Improvement of the Jet Calibration Techniques for the ATLAS Experiment in LHC Run II

Bachelorthesis

UNIVERSITY OF FREIBURG

Department of Mathematics and Physics

Manuel Guth

Supervisor: Prof. Dr. Gregor Herten

Submission date: September 1 2015

Acknowledgement

The first scientific thesis is a challenge which requires the support of experienced people. I would like to give my sincere gratitude to Prof. Dr. Gregor Herten, who gave me the opportunity to write my Bachelorthesis in his group. In addition, I would like to thank everyone in the analysis team: The two guys in my office, Fabio Cardillo and Peter Tornambè, who always helped me solving some programming problems as well as Manfredi Ronzani. In addition I would like to thank Dr. Andrea Di Simone for his support. Foremost, I would like to give special thanks to Valerio Consorti and Zuzana Rúriková. They guided me through the whole work for this thesis. Especially Valerio Consorti was a great help during the last three months. Thanks also to all who reviewed the thesis and gave me helpful suggestions.

Abstract

Many analyses in particle physics need jets, clustered with different parameters or even with different algorithms. But it is not feasible to provide in-situ correction factors for each case. In this thesis a new calibration method is tested which is called "R-scan calibration". This calibration uses only one fully calibrated jet collection. By building ratios between the fully calibrated jets and partially calibrated jets, it is possible to derive in-situ correction factors for the probe jet. Especially for the new run of the LHC a quick inter-calibration like the R-scan is very useful. Therefore, during this thesis the R-scan method is designed and improved with the data acquired during Run I. After a validation of this method it is applied to the new data format of Run II.

Kurz-Zusammenfassung

Viele Analysen in der Teilchenphysik benötigen Jets, die mit verschieden Parametern oder sogar mit unterschiedlichen Algorithmen rekonstruiert wurden. Jedoch ist es nicht möglich für jeden Fall eine in-situ Kalibration bereitzustellen. In dieser Arbeit wurde eine neue Kalibrationsmethode entwickelt und getestet, die sogenannte "R-scan Kalibration". Diese Kalibration benötigt nur eine vollständig kalibrierte Jet-Kollektion. Durch das Bilden von Verhältnissen zwischen vollständig kalibrierten Jets und teilweise kalibrierten Jets ist es möglich, in-situ Korrekturfaktoren zu bestimmen. Insbesondere für den neuen Durchlauf des LHC ist eine schnelle Inter-Kalibration wie die R-scan Methode sehr nützlich. Auf Grund dessen wird die R-scan Methode in dieser Arbeit mit Hilfe der Daten aus dem ersten Durchlauf des LHC konzipiert und verbessert, um nach der Validierung der Methode auf das neue Datenformat angewendet zu werden.

Contents

1 Introduction

"There are still many open questions that need to be answered. We have every possible idea - but the final answer can only come from experimentation. And the only experiments that can provide answers are those taking place at the Large Hadron Collider. Which makes this an exciting moment - we may get new clues from outside the Standard Model to answer the questions we still have."
Hubert Kroha of the Max Planck Institute for Physics

Physics is the science looking for an explanation of the phenomena in nature and physics also strives to understand nature itself. Especially particle physics grants an insight to the most fundamental level of nature. In order to get a glimpse behind the secrets of the impressive world of elementary physics, vast experiments are necessary.

The Large Hadron Collider (LHC) is the largest particle accelerator in the world. It is designed to accelerate protons in a 27 km collider ring up to a centre of mass energy of 14 TeV. The LHC is a part of the accelerator complex at CERN (Organisation Européenne pour la Recherche Nucléaire) in Geneva. Along the LHC collider ring, there are four main experiments located. The two largest experiments are the ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid) experiments, they are general-purpose detectors and carry a wide range of physics. They are very similar, but they use a different detector architecture, especially the magnet system of the ATLAS detector is more complex. The LHCb experiment (The Large Hadron Collider Beauty experiment) is a particle detector focused on b-physics measurements. Furthermore the ALICE experiment (A Large Ion Collider Experiment) studies the quark-gluon plasma in heavy ion collisions.

This thesis is focused on the ATLAS experiment. It is the largest particle detector ever constructed: 46 m long, 25 m high, 25 m wide and weighing 7000 tonnes. The scientific goals of the ATLAS experiment are highly diversified: from proving the Standard Model of particle physics to extra dimensions and the search for components of dark matter. The detector is divided in several subsystems which record the path, the momentum and the energy of the passing particles. Because of the enormous amounts of data, a complex trigger and data acquirement system is necessary.
The greatest success in the last years was the discovery of the Higgs boson in 2012 [1] during the first run of the LHC with a centre of mass energy of 7 TeV and 8 TeV. It was the last piece of the puzzle of the current Standard Model.

After a long shut-down since February 2013, the second run of the LHC with a centre of mass energy of $\sqrt{s} = 13\,\text{TeV}$ has already started in May 2015. It provides new opportunities to search for new physics. Nevertheless the data of the current LHC run represent a new challenge for the physicists. The analyses are pushed to be more efficient and more precise to discover new physical phenomena. The acquired data from the ATLAS detector cannot be directly used for any analysis. Most of the analyses are interested in the original parton interaction but the detector measures the final state particles produced for example in hadronisations. Therefore the information about each single particle is not always important for the analysis. For this reason a tool is used to cluster these particles to an object called "jet". There are many different jet clustering methods, which differ in their algorithms or in the parameters of the algorithm like the jet radii. Each analysis needs its specific jet calibration to be able to interpret the results physically correct. Therefore it is essential to provide very good calibrations. Nevertheless, the calibration process should be as fast as possible.
The common calibration method is divided into different parts: based on simulations, considering the response of the detector and data driven methods. This thesis focuses on the data-driven methods, the so called in-situ correction. Due to the many different jet clustering methods it is not feasible to provide in-situ correction for each case. The new R-scan method, which is introduced in this thesis, provides a technique to inter-calibrate jets with different clustering methods and determines the data-driven correction factors in a less complicated and faster way.
This analysis uses ratios between fully calibrated jets and jets with missing in-situ corrections to determine the correction factors. During this thesis the R-scan method is designed, improved and validated for an usage in further analyses.

2 Theoretical Overview

In order to understand the main aspects of this thesis, it is important to know some basics about the underlying theory of particle physics. The current theory of the modern particle physics is the Standard Model (SM). This model, together with its predictions, has been tested for decades by several experiments. There are two baseline principles of the SM: first, its simplicity, in the sense that there is a minimal number of parameters; second, symmetries which define the fundamental structure of the model. All forces in nature are described in this model with the exception of gravitation. There are many indications, for example from astrophysics, that there are phenomena not described in the SM. To search for new physics, experiments at hadron colliders are currently promised to be most successful.

2.1 The Standard Model

The Standard Model of particle physics (SM) is a quantum field theory which describes the elementary particles and the fundamental interactions. Therefore it provided the basis for discovering new particles, such as the Higgs boson.
The first steps towards particle physics was the postulation of Paul Dirac that each particle has an antiparticle (1928) and Pauli's prediction of the neutrino (1930). After the discovery of the anti-electron, proton, neutron etc., the first step towards the Standard Model as it is known today was the electroweak unification (see section 2.1.2) of Sheldon Glashow (1961). In 1967, Steven Weinberg and Abdus Salam included the Higgs mechanism to the electroweak theory and this is the basic form in which the SM is known today.

2.1.1 Particles of the Standard Model

The particles of the Standard Model are divided in two basic types, called fermions (half-integer spin particles) and bosons (integer spin particles). Fermions are further subdivided into quarks and leptons.
By 1980, the three generations of the leptons were discovered as well as the existence of the first two generations of quarks were proved. Therefore leptons and quarks consist of six particles and their antiparticles. Additionally, they are grouped in three different generations, in which they will always be found in pairs (see table 2.1). Each generation has two pairs of fermions, a lepton pair, which consists of an electron-like particle and its associated neutrino, and a quark pair, consisting of an up-like and a down-like quark. Electron-like particles have an electric charge and a substantial mass, whereas neutrinos

are electrically neutral and their masses are extremely low and not yet measured: only upper limits have been set so far. At that time the third generation was not complete, the sixth quark and the corresponding neutrino of the τ lepton were not yet observed. More than ten years later, in 1995 the top quark was discovered at the Fermilab [2]. The last fermion, the τ neutrino, of the modern SM was found in 2000 also at Fermilab.
Considering the bosons, around the 1980's the self coupling of the gluon has been observed at DESY [3], the German Electron Synchrotron, by observing three-jet events and the spin 1 nature has been proved. In 1983 a big step to prove the Standard Model was done by the discovery of the $W^{\pm}$ and Z bosons, the weak mediators, at CERN [4].

	Leptons			Quarks		
Generation	Flavor	Charge [e]	Mass [MeV]	Flavor	Charge [e]	Mass [MeV]
I	e	-1	0.511	u	$^2/_3$	$2.3^{+0.7}_{-0.5}$
	ν_e	0	$< 2 \cdot 10^{-6}$	d	$-^1/_3$	$4.8^{+0.5}_{-0.3}$
II	μ	-1	105.7	c	$^2/_3$	95 ± 5
	ν_μ	0	< 0.19	s	$-^1/_3$	$1,275 \pm 25$
III	τ	-1	$1,776.8$	t	$^2/_3$	$4,180 \pm 30$
	ν_τ	0	< 18.2	b	$-^1/_3$	$173,210 \pm 510$

Table 2.1: *The mass eigenstates of the SM fermions grouped in three generations and divided in leptons and quarks [5].*

The first generation forms the ordinary matter with the lightest and stable quarks and the electron. The fermions of the second and third generations are heavier copies of the first generation's particles, with the difference that they decay very quickly in lighter particles, except the neutrinos. The leptonic number is the preserved quantum number for each single generation. Particles have the leptonic number 1 and antiparticles -1. As a consequence of this lepton conservation law, the neutrinos are stable.

In nature there are four fundamental forces: the electromagnetic force, the strong force, the weak force and the gravitational force. The Standard Model includes only three of them: the gravitation is not a part of this model. The interactions are described as the exchange of bosonic particles, called gauge bosons. Each interaction has its specific force-carrier. The mediator of the electromagnetic force is the photon (γ) while the strong force is carried by the gluon (g) which has eight different colour states. These particles have no mass, in contrast to the force carrier particles of the weak force, the $\mathrm{W}^{\pm}$ and Z^0 bosons which are massive. A overview of these bosons is shown in table 2.2. All fermions can interact weakly. Every charged particle is able to interact through the electromagnetic force. Furthermore strong interactions are only possible between quarks.

Gauge bosons					
	Mass[GeV]	Spin	Charge [e]	Coupling	Range [fm]
gluon g	0	1	0	$\alpha_S \sim 1$	~ 1
photon γ	$< 10^{-24}$	1	$< 10^{-35}$	$\alpha_{EM} \sim 1/137$	∞
$W^{\pm}$	80.385 ± 0.015	1	± 1	$G_F \sim 1.01 \cdot 10^{-5}$	$< 2 \cdot 10^{-3}$
Z^0	91.1876 ± 0.0021	1	0		

Table 2.2: *Gauge bosons of the Standard Model [5].*

2.1.2 Electroweak Unification

In 1961 Glashow combined the electromagnetic and the weak theories to the electro-weak theory [6]. Beginning with the symmetry of the Lagrangians of the electromagnetic theory and the weak theory he formed a Lagrangian with the combination of both symmetries:

$$\begin{cases} U(1)_{EM} \\ SU(2)_W \end{cases} \Rightarrow SU(2)_L \otimes U(1)_Y. \tag{2.1}$$

As a consequence of the $SU(2)$ symmetry group, the existence of a triplet of mediator particles is predicted which can couple to left-handed fermions:

$$W_\mu^1,\ W_\mu^2,\ W_\mu^3. \tag{2.2}$$

Additionally, the B_μ mediator originate from the $U(1)$ symmetry and it couples to the hypercharge Y. The hypercharge is a conserved quantum number relating the third component of the weak isospin T_3 and the electric charge, described through the *Gell-Mann-Nishijima* relation:

$$Y = 2Q - 2T_3, \tag{2.3}$$

with Q as electric charge.

Left-handed fermions are arranged in doublets and the right-handed fermions in singlets (see table 2.3). The left-handed down-like quarks are labelled differently because these are electroweak eigenstates, not to be confused with mass eigenstates here.
The mixing of the introduced mediators yields to the physical mediators:

$$\begin{aligned} W^{\pm}: &\quad W_\mu^{\pm} := \tfrac{1}{\sqrt{2}}\left(W_\mu^1 + W_\mu^2\right) \\ \gamma: &\quad A_\mu := B_\mu \cdot \cos\theta_W + W_\mu^3 \cdot \cos\theta_W \\ Z^0: &\quad Z_\mu := -B_\mu \cdot \sin\theta_W + W_\mu^3 \cdot \cos\theta_W, \end{aligned} \tag{2.4}$$

whereby θ_W is the weak mixing angle which has to be determined experimentally. It is nice to see that all physical mediators have a part of the electromagnetic and of the weak

Generations			Charges		
I	II	III	T_3	Y	Q
$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	$+1/2$ $-1/2$	-1 -1	0 -1
e_R	μ_R	τ_R	0	-2	-1
$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$+1/2$ $-1/2$	$1/3$ $1/3$	$+2/3$ $-1/3$
u_R	c_R	t_R	0	$4/3$	$+2/3$
d_R	s_R	b_R	0	$-2/3$	$-1/3$

Table 2.3: *The fermions are grouped in singlets according to the $U(1)$ symmetry group and in doublets corresponding to the $SU(2)$ symmetry group. The shown quantum numbers are the third component of the weak isospin (T_3), the hypercharge (Y) and the electric charge (Q).*

symmetry.

This model predicts massless mediators. If they had mass, the symmetry would be broken. Because of the experimentally observed non - zero mass of the Z^0 and $W^{\pm}$, a new mechanism has to be introduced.
The Brout-Englert-Higgs mechanism solves this problem by introducing a spontaneous symmetry breaking. The mechanism introduces the Higgs field which permeates all space. All fields can interact with the Higgs field and it can also interact with itself. By means of this mechanism the gauge bosons can reach mass. This mechanism was developed by Brout, Englert [7] and Higgs [8], and included in the Standard model by Salam [9] and Weinberg [10] in the late 1960.
In 2012 the Higgs like particle was discovered at CERN with a mass around 125 GeV [1]. If this Higgs particle is really the SM Higgs is not yet clear. The results of the discovered Higgs like particle are shown in the figures 2.1 and 2.2.
Cronin and Fitch (1964) observed a charge parity (CP) violation in neutral Kaon ($d\bar{s}$,$\bar{d}s$) systems. This indicates that quark flavour oscillation happens. This phenomenon can be explained with the hypothesis that two different quark eigenstates exist, depending on the interaction where quarks are involved. The relation between the mass eigenstates and the electroweak eigenstates is described through the *Cabibbo-Kobayashi-Maskawa-Matrix* [11]:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} 0.974 & 0.227 & 0.004 \\ 0.227 & 0.973 & 0.042 \\ 0.008 & 0.042 & 0.999 \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}.$$

According to this matrix it is visible for example that strangeness-changing processes are permitted.

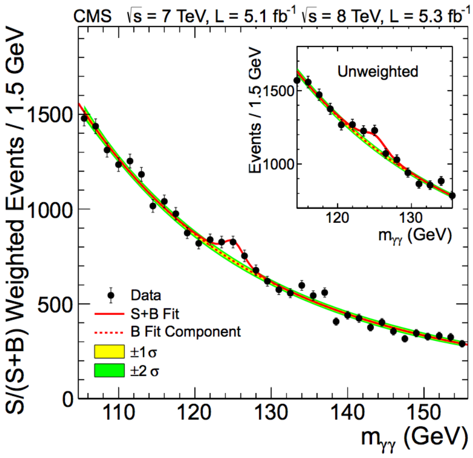


Figure 2.1: *The diphoton invariant mass distribution measured by the CMS collaboration. The red dotted line shows the Standard Model background with the uncertainties of one σ (yellow) and two σ (green). The black data points show a deviation from the Standard model trend at ~125 GeV fitted with the Higgs mechanism yields to the red line [12].*

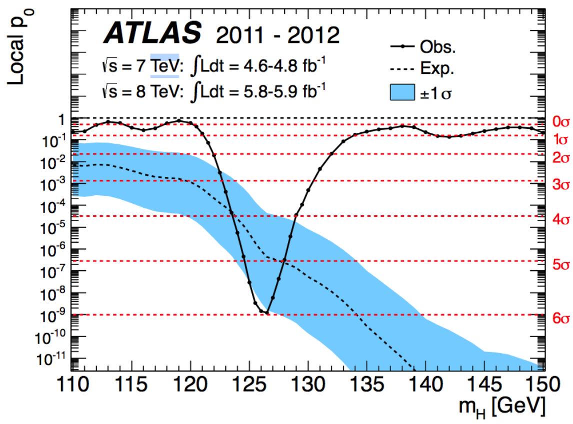


Figure 2.2: *The local p_0-value as a function of the Higgs mass, where the p-value represents the probability that the observed signal matches the underlying model. The dashed red lines indicate the significances of the p-values. The dashed black curve shows the expected p-value considering the SM Higgs and the blue band represents the $\pm 1\sigma$ uncertainty on the modelling [1].*

2.1.3 Quantum Chromodynamics

The strong force in the SM is described by the Quantum Chromodynamics (QCD). It is structured analogous to the Quantum Electro Dynamics (QED). Comparable to the electric charge, in the QCD exist the colour charge (red, green, blue) does exist. This theory is described by the $SU(3)$ symmetry group. The involved particles are quarks which carry colour charge, and anti-quarks carrying anti - colour. As mentioned above the gluon is the mediator particle of the strong force. In fact there are eight different gluons. Each of them carry a colour (r, g, b) and an anti colour $(\overline{r}, \overline{g}, \overline{b})$. As a consequence of the fact that gluons are colour charged, they can couple with themselves, the leading order yield to the vertices as shown in figure 2.3. This behaviour is not possible in the QED since the photon does not have an electric charge. Furthermore, there is another difference with respect to the QED: the potential of the QED is inverse proportional to r whereby the QCD potential is directly proportional to r. As consequence, if the distance between two quarks becomes too large, a new quark anti quark pair will be formed. This phenomenon, known as asymptotic

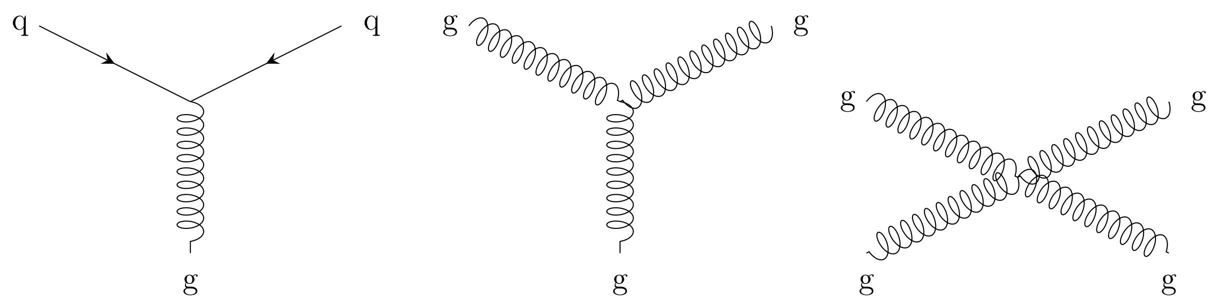


Figure 2.3: *Fundamental leading order vertices of the QCD.*

freedom, explains the behaviour that quarks cannot exist as free particles.

By including the QCD, the complete symmetry group of the Standard Model is:

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y. \tag{2.5}$$

2.2 Physics beyond the Standard Model

Even though the Standard Model is the most successful model to describe the ordinary matter, there are some observed phenomena that cannot be explained by the SM.

- Unification of the four fundamental forces
  The grand unification theory predicts that by extrapolating the gauge coupling constants to higher energies they should intersect at one point. But do they? Additionally, the gravitation is not included in the SM.
- Hierarchy problem
  It is not clear why the masses of the SM particles differ in several orders of magnitude.
- Why are there three generations?
  There is no explanation why there are three observed generations, and also the fact that quarks and leptons both have three generations is not clear because they have no other relation to each other.
- Matter-Antimatter asymmetry
  CP violation cannot explain the huge discrepancy between the matter and anti-matter. Where is the whole anti-matter?
- Dark matter
  Considering the results of astrophysics, only 4% of the universe consist of ordinary matter. Which physics hides behind dark matter?

- Dark energy
  There are also measurements in astrophysics which predict dark energy because of the expansion of the universe.

There are several models to explain these phenomena but they have not been proved yet. At the LHC, specially at ATLAS and CMS, the search for Super Symmetric (SUSY) models takes a main part of the analysis. SUSY is a new symmetry which is able to transform bosons to fermions and vice-verse. Each particle of the SM has a super-symmetric partner which differs by a half-integer spin. Some of the SUSY-Models can naturally predict for example the grand unification theory and unify the three gauge coupling constants at high energies. The lightest SUSY particles cannot further decay and therefore they are stable in some SUSY models and can explain dark matter.

2.3 Physics at Hadron Colliders

Modern high energy physics uses hadron colliders to search for new phenomena on the level of elementary particles. In contrast to electron anti-electron colliders where elementary particles are collided, hadron collisions show different behaviour because of the substructure of hadrons.

