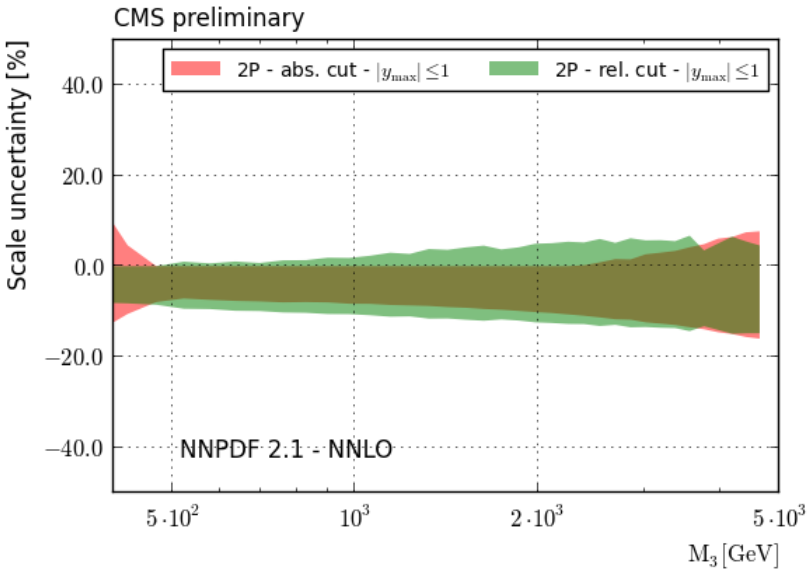
$$(\mu_R, \mu_F) = (2 \times \mu, 2 \times \mu), (\mu/2, \mu/2)$$

The six point variation takes additional permutations into account, while avoiding relative scale factors of 4 between the two scales:

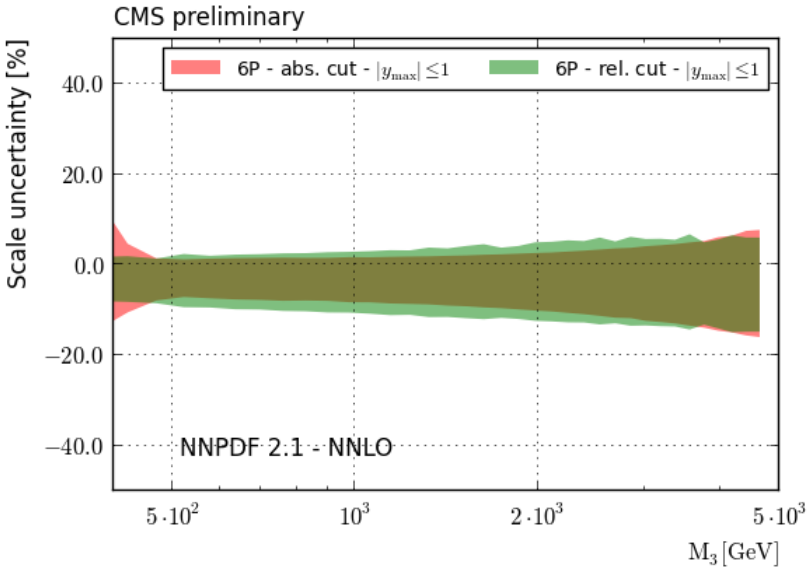
$$(\mu_R, \mu_F) = (2 \times \mu, \mu), (2 \times \mu, 2 \times \mu), (\mu, 2 \times \mu), (\mu/2, \mu), (\mu/2, \mu/2), (\mu, \mu/2)$$

The difference between the two variation methods is shown in figure 5.7 for the inner rapidity bin. In case of the two-point scale variation of the prediction with the absolute cut in the inner rapidity bin, it is apparent that both scale variations cause a downward fluctuation in the calculation, leading to an upper uncertainty of 0%. However with the inclusion of the 4 other scale variations, a component influencing the upper uncertainty is introduced. The final results for both cut scenarios with the Anti-k_T 0.5 jet algorithm are presented in figure 5.8. It is remarkable that the scale uncertainty always has a small upper boundary of up to 10%, while the lower boundary is much larger and on the order of up to 20%. In addition, the

uncertainties for different cut scenarios, rapidities and PDFs are very close together. However in general, the scale uncertainty of the inner rapidity bin is smaller than the uncertainty of the outer rapidity bin. Figure 5.9 presents the differences between the two available jet sizes. It can be observed that the NLO prediction for the larger jet size also exhibits a larger scale uncertainty. The rapidity-dependent behaviour however stays the same. The results of figure 5.10 further demonstrate that using $m_3/2$ as the default choice for the theory prediction is a good decision. Especially in the outer rapidity region, the prediction using $p_{T,\mathrm{avg}}$ is afflicted with very large, asymmetric uncertainties.

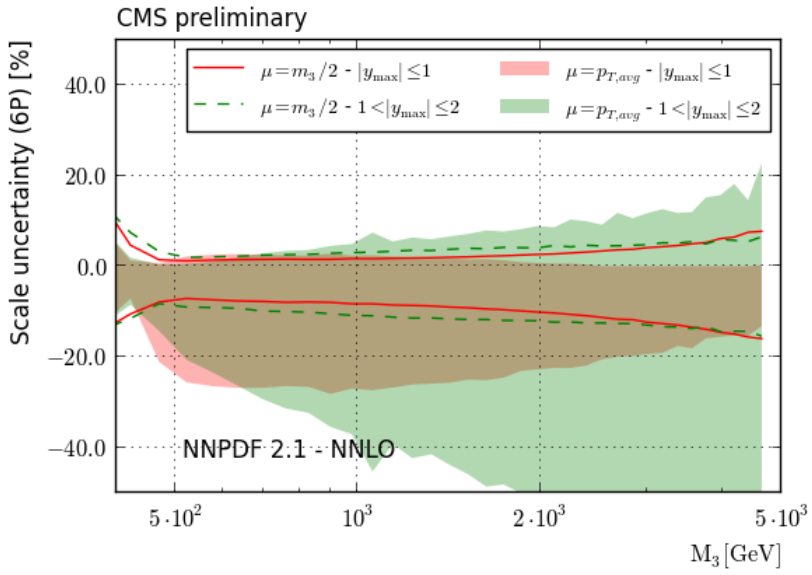


(a)

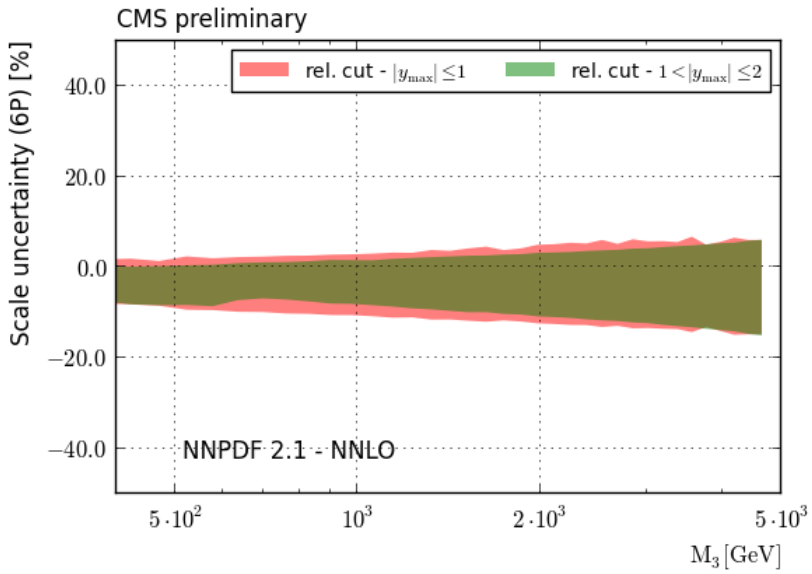


(b)

Figure 5.7: This figure shows the scale uncertainty of the three-jet mass distribution for the NNPDF 2.1 - NNLO PDF with Anti-k_T 0.5 jets for the inner rapidity region. The uncertainties were calculated using both the two- (top) and six-point (bottom) variation to show the difference between the two methods. While both cut scenarios are used, only results from the inner rapidity bin are presented. The six-point variation also introduces addition upper uncertainties in the lower three-jet mass range, which are not present in the two-point variation.

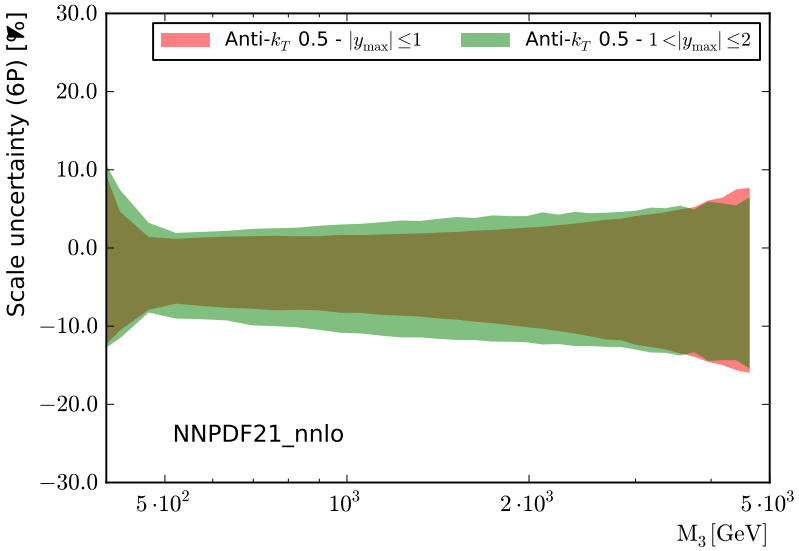


(a) Absolute cut

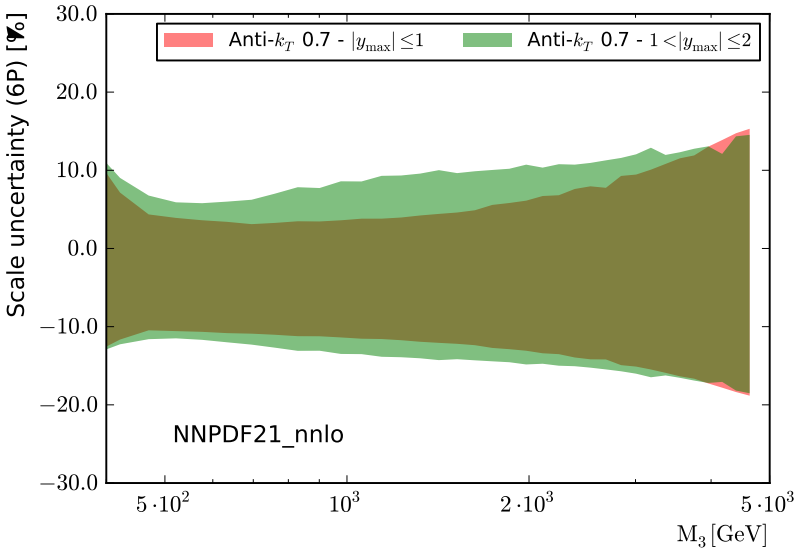


(b) Relative cut

Figure 5.8: Scale uncertainty of the three-jet mass distribution for the relative cut and absolute cut with Anti-k_T 0.5 jets. The absolute cut scenario, the uncertainty in the outer rapidity region is larger than the inner region. For the relative cut scenario, the scale uncertainty relationship is swapped - with the uncertainty in the inner region being larger than in the outer region.

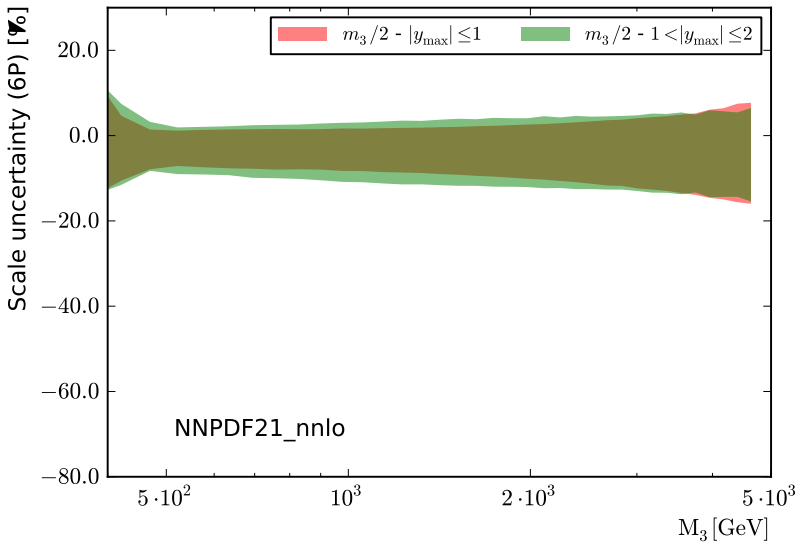


(a) Anti-k_T 0.5

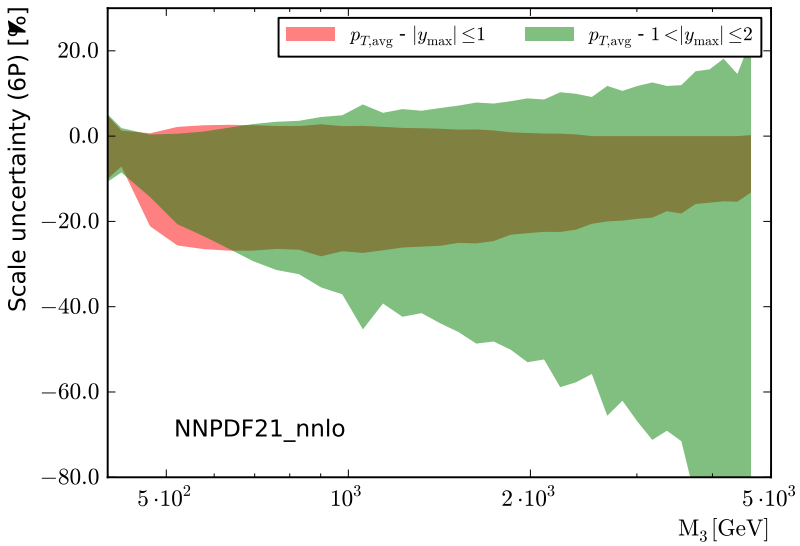


(b) Anti-k_T 0.7

Figure 5.9: Overview of the differences in the scale uncertainty of the three-jet mass distribution for Anti-k_T 0.5 and 0.7 jets. In both cases, the absolute cut scenario was applied. The scale uncertainties for the bigger jet size are larger for both rapidity regions.



(a) $m_3/2$



(b) $p_{T,\text{avg}}$

Figure 5.10: This figure shows the differences in the scale uncertainty of the three-jet mass distribution for Anti-k_T 0.5 jets for different choices of the scale. In both cases, the absolute cut scenario was applied. The average p_T scale is afflicted with very large scale uncertainties, while the $m_3/2$ scale has much smaller scale uncertainties.

5.2.3 PDF uncertainties

The parton distribution functions entering the theory calculation are based on global fits to experimental data. Each dataset has its own set of uncertainties. For recent PDFs, these experimental uncertainties are propagated to uncertainties on the PDF. In addition, there are modelling and parametrisation uncertainties caused by assumptions about the shape and evolution of the PDF distributions.

The following set of PDFs was studied:

CTEQ6.6[95]

The CTEQ 6.6 PDF was obtained from a global fit of hard scattering data from 2008 and uses $\alpha_s(M_Z) = 0.118$ for the α_s evolution. The uncertainties are given in the form of 22 eigenvectors, with an up and downwards fluctuating PDF member along each eigenvector axis. These variations describe the 90% confidence intervals and are rescaled in order to arrive at a common 68% uncertainty band for all PDFs. The rescaling factor can be calculated with the inverse error function using $s = \sqrt{2}\,\mathrm{erf}^{-1}(0.90) \approx 1.645$.

CT10[92]

Compared with CTEQ6.6, the CT10 PDF was extracted from a more recent dataset available in 2010 and includes additional data from the HERA experiment. The eigenvectors are rescaled to the 68% confidence interval like done for CTEQ6.6.

HERAPDF 1.5[96]

The HERAPDF 1.5 PDF is based on data measured by the HERA experiments H1 and ZEUS. The best fit for the strong coupling constant is using $\alpha_s(M_Z) = 0.1176$ with the α_s uncertainty determined from variations with $\alpha_s(M_z) = 0.1156$ and 0.1196. The experimental uncertainties are encoded in the form of 10 eigenvectors for 68% confidence intervals. In addition to these experimental uncertainties, 8 model error PDFs are available, for which the positive and negative difference to the central value is added in quadrature to give the positive and negative model error. The envelope of 4 PDF variations is used to describe the maximal parametrisation variation. This parametrisation variation is quadratically added to the model errors to form

the parametrisation envelope:

$$
\begin{aligned}
(\Delta m_{3,i}^{\text{up}})^2 &= \sum_{k=0}^{8} \left[\max\left(m_{3,i}^{\text{Model}(k)} - m_{3,i}^{\text{central}}, 0\right)\right]^2 \\
&\quad + \left(\max_{k=0}^{4}\left(m_{3,i}^{\text{Param}(k)} - m_{3,i}^{\text{central}}, 0\right)\right)^2 \\
(\Delta m_{3,i}^{\text{down}})^2 &= \sum_{k=0}^{8} \left[\max\left(m_{3,i}^{\text{central}} - m_{3,i}^{\text{Model}(k)}, 0\right)\right]^2 \\
&\quad + \left(\max_{k=0}^{4}\left(m_{3,i}^{\text{central}} - m_{3,i}^{\text{Param}(k)}, 0\right)\right)^2
\end{aligned}
$$

MSTW 2008[97]

MSTW 2008 is the result of a global PDF fit to data from many different sources. The underlying dataset contains inclusive deep inelastic scattering structure functions, low-energy Drell-Yan production, W and Z measurements at the Tevatron experiments and inclusive jet data from HERA and Tevatron. While the MSTW 2008 PDFs are available at LO, NLO and NNLO, only the NLO PDFs were used in this study. The central value of the PDF is is using $\alpha_s(M_z) = 0.12018$. There are PDFs available for both one-sigma (68%) and 90% confidence level limits. However for consistency with the other PDFs, the 68% PDF set was chosen.

NNPDF 2.1[98]

The NNPDF 2.1 PDF fit uses data from deep inelastic scattering, Drell-Yan and Jet data from HERA, Tevatron and other experiments. In this analysis, two PDFs from the NNPDF2.1 family are used, which differ in the perturbative order (NLO and NNLO). Both the NLO and NNLO PDF central prediction use $\alpha_s(M_z) = 0.119$ for the strong coupling constant. However the authors also supply PDFs with a large range of different values for α. The fact that there are variations from $\alpha_s(M_z) = 0.124$ down to $\alpha_s(M_z) = 0.106$ available, makes this PDF very valuable for later data-theory comparisons.

Compared with the other studied PDFs, NNPDF uses a very different approach to encode the PDF uncertainties. The uncertainties are given in the form of an ensemble of n PDF members, also called PDF replicas. The PDFs from the NNPDF 2.1 family usually contain $n = 100$ members. However the central value for the NNLO PDF uses $n = 1000$ and the NNLO PDFs for $\alpha_s < 0.11$ just contain between 9 and 25 members.

The PDF uncertainties can be calculated, taking into account that the central value was determined from the same set of PDF members. The NNPDF uncertainties are defined with the following formula:

$$(\Delta m_{3,i}^{\mathrm{up}})^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left[\max\left(m_{3,i}^{(k)} - m_{3,i}^{\mathrm{central}}, 0\right)\right]^2$$
$$(\Delta m_{3,i}^{\mathrm{down}})^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left[\max\left(m_{3,i}^{\mathrm{central}} - m_{3,i}^{(k)}, 0\right)\right]^2$$

which describes the asymmetric spread of the PDF replicas around the mean. The covariance matrix is calculated from the PDF ensemble like the sample covariance

$$\mathrm{cov}\left(m_{3,i}; m_{3,j}\right) = \frac{1}{n-1}\sum_{k=1}^{n}\left(m_{3,i}^{(k)} - m_{3,i}^{\mathrm{central}}\right)\left(m_{3,j}^{(k)} - m_{3,i}^{\mathrm{central}}\right)$$

PDF uncertainties from eigenvector sets

For the PDFs that encode their uncertainties using PDF eigenvectors, there is a simple recipe for the uncertainty determination. In addition to the central value m_3^{central}, the three-jet mass cross section $m_3^{\mathrm{up/down}(k)}$ is calculated for every PDF eigenvector k. Each pair of PDF eigenvectors represent independent uncertainty sources. The asymmetric PDF uncertainty $\Delta m_{3,i}^{\mathrm{up/down(k)}}$ due to a single pair of PDF eigenvectors k is given by the asymmetric deviation from the central prediction:

$$\begin{aligned}
\Delta m_{3,i}^{\mathrm{up(k)}} &= \max\left(\frac{m_{3,i}^{\mathrm{up}(k)} - m_{3,i}^{\mathrm{central}}}{s}, \frac{m_{3,i}^{\mathrm{down}(k)} - m_{3,i}^{\mathrm{central}}}{s}, 0\right) \\
\Delta m_{3,i}^{\mathrm{down(k)}} &= \max\left(\frac{m_{3,i}^{\mathrm{central}} - m_{3,i}^{\mathrm{up}(k)}}{s}, \frac{m_{3,i}^{\mathrm{central}} - m_{3,i}^{\mathrm{down}(k)}}{s}, 0\right)
\end{aligned}$$

Therefore, the total PDF uncertainty is simply given by the quadratic sum of uncertainties described by each pair of PDF eigenvectors. With this information, it is possible to determine the asymmetric PDF errors $\Delta m_{3,i}^{\mathrm{up/down}}$on the three-jet mass in bin i using:

$$(\Delta m_{3,i}^{\mathrm{up}})^2 = \sum_{k=0}^{n_{\mathrm{EV}}}\left(\Delta m_{3,i}^{\mathrm{up(k)}}\right)^2, \qquad (\Delta m_{3,i}^{\mathrm{down}})^2 = \sum_{k=0}^{n_{\mathrm{EV}}}\left(\Delta m_{3,i}^{\mathrm{down(k)}}\right)^2$$

with the scaling factor s. The scaling factor is 1 for all PDFs except CTEQ6.6 and CT10 as previously discussed. The three-jet mass bin covariance for the eigenvector based uncertainties is based on the symmetrised and rescaled errors $\Delta m_{3,i}^{\mathrm{sym(k)}}$ from the eigenvector variations:

$$\Delta m_{3,i}^{\mathrm{sym(k)}} = \frac{m_{3,i}^{\mathrm{up}(k)} - m_{3,i}^{\mathrm{down}(k)}}{2s}$$

$$\mathrm{cov}\left(m_{3,i}; m_{3,j}\right) = \sum_{k=0}^{n_{\mathrm{EV}}} \left[\frac{m_{3,i}^{\mathrm{up}(k)} - m_{3,i}^{\mathrm{down}(k)}}{2s}\right] \left[\frac{m_{3,j}^{\mathrm{up}(k)} - m_{3,j}^{\mathrm{down}(k)}}{2s}\right]$$

Strong coupling constant

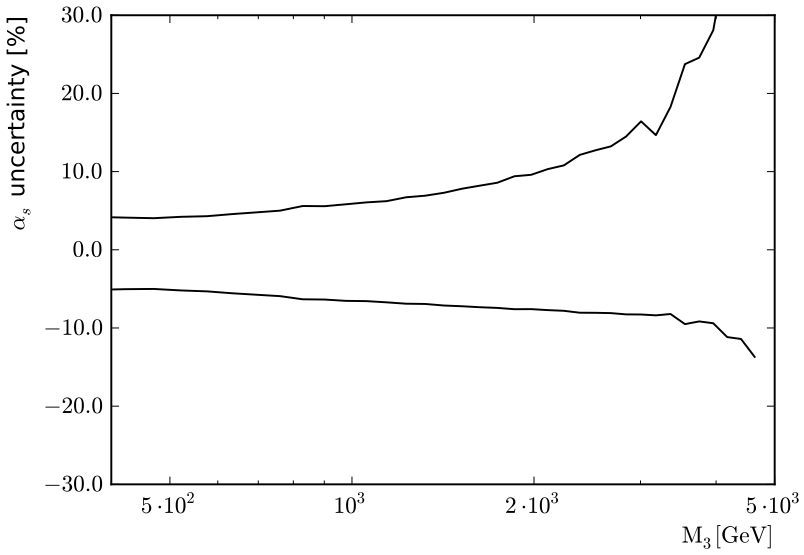
Another source of uncertainty inherent to PDFs is the choice of the strong coupling constant when fitting the PDF. In order to quantise the size of this uncertainty, the PDFs provide a set of α_s variations. This allows to derive the α_s uncertainty of the PDF from the variations of $\alpha_s(M_Z)$ within the uncertainties. These variations use the best α_s fit value from the PDF and the one sigma uncertainty on this value. An example for this uncertainty source is given in figure 5.11 for HERAPDF 1.5, which uses $\alpha_s = 0.1176 \pm 0.002$.

Overview

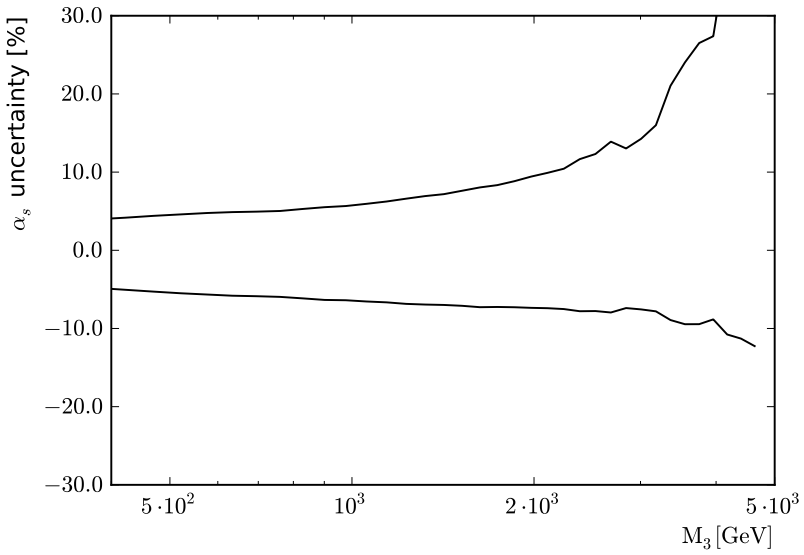
Some PDF uncertainties for the three-jet mass are shown in figure 5.13. As shown in figure 5.12, the theory calculations can show quite large differences for different PDFs. All PDFs except HERAPDF seem to predict the same shape up to a three-jet-mass of about $m_3 = 1000\,\mathrm{GeV}$. Above this threshold, the PDFs predict different shapes for the three-jet-mass distribution, however the shape variation is covered by the large PDF uncertainties in this region.

5.3 Non-perturbative corrections

The result for the three-jet mass distribution from FastNLO and NLOJet++ gives the theory prediction at next-to-leading order of perturbative QCD. However these calculations do not include the effects from hadronization and multiple parton interactions, which can not be described using perturbative techniques. In contrast, the measurement in nature and corrected for detector effects, does contain these effects. To allow a comparison between data and theory, the theory has to be corrected for these influences.

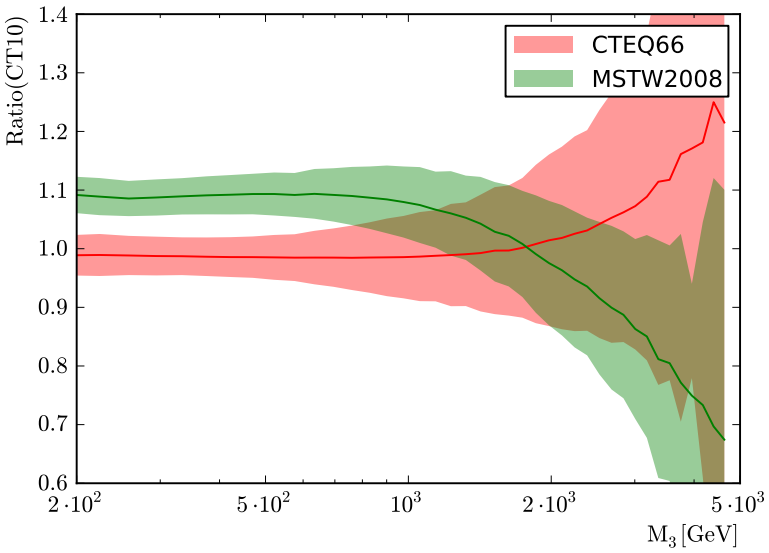


(a) Absolute cut

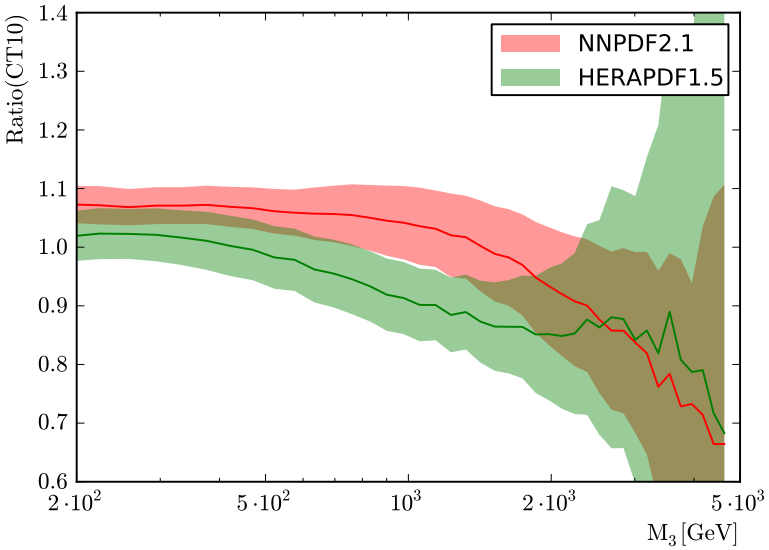


(b) Relative cut

Figure 5.11: Size of the 68 percent confidence interval of the α_s uncertainty for HERAPDF 1.5 with the absolute cut (top) and relative cut (bottom). The uncertainty in both cases shows a small slope, increasing the uncertainty from below five percent at 200 GeV up to ten percent at 2000 GeV. At the same time, the uncertainties are increasingly asymmetric, with large upwards and a small downwards fluctuation for three-jet masses between 3000 GeV and 5000 GeV.

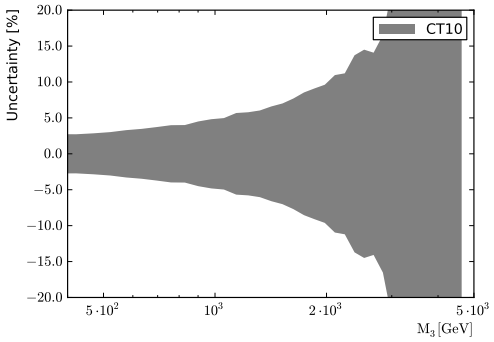


(a)

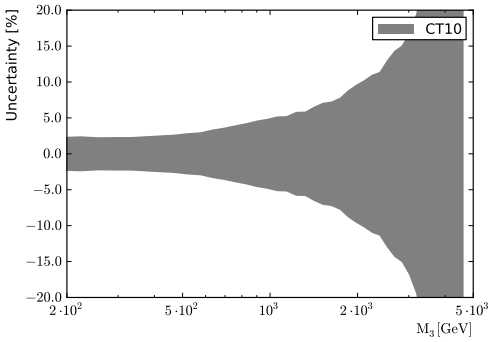


(b)

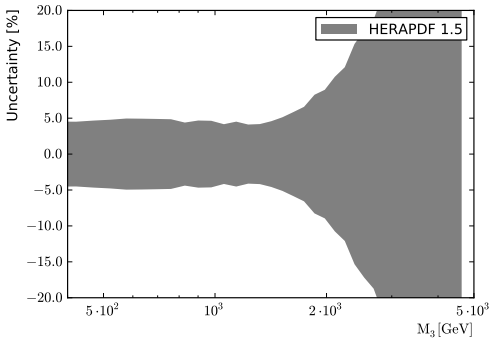
Figure 5.12: Comparison between the PDFs used in this analysis. The PDFs are normalised to the CT10 cross section and the uncertainty band contains the propagated PDF uncertainties on the mean. Except for the HERAPDF result, all PDFs give a similar shape up to a three-jet-mass of about $m_3 = 1000\,\text{GeV}$.



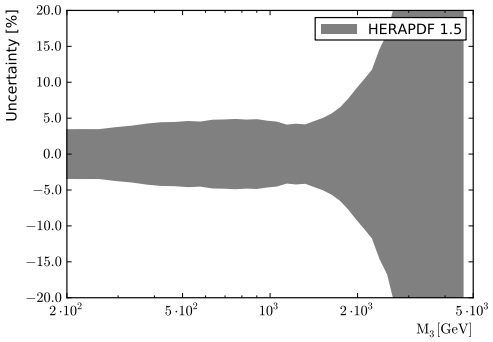
(a) Abs. cut, CT10



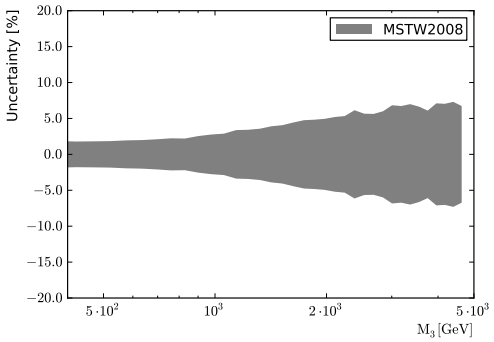
(b) Rel. cut, CT10



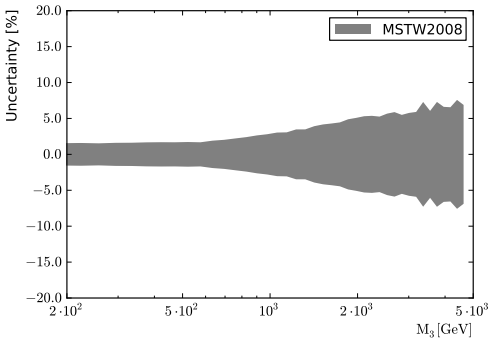
(c) Abs. cut, HERAPDF 1.5



(d) Rel. cut, HERAPDF 1.5



(e) Abs. cut, MSTW2008



(f) Rel. cut, MSTW2008

Figure 5.13: The PDF uncertainties for the absolute cut (left) and the relative cut (right) are shown for CT10 (top), HERAPDF 1.5 (middle) and MSTW2008 (bottom). In general, the PDF uncertainties for the absolute cut scenario are larger than the uncertainties for the relative cut.

Previous studies, like the inclusive jet spectrum measurement[99], estimated the size of these corrections using the Monte Carlo event generators Pythia and Herwig++. However for the study of the three-jet mass, these generators are not able to give a leading-order result due to missing matrix elements. Therefore, to derive correction factors to take care of these non-perturbative effects, the Monte Carlo event generator Sherpa was used. Sherpa is able to generate three-jet events at leading order and simulates hadronization and multiple parton interactions using the built-in models. A comparison of the Sherpa prediction with unfolded data is shown in figure 6.15.

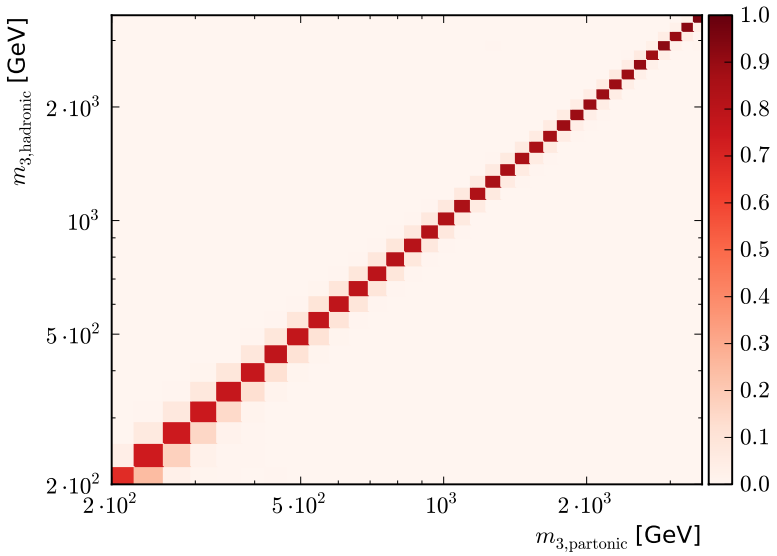
The usual way to derive these correction factors is to look at the ratio between the prediction applying a hadronization and multiple parton interactions model and the prediction without these two effects. This procedure gives bin-by-bin correction factors, which are later applied to the next-to-leading order theory prediction.

A more general approach is to perform the comparison on an event-by-event basis. The three jet mass is determined for a given event before the the hadronization and multiple parton interaction is simulated and after these two steps are done. This enables the creation of a transfer matrix which contains the probability distributions that a partonic three-jet mass $m_{3,p}$ matches a hadronic three-jet-mass $m_{3,h}$. Such a transfer matrix is shown in figure 5.14(a). Naturally, this matrix contains more information about the processes simulated by the non-perturbative models than the bin-by-bin corrections. The first off-diagonal elements are between 2 and 10%. Other elements of the matrix are much smaller than 1%. This shows that the bin correlations due to the non-perturbative corrections are quite small, but they do not vanish.