2.3.1 Parton Distribution Function

Hadrons, such as the proton, are particles composed of quarks. Further the proton does not only consist of the three valence-quarks (u, u, d), but it also contains gluons as well as sea quarks. All these constituents are called partons. Consequently, in proton-proton collisions any parton can interact. The probability that a parton carries a certain fraction of the proton momentum is described by the parton distribution function (PDF).
The PDF $f_i(x_1, Q^2)$ describes the momentum distribution of the parton i from proton 1 at an energy scale Q^2, where x_1 is the momentum fraction p_i/p_1. Figure 2.4 shows the distribution functions of the individual partons plotted against x considering two different values for Q^2.
The whole cross section of a proton-proton collision is given as:

$$\sigma = \sum_{i,j} \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 f_i(x_1, Q^2) f_j(x_2, Q^2) \hat{\sigma}_{i,j}, \tag{2.6}$$

where $\hat{\sigma}_{i,j}$ is the probability that parton i interacts with parton j. In the further analysis the parton distribution function: NNPDF2.3 leading order is used [13].

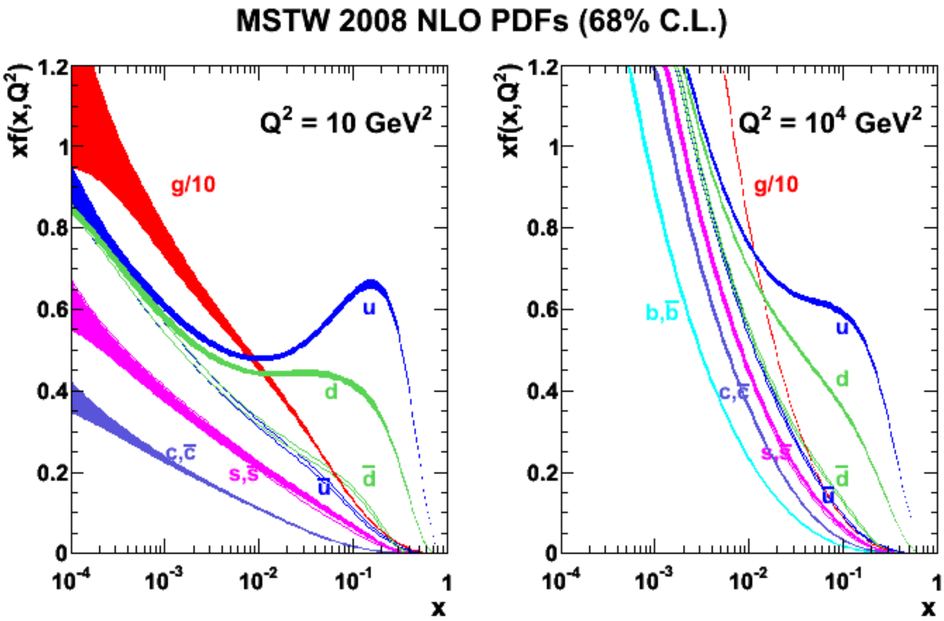


Figure 2.4: *Proton distribution functions of gluons, valence and sea quarks inside the proton depending on two different values of Q^2 [14].*

2.3.2 Monte Carlo Simulation

Monte Carlo (MC) is the tool known from mathematics based on random event generation. It is used as a method to simulate the energy and direction of particles produced in hadron collisions, using the predictions of the Standard Model. There are several steps of the MC simulation. First the hard scattering of the partons is simulated by matrix element generators considering the parton distribution functions. As shown in figure 2.5 there are initial state radiations (ISR) of gluons and photons before the hard scattering takes place, and also after the hard parton interaction, they are called final state radiations (FSR). These radiations produce hadron showers which have to be considered in the simulation.
The next step of the simulation is the modelling of the hadronisation. Apart from the hard scattering there are partons of the collided protons which produce underlying events. Furthermore, the hadron decay of unstable particles has to be simulated. Finally some QED radiations have to be done, like bremsstrahlung. The procedure is shown as graph in figure 2.6.
These are all simulated truth events, i.e. the simulation is done without considering any response of the detector. An additional step is to simulate the reconstructed events. This is done by simulating the response of the detector to the truth event.

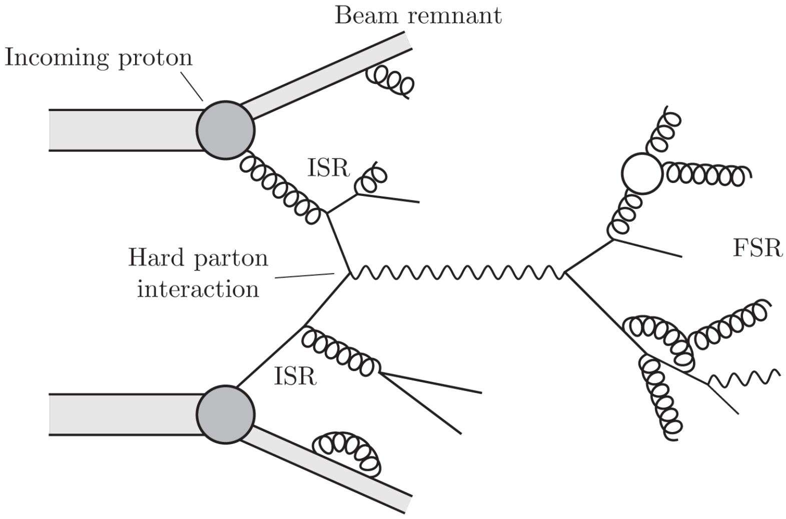


Figure 2.5: *Schematic overview of a pp-collision [15].*

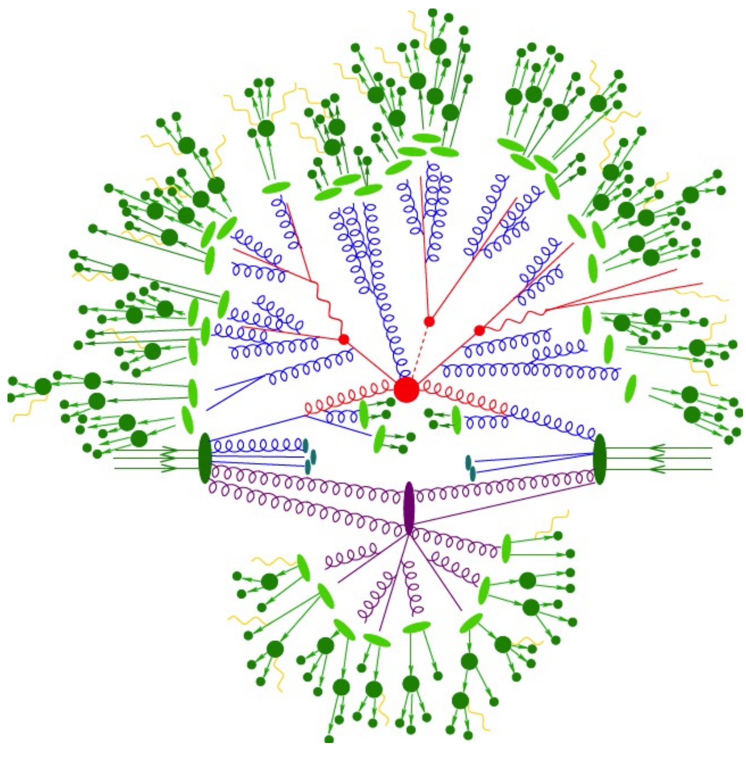


Figure 2.6: *MC collision simulation. The hard scattering process is simulated by matrix element generators shown in red. Parton showers are highlighted in the blue lines. Any secondary interactions - or rather underlying events - are marked in purple. The hadronisation is shown in green and hadron decays in dark green, the QED bremsstrahlung in yellow [16].*

2.3.3 Jets

QCD allows only colourless states of free particles; consequently, quarks cannot exist as single particles. Therefore a new quark anti-quark pair will be produced, if the distance between two quarks becomes too large. This is possible because of the potential of the strong force:

$$V(r) \sim r. \tag{2.7}$$

If this procedure happens several times a jet will be formed. Thus, a jet is a spray of collimated particles originating from the hadronisation of a parton produced in a hard scattering process [18].

The functional diagram is shown in figure 2.7. In this case there is no difference between an electron anti-electron collision ($e^+e^- \to q\bar{q}$) and a hadron collision.
A simple example of the fact that jets get more collimated at higher energies, is the two body decay:

$$A \to B_1 B_2. \tag{2.8}$$

Assuming that in the final state two identical massless particles are created as shown in figure 2.8, the four-momentum of the particles is given as:

$$P_A = (E_A/c, |\vec{p}_A|, 0, 0) \tag{2.9}$$
$$P_{B_{1,2}} = (E_{B_{1,2}}/c, \pm|\vec{p}_{B_{1,2}}|\cos\theta, 0, \pm|\vec{p}_{B_{1,2}}|\sin\theta). \tag{2.10}$$

The following equation is fulfilled in the centre of mass system:

$$m_A c^2 = 2E_{B_{1,2}} \tag{2.11}$$

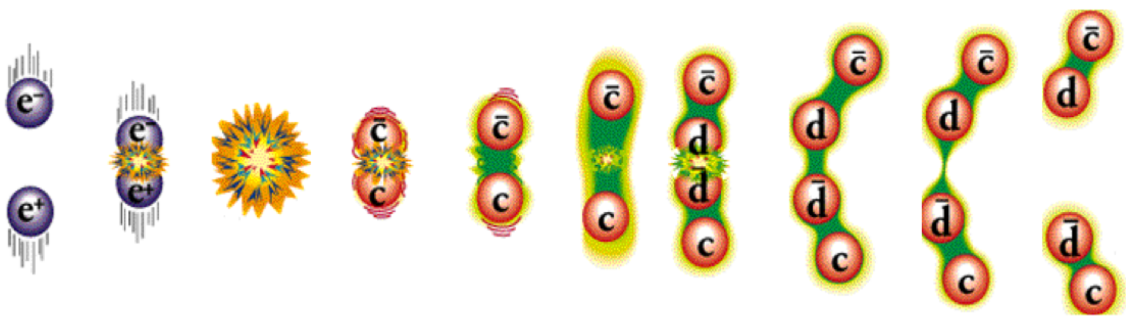


Figure 2.7: *Principle of hadronisation. After the hard scattering boosted quarks fly off in different directions. Because of the asymptotic freedom, a new quark anti-quark pair will be formed in between if the distance is too large. [17]*

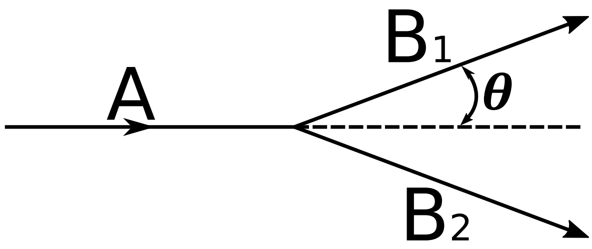


Figure 2.8: *Scheme of the two body decay with two identical massless particles in the final state.*

Therefore the four-momentum can be written as:

$$P_{B_{1,2}} = \frac{m_A c}{2}(1, \pm\cos\theta, 0, \pm\sin\theta). \tag{2.12}$$

For the laboratory - system, the Lorentz - transformation yields:

$$P_{B_{1,2}} = \frac{m_A c}{2}(\gamma \mp \beta\gamma\cos\theta, -\beta\gamma \pm \gamma\cos\theta, 0, \pm\sin\theta) \tag{2.13}$$

$$\Rightarrow E_{B_{1,2}} = \frac{E_A}{2} \mp |\vec{p_A}|c\cos\theta. \tag{2.14}$$

And the angle θ is now given as:

$$\theta = \arccos\left[\frac{E_A - 2E_{B_{1,2}}}{|\vec{p_A}|c}\right]. \tag{2.15}$$

By plotting θ against p_A as shown in figure 2.9, θ decreases with higher momentum of the initial particle.

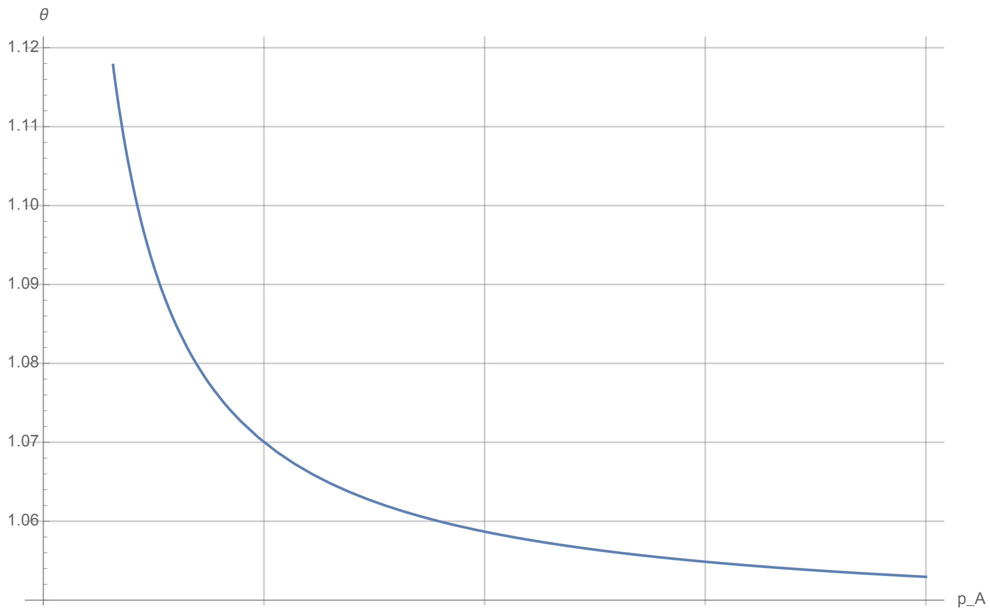


Figure 2.9: *Dependency of* θ *and* p_A.

2.3.3.1 Jet Algorithms

Jets are clustered and reconstructed by using different Jet-Algorithms depending on the physics processes. It is important to mention that jets are not physical objects, they are defined by their algorithms. In modern physics mostly iterative cone algorithms are used. As a starting point there is only information from the detector of single particles which deposed energy in the calorimeters. The energy clusters are called initial proto-jets. They will be gradually be combined as new protojets and finally to jets.
Each proto-jet is characterised by its transverse momentum $k_{T,i}$, its azimuthal angle ϕ_i and its pseudo-rapidity η_i, defined as:

$$\eta_i = -\ln(\tan(\theta_i/2)), \tag{2.16}$$

where θ is the polar angle of the detector. The advantage of using η instead of θ is due to the fact that the pseudo-rapidity is a high energy approximation of the rapidity which is Lorentz invariant. For a more flexible definition ζ is an additional parameter which allows to switch between different jet algorithms ($\zeta = 1$: inclusive k_t algorithm [19]; $\zeta = 0$: Cambridge/Aachen algorithm [20]; $\zeta = -1$ anti-k_t algorithm [21]). Basically, the sequential recombined jet-algorithms work according to the following scheme [19]:

1. For each protojet, define the distance between particle i and the beam (B):

$$d_{iB} = k_{T,i}^{2\zeta} \tag{2.17}$$

and for each pair of protojets define the distance:

$$d_{ij} = \min(k_{T,i}^{2\zeta}, k_{T,j}^{2\zeta})\frac{R_{ij}^2}{R^2}, \tag{2.18}$$

where

$$R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \tag{2.19}$$

describes the angular distance between two objects. R is the maximal allowed distance. For example in ATLAS the most common values of the parameter R are: 0.4 (used in this analysis), 0.6 or 1.0.

2. Find the smallest of all d_{iB} and d_{ij} and label it d_{min}.

3. If d_{min} is a d_{ij}, merge protojets i and j into a new protojet l with:

$$k_{T,l} = k_{T,i} + k_{T,j} \tag{2.20}$$

and

$$\eta_l = (k_{T,i}\eta_i + k_{T,j}\eta_j)/k_{T,l} \tag{2.21}$$
$$\phi_l = (k_{T,i}\phi_i + k_{T,j}\phi_j)/k_{T,l} \tag{2.22}$$

4. If d_{min} is a d_{iB}, the corresponding protojet i is "not mergable". Remove it from the list of protojets and add it to the list of jets.

5. Go to step 1.

Most ATLAS analyses use the anti-k_t jet clustering algorithm [21]. The anti-k_t algorithm is characterised by setting $\zeta = -1$ in equation (2.18). The advantage of the anti-k_t algorithm: it is very safe against infra-red effects and collinear gluon radiations, and it is also a flexible adaptation to hard substructures in the jets.

2.3.3.2 Jet Calibration

In order to be able to analyse the physics behind the events produced by pp-collisions, it is necessary to calibrate the energy and the momentum of the resulting jets. Both are reconstructed from the energy deposition in the calorimeters. In ATLAS the mostly common method is a combination of Monte Carlo simulation and data-driven techniques. There are a lot of effects which have to be considered. The detector-based corrections like dead material, where the energy gets lost in inactive parts of the detector. Also corrections for the different energy scales measured from electromagnetic and hadronic showers are needed. Another issue is represented by the showers reaching the outer edge of the calorimeter. If the energy of some particles is included in the truth jet[1] but not in the

[1] A truth jet is a modelled jet without considering the response of the detector.

reconstructed jet it has to be adjusted. Moreover, clusters are only formed if the deposited energy is well above the background noise. Therefore particles with too small depositions are lost, and this effect must be corrected. It is also important to correct the energy against the effect of pile-up: because of multiple pp-collisions the energy deposition can originate from different bunch crossings or different pp-collisions within the same bunch crossing. Finally the in-situ correction compares the p_T of reference objects, like photons, Z-bosons or other jets, and the jets being calibrated. This is done for Monte Carlo(MC) and data, the correction factor looks like this [22]:

$$\left.\frac{p_T^{jet}}{p_T^{ref}}\right|_{data} \Big/ \left.\frac{p_T^{jet}}{p_T^{ref}}\right|_{MC} . \tag{2.23}$$

The logical scheme of the whole calibration procedure in ATLAS is shown in figure 2.10. First it has to be decided whether the calibration uses EM (electromagnetic scale) or the local cell signal weighting method (LCW). LCW corrects the signals from hadronic shower fluctuations, and thereby the fluctuations due to the non-compensating response of the ATLAS calorimeter are reduced [23]. The next important step is the pile-up correction as mentioned above. Another calibration is the jet energy scaling (JES), whereby the energy and the direction are calibrated. To correct flavour dependence and energy leakage effects the global sequential calibration (GSC) is done. At last the in-situ correction factors will be determined.

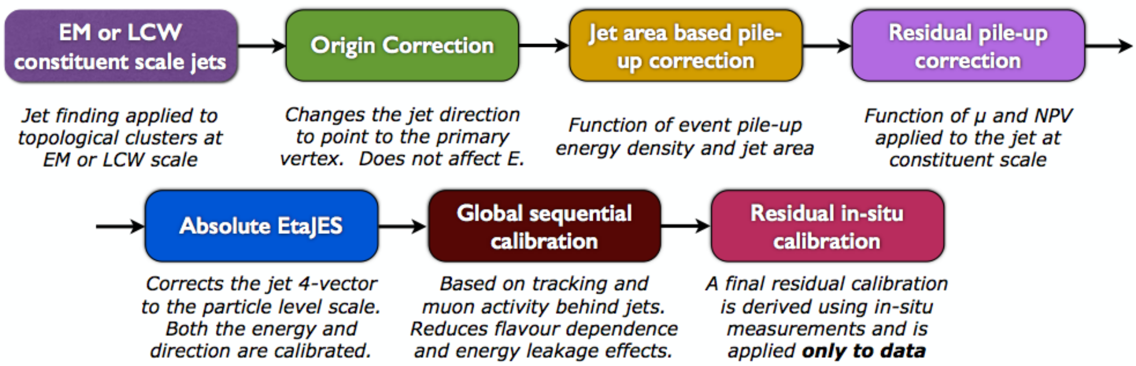


Figure 2.10: *Logical diagram of the calibration procedure used in ATLAS [22].*

3 The ATLAS Experiment at the Large Hadron Collider

In order to be able to search for new physics at a high energy scale, complex experiments are needed. Only large collaborations can manage to maintain experiments of these dimensions. One of them is the ATLAS experiment. It is a particle detector placed at the Large Hadron Collider (LHC) at CERN in Geneva.

3.1 CERN

The European Organization for Nuclear Research, CERN, is one of the largest institutions for high energy physics, based in Meyrin, Geneva. It was founded in 1954 by 12 states and was one of the first joint research organisations of Europe. Today there are 21 member states and over 600 institutes and universities around the world which participate in this project [24].
The main part of the research at CERN concerns in particle accelerators and particle detectors. The aim of this research is to get new insights about the elementary structure of the matter which is around us and to find new physics, such as dark matter.

3.2 The Large Hadron Collider

The Large Hadron Collider (LHC) represents the centrepiece of CERN's accelerator complex. It is currently the largest and most powerful particle accelerator in the world. The 26.7 km collider ring is designed to collide protons at a centre of mass energy up to 14 TeV. Additionally, lead (Pb) can be accelerated, used in heavy ion experiments. The collider ring is placed in the same tunnel as the previous collider LEP (Large Electron-Positron Collider).
Before the protons can be injected to the LHC, they have to be accelerated in several steps. The first step where the protons receive an initial boost of 50 MeV is the linear accelerator *LINAC2* as shown in figure 3.1. Next, the protons pass in the *Proton Synchrotron Booster* (PSB) after which they will be transferred to the *Proton Synchrotron* (PS) with an energy of 1.4 GeV. The PS accelerates them up to an energy of 25 GeV. At an energy of 450 GeV, which will be reached in the *Super Proton Synchrotron* (SPS). At that point the protons are ready to be injected into the LHC, the main accelerator ring. After 20 minutes in the LHC they finally reach their maximum energy which currently is $\sqrt{s} = 13\,\text{TeV}$. To

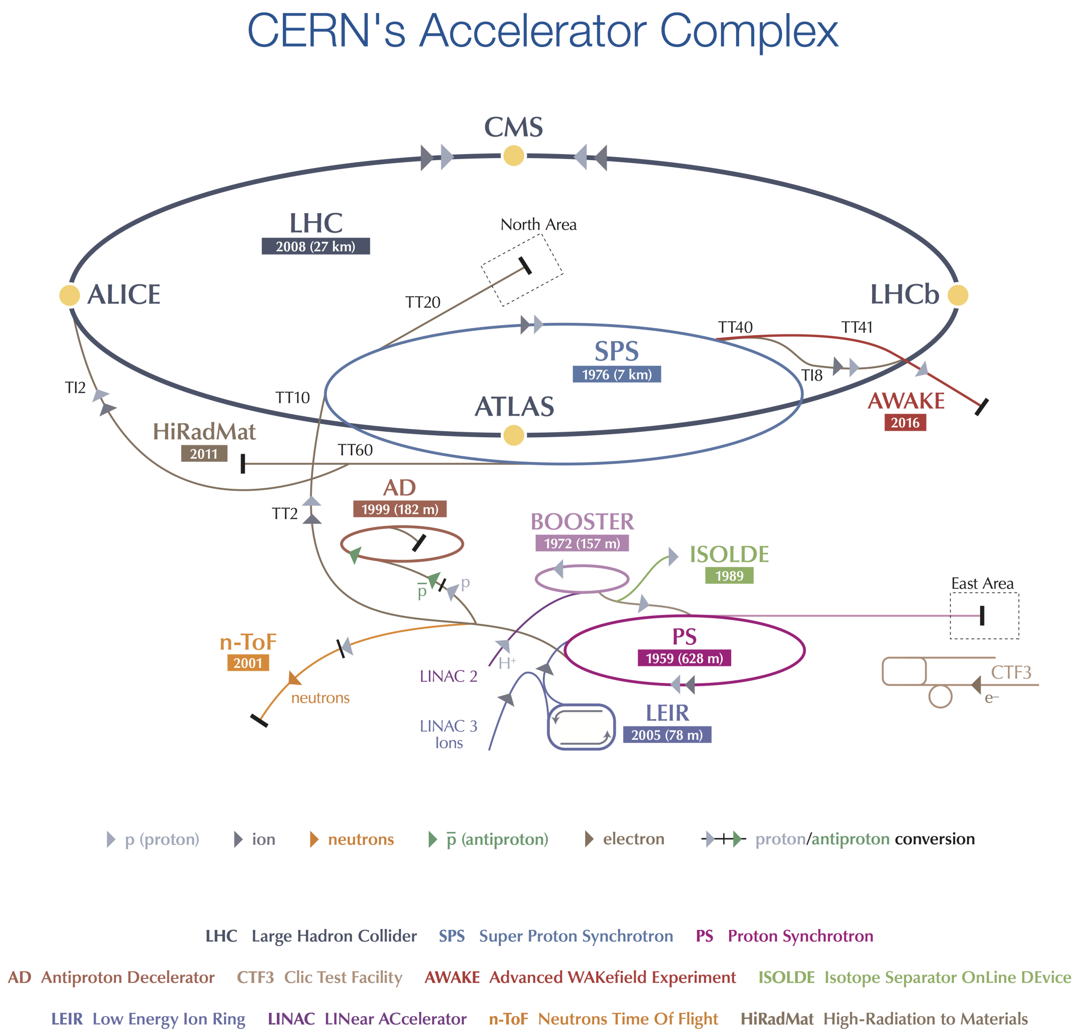


Figure 3.1: *Accelerator complex and detectors at CERN in Geneva [26].*

accelerate the protons to these high energies there are 16 superconductive radio-frequency cavities which generate a strong electric field. Furthermore, along the LHC ring 9593 super conductive magnets are installed which are responsible to hold the proton bunches in the beam pipe and to focus the beam. The magnets are cooled to a temperature of 1.9 K by a liquid helium system [25].

Along the collider ring there are four main particle detectors (see figure 3.1):

- ATLAS (*A Toroidal LHC ApparatuS*) and CMS (*Compact Muon Solenoid*):
  General-purpose detectors with a wide range of physics (from Standard Model physics to the search of dark matter).

- ALICE (*A Large Ion Collider Experiment*):
  A detector to study the quark-gluon plasma, where quarks and gluons behave as free

particles.

- LHCb (*The Large Hadron Collider Beauty experiment*):
  A forward detector focused on b-physics measurements.

3.3 The ATLAS Experiment

The ATLAS experiment is one of the two major experiments at the LHC. The heart of this experiment is the ATLAS detector, a general-purpose particle detector. It is designed to cover a wide range of physics, from determining properties of SM particles to the search for evidence of dark matter. In order to ensure this, a huge hardware and software system is necessary. The dimensions of the detector are: 46 m long, 25 m high, 25 m wide and weighing 7000 tonnes. The ATLAS detector is the largest particle detector ever constructed and it is situated 100 m below ground near Meyrin, Switzerland.
The detector is structured in four main subsystems. The whole detector has a cylindrical symmetry (a overview is shown in figure 3.2). These systems record path, momentum and energy of the particles, whereby several magnets are needed to measure the momentum of the particles. These measurements produce an enormous flow of data wherefore a complex trigger system is indispensable.
In the following, the different parts are described in detail.

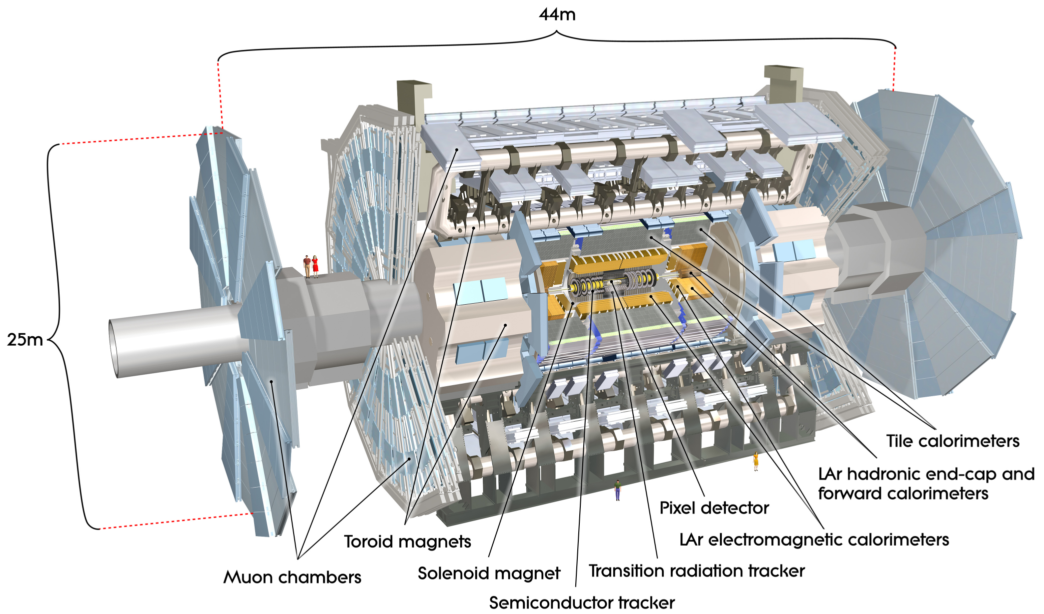


Figure 3.2: *Schematic overview of the ATLAS detector [27].*

Magnet system

The magnet system in ATLAS is needed to bend the particle tracks in the tracking region of the detector to determine their momentum from their curvature. It is divided in two subsystems. First the central solenoid magnet which is located between the inner detector and the calorimetry. This super-conductive magnet system operates at a temperature of 4.5 K and is responsible for the momentum reconstruction of low energy and fast decaying particles in the tracking section. It is able to create a magnetic field up to 2 T. The second magnet system is located at the outmost part of the detector. It consists of three super-conductive air-core toroid magnets which consist of eight coils each. The super-conductive toroids cover a large volume and generate a magnetic field up to 3.9 T at the barrel region and at the end-caps up to 4.1 T.

Inner Detector

The inner detector encloses directly the interaction point where the collisions take place. Its architecture is divided into three subsystems: Pixel Detector, Silicon Micro-Strip Detector (SCT) and Transition Radiation Tracker (TRT), as shown in figure 3.3.

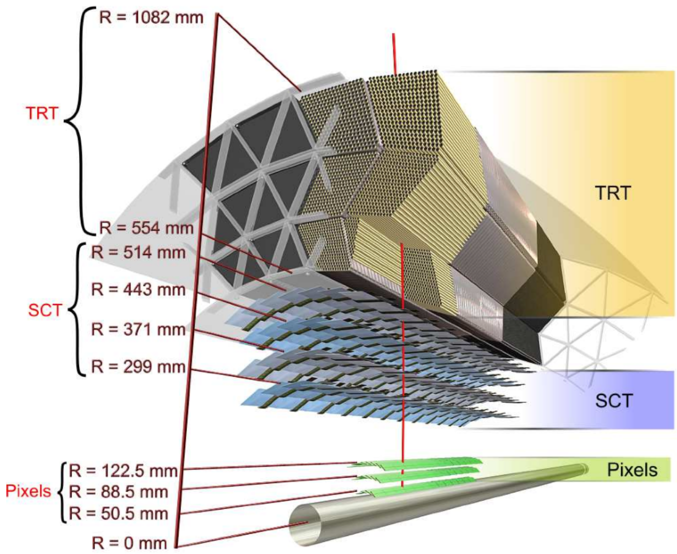


Figure 3.3: *The inner detector with its three sub-detectors (Pixels, SCT and TRT) [28].*

The pixel detector consists of three cylindrical silicon sensor layers and the new Insertable B-Layer (IBL) which reduces the distance between the interaction point and the first tracking layer [29]. Each sensor layer has a lot of single silicon detectors arranged as pixels. Altogether they have a large number of channels ($\sim 10^8$). Hence this results in a very good spatial resolution of about 10 µm to 115 µm. Secondly, the silicon micro-strip detector is arranged in concentric cylinders around the pixel detector and delivers good spatial charged track information for the momentum reconstruction. The transition radiation tracker forms the outer part of the inner detector and records around 30 hits per track. Its purpose is to measure the photons from transition radiation to provide additional points for fitting the trajectory of the particles. In addition, hadron and electron decay products can be distinguished.

Calorimetry System

The calorimetry system is divided into three different parts: the electromagnetic calorimeter, the hadronic calorimeter and the forward calorimeter (see figure 3.4). They are mainly based on liquid argon technology to ensure a long lifetime in such a radioactive environment.
The electromagnetic calorimeter is the inner part of the calorimetry system and it measures the energy and position of electromagnetic showers. It shows the typical cylindrical geometry at the central part, the so called barrel, which covers a pseudo-rapidity range of $0 < |\eta| < 1.475$. Additionally, end-caps are disposed at both ends. They ensure a pseudo-rapidity range up to $|\eta| = 3.2$.

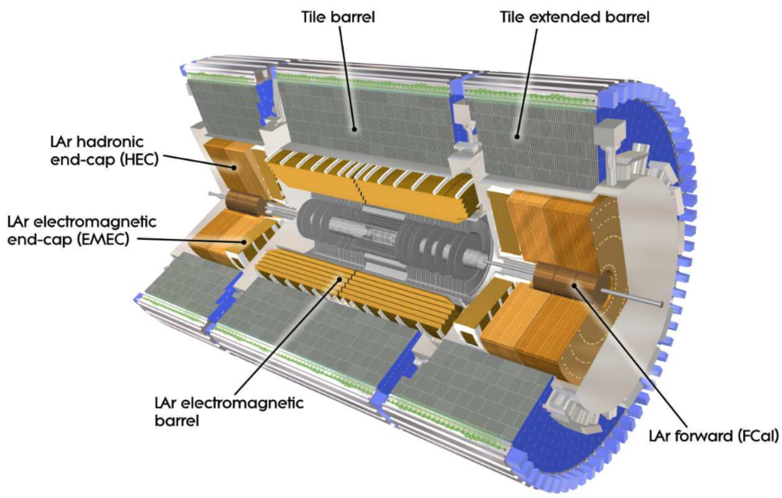


Figure 3.4: *The calorimetry system of the ATLAS detector [28].*

The hadronic calorimeter is directly adjacent to the electromagnetic calorimeter. However, it uses a different technology in the barrel region. There it consists of plastic scintillators spaced out with iron absorbers. Instead the hadronic end-caps are again based on liquid argon technology. The main task of this calorimeter section is to exactly determine the energy of hadronic showers and nuclear fragments. The granularity of the samplings increases in high η. In the central region ($\eta < 2.5$) it has the finest granularity is $\Delta\eta \times \Delta\phi \approx 0.003 \times 0.1$, in order to have the best match with ID tracks. At high η the size of the cells increases up

to $\Delta\eta \times \Delta\phi \approx 0.025 \times 0.25$.
Finally the forward calorimeter (FCal) is responsible for the pseudo-rapidity range from $|\eta| = 3.1$ to $|\eta| = 4.9$. It is a combination of electromagnetic and hadronic calorimeters. The FCal directly surrounds the beam pipe at the ends of the detector.

Muon Spectrometer

The outermost hardware part of the ATLAS detector is the muon spectrometer (MS). As the name suggest, this spectrometer measures the coordinates, charges and momenta of muon tracks. In order to provide the momentum of a track, the muon system is embedded in three large superconducting air-core toroids magnets to deflect the trajectory of the track. The MS is divided into two main subsystems: the precision chambers and the trigger chambers.

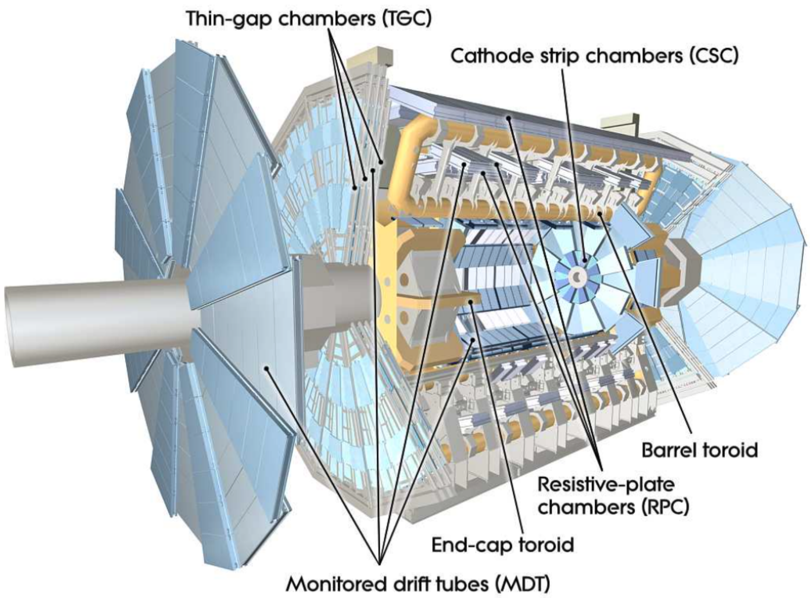


Figure 3.5: *A Cut-away view of the ATLAS muon system [28]*

The first system, the precision chambers, consists of Monitored Drift Tubes (MDT) which measure the track coordinates very precisely. Furthermore, at the forward region there are Cathode-Strip Chambers (CSC). The trigger chambers are the second part of the MS. In the barrel they are built using Resistive Plate Chambers (RPC's) while at the end-caps Thin Gap Chambers (TGC) are used to cope with a more busy environment. These trigger chambers identify muons flying through the MS and enable the data acquisition in the region of interest. It also measures the track coordinates orthogonal to the precision chambers that yield to a better track reconstruction.

Trigger System and Data Acquisition

Recording the large amount of data is one of the greatest challenges of ATLAS. On the one hand, it is not optimal to lose any information about the processes happening in the detector, on the other hand, it is not possible to acquire and record all of them. In addition, not every event is physically important (soft scattering or radiation processes). As a consequence, a trigger system is needed that selects the important events and records them for further analysis.

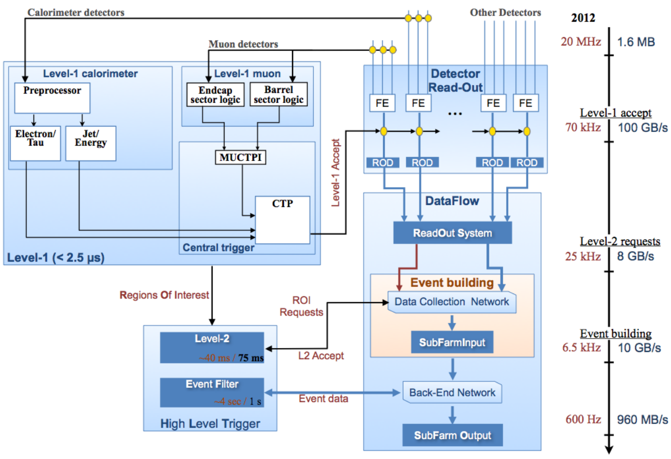


Figure 3.6: *Schematic overview of the Trigger and DAQ system of RUN I [30].*

The original ATLAS trigger system was based on a three-level system: Level 1 (hardware trigger), Level 2 (software trigger) and an Event Filter as shown in figure 3.6. Since the second run of the LHC has started the ATLAS detector has a new trigger management because of the expected increasing of the rates by a factor of 5-6. The Level 1 trigger is still there, fed by the information of the calorimeter and muon detectors it decides if the events are relevant or not (requires about 2 μs) based on a region of interest (RoIs) mechanism.

The Level 2 and the Event Filter are merged to the new High-Level Trigger (HLT) with a rate reduction from 100 kHz to 1 kHz. At this point the information of the Level 1 trigger and of the inner detector are compared and it can be chosen if the HLT runs either on the RoIs only or on the full event information. This method allows a reduced complexity and it is simultaneously more flexible. In addition, it reduces duplicate data-fetching and is therefore more efficient resource-wise. The schematic overview is shown in figure 3.7.

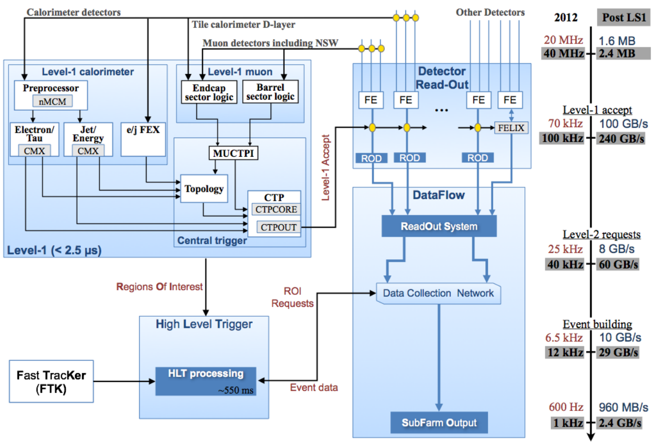


Figure 3.7: *Schematic overview of the Trigger and DAQ system of RUN II [30].*

4 R-scan calibration

There are many analyses which need calibrated jets with different radii but it's not feasible to provide in-situ calibrations for each jet-collection. A new method is tested which is able to inter-calibrate jets with different radii very fast, avoiding an ad-hoc in-situ calibration. This method is called R-scan since it aims to provide calibration factors for many jet collections with different jet radii. The correction factors can be derived by building ratios between fully calibrated reference jets and partially calibrated probe jets. In the following analysis the R-scan method is tested with the data acquired during the LHC run I ($\sqrt{s} = 8\,\text{TeV}$).