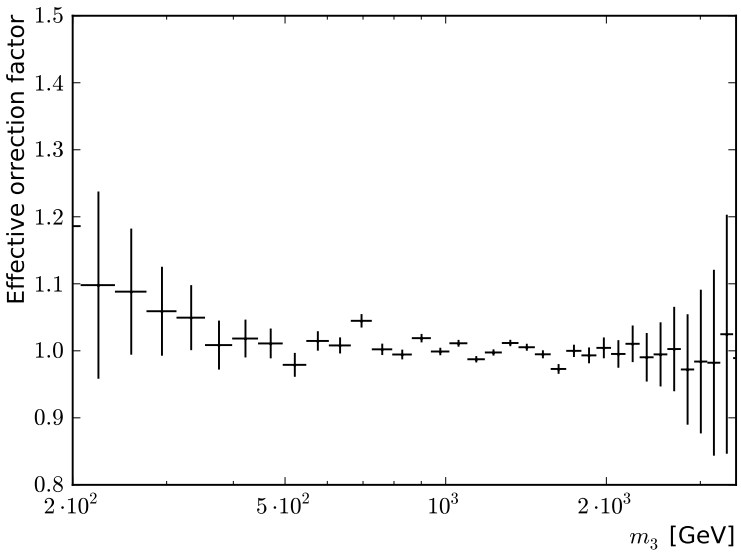
In order to allow this event-by-event comparison, the generator interface in CMSSW was modified to re-initialise the random number generator for each event, so that identical events are generated before the differences in the simulation cause a deviation between the consumed random numbers. The two relevant settings in Sherpa for the determination of the non-perturbative corrections are responsible for the hadronization effects and multiple parton interactions. The exact configuration of these parameters is described in appendix B.4.

Figure 5.14(b) shows the effective bin-by-bin correction factor determined from this new matrix method. The correction factors can be quite large with a 20% correction for the very low three-jet-masses. However above 400 GeV, the corrections are very close to unity. The uncertainty on this effective correction factor is given by the statistical uncertainty, which dominates the large three-jet masses and an estimate of the uncertainty due to the modelling

of the hadronization and underlying event simulation. This uncertainty was conservatively estimated to be half the size of the correction. Therefore this term dominates for the low three-jet masses.



(a)



(b)

Figure 5.14: Overview of non-perturbative corrections. From an event-by-event matching between datasets with different simulation setup, the relationship between the partonic and hadronic final state is derived (top) using the Sherpa model. With this transfer matrix, a set of effective bin-by-bin correction factors can be calculated (bottom). The uncertainty on this result is derived from the statistical uncertainties and an estimate of the modelling systematic.

Chapter 6

Measurement of the three-jet mass

The previous chapters showed that the study of multi-jet events is one of the primary ways to gain a greater understanding of quantum chromodynamics[100, 101]. In particular, events containing three jets can be used to investigate fundamental properties of the theory [102, 103, 104] like the running of the strong coupling constant [105, 106] and it's value at certain scales. The preceding chapter introduced one of the key measurements using these three-jet events: The determination of the cross-section as a function of the invariant mass of the three-jet system.

The chapter begins with the introduction of the datasets and event selection to get clean three-jet events. Using these events, the three-jet mass is measured in data and compared to Monte-Carlo simulation and the next-to-leading order theory predictions derived in chapter 5. In order to allow this comparison, several analysis steps are necessary. In a first step, the three-jet mass resolution is measured to determine an optimal binning. Detector effects influencing the three-jet mass measurement are removed using an unfolding technique to facilitate the comparison with different experiments. The measurement is completed with a study of the systematic uncertainties on the three-jet mass. The chapter closes with a discussion of the differences between data and theory.

6.1 Datasets

This analysis is based on proton-proton collisions collected by the CMS detector at the LHC accelerator. Since the study works with jet objects, the "Jet" primary dataset (PD) was chosen as basis. This dataset contains all events

Table 6.1: Overview of the used datasets

Sample name	Run range	Dataset Bookkeeping System entry
Data - 1	160329-163869	`/Jet/Run2011A-May10ReReco-v1/AOD`
Data - 2	165088-168229	`/Jet/Run2011A-PromptReco-v4/AOD`
Data - 3	170053-172619	`/Jet/Run2011A-05Aug2011-v1/AOD`
Data - 4	172620-175770	`/Jet/Run2011A-PromptReco-v6/AOD`
Data - 5	175832-180296	`/Jet/Run2011B-PromptReco-v1/AOD`

which triggered one of the CMS jet triggers and is split into multiple samples corresponding to different periods of data taking. There are three samples which utilise the reconstruction done at the detector site during data taking and two reprocessed samples. The two reprocessed samples profit from improved understanding of the detector alignment and calibration. Together, these five samples cover the whole data taking period between March and October 2011 when the LHC was running with a centre-of-mass energy of 7000 GeV.

The same dataset was also the subject of detailed detector performance studies, which gave rise to precise uncertainty estimates used later in the analysis. The list of used datasets is given on table 6.1.

6.1.1 Monte Carlo datasets

For Monte-Carlo studies, three different event generators were employed in this analysis. An overview of the used Monte-Carlo samples is given in section B.1.1. The determination of the non-perturbative corrections is based on private samples produced with the Sherpa Monte-Carlo generator. This program uses $2 \to 2$ and $2 \to 3$ leading order QCD matrix elements with a matched parton shower algorithm to generate events. Further details about the setup of the generator is given in section 5.3.

For detector response measurements, two samples from the official CMS production with the Pythia 6 and Herwig++ multi-purpose event generators are chosen. Both Monte-Carlo event programs generate the hard interaction with a $2 \to 2$ leading order QCD matrix element and use a parton shower algorithm to simulate higher jet multiplicities. The generators are configured to apply a phase-space reweighting on matrix-element level w_{ME} during the generation process. This allows to cover the complete phase space with a single sample. However during the analysis, the individual events have to be weighted by $1/w_{\mathrm{ME}}$ to reproduce the physical spectrum. The reweighting

factor is a function of the transverse momentum $\hat{p}_T$ of the outgoing two partons in the rest-frame and described by the following formula:

$$w_{\mathrm{ME}} = \left(\frac{\hat{p}_T}{15\,\mathrm{GeV}} \right)^{4.5}$$

To avoid issues with large matrix element weights, the two samples also apply a phase space cut of $15\,\mathrm{GeV} < \hat{p}_T < 3000\,\mathrm{GeV}$.

For the simulation of the underlying event, the Pythia6 sample uses the Z2 tune[107], while the Herwig++ sample uses the default tune[108]. The generated events are processed through the full CMS detector simulation with detector conditions closely resembling the observed state instead of the design specifications. These detector conditions also contain a model for pile-up collisions, where the simulation introduces additional scattering events according to a predetermined pile-up profile.

In an ideal case, this simulated pile-up profile ($N_{\mathrm{MC}}(N_{\mathrm{PU,truth}})$) is identical to the estimated pile-up profile $N_{\mathrm{Data}}(N_{\mathrm{PU,est}})$ in data. However as shown in figure 6.1, there are differences between these two profiles. It is therefore necessary to apply a weight w_{PU} to match the pile-up distribution in the simulation to the profile in data. The pile-up distribution in data is estimated based on the instantaneous luminosity averaged over all colliding bunches, the total luminosity and the standard deviation of the individual bunch luminosities during each luminosity section. Since the true value of pile-up events in the simulation is known, a simple bin-by-bin reweighting of the pile-up distributions is sufficient.

$$w_{\mathrm{PU}}(N_{\mathrm{PU,truth}}) = \frac{N_{\mathrm{Data}}(N_{\mathrm{PU,est}}) / \sum N_{\mathrm{Data}}}{N_{\mathrm{MC}}(N_{\mathrm{PU,truth}}) / \sum N_{\mathrm{MC}}}$$

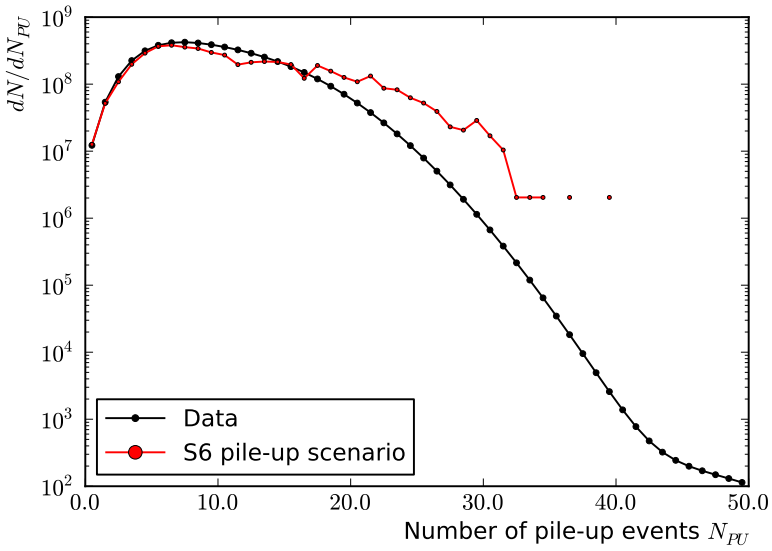


Figure 6.1: Pile-up distribution in simulation in comparison to the estimate in data. The simulated curve is normalised to the estimated data distribution. The simulated events are reweighted to reproduce the estimated distribution. The plot shows that the default pile-up scenario used for the simulation cannot be used to describe data with more than 35 pile–up collisions.

6.2 Event selection

6.2.1 Data selection

The first step in the event selection is to pick good runs and luminosity sections (LS) from the recorded data, in which all detector subsystems are properly setup and work reliably. During a data taking period, some sub-detectors can encounter problems which makes it necessary to reject all recorded data within a luminosity section. These problematic luminosity sections are identified by the data quality management team of the CMS collaboration during the data validation process. The list of validated runs and luminosity sections is made public in the form of a JSON[1] file and was used as the basis for the further steps.

Since this study uses the new pixel reconstructed CMS luminosity measurement (detailed in [109, 110]) in contrast to the old HF based luminosity measurement, the $\approx$ 156.000 certified luminosity sections were exposed to additional checks. The first check is to scan the pixel luminosity database for invalid or empty luminosity entries. 1990 entries were removed from the certificated set of luminosity sections because of this cross check.

Another cross check is performed by comparing the trigger information as stored in the event data files with the trigger stored in the luminosity database. This check uncovered 192 luminosity sections with inconsistent trigger data for the single-jet trigger paths. To avoid any ambiguity, these luminosity sections were removed from the list of analysed luminosity sections as well.

The complete list of selected runs and their luminosity sections is given in appendix B.2.

6.2.2 Trigger

In order to reconstruct the three-jet mass spectrum, it is necessary to select an appropriate set of triggers. For this study the single-jet triggers were used. The family of single-jet triggers give a positive trigger decision when the jets reconstructed by the L1 and HLT part of the data acquisition system overstep certain transverse momentum thresholds. To illustrate the single-jet trigger decision, consider a jet object reconstructed by the L1 trigger system with $p_{T,\mathrm{L1}} = 24\,\mathrm{GeV}$. This object is above the p_T threshold of the L1SingleJet16 trigger ($p_T > 16\,\mathrm{GeV}$) and therefore the event is forwarded to the HLT reconstruction. In this example, the much more detailed reconstruction at this level yields a jet object with $p_{T,\mathrm{HLT}} = 31\,\mathrm{GeV}$. Since this HLT trigger

[1] JavaScript Object Notation

object is above the threshold of the HLT_Jet30 trigger path, a positive trigger decision is given and the event is recorded for further processing.

Trigger thresholds

In order to combine the data recorded by different triggers, mutually exclusive regions of the phase space are identified in which a single trigger path is fully efficient with respect to its definition. To define this active domain for each trigger path, it is necessary to determine the trigger efficiency. The first step is to measure the distribution of a relevant observable for data taken from each trigger path. The next task for the evaluation of the trigger efficiency is to look at the ratios of distributions belonging to two trigger paths. This allows to define a so called turn-on point, at which the efficiency is above a certain threshold. However for the jet triggers, prescale differences between adjacent trigger paths can be between one and two orders of magnitude in size. In addition, the two samples are in general statistically independent and the relevant area of phase space only contains a small number of events for the lower trigger path. Due to these issues, statistical fluctuations can play an important role in the trigger efficiency measurement. However such large statistical uncertainties can poses a problem for the determination of the trigger turn-on points.

To avoid these problems, the trigger thresholds of the single-jet triggers were determined using a technique which measures the trigger efficiency from a set of events $S_1 = \{E_i | T_\mathrm{A}(E_i) = \mathrm{true}\}$ which was triggered by a single trigger path T_A. From this set of events S_1, the subset $S_2 = \{E_i | T_\mathrm{A}(E_i) \wedge T_\mathrm{B}(E_i)\}$ of events also belonging to the next trigger T_B with a higher threshold is determined. This means that the analysis has to emulate the trigger decision to determine this subset $S_2 \subseteq S_1$. The advantage of this method is that the prescale of the trigger T_B with the higher threshold does not enter the determination of the trigger prescale at all, since only events from trigger path T_A are used. Mathematically, the efficiency can be expressed using some distribution $N(S, x)$ of observable x for events from set S :

$$f_\mathrm{eff}(x) = \frac{N(\{E_i | T_\mathrm{A}(E_i) \wedge T_\mathrm{B}(E_i)\}, x)}{N(\{E_i | T_\mathrm{A}(E_i) = \mathrm{true}\}, x)}$$

In order to reproduce the trigger decision for other trigger paths, the L1 and HLT trigger objects, which were used in the original trigger decision are accessed. Since these trigger collections were changed over time, great care has to be taken to track the correct trigger object collection. For each event, it is checked whether these trigger objects would also fire the next highest trigger path. With the two sets of events available for a given trigger

Trigger	L1 threshold	HLT threshold	$p_{T,99\%}$
HLT_Jet30	16 GeV	30 GeV	50 GeV
HLT_Jet60	36 GeV	60 GeV	85 GeV
HLT_Jet80	52 GeV	80 GeV	110 GeV
HLT_Jet110	68 GeV	110 GeV	144 GeV
HLT_Jet150	92 GeV	150 GeV	192 GeV
HLT_Jet190	92 GeV	190 GeV	238 GeV
HLT_Jet240	92 GeV	240 GeV	294 GeV
HLT_Jet300	92 GeV/128 GeV	300 GeV	355 GeV
HLT_Jet370	92 GeV/128 GeV	370 GeV	435 GeV
HLT_Jet800	92 GeV/128 GeV	800 GeV	873 GeV

Table 6.2: Overview of the trigger thresholds and turn-on points $p_{\mathrm{T},99\%}$ for the single-jet trigger paths. In some trigger menus, the L1 trigger thresholds for the highest single jet triggers were increased from 92 GeV to 128 GeV.

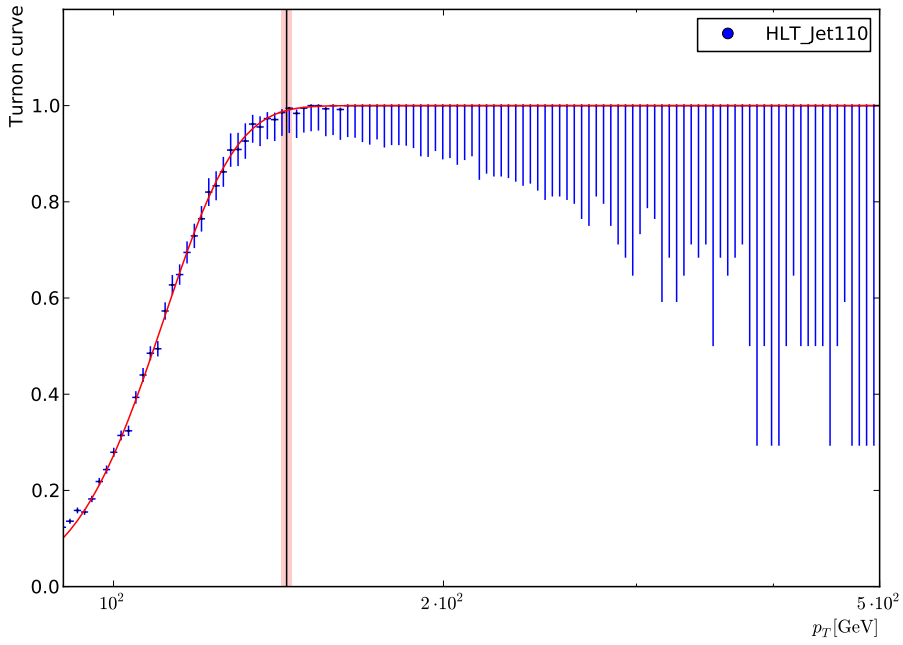
path, the trigger efficiency can be calculated using binomial statistics. The errors on the trigger efficiency are given by the Clopper-Pearson confidence intervals. The resulting trigger efficiency is then fitted with the function:

$$f_{\mathrm{eff}}^{\mathrm{fit}}(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right)$$

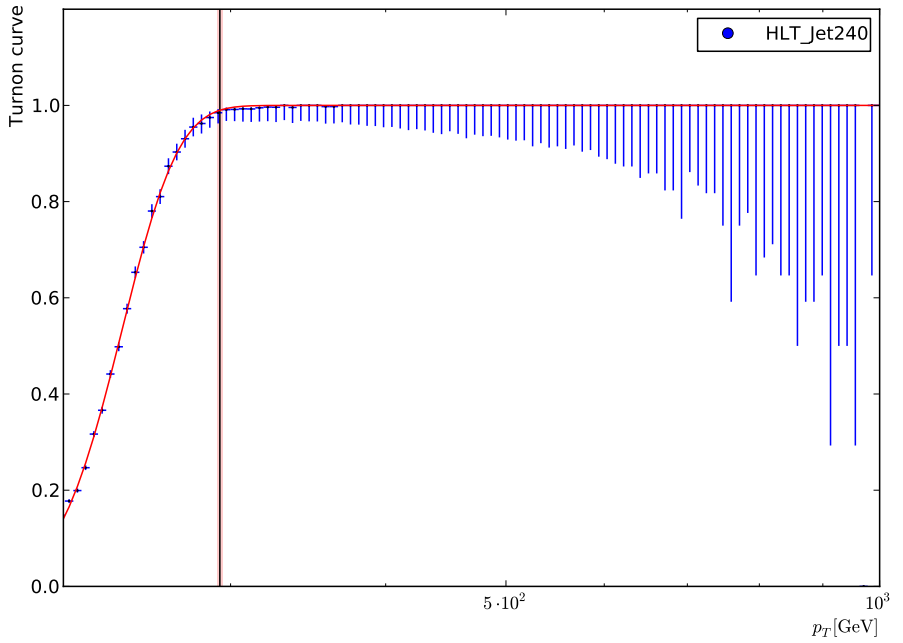
The start values of the fit for μ are determined from the point x_{50} where the trigger efficiency reaches the 50% threshold, while σ is set to 1. The fit is performed in two steps to increase numerical stability. The first step fits the function in the range between x_{50} and infinity. In the second iteration the fit is performed with a lower bound of $\mu - \sigma$. From the resulting efficiency curve, the 99% threshold is calculated and used as turn-on point for the trigger. An overview for the transverse momentum turn-on points of the single-jet trigger is presented in table 6.2. The fit of the turn-on curve is shown in figure 6.2 for two exemplary trigger paths.

Trigger composition

As shown above, the single-jet trigger paths are able to reach full efficiency to measure jets with transverse momentum above a certain threshold. However an individual single-jet trigger path is unable to reach full efficiency as a function of the three-jet mass. In the case of the lowest trigger HLT_Jet30, this is due to prescales which cause the trigger to record very few events



(a)



(b)

Figure 6.2: Turn-on curves for the HLT_Jet110 and HLT_Jet240 trigger path with trigger efficiency errors based on Clopper-Pearson confidence intervals. The trigger turn-on (error bars) is fitted with an analytical function (line) to determine the 99% trigger threshold. This trigger threshold is indicated by the vertical line. The light grey band indicates the uncertainty on the turn-on point.

with large three-jet masses. The other single-jet triggers, by definition, apply transverse momentum cuts which are larger than the analysis cuts and thereby are unable to reach full efficiency. It is therefore necessary to combine multiple single-jet triggers to reconstruct the complete three-jet mass distribution. The trigger path composition of the final three-jet mass measurement is shown in figure 6.3.

The combined trigger efficiency can be determined using a Monte-Carlo study, which uses the trigger turn-ons and trigger composition found in data. This study works by using the trigger composition to determine the probability for an event to belong to a certain trigger path. The fitted trigger turn-on for this path gives the probability for accepting the leading jet of the event. A comparison between the three-jet mass distribution with and without these two steps gives the combined trigger efficiency as a function of the three-jet mass. This efficiency is above 99% for all three-jet masses above the theoretically motivated lower three-jet mass cut $m_3 > 150\,\text{GeV}$.

Trigger menu adaptation

The CMS trigger is very flexible and allows the trigger menu to change between runs and prescales between luminosity sections. This implies that not all runs are taken with the same trigger menu. In order to benefit from trigger paths with low prescales, the analysis adapts the used paths on a luminosity section by luminosity section basis to use all available single-jet triggers. During the 2011 data-taking period, 10 different trigger path combinations were used. The information about the trigger paths and the corresponding trigger menus is given in section B.3.2. In total $26,690,474$ events corresponding to $4,984\,\text{pb}^{-1}$ pass the single-jet HLT selection.

6.2.3 Beam scraping filter

Interactions between the circulating proton bunches and both residual gas particles and the beam collimators can produce showers of secondary particles which are registered in the CMS detector as events with heavy, isotropic noise.

To remove this effect, for all events with more than 10 tracks, the purity of the tracks is required to meet a certain threshold. This purity requirement is based on the ratio of tracks classified as high-purity tracks[111] over the total number of reconstructed tracks and must exceed 25%. The discriminating variables of this filter are shown in figure 6.4.

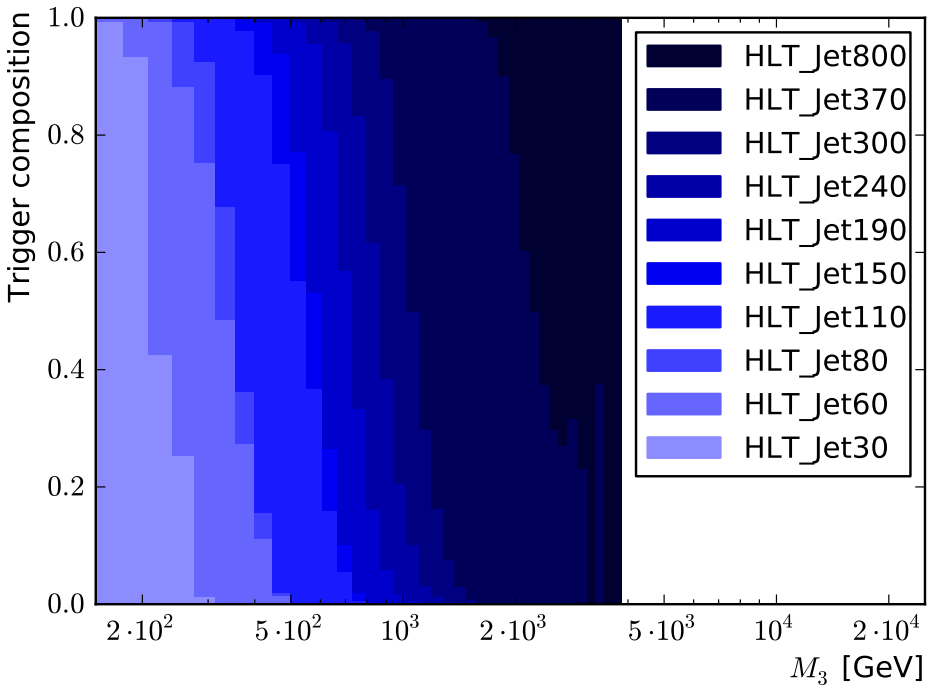
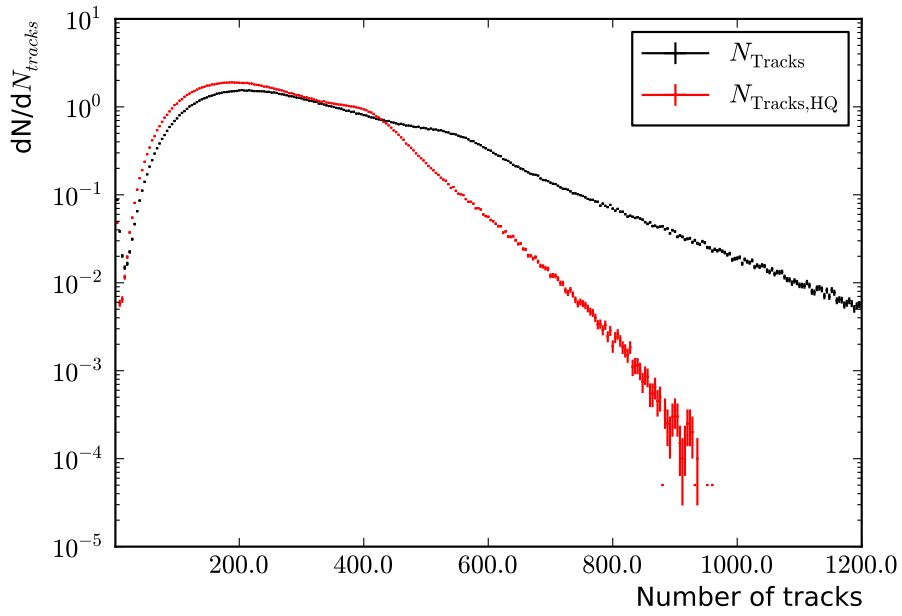


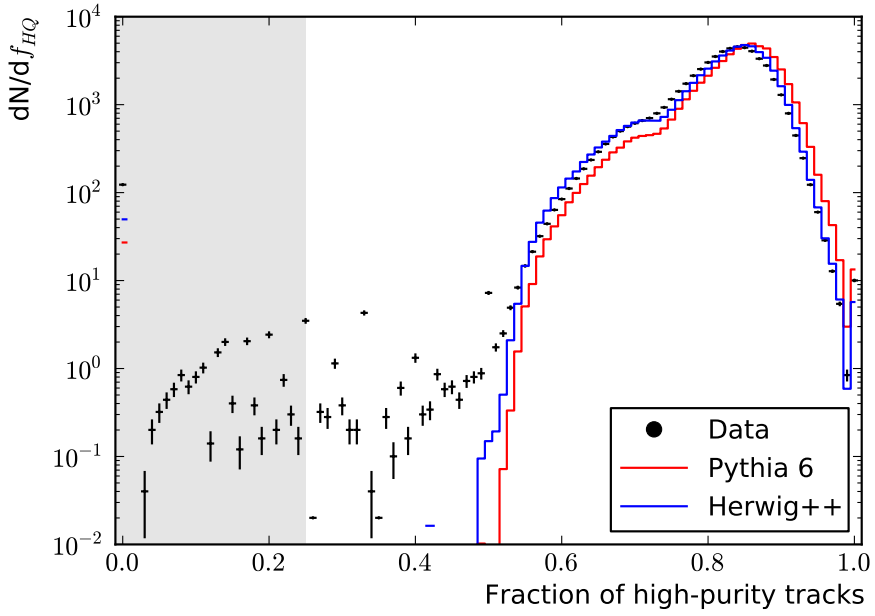
Figure 6.3: Trigger composition of the final three-jet mass measurement. For each three-jet mass bin, the fraction of events coming from a certain trigger path is shown. An individual single-jet trigger path is unable to reach full efficiency as a function of the three-jet mass, either due to prescales or the inherent trigger cuts.

Dataset	Data - 1	Data - 2	Data - 3	Data - 4	Data - 5
N_{rec}	9,166,798	16,501,507	3,764,807	4,996,600	13,358,914
N_{run}	7,219,128	11,649,876	2,422,639	4,541,148	11,964,609
R	122	103	47	40	147
LS	26,679	46,822	12,343	20,757	49,694
L	$215.5\,\mathrm{pb}^{-1}$	$954.7\,\mathrm{pb}^{-1}$	$389.2\,\mathrm{pb}^{-1}$	$703.3\,\mathrm{pb}^{-1}$	$2721.7\,\mathrm{pb}^{-1}$
N_{trg}	5,078,401	8,273,797	1,637,747	3,053,394	8,647,135

Table 6.3: Cut flow for the different data samples. N_{rec} : Number of recorded events, N_{run} : Number of events passing the run and luminosity section selection, R : Number of selected runs, LS : Number of selected luminosity sections, L : integrated luminosity, N_{trg} : Number of events passing any single-jet trigger.



(a)



(b)

Figure 6.4: In the top figure, the total number of tracks and high-purity tracks for events triggered by HLT_Jet30 path is shown. The bottom plot shows the fraction of high-purity tracks, which is used in the beam scraping filter. The region removed by the filter is shaded.

6.2.4 Vertex selection

To reject further beam backgrounds and off-centre parasitic bunch crossings, the recommended vertex selection cuts are applied. To pass the step of the selection process, an event has to contain at least one well reconstructed primary vertex (PV), within a distance of $|z(\mathrm{PV})| < 24\,\mathrm{cm}$ between the primary vertex and the nominal interaction point (IP) of the detector. The radial distance $\rho(\mathrm{PV})$ of the vertex from the z axis though the IP is required to be smaller than 2 cm, corresponding to the size of the beam pipe. Additionally, the fit used to determine the vertex needs to have at least 4 degrees of freedom. For an unconstrained vertex fit, which has $2\sum_i w_i - 3$ degrees of freedom ($w_i \leq 1$), this means at least four tracks with weight ≈ 1 are required. The influence of the different steps of the vertex selection is shown in table 6.5.

6.2.5 Anomalous HCAL noise filter

The so called HBHE filter removes anomalous signal events caused by HCAL noise. Anomalous HCAL noise does not arise from electronic (pedestal) noise but rather from instrumentation issues associated with the photo-diodes and readout boxes. Since the noise manifests in very high energy and occupies multiple towers, it is recommended to reject these events instead of trying to just remove the noisy towers. The filter to remove this kind of events is based on timing information, the number of hits per readout box and photo-diode, the pulse shape and isolation criteria. The set of applied cuts is shown on table 6.4. A detailed discussion for the causes of this noise and the filter strategy can be found in [112]. Starting at 0.1% for HLT_Jet30, this filter removes an increasing amount of events for higher single-jet thresholds of up to 8% of events for the HLT_Jet370 trigger. With a 72% rejection rate, this noise filter removes a very large amount of events triggered by the HLT_Jet800 path. However this trigger path in particular was active only in a small number of runs and therefore the large noise ratio has very little impact on the final result.

6.2.6 Missing transverse energy cut

To further enhance the QCD event purity, events where the missing transverse energy $\not\!\!E_T$ constitutes a large fraction of the measured total transverse energy, are removed in analogy to other QCD analyses[99]:

$$\not\!\!E_T / \sum_i E_{T,i} < 0.5$$

min E^2over 10TS	$> -999\,\mathrm{GeV}^2/\mathrm{TS}$	min 25 GeV hit time	$> -9999\,\mathrm{GeV}$
max E^2 over 10TS	$< 999\,\mathrm{GeV}^2/\mathrm{TS}$	max 25 GeV hit time	$< 9999\,\mathrm{GeV}$
N(HPD hits)	< 17	N(isolated noise channels)	< 10
N(HPD without hits)	< 10	$\sum_{\text{isolated noise}} E$	$< 50\,\mathrm{GeV}$
N(Zeros)	< 10	$\sum_{\text{isolated noise}} E_T$	$< 25\,\mathrm{GeV}$

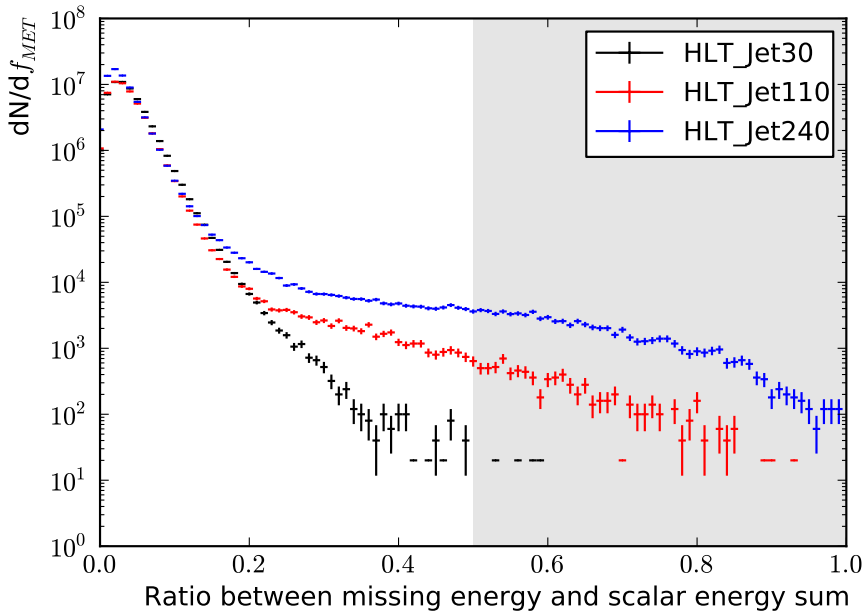
Table 6.4: List of the discriminating variables and their thresholds used by the HCAL noise filter.

This cut primarily removes normal calorimeter noise, beam halo and cosmic-ray backgrounds mimicking real jets. As shown in figure 6.5, this background increases for higher energy scales. Depending on the trigger path, this selection step removes between 0.1% and 0.6% of events compared to the previous step for the single-jet triggers up to HLT_Jet370. For the highest jet trigger it removes 6% of events.

6.2.7 Jet ID selection

The jet identification (jet ID) selection criteria[113] were developed in order to reject pure noise or noise enhanced jets, while keeping more than 99% of physics jets above 10 GeV. Contrary to the previous selection criteria, which rejected complete events, the jet ID operates on individual jet objects. While it is possible that all jets in an event fail the jet ID criteria, having jets which fail the jet identification does not imply that the event should be discarded. For the jet identification, the official loose jet ID for particle flow jets is applied. The algorithm operates on the particle flow jet characteristics, which reflect the properties of the underlying particles. The particle flow jet ID cuts were designed with the following in mind:

Jet properties require that the jet cannot consist of neutral hadrons or photon like objects alone - in addition ECAL noise tends to get reconstructed by the particle flow algorithm as photon like objects (neutral EM fraction < 0.99) and HCAL noise as neutral hadrons (neutral hadron fraction < 0.99). In addition the jet should consist of more than one reconstructed object (number of constituents > 1). Within the rapidities covered by the CMS tracking system ($|y| < 2.4$), additional cuts using tracker information are required by the jet ID. The jet has to contain at least one charged particle (number of charged particles > 0) and at least one charged hadron (charged hadron fraction > 0). For these inner rapidities, it is also possible to directly remove jets which only contain electron like objects (charged EM fraction < 0.99).



(a)

Figure 6.5: Discriminating variable of the $\not{E}_T / \sum E_T$ selection for different trigger paths. Higher trigger paths show wider distributions with more missing energy. The region removed by the filter is shaded.

HLT path	HLT_Jet30	HLT_Jet60	HLT_Jet80	HLT_110
Triggered events	2,741,044	2,591,154	1,491,011	2,574,451
Beam scraping filter	2,740,566	2,590,975	1,490,620	2,573,587
Vertex selection	2,732,261	2,585,365	1,479,547	2,552,687
HCAL noise filter	2,729,723	2,579,573	1,475,476	2,542,735
$\not\!\!E_T/\sum E_T$ selection	2,728,556	2,578,149	1,473,745	2,538,842

HLT path	HLT_Jet150	HLT_Jet190	HLT_Jet240
Triggered events	2,572,083	3,533,874	3,629,577
Beam scraping filter	2,570,941	3,532,315	3,626,168
Vertex selection	2,541,653	3,486,989	3,543,538
HCAL noise filter	2,524,482	3,445,296	3,460,556
$\not\!\!E_T/\sum E_T$ selection	2,518,498	3,435,789	3,446,276

HLT path	HLT_Jet300	HLT_370	HLT_Jet800
Triggered events	9,785,529	3,129,458	54,824
Beam scraping filter	9,784,720	3,128,172	54,614
Vertex selection	9,752,689	3,092,982	49,587
HCAL noise filter	9,325,702	2,839,361	13,918
$\not\!\!E_T/\sum E_T$ selection	9,286,361	2,822,310	13,051

Table 6.5: Cut flow for the different trigger paths. This table summarises the results from dataset 1-5

6.2.8 Jet correction step

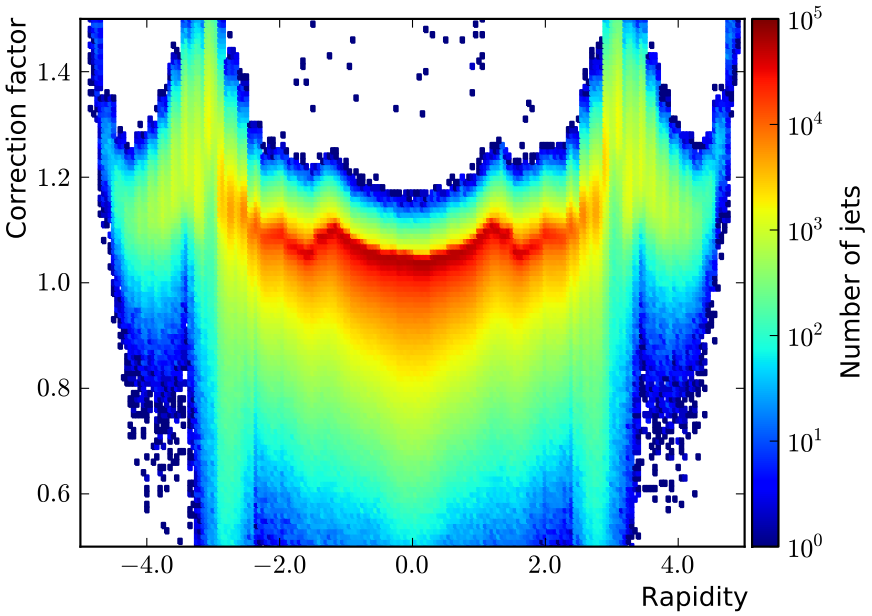
After the jet ID selection is passed, the jet energy corrections described in section 4.3 are applied to allow further selection steps to operate on the corrected energy and momentum of the jet objects. An example for the used jet energy correction factors are shown in figure 6.6.