4.1 Method

The R-scan calibration uses samples of dijet events. This has the reason that the kinematic of dijet events is well defined and the transverse momentum (p_T) of the two jets is balanced. At least one fully calibrated jet-collection is required for the R-scan calibration. The fully calibrated jets are used as reference to calibrate jets with missing in-situ corrections. The aim of this calibration is to determine exactly these in-situ correction factors for the probe jets.
In this analysis jets are clustered with the anti-kt algorithm. The reference jets are clustered with the parameter R equal to 0.4. Since the R-scan aims to calibrate many jet collections with different cone radii, in the following, to explain the method, the probe jet will be denoted with the generic R parameter equal to "X". In the following the probe jets are also referred to as R-scan jets. The calibration factors are derived by using two ratios. The first:

$$R_{MC} = \frac{(p_T^X)_{MCC+JES+GSC}}{(p_T^{0.4})_{MCC+JES+GSC}} \tag{4.1}$$

is used to model the dependency on the two different jet radii. Jets with different radii have a different geometry and the energy can be differently clustered. Both MC samples are fully calibrated. As described in section 2.3.3.2, JES and GSC are calibration methods, and MCC summarizes the MC based calibrations: origin correction and pile-up corrections. In Monte Carlo the in-situ calibration is not applied because the in-situ calibration is a data driven correction based on well-known events and therefore the difference between MC and data is the in-situ calibration.
The second ratio is:

$$R_{DATA} = \frac{(p_T^X)_{MCC+JES+GSC}}{(p_T^{0.4})_{MCC+JES+GSC+InSitu}}. \tag{4.2}$$

The numerator and the denominator differ by the in-situ calibration and the radii. Since the MC ratio (R_{MC}) handles the geometry difference between the two jet collections, the data ratio (R_{DATA}) takes into account of the difference arising from the in-situ calibration. For this reason only the two ratios together can provide an inter-calibration factor.
The two ratios shown in the equations (4.1) and (4.2) are mapped as a function of η and of the p_T - average, where p_T^{avg} is defined as the average value of the leading jet (LJ) and the sub-leading jet (SJ):

$$p_T^{avg} = \frac{p_T^{LJ} + p_T^{SJ}}{2}. \tag{4.3}$$

The average value of p_T is only calculated for the reference-jet collection. This is necessary since the trigger thresholds are parametrised in p_T^{avg}. The reason to use the average value of the transverse momentum in a dijet analysis is to minimize the impact of energy fluctuations related to detector resolution effects. The p_T of the leading and sub-leading jet in a dijet event should be the same. Due to the energy fluctuations it is possible that either the leading jet p_T is overestimated or the sub-leading jet p_T is underestimated. However, the average value mainly compensates this effect.

The calibration factor shows a different behaviour in the forward (high $|\eta|$) and the central regions (low $|\eta|$), because there are different detector technologies and also different physics processes. Additionally, the jet-geometry changes with higher p_T, the jets get more collimated, as described in section 2.3.3. Therefore there are different corrections needed (i.e. the correction related to the "out of cone" energy or the pile-up subtraction). As a consequence a p_T - dependency of the calibration factors is expected. Therefore the analysis is divided into different η - and p_T - slices. The resulting ratios R_{MC} and R_{DATA} show a normal distribution which is fitted with a Gaussian function:

$$f_i(x) = a \cdot \exp\left[-\frac{(x-\mu_i)^2}{2\sigma^2}\right], \tag{4.4}$$

where i is either MC or DATA.
However, the MC and data ratio have no relation yet. In order to compare the results of the fitted distributions a double ratio is built:

$$R_{final}(p_T^{avg}, \eta) = \frac{\mu_{DATA}(p_T^{avg}, \eta)}{\mu_{MC}(p_T^{avg}, \eta)} \tag{4.5}$$

This final ratio already contains the information about the calibration factors for the probe jet collection.
Since the results depend on p_T^{avg}, a mapping back to p_T^{ref} has to be done. The mapping is necessary because the calibration factor has to be applied for each jet and not for an average value of two jets. After the mapping, the double ratio $R_{final}(p_T^{ref}, \eta)$ provides the in-situ correction factors for the $R = X$ probe-jet collection.

4.2 Analysis

The preliminary analysis uses the full luminosity acquired during the first run of LHC with $\sqrt{s} = 8\,\text{TeV}$. The data was taken in 2012 at the period E-L which corresponds to an overall luminosity of $20.3\,\text{fb}^{-1}$. In run I, two fully calibrated data samples with different jet radii are available, therefore it is possible to test the R-scan method and validate it. The fully calibrated reference-jet collection is the Anti-k_t4TopoEM. This means that the jets are clustered with the anti-k_t jet algorithm by setting the parameter $R = 0.4$ (see section 2.3.3), and the calibration is based on the electro-magnetic scale. As R-scan collection the Anti-k_t6TopoEM is used.

The analysis is structured in different parts. An important task of a calibration is to cover a large p_T-range. For this reason a complex trigger strategy must be used in this analysis. Apart from the events resulting directly from the hard scattering, there are a lot of effects like ISR, FSR, pile-up events (described in section 2.3) or soft scattering processes of protons. In addition, the non-collision background can contaminate or fake physics collisions. The non-collision backgrounds includes for example cosmic particles (mostly muons), detector noise or beam-backgrounds (i.e. interactions with the residual gas in the beam pipe). In order to suppress this background a proper event selection is necessary. The event selection provides a sample of dijet events adapted to the analysis. Since the reference and probe jets differ only in the radii of the clustering algorithm, they should propagate in the same direction within a certain angular distance to each other. If the jets fulfil the matching criteria, the ratios as shown in the equations (4.1) and (4.2) are built. Then the ratio distributions, subdivided into different p_T and η regions, can be fitted. The fit results are plotted as function of p_T^{avg} and η. As described in section 4.1 the results have to be mapped to p_T^{ref}. For the p_T-mapping, quadratic and linear interpolation among different p_T^{avg} bins are tested. Finally the results are compared with a given calibration to validate the R-scan method.

η - bin	$\lvert\eta\rvert$
1	0.0 – 0.8
2	0.8 – 1.2
3	1.2 – 2.1
4	2.1 – 2.8
5	2.8 – 3.2
6	3.2 – 3.6
7	3.6 – 4.5

Table 4.1: η - *binning*

As described above the analysis is divided in different η-slices because of different detector technologies and different physical processes in the central and forward region.

The η-binning has to be thorough to such an extent that the different regions can be distinguished but if the binning is too fine there is too less statistic and therefore both criteria have to be considered. In this analysis the η - binning in table 4.1 is chosen.

The dijet events are simulated with the MC generator Pythia 8 [31]. It simulates pp-collisions from hard process to a complex multi-hadronic final state. The procedure of Monte Carlo simulations has been already described in section 2.3.2.
This analysis covers a wide p_T range. The number of events decreases exponentially with the jet p_T. To achieve a high statistic in the tails of the p_T distribution, an enormous number of events would have to be simulated. Therefore the MC simulation is divided into different p_T samples (J0-J7) as shown in figure 4.1. For each sample the same number of events are generated and afterwards are renormalised.

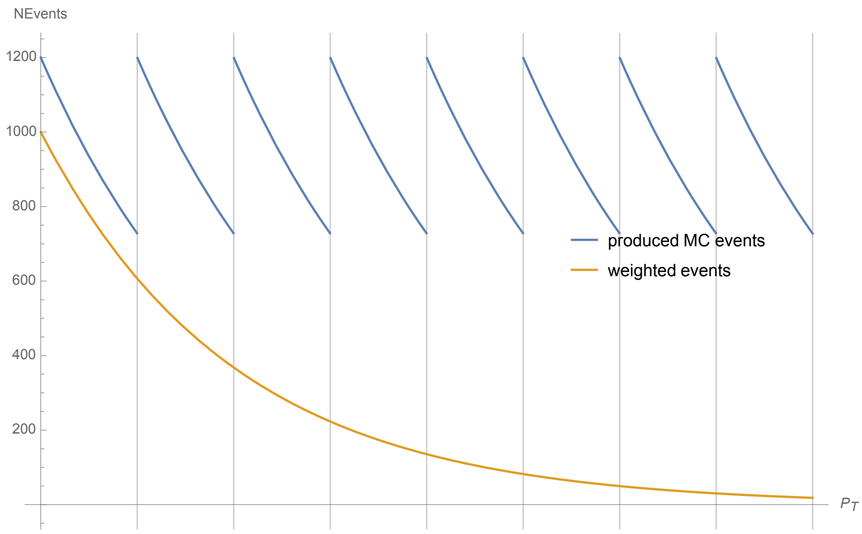


Figure 4.1: *For each defined p_T regions J0-J7, Monte Carlo samples are produced. For each sample the same number of events are generated (blue line) for a better statistic. After the generation norm - factors are applied for each slice (orange line).*

Detector effects like smearing, leakages or interaction with the detector material have to be simulated, too. The simulated pile-up of MC does not match the data. Therefore the pile-up has to be re-weighted. This is for example done by comparing the number of primary vertices (NPV) distribution bin by bin. Additionally, each trigger has a different pile-up weight but in MC there are no triggers applied and this analysis uses several triggers. For this reason, the applied pile-up weights ρ are based on the $p_{T,avg}$ of the dijet-system which corresponds to the given triggers. This is possible because each trigger is defined in its specific p_T^{avg} bin and therefore a clearly assignment can be done to apply the pile-up weight. The normalisation factor β for MC is given as:

$$\beta = \frac{\sigma \epsilon \rho}{N}, \tag{4.6}$$

where σ is the cross-section, ϵ the filter-efficiency of generation process and N the number of events. In addition, MC is normalised to the whole luminosity of the data-samples.

4.2.1 Trigger Strategy

In this analysis, single jet triggers are used. As mentioned above, this analysis covers a wide p_T - range. In low p_T - regions there are a lot of events and it is not possible to record each of them. Therefore low p_T trigger have pre-scale factors. These factors define that only a fraction of the events passing the trigger is actually recorded. Due to the different pre - scale factors, the data collected by the different triggers corresponds to different luminosities. In order to avoid double counting of the events, the triggers are further selected by comparing these luminosities. For each event the trigger with the highest luminosity $\mathcal{L}$ is chosen as the best because of a better statistic. The weight ν for the data is directly proportional to the pre - scale γ:

$$\nu = \gamma \cdot \xi = \frac{1}{\mathcal{L}}, \tag{4.7}$$

where ξ is the online trigger efficiency.
The efficiency of a perfect trigger could be described with a heaviside step function. But physical triggers have a non negligible turn-on as shown in figure 4.2. In this analysis the thresholds are defined as the point where the trigger efficiency exceeds 99%. According to these thresholds, the p_T^{avg} - binning is adapted to minimise the impact of energy fluctuations as described in section 4.1. The triggers are mainly divided into central and forward triggers because of different detector technologies, the very different rates and also different physics processes in these two regions. Each trigger is defined in its specific $p_{T,avg}$ range where the minimum threshold is given by the turn-on of the trigger and the maximum threshold by the turn-on of the next higher trigger. The triggers used with their corresponding p_T^{avg} are shown in table 4.2. Up to 270 GeV each p_T^{avg} bin has one corresponding trigger while at higher p_T^{avg} only three trigger exist.
For each event several triggers are firing. The online trigger selects the events very fast by clustering the jets with the anti - k_t jet algorithm with $R = 0.4$ at hadronic scale. Afterwards the offline trigger simulation checks which trigger should have fired, these triggers are called emulated trigger.
Therefore the trigger selection for each event is divided into several steps: A match between online and emulated trigger is required. Then the trigger with the highest p_T threshold is chosen. If there are two triggers left, i.e. central and forward trigger, the forward trigger is chosen because the forward triggers have a higher luminosity.

4.2.2 Event Selection

In order to guarantee a good quality of the data, only data sets from the good run list (GRL) are taken. The GRL comprises information about the status of the LHC (i.e. if there is a stable beam) and about lots of parameters of the ATLAS detector during the data taking. The runs with optimal conditions are called physics run.

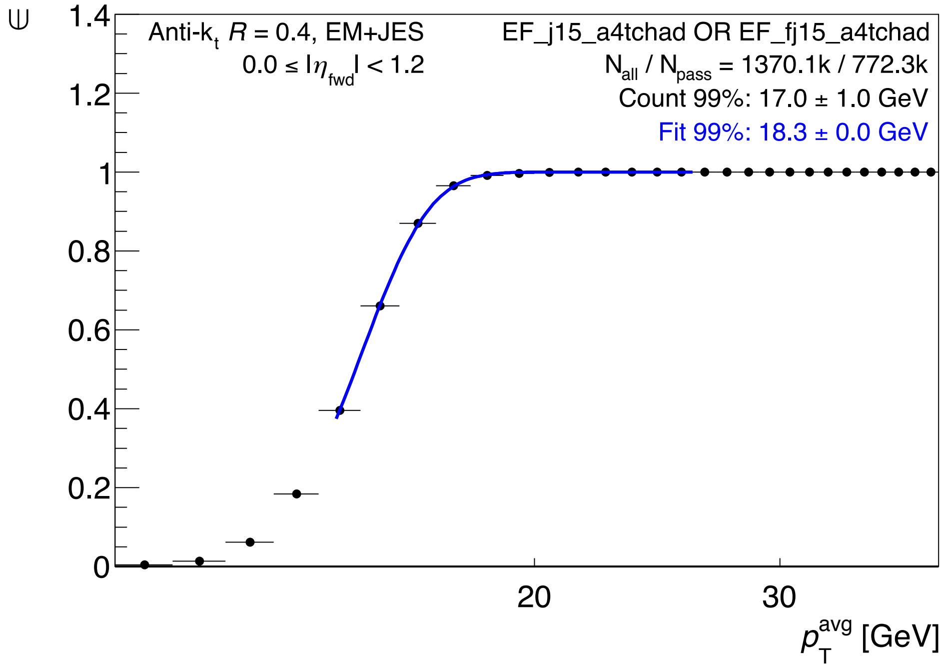


Figure 4.2: *Efficiency of the EF_j15_a4tchadOREF_fj15_a4tchad trigger as function of* p_T^{avg} *[32].*

The dijet analysis needs two reference jets and two probe jets in each event. The p_T-threshold of each jet is:

$$p_T = 25\,\text{GeV} \tag{4.8}$$

and the upper η-limit is:

$$|\eta| < 4.5. \tag{4.9}$$

But the recorded events are rarely perfect dijet events. The two highest p_T jets are mostly produced with a recoil system composed of multiple lower p_T jets (mainly from soft quark and gluon radiation). Recoil jets are allowed if they do not influence the balanced transverse momentum of the leading dijet system as described in section 4.1. The events are rejected if:

$$p_T^{jet3} > \text{Max}[0.4 \cdot p_T^{avg}, 12\,\text{GeV}]. \tag{4.10}$$

In addition, the leading and sub-leading jets have to be approximately back-to-back. Therefore a $\Delta\phi$-cut is done between the leading and sub-leading jet:

$$\Delta\phi > 2.5. \tag{4.11}$$

The jets are required to fulfil some quality criteria that remove fake jets originating from the background mentioned in section 4.2. Therefore the jets have to originate from one primary vertex. The jet vertex fraction (JVF) is defined as the sum of p_T of tracks associated to the

P_T^{avg}-bin	P_T^{avg} slice [GeV]	Central Trigger – $\lvert\eta\rvert < 3.1$ Name	Pre-scale γ	Forward Trigger – $\lvert\eta\rvert > 2.8$ Name	Pre-scale γ
1	$25 - 40$	EF_j15_a4tchad	$1.4 \cdot 10^6$	EF_fj15_a4tchad	241,610
2	$40 - 55$	EF_j25_a4tchad	258368	EF_fj25_a4tchad	27,748
3	$55 - 65$	EF_j35_a4tchad	44768	EF_fj35_a4tchad	8,681
4	$65 - 85$	EF_j45_a4tchad	112573	EF_fj45_a4tchad	5,339
5	$85 - 115$	EF_j55_a4tchad	45888	EF_fj55_a4tchad	4,965
6	$115 - 145$	EF_j80_a4tchad	8752	EF_fj80_a4tchad	243
7	145-175	EF_j110_a4tchad	2066	EF_fj110_a4tchad	13
8	175-220	EF_j145_a4tchad	559	EF_fj145_a4tchad	2
9	220-270	EF_j180_a4tchad	257	EF_fj180_a4tchad	1
10	270-330	EF_j220_a4tchad	78		
11	330-400				
12	400-525	EF_j360_a4tchad	1	EF_fj220_a4tchad	1
13	525-760				
14	760-1200				
15	1200-1500				

Table 4.2: *The trigger list of the used triggers in this analysis for the data of run I with their pre-scale factors. They are divided in central and forward triggers and each trigger has its corrsponding p_T slice.*

jet that originate from the primary vertex, divided by the sum of p_T of associated tracks from all vertices. This cut is important to suppress the effect of pile-up which is very high in low p_T regions. Any leading or sub-leading jet with $|\eta| < 2.4$ or $p_T < 50\,\text{GeV}$ has to fulfil:

$$\text{JVF} > 0.25. \tag{4.12}$$

This condition is also applied to the third jet, if it is present. This selection is applied to MC and data.
In addition to the trigger analysis described in section 4.2.1, an additional trigger check is implemented. Since the triggers are defined in different p_T ranges, it is required that the p_T^{avg} matches the corresponding range. In addition, if the trigger is a forward trigger, the η range is checked, too.

4.2.3 Matching

After the event-selection there is a sample with high quality dijet events but it is not yet checked how the R-scan and reference jets match. These jets differ in their radii but the direction of their propagation vector should not be very different. If the jets match within a defined angular distance, they can be compared and can be used in this analysis. The schematic view of a dijet - system is shown in figure 4.3. The allowed angular distance $R_{i,j}$ between the reference and R-scan jet is:

$$R_{i,j} < 0.4, \tag{4.13}$$

where $R_{i,j}$ is defined in equation (2.19).
Because of the energy fluctuation (see section 4.1), it is not obvious that the leading reference jet matches the leading R-scan jet and likewise with the sub-leading jets. Therefore all four matching possibilities have to be checked and also if these matchings are possible. For example, if the R-scan leading jet matches the leading reference jet, then the R-scan sub-leading jet cannot match the reference leading jet as well. If this happens, the event is rejected.
The ratios as shown in the equations (4.1) and (4.2) are calculated depending on which jets are matched. In figure 4.3, for example, the leading reference jet and the sub-leading

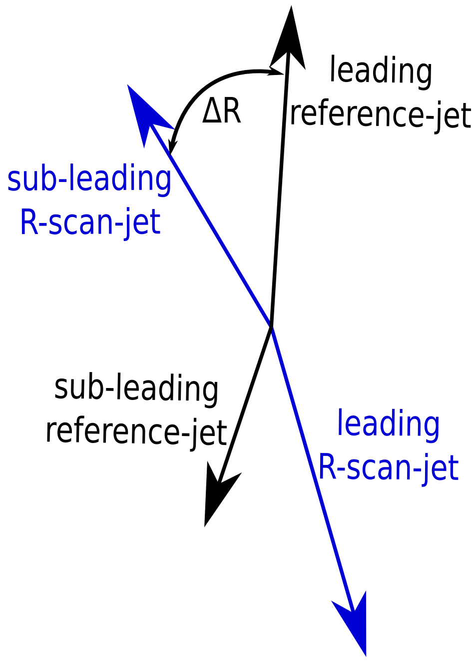


Figure 4.3: *The dijet system with R-scan and reference jets.*

R-scan jet match within the defined radii, thereby the ratio is given as:

$$R_l = \frac{p_T^{R-scan,SJ}}{p_T^{ref,LJ}} = \frac{p_T^{0.6,SJ}}{p_T^{0.4,LJ}} \tag{4.14}$$

and the second ratio is:

$$R_l = \frac{p_T^{R-scan,LJ}}{p_T^{ref,SJ}} = \frac{p_T^{0.6,LJ}}{p_T^{0.4,SJ}}, \tag{4.15}$$

where l is either DATA or MC.
These ratios are weighted (see eq. (4.6) & (4.7)) and mapped together as a function of η in different p_T^{avg} - slices. The upper plot in figure 4.4 shows the resulting ratio distribution of data for p_T^{avg} bin $270\,\mathrm{GeV} < p_T^{avg} < 330\,\mathrm{GeV}$, which corresponds to the central trigger EF_j220_a4tchad. And the lower plot in figure 4.4 shows the MC ratio distribution for the same p_T^{avg} bin.

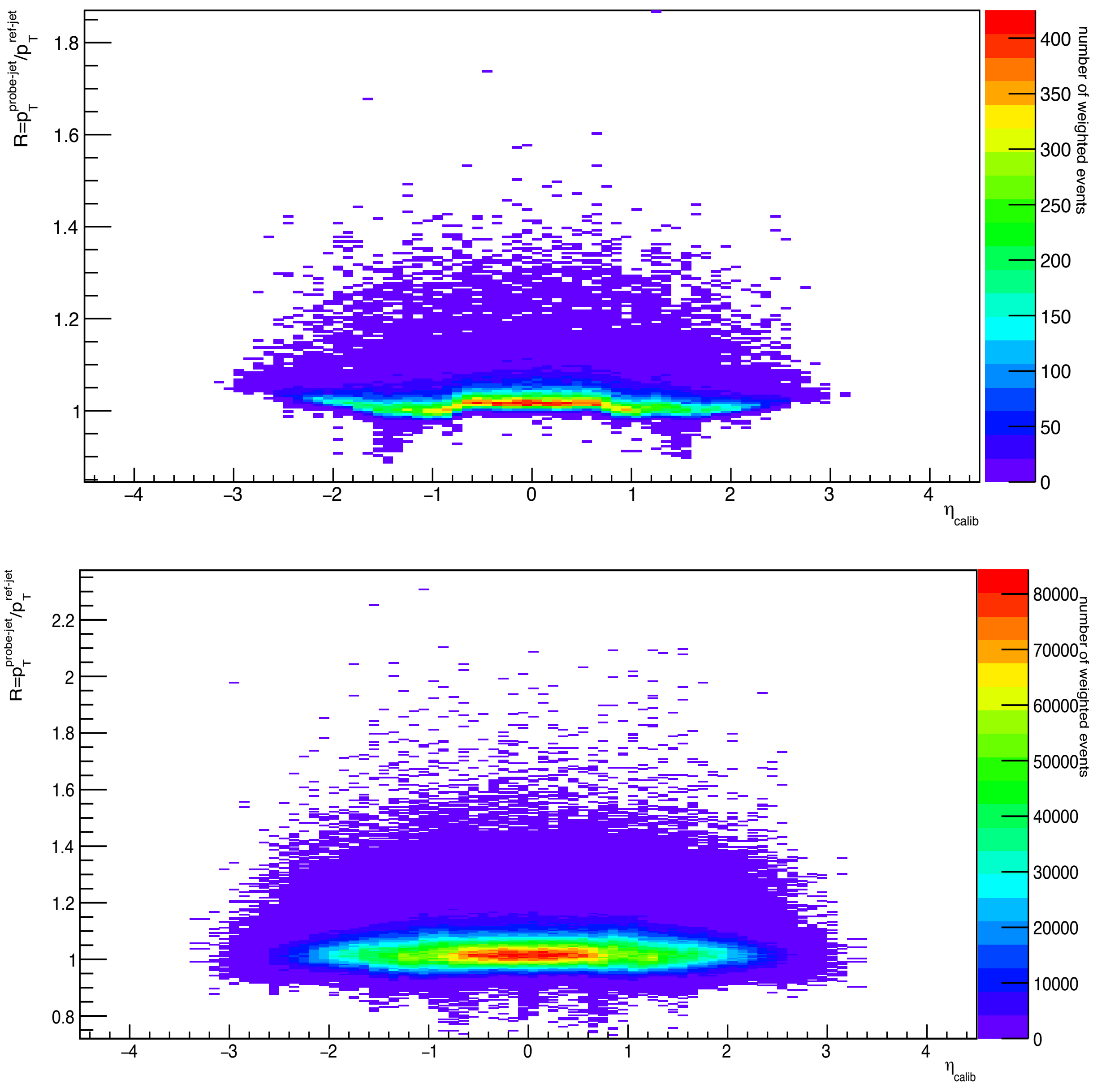
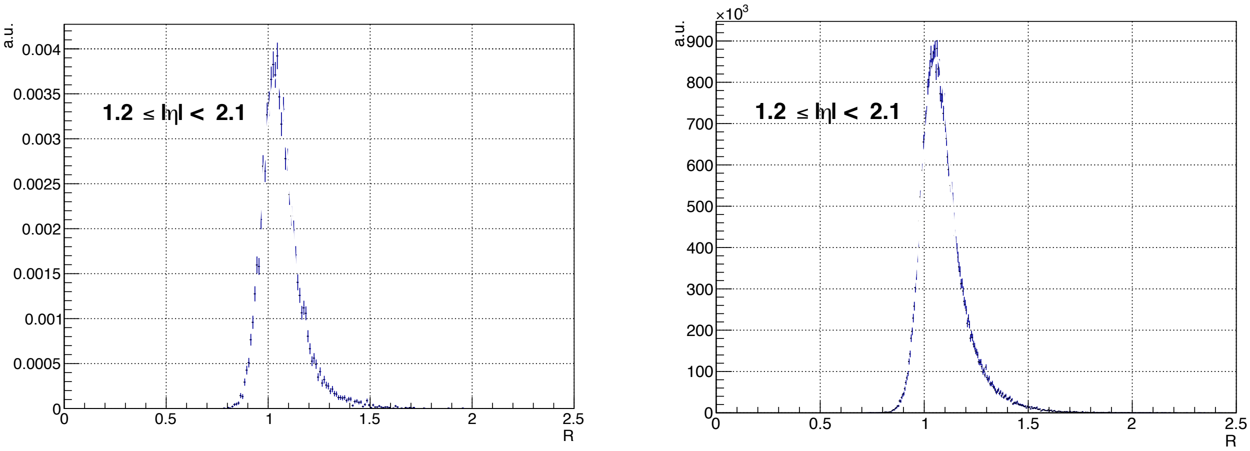


Figure 4.4: *These two histograms show the p_T ratio mapped as a function of η in the p_T^{avg}-region:* $270\,\mathrm{GeV} < p_T^{avg} < 330\,\mathrm{GeV}$. *The upper plot shows the data ratio and the lower one the MC ratio.*

4.2.4 Fitting

The distribution shown in figure 4.4 provides the main information for the rest of the analysis. In order to split the analysis into different $|\eta|$ regions, the 2-D histograms are projected in the different η-bins.
The resulting projections are distributed as shown in figure 4.5.



DATA *MC*

Figure 4.5: *Data and MC ratio distributions for the η-region $1.2 \leq |\eta| < 2.1$ and a p_T^{avg}-range of* $65\,\text{GeV} < p_T^{avg} < 85\,\text{GeV}$.

As expected, the ratios are dependent on p_T and $|\eta|$. The higher is the p_T, the narrower is the corresponding ratio distribution. This behaviour is shown in the figure 4.6. Their full width at half maximum differ by a factor 4.5. In addition, the higher $|\eta|$ or p_T the lower the statistics. In order to ensure a better statistic, the histograms are rebinned with Scott's normal reference rule [33], which minimises the integrated mean squared error of the density estimate. All histograms which have no content are directly rejected. Because of the decreasing statistics in higher $|\eta|$ or p_T most of the empty histograms are in these regions.
Most of the distributions show a small tail, but nevertheless a Gaussian fit is justified, because the impact of these tails should not influence the fitting result too much. The fitting function is given as:

$$f(x) = a \cdot \exp\left[-\frac{(x-\mu_i)^2}{2\sigma^2}\right], \tag{4.16}$$

where a is a normalisation factor - neglected in this analysis - μ the mean value and σ the standard deviation. The important fit value is the mean value μ of the distribution and, of course, its error.
The fitting is divided into two steps: first, the mean and RMS (root mean squared) of the

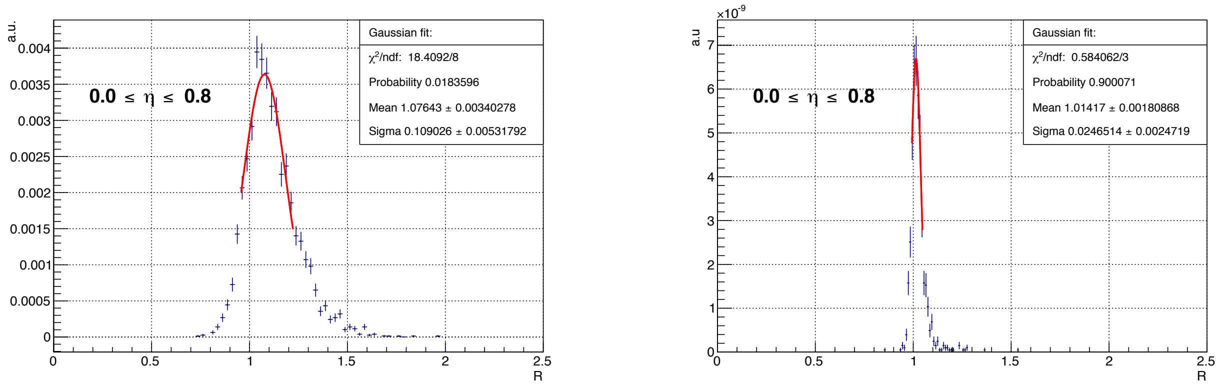


Figure 4.6: *Data ratio distribution in the* $|\eta|$ *slice:* $0 \leq |\eta| < 0.8$*. The right plot has the range:* $40\,\mathrm{GeV} < p_T^{avg} < 55\,\mathrm{GeV}$*; and the left plot:* $1200\,\mathrm{GeV} < p_T^{avg} < 1500\,\mathrm{GeV}$*.*

histogram are used as parameters to initialise the first fit. But this fit is not very stable and does not guarantee high quality results. Therefore a second fit is done by using the parameters of the first Gaussian fit to initialise the second fit. An asymmetric fit range is used to cut off the mentioned tails and improves the stability of the fit. The Gaussian fits are shown, for example, in figure 4.6. If the number of degrees of freedom (ndf) of the fit is ≤ 1, the fit has no significance and the mean value of the histogram and its statistical uncertainty are used for further analysis. This case is shown in figure 4.7: The fit failed but, since the distribution has a significant number of entries, the mean value provides the best result in this case.

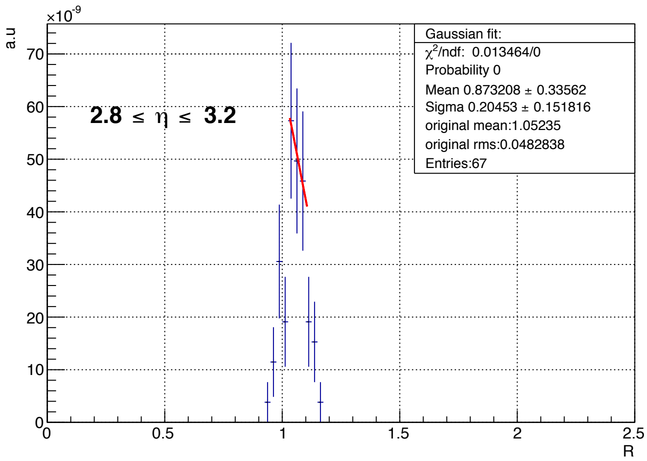


Figure 4.7: *Failed fit of the data ratio with* $2.8 \leq |\eta| < 3.2$ *and* $270\,\mathrm{GeV} < p_T^{avg} < 330\,\mathrm{GeV}$

The results of the Gaussian fits, i.e. the mean value μ of the ratio distributions, are plotted against p_T^{avg}, separated in MC and data and the plots are divided into different $|\eta|$ slices.

In addition, the ratio:

$$R_{final}(p_T^{avg},\eta) = \frac{\mu_{DATA}(p_T^{avg},\eta)}{\mu_{MC}(p_T^{avg},\eta)} \tag{4.17}$$

is built and plotted against p_T^{avg} for each $|\eta|$ bin.

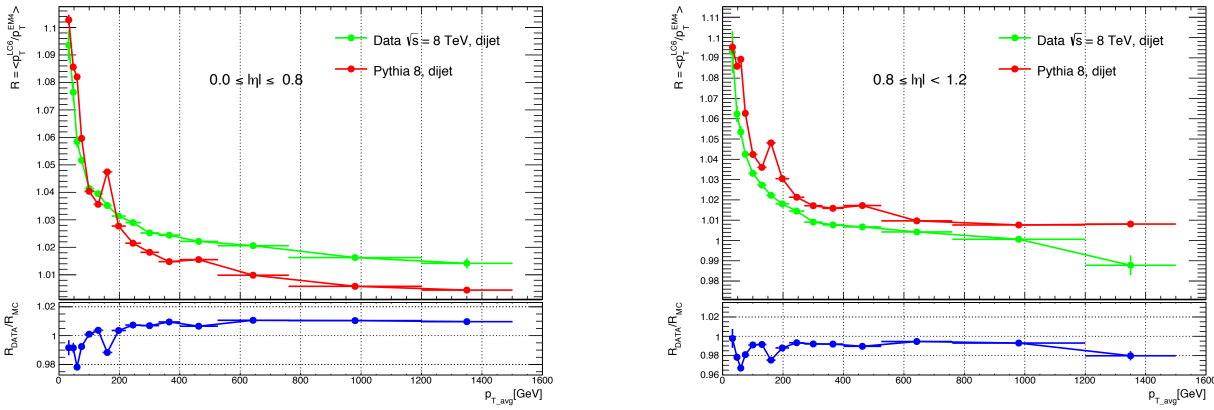


Figure 4.8: *MC and data ratio and the* R_{final} *plotted against* p_T^{avg}*, the left plot in the* $|\eta|$*-slice* $0 \leq |\eta| < 0.8$ *and the right plot in the* $|\eta|$*-slice* $0.8 \leq |\eta| < 1.2$.

The resulting plots are presented in the figures 4.8 - 4.11. The big error bars in p_T^{avg} are used because the ratio fits were only divided into p_T^{avg} bins and therefore the points in the plots are set in the middle of the bin and the error bars cover the whole bin size. The statistical uncertainties of the ratios R_{MC} and R_{DATA} are mostly negligible, except in the forward region (see figure 4.10-4.11).

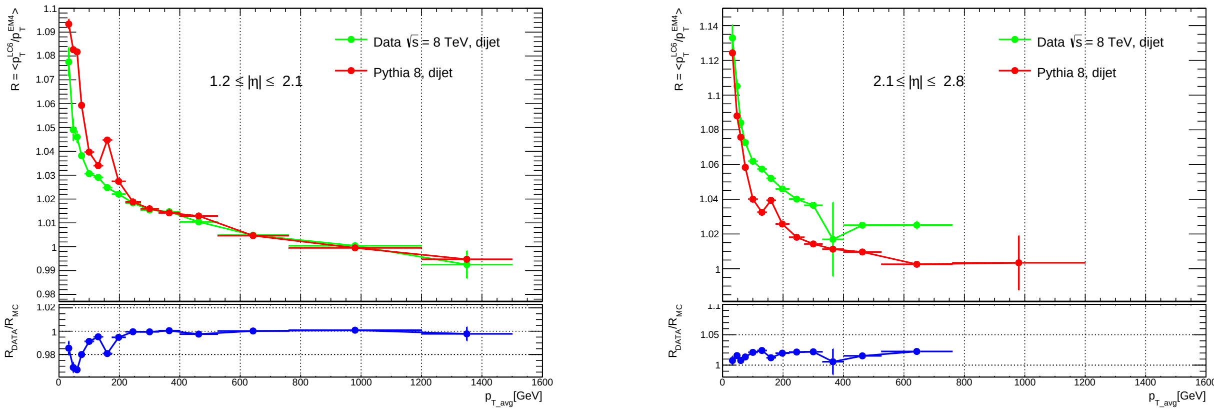


Figure 4.9: *MC and data ratio and the* R_{final} *plotted against* p_T^{avg}*, the left plot in the* $|\eta|$*-slice* $1.2 \leq |\eta| < 2.1$ *and the right plot in the* $|\eta|$*-slice* $2.1 \leq |\eta| < 2.8$.

At this point a trend is visible that the data and MC ratios decrease with higher p_T. In most cases data has higher values except in the $|\eta|$-slice $0.8 \leq \eta < 1.2$. In the central region (figure 4.8 and 4.9), $R_{final}(p_T^{avg}, \eta)$ fluctuates around 1, while in the forward region (figure 4.10 and 4.11) there is a difference of $\sim 5\%$ visible.

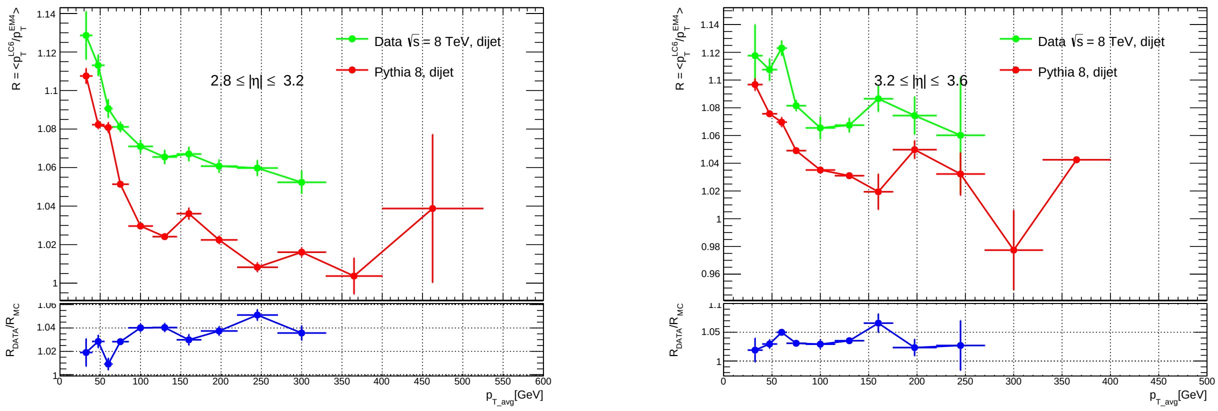


Figure 4.10: *MC and data ratio and the* R_{final} *plotted against* p_T^{avg}*, the left plot in the* $|\eta|$*-slice* $2.8 \leq |\eta| < 3.2$ *and the right plot in the* $|\eta|$*-slice* $3.2 \leq |\eta| < 3.6$.

Considering only the global curvature trend of the first three $|\eta|$ bins, the analysis seems to be roughly stable. But in each $|\eta|$ slice there is an outlier in R_{MC}, always in the same p_T^{avg} region between 145 GeV and 175 GeV. The MC ratio distribution and the fit is shown in figure 4.12; this is a very asymmetric distribution, and because of the tail the value of μ is overestimated. The mean value of the fit does not match the high point of the distribution.

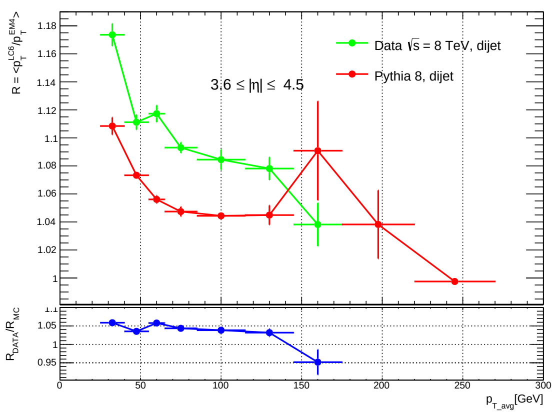


Figure 4.11: *MC and data ratio and the* R_{final} *plotted against* p_T^{avg} *in the* $|\eta|$*-slice* $3.6 \leq |\eta| < 4.5$.

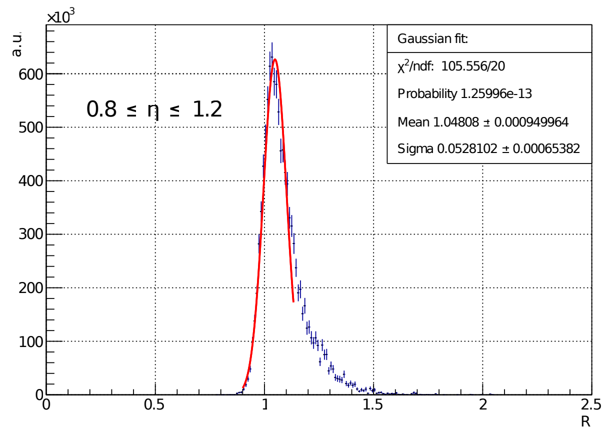


Figure 4.12: *MC ratio distribution:* $0.8 \leq |\eta| < 1.2$ *and* $145\,\text{GeV} < p_T^{avg} < 175\,\text{GeV}$ *with a dominating tail falsifying the fit.*

Therefore a better and more stable fitting process is necessary. For this reason a different method is used. This is again an iterative fitting method. The first step is the same as before, the mean, RMS and the norm are taken from the histogram and a first Gaussian fit is done with these parameters in a range of three standard deviations in each direction from the mean value. Then a second fit is done with the parameters of the first fit with a range limited to two standard deviations. If the second fit converges, these parameters are used for the final fit with a range of one standard deviation, else the original mean, RMS and norm from the histogram are taken as parameters for this fit. In the worst case scenario where the final fit fails, the mean value from the histogram is taken for the further analysis.

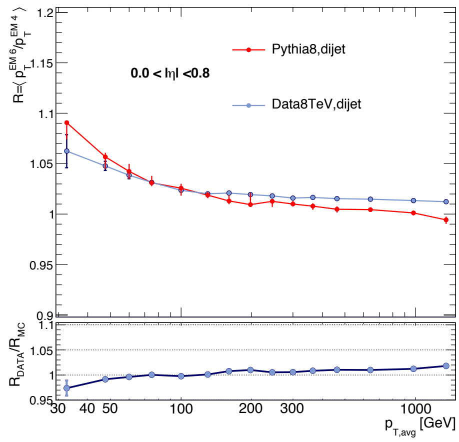


Figure 4.13: *MC and data ratio and the R_{final} plotted against p_T^{avg} in the η-slice $0 < |\eta| < 0.8$ – new fitting method.*

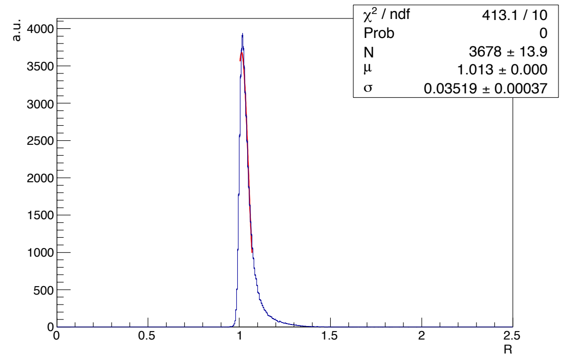


Figure 4.14: *MC ratio distribution: $0 \leq |\eta| < 0.8$ and $145\,\text{GeV} < p_T^{avg} < 175\,\text{GeV}$ with the new fitting method.*

Two main systematic effects are considered in this analysis. The first is related to the pile-up JES components and the second is related to JES itself. These uncertainties are varied up-wards and down-wards and all systematics are added in quadrature. They are applied to the reference and probe collection as well as to MC and data. Because of varying in both directions there are asymmetric error bars.
The resulting plots are shown in figure 4.13 and 4.15. This new procedure largely improves the stability of the fits. For example, the previous outliers in the seventh p_T^{avg} bin are now showing a smoother behaviour as shown in figure 4.14. Also, the statistical fluctuations in high $|\eta|$ regions disappear except for one big outlier in the highest $|\eta|$ bin caused by a bad fit. Even though, this outlier has large enough uncertainties, within which it is comparable to the trend outlined by the rest of the points.
The statistic uncertainties of the Gaussian fits are mostly negligible. The visible errors in the ratio plots are systematic uncertainties. These systematic errors dominate over the statistical errors from the Gaussian fits with very few exception.