6.2.9 Acceptance cuts

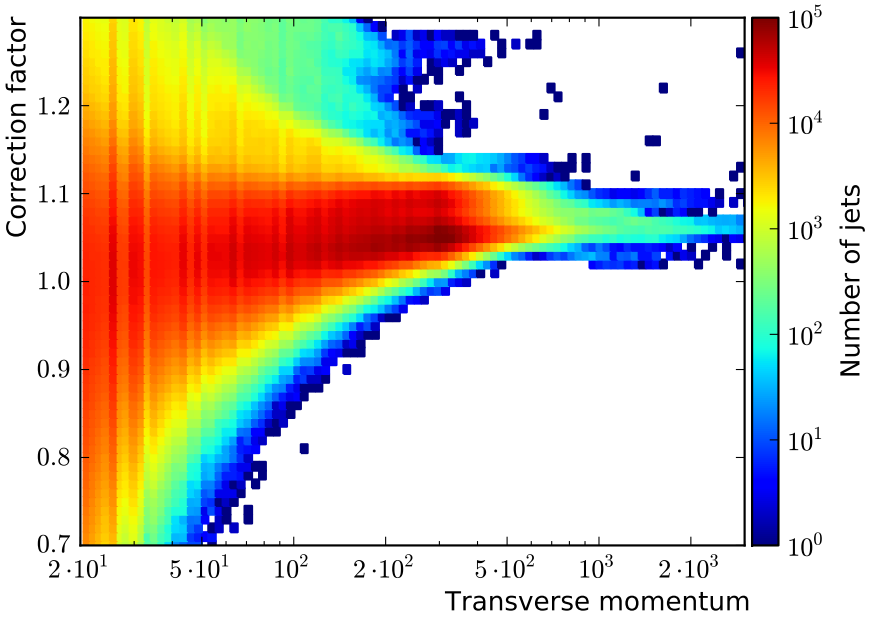
In order to synchronise the accessible phase space of the NLO theory calculation and the detector reconstruction, a filter is applied which removes all jets from the event which fail certain acceptance criteria. To fall within the detector acceptance, jets are required to have a transverse momentum exceeding 50 GeV and lie in the rapidity region between -3 and 3. The whole event is rejected if less than three jets fulfil the jet ID and detector acceptance requirements.

6.2.10 Three-jet mass cut scenarios

The final step in the event selection is the application of one of the two cut scenarios used in the analysis. The absolute cut scenario applies a cut on the transverse momentum of the third jet ($p_{T,3} > 100\,\text{GeV}$).The relative cut scenario applies a cut on the ratio between the transverse momentum of the third jet and the average transverse momentum of the leading two jets ($p_{T,3} > 0.25\langle p_{T,1}, p_{T,2}\rangle$). The differences between the two cut scenarios are discussed in section 5.0.1. Unless stated otherwise, all plots in the following sections show the results for the anti-k$_T$ 0.5 jet algorithm in the innermost rapidity bin $|y_{\max}| \leq 1$ for the relative cut scenario.



(a)



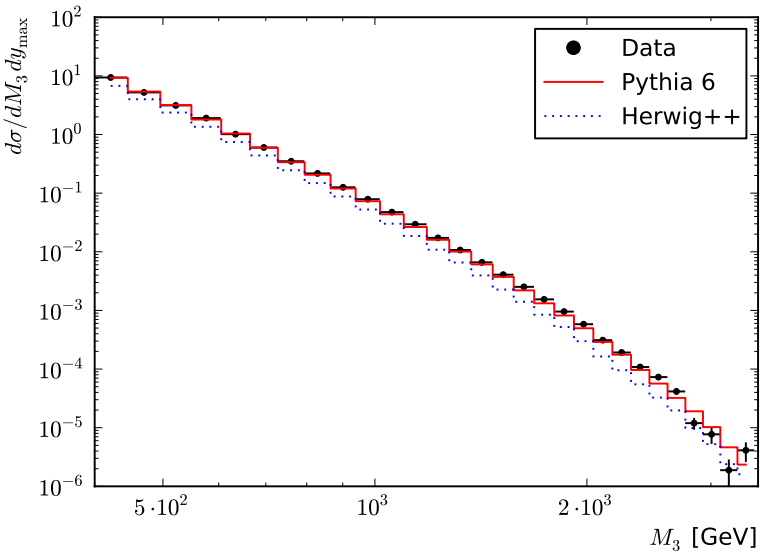
(b)

Figure 6.6: Overview of the applied jet energy correction. The distribution of the jet energy correction factors is shown as a function of pseudo-rapidity for jets with $p_T > 20\,\text{GeV}$ (top) and transverse momentum over the whole rapidity region (bottom).

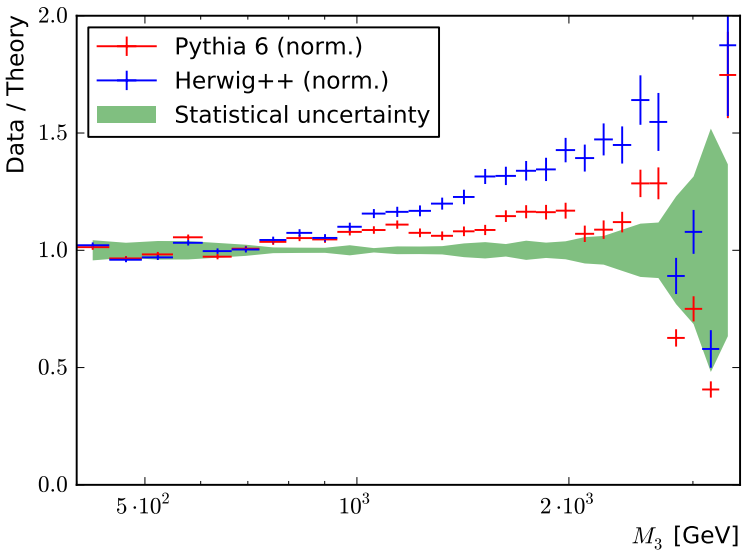
6.3 Comparison with simulated events

The measured data distribution can be directly compared with Monte-Carlo simulated events. For the comparison in figure 6.7, the Monte-Carlo event generated by Pythia 6 and Herwig++ and simulated by the full CMS detector simulation were used. As previously mentioned, these two Monte-Carlo generators do not perform a leading order calculation for the three-jet mass, but rather use a shower approximation as described in chapter 1. However the two generators are still able to follow the three-jet mass distribution over seven orders of magnitude very closely. Pythia 6 is very well tuned and matches the overall normalisation within 5%. Herwig++ on the other hand is off by a factor of 2.7. When the Monte-Carlo distributions are normalised to the data, shape differences become visible. Up to a three-jet mass of 1000 GeV, the two generators are within 10% of the data distribution. For larger three-jet masses, Pythia 6 and Herwig++ increasingly deviate from the data distribution.

However, for the purpose of detector studies and response measurements in particular, the results of the two generators are still well suited.



(a) Measured and simulated three-jet mass distribution



(b) Ratio between data and normalised Monte-Carlo prediction

Figure 6.7: Comparison between the three-jet mass distribution in data and events produced by the two Monte-Carlo generators Herwig++ and Pythia 6 in conjunction with a full detector simulation. The results are for Anti-k_T 0.5 jets in the innermost rapidity bin in the relative cut scenario. The figure shows the statistical uncertainty on the data as a green band, while the statistical uncertainties on the Monte-Carlo are expressed by the error bars. In contrast to Herwig++, Pythia gives a good estimate of the overall normalisation. When both Monte-Carlo distributions are normalised to data, shape differences, which increase for larger three-jet masses are observed.

6.4 Three-Jet mass resolution

The measurement of the three-jet mass resolution is based on the truth information in Monte-Carlo generated events, which were processed in the full detector simulation of CMS. The previously discussed event selection criteria are applied and the three-jet mass is calculated.

This calculation is performed for both the particle jets clustered from the stable generator particles and the particle flow jets which were determined from the simulated and reconstructed detector output, yielding the particle level three-jet mass $m_{3,\mathrm{gen}}$ and the particle flow reconstructed three-jet mass $m_{3,\mathrm{reco}}$. From these matching masses, the three-jet mass response

$$R = \frac{m_{3,\mathrm{reco}}}{m_{3,\mathrm{gen}}}$$

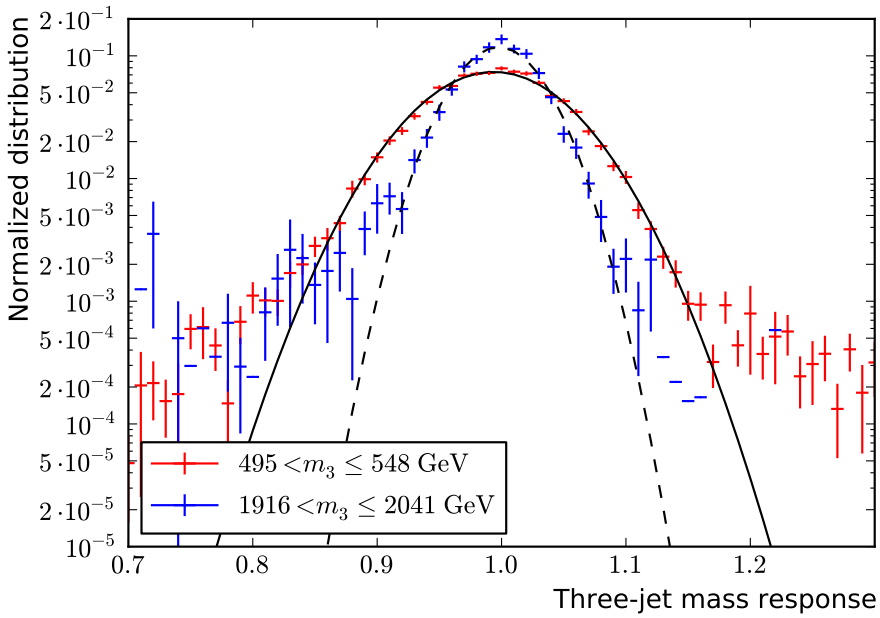
is calculated on an event-by-event basis and recorded in bins of $m_{3,\mathrm{gen}}$. In the next step, $R\left(m_{3,\mathrm{gen}}\right)$ is fitted with a Gaussian function for each $m_{3,\mathrm{gen}}$ to determine the mean $\mu_R\left(m_{3,\mathrm{gen}}\right)$ and width $\sigma_R\left(m_{3,\mathrm{gen}}\right)$ of the three-jet mass response distribution. The result of this fit is demonstrated in figure 6.8(a). The width is then interpreted as the three-jet mass resolution $\sigma_R\left(m_{3,\mathrm{gen}}\right) = \sigma_{m3}/m_3$. The measurement of $\sigma_R\left(m_{3,\mathrm{gen}}\right)$ is shown in figure 6.8(b) for different bins of $y_{\max}$, which demonstrates how the three-jet mass resolution degrades for higher rapidities.

In order to parametrise the measured resolution, a modified version of the NSC formula was used:

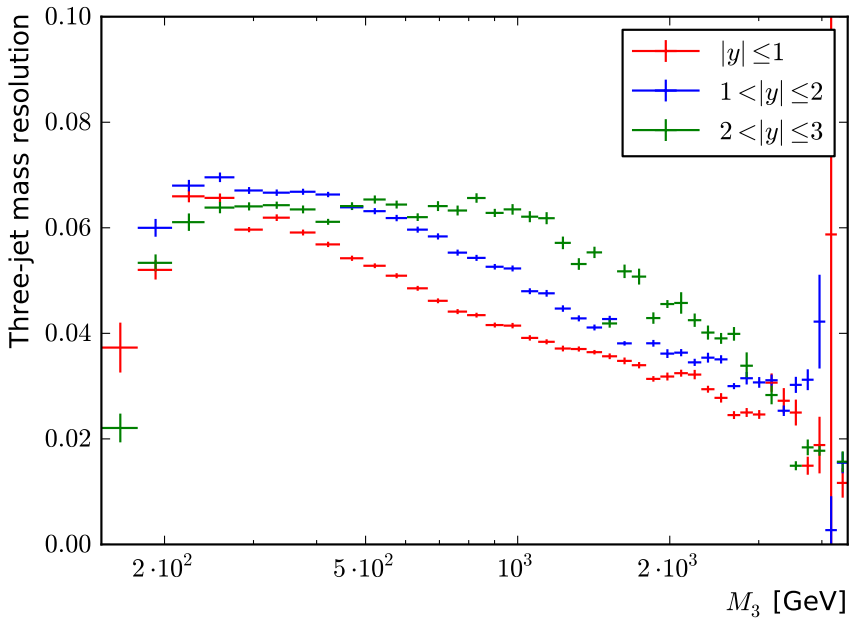
$$\frac{\sigma_{m3}}{m_3}\left(m_3\right) = \sqrt{\mathrm{sgn}\left(N\right)\left(\frac{N}{m_3}\right)^2 + \left(m_3/\,\mathrm{GeV}\right)^s \frac{S^2}{m_3} + C^2} \tag{6.1}$$

This formula was introduced in [114] to better describe the behaviour of the jet energy resolution of particle flow jets at low transverse momenta which include tracking information. The first term is modified to include the sign of the noise parameter N, while the stochastic term now contains the parameter s in addition to the stochastic noise parameter S. The new parameter s can be used to change the three-jet mass dependence of the stochastic term and allows to improve the fit for low three-jet masses. The constant term is left unchanged.

These changes to the resolution function are motivated by the fact that the resolution is guided by two very different detector subsystem. The superior resolution of the tracker dominates the measurement for low energy scales, while for increasing energy scales, the resolution converges towards the calorimeter resolution.



(a)



(b)

Figure 6.8: The three-jet mass response distribution for $|y_{\max}| \leq 1$ is shown for two different three-jet mass bins in the top plot. The error-bars are given by the statistical error. The kernel of the distribution can be approximated with a fitted Gaussian function. The width of the distribution as a function of the three-jet mass is shown for different $|y_{\max}|$ bins in the bottom plot. The shown uncertainties on the data points are the error on the width of the distribution. In both cases the relative cut scenario with Anti-k_T 0.5 jets was used

Even though formula 6.1 is able to describe the three-jet mass resolution much better than the unmodified formula 2.1, the rather flat behaviour at low masses (150 GeV − 300 GeV) leaves room for improvement. However since this mass region is of little importance for the following studies, no additional modifications to the resolution function are proposed.

The result of fitting 6.1 to the measured resolution is shown in figure 6.9(a). This step is repeated for the two different Monte-Carlo generators Pythia and Herwig++ and the mean value of the two resolutions is taken as the final result. A comparison of the two generators is shown in figure 6.9(b). The envelope of the two input resolutions is taken as the Monte-Carlo uncertainty, which takes into account the different fragmentation, hadronization and underlying event models used in the simulation. This uncertainty is added in quadrature to the parametrisation error, which contains the propagated, correlated uncertainties from the final resolution fit of the average. The errors assigned to the average of the two generators during the fit is the quadratic sum of the errors of the original data distributions. The final result for $|y_{\max}| \leq 1$ is shown in figure 6.10(a), while the parametrisation is listed in table 6.6. The detailed results for other rapidities $y_{\max}$ are given in appendix C.1.

In general, the two scenarios have comparable three-jet resolutions. However the absolute scenario provides a slightly better resolution in the lower mass region, while the relative scenario surpasses the resolution of the absolute scenario for high three-jet masses.

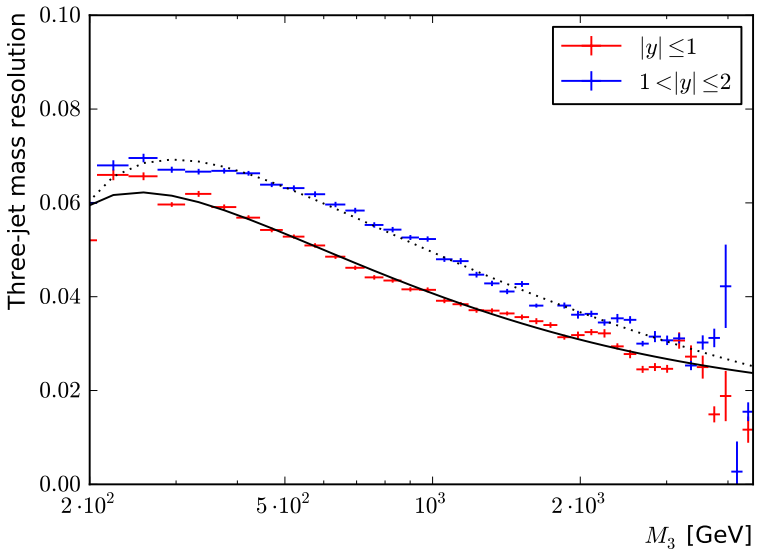
A comparison how the three-jet mass resolution behaves for different jet resolution parameters R is shown in figure 6.10(b). For low three-jet masses $m_3 < 500\,\text{GeV}$, the jet algorithm with the larger cone size shows a worse three-jet mass resolution. However for larger three-jet masses, the resolution function rapidly converges towards the values of the anti-k$_T$ jet algorithm with $R = 0.5$. For $m_3 > 800\,\text{GeV}$ there is no observable difference between two jet reconstruction algorithms.

6.4.1 Three-jet mass binning

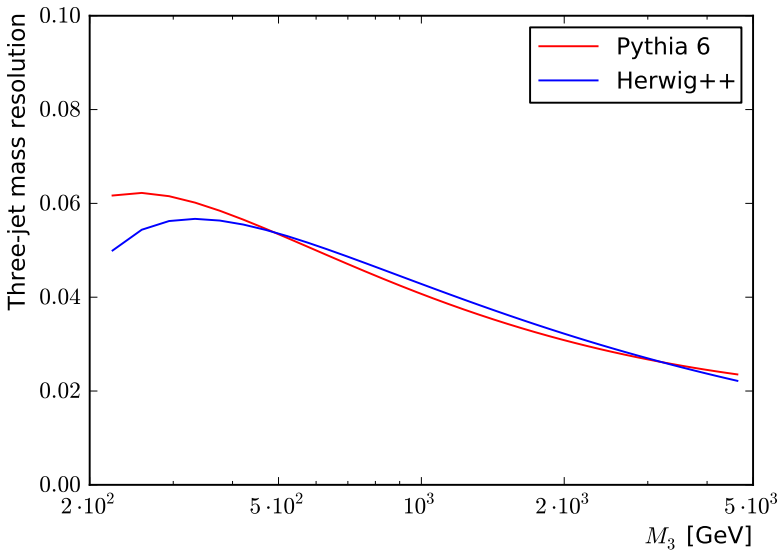
In a final step, the bin borders $m_{3,n}$ for the three-jet mass measurement were chosen in an iterative process using the measured three-jet mass resolution.

$$m_{3,n+1} = m_{3,n} + 2\sigma_{m3}\left(m_{3,n}\right)$$

This resolution-driven binning of the three-jet mass constrains the amount of bin-to-bin migration to some degree. With the approximation that the three-jet mass resolution within a bin can be assumed constant, this can

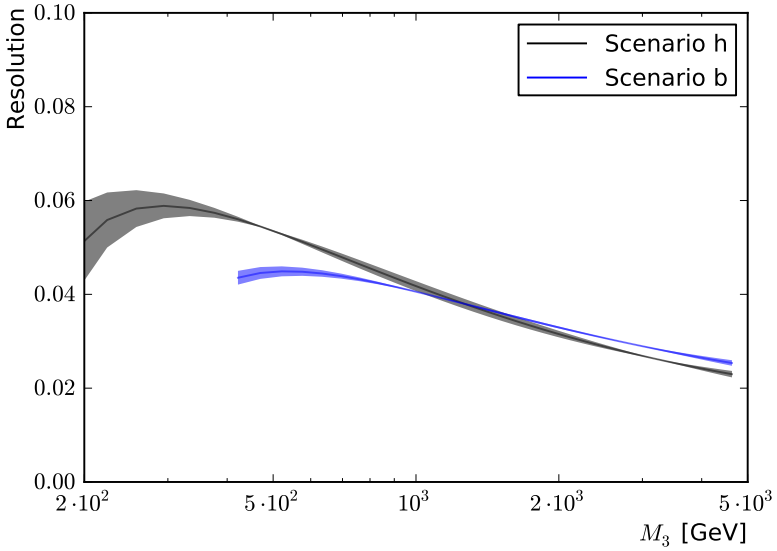


(a)

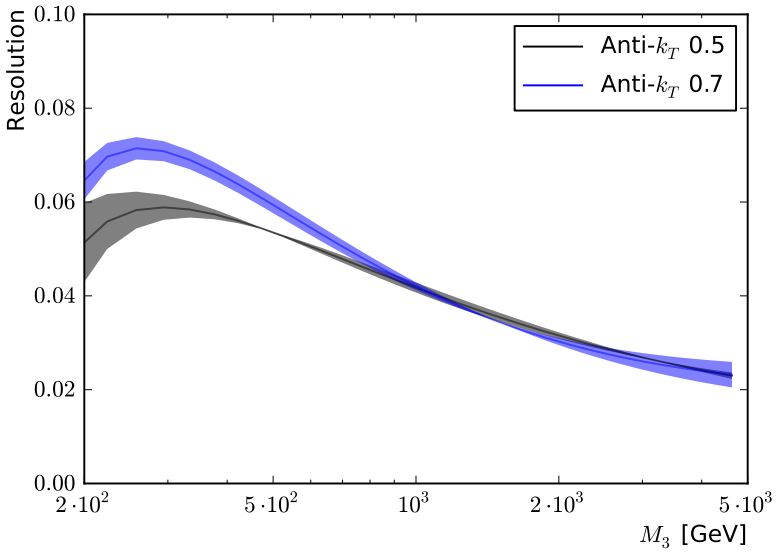


(b)

Figure 6.9: The three-jet mass resolution is parametrised for two $|y_{\max}|$ bins (top). Compared with the inner rapidity bin, the outer rapidity bin shows a slightly worse three-jet mass resolution. The procedure is repeated for two different Monte-Carlo generators and the differences are presented on the bottom plot for a single rapidity bin $|y_{\max}| \leq 1$. In both cases the relative cut scenario was used. Except for small differences, especially in the low mass region, the two generators show general agreement for the three-jet mass resolution.



(a)



(b)

Figure 6.10: The final result of the parametrised three-jet mass resolution with Monte-Carlo and propagated parametrisation errors is shown for $|y_{\max}| \leq 1$ in the relative (labelled h) and absolute cut scenario (labelled b) in the top plot. A comparison for different jet sizes shows a worse resolution at low masses for the algorithm with larger jet size in the relative cut scenario (bottom).

	Absolute cut	Relative cut
N [GeV]	-21.481 ± 0.55	-18.252 ± 1.1
S [$\sqrt{\text{GeV}}$]	1.1457 ± 0.031	1.8398 ± 0.36
C [1]	0.014982 ± 0.00083	0.013592 ± 0.0024
s [1]	0.052171 ± 0.0042	-0.084198 ± 0.06

Table 6.6: Table of parameters for the three-jet mass resolution function 6.1 (Anti-k$_T$ 0.5, innermost rapidity bin $|y_{\max}| \leq 1$). Parameters and covariance matrices for all scenarios are given in appendix C.1.

150	177	207	240	275	313	354	398	445	495	548
604	664	727	794	864	938	1016	1098	1184	1274	1369
1469	1573	1682	1796	1916	2041	2172	2309	2452	2602	2758
2921	3092	3270	3456	3650	3852	4063	4283	4513	4753	∞

Table 6.7: Table of resolution driven three-jet mass bins used in this analysis

be shown by looking at some true three-jet mass m_{3t} and calculating the probability of measuring the smeared three-jet mass m_{3s} in the given bin. From the cumulative distribution of a Gaussian resolution function follows that for a true three-jet mass at the bin edges, the measured three-jet mass stays with a probability of 48% within the bin. For true value m_{3t} at the centre of the bin, the measured value doesn't migrate out of the bin with a probability of 68%. This yields an overall probability of more than 55% for some true value to get measured in the appropriate bin.

The bin edges produced by this algorithm and used in the analysis are given in table 6.7. It should be mentioned that this binning was produced at an earlier stage of the analysis. Refinements to the three-jet mass resolution measurement have lead to small differences between this binning and the binning determined from the final resolution results. However these differences are smaller than 1% and within the error introduced by the integerisation of the bins boundaries for three-jet masses above 800 GeV.

6.5 Unfolding

Detector effects like limited acceptance and finite resolution cause differences between the measured and true three-jet mass distribution. In order to remove these influences, the so-called unfolding problem must be solved. A detailed introduction to the unfolding problem is given in appendix A. The in general very complex relationship between measured and true distribution can be expressed in the case of the three-jet mass measurement in the form of a simple matrix equation:

$$\vec{y} = \mathbf{A}\vec{x} + \vec{b} \tag{6.2}$$

with matrix $\mathbf{A}$ describing the detector response and vectors representing the histogram entries of the measured distribution $\vec{y}$, background distribution $\vec{b}$ and true distribution $\vec{x}$.

Since events at a hadron collider are dominated by QCD interactions, the contribution of background processes are insignificant. Therefore the background is neglected in the following discussions. In order to avoid issues related to the limited detector acceptance in the three-jet mass measurement, the acceptance is artificially limited in the theory prediction and the same acceptance cuts were applied on data during the event selection. In addition, the basic analysis steps for the three-jet mass are the same for both data and theory. This means that there are no transformative changes between the three-jet mass in data and theory and that the true distribution from the unfolding procedure can be directly compared with the theory prediction.

Without acceptance induced issues, the transformative changes modelled by the detector response can be traced back to differences between the reconstructed and true jet objects. The application of jet energy corrections take care of the non-linear detector response and thereby remove the bias from the reconstructed input objects. However the influence of the finite detector resolution on the input objects cannot be solved with a similar technique. Therefore the detector response mainly models the finite detector resolution.

Several established unfolding methods were studied to remove these detector resolution effects from the three-jet mass measurement and get the true three-jet mass distribution.

6.5.1 Bin-by-bin method

A simple correction method, which is often used in experiments is the bin-by-bin method. The bin-by-bin method is based on the assumption, that there are no correlations between neighbouring bins. In this case, the detector

response is a simple diagonal matrix and the unfolding algorithm is reduced to the application of multiplicative correction factors to each bin.

Unless the detector response can be measured in a data-driven way, the diagonal matrix has to be determined from a Monte-Carlo simulation. However this might introduce biases into the correction factor, which are not accounted for in the unfolding algorithm. In addition, the assignment of errors on the final result might be problematic, since the correlations between bins are completely ignored. Therefore the bin-by-bin method should only be used to account for the effects introduced by limited acceptance, since in that case the consequences of disregarding bin correlations are usually less severe.

All these facts show that the bin-by-bin method is inappropriate for the unfolding of the three-jet mass.

6.5.2 Matrix inversion

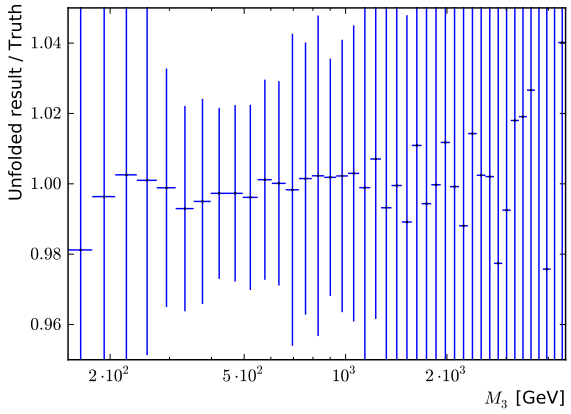
An obvious method to solve equation 6.2 is based on simple linear algebra. For non-singular, rectangular matrices $\mathbf{A}$, it is possible to calculate the inverse matrix $\mathbf{A}^{-1}$ and determine the solution for $\vec{x}$ with:

$$\vec{x} = \mathbf{A}^{-1}\vec{y}$$

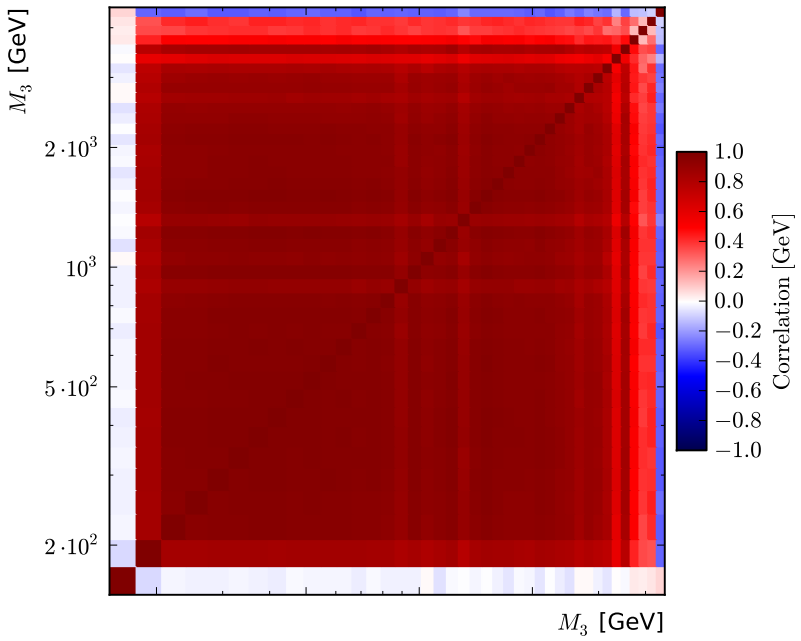
The matrix inversion method also gives a straightforward method to calculate the covariance matrix cov_x of the true distribution $\vec{x}$ from the covariance matrix cov_y of the measured distribution with standard error propagation:

$$\mathrm{cov}_x = \mathbf{A}^{-1}\mathrm{cov}_y \left(\mathbf{A}^{-1}\right)^{\mathrm{T}}$$

It can be shown that this solution is unbiased and exhibits minimal covariance [115]. When the method is applied in the context of a Monte-Carlo study to the same simulated distribution, which was used to determine the detector response, the matrix inversion yields the exact result for the true distribution. However as soon as there are variations, which break the definite relationship of equation 6.2, instabilities of the solution are evident. An example for this behaviour is shown in figure 6.11(a), where the Monte-Carlo simulated three-jet mass distribution is varied within the statistical uncertainties and the unfolding result is compared with the Monte-Carlo truth. The instabilities of the method are caused by large correlations between different, unfolded bins as shown in figure 6.11(b). Large, negative correlations can give rise to oscillations in the unfolded result, while positive correlations can cause small statistical fluctuations in one bin to change the solution in an inappropriately large way. The large correlations make this method inadequate for this study.



(a)



(b)

Figure 6.11: In a toy experiment, the simulated Monte-Carlo distribution, which was used to determine the detector response matrix, is varied within the statistical uncertainties. The modified distributions are unfolded using the matrix inversion method and compared with the Monte-Carlo truth distribution (top). This closure plot shows increasing fluctuations for higher three-jet masses. The bin correlations, as determined from the toy ensemble, show a large correlation between all bins of the unfolded distribution (bottom).

6.5.3 Singular value decomposition

The singular value decomposition (SVD) unfolding algorithm is based on the SVD decomposition (explained in appendix A) of the detector response. Similar to the previously discussed matrix inversion method, the unfolding solution is given by

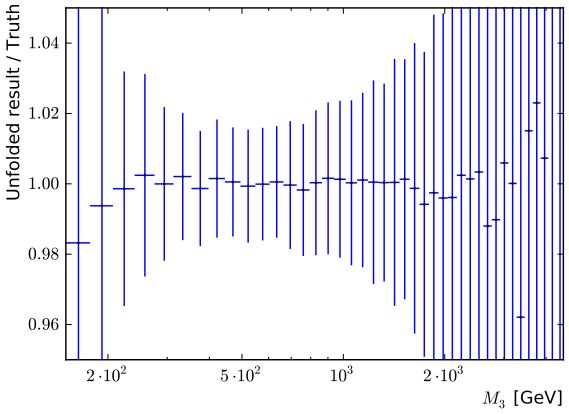
$$\vec{x} = \sum_{i=1}^{n} \frac{\vec{v}_i \vec{u}_i^{\mathbf{T}} \vec{y}}{\sigma_i} \tag{6.3}$$

with orthonormal vectors $\vec{u}$ and $\vec{v}$ and singular values σ_i from the SVD decomposition. An analysis of the numerical instabilities[116] shows that the spectral properties of the detector response are the reason that the naive unfolding methods run into problems. Since the measured distribution $\vec{y}$ is divided by the singular values σ_i of the detector response, small singular values can cause large fluctuations and uncertainties in the unfolded distribution $\vec{x}$.

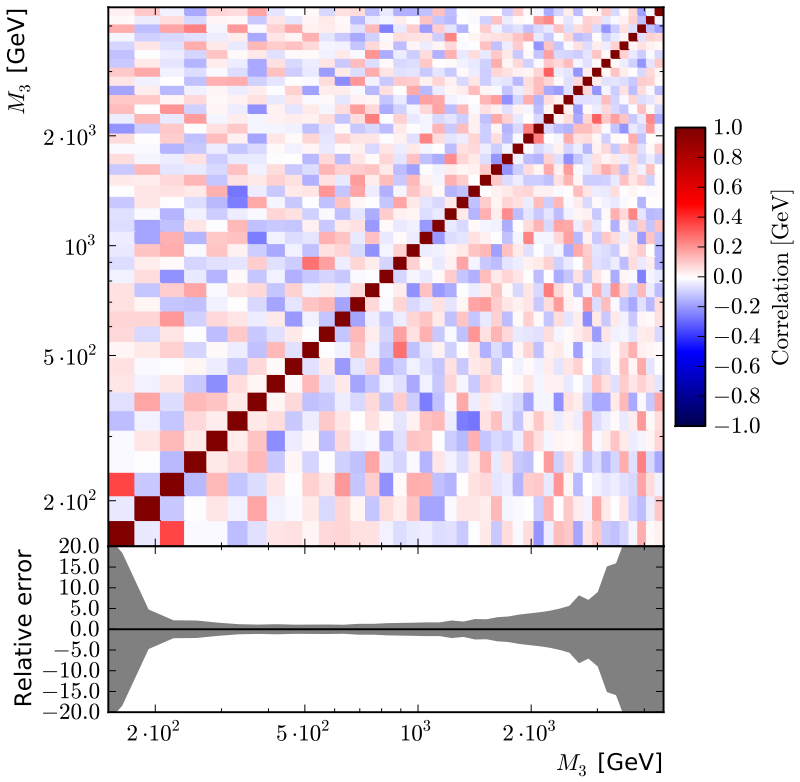
The solution to this problem is to impose a measure of smoothness on the unfolded result in a process called regularisation. It is evident that this unavoidably introduces some bias. Every regularisation method usually is expressed as a function of some parameter, which influences the magnitude of the regularisation bias.

The SVD method can be regularised in a number of different ways. The method used by the RooUnfold package is based on an effective cut-off for small singular values, parametrised by the integer value k. For the ordered list of singular values $\sigma_1 \geq \cdots \geq \sigma_k \geq \cdots \geq \sigma_n \geq 0$, the cut-off parameter k determines the smallest singular value σ_k, which is used in the determination of the true distribution in equation 6.3. It is therefore possible to choose between an unregularised unfolding ($k = n$) and a completely regularised unfolding ($k = 1$).

When the SVD method is applied on the three-jet mass measurement in Monte-Carlo, very reasonable results can be derived as shown in figure 6.12(a) for the generators Herwig++. The algorithm also gives a quite low unfolding error. However when the algorithm is used to unfold a simulated distribution from Herwig++ with a detector response matrix determined from Pythia 6, large differences are uncovered (see figure 6.13(a)), which point towards the fact that the SVD method shows a large bias towards the Monte-Carlo distribution underlying the detector response. This is true for a large number of cut-off parameters k.

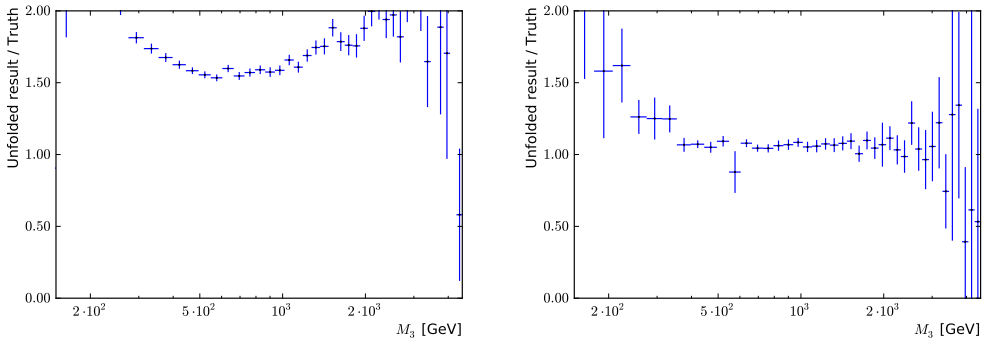


(a)



(b)

Figure 6.12: In an toy experiment, the simulated Monte-Carlo distribution, also containing unmatched events, is varied within the statistical uncertainties. The modified distributions are unfolded using the matrix inversion method and compared with the Monte-Carlo truth distribution (top). The bin correlations, as determined from the toy ensemble, show quite small correlations between the bins of the unfolded distribution (bottom). Two different Monte-Carlo programs were used to simulate the events. These plots show the results for Herwig++, however the Pythia 6 results are very similar (see A.



(a) SVD method, $k = 20$ (b) Bayes method, $s = 4$

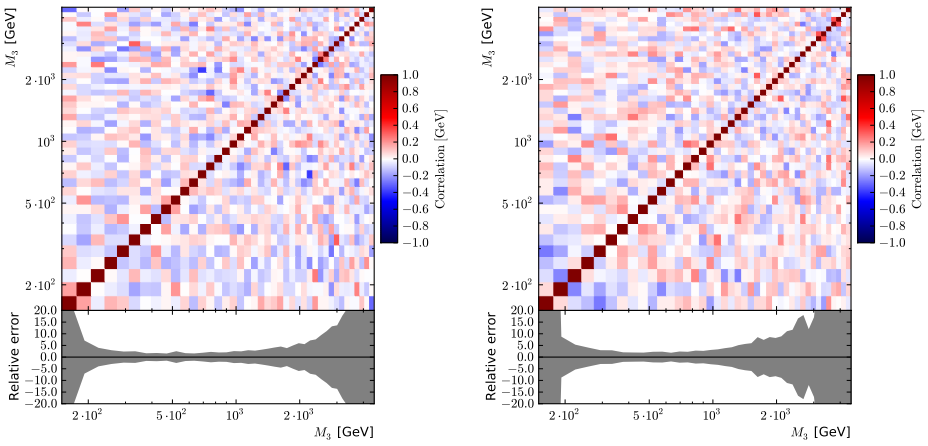
Figure 6.13: In order to check for Monte-Carlo dependencies of the unfolding algorithm, the SVD (left) and Bayesian (right) unfolding algorithm are given detector responses derived from Pythia 6 and simulated distributions from Herwig++. The result is compared with the Monte-Carlo truth distribution of Herwig++. The SVD method shows a large Monte-Carlo bias, while the Bayesian unfolding method only shows differences in the low three-jet mass region where the two Monte-Carlo generators already showed different resolution modelling.

6.5.4 Bayesian unfolding

The Bayesian unfolding method[117] is an example for an iterative unfolding algorithm. Details of this method are discussed in appendix A. The algorithm formulates the unfolding problem in terms of causes C_i and effects E_j. In each step s, the probability $P_s(C_i|E_j)$ is determined with Bayesian inference from results from the previous iteration step. This allows to calculate for a number $n(E_j)$ of observed events of a certain effect E_j, how many events $\hat{n}_s(C_i)$ can be assigned to a certain cause C_i after s iteration steps.

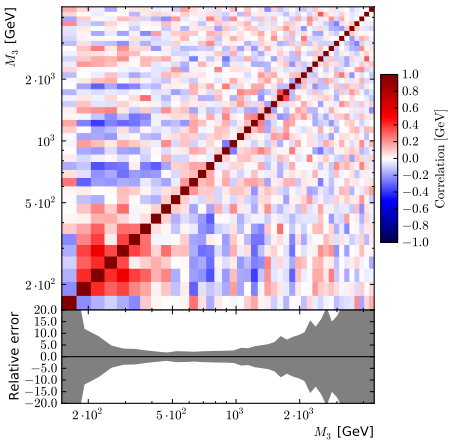
Usually, the method already yields good results after a small number of steps s. In the original example in [117] $3-4$ iterations were sufficient. However the algorithm converges for $s \to \infty$ towards the unregularised matrix inversion result, with the same problems as demonstrated earlier. The number of iterations can therefore be identified as a kind of regularisation parameter.

The original publication suggests to do a χ^2 comparison between $\hat{n}(C_i)$ and $\hat{n}_0(C_i)$ to determine the break condition for the iteration process. However it is not always clear when the χ^2 is "small enough"[117] to stop the iteration, because with an increasing number of iterations both the error and the fluctuations increase, leading to smaller values of χ^2.

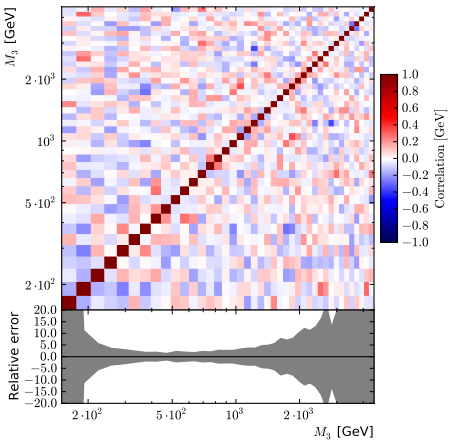


(a) $s = 1$ iteration

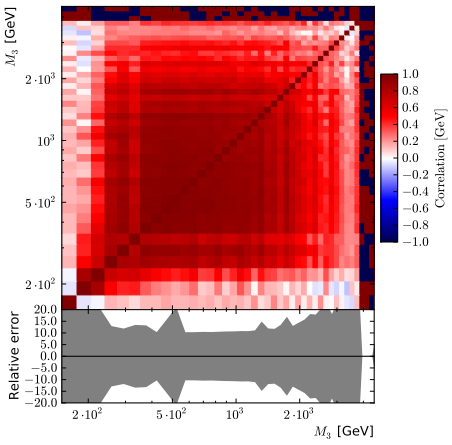
(b) $s = 2$ iterations



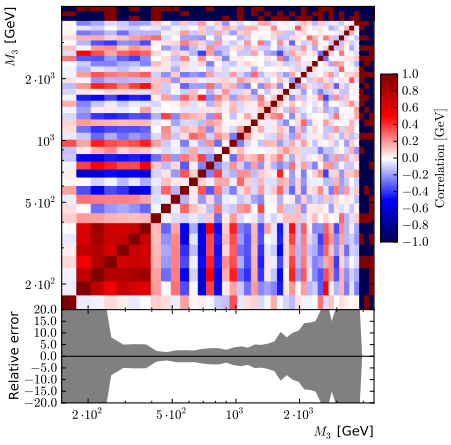
(c) $s = 3$ iterations



(d) $s = 4$ iterations



(e) $s = 5$ iterations



(f) $s = 6$ iterations

Figure 6.14: Three-jet mass bin correlation matrices and relative unfolding errors determined for different iterations of the Bayesian unfolding method. The three-jet mass distribution measured in data with Anti-k_T 0.5 jets, $|y_{max}| \leq 1$ and relative cut scenario is used as input. Starting with $s = 5$ iterations, the algorithm introduces large (anti-)correlation patterns in the matrix.

For the unfolding of the three-jet mass, the number of iterations is chosen based on the behaviour of the bin correlations. For $s > 4$, patterns of large bin correlations and anti-correlations start to appear and the unfolding result shows increasing fluctuations as shown in figure 6.14.

Unlike the previously discussed algorithms, this unfolding method doesn't show any problematic behaviour related to the unfolding of Monte-Carlo distributions as shown in figure 6.13(b). Therefore the $s = 4$ Bayesian unfolding algorithm is chosen as the primary unfolding method for the three-jet mass measurement.

6.6 Comparison with unfolded events

The unfolded data distribution can now be compared with the particle level prediction of Monte-Carlo event generators. Figure 6.15 shows such a comparison between data and the Monte-Carlo generators Pythia 6, Herwig++ and Sherpa.

The already discussed deviation between data and the Pythia 6 and Herwig++ prediction (see section 6.3) is left unchanged by the unfolding procedure and the removal of the detector simulation. However the unfolding makes it possible to perform a comparison between data and the Monte-Carlo generator Sherpa.

Sherpa generated events, which were never processed by the full detector simulation, show a much better shape agreement with data. This can be traced back to the fact that Sherpa is able to calculate the three-jet QCD matrix elements in leading order, before applying the shower approximation.

While there are small deviations for low three-jet masses, the deviation between data and Sherpa prediction is very flat and within the total data and theory uncertainties.

Therefore, Sherpa is well suited for the study of the non-perturbative corrections presented in section 5.3.

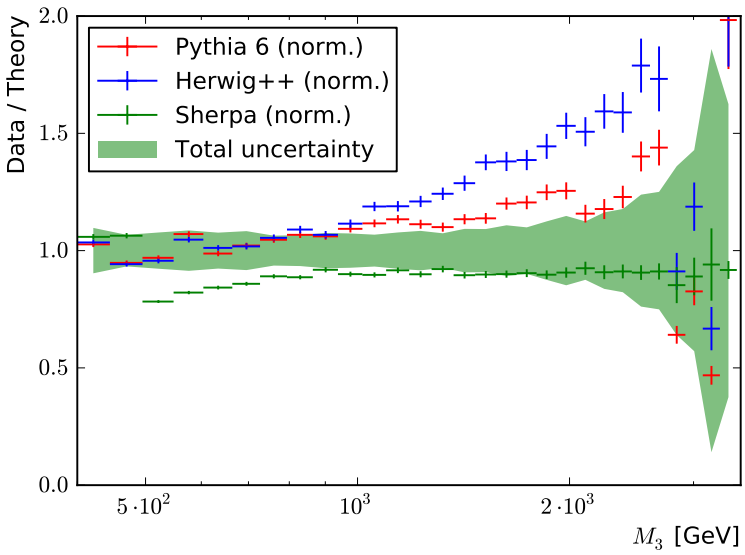


Figure 6.15: Comparison between unfolded data and the particle level results of three Monte-Carlo generators Herwig++, Pythia 6 and Sherpa for Anti-k_T 0.5 jets in the innermost rapidity bin in the relative cut scenario. The figure shows the total uncertainty on the unfolded data as a green band, while the statistical uncertainties on the Monte-Carlo are expressed by the error bars. The two generators Herwig++ and Pythia 6, which use the a shower algorithm to generate the third jet, show a large shape difference. The Monte-Carlo Sherpa, which calculates the third jet from the leading-order matrix element, shows a good agreement between data and simulation.

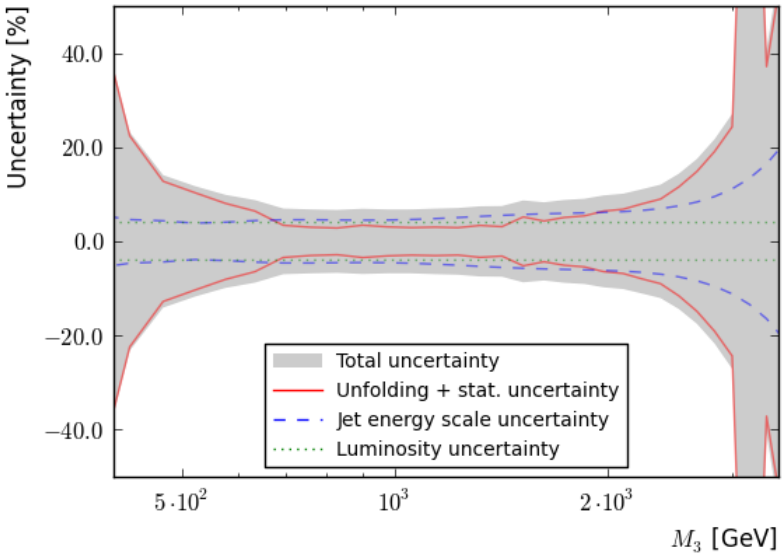


Figure 6.16: Overview of the measurement uncertainties for Anti-k$_T$ 0.5 jets in the innermost rapidity region ($|y_{\max}| \leq 1$) with the absolute cut applied. The statistical and unfolding uncertainties dominate the low and high three-jet mass regions, while over a large range of three-jet masses, the jet energy scale uncertainty is slightly larger. The measurement uncertainties are relatively small, which will allow to use this measurement to improve theory predictions in the future.

6.7 Uncertainties

In the following, the four major uncertainty sources for the three-jet mass measurement are presented. The first two sources to be discussed are the statistical and unfolding uncertainties, which are intrinsically linked through the unfolding algorithm. These unfolding uncertainties also have a Monte-Carlo component, since the three-jet mass detector response is determined from simulated events. The luminosity uncertainty is independent of the three-jet mass and only influences the overall normalisation of the measurement. The most complex uncertainty source is the jet energy uncertainty. This uncertainty is broken up into 16 constituent uncertainty sources, which are independently analysed.

For all these uncertainty sources, the complete covariance matrix is available in appendix D. A general overview of the measurement uncertainties is given in table 6.8.

$500 - 2000\,\text{GeV}$	Min	Average	Max
Luminosity	4%	4%	4%
Unfolding and statistical	2%	3%	10%
Statistical	1%	1%	5%
Jet energy scale	4%	6%	8%
Absolute scale	2%	3%	4%
Flavour	2%	3%	5%
High-p_T extrapolation	2%	3%	5%
Pile-up	1%	1%	1%

Table 6.8: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 2000 GeV.

6.7.1 Statistical and unfolding uncertainties

The uncertainties of the previously discussed unfolding algorithms are determined from toy experiments. The toy experiments are organised into toy scenarios each consisting of $n = 1000$ toy experiments. Each toy scenario uses a particular unfolding algorithm, regularisation parameter, detector response matrix and toy variation strategy. For each toy scenario, the unfolding result is determined from the sample mean, while the uncertainties for the toy scenario are given by the sample covariance:

$$\bar{m}_{3,i} = \frac{1}{n}\sum_{k=1}^{n} m_{3,i}^{(k)}$$

$$\text{cov}_{\text{unf}}\left(m_{3,i}; m_{3,j}\right) = \frac{1}{n-1}\sum_{k=1}^{n}\left(m_{3,i}^{(k)} - \bar{m}_{3,i}\right)\left(m_{3,j}^{(k)} - \bar{m}_{3,j}\right)$$

In these formula, $m_{3,i}^{(k)}$ is the three-jet mass in bin i of toy sample k. The toy sample mean for the three-jet mass bin i is designated $\bar{m}_{3,i}$.The covariance between the two bins i and j is called $\text{cov}\left(m_{3,i}; m_{3,i}\right)$ and takes into account that the sample mean and the sample covariance are derived from the same toy scenario by applying Bessel's correction $n/\left(n-1\right)$.

The following steps are performed for all unfolding algorithms and regularisation parameters. However as discussed earlier, only the output of the Bayesian unfolding method with 4 iterations is relevant for the final result.

The statistical error component is determined from toy scenarios which change just the input distributions within their uncertainties. The total unfolding uncertainty is taken from toy scenarios which vary both the input histogram and the response matrix within the limits of their uncertainties.

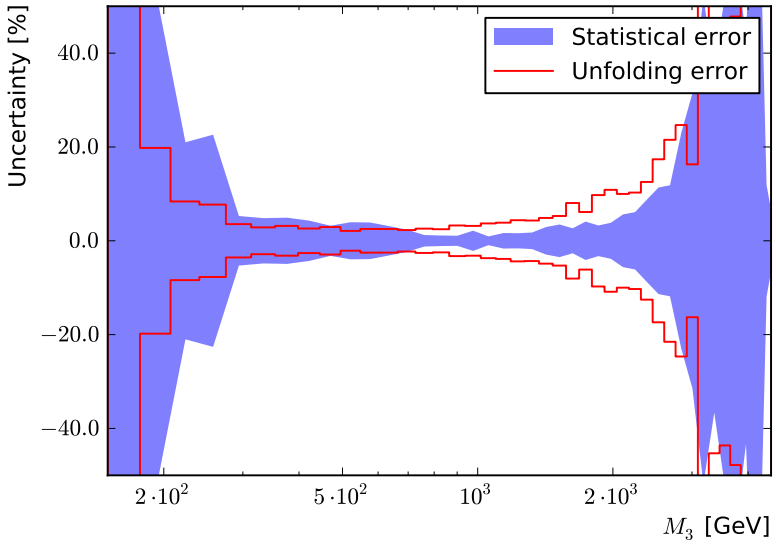


Figure 6.17: Overview of statistical (band) and algorithmic unfolding uncertainty (line) for anti-k_T 0.5 jets in the innermost rapidity region with the relative cut scenario. These two uncertainties are closely related via the unfolding algorithm and show similar three-jet mass dependence and size. Due to their relationship, only the combination (called total unfolding uncertainty) of these two uncertainties is relevant for this analysis.

The algorithmic unfolding uncertainty is then defined as

$$\sigma_{\text{unf}} = \sqrt{\sigma^2_{\text{total}} - \sigma^2_{\text{stat}}} \tag{6.4}$$

For the determination of the total uncertainty and the final unfolding result, the two toy scenarios using the Pythia 6 and Herwig++ derived detector response are merged to form a single toy scenario. The unfolding result is therefore the average of the Herwig++ and Pythia 6 toys, while any systematic differences between the generators are accounted for in the variance of the toys. It is easy to calculate that the merged variance incorporates the average variance of the two sub-samples for Herwig++ and Pythia plus half the difference between the respective mean values squared.

As shown in figure 6.18, large differences in the uncertainty composition for the two cut scenarios can be observed. For $|y_{\text{max}}| \leq 0$, the absolute cut scenario has a quite constant unfolding uncertainty of 3%, which slowly increases till it reaches at $m_3 \approx 2500\,\text{GeV}$, the same order of magnitude as the statistical uncertainty. In the same rapidity region, the relative cut scenario exhibits an unfolding uncertainty of around 8%, which reaches the level of

the statistical uncertainty around the same time as the absolute cut scenario. For higher rapidity bins, the decreasing statistics in the Monte-Carlo becomes more significant, leading to larger fluctuations in the distributions and uncertainties.

6.7.2 Luminosity uncertainty

The luminosity enters the measurement as a scaling factor affecting the overall normalisation of the three-jet mass distribution. For the selected run and luminosity sections, the integrated luminosity was determined to be $4,984.49\,\mathrm{pb}^{-1}$. This measurement was performed using the pixel reconstruction and corrected for afterglow effects[109]. The uncertainty on this number is 4%, which corresponds to $\pm 199\,\mathrm{pb}^{-1}$.

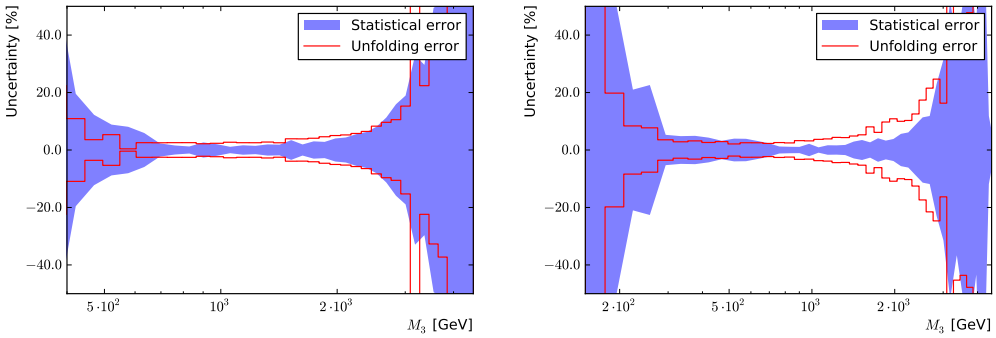
6.7.3 Jet energy scale uncertainties

As discussed previously, the jets reconstructed by the jet algorithm need additional corrections due to detector effects. These correction factors are afflicted with several sources of uncertainties. These uncertainties are provided by the JetMET group in the form of distinct jet energy uncertainty sources. Since the uncertainty sources are parametrised as functions of p_T and η, it is convenient to write the uncertainty from source i in the form $\sigma_i(p_T, \eta)$. All uncertainties $\sigma_i(p_T, \eta)$ were determined in such a way that they are fully uncorrelated from a second $(i \neq j)$ uncertainty source $\sigma_j(p_T, \eta)$. To allow for asymmetric errors, each uncertainty source carries information about the upward and downward fluctuation $\sigma_i^{\mathrm{up}}(p_T, \eta)$ and $\sigma_i^{\mathrm{down}}(p_T, \eta)$. Therefore, the way these uncertainties are used is analogous to they way PDF eigenvectors are processed to calculate the uncertainties of the theoretical prediction on page 84.

The full list of $n_{\mathrm{JEC}} = 16$ uncertainty sources used in the analysis is given below together with the name and description of the uncertainty[88]:

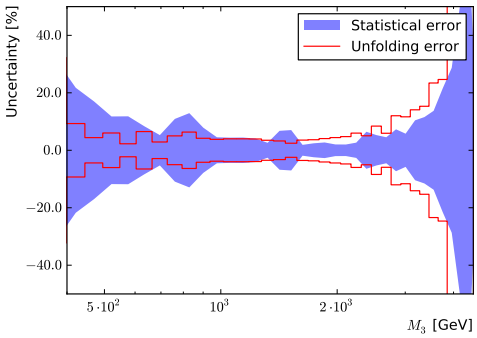
- **Absolute**: Uncertainty on the absolute jet energy scale. It derives mainly from the combined photon and $Z \to \mu\mu$ reference scale and from corrections for FSR and ISR in these processes.

- **HighPtExtra**: Uncertainty due to the high-p_T extrapolation of the corrections using the two generators Herwig++ and Pythia 6. The differences between the two generators can be traced back to to differences in the fragmentation and underlying event models.

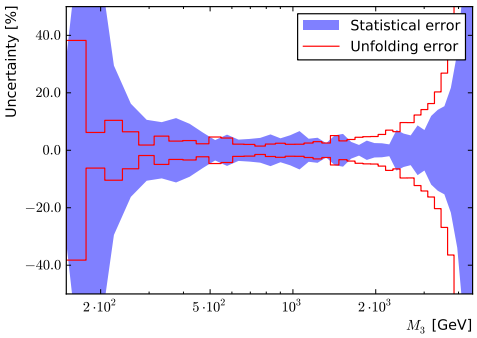


(a) Abs. cut, $|y_{max}| \leq 1$

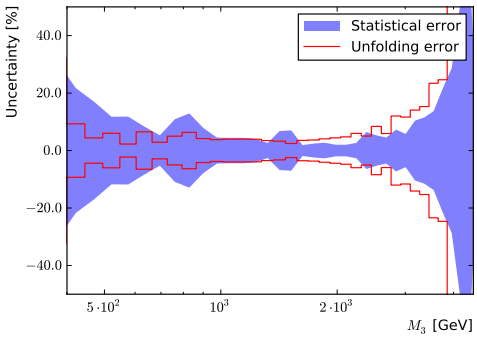
(b) Rel. cut, $|y_{max}| \leq 1$



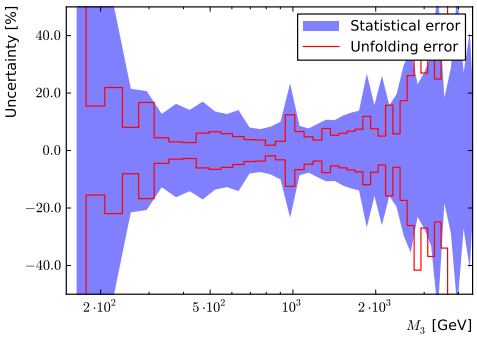
(c) Abs. cut, $1 < |y_{max}| \leq 2$



(d) Rel. cut, $1 < |y_{max}| \leq 2$



(e) Abs. cut, $2 < |y_{max}| \leq 3$



(f) Rel. cut, $2 < |y_{max}| \leq 3$

Figure 6.18: Overview of statistical (band) and algorithmic unfolding uncertainties (line) determined for different rapidity bins and cut scenarios. Differences in the uncertainty composition for the two cut scenarios can be observed, which can be explained by the hardness of the jet cuts. In general, the size of both the statistical and algorithmic unfolding uncertainties get larger for increasing rapidities. This is due to a decreasing number of events in both data and Monte-Carlo determined detector response matrix.

- **SinglePion**: This uncertainty is derived from the propagation of a 3% variation of the single particle response to particle flow jets.

- **Flavour**: Based on the differences between Herwig++ and Pythia 6 quark and gluon responses in a QCD flavour mixture.

- **Time**: The jet energy scale shows time-dependent variations in the end-cap region and therefore change over time. This shift is probably due to ECAL laser calibration instabilities.

- **RelativeJER - EC1,EC2,HF**: Rapidity dependent uncertainty due to the jet energy resolution, split into three regions in which the jet energy resolution is assumed to be fully correlated. The regions are the end-cap covered by tracking (**EC1**), the end-cap outside the tracking system (**EC2**) and the region covered by the hadronic forward calorimeter (**HF**).

- **RelativeFSR**: Rapidity dependent uncertainty caused by final state radiation corrections. The uncertainty is correlated over the complete detector and increases for large rapidities.

- **RelativeStat - EC2,HF**: Statistical uncertainty in the determination of the L2 correction factor. This error only plays a role in the end-cap region not covered by the tracker (**EC2**) and in the HF calorimeter (**HF**)

- **PileUp - DataMC, OOT, Pt, Bias, JetRate**: Uncertainties on the pile-up corrections. **DataMC** is determined from data and Monte-Carlo differences in ZeroBias data. The **OOT** component estimates residual out-of-time pile-up for prescaled triggers when reweighting Monte-Carlo to unprescaled data. The **Pt** uncertainty is due to the fact that the offset correction is calibrated for jet between 20 and 30 GeV and that there is an offset dependence on the jet p_T. **Bias** is an uncertainty determined from an observed difference between the offset in ZeroBias and QCD Monte-Carlo. The **JetRate** uncertainty component accounts for jet rate variations as a function of the number of vertices in singlet-jet trigger data.

These uncertainties are related to the total jet energy scale uncertainty via the formula:

$$\sigma_{\text{Total}}\left(p_T, \eta\right) = \sqrt{\sum_{i=0}^{n_{\text{JEC}}} \left(\sigma_i\left(p_T, \eta\right)\right)^2}$$

In order to propagate the uncertainties to the three-jet mass distribution, these fluctuations are used directly after the jet energy scale corrections are applied. For each corrected jet four-vector p_i and uncertainty source j, two variations are calculated:

$$p_{i,\mathrm{up}} = \left(1 + \sigma_j^{\mathrm{up}}\left(p_T, \eta\right)\right) p_i, \quad p_{i,\mathrm{down}} = \left(1 - \sigma_j^{\mathrm{down}}\left(p_T, \eta\right)\right) p_i$$

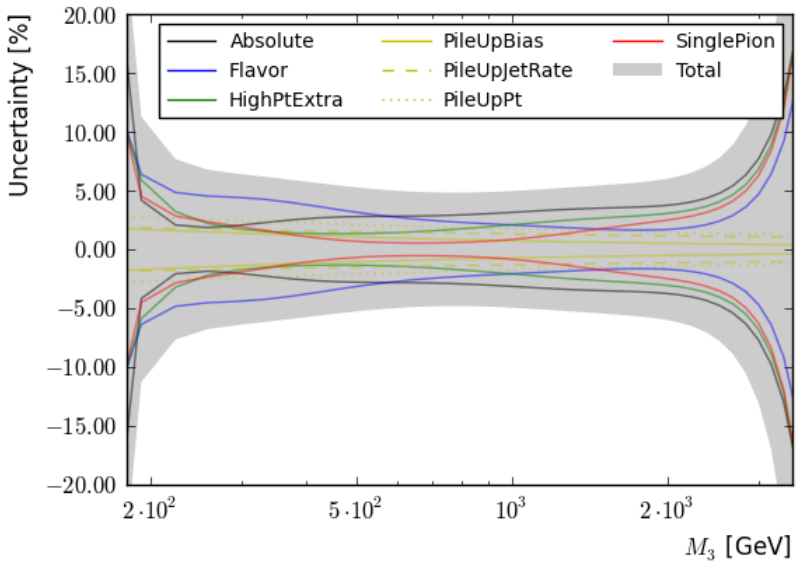
The resulting jet four-vectors are processed in pairs by the analysis, yielding two three-jet mass distributions $m_3^{\mathrm{up/down}}$ per jet energy uncertainty source in addition to the central value m_3^{centre} without any fluctuations. Since the uncertainty sources are fully uncorrelated between themselves, but describe jet energy correction variations that are fully correlated within a given source, the asymmetric error $\Delta m_{3,i}^{\mathrm{up/down}}$ due to the jet energy uncertainty source k on the three-jet mass distribution $m_{3,i}^{(k)}$ in bin i can be calculated with the following formula

$$\begin{aligned}
\Delta m_{3,i}^{\mathrm{up}} &= \sqrt{\sum_{k=0}^{n_{JEC}} \left[\max\left(m_{3,i}^{\mathrm{up}(k)} - m_{3,i}^{\mathrm{centre}}, m_{3,i}^{\mathrm{down}(k)} - m_{3,i}^{\mathrm{centre}}, 0\right)\right]^2} \\
\Delta m_{3,i}^{\mathrm{down}} &= \sqrt{\sum_{k=0}^{n_{JEC}} \left[\max\left(m_{3,i}^{\mathrm{centre}} - m_{3,i}^{\mathrm{up}(k)}, m_{3,i}^{\mathrm{centre}} - m_{3,i}^{\mathrm{down}(k)}, 0\right)\right]^2}
\end{aligned}$$

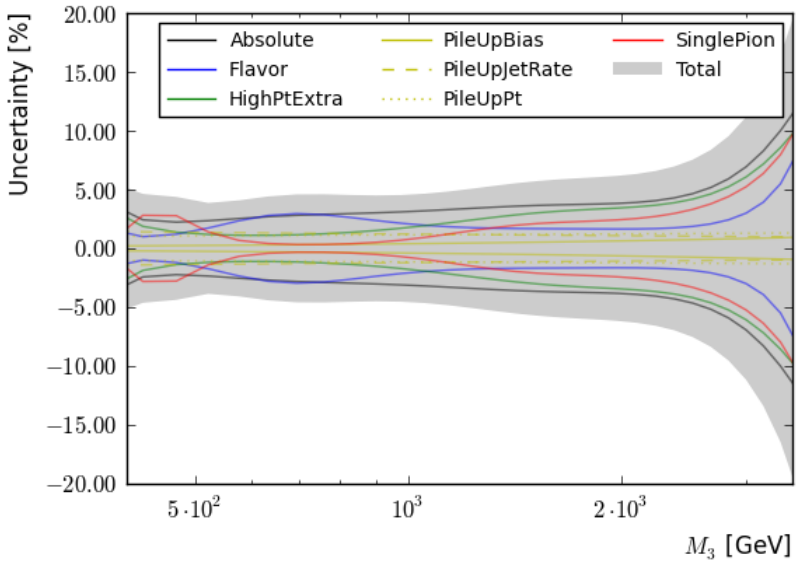
The covariance matrix describing the jet energy scale uncertainty is based on the symmetrised errors introduced by the uncertainty sources:

$$\mathrm{cov}_{\mathrm{JEC}}\left(m_{3,i}; m_{3,j}\right) = \sum_{k=0}^{n_{JEC}} \left[\frac{m_{3,i}^{\mathrm{up}(k)} - m_{3,i}^{\mathrm{down}(k)}}{2}\right] \left[\frac{m_{3,j}^{\mathrm{up}(k)} - m_{3,j}^{\mathrm{down}(k)}}{2}\right]$$

The parametrised jet energy uncertainties on the three-jet mass are presented in figure 6.19, 6.20 and 6.21. While the uncertainty of the central rapidity bin is dominated by the absolute and flavour uncertainty, the pile-up related uncertainty become more important for larger rapidities. Similarly, jet algorithms with larger cone sizes show a considerable impact from the pileup uncertainties for small three-jet masses. However this contribution gets smaller as the three-jet mass increases. The three-jet mass uncertainties for the two different cut scenarios exhibit behaviour which closely resemble each other above 500 GeV, where the different three-jet mass thresholds do not play a role anymore. For a large range of three-jet masses, the total jet energy scale uncertainty is between 6 and 7% for the exemplary, relative cut scenario. A detailed tabular overview of the main contributions to the final jet energy uncertainty is given in appendix D.3.1.

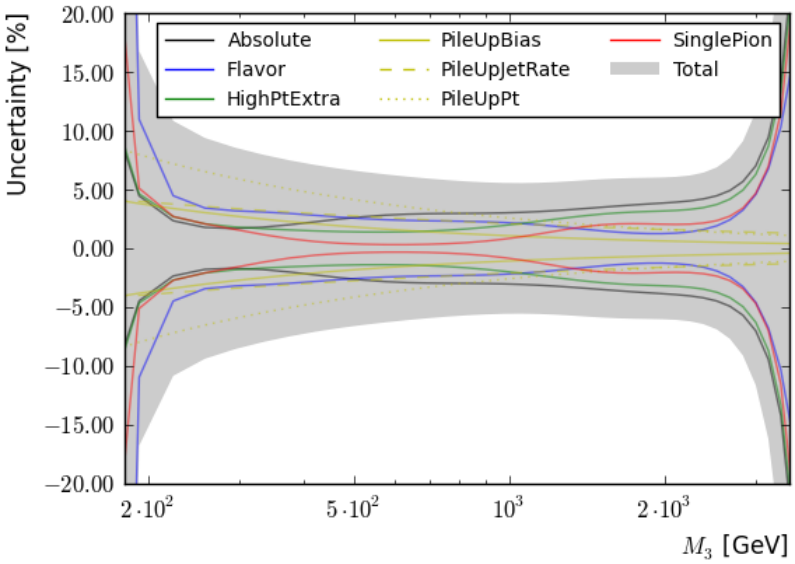


(a) Rel. cut, $|y_{\max}| < 1$, Anti-k_T 0.5

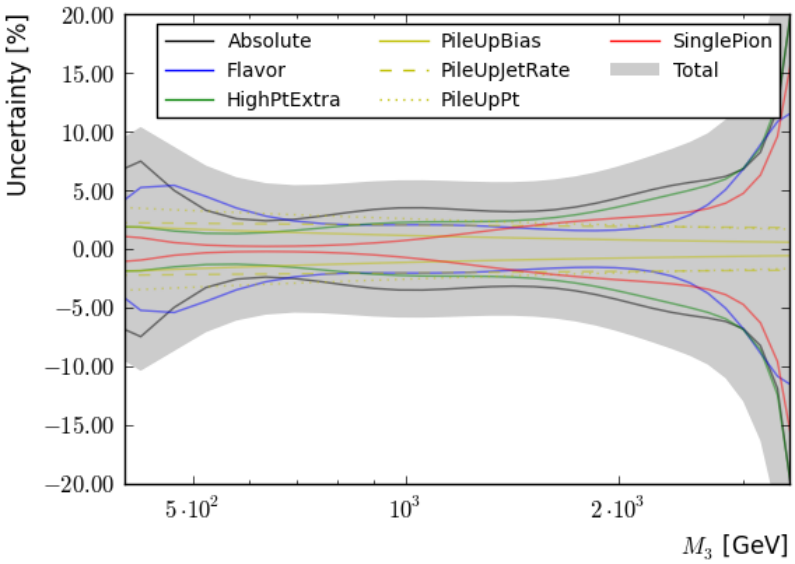


(b) Abs. cut, $|y_{\max}| < 1$, Anti-k_T 0.5

Figure 6.19: Jet energy scale uncertainty of the three-jet mass distribution with major uncertainty sources shown separately for the central rapidity bin ($|y_{\max}| < 1$). The uncertainties are presented for the relative cut scenario, absolute scenario and jet cone size $R = 0.5$. The three-jet mass uncertainties for the two different cut scenarios show very similar behaviour above 500 GeV, where the different three-jet mass thresholds are negligible.