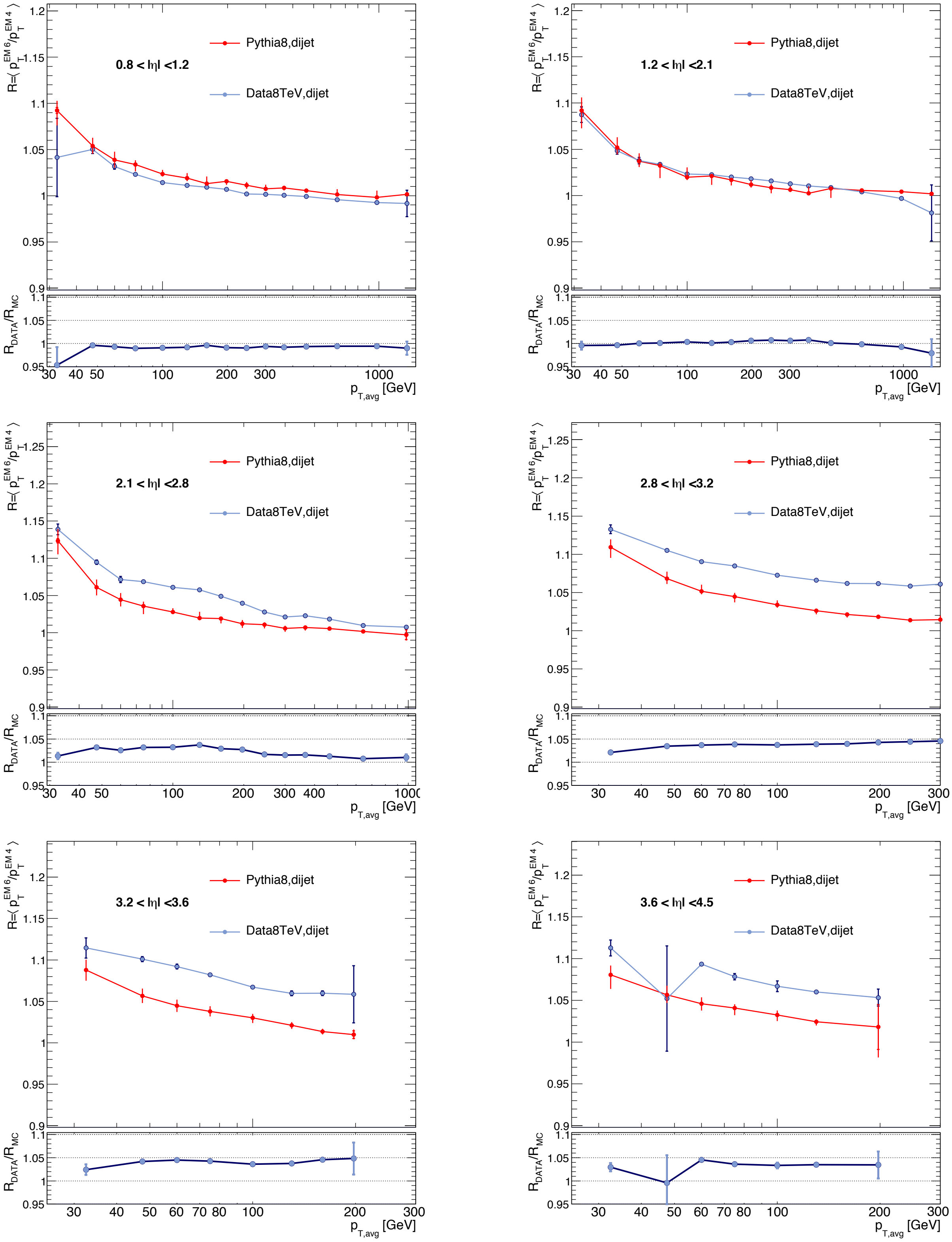


Figure 4.15: R_{MC}, R_{DATA} and R_{final} plotted against p_T^{avg} in different η-slices.

4.2.5 Mapping to p_T^{ref}

The binning in p_T^{ref} is more appropriate for the R-scan calibration than binning in p_T^{avg}. For each p_T^{avg} slice there is a p_T^{ref}-η distribution as shown in figure 4.16, which allows to determine p_T^{ref}.

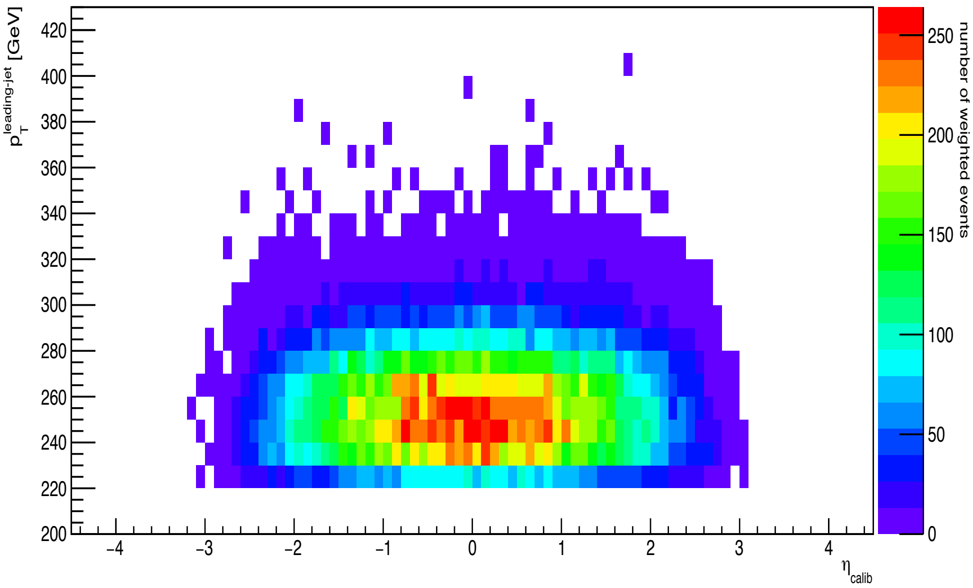


Figure 4.16: p_T^{ref} *distribution as function of* η *in the* p_T^{avg} *bin* $220\,\mathrm{GeV} < p_T^{avg} < 270\,\mathrm{GeV}$.

The approach is the same as the one applied to determine the data and MC ratios: For each p_T^{avg} bin the 2-D histogram is projected in the different η bins as shown in figure 4.17. In order to determine the p_T^{ref}, the mean value and its error of the projected histograms are taken. The mapping independently shifts the data and MC distributions and the double ratio cannot be built directly any more. This behaviour is shown in figure 4.18. The left plot shows the unshifted case where the ratio depends on p_T^{avg}. At the left plot, a global shift to lower p_T values is visible as well as the independent shift of MC and data. Before, the double ratio was calculated point by point because the points were always in the middle of the p_T^{avg} bin.
A global fit of the whole ratio is very difficult because of the large p_T value in contrast to the small difference in the ratios (R_{MC} and R_{DATA}). A polynomial of a higher order would describe the shape of the plots, but the higher orders have to be rather suppressed because of this tiny y-range. Even very complex functions which suppress higher orders were tested and they resulted very unstable. For this reason in this analysis an interpolation is used. Considering the curvature of the ratios, either a linear or a quadratic interpolation/extrapolation is appropriated. Afterwards it is tested which of the two interpolations is the best.

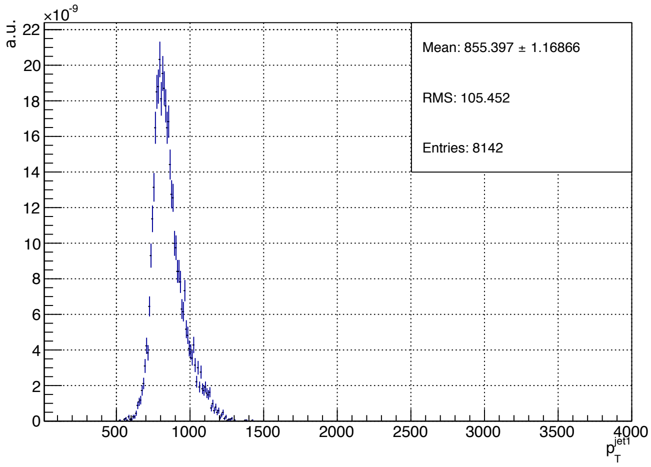


Figure 4.17: Projected p_T^{ref} distribution as function of η in the p_T^{avg} bin: 760 GeV $< p_T^{avg} <$ 1200 GeV.

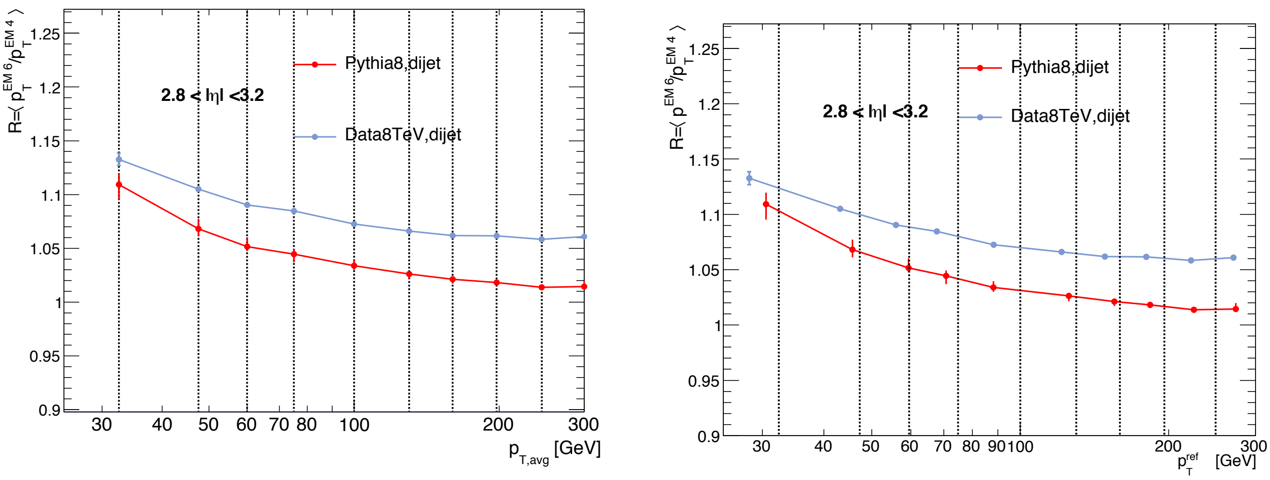


Figure 4.18: *The left plot shows the ratios depending on p_T^{avg}. Every MC and data point has the same p_T value highlighted with the vertical dashed lines. The right plot shows the shifted MC and data ratio. The vertical dashed lines are at the same position as they are in the left plot. It is visible that all points are shifted to lower p_T values, but data and MC are shifted independently.*

4.2.5.1 Linear Interpolation

The linear interpolation provides a very simple method to determine the shifted or missing points. The linear equation is given as:

$$f(x) = a \cdot x + b.$$

The first order polynomial has two degrees of freedom and needs therefore at least two points to determine its parameters. The interpolation is always done between two neighbouring points. Therefore it is assumed that the curvature is linear section by section. The solution of the equation system:

$$\begin{aligned} \mathrm{I}: &\quad ax_1 + b = y_1 \\ \mathrm{II}: &\quad ax_2 + b = y_2 \end{aligned}$$

delivers the slope a and the y-offset b:

$$\begin{aligned} a &= \frac{y_1 - y_2}{x_1 - x_2} \\ b &= y_1 - \frac{y_1 - y_2}{x_1 - x_2} \cdot x_1 = y_1 - a \cdot x_1. \end{aligned}$$

Considering the results of the equation system, the linear equation yields:

$$y = a \cdot x + b = \left(\frac{y_1 - y_2}{x_1 - x_2} \right) (x - x_1) + y_1. \tag{4.18}$$

This equation defines the interpolated points between two neighbour points. However, if the last point or the first point has to be interpolated, there are no two neighbouring points. In this case an extrapolation is done with the two right points if it is the first point, or with the two left points if it is the last point. At the beginning the two first points are taken to determine the parameters for equation (4.18) to extrapolate the data. The extrapolation at the end is done with the two last points.
In the figures 4.19 and 4.20 the interpolation and extrapolation are well visible but mostly there are no big shifts.

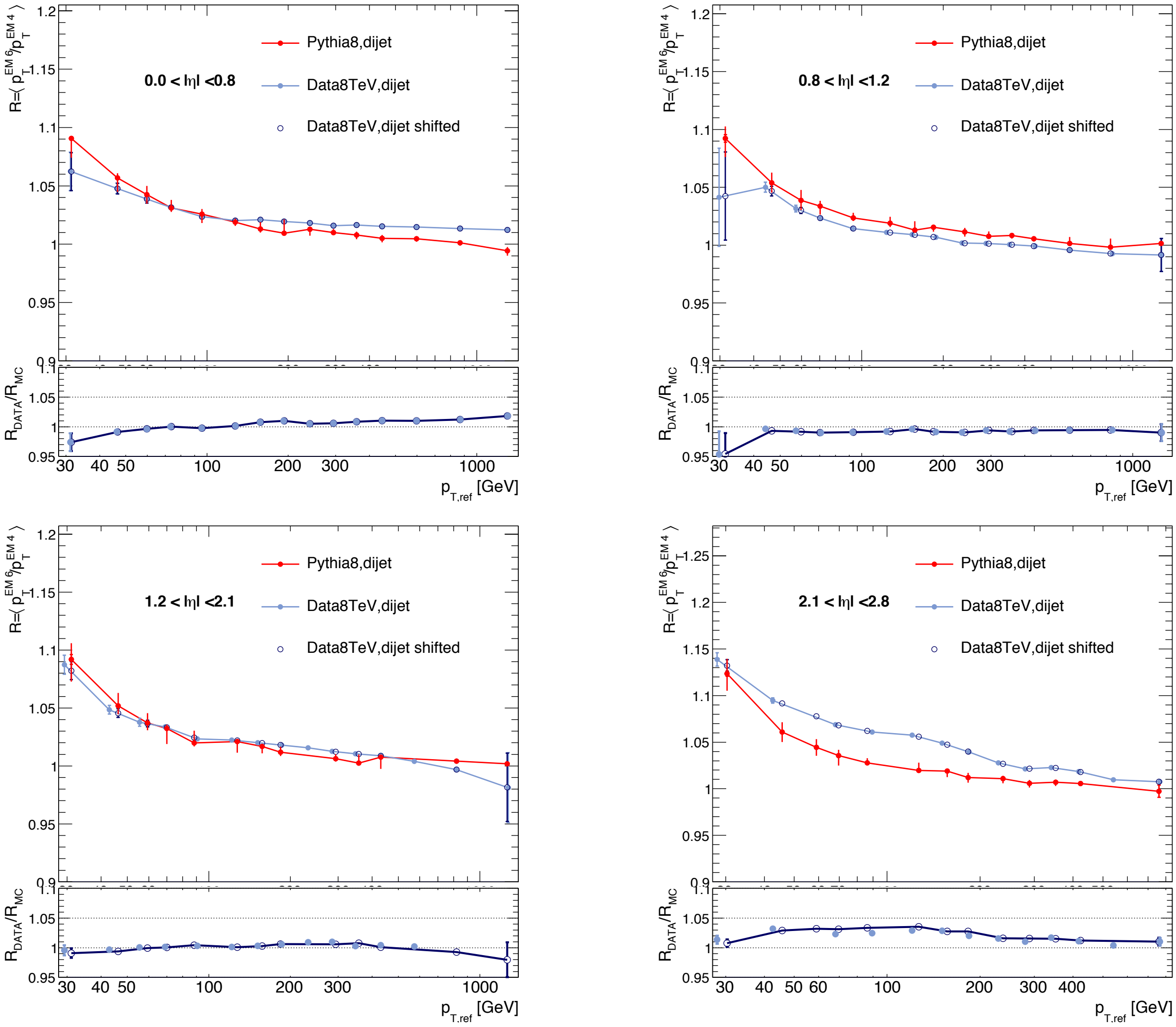


Figure 4.19: *MC and data ratios and the* R_{final} *for the central* η *regions, plotted against* p_T^{ref} *in different* η*-slices.*

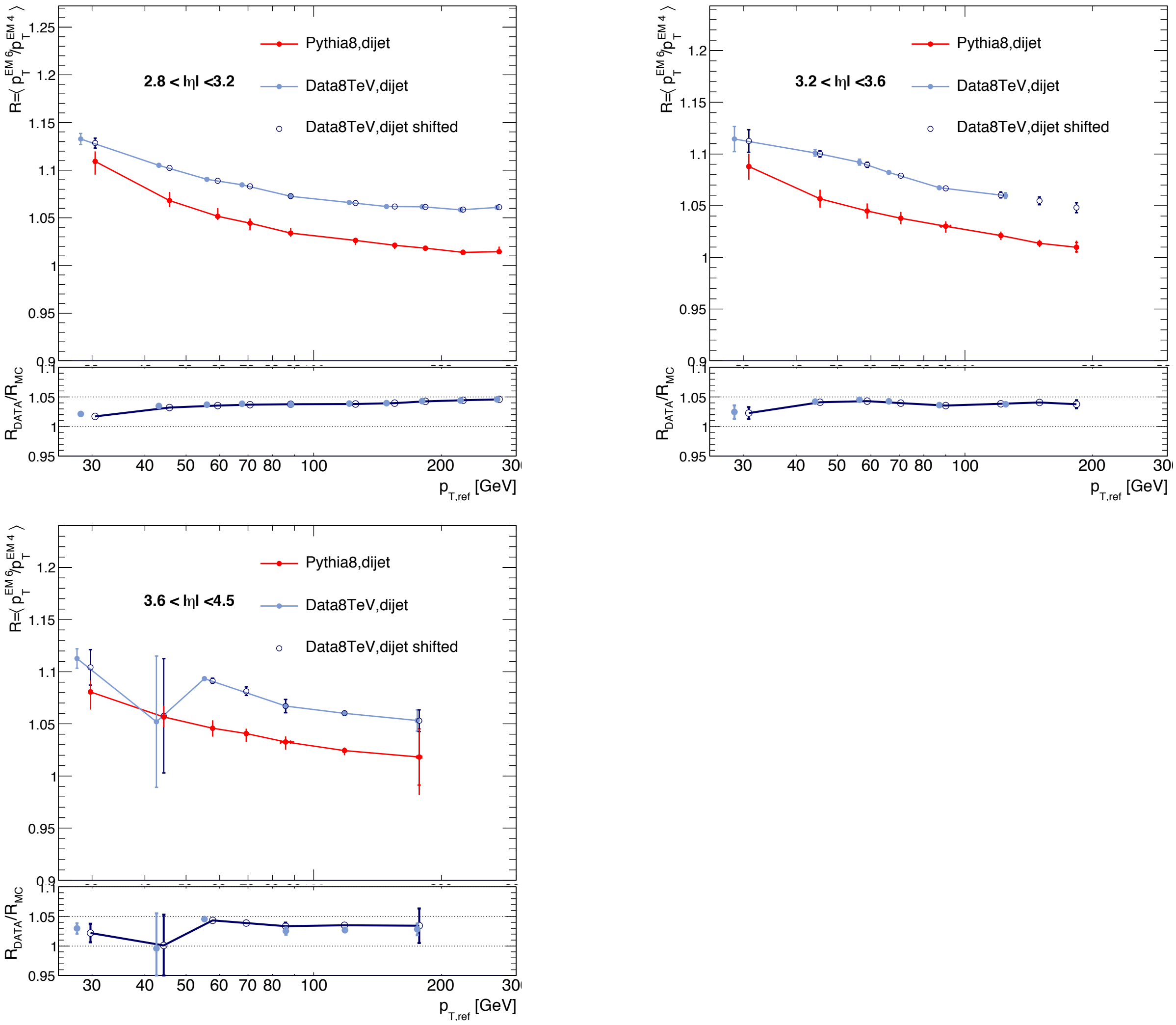


Figure 4.20: *MC and data ratios and the R_{final} for the forward η regions, plotted against p_T^{ref} in different η-slices.*

4.2.5.2 Quadratic Interpolation

A quadratic interpolation is also tested. The approach is the same as for the linear interpolation. The quadratic function is given as:

$$f(x) = a \cdot x^2 + b \cdot x + c.$$

A quadratic polynomial needs at least three points to determine all its parameters. Therefore always three neighbouring points are used to interpolate and extrapolate.
Again, an equation system:

$$\mathrm{I}: \quad ax_1^2 + bx_1 + c = y_1 \tag{4.19}$$

$$\mathrm{II}: \quad ax_2^2 + bx_2 + c = y_2 \tag{4.20}$$

$$\mathrm{III}: \quad ax_3^2 + bx_3 + c = y_3 \tag{4.21}$$

is solved and the solution returns the parameters a, b and c:

$$a = \frac{1}{x_1 - x_3}\left(\frac{y_1 - y_2}{x_1 - x_2} - \frac{y_2 - y_3}{x_2 - x_3}\right)$$

$$b = \frac{y_2 - y_3}{x_2 - x_3} - a(x_2 + x_3)$$

$$c = y_3 - ax_3^2 - bx_3.$$

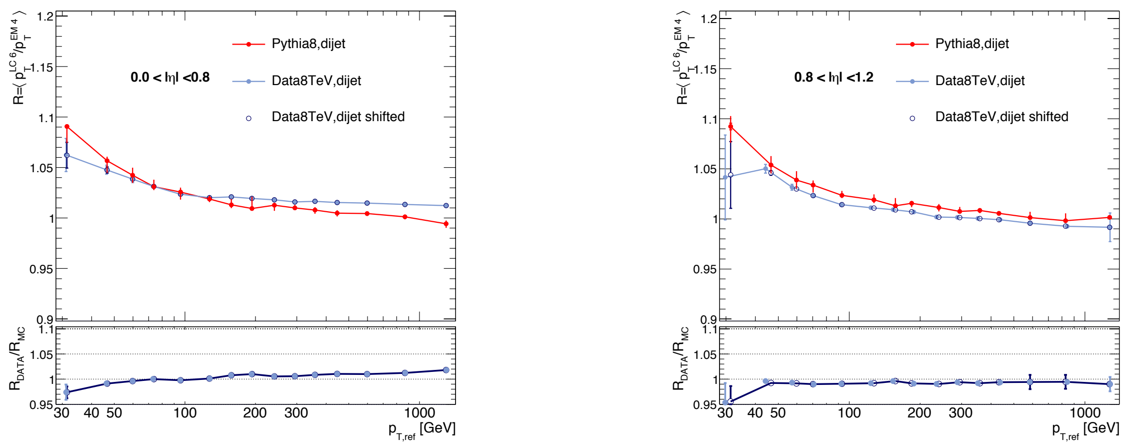


Figure 4.21: *MC and data ratio and the R_{final} plotted against p_T^{ref} in the η-slice $0 < |\eta| < 0.8$ (left plot) and $0.8 < |\eta| < 1.2$ (right plot).*

If the interpolation is done with three points, there is an ambiguity concerning the choice which three points should be taken. In this analysis the middle point of the interpolation function (corresponding to equation (4.20)) is chosen so that the point which has to be interpolated is as near as possible to this middle point. After the interpolation the ratios can easily be determined.

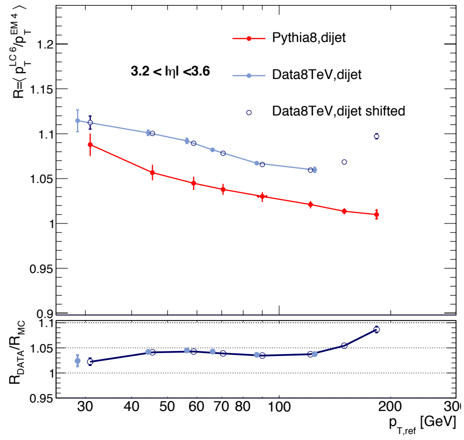


Figure 4.22: *MC and data ratio and the* R_{final} *plotted against* p_T^{ref} *in the* η*-slice* $3.2 < |\eta| < 3.6$

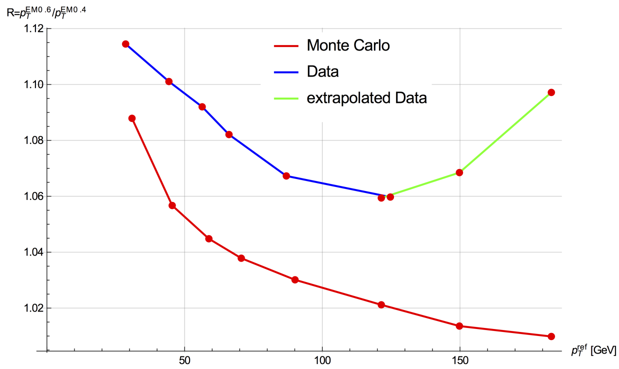


Figure 4.23: *MC and data ratio plotted against* p_T^{ref} *on a linear scale in the* η*-slice* $3.2 < |\eta| < 3.6$. *The quadratic interpolation is shown in green.*

The quadratic interpolation reflects the curve shape very well as shown in the figure 4.21. But in figure 4.22 there is a huge deviation from the expected shape given by the extrapolation. Additionally, in figure 4.23 this behaviour is plotted on a linear scale. The quadratic extrapolation is very sensitive to small fluctuations.

4.2.5.3 Comparison of linear and quadratic interpolation

In the following, the two methods are compared to decide which is the best for the R-scan analysis. The interpolations are done simultaneously as described in section 4.2.5.1 and 4.2.5.2. The direct comparisons of the two interpolation methods are shown in figure 4.24. The R_{MC} and R_{DATA} are the same ratios as described in the previous analysis. In the second pad, there is the double ratio based on the quadratic interpolation.
Furthermore, the bottom pad shows the triple ratio:

$$R_{comp} = \frac{R_{final}^{quadratic}(p_T^{avg}, \eta)}{R_{final}^{linear}(p_T^{avg}, \eta)}. \tag{4.22}$$

If the linear and quadratic interpolation are approximately equivalent, the R_{comp} ratio should be 1. In every $|\eta|$ slice the difference between the two interpolations is less than 0.1% with the exception of the one quadratic extrapolation already mentioned above (see figure 4.21 on the bottom right plot and 4.23). In addition, the direct comparison between the linear and quadratic extrapolation, shown in figure 4.24 (on the bottom right plot), confirms a not negligible difference.

It is recognisable that the linear extrapolation (see figure 4.19) describes the expected trend better than the quadratic extrapolation. By comparing the behaviour of the other $|\eta|$ slices in high p_T there is always a decrease and therefore the increase of the data ratio R_{DATA} in this case is not justified. As a consequence of the high sensibility to little fluctuations for an extrapolation, the linear interpolation is the better choice for the p_T mapping.

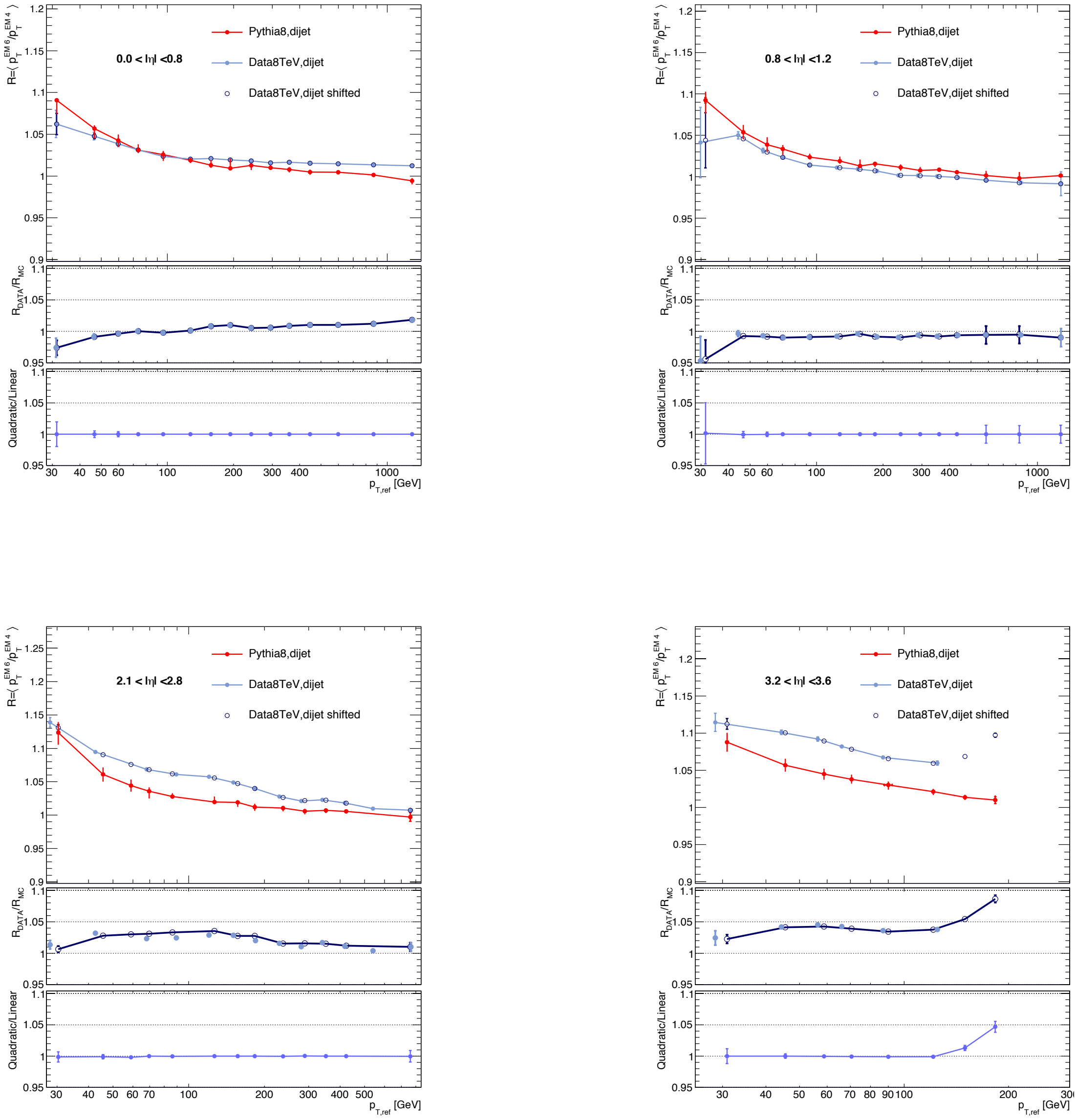


Figure 4.24: *The upper pad shows the MC and data ratios. In the middle, the* R_{final} *is plotted considering quadratic interpolation. And in the bottom pad: the comparison of the double ratio built once with the quadratic interpolation and once with the quadratic interpolation. They are plotted against* p_T^{ref} *in different* η*-slices.*

4.2.6 Validation

As already mentioned above, for both jet collections which are used in this analysis a full in-situ calibration exists. Therefore it is possible to compare the results of the R-scan method with the fully calibrated results. It is expected that if the in-situ calibrations are applied to both data jet collections, the R_{MC} and R_{DATA} curves should be approximately the same. The analysis is the same as above except that the probe collection is fully calibrated including in-situ corrections.

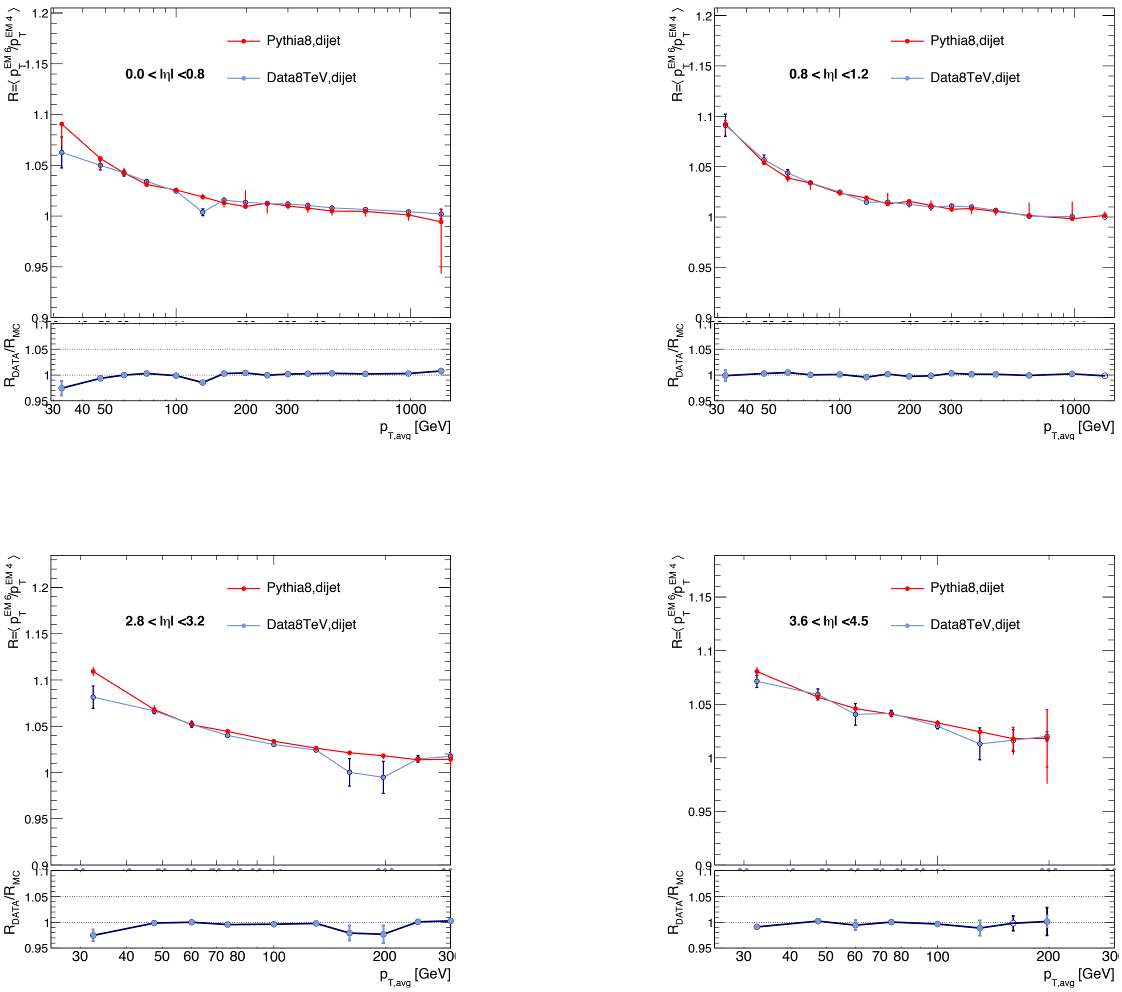


Figure 4.25: *MC and data ratio of two fully calibrated jet collections and the R_{final} plotted against p_T^{avg} in different $|\eta|$ slices.*

The resulting plots are presented in the figure 4.25, depending on p_T^{avg} and $|\eta|$. In this case it is not necessary to map back to p_T^{ref} because the final ratio is not needed to calibrate the probe jet. There are small deviations visible, but they are randomly distributed and they correspond to points with large uncertainties. The double ratio R_{DATA}/RMC is almost one except for the already mentioned fluctuations. In addition, there is no difference between forward and central region recognisable as well as no differences between any p_T regions. Consequently, the R-scan method is able to determine inter - calibration factors which correspond nicely to the in-situ corrections.

5 Applying the R-scan Method to the new Data from LHC Run II

The new run of the LHC provides a centre of mass energy up to 13 TeV. In the following, preliminary results based on the first data acquired in 2015 with $\sqrt{s} = 13\,\mathrm{TeV}$ are presented.
At the moment there is only one fully calibrated jet collection available, including in-situ corrections: Anti-k_t4LCTopo, where the jets are clustered with the anti-k_t algorithm with the radii $R = 0.4$ and calibrated with local cell signal weighting method (LCW). The LCW is used because most analyses of the new data are based on this energy scale (described in section 2.3.3.2). Furthermore, the probe jets (Anti-k_t6LCTopo) are not calibrated at all, neither are their MC samples. There are no other samples available yet. Therefore, the quality of the reference collection is checked, and the analysis is improved to be able to manage the new data format. Additionally, in order to test some preliminary results of the R-scan method, an analysis of the current available data is shown. The analysed data corresponds to an overall luminosity of $78.35\,\mathrm{pb}^{-1}$.

5.1 Changes with respect to the Analysis of Run I

Each LHC run has its specific peculiarities. The most relevant change for this analysis is the new trigger management as described in section 3.3. The new high level triggers are shown in table 5.1. For example "HLT_j15_320eta490" means that it is a high level trigger (HLT), j15 describes the online threshold (there is no turn-on considered) and 320eta490 defines the $|\eta|$ range of the forward trigger which is $3.2 < |\eta| < 4.9$. They are also single jet triggers and divided into central and forward triggers. There are less triggers than in the previous data analysis and also the trigger thresholds are different. The thresholds correspond again to a trigger efficiency of 99% and the p_T binning is adjusted to it.

The event selection is almost the same as in the previous analysis, but the JVF cut is now replaced by a cut on the jet vertex track (JVT). The JVT has been proved to be more efficient against pile-up than JVF. The JVT is a variable which indicates how likely it is that the jet comes from the primary vertex. If the jet p_T is smaller than 50 GeV and the jet $|\eta| < 2.4$, then this condition has to be fulfilled:

$$JVT > 0.64. \tag{5.1}$$

P_T^{avg} bin [GeV]	Central trigger	Forward trigger
25 - 40	HLT15	HLT_j15_320eta490
40 - 55	HLT25	HLT_j25_320eta490
55 - 85	HLT35	HLT_j35_320eta490
85 - 90	HLT55	HLT_j55_320eta490
90 - 115	HLT60	HLT_j60_320eta490
115 - 145	HLT85	HLT_j85_320eta490
145 - 220	HLT110	HLT_j110_320eta490
220 - 330	HLT175	HLT_j175_320eta490
330 -	HLT260	HLT_j260_320eta490

Table 5.1: The trigger list of the used triggers in this analysis for the data of run II. They are divided in central and forward triggers and each trigger has its corrsponding p_T slice.

5.2 Data Quality

Data of run 2 are very new and not yet well understood. In addition, only the reference collection is thoroughly checked because it does not make any sense to check a collection which is not calibrated at all. The variables which are used for the event selection (see section 4.2.2) are analysed to test the quality of the data.

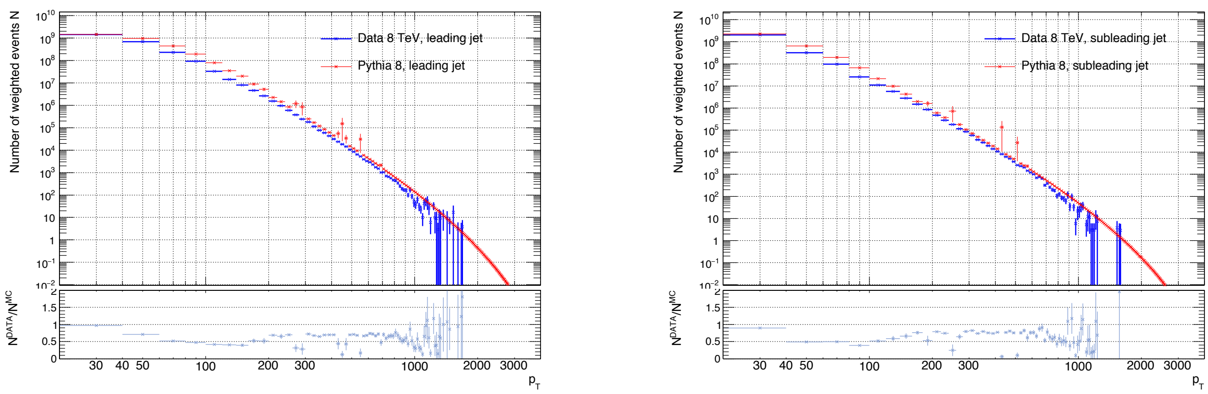


Figure 5.1: *The p_T distribution of the leading reference-jet on the left and the subleading reference-jet on the right: In the upper pad, the red points are from MC and the blue points are from data. The bottom pad shows the ratio of MC and data.*

Figure 5.1 shows the p_T distribution of the leading and subleading reference-jet after the JVT cut. The curve follows the expected exponential p_T distribution. In the plot showing the data/MC ratio, there are several steps visible in different p_T regions. The ratio is not completely flat. It varies a lot in the first few bins: from a value close to 1 in the first

bin to about 0.5 in the third bin. From 60 GeV to ∼150 GeV, the ratio is approximately constant. The region between 300 GeV and 900 GeV has also a nearly flat ratio but it is higher than in the previous region. At higher p_T, the statistic is too low for a comparison. The η distributions of the leading reference-jet and the subleading reference-jet are shown in the figure 5.2. In the central region there is a very flat ratio between MC and data but MC and data differ by a factor about 1.5. Between the central and the forward region there is a step visible ($\sim |\eta| = 3.4$). Additionally, in the forward region there are some fluctuations in the ratio, mainly caused by statistics.
These steps in the p_T and η ratios can indicate that there is a trigger dependency. Either the trigger weight or the pile-up weight have not been correctly applied yet. If the trigger weight were responsible for this behaviour, the p_T distribution of the data would not be as smooth as it is shown in figure 5.1. Therefore a possible option is that it is an impact of the applied pile-up weight.