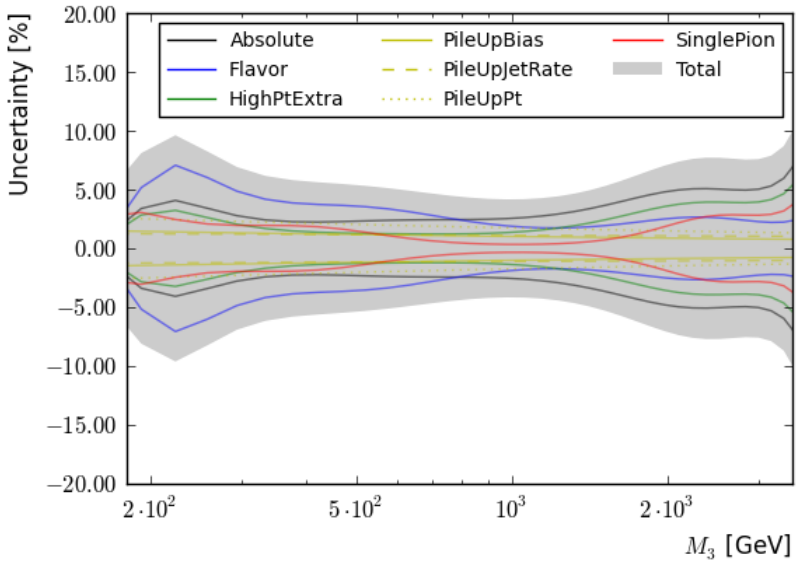


(a) Rel. cut, $|y_{\max}| < 1$, Anti-k_T 0.7

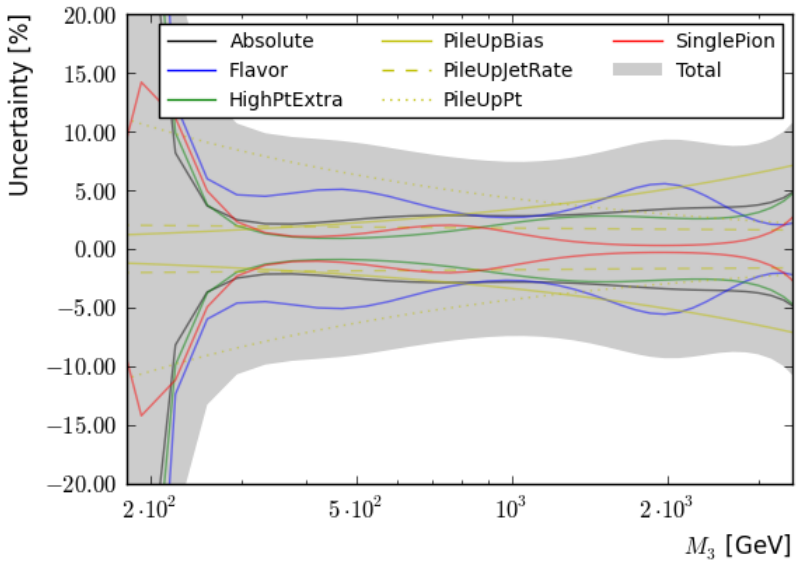


(b) Abs. cut, $|y_{\max}| < 1$, Anti-k_T 0.7

Figure 6.20: Jet energy scale uncertainty of the three-jet mass distribution with major uncertainty sources shown separately for the central rapidity bin ($|y_{\max}| < 1$). The uncertainties are presented for the relative cut scenario, absolute scenario and two different jet cone size $R = 0.7$. Comparison with figure 6.19 shows that jet algorithms with larger cone sizes show a considerable impact from the pileup uncertainties for small three-jet masses. However this contribution gets smaller as the three-jet mass increases.



(a) $1 < |y_{\max}| \leq 2$



(b) $2 < |y_{\max}| \leq 3$

Figure 6.21: Jet energy scale uncertainty of the three-jet mass distribution with major uncertainty sources shown separately. The uncertainties are presented for Anti-k_T 0.5 jets in the relative cut scenario in the outer two rapidity bins $1 < |y_{\max}| \leq 2$ (top) and $2 < |y_{\max}| \leq 3$ (bottom). Together with figure 6.19, this shows that the uncertainty of the central rapidity bin is dominated by the absolute and flavour uncertainty. However the pile-up related uncertainty become more important for larger rapidities.

6.8 Comparison with NLO predictions

With the unfolded distributions and uncertainties from the previous sections and the theory prediction from chapter 5, it is possible to look at the differences between NLO theory and measured data. A general overview of the measured distributions in different rapidity bins and the available NLO predictions are shown in figure 6.22.

Depending on the used parton distribution function, cut scenario and jet size, there is a difference of up to 20 percent between the theory prediction and data as shown in figures 6.23, 6.24 and 6.25. However, taking the uncertainties into consideration, the measurements are in general agreement with the theoretical predictions. It should be noted that the differences take the form of a flat offset and that no large differences between the data and theory shapes are observed.

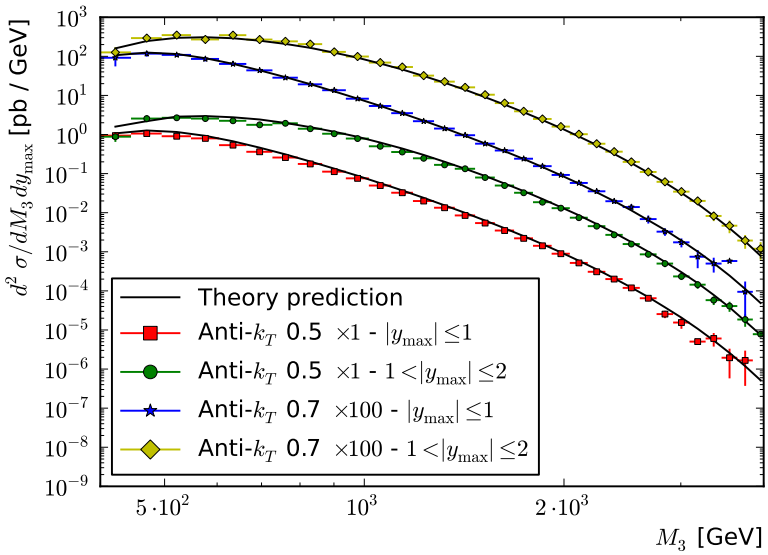
The behaviour that the cross-sections predicted by the NLO theory are higher than the measured cross-section is also observed in several other studies. The inclusive jet spectrum measurement by CMS[99] for example observed a flat offset of 15 to 20 percent.

Such a flat shift in the ratio of data and theory can be easily caused by the luminosity normalisation, which is afflicted with a 4% uncertainty. From the theory side, the parton distribution functions could cause the observed deviations. Since the measurement covers parts of the phase-space where the PDFs are not well measured, this measurement can also serve as input for the PDF fits. The offset can also be caused by missing terms in the perturbative series. Unfortunately, there is no NNLO three-jet mass prediction available to confirm this or give a better estimate. However the size of the K-factors discussed in 5 is between -50% and $+70\%$, leaving room for further NNLO contributions.

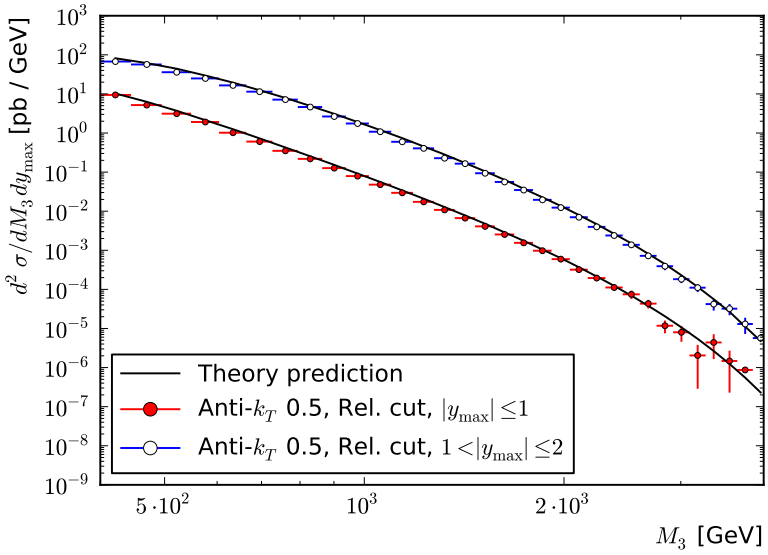
In order to exactly quantise the deviation between data and theory, normalisation factors n are fitted to the data-theory ratios. The best agreement between data and theory is observed by using the absolute cut scenario applied to Anti-k_T 0.7 jets in the innermost rapidity region and by calculating the theory prediction with the NNPDF 2.1 - NNLO parton density function. In this case the normalisation factor is given by:

$$n = 0.990^{+0.074}_{-0.066}$$

and the observed fit quality is $\chi^2/nDoF = 0.490/19$, which points to a large overestimation of the uncertainties for this particular set of jets, cuts and theory parameters. The very small residuals also show that the hypothesis of a flat offset is well motivated. The normalisation factors for the other measurements are summarised in appendix C.

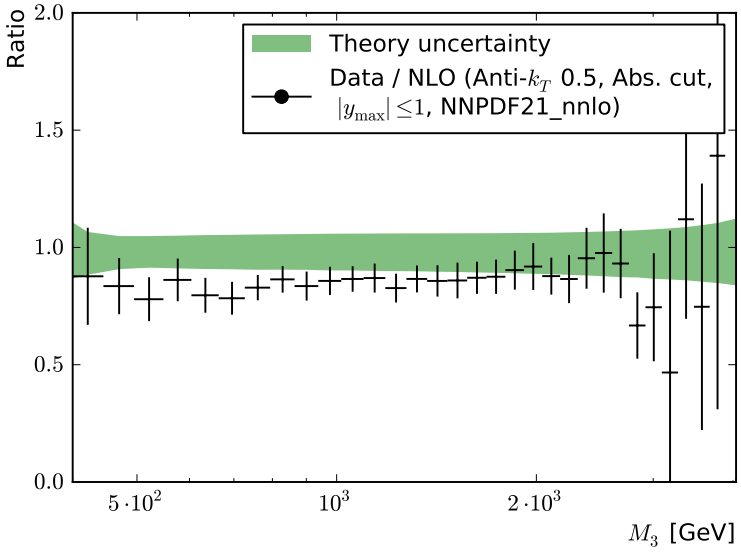


(a) Absolute cut scenario

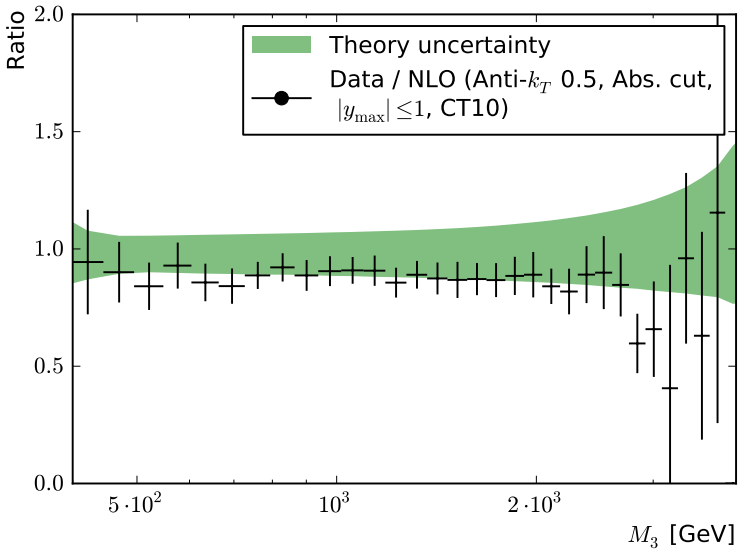


(b) Relative cut scenario

Figure 6.22: Unfolded three-jet mass distribution for different rapidity bins $y_{\max}$ from data together with the available NLO prediction. For the absolute cut scenario, the results for both Anti-k_T 0.5 and 0.7 jets are presented. In order to increase readability, the cross-section of for the larger jet size was multiplied by 100. The theory prediction shows a very good agreement for the double-differential three-jet mass cross-section with data over several orders of magnitude.

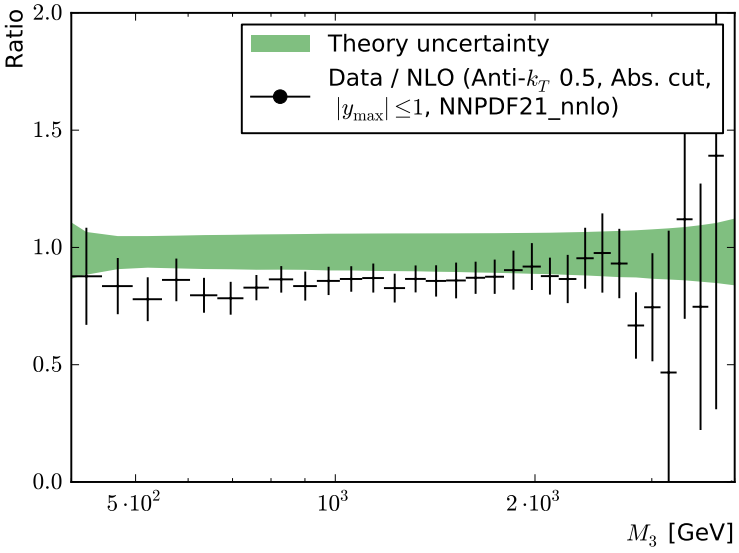


(a) NNPDF 2.1 - NNLO

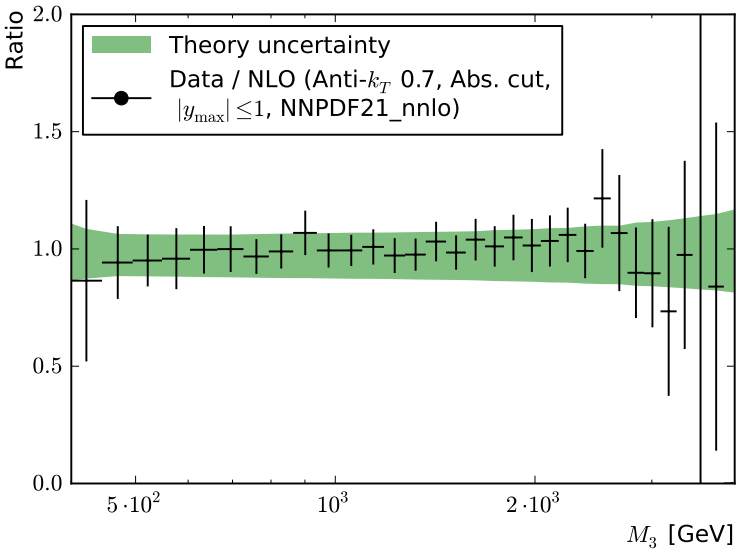


(b) CT10 - NLO

Figure 6.23: Ratio of the unfolded three-jet mass distribution and the NLO predictions based on the NNPDF2.1 NNLO (top) and the CT10 (bottom) parton distribution functions. The absolute cut scenario was applied to Anti-k_T 0.5 jets. The green band represents the total theory uncertainty, while the error bars are given by the total measurement uncertainty. Both PDFs show an agreement within the quoted uncertainties.

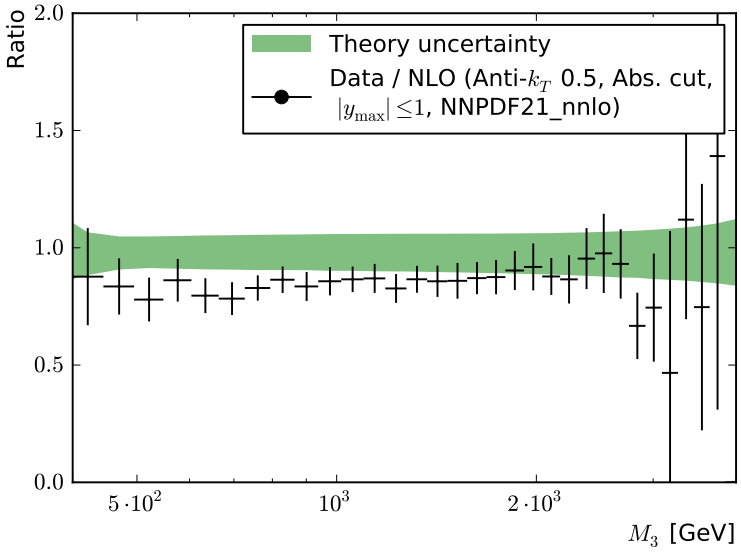


(a) Anti-k_T 0.5

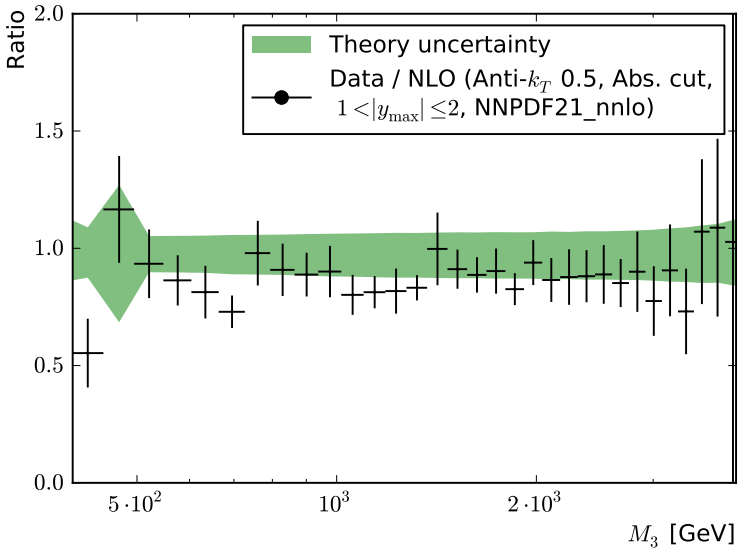


(b) Anti-k_T 0.7

Figure 6.24: Ratio of the unfolded three-jet mass distribution and the NLO predictions for the absolute cut scenario based on the NNPDF2.1 NNLO for Anti-k_T 0.5 (top) and 0.7 jets (bottom). The total theory uncertainty is given by the green band, while the error bars represent the total measurement uncertainty. With the larger jet size, a much better agreement between data and theory prediction can be observed.



(a) $|y_{\max}| < 1$



(b) $1 < |y_{\max}| \leq 2$

Figure 6.25: Comparison between the unfolded three-jet mass distribution and the NLO predictions for the inner (top) and outer (bottom) rapidity region based on the NNPDF2.1 NNLO PDF for Anti-k_T 0.5 jets in the absolute cut scenario. The error bars are given by the total measurement uncertainty, while the green band represents the total theory uncertainty. While the data in the outer rapidity bin suffers from a smaller number of available events, leading to larger uncertainties and statistical fluctuations, the prediction still shows a good agreement in this challenging phase-space region.

Chapter 7

Conclusion and Outlook

The successful start of the LHC has opened up a new era in particle physics. In 2011, the accelerator delivered collisions with a centre-of-mass energy of $\sqrt{s} = 7000\,\text{GeV}$, which allow access to new regions of the QCD phase-space.

The dataset collected by the CMS experiment is successfully used in this study to measure the double differential three-jet mass cross-section $d^2\sigma/dm_3dy_{\max}$ from the selected three-jet events.

In addition, detailed detector studies on the performance of the individual sub-detectors and the jet reconstruction allow to determine precise uncertainties on this measurement. In order to remove detector effects, like the influence of to the finite three-jet mass resolution, several unfolding algorithms were studied. For this study the iterative Bayesian unfolding algorithm yielded the best results. Applying this unfolding method allows to directly compare the results of this study with an identical analysis on data from another experiment.

In order to investigate differences between the measured data and theory, theory predictions for the three-jet mass are calculated at next-to-leading order precision. The uncertainties due to scale choice and value, non-perturbative corrections, PDF and α_s are studied.

The comparison between data and theory shows a general overestimation of the cross-section in the theory calculation. However for certain choices for the jet size, cut scenario and theory parameters, the agreement between data and theory is on the percent level, with a flat offset given by the following normalisation factor:

$$n = 0.990^{+0.074}_{-0.066}$$

The results of this measurement could help to further reduce the observed differences by supplying data to the PDF fit community. Another application of this measurement is given in the following outlook.

7.1 Outlook: Sensitivity to α_s

One of the major standard model parameters entering the theory predictions is the value of the strong coupling constant α_s. Since a direct measurement of α_s must be done in conjunction with a complex PDF fit, which is outside the scope of this study, only a simple comparison between data and NLO calculations using different values of α_s, is performed. The value of α_s, which shows the best compatibility can be determined by quantising the difference between the measured distribution $m_{3,i}^{\text{Data}}$ and the different NLO theory predictions $m_{3,i}^{\text{Theory}}(\alpha_s)$ with a χ^2 test. This test is based on the following formula:

$$\chi^2 = \sum_{i,j=n_A}^{n_B} \left(m_{3,i}^{\text{Data}} - m_{3,i}^{\text{Theory}}\right) \left[\text{cov}_{\text{sum}}^{-1}(m_{3,i}; m_{3,j})\right] \left(m_{3,j}^{\text{Data}} - m_{3,j}^{\text{Theory}}\right)$$

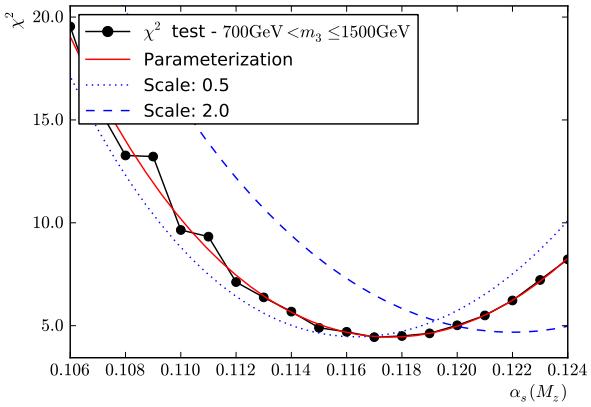
The inverse covariance matrix $\text{cov}_{\text{sum}}^{-1}$ is determined from the matrix inversion of the total three-jet mass covariance cov_{sum}. The total three-jet mass covariance is the sum of the covariances due to jet energy scale, unfolding and luminosity uncertainty on the measurement on one hand and the theory covariances on the other hand:

$$\text{cov}_{\text{sum}} = \text{cov}_{\text{JEC}} + \text{cov}_{\text{unf}} + \text{cov}_{\text{lumi}} + \text{cov}_{\text{NLO}}$$

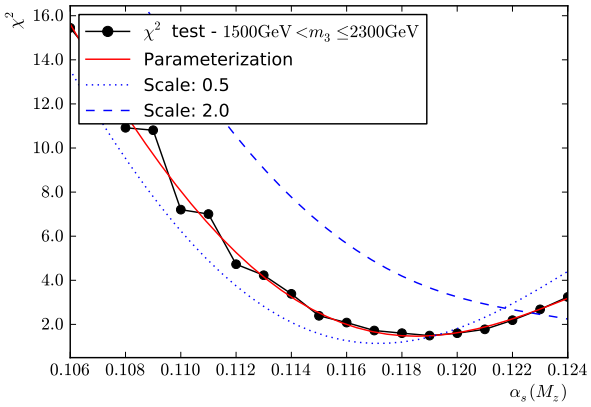
The theory covariance includes all discussed theory uncertainty sources except the scale uncertainty, which is handled separately. In order to study the running of the strong coupling constant, the comparison between data and theory is performed for $|y_{\max}| < 1$ in three different three-jet mass regions shown in table 7.1, each with a three-jet mass window width of 800 GeV. The table contains for both cut scenarios the weighted average of the three-jet mass in each region together with the three jet mass resolution at this value.

For the theory calculations in this comparison, the NNPDF2.1 NNLO PDF set was used. This PDF is available with 19 different α_s variations, ranging between $a_s(M_Z) = 0.106$ and 0.124. Figure 7.1 shows the result of the χ^2 test for the different values of $\alpha_s(M_Z)$. In order to take the scale uncertainty into consideration, a two point variation of the scale is performed in this fit.

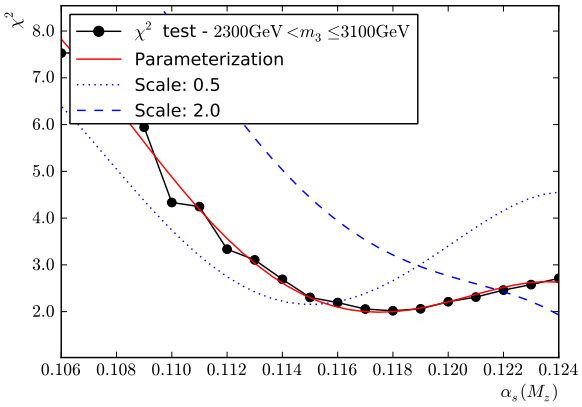
The minimum is determined from a parametrisation of the χ^2 distribution using a polynomial $p(\alpha_s(M_Z))$ of degree 4. The value of the strong coupling constant at the minimum point $(\alpha_{s,0}(M_Z), p(\alpha_{s,0}(M_Z)))$ of this polynomial is taken as the central value for α_s in each three-jet mass region. The asymmetric error on this value is determined from the position of the solutions of equation $p(\alpha_s(M_Z)) = p(\alpha_{s,0}(M_Z)) + 1$.



(a)



(b)



(c)

Figure 7.1: Result of the χ^2 comparison between data and a set of theory predictions with different values of $\alpha_s(M_Z)$. The comparison is performed in three different three-jet mass regions (700 GeV − 1500 GeV, 1500 GeV − 2300 GeV and 2300 GeV − 3100 GeV). In order to determine the minimum of the χ^2 distribution, the curve is parametrised using a polynomial of degree 4. In order to quantise the scale dependence, this α_s fit procedure is performed for three different scales variations ($0.5\mu, \mu$ and 2μ)

Three-jet mass range	$700 - 1500\,\text{GeV}$	$1500 - 2300\,\text{GeV}$	$2300 - 3100\,\text{GeV}$
$\bar{m}_{3h}$ Absolute cut	$866 \pm 38\,\text{GeV}$	$1670 \pm 57\,\text{GeV}$	$2518 \pm 72\,\text{GeV}$
$\bar{m}_{3b}$ Relative cut	$886 \pm 37\,\text{GeV}$	$1680 \pm 58\,\text{GeV}$	$2517 \pm 77\,\text{GeV}$
N_{obs}	9	7	5

Table 7.1: Overview of the three-jet mass ranges for the comparison between data and theory. The weighted average $\bar{m}_3$ of the three-jet mass is presented for the two cut scenarios. The uncertainty given for this average is determined from the three-jet mass resolution. N_{obs} gives the number of three-jet mass bins within in each mass range.

Three-jet mass range	
$700 - 1500\,\text{GeV}$	$\alpha_s\,(886\,\text{GeV}/2) = 0.0951^{+0.0035\ +0.0028}_{-0.0021\ -0.0008}$
$1500 - 2300\,\text{GeV}$	$\alpha_s\,(1680\,\text{GeV}/2) = 0.0888^{+0.0022\ +0.0027}_{-0.0019\ -0.0008}$
$2300 - 3100\,\text{GeV}$	$\alpha_s\,(2517\,\text{GeV}/2) = 0.0844^{+0.0032\ +0.0019}_{-0.0023\ -0.0013}$

Table 7.2: Value of the strong coupling constant α_s as measured from the three-jet mass distribution in the inner rapidity region for the Anti-k_T 0.7 jet algorithm at three different scales. The quoted uncertainties are the total measurement uncertainties and the scale uncertainties.

In order to get the value of α_s at the scale of the measurement, the fitted $\alpha_s\,(M_Z)$ values and uncertainties are propagated using an α_s evolution code (GRV[118]). The scale of the measurement is chosen to be the same scale used in the theory prediction $m_3/2$. The central value for this scale is the average three-jet mass in each bin as shown in table 7.1

The result of this procedure is shown in table 7.2 and figure 7.2. The correlated and statistical uncertainties are separately shown. The correlated uncertainty is derived from the propagation of the normalisation uncertainty. The running of α_s is very well described by the measurement. Figure 7.3 shows how this result fits into the established set of measurements.

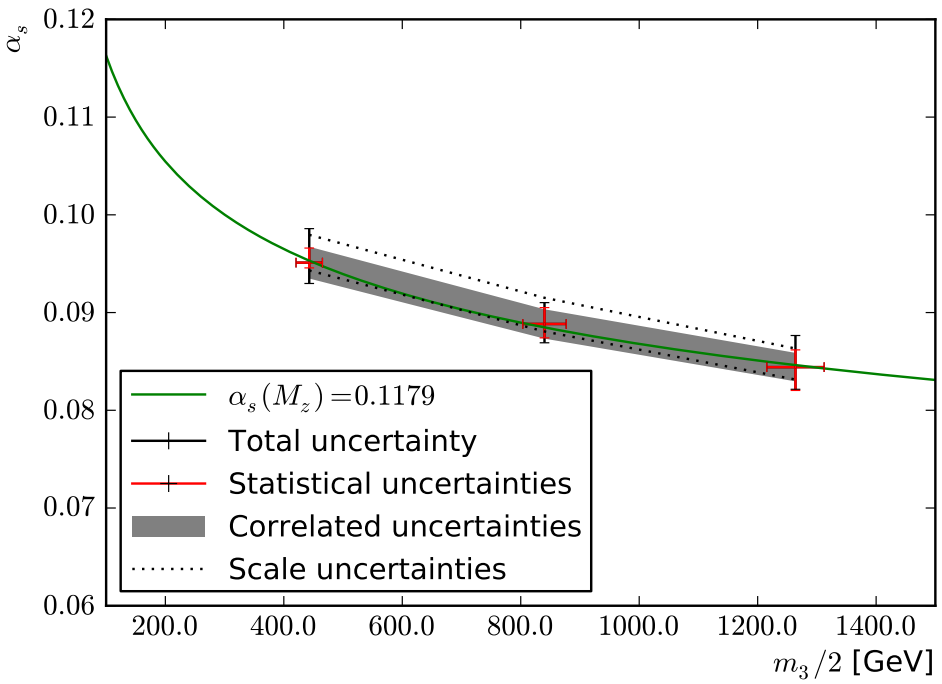


Figure 7.2: Comparison between the running of the strong coupling constant as predicted using the world average for $\alpha_s(M_Z) = 0.1179$ and the running as measured from the three-jet mass distribution. The measured values are compatible with the world average within the uncertainties. Two major components of the total uncertainty is the uncorrelated, statistical uncertainty (red) and the correlated uncertainty (grey band). Another uncertainty not included in the total uncertainty is the scale uncertainty, which is also shown (dashed line).

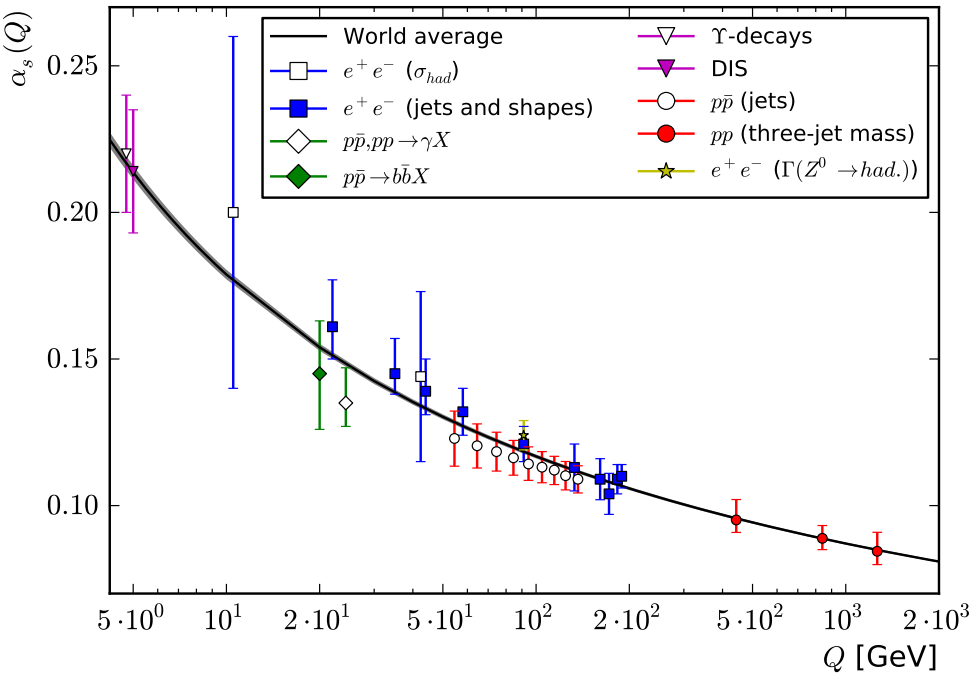


Figure 7.3: Overview of measurements of the running of the strong coupling constant α_s, including the world average for $\alpha_s(M_Z) = 0.1179$ and its uncertainty (band). The presented three-jet mass measurement extends the covered phase-space at high scales. The uncertainties on the measurements include all experimental and theory uncertainties. (Sources: [119], [120])

Appendix A

Unfolding

This appendix gives a short introduction to the unfolding problem. A detailed overview of the subject is given in [116, 121]. The unfolding problem appears already in the basic measurement of the distribution $f(x)$ of some kinematical quantity x. With access to an ideal detector, this task is reduced to a simple measurement of the quantity x and obtaining $f(x)$ from a histogram. Unfortunately, ideal detectors do not exist in nature so a real detector will influence all performed measurements to varying degree due to several effects. The most important issues are limited detector acceptance, finite resolution, transformed quantities and backgrounds.

Since the probability to observe an event, which is called the detector acceptance, is less than one for a real detector, a certain loss of signal events will occur. Usually, this detector acceptance also depends on the quantity x. However in many cases x is not directly accessible with the detector, so that instead of the quantity x, only a correlated observable y can be measured in the detector. The relationship between the quantities x and y are described via some, not necessarily bijective, transformation. The measured quantity y is additionally smeared due to the finite resolution inherent to any real detector. This means that the measured quantity and the true quantity x are only linked via a statistical relationship, further breaking any bijective properties of the mapping between x and y. It is also possible to measure some quantity y in the detector, which has no relationship to x since it was caused by some background process.

The relationship between the true distribution $f(x)$ and the measured distribution $g(y)$ with some background distribution $b(y)$ can be expressed mathematically with the following integral equation:

$$g(y) = \int A(y,x)\ f(x)dx + b(y) \tag{A.1}$$

The function $A(y, x)$ is called the transfer function or detector response. It contains the information about the detector resolution, acceptance and other transformations between true quantity x and measured quantity y in the form of probabilities. For a given true quantity x_0, the detector response $A(y, x_0)$ gives the probability density function of measuring the quantity y. The integral equation A.1 is a so-called Fredholm integral equation of the first kind and the process of solving the inverse problem of reconstructing the true distribution $f(x)$ from the measured distribution $g(y)$ is called unfolding, although the detector response is allowed to be more general than a convolution integral[1].

When the goal is just to compare some theory prediction $f_{\text{theory}}(x)$ to the measured distribution $g(y)$, it is sufficient to apply equation A.1 to the theory distribution to get $g_{\text{theory}}(y)$, which can directly be used for the comparison. However it is preferable to supply the measured distribution in unfolded form, since it allows third-parties to easily make comparisons between the measured $f(x)$ distribution and some new theory $f'_{\text{theory}}(x)$. In addition, to compare results measured with two different detectors, it is necessary that at least one of the measurements is unfolded.

In order to understand the general principles behind trying to solve the unfolding problem, assume that equation A.1 can successfully be linearised and discretised. In this case the problem is equivalent to solving the matrix equation

$$\vec{y} = \mathbf{A}\vec{x} + \vec{b} \tag{A.2}$$

with two m-dimensional vectors $\vec{y}$ and $\vec{b}$ representing the histogram entries of the measured and background distribution, an n-dimensional vector $\vec{x}$ representing the true distribution, and $m \times n$ Matrix $\mathbf{A}$ describing the detector response.

From here it is easy to show that the unfolding problem, like many inverse problems, is a so-called "ill posed" problem. Neither uniqueness nor existence of a solution is guaranteed, as shown by the following example:

$$\mathbf{A} = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}, \quad \vec{x}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \vec{x}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

In this case it is not possible to answer the question, which true distribution $\vec{x}_i$ is the correct one after measuring $\vec{y}$. Therefore additional constraints on the detector response $\mathbf{A}$ have to be introduced.

[1] A convolution integral imposes on the detector response $A(y, x) = A(y - x)$ and for the background $b(y) = 0$

A.1 Singular value decomposition

The singular value decomposition theorem states that every $m \times n$ matrix $\mathbf{M}$ can be decomposed into three matrices $\mathbf{U}, \mathbf{V}$ and $\mathbf{\Sigma}$:

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$

These matrices fulfil the following properties:

- $\mathbf{U}$ is a unitary $m \times m$ matrix,
- $\mathbf{V}^*$ is the conjugate transpose of the unitary $n \times n$ matrix $\mathbf{V}$,
- and $\mathbf{\Sigma}$ is a rectangular $m \times n$ diagonal matrix.

For unfolding problems, $\mathbf{M}$ is a real matrix and it can be shown that in this case $\mathbf{U}$ and $\mathbf{V}$ are also real matrices. The positive diagonal entries σ_i of the matrix $\mathbf{\Sigma}$ are called the singular values of the matrix $\mathbf{M}$ and it is common to list them in descending order. The singular values are the square roots of the non-zero eigenvalues of $\mathbf{M}\mathbf{M}^*$ and $\mathbf{M}^*\mathbf{M}$.

Applying this decomposition on the (real-valued) detector response $\mathbf{A}$, it is possible to define a generalised inverse matrix $\mathbf{A}^\#$

$$\begin{aligned} \mathbf{A} &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}} \\ \mathbf{A}^\# &= \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^{\mathrm{T}} \end{aligned}$$

which is equivalent to the inverse matrix $\mathbf{A}^{-1}$ for square matrices.