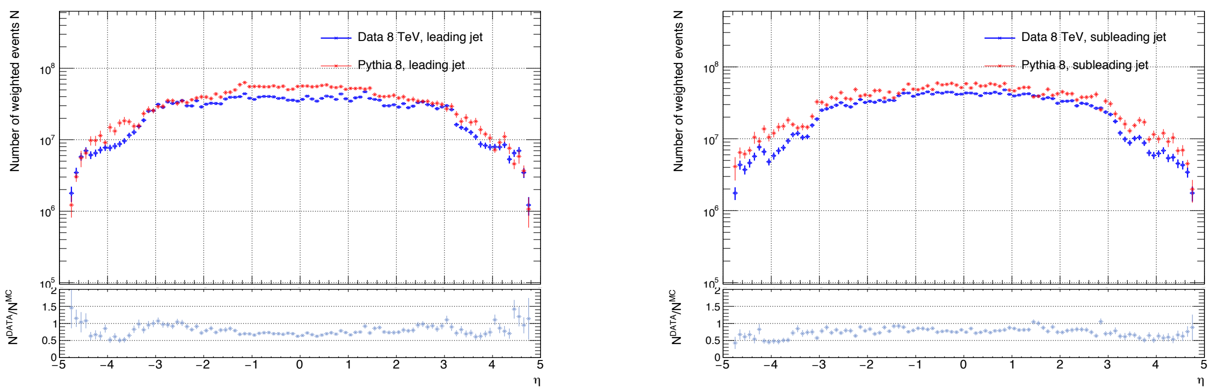


Figure 5.2: *The η distribution of the leading reference-jet on the left and the subleading reference-jet on the right: In the upper pad, the red points are from MC and the blue points are from data. The undermost pad shows the ratio of MC and data.*

Figure 5.3 shows the distribution of primary vertices (NPV). The ratio of MC and data shows a very abnormal shape. Normally the pile-up weights are applied to MC so that the ratio is equal to one. This confirms the behaviour of the previous distributions. It is probably related to a mismodelling of the pile - up in the simulation not properly corrected to reproduce the observed data. Therefore the pile-up weights have to be adapted to this behaviour.

In order to consider the impact of the pile-up weight, the cut on JVT in the low p_T region is applied. The figure 5.4 shows the JVT distributions. In the subleading distribution the cut is nicely visible at JVT=0.64.
The impact of the JVT cut is recognisable by comparing the p_T distribution with a JVT cut applied on the first three jets (see figure 5.1) and without this cut to the third jet as

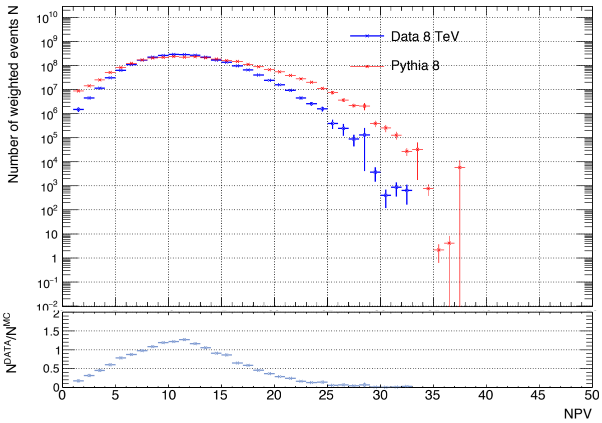


Figure 5.3: *The NPV distribution: In the upper pad, the red points are from MC and the blue points are from data. The bottom pad shows the ratio of MC and data.*

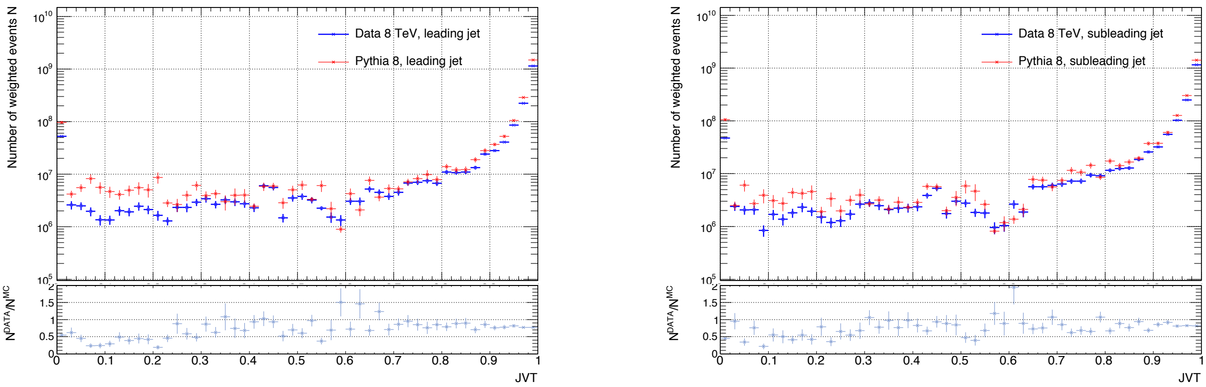


Figure 5.4: *The JVT distribution of the leading reference-jet on the left and the subleading reference-jet on the right: In the upper pad, the red points are from MC and the blue points are from data. The undermost pad shows the ratio of MC and data.*

shown in figure 5.5. Without the cut on the third jet, data and MC differ by a factor 2, whereas with the cut on the third jet the ratio is roughly one. Therefore the JVT cut suppresses the impact of the pile - up weight to a negligible level.
Consequently, the current data is not sufficient for using it for the R-scan calibration. It is expected that the R-scan method cannot deliver optimal results when using these data samples.

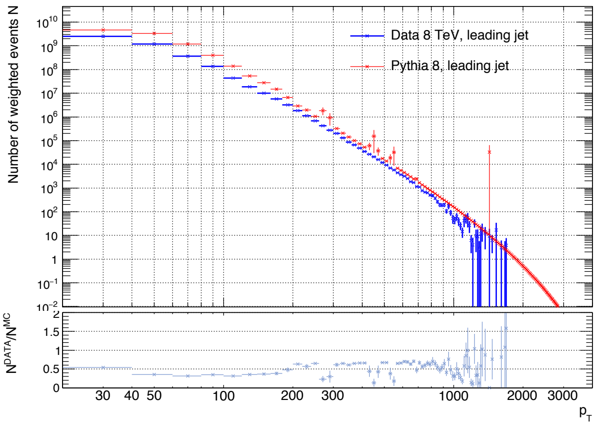


Figure 5.5: *The p_T distribution of the leading reference-jet without a cut on JVT of the third jet. The bottom pad shows the ratio of MC and data.*

5.3 Preliminary Results of the LHC Run II

The approach is the same as for the previous analysis in section 4.2. After the event selection and the matching, the ratios between the reference and R-scan jet (see section 4.2.3) are mapped as function of p_T^{avg} and η as shown in figure 5.6. By comparing this ratio distribution with a previous one, for example in figure 4.4, it is evident that this distribution is very asymmetric. This effect can be seen better after the splitting into different $|\eta|$ bins.
The 2-D distributions are divided into different $|\eta|$ regions by projecting the 2-D histograms. The resulting distributions are shown in the figures 5.7 and 5.8. As expected, they are very asymmetric. Additionally, in figure 5.8 there is a second peak in the tail. These effects influence the further analysis significantly. This asymmetry can be caused by the missing calibration of the R-scan jets and also because of differences in MC and data of the variables of the reference jets as described above.

The resulting ratio plots are, as expected, not significant for a calibration (see figure 5.9). The "best" result can be seen in the lowest $|\eta|$ region (figure 5.9, left plot) because the impact of the calibration is not as big in the central region. The global trend of the MC ratio (R_{MC}) corresponds roughly to the one observed in the previous data analysis. The data ratio seems to be roughly constant in the first $|\eta|$ region. But with higher $|\eta|$ it develops a slope. All in all, there is no significant trend visible and the statistic uncertainties are very large (systematics are not yet implemented in MC of the new data). Therefore, the current status of the data samples does not allow a a calibration with the R-scan method.

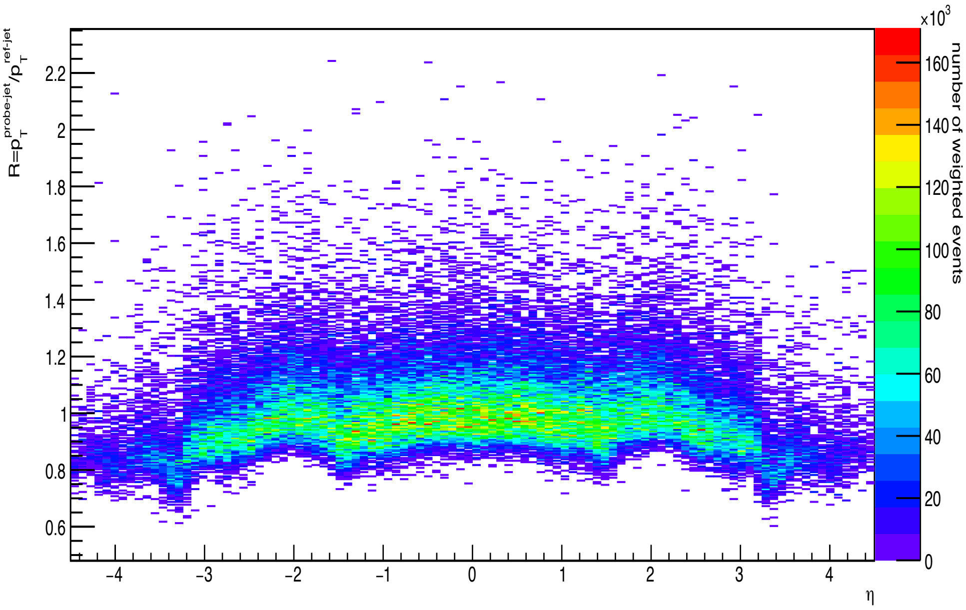


Figure 5.6: *Ratio plot of the new data mapped as function of η in the p_T^{avg} region: 55 GeV - 85 GeV*

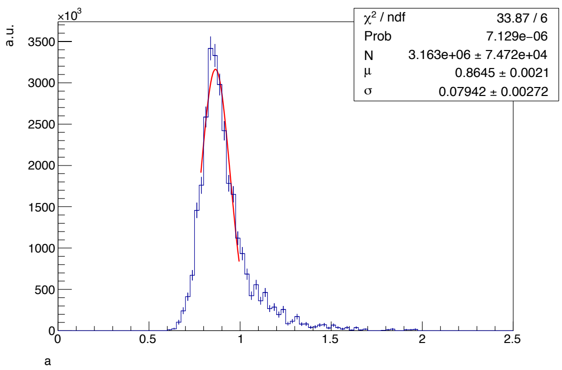


Figure 5.7: *Data ratio distribution, new data:* $1.2 \leq |\eta| < 2.1$ *and* $55\,\mathrm{GeV} < p_T^{avg} < 85\,\mathrm{GeV}$

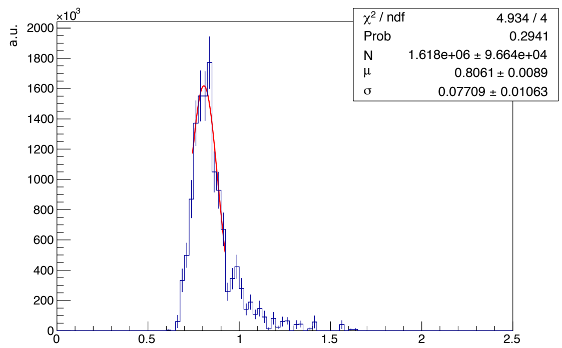


Figure 5.8: *Data ratio distribution, new data:* $3.2 \leq |\eta| < 3.6$ *and* $40\,\mathrm{GeV} < p_T^{avg} < 55\,\mathrm{GeV}$

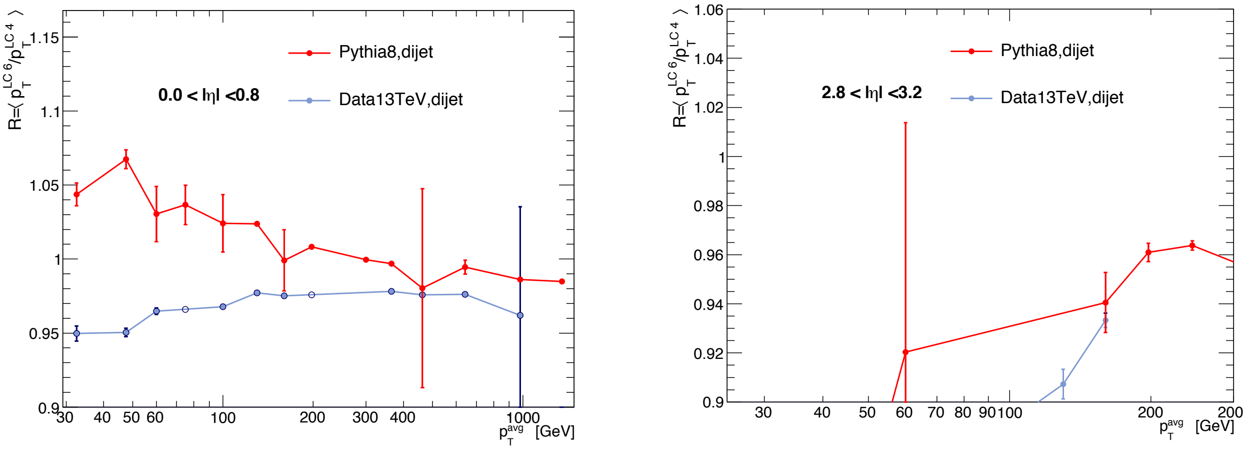


Figure 5.9: *MC and data ratio plotted against p_T^{ref} in the η-slice $0 < |\eta| < 0.8$ (left plot) and $2.8 < |\eta| < 3.2$ (right plot)*

6 Conclusion

The progress in modern particle physics is lead by large particle accelerators, like the Large Hadron Collier (LHC) at CERN. After a successful first run of the LHC including the discovery of the Higgs Boson, the second run has already started. It is a new challenge for experiments like ATLAS to handle the new energy scale of 13 TeV and the increased luminosity. Besides hardware modifications, the analysis has to be adjusted too.
In order to be prepared for the new LHC run, the new analysis methods have been improved. One big task is to provide calibrations for each analysis. There are many analyses which are interested in differently clustered jets, that means for example different jet radii. However, it is not possible to provide in-situ corrections for each jet definition.
A new option is the R-scan calibration. This method is tested and improved with the data acquired during Run I of the LHC. It is a calibration technique which is based on di-jet events which avoids an ad-hoc in-situ calibration for each differently clustered jet collection. One fully calibrated jet collection is used as reference to calibrate jets with missing in-situ corrections. The MC ratio of these two collections considers the differences in their geometry and the data ratio takes into account of the differences arising from the in-situ calibration. These two ratios are mapped as a function of p_T^{avg} and η. This fact allows a splitting in the different p_T^{avg} and $|\eta|$ regions since the correction factor depends on them. The resulting ratio distributions are fitted with a Gaussian function and the results are plotted against p_T^{avg} in different $|\eta|$ slices. The resulting trend indicates that the ratios decrease with p_T. In order to consider the geometry of the jet and the different calibration, the final double ratio between the MC ratio and the data ratio provides the correction factors. The double ratio in the central region is roughly one whereas in the forward region, it is about 5% higher.
Due to the fact that the correction factor depends on p_T^{avg} and the binning in p_T^{ref} is more appropriate for the R-scan method, a mapping to p_T^{ref} has to be done. However, the mapping shifts the MC and data points independently. Therefore, the double ratio cannot be directly calculated anymore and an interpolation is necessary. A quadratic and a linear interpolation are tested. They deliver almost the same results except for the extrapolation. For this reason, the linear interpolation is the better choice.
In addition, there are two fully calibrated jet collections and therefore it is possible to cross-check the results. The two jets only differ in their geometry, which is compensated in the analysis. The final double ratio is compatible across the whole p_T and η range. Consequently, the R-scan method is able to determine in-situ correction factors which are compatible to the official in-situ corrections.

Additionally, the R-scan method is applied to the new data from LHC run II. There is

only one calibrated jet collection available. Since the R-scan method takes two calibrated jet collections which only differ in the in-situ calibration, the current status of the data is not sufficient for the R-scan calibration. Therefore the data quality has been tested. As a result of the quality check, there is a discrepancy in the calibrated jets between MC and data. After checking the variables p_T, η and JVT of the leading and subleading jets of the reference collection and NPV, a possible explanation for the discrepancies is the presence of a mismodeling of the pile-up.
As expected, the results of the R-scan calibration with this data set are not satisfying. A next step would be to improve the event selection and the MC modelling.
The R-Scan method has been successfully tested and it is now a reliable and quick method to determine the in-situ calibration for a wide variety of jet collections used by all the analyses of the ATLAS collaboration. The new early data have been already analysed and soon the R-Scan analysis will be fully optimised for the new data acquisition conditions, providing a first set of quick and reliable calibration constants.

A Bibliography

[1] Georges Aad et al.
Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.
Phys.Lett., B716:1–29, 2012.

[2] F. Abe and Akimoto et al.
Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab.
Phys. Rev. Lett., 74:2626–2631, Apr 1995.

[3] Bruno R. Stella and Hans-Jurgen Meyer.
Y(9.46 GeV) and the gluon discovery (a critical recollection of PLUTO results).
Eur. Phys. J., H36:203–243, 2011.

[4] G. Arnison et al.
Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 gev.
Physics Letters B, 122(1):103 – 116, 1983.

[5] Particle data group.
Official website: http://pdg.lbl.gov/.

[6] Sheldon L. Glashow.
Partial-symmetries of weak interactions.
Nuclear Physics, 22(4):579 – 588, 1961.

[7] F. Englert and R. Brout.
Broken symmetry and the mass of gauge vector mesons.
Phys. Rev. Lett., 13:321–323, Aug 1964.

[8] Peter W. Higgs.
Broken symmetries and the masses of gauge bosons.
Phys. Rev. Lett., 13:508–509, Oct 1964.

[9] Abdus Salam.
Gauge unification of fundamental forces.
Rev. Mod. Phys., 52:525–538, Jul 1980.

[10] Steven Weinberg.
A model of leptons.
Phys. Rev. Lett., 19:1264–1266, Nov 1967.

[11] D. Griffiths.
Introduction to Elementary Particles.
Physics textbook. Wiley, 2008.

[12] Serguei Chatrchyan et al.
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC.
Phys. Lett., B716:30–61, 2012.

[13] Richard D. Ball et al.
Parton distributions with {LHC} data.
Nuclear Physics B, 867(2):244 – 289, 2013.

[14] https://mstwpdf.hepforge.org/plots/mstw2008nlo68cl_allpdfs.png.

[15] Chris Blanks.
Strangeness at lhcb.
http://imperialhep.blogspot.de/2011/08/strangeness-at-lhcb.html, August 2011.

[16] http://www.isgtw.org/sites/default/files/img_2014/sherpa_sim_0.jpeg.

[17] Prof. Karl Jakobs.
Skript – experimentalphysik v - kern- und teilchenphysik.
WS 14/15.

[18] ATLAS Collaboration.
Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV.
Technical Report ATL-COM-PHYS-2015-590, CERN, Geneva, Jun 2015.
https://cds.cern.ch/record/2028594.

[19] Stephen D. Ellis and Davison E. Soper.
Successive combination jet algorithm for hadron collisions.
Phys. Rev., D48:3160–3166, 1993.

[20] M. Wobisch and T. Wengler.
Hadronization corrections to jet cross-sections in deep inelastic scattering.
In *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, 1998.

[21] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez.
The Anti-k(t) jet clustering algorithm.
JHEP, 04:063, 2008.

[22] Craig Sawyer, Christopher John Young, Cigdem Issever, and David Lopez Mateos.
Monte Carlo Calibration and Combination of In-situ Measurements of Jets in ATLAS.
Technical Report ATL-COM-PHYS-2015-071, CERN, Geneva, Jan 2015.

[23] Jet global sequential corrections with the ATLAS detector in proton-proton collisions at sqrt(s) = 8 TeV.

Technical Report ATLAS-CONF-2015-002, CERN, Geneva, Mar 2015.
https://cds.cern.ch/record/2001682.

[24] Cern - accelerating science.
Official webpage: http://home.web.cern.ch/.

[25] LHC Season 2: facts & figures.
http://press.web.cern.ch/backgrounders/lhc-season-2-facts-figures.

[26] Julie Haffner.
The CERN accelerator complex. Complexe des accélérateurs du CERN.
https://cds.cern.ch/record/1621894, Oct 2013.
General Photo.

[27] Joao Pequenao.
Computer generated image of the whole ATLAS detector.
https://cds.cern.ch/record/1095924, Mar 2008.

[28] The ATLAS Collaboration and G Aad et al.
The atlas experiment at the cern large hadron collider.
Journal of Instrumentation, 3(08):S08003, 2008.

[29] Alignment of the ATLAS Inner Detector with the initial LHC data at $\sqrt{s} = 13$ TeV.
Technical Report ATL-PHYS-PUB-2015-031, CERN, Geneva, Jul 2015.

[30] R (SLAC) et al. Bartoldus.
Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System.
Technical Report CERN-LHCC-2013-018. ATLAS-TDR-023, CERN, Geneva, Sep 2013.
Final version presented to December 2013 LHCC.

[31] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands.
A Brief Introduction to PYTHIA 8.1.
Comput. Phys. Commun., 178:852–867, 2008.

[32] Michaela Queitsch-Maitland.
Jet trigger emulation and efficiency study.
https://indico.cern.ch/event/247353/contribution/0/attachments/425112/590182/SingleJetTriggerOR_Effs2013_allPeriods.pdf, 15/04/2013.

[33] David W. Scott.
On optimal and data-based histograms.
Biometrika, 66(3):pp. 605–610, 1979.

List of Figures

List of Tables