A.2 Bayesian unfolding

The Bayesian unfolding method applies Bayesian inference to the problem of unfolding and is an example for an iterative unfolding algorithm. Following the paper[117], which introduced the method, the algorithm is explained in terms of causes C_i and effects E_j. Causes can be identified with the true values $\vec{x}$ to be determined, while the effects are the measured $\vec{y}$. The detector response $\mathbf{A}$ takes the form of the probability $P(E_j|C_i)$ to observe an effect E_j from a cause C_i. To solve the unfolding problem, it is necessary to determine the probability $P(C_i|E_j)$, which allows to calculate for a number $n(E_j)$ of observed events, how many events $\hat{n}(C_i)$ can be assigned to a certain cause C_i. Bayes theorem states that

$$P(C_i|E_j) = \frac{P(E_j|C_i) \cdot P(C_i)}{P(E_j)} = \frac{P(E_j|C_i) \cdot P(C_i)}{\sum_{k=1}^{N_c} P(E_j|C_k) \cdot P(C_k)}$$

with the number of causes N_c. Taking the efficiency $\varepsilon_i \neq 0$ to detect a cause C_i in any possible effect into account, the expected number of events $\hat{n}(C_i)$ for cause C_i is given by

$$\hat{n}(C_i) = \frac{1}{\varepsilon_i} \sum_{j=1}^{N_E} n(E_j) \cdot P(C_i|E_j)$$

with the number of effects N_E. It is now necessary to choose an initial distribution of $P_0(C_i)$. This can be flat prior $P_0(C_i) = 1/n_C$, or in case of the RooUnfold implementation, the truth distribution from Monte-Carlo simulation. This initial estimate $P_0(C_i)$ of $P(C_i)$ is used to first calculate $P(C_i|E_j)$, then $\hat{n}(C_i)$ and $P_1(C_i)$:

$$P_1(C_i) = \frac{\hat{n}(C_i)}{\sum_{k=1}^{N_C} \hat{n}(C_k)}$$

This result for $P_1(C_i)$ can be used in another iteration in place of $P_0(C_i)$, which allows to do the calculation for any number of iterations.

Appendix B

Technical details

B.1 Datasets

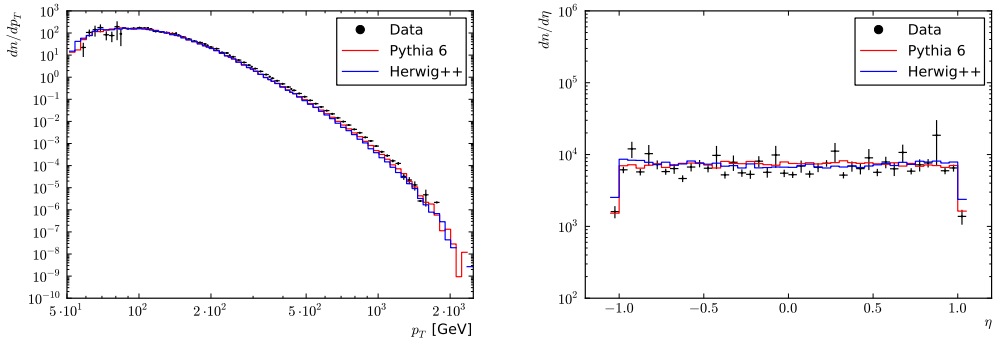
This analysis is based on five datasets recorded by the CMS detector. In addition, two sets of Monte-Carlo datasets, including a full detector simulation, were used. These two samples were produced by the official CMS Monte-Carlo production team. Another Monte-Carlo sample was privately produced with the event generator Sherpa (B.4). In order to confirm the validity of the Monte-Carlo simulation, basic event variables are compared with data.

Table B.1: Overview of the used datasets

Sample name	Run range	Dataset Bookkeeping System entry
Data - 1	160329-163869	`/Jet/Run2011A-May10ReReco-v1/AOD`
Data - 2	165088-168229	`/Jet/Run2011A-PromptReco-v4/AOD`
Data - 3	170053-172619	`/Jet/Run2011A-05Aug2011-v1/AOD`
Data - 4	172620-175770	`/Jet/Run2011A-PromptReco-v6/AOD`
Data - 5	175832-180296	`/Jet/Run2011B-PromptReco-v1/AOD`

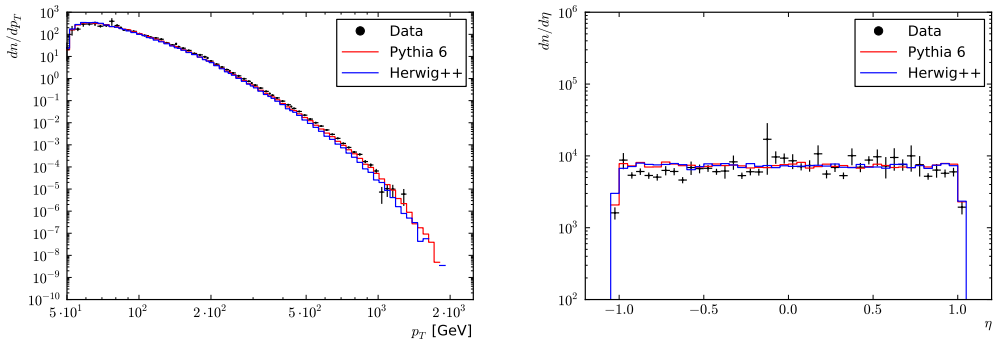
B.1.1 Overview of the used Monte-Carlo samples

- Pythia 6.424
  - Tune Z2
  - Phase-space: reweighted, $15\,\text{GeV} < \hat{p}_T < 3000\,\text{GeV}$
  - DBS dataset path:
    /QCD_Pt-15to3000_TuneZ2_Flat_7TeV_pythia6/
    Fall11-PU_S6_START42_V14B-v1/AODSIM
- Herwig++ 2.4.2
  - Default tune
  - Phase-space: reweighted, $15\,\text{GeV} < \hat{p}_T < 3000\,\text{GeV}$
  - DBS dataset path:
    /QCD_Pt-15to3000_Tune23_Flat_7TeV_herwigpp/
    Fall11-PU_S6_START42_V14B-v2/AODSIM
- Sherpa 1.2.2
  - Default tune (UE/Hadronisation setting varied for NP corrections)
  - Phase-space: $p_T > 30, 140, 230, 640, 1070$ and $1790\,\text{GeV}$
  - DBS dataset path: private sample



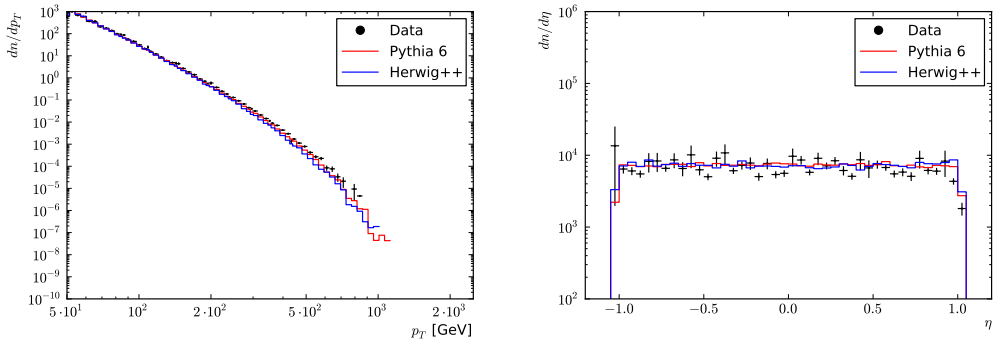
(a) Jet p_T (b) Jet η

Figure B.1: Validation of basic Monte-Carlo variables - leading jet



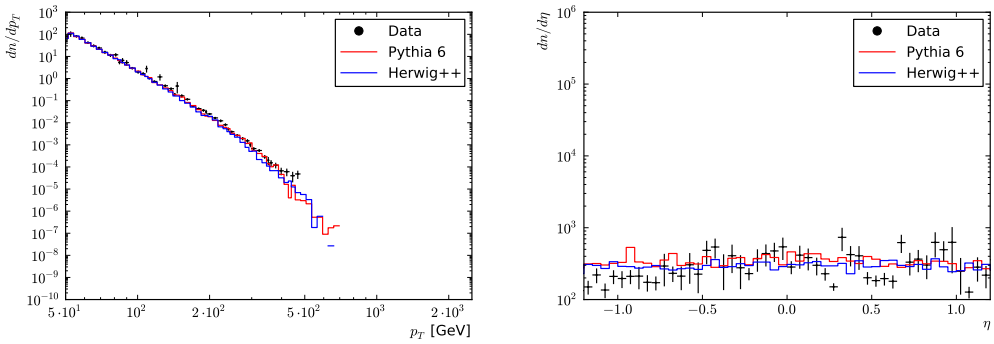
(a) Jet p_T (b) Jet η

Figure B.2: Validation of basic Monte-Carlo variables - second jet



(a) Jet p_T (b) Jet η

Figure B.3: Validation of basic Monte-Carlo variables - third jet



(a) Jet p_T (b) Jet η

Figure B.4: Validation of basic Monte-Carlo variables - fourth jet

B.2 Run and luminosity sections

Since this analysis applies additional run and lumi section selection criteria, it is necessary to give the list of used run and luminosity sections. Since the list of luminosity sections is too long to show here, the program to calculate the list is given instead.

```
#!/bin/env python

import os, sys, gzip

# RUN == STRING, LUMI == INT

def readRecorded(fnRecorded):
  recorded = {}
  fp_rec = open(fnRecorded)
  print fnRecorded
  for line in fp_rec.readlines():
    tmp = map(str.strip, line.split(","))
    if not tmp:
      continue
    run, lumi, rec = tmp
    recorded.setdefault(run, dict())[int(lumi)] = float(rec)
  return recorded

def checkLumi(fnCert, fnRecorded, fnBrokenPSList):
  broken = {}
  for fnBrokenPS in fnBrokenPSList:
    fp_broken = open(fnBrokenPS)
    for line in fp_broken:
      tmp = map(str.strip, line.split(","))
      if not tmp:
        continue
      run, lumi = tmp
      broken.setdefault(run, list()).append(int(lumi))

  recorded = []
  for fn in fnRecorded:
    recorded.append((readRecorded(fn), fn))

  fp_miss = open(fn + ".missing", 'w')
```

```
found = {}
lost = {}
included = {}
fp_cert = open(fnCert)
certDict = eval(fp_cert.read())
for run in certDict:
  for lr in certDict[run]:
    ls, le = lr
    for lumi in range(ls, le + 1):
      missing = False
      if lumi in broken.get(run, []):
        print "BROKEN PS", run, lumi
        missing = True
      for (rec, src) in recorded:
        try:
          rec[run][lumi]
        except:
          fp_miss.write('%8d %5d %s\n' % (int(run), int(lumi), src))
          print "MISSING", run, lumi, src
          missing = True
      if not missing:
        found.setdefault(run, list()).append([lumi, lumi])
        for (rec, src) in recorded:
          included.setdefault(src, 0)
          included[src] += rec[run][lumi] / 1e6
      else:
        for (rec, src) in recorded:
          lost.setdefault(src, 0)
          try:
            lost[src] += rec[run][lumi] / 1e6
          except:
            pass

print fnCert
print "Included", included
print "Lost", lost

fp_out = open('output.json', 'w')
fp_out.write(str(found))
```

```
def doAllCerts(fnRec):
  for fn in os.listdir("."):
    if fn.startswith("Cert") and fn.endswith(".txt.nice"):
      checkLumi(fn, fnRec, ["brokenPS.csv", "excluded.csv"])
      fnNew = fn + ".haslumi" #+ fnRec.split(".")[0]
      os.system("~/gc/scripts/lumiInfo.py -J output.json > %s" % fnNew)

doAllCerts(["lumiCalc2.recorded.dat", "pixelCalc.recorded.dat"])
```

B.3 HLT menus

The analysis uses an adaptive algorithm to choose the optimal set of HLT paths in each lumi section.

B.3.1 HLT menu ranges

The following list shows the run ranges, in which a certain HLT menu was active:

/cdaq/physics/Run2011/5e33/v1.4/HLT/V3 178420-178479
/cdaq/physics/Run2011/5e32/v5.3/HLT/V1 160955-161016
/cdaq/physics/Run2011/3e33/v2.3/HLT/V2 176467-176469
/cdaq/physics/Run2011/5e32/v5.2/HLT/V7 160939-160943
/cdaq/physics/Run2011/3e33/v2.2/HLT/V3 176286-176309
/cdaq/physics/Run2011/3e33/v4.0/HLT/V5 177222, 177317, 177318, 177319, 177449, 177452, 177730, 177776, 177782, 177783, 177788, 177789, 177790, 177791, 177875, 177878
/cdaq/physics/Run2011/5e33/v1.4/HLT/V5 178712-179889
/cdaq/physics/Run2011/3e33/v3.1/HLT/V1 176697-177053
/cdaq/physics/Run2011/5e33/v1.4/HLT/V4 178703-178708
/cdaq/physics/Run2011/1.4e33/v1.2/HLT/V3 167551-167913
/cdaq/physics/Run2011/3e33/v1.2/HLT/V1 173657-173692
/cdaq/physics/Run2011/5e32/v5.3/HLT/V2 161103-161176
/cdaq/physics/Run2011/1e33/v1.3/HLT/V12 165364-165364
/cdaq/physics/Run2011/1e33/v1.3/HLT/V13 165402-165633
/cdaq/physics/Run2011/5e33/v2.2/HLT/V4 180241-180252
/cdaq/physics/Run2011/5e32/v8.1/HLT/V8 163296-163334
/cdaq/physics/Run2011/3e33/v4.0/HLT/V6 177718-177719
/cdaq/physics/Run2011/3e33/v5.0/HLT/V1 178098-178380
/cdaq/physics/Run2011/1.4e33/v1.1/HLT/V1 167039-167043
/cdaq/physics/Run2011/5e32/v8.3/HLT/V4 163757-163869
/cdaq/physics/Run2011/5e32/v8.3/HLT/V2 163738-163738
/cdaq/physics/Run2011/5e32/v8.1/HLT/V5 163270-163270
/cdaq/physics/Run2011/5e32/v8.1/HLT/V6 163286-163289
/cdaq/physics/Run2011/1e33/v2.4/HLT/V4 166374-166512
/cdaq/physics/Run2011/1e33/v2.4/HLT/V5 166514-166782
/cdaq/physics/Run2011/1e33/v2.4/HLT/V6 166784-166787
/cdaq/physics/Run2011/1e33/v2.4/HLT/V2 166161-166164
/cdaq/physics/Run2011/5e32/v6.2/HLT/V3 162803-162909
/cdaq/physics/Run2011/5e32/v6.2/HLT/V2 162762-162765
/cdaq/physics/Run2011/3e33/v4.0/HLT/V3 177201-177201
/cdaq/physics/Run2011/3e33/v2.1/HLT/V2 176161-176207
/cdaq/physics/Run2011/1e33/v2.4/HLT/V8 166839-166967

/cdaq/physics/Run2011/3e33/v4.0/HLT/V2 177074-177184
/cdaq/physics/Run2011/5e32/v6.2/HLT/V4 163046-163261
/cdaq/special/Tests/HighRateTest/v1.0/HLT/V1 175872-175874
/cdaq/physics/Run2011/5e33/v2.2/HLT/V2 179959-180093
/cdaq/physics/Run2011/1e33/v2.3/HLT/V1 165970-165993
/cdaq/physics/Run2011/1e33/v2.3/HLT/V3 166011-166150
/cdaq/physics/Run2011/5e32/v5.1/HLT/V3 160871-160874
/cdaq/physics/Run2011/5e32/v6.1/HLT/V3 161217-161217
/cdaq/physics/Run2011/3e33/v3.0/HLT/V2 176545-176548
/cdaq/physics/Run2011/5e32/v6.1/HLT/V6 161310-161312
/cdaq/physics/Run2011/5e32/v6.1/HLT/V5 161222-161233
/cdaq/physics/Run2011/1e33/v2.5/HLT/V1 166346-166346
/cdaq/physics/Run2011/1.4e33/v1.2/HLT/V1 167078-167284
/cdaq/physics/Run2011/5e32/v4.2/HLT/V7 160577-160578
/cdaq/physics/Run2011/5e32/v4.2/HLT/V6 160431-160431
/cdaq/physics/Run2011/3e33/v1.1/HLT/V4 173380-173439
/cdaq/physics/Run2011/3e33/v2.0/HLT/V7 175860, 175863, 175865, 175866, 175877, 175881, 175886, 175887, 175888, 175906, 175910, 175921
/cdaq/physics/Run2011/1e33/v1.3/HLT/V6 165205-165205
/cdaq/physics/Run2011/1e33/v1.3/HLT/V7 165208-165208
/cdaq/physics/Run2011/3e33/v2.1/HLT/V1 175973-176023
/cdaq/physics/Run2011/3e33/v1.1/HLT/V1 173236-173241
/cdaq/physics/Run2011/1e33/v1.3/HLT/V2 165088-165121
/cdaq/physics/Run2011/3e33/v1.1/HLT/V3 173243-173243
/cdaq/physics/Run2011/2e33/v1.2/HLT/V7 172478-173198
/cdaq/physics/Run2011/2e33/v1.2/HLT/V4 171219-172033
/cdaq/physics/Run2011/2e33/v1.2/HLT/V5 172163-172411
/cdaq/physics/Run2011/5e32/v8.2/HLT/V3 163337-163668
/cdaq/physics/Run2011/2e33/v1.2/HLT/V1 170826-171178

B.3.2 HLT menu trigger list

The following list shows which single jet HLT paths are active as a function of the HLT menu name. In case all versions of a certain menu is using the same set of HLT paths, this is marked with a wild-card:

/cdaq/physics/Run2011/3e33/v3.0/HLT/V2, /cdaq/physics/Run2011/3e33/v4.0/HLT/V5, /cdaq/physics/Run2011/3e33/v4.0/HLT/V6, /cdaq/physics/Run2011/3e33/v4.0/HLT/V2, /cdaq/physics/Run2011/3e33/v4.0/HLT/V3, /cdaq/physics/Run2011/3e33/v5.0/HLT/V1, /cdaq/physics/Run2011/3e33/v3.1/HLT/V1

HLT_Jet30_v6, HLT_Jet60_v6, HLT_Jet110_v6, HLT_Jet190_v6, HLT_Jet240_v6, HLT_Jet300_v6, HLT_Jet370_v7, HLT_Jet800_v2

/cdaq/physics/Run2011/3e33/v2.3/HLT/V2, /cdaq/physics/Run2011/3e33/v1.2/HLT/V1, /cdaq/physics/Run2011/3e33/v2.1/HLT/V2, /cdaq/physics/Run2011/3e33/v1.1/HLT/V4, /cdaq/physics/Run2011/3e33/v2.0/HLT/V7, /cdaq/special/Tests/HighRateTest/v1.0/HLT/V1, /cdaq/physics/Run2011/3e33/v2.2/HLT/V3, /cdaq/physics/Run2011/3e33/v2.1/HLT/V1, /cdaq/physics/Run2011/3e33/v1.1/HLT/V1, /cdaq/physics/Run2011/3e33/v1.1/HLT/V3

HLT_Jet30_v6, HLT_Jet60_v6, HLT_Jet110_v6, HLT_Jet190_v6, HLT_Jet240_v6, HLT_Jet300_v5, HLT_Jet370_v6, HLT_Jet800_v1

/cdaq/physics/Run2011/1e33/v2.4/HLT/V4, /cdaq/physics/Run2011/1e33/v2.4/HLT/V5, /cdaq/physics/Run2011/1e33/v2.4/HLT/V6, /cdaq/physics/Run2011/1e33/v2.4/HLT/V2, /cdaq/physics/Run2011/1e33/v2.4/HLT/V8, /cdaq/physics/Run2011/1e33/v2.3/HLT/V1, /cdaq/physics/Run2011/1e33/v2.3/HLT/V3

HLT_Jet30_v4, HLT_Jet60_v4, HLT_Jet80_v4, HLT_Jet110_v4, HLT_Jet150_v4, HLT_Jet190_v4, HLT_Jet240_v4, HLT_Jet300_v3, HLT_Jet370_v4

/cdaq/physics/Run2011/5e33/*

HLT_Jet30_v9, HLT_Jet60_v9, HLT_Jet110_v9, HLT_Jet190_v9, HLT_Jet240_v9, HLT_Jet300_v9, HLT_Jet370_v10, HLT_Jet800_v5

/cdaq/physics/Run2011/1e33/v2.5/HLT/*

HLT_Jet30_v5, HLT_Jet60_v5, HLT_Jet80_v5, HLT_Jet110_v5, HLT_Jet150_v5, HLT_Jet190_v5, HLT_Jet240_v5, HLT_Jet300_v4, HLT_Jet370_v5

/cdaq/physics/Run2011/5e32/v5.2/HLT/V7, /cdaq/physics/Run2011/5e32/v5.3/HLT/V1, /cdaq/physics/Run2011/5e32/v5.3/HLT/V2, /cdaq/physics/Run2011/5e32/v6.2/HLT/V3, /cdaq/physics/Run2011/5e32/v6.2/HLT/V2, /cdaq/physics/Run2011/5e32/v4.2/HLT/V7, /cdaq/physics/Run2011/5e32/v4.2/HLT/V6, /cdaq/physics/Run2011/5e32/v5.1/HLT/V3, /cdaq/physics/Run2011/5e32/v6.1/HLT/V3, /cdaq/physics/Run2011/5e32/v6.2/HLT/V4, /cdaq/physics/Run2011/5e32/v6.1/HLT/V6, /cdaq/physics/Run2011/5e32/v6.1/HLT/V5

HLT_Jet30_v1, HLT_Jet60_v1, HLT_Jet80_v1, HLT_Jet110_v1, HLT_Jet150_v1, HLT_Jet190_v1, HLT_Jet240_v1, HLT_Jet370_v1

/cdaq/physics/Run2011/5e32/v8.1/HLT/V8, /cdaq/physics/Run2011/5e32/v8.1/HLT/V6, /cdaq/physics/Run2011/5e32/v8.3/HLT/V4, /cdaq/physics/Run2011/5e32/v8.3/HLT/V2, /cdaq/physics/Run2011/5e32/v8.1/HLT/V5, /cdaq/physics/Run2011/5e32/v8.2/HLT/V3

HLT_Jet30_v2, HLT_Jet60_v2, HLT_Jet80_v2, HLT_Jet110_v2, HLT_Jet150_v2, HLT_Jet190_v2, HLT_Jet240_v2, HLT_Jet300_v1, HLT_Jet370_v2

/cdaq/physics/Run2011/1e33/v1.3/HLT/*
HLT_Jet30_v3, HLT_Jet60_v3, HLT_Jet80_v3, HLT_Jet110_v3, HLT_Jet150_v3, HLT_Jet190_v3, HLT_Jet240_v3, HLT_Jet300_v2, HLT_Jet370_v3
/cdaq/physics/Run2011/1.4e33/v1.2/HLT/V3, /cdaq/physics/Run2011/1.4e33/v1.2/HLT/V1, /cdaq/physics/Run2011/2e33/v1.2/HLT/V7, /cdaq/physics/Run2011/2e33/v1.2/HLT/V4, /cdaq/physics/Run2011/2e33/v1.2/HLT/V5, /cdaq/physics/Run2011/2e33/v1.2/HLT/V1
HLT_Jet30_v6, HLT_Jet60_v6, HLT_Jet80_v6, HLT_Jet110_v6, HLT_Jet150_v6, HLT_Jet190_v6, HLT_Jet240_v6, HLT_Jet300_v5, HLT_Jet370_v6, HLT_Jet800_v1
/cdaq/physics/Run2011/1.4e33/v1.1/HLT/*
HLT_Jet30_v4, HLT_Jet60_v4, HLT_Jet80_v4, HLT_Jet110_v4, HLT_Jet150_v4, HLT_Jet190_v4, HLT_Jet240_v4, HLT_Jet300_v3, HLT_Jet370_v4, HLT_Jet800_v1

B.4 Monte Carlo configuration

B.4.1 Sherpa

There are two relevant settings in Sherpa for the determination of the non-perturbative corrections. The first is the `FRAGMENTATION` setting which is either `Off` or `Ahadic`, the name of the Sherpa module responsible for the hadronization effects. The second is parameter is `MI_HANDLER`, which is either `None` or `Amisic`, where `Amisic` is the name of the multiple parton interaction implementation in Sherpa. To generate the non-perturbative correction samples, the events are simulated once with both parameters switched on and off. The phase space The general configuration of the generator used the following settings:

```
(run){
 EVENTS = 100000
 EVENT_MODE = HepMC
! HEPMC2_GENEVENT_OUTPUT = hepmc2_genevent_out
 WRITE_MAPPING_FILE 3;
}(run)

(beam){
 BEAM_1 = 2212; BEAM_ENERGY_1 = 3500;
 BEAM_2 = 2212; BEAM_ENERGY_2 = 3500;
}(beam)

(processes){
 Process 93 93 -> 93 93 93{1}
 Max_Order_EW 0;
! Order_EW 0;
 CKKW sqr(30/E_CMS)
 Integration_Error 0.01 {5};
 End process;
}(processes)

(selector){
 NJetFinder 2 640. 0 0.4
}(selector)

(isr){
 PDF_LIBRARY     = LHAPDFSherpa
```

```
 PDF_SET         = cteq6ll.LHpdf
 PDF_SET_VERSION = 1
 PDF_GRID_PATH   = PDFsets
}(isr)

(me){
 ME_SIGNAL_GENERATOR = Internal Comix
}(me)
```

B.4.2 Pythia

The Pythia samples are taken from the official production of the CMS collaboration. The parameters of the underlying event are set to the Z2 tune. The Z2 tune is the official Pythia tune of CMS and is described by the parameters below:

```
pythiaUESettings = cms.vstring(
  'MSTJ(11)=3      ! Choice of the fragmentation function',
  'MSTJ(22)=2      ! Decay those unstable particles',
  'PARJ(71)=10.    ! for which ctau  10 mm',
  'MSTP(2)=1       ! which order running alphaS',
  'MSTP(33)=0      ! no K factors in hard cross sections',
  'MSTP(51)=7      ! structure function chosen',
  'MSTP(81)=1      ! multiple parton interactions 1 is Pythia default',
  'MSTP(82)=4      ! Defines the multi-parton model',
  'MSTU(21)=1      ! Check on possible errors during program execution',
  'PARP(82)=1.9409 ! pt cutoff for multiparton interactions',
  'PARP(89)=1960.  ! sqrts for which PARP82 is set',
  'PARP(83)=0.5    ! Multiple interactions: matter distr. parameter',
  'PARP(84)=0.4    ! Multiple interactions: matter distr. parameter',
  'PARP(90)=0.16   ! Multiple interactions: rescaling power',
  'PARP(67)=2.5    ! amount of initial-state radiation',
  'PARP(85)=1.0    ! gluon prod. mechanism in MI',
  'PARP(86)=1.0    ! gluon prod. mechanism in MI',
  'PARP(62)=1.25   ! ',
  'PARP(64)=0.2    ! ',
  'MSTP(91)=1      ! ',
  'PARP(91)=2.1    ! kt distribution',
  'PARP(93)=15.0   ! '
)
```

```
processParameters = cms.vstring(
  'MSEL=1          ! QCD high pT processes',
  'CKIN(3)=1000.   ! minimum pt hat for hard interactions',
  'CKIN(4)=1400.   ! maximum pt hat for hard interactions'
)

parameterSets = cms.vstring(
  "pythiaUESettings",
  "processParameters"
)
```

B.4.3 Herwig++

The configuration used for the generation of the Herwig++ Monte-Carlo sample is shown below:

```
basicSetup = cms.vstring(
  'cd /Herwig/Generators',
  'set LHCGenerator:NumberOfEvents 10000000',
  'set LHCGenerator:DebugLevel 1',
  'set LHCGenerator:PrintEvent 0',
  'set LHCGenerator:MaxErrors 10000',
  'cd /',
),

cm7TeV = cms.vstring(
  'cd /Herwig/Generators',
  'set LHCGenerator:EventHandler:LuminosityFunction:Energy 7000.0',
  'set /Herwig/Shower/Evolver:IntrinsicPtGaussian 2.2*GeV'
),

QCDParameters = cms.vstring(
  'cd /Herwig/MatrixElements/',
  'insert SimpleQCD:MatrixElements[0] MEQCD2to2',
  'cd /',
  'set /Herwig/Cuts/JetKtCut:MinKT 15*GeV',
  'set /Herwig/Cuts/JetKtCut:MaxKT 3000*GeV',
  'set /Herwig/UnderlyingEvent/MPIHandler:Algorithm 1',
),
parameterSets = cms.vstring(
  '+basicSetup',
```

```
  '+cm7TeV',
  '+QCDParameters',
  '+setParticlesStableForDetector',
)
```

Appendix C

Results for different jet algorithms

C.1 Three-jet mass resolution

This section contains the tables with the fitting parameters and covariance matrices for the three-jet mass resolution function 6.1.

	Absolute cut	Relative cut
N [GeV]	-21.481 ± 0.55	-18.252 ± 1.1
S [$\sqrt{\text{GeV}}$]	1.1457 ± 0.031	1.8398 ± 0.36
C [1]	0.014982 ± 0.00083	0.013592 ± 0.0024
s [1]	0.052171 ± 0.0042	-0.084198 ± 0.06

Table C.1: Table of parameters for the three-jet mass resolution function 6.1 (Anti-k$_T$ 0.5, $|y_{\max}| \leq 1$).

	N	S	C	s
	1.24752	-0.401196	-0.00246858	0.0652372
	-0.401196	0.132398	0.000848445	-0.0216427
	-0.00246858	0.000848445	$5.95342e-06$	-0.00014007
	0.0652372	-0.0216427	-0.00014007	0.00354213

Table C.2: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.5, $|y_{\max}| \leq 1$, Relative cut).

	N	S	C	s
	0.298062	-0.0157618	0.000415947	0.00180102
	-0.0157618	0.00097333	$-2.54503e-05$	-0.000124774
	0.000415947	$-2.54503e-05$	$6.82167e-07$	$3.14602e-06$
	0.00180102	-0.000124774	$3.14602e-06$	$1.77172e-05$

Table C.3: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.5, $|y_{\max}| \leq 1$, Absolute cut).

	Absolute cut	Relative cut
N [GeV]	-29.396 ± 2.5	-23.517 ± 1.1
S [$\sqrt{\text{GeV}}$]	1.6437 ± 0.66	2.3648 ± 0.36
C [1]	0.007877 ± 0.011	0.0033116 ± 0.012
s [1]	0.010764 ± 0.12	-0.090331 ± 0.046

Table C.4: Table of parameters for the three-jet mass resolution function 6.1 (Anti-k$_T$ 0.5, $1 < |y_{\max}| \leq 2$).

N	S	C	s
1.1634	−0.387275	−0.0117481	0.0489016
−0.387275	0.132296	0.00418067	−0.0167969
−0.0117481	0.00418067	0.00014442	−0.000536134
0.0489016	−0.0167969	−0.000536134	0.00213532

Table C.5: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.5, $1 < |y_{\max}| \leq 2$, Relative cut).

N	S	C	s
6.02934	−1.59195	−0.0254264	0.282989
−1.59195	0.429647	0.007057	−0.0766674
−0.0254264	0.007057	0.000121102	−0.00126614
0.282989	−0.0766674	−0.00126614	0.0136905

Table C.6: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.5, $1 < |y_{\max}| \leq 2$, Absolute cut).

	Absolute cut	Relative cut
N [GeV]	-30.483 ± 0.61	-20.419 ± 1.1
S [$\sqrt{\text{GeV}}$]	-1.3501 ± 0.031	0.97715 ± 0.21
C [1]	0.0026182 ± 0.011	-0.0020661 ± 0.059
s [1]	0.069384 ± 0.0068	0.18785 ± 0.072

Table C.7: Table of parameters for the three-jet mass resolution function 6.1 (Anti-k$_T$ 0.5, $2 < |y_{\max}| \leq 3$).

N	S	C	s
1.21843	−0.230707	0.0573999	0.0766671
−0.230707	0.0461294	−0.0121206	−0.0154745
0.0573999	−0.0121206	0.00342727	0.00411211
0.0766671	−0.0154745	0.00411211	0.00520067

Table C.8: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.5, $2 < |y_{\max}| \leq 3$, Relative cut).

N	S	C	s
0.368765	0.00981033	0.00537498	−0.000672039
0.00981033	0.000964872	$6.00452e-05$	0.000156842
0.00537498	$6.00452e-05$	0.000122844	$-3.86844e-05$
−0.000672039	0.000156842	$-3.86844e-05$	$4.68819e-05$

Table C.9: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.5, $2 < |y_{\max}| \leq 3$, Absolute cut).

	Absolute cut	Relative cut
N [GeV]	-24.581 ± 0.67	-44.721 ± 3.4
S [$\sqrt{\text{GeV}}$]	3.568 ± 0.089	22.135 ± 3.5
C [1]	-0.022794 ± 0.00048	-0.019745 ± 0.00056
s [1]	-0.28388 ± 0.0069	-0.72045 ± 0.034

Table C.10: Table of parameters for the three-jet mass resolution function 6.1 (Anti-k$_T$ 0.7, $|y_{\max}| \leq 1$).

N	S	C	s
11.4599	−11.9498	0.00130314	0.115036
−11.9498	12.5253	−0.00144303	−0.12098
0.00130314	−0.00144303	$3.14947e-07$	$1.44524e-05$
0.115036	−0.12098	$1.44524e-05$	0.00117107

Table C.11: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.7, $|y_{\max}| \leq 1$, Relative cut).

N	S	C	s
0.442225	−0.0317081	−0.000237715	−0.000354317
−0.0317081	0.00794517	$5.24218e-06$	−0.000487362
−0.000237715	$5.24218e-06$	$2.28935e-07$	$1.52921e-06$
−0.000354317	−0.000487362	$1.52921e-06$	$4.79459e-05$

Table C.12: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.7, $|y_{\max}| \leq 1$, Absolute cut).

	Absolute cut	Relative cut
N [GeV]	-20.789 ± 0.78	-32.111 ± 1.6
S [$\sqrt{\text{GeV}}$]	2.7029 ± 0.074	5.0552 ± 0.78
C [1]	0.027475 ± 0.0005	0.0015074 ± 0.021
s [1]	-0.20417 ± 0.0078	-0.27387 ± 0.044

Table C.13: Table of parameters for the three-jet mass resolution function 6.1 (Anti-k$_T$ 0.7, $1 < |y_{\max}| \leq 2$).

N	S	C	s
2.57861	-1.24651	-0.0305616	0.0691266
-1.24651	0.611605	0.0155566	-0.0340538
-0.0305616	0.0155566	0.000443169	-0.00087561
0.0691266	-0.0340538	-0.00087561	0.00189829

Table C.14: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.7, $1 < |y_{\max}| \leq 2$, Relative cut).

N	S	C	s
0.604152	-0.0301375	0.000298259	-0.000567484
-0.0301375	0.00548686	$-4.45512e-06$	-0.000453962
0.000298259	$-4.45512e-06$	$2.50008e-07$	$-1.85652e-06$
-0.000567484	-0.000453962	$-1.85652e-06$	$6.06485e-05$

Table C.15: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.7, $1 < |y_{\max}| \leq 2$, Absolute cut).

	Absolute cut	Relative cut
N [GeV]	-36.225 ± 4.4	-25.607 ± 1.6
S [$\sqrt{\text{GeV}}$]	2.1869 ± 1.3	1.8601 ± 0.46
C [1]	0.0039638 ± 0.037	0.018807 ± 0.0064
s [1]	-0.039513 ± 0.18	0.02616 ± 0.078

Table C.16: Table of parameters for the three-jet mass resolution function 6.1 (Anti-k$_T$ 0.7, $2 < |y_{\max}| \leq 3$).

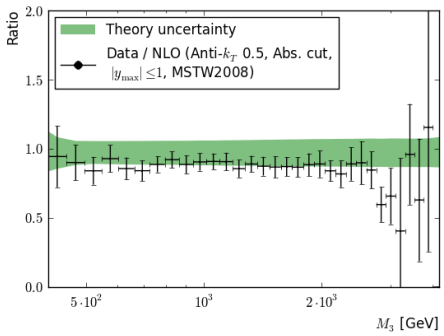
N	S	C	s
2.68548	−0.746476	−0.00960677	0.125045
−0.746476	0.214012	0.00286664	−0.0360625
−0.00960677	0.00286664	$4.12616e-05$	−0.000487412
0.125045	−0.0360625	−0.000487412	0.0060844

Table C.17: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.7, $2 < |y_{\max}| \leq 3$, Relative cut).

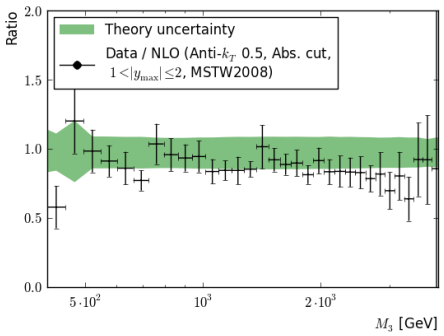
N	S	C	s
19.2301	−5.85842	−0.149916	0.767695
−5.85842	1.82127	0.0480902	−0.239572
−0.149916	0.0480902	0.00134168	−0.00636606
0.767695	−0.239572	−0.00636606	0.0315374

Table C.18: Table of covariances for the three-jet mass resolution parameters (Anti-k$_T$ 0.7, $2 < |y_{\max}| \leq 3$, Absolute cut).

C.2 Three-jet mass comparison

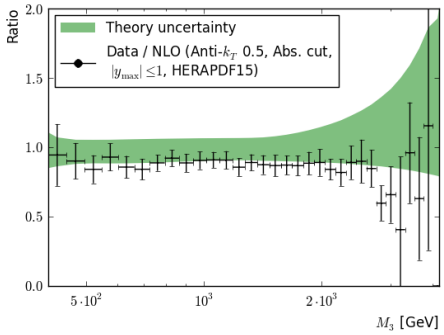


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $1.1715^{+0.0692}_{-0.0765}$.

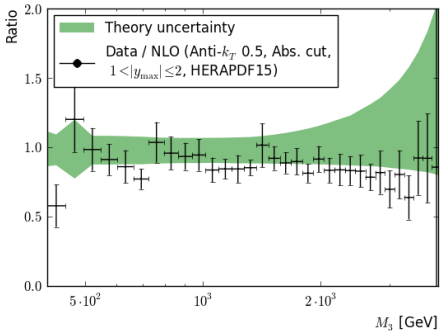


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $1.1519^{+0.1034}_{-0.0839}$.

Figure C.1: Theory data comparison (Anti-k$_T$ 0.5, Absolute cut, MSTW2008)

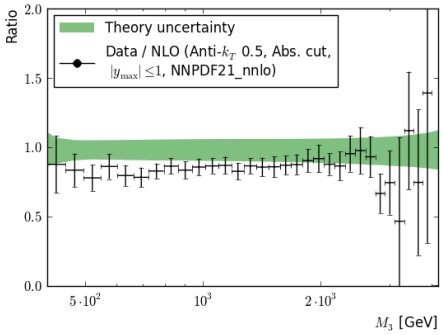


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $1.0541^{+0.0869}_{-0.0746}$.

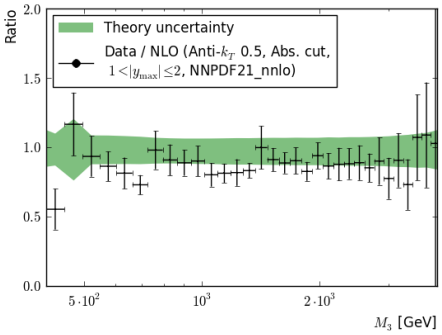


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $1.0606^{+0.0956}_{-0.0815}$.

Figure C.2: Theory data comparison (Anti-k$_T$ 0.5, Absolute cut, HERAPDF15)

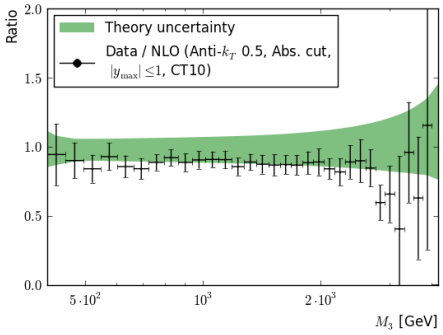


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $1.1551^{+0.0976}_{-0.0776}$.

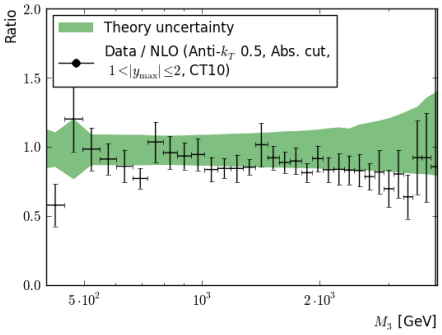


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $1.1434^{+0.1182}_{-0.0839}$.

Figure C.3: Theory data comparison (Anti-k$_T$ 0.5, Absolute cut, NNPDF 2.1 - NNLO)

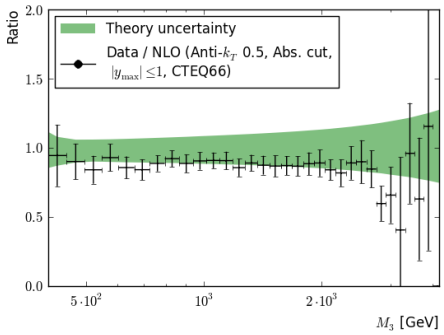


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $1.1510^{+0.1049}_{-0.0826}$.

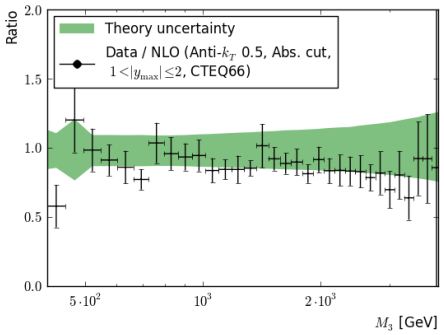


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $1.1478^{+0.1107}_{-0.0888}$.

Figure C.4: Theory data comparison (Anti-k$_T$ 0.5, Absolute cut, CT10)

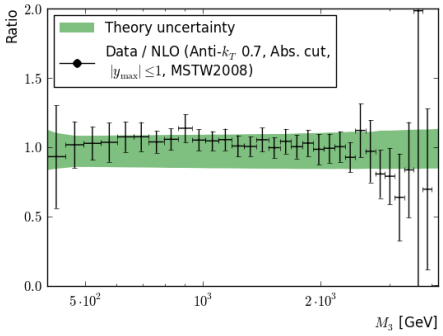


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $1.1513^{+0.1044}_{-0.0839}$.

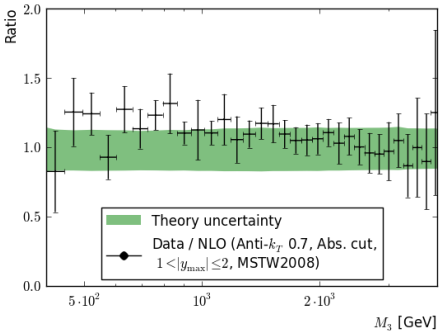


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $1.1537^{+0.1006}_{-0.0901}$.

Figure C.5: Theory data comparison (Anti-k$_T$ 0.5, Absolute cut, CTEQ66)

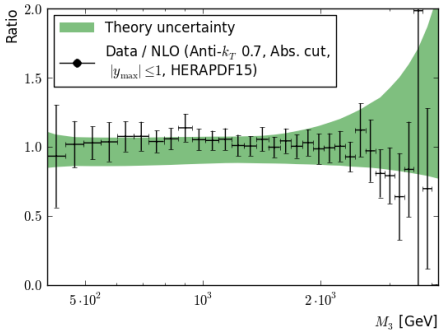


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $1.0043^{+0.0765}_{-0.0656}$.

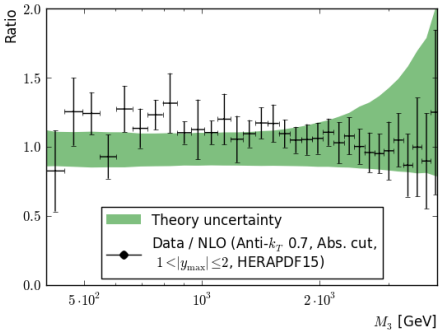


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $0.9194^{+0.0711}_{-0.0617}$.

Figure C.6: Theory data comparison (Anti-k$_T$ 0.7, Absolute cut, MSTW2008)

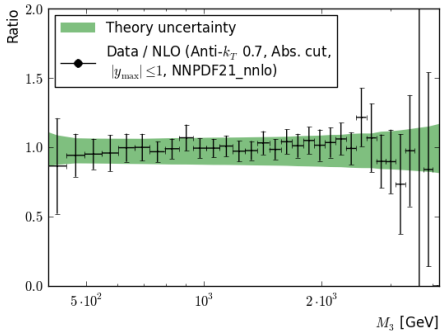


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $0.8998^{+0.0679}_{-0.0595}$.

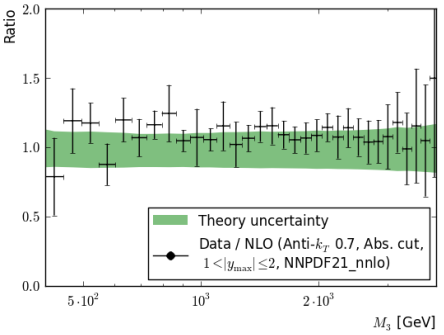


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $0.8457^{+0.0682}_{-0.0602}$.

Figure C.7: Theory data comparison (Anti-k$_T$ 0.7, Absolute cut, HERA-PDF15)

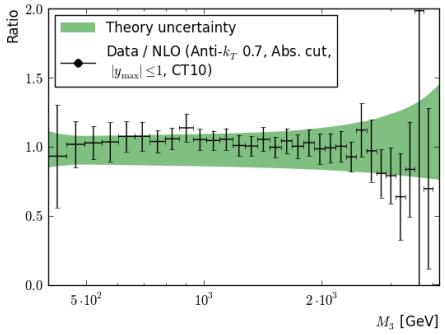


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $0.9877^{+0.0747}_{-0.0641}$.

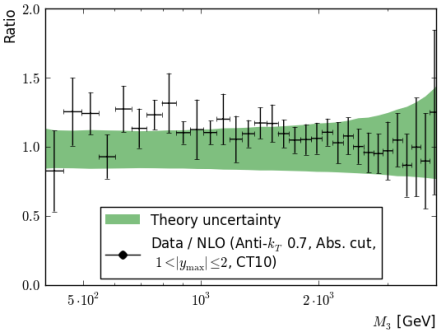


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $0.9128^{+0.0703}_{-0.0612}$.

Figure C.8: Theory data comparison (Anti-k$_T$ 0.7, Absolute cut, NNPDF 2.1 - NNLO)

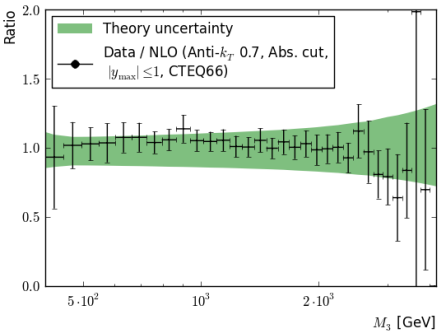


(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $0.9821^{+0.0791}_{-0.0674}$.

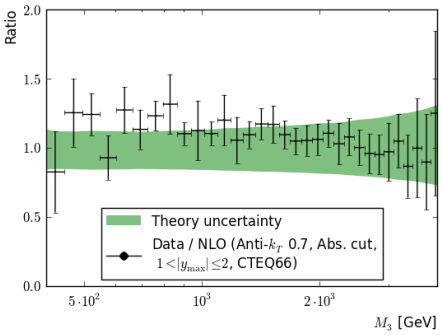


(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $0.9146^{+0.0751}_{-0.0648}$.

Figure C.9: Theory data comparison (Anti-k$_T$ 0.7, Absolute cut, CT10)



(a) Rapidity bin $|y_{\max}| \leq 1$ normalisation factor $0.9823^{+0.0806}_{-0.0685}$.



(b) Rapidity bin $1 < |y_{\max}| \leq 2$ normalisation factor $0.9199^{+0.0769}_{-0.0661}$.

Figure C.10: Theory data comparison (Anti-k$_T$ 0.7, Absolute cut, CTEQ66)

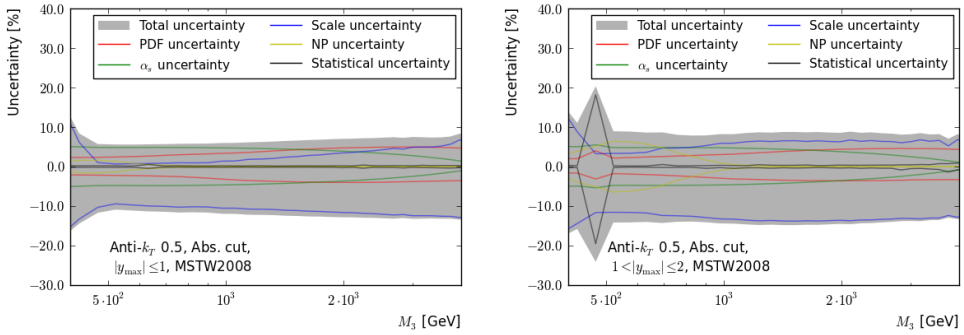
Appendix D

Uncertainties

This appendix gives an overview of the theory and measurement uncertainties for the different jet sizes, rapidity regions and cut scenarios. The

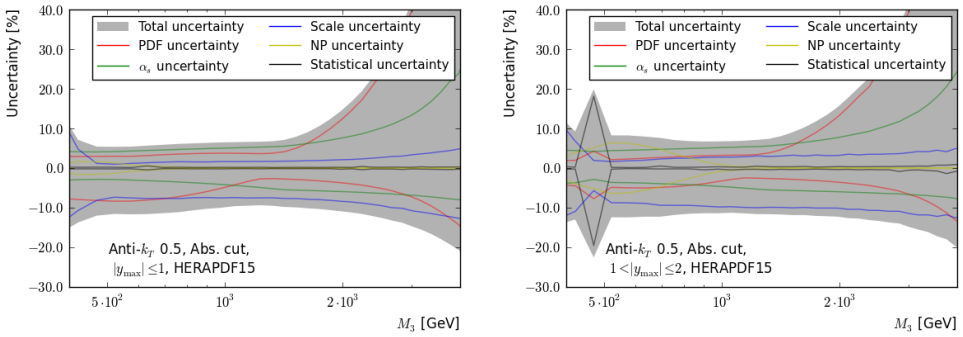
D.1 Theory Uncertainties

In the following, the theory uncertainties are presented. While the statistical and non-perturbative uncertainties are the same for the different cut scenarios, rapidity bins and jet sizes, the PDF, α_s and scale uncertainties are determined for each PDF with the appropriate recipe.



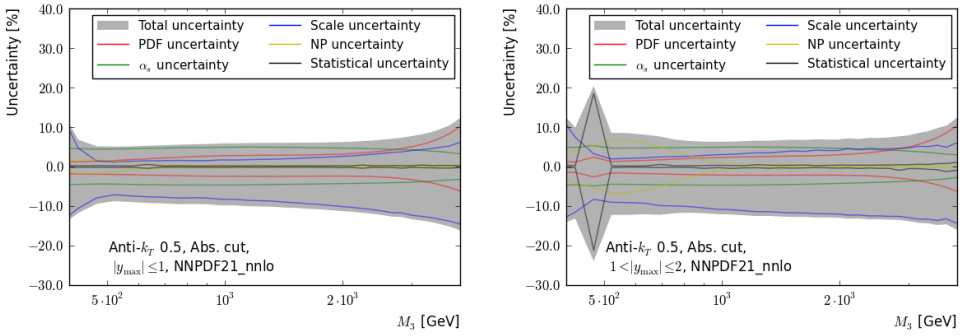
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.1: Theory uncertainties (Anti-k$_T$ 0.5, Absolute cut, MSTW2008)



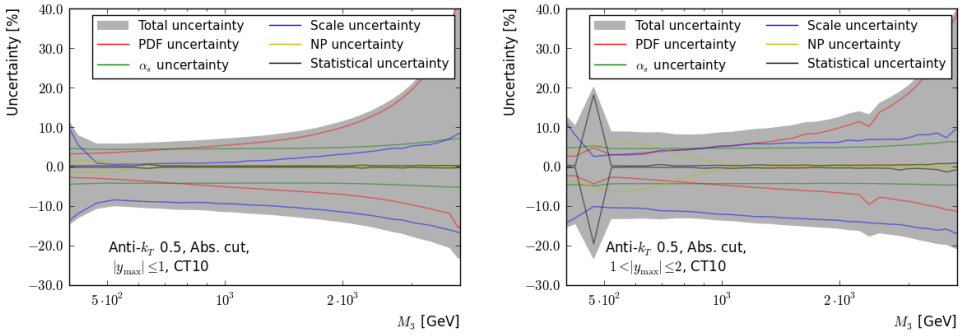
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.2: Theory uncertainties (Anti-k$_T$ 0.5, Absolute cut, HERAPDF15)



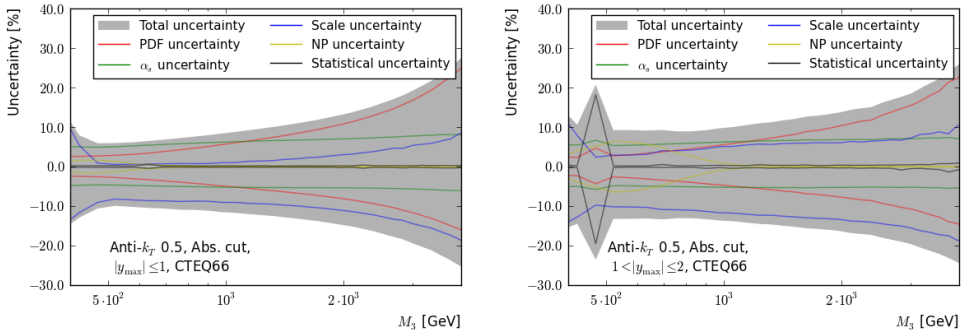
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.3: Theory uncertainties (Anti-k$_T$ 0.5, Absolute cut, NNPDF 2.1 - NNLO)



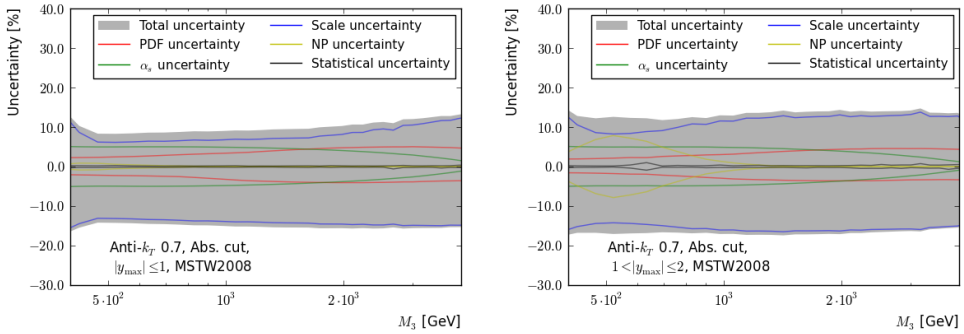
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.4: Theory uncertainties (Anti-k$_T$ 0.5, Absolute cut, CT10)



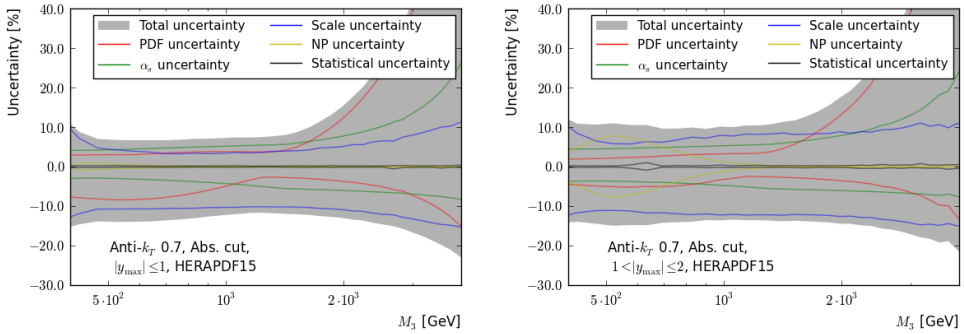
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.5: Theory uncertainties (Anti-k$_T$ 0.5, Absolute cut, CTEQ66)



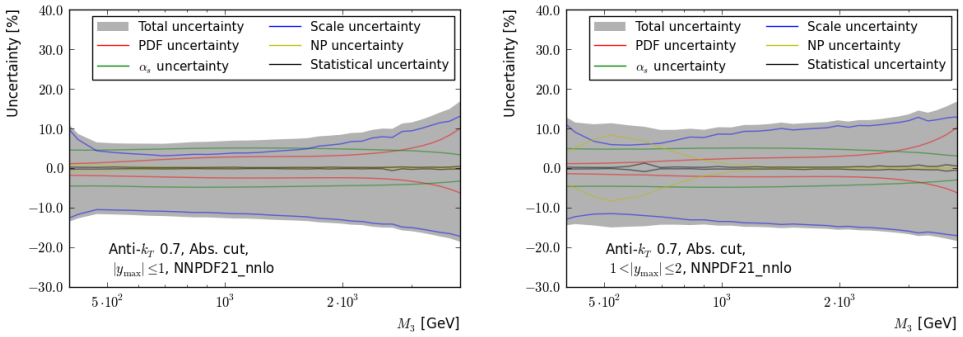
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.6: Theory uncertainties (Anti-k$_T$ 0.7, Absolute cut, MSTW2008)



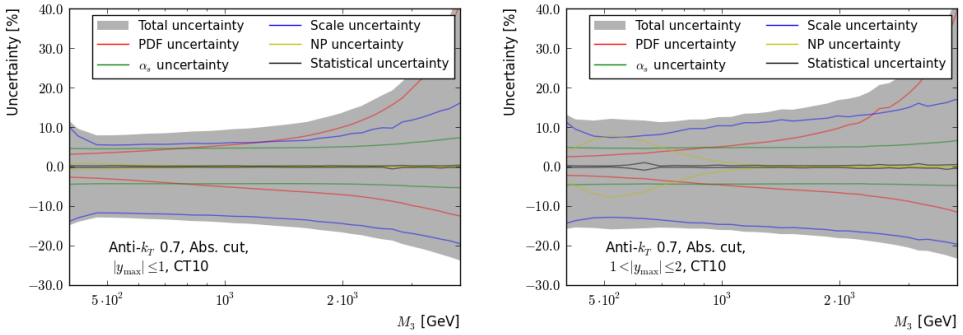
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.7: Theory uncertainties (Anti-k$_T$ 0.7, Absolute cut, HERAPDF15)



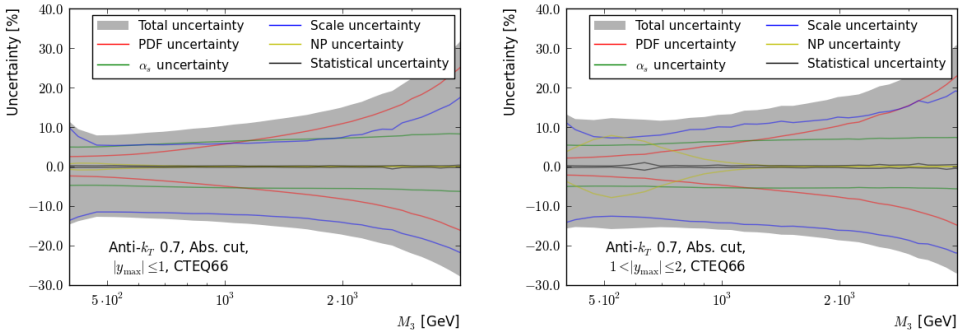
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.8: Theory uncertainties (Anti-k$_T$ 0.7, Absolute cut, NNPDF 2.1 - NNLO)



(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.9: Theory uncertainties (Anti-k$_T$ 0.7, Absolute cut, CT10)



(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.10: Theory uncertainties (Anti-k$_T$ 0.7, Absolute cut, CTEQ66)

D.2 Measurement and covariance matrices

Numerical results of the three-jet mass measurements will be made available on the Durham HepData[122] web-page in tabular form together with covariance matrices. The covariance matrices will contain information about the measurement uncertainties and the correlation between three-jet mass bins. This will be done as part of the publication process of the CMS physics analysis summary [123], which is derived from the work presented in this thesis.

D.3 Measurement uncertainties

This section gives an overview of the measurement uncertainties. The major uncertainties are the jet energy scale uncertainty, the unfolding uncertainty (which includes both the statistical uncertainty and the algorithmic unfolding uncertainty) and the luminosity uncertainty. The uncertainties are shown in a figure as a function of the three-jet mass range and as a table, which contains the average, minimum and maximum value for each uncertainty source in the three-jet mass range between 500 and 3000 GeV.

$500 - 3000\,\text{GeV}$	Min	Average	Max
Total	6.4%	12.0%	36.0%
Luminosity	4.0%	4.0%	4.0%
Unfolding and statistical	2.7%	8.9%	33.7%
Statistical	0.95%	4.0%	22.8%
Jet energy scale	2.8%	6.0%	12.0%

Table D.1: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 3000 GeV - (Anti-k$_T$ 0.5, $|y_{\text{max}}| \leq 1$, Relative cut).

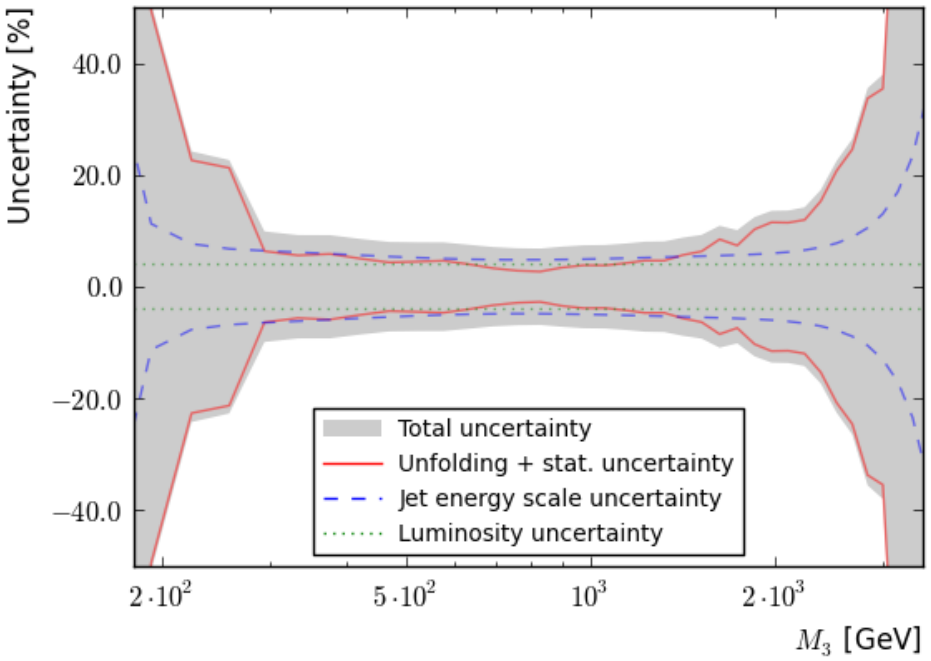


Figure D.11: Overview of the measurement uncertainties (Anti-k$_T$ 0.5, $|y_{\text{max}}| \leq 1$, Relative cut)

$500 - 3000\,\text{GeV}$	Min	Average	Max
Total	6.3%	9.9%	21.2%
Luminosity	4.0%	4.0%	4.0%
Unfolding and statistical	2.8%	6.3%	19.1%
Statistical	0.7%	3.6%	15.6%
Jet energy scale	3.9%	5.9%	12.2%

Table D.2: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 3000 GeV - (Anti-k$_T$ 0.5, $|y_{\text{max}}| \leq 1$, Absolute cut).

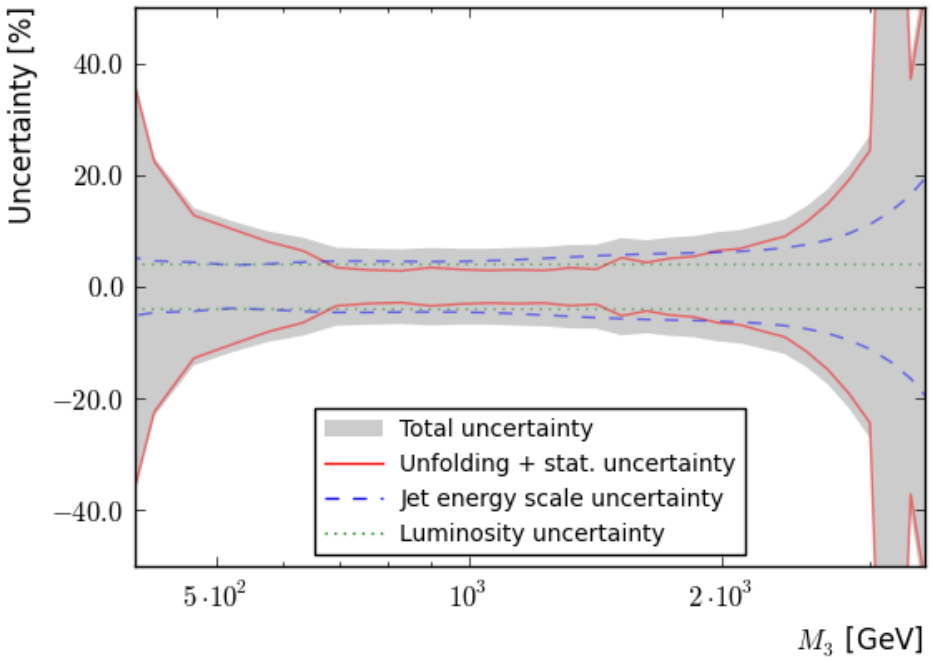


Figure D.12: Overview of the measurement uncertainties (Anti-k$_T$ 0.5, $|y_{\text{max}}| \leq 1$, Absolute cut)

$500 - 3000\,\text{GeV}$	Min	Average	Max
Total	5.8%	11.0%	20.9%
Luminosity	4.0%	4.0%	4.0%
Unfolding and statistical	3.5%	6.6%	15.0%
Statistical	1.7%	3.4%	5.6%
Jet energy scale	2.4%	7.6%	14.0%

Table D.3: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 3000 GeV - (Anti-k$_T$ 0.5, $1 < |y_{\max}| \leq 2$, Relative cut).

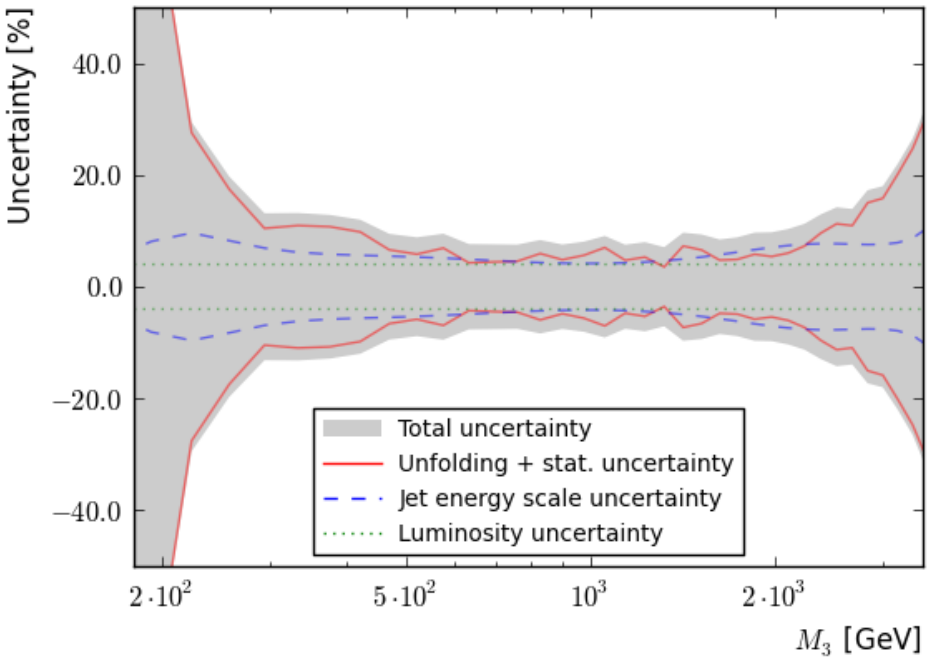


Figure D.13: Overview of the measurement uncertainties (Anti-k$_T$ 0.5, $1 < |y_{\max}| \leq 2$, Relative cut)

$500 - 3000\,\text{GeV}$	Min	Average	Max
Total	6.5%	11.8%	19.0%
Luminosity	4.0%	4.0%	4.0%
Unfolding and statistical	4.0%	7.7%	14.1%
Statistical	1.6%	4.5%	10.4%
Jet energy scale	1.8%	7.2%	13.0%

Table D.4: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 3000 GeV - (Anti-k$_T$ 0.5, $1 < |y_{\text{max}}| \leq 2$, Absolute cut).

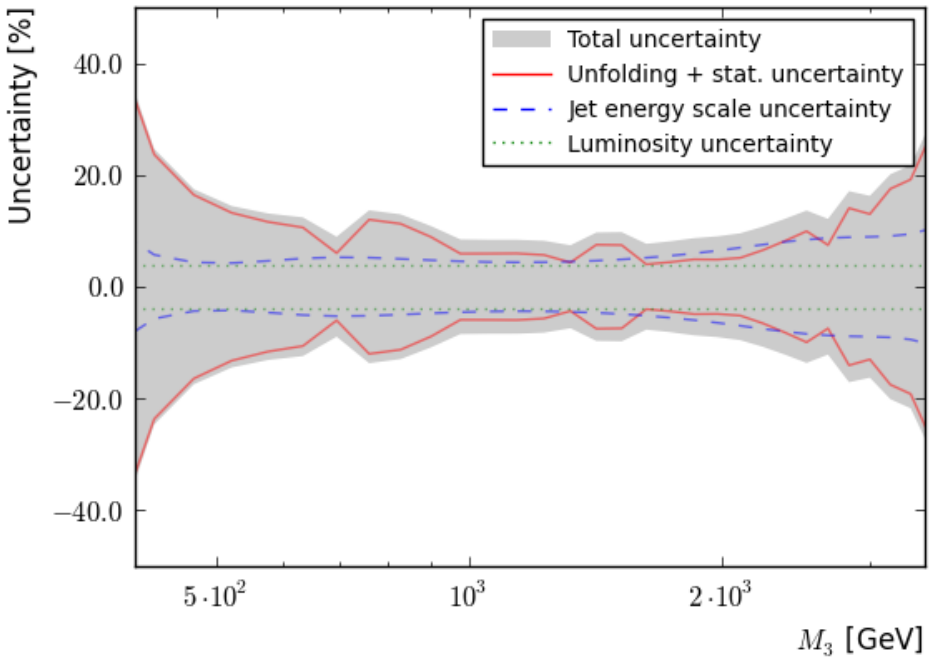


Figure D.14: Overview of the measurement uncertainties (Anti-k$_T$ 0.5, $1 < |y_{\text{max}}| \leq 2$, Absolute cut)

$500 - 3000\,\text{GeV}$	Min	Average	Max
Total	7.1%	12.9%	35.7%
Luminosity	4.0%	4.0%	4.0%
Unfolding and statistical	2.9%	9.6%	30.0%
Statistical	0.95%	3.9%	17.2%
Jet energy scale	3.0%	6.6%	25.5%

Table D.5: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 3000 GeV - (Anti-k$_T$ 0.7, $|y_{\text{max}}| \leq 1$, Relative cut).

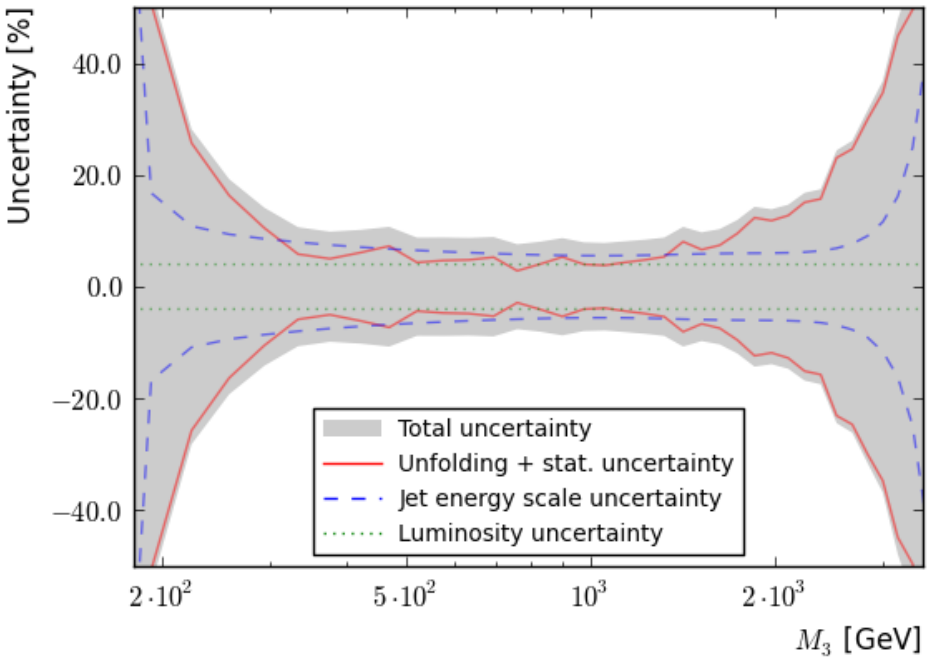


Figure D.15: Overview of the measurement uncertainties (Anti-k$_T$ 0.7, $|y_{\text{max}}| \leq 1$, Relative cut)

$500 - 3000\,\text{GeV}$	Min	Average	Max
Total	6.7%	10.6%	23.2%
Luminosity	4.0%	4.0%	4.0%
Unfolding and statistical	2.6%	6.5%	19.1%
Statistical	0.7%	3.7%	13.6%
Jet energy scale	4.3%	6.8%	17.5%

Table D.6: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 3000 GeV - (Anti-k$_T$ 0.7, $|y_{\text{max}}| \leq 1$, Absolute cut).

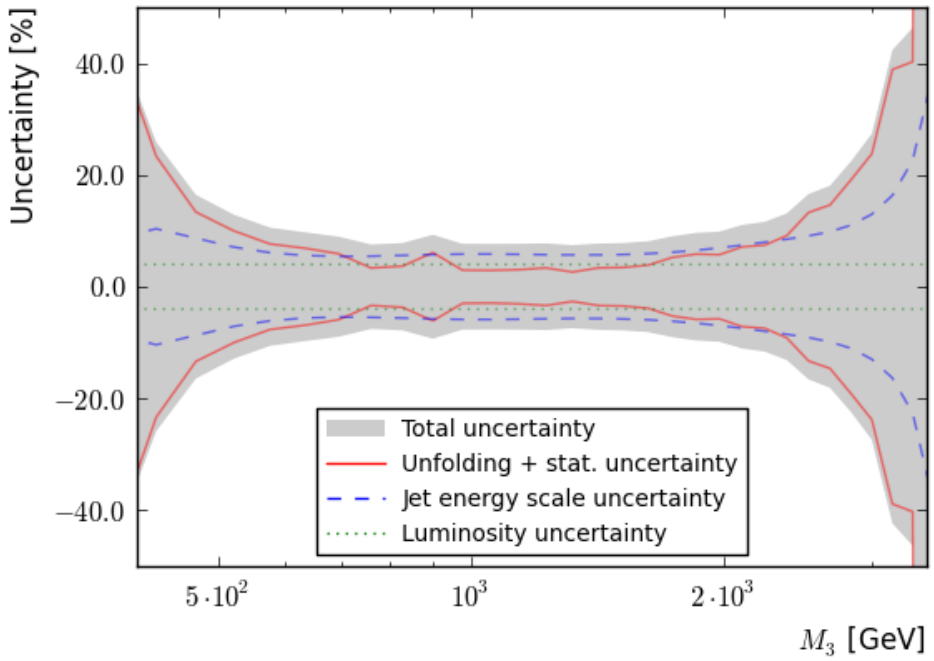


Figure D.16: Overview of the measurement uncertainties (Anti-k$_T$ 0.7, $|y_{\text{max}}| \leq 1$, Absolute cut)

$500 - 3000\,\text{GeV}$	Min	Average	Max
Total	7.3%	12.2%	22.1%
Luminosity	4.0%	4.0%	4.0%
Unfolding and statistical	4.3%	7.2%	14.4%
Statistical	1.6%	3.4%	5.8%
Jet energy scale	3.1%	8.6%	17.8%

Table D.7: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 3000 GeV - (Anti-k$_T$ 0.7, $1 < |y_{\text{max}}| \leq 2$, Relative cut).

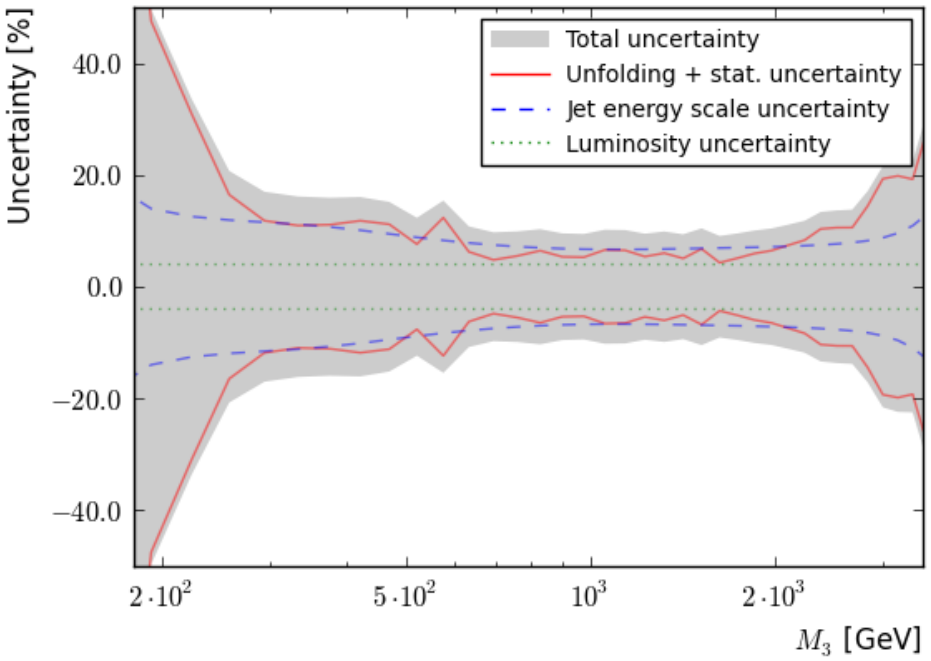


Figure D.17: Overview of the measurement uncertainties (Anti-k$_T$ 0.7, $1 < |y_{\text{max}}| \leq 2$, Relative cut)

$500 - 3000\,\text{GeV}$	Min	Average	Max
Total	7.4%	12.2%	19.3%
Luminosity	4.0%	4.0%	4.0%
Unfolding and statistical	3.5%	7.3%	13.2%
Statistical	1.5%	4.1%	8.5%
Jet energy scale	2.2%	8.4%	17.8%

Table D.8: Tabular overview of the measurement uncertainties for the three-jet mass bins between 500 and 3000 GeV - (Anti-k$_T$ 0.7, $1 < |y_{\text{max}}| \leq 2$, Absolute cut).

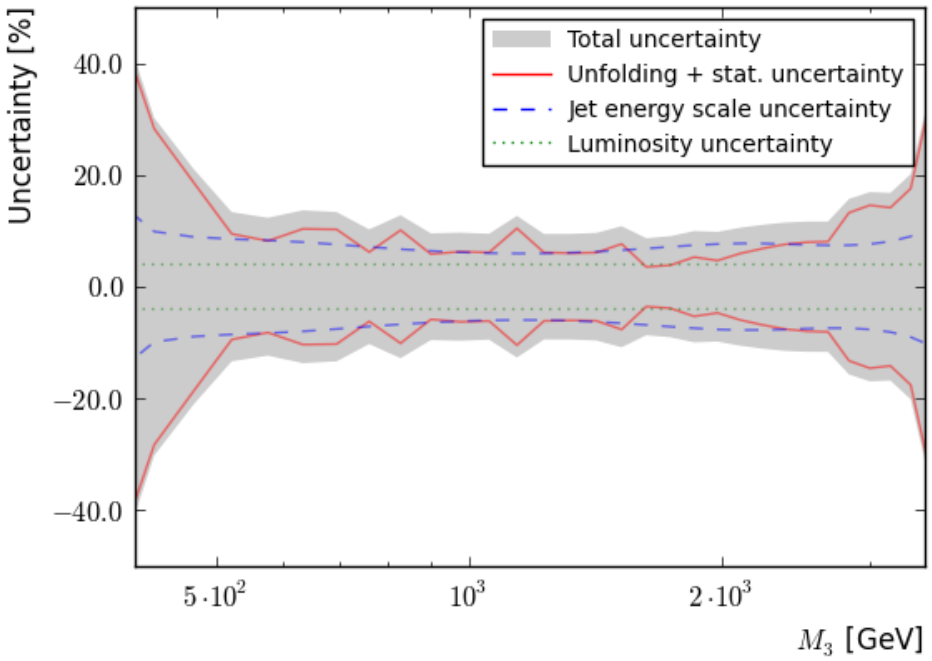
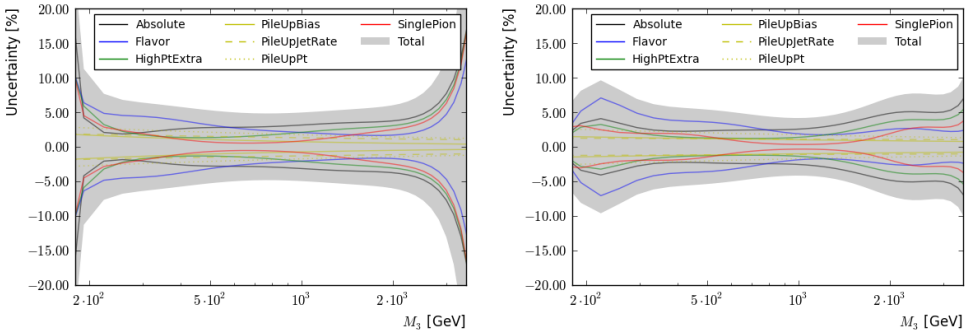


Figure D.18: Overview of the measurement uncertainties (Anti-k$_T$ 0.7, $1 < |y_{\text{max}}| \leq 2$, Absolute cut)

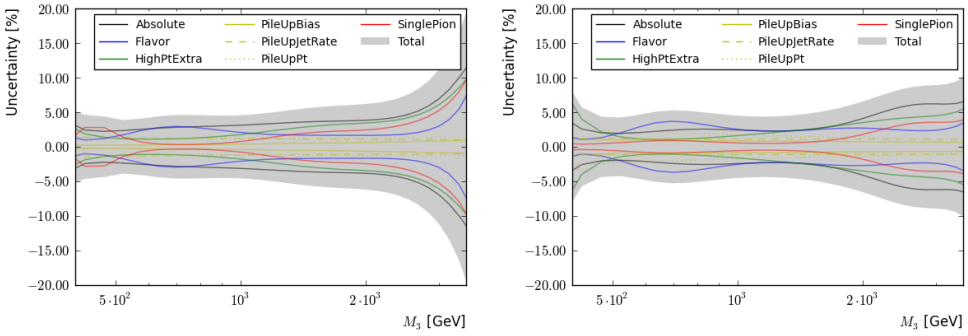
D.3.1 Jet Energy Uncertainties

The shown jet energy uncertainties in this analysis are smoothed with a polynomial interpolation of degree seven. The smoothing polynomial is fitted as a function of the logarithmic three-jet mass to the uncertainties. This is done since the procedure, which was used to derive the uncertainties can be affected by binning effects. Since the three-jet mass bins are both sparsely and unevenly populated, a small jet energy scale variation due to one of the uncorrelated jet energy scale uncertainty sources can sometimes lead to a step-like behaviour at the bin edges or even an unchanged bin content for events populating the bin centre.



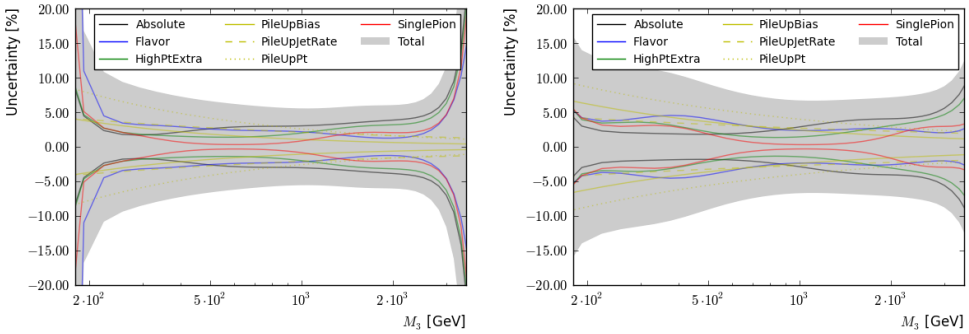
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.19: Overview of the dominant jet energy scale uncertainties (Anti-k_T 0.5, Relative cut)



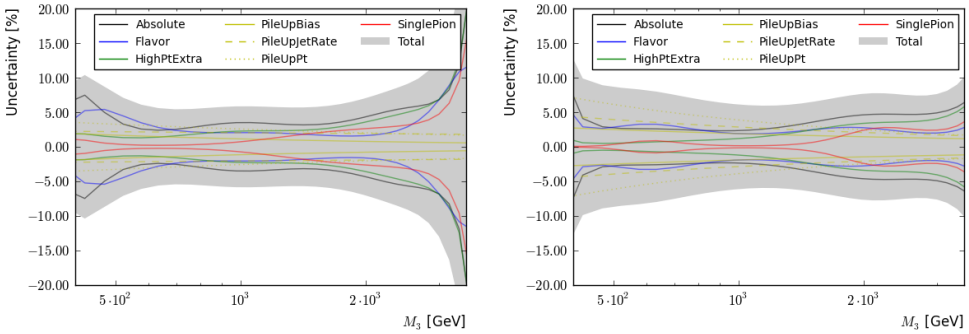
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.20: Overview of the dominant jet energy scale uncertainties (Anti-k_T 0.5, Absolute cut)



(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.21: Overview of the dominant jet energy scale uncertainties (Anti-k_T 0.7, Relative cut)

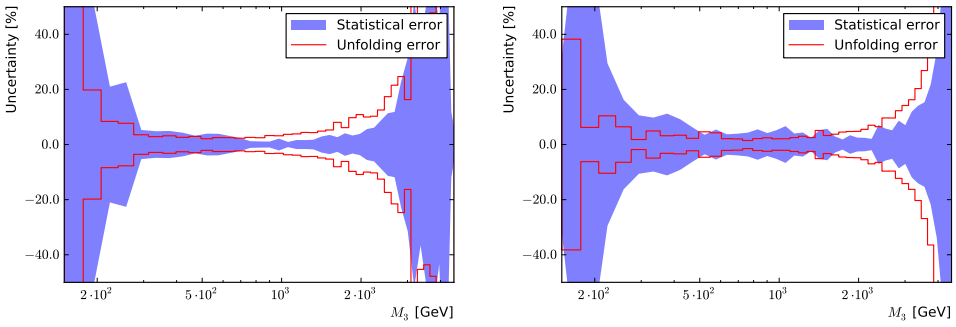


(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.22: Overview of the dominant jet energy scale uncertainties (Anti-k_T 0.7, Absolute cut)

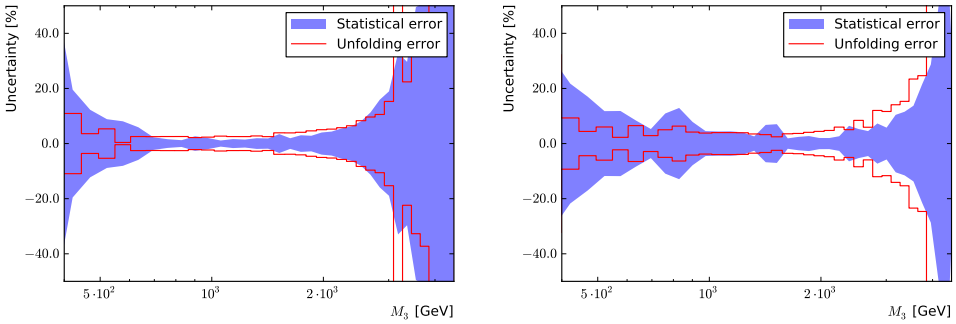
D.3.2 Unfolding Uncertainties

The unfolding uncertainties also include the statistical uncertainties. The results are taken from unfolding toy samples, which independently varied the input histograms and the response matrix of the detector. Using these samples, both the propagated statistical uncertainty and the algorithmic uncertainty for the unfolding method can be determined.



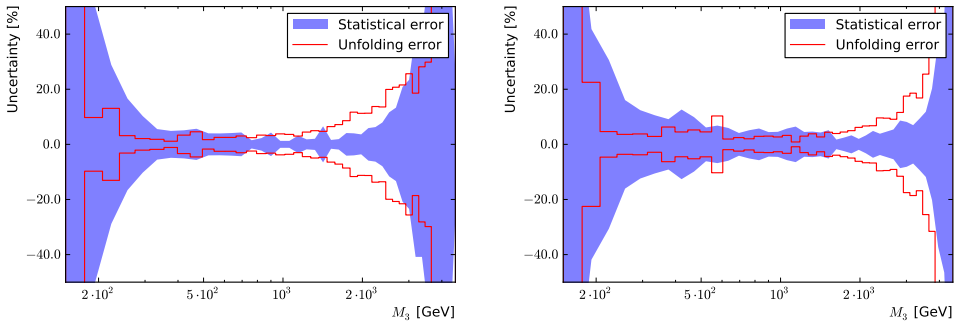
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.23: Statistical and algorithmic unfolding uncertainty (Anti-k$_T$ 0.5, Relative cut)



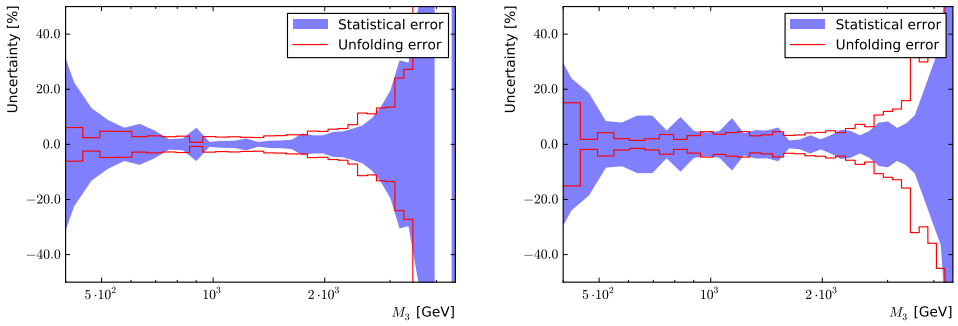
(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.24: Statistical and algorithmic unfolding uncertainty (Anti-k$_T$ 0.5, Absolute cut)



(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.25: Statistical and algorithmic unfolding uncertainty (Anti-k$_T$ 0.7, Relative cut)



(a) $|y_{\max}| \leq 1$ (b) $1 < |y_{\max}| \leq 2$

Figure D.26: Statistical and algorithmic unfolding uncertainty (Anti-k$_T$ 0.7, Absolute cut)

List of Figures

List of Tables

Bibliography

[1] D. H. Perkins, "Introduction to high energy physics". Cambridge University Press, 4th edition edition, 2000.

[2] S. Weinberg, "The Quantum Theory of Fields, Volume 1: Foundations". Cambridge University Press, 2005.

[3] M. Burgess, "Classical covariant fields". Cambridge University Press, 2002.

[4] M. E. Peskin and D. V. Schroeder, "An Introduction to quantum field theory". Addison-Wesley, 1995.

[5] U. Nierste and F. Stober, " Skript zur Vorlesung Theoretische Teilchenphysik I-III". http://fredstober.de/physik/TTP.pdf.

[6] F. Stober, "Study of Three-Jet Topologies and first Event Simulation with the Event Generator Herwig++ for the CMS experiment at the LHC". CERN-THESIS-2008-114, CMS-TS-2010-002, IEKP-KA-2008-28.

[7] C. N. Yang and R. L. Mills, "Conservation of Isotopic Spin and Isotopic Gauge Invariance", *Phys. Rev.* **96** (Oct, 1954) 191–195, doi:10.1103/PhysRev.96.191.

[8] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys.Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[9] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys.Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[10] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries", *Phys. Rev.* **127** (Aug, 1962) 965–970, doi:10.1103/PhysRev.127.965.

[11] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[12] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction", *Progress of Theoretical Physics* **49** (1973), no. 2, 652–657.

[13] B. Webber, "QCD Phenomenology at High Energy". CERN Academic Training Lectures, 2008.

[14] G. Sterman et al., "Handbook of perturbative QCD", *Reviews of Modern Physics* **67** (January, 1995) 157–248, doi:10.1103/revmodphys.67.157.

[15] J. M. Campbell, J. Huston, and W. Stirling, "Hard Interactions of Quarks and Gluons: A Primer for LHC Physics", *Rept.Prog.Phys.* **70** (2007) 89, doi:10.1088/0034-4885/70/1/R02, arXiv:hep-ph/0611148.

[16] Particle Data Group Collaboration, "Review of particle physics", *Phys. Lett.* **B667** (2008) 1, doi:10.1016/j.physletb.2008.07.018.

[17] R. K. Ellis, W. J. Stirling, and B. R. Webber, "QCD and Collider Physics". Cambridge University Press, December, 2003.

[18] S. Bethke, "Determination of the QCD coupling", *Journal of Physics G: Nuclear and Particle Physics* **26** (2000), no. 7, R27–R66.

[19] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language", *Nucl.Phys.* **B126** (1977) 298, doi:10.1016/0550-3213(77)90384-4.

[20] L. Lonnblad, "Basics of Event Generators". Terascale Monte Carlo School, 2008.

[21] M. Bähr et al., "Herwig++ Physics and Manual", *Eur. Phys. J. C* **58** (2008) 639, doi:10.1140/epjc/s10052-008-0798-9.

[22] L. Lonnblad, "ARIADNE version 4: A Program for simulation of QCD cascades implementing the color dipole model", *Comput. Phys. Commun.* **71** (1992) 15–31, doi:10.1016/0010-4655(92)90068-A.

[23] S. Catani et al., "QCD Matrix Elements + Parton Showers", *Journal of High Energy Physics* **2001** (2001), no. 11, 063.

[24] B. Andersson et al., "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.

[25] B. Povh, ed., "Teilchen und Kerne : eine Einfhrung in die physikalischen Konzepte". Springer-Lehrbuch. Springer, Berlin, 8. aufl. edition, 2009.

[26] P. Eden and G. Gustafson, "Baryon production in the string fragmentation picture", *Z. Phys.* **C75** (1997) 41–49, doi:10.1007/s002880050445, arXiv:hep-ph/9606454.

[27] M. Bahr and S. Gieseke and M. A. Gigg and D. Grellscheid and K. Hamilton and O. Latunde-Dada and S. Platzer and P. Richardson and M. H. Seymour and A. Sherstnev and B. R. Webber, "Herwig++ Physics and Manual", 2008.

[28] T. Gleisberg et al., "Event generation with SHERPA 1.1", *Journal of High Energy Physics* **2009** (2009), no. 02, 007.

[29] S. Gieseke, "LHC Event Generation with Herwig++", *Acta Physica Polonica B* **38** (July, 2007) 2261–+.

[30] UA1 Collaboration, "Experimental observation of lepton pairs of invariant mass around 95-GeV/c**2 at the CERN SPS collider", *Phys. Lett.* **B126** (1983) 398–410, doi:10.1016/0370-2693(83)90188-0.

[31] UA2 Collaboration, "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN anti-p p collider", *Phys. Lett.* **B122** (1983) 476–485, doi:10.1016/0370-2693(83)91605-2.

[32] CDF Collaboration, "Observation of Top Quark Production in Pbar-P Collisions", *Physical Review Letters* **74** (1995) 2626.

[33] L. Evans and P. Bryant, "LHC Machine", *Journal of Instrumentation* **3** (2008), no. 08, S08001.

[34] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08002.

[35] ALICE Collaboration, "Elliptic Flow of Charged Particles in Pb-Pb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV", *Phys. Rev. Lett.* **105** (Dec, 2010) 252302, doi:10.1103/PhysRevLett.105.252302.

[36] LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08005.

[37] CMS Collaboration, "The CMS experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08004.

[38] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08003.

[39] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08007.

[40] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08006.

[41] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group Collaboration, "Precision electroweak measurements on the Z resonance", *Phys.Rept.* **427** (2006) 257–454, doi:10.1016/j.physrep.2005.12.006, arXiv:hep-ex/0509008.

[42] LHC Incident report https://edms.cern.ch/document/973073/1.

[43] CERN Communication Group, "CERN FAQ - LHC the Guide". CERN–Brochure-2008-001-Eng, January, 2008.

[44] LHC Programme Coordination, http://lpc.web.cern.ch/lpc/.

[45] LHC machine commissioning and the near future, http://lhc-commissioning.web.cern.ch/lhc-commissioning/presentations/2010/lhc-status-split-oct10.pdf.

[46] LHC operation wiki, http://op-webtools.web.cern.ch/op-webtools/dokuwiki.

[47] LHC OP Vistars, `http://op-webtools.web.cern.ch/op-webtools/Vistar/`.

[48] LHC Machine Parameters `http://lhccwg.web.cern.ch/lhccwg/Bibliography/TargetMachineParameters.pdf`.

[49] "LHC Lumi Days: LHC Workshop on LHC Luminosity Calibration". CERN, Geneva, (2011).

[50] S. van der Meer, "Calibration of the effective beam height in the ISR.", Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, (1968).

[51] Public CMS luminosity measurements `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[52] CMS Collaboration, "Transverse momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt(s) = 0.9$ and 2.36 TeV", *JHEP* **1002** (2010) 041, `doi:10.1007/JHEP02(2010)041`, `arXiv:1002.0621`.

[53] P. Adzic, "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter", *Journal of Instrumentation* **2** (2007), no. 04, P04004.

[54] CMS Collaboration, "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data", *Journal of Instrumentation* **5** (2010), no. 03, T03012.

[55] C. M. G. Lattes, Y. Fujimoto, and S. Hasegawa, "Hadronic Interactions of High-Energy Cosmic Ray Observed by Emulsion Chambers", *Phys. Rept.* **65** (1980) 151, `doi:10.1016/0370-1573(80)90165-9`.

[56] V. Kopenkin, Y. Fujimoto, and T. Sinzi, "Solution to the Centauro puzzle", *Phys. Rev. C* **68** (September, 2003) 052007–+.

[57] CMS Collaboration, "The TriDAS Project Technical Design Report, Volume 2: Data Aquisition and High-Level Trigger". CERN, Geneva, 2002.

[58] CMS Collaboration, "CMS TriDAS project: Technical Design Report". CERN-LHCC-2000-038, 1997.

[59] GLITE 3.1 USER GUIDE
https://edms.cern.ch/file/722398//gLite-3-UserGuide.pdf.

[60] CRAB, twiki.cern.ch/twiki/bin/view/CMS/CRAB.

[61] Grid-Control,
ekptrac.physik.uni-karlsruhe.de/trac/grid-control.

[62] 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.

[63] M. Bahr, S. Gieseke, and M. H. Seymour, "Simulation of multiple partonic interactions in Herwig++", *JHEP* **07** (2008) 076,
doi:10.1088/1126-6708/2008/07/076, arXiv:0803.3633.

[64] L. Lonnblad, "Development strategies for PYTHIA version 7", *Comput. Phys. Commun.* **118** (1999) 213–228,
doi:10.1016/S0010-4655(98)00200-8, arXiv:hep-ph/9810208.

[65] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029.
hep-ph/0204244.

[66] S. Catani and M. Seymour, "A General algorithm for calculating jet cross-sections in NLO QCD", *Nucl.Phys.* **B485** (1997) 291–419,
doi:10.1016/S0550-3213(96)00589-5, arXiv:hep-ph/9605323.

[67] R. Brun and F. Rademakers, "ROOT – An object oriented data analysis framework", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389** (Apr, 1997) 81–86.

[68] G. van Rossum, "Python reference manual", Report CS-R9525, CWI, (1995).

[69] E. Jones et al., "SciPy: Open source scientific tools for Python", 2001.

[70] J. D. Hunter, "Matplotlib: A 2D graphics environment", *Computing In Science & Engineering* **9** (May-Jun, 2007) 90–95.

[71] Grid Control - An Alternative Grid Job Submission Tool,
http://indico.desy.de/contributionDisplay.py?contribId=5&confId=2295.

[72] CMS Collaboration, The CMS Collaboration, " CMS Computing Technical Design Report". CERN-LHCC-2005-023, ISBN 92-9083-252-5, 2005.

[73] CMSSW Application Framework `twiki.cern.ch/twiki/bin/view/CMS/WorkBookCMSSWFramework`.

[74] S. Agostinelli et al., "GEANT4—a simulation toolkit", *Nucl. Inst. Meth. A* **506** (2003) 250, `doi:10.1016/S0168-9002(03)01368-8`.

[75] M. Cacciari and G. P. Salam, "Dispelling the N**3 myth for the k(t) jet-finder", *Phys. Lett.* **B641** (2006) 57–61, `doi:10.1016/j.physletb.2006.08.037`, `arXiv:hep-ph/0512210`.

[76] M. Dobbs and J. B. Hansen, "The HepMC C++ Monte Carlo event record for High Energy Physics", *Computer Physics Communications* **134** (Feb, 2001) 41–46.

[77] C. M. Harris, P. Richardson, and B. R. Webber, "CHARYBDIS: A black hole event generator", *JHEP* **08** (2003) 033, `arXiv:hep-ph/0307305`.

[78] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET",. CMS PAS PFT-09-001.

[79] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector",. CMS PAS PFT-10-001.

[80] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J.* **C72** (2012) 1896, `doi:10.1140/epjc/s10052-012-1896-2`, `arXiv:1111.6097`.

[81] M. Cacciari, "FastJet: A Code for fast k(t) clustering, and more",. hep-ph/0607071.

[82] M. Cacciari and G. P. Salam, "Dispelling the N^3 myth for the Kt jet-finder". Phys.Lett. B641 (2006) 57-61 [arXiv:hep-ph/0512210v2], 2006.

[83] J. M. Butterworth et al., "KtJet: A C++ implementation of the k_{T} clustering algorithm", *Comput. Phys. Commun.* **153** (2003) 85–96. hep-ph/0210022.

[84] Gavin Salam: Jets, CTEQ MCNet school http://www.lpthe.jussieu.fr/~salam/repository/talks/2008-cteq-mcnet.pdf.

[85] G. Soyez, "The SISCone and anti-kt jet algorithms", doi:10.3360/dis.2008.178. arXiv:0807.0021.

[86] G. P. Salam and G. Soyez, "A practical Seedless Infrared-Safe Cone jet algorithm", *JHEP* **05** (2007) 086, arXiv:0704.0292.

[87] A. Bhatti et al., "Performance of the SISCone Jet Clustering Algorithm". CMS AN-2008/002, 2008.

[88] CMS Collaboration, "Jet Energy Calibration and Transverse Momentum Resolution in CMS". CMS-PAS-JME-10-011, 2011.

[89] M. Cacciari, G. P. Salam, and G. Soyez, "The Catchment Area of Jets", *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005. arXiv:0802.1188.

[90] G. Salam, "Towards jetography", *The European Physical Journal C - Particles and Fields* **67** (2010) 637–686. 10.1140/epjc/s10052-010-1314-6.

[91] A. Bhatti et al., "Plans for Jet Energy Corrections at CMS". CMS AN-2007/055, 2007.

[92] H.-L. Lai et al., "New parton distributions for collider physics", *Phys. Rev. D* **82** (2010) 074024, doi:10.1103/PhysRevD.82.074024.

[93] Z. Nagy, "Next-to-leading order calculation of three jet observables in hadron hadron collision", *Phys.Rev.* **D68** (2003) 094002, doi:10.1103/PhysRevD.68.094002, arXiv:hep-ph/0307268.

[94] T. Kluge, K. Rabbertz, and M. Wobisch, "FastNLO: Fast pQCD calculations for PDF fits", arXiv:hep-ph/0609285.

[95] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys.Rev.* **D78** (2008) 013004, doi:10.1103/PhysRevD.78.013004, arXiv:0802.0007.

[96] H1 Collaboration, ZEUS Collaboration Collaboration, "Hera Precision Measurements and Impact for LHC Predictions", arXiv:1107.4193.

[97] A. Martin et al., "Parton distributions for the LHC", *Eur.Phys.J.* **C63** (2009) 189–285, doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.

[98] R. D. Ball et al., "Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology", *Nucl.Phys.* **B849** (2011) 296–363, doi:10.1016/j.nuclphysb.2011.03.021, arXiv:1101.1300.

[99] CMS Collaboration, "Measurement of the Inclusive Jet Cross Section in pp Collisions at sqrt(s) = 7 TeV", *Phys.Rev.Lett.* **107** (2011) 132001, doi:10.1103/PhysRevLett.107.132001, arXiv:1106.0208.

[100] D0 and CDF Collaboration, "QCD measurements at the Tevatron", arXiv:1112.0051.

[101] L. Sawyer, "Multijet Measurements with the D0 Detector", arXiv:1110.3215.

[102] D0 Collaboration, "Measurement of three-jet differential cross sections at Tevatron", *PoS* **EPS-HEP2011** (2011) 279.

[103] D0 Collaboration, "Measurement of the ratio of three-jet to two-jet cross sections in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys.Lett.B* (2012) arXiv:1209.1140.

[104] D0 Collaboration, "Measurement of three-jet differential cross sections $d\sigma_{3\text{jet}}/dM_{3\text{jet}}$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV", *Phys.Lett.* **B704** (2011) 434–441, doi:10.1016/j.physletb.2011.09.048, arXiv:1104.1986.

[105] J. Schieck, "Measurement of the strong coupling α_S in e^+e^- annihilation using the three-jet rate", arXiv:1208.4588.

[106] JADE Collaboration, "Measurement of the strong coupling from the three-jet rate in ee annihilation using JADE data", arXiv:1205.3714.

[107] The PYTHIA6 Z2 tune is identical to the Z1 tune described in [124] except that Z2 uses the CTEQ6L PDF while Z1 uses CTEQ5L.

[108] S. Gieseke, C. Rohr, and A. Siodmok, "Colour reconnections in Herwig++", arXiv:1206.0041.

[109] CMS Collaboration, "Absolute luminosity normalization". CMS-DP-2011-002, 2011.

[110] CMS Collaboration, "Measurement of CMS Luminosity". CMS PAS EWK-10-004, 2010.

[111] CMS Collaboration, "Tracking and Vertexing Results from First Collisions". CMS-TRK-10-001, 2010.

[112] Description of the HCAL Noise filter `https://twiki.cern.ch/twiki/bin/viewauth/CMS/HcalNoiseInfoLibrary`.

[113] CMS Collaboration, "Calorimeter Jet Quality Criteria for the First CMS Collision Data". CMS-PAS-JME-09-008, Apr, 2010.

[114] CMS Collaboration, "Jet Energy Resolution in CMS at 7 TeV". CMS-PAS-JME-10-014, 2011.

[115] G. Cowan, "Statistical Data Analysis". Clarendon Press, Oxford, 1998.

[116] V. Blobel, "An Unfolding method for high-energy physics experiments", `arXiv:hep-ex/0208022`.

[117] G. D'Agostini, "A multidimensional unfolding method based on Bayes' theorem", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **362** (1995), no. 23, 487 – 498, `doi:10.1016/0168-9002(95)00274-X`.

[118] M. Gluck, E. Reya, and A. Vogt, "Dynamical parton distributions revisited", *Eur.Phys.J.* **C5** (1998) 461–470, `doi:10.1007/s100520050289`, `arXiv:hep-ph/9806404`.

[119] S. Bethke, "Determination of the QCD coupling α_s", *J.Phys.G* **G26** (2000) R27, `doi:10.1088/0954-3899/26/7/201`, `arXiv:hep-ex/0004021`.

[120] CDF Collaboration, "Measurement of the Strong Coupling Constant from Inclusive Jet Production at the Tevatron $\overline{p}p$ Collider", *Phys. Rev. Lett.* **88** (Jan, 2002) 042001, `doi:10.1103/PhysRevLett.88.042001`.

[121] V. Blobel and E. Lohrmann, "Statistische und numerische Methoden der Datenanalyse". Teubner, 1998.

[122] Durham University - Online Plotting and Calculation http://durpdg.dur.ac.uk/hepdata/pdf3.html.

[123] CMS Collaboration, "Measurement of the 3-jet mass cross section at 7 TeV". CMS PAS SMP-12-027, tbp.

[124] R. Field, "Early LHC Underlying Event Data-Findings and Surprises", (2010). arXiv:1010.3558v1. preprint.

Danksagung

An erster Stelle möchte ich mich bei Herrn Prof. Dr. Günter Quast für die Möglichkeit, diese Arbeit anzufertigen bedanken. Die hervorragende Betreuung, die sich in vielen, persönlichen Diskussionen ausgedrückt hat, war sehr hilfreich für das zustandekommen dieser Arbeit. Für die Möglichkeit, wertvolle Forschungszeit am CERN zu verbringen, möchte ich Ihm hiermit auch nochmals danken. Die dortige, enge Zusammenarbeit mit den Experten hat viel zu den Ergebnissen dieser und anderer meiner Arbeiten beigetragen. Auch für sein Verständnis und seine Geduld bei persönlich Notlagen bin ich sehr dankbar.

Ebenso danke ich Herrn Prof. Dr. Wim de Boer für die Übernahme des Korreferats dieser Arbeit und die hilfreichen Diskussionen.

Herrn Dr. Klaus Rabbertz hat mit der Berechnung der fastNLO Tabellen einen enorm wichtigen Beitrag zu dieser Arbeit geleistet. Ich möchte Ihm auch danken für die sehr gute Betreuung, die wertvollen Anregungen und hilfreichen Korrekturen. Auch gemeinsam verbrachte Zeit am CERN werde ich immer in guter Erinnerung behalten.

Bei den jetzigen und ehemaligen Doktoren und Doktoranden meiner Arbeitsgruppe: Joram Berger, Dr. Volker Büge, Dr. Christoph Hackstein, Dr. Oliver Oberst, Dr. Andreas Oehler, Dr.Armin Scheurer und Dr. Manuel Zeise möchte ich mich besonders für das Training und die vielen, vielen Kommentare für diverse Talks und insbesondere diese Arbeit bedanken.

Mein Dank geht auch an alle anderen Kollegen aus dem Büro für die angenehme Arbeitsatmosphäre und die interessanten Diskussionen. Ein besonderes Dankeschön gilt auch meinem Rettungswagenfahrer Dr. Hauke Held.

Zudem danke ich allen Kollegen am EKP, die für die Bereitstellung der intensiv genutzen Computing Ressourcen zuständig sind. Auch dem gesamten Institut möchte ich für die gute Atmosphäre ein herzliches Dankeschön aussprechen.

Meinem Bruder danke ich sehr für das Korrekturlesen dieser Arbeit.

Besonders bedanken möchte ich mich an dieser Stelle bei meinen Eltern. Ihre unermüdliche Unterstützung und Förderung während meiner gesamten Ausbildung hat diese Arbeit letztlich erst ermöglicht. Vielen lieben Dank!