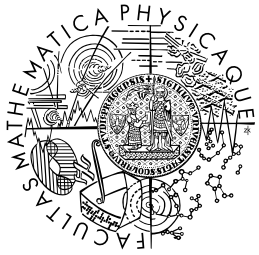

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Patrik Novotný

Exploring jet calibration with machine learning techniques

Institute of Particle and Nuclear Physics

Supervisor of the bachelor thesis: Mgr. Martin Rybář, Ph.D
Study programme: Physics
Study branch: General Physics

Prague 2021

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

It is difficult to express in words gratitude to my supervisor, Martin Rybář, for his enormous patience, infectious enthusiasm, and valuable advice. Undoubtedly, there are scientists from the HEP community, who taught me the skills needed to complete the thesis. Daniel Scheirich and Oldřich Kepka, who taught me to work with the ROOT framework. Zdeněk Doležal, who introduced me into the physics of detectors and accelerators. Lastly, Jiří Kroll, who does not hesitate to discuss numerous physical topics with me in a friendly manner. I have to also mention my family and fiancée Anna for their endless support, understanding, and care, thanks to which I can study physics in a difficult and changing world.

Title: Exploring jet calibration with machine learning techniques

Author: Patrik Novotný

Institute: Institute of Particle and Nuclear Physics

Supervisor: Mgr. Martin Rybář, Ph.D, Institute of Particle and Nuclear Physics

Abstract: Jets, collimated sprays of particles, are considered to be a perfect probe of the matter created at heavy-ion collisions. The work explores the possibility of using machine learning techniques to improve the overall calibration of jet energy scale and its resolution for ATLAS experiment calorimeters, which are used in the research of lead nucleus collisions. First, the current performance of jet reconstruction quantified by mean response and energy resolution dependence on the value of transverse momentum, pseudorapidity, and centrality of collisions is shown. It is further discussed that the current calibration does not take into account whether the jets are induced by a quark or gluon. Subsequently, four variables are selected, which might be used to distinguish between these two groups of jets. The dependence of the response on those quantities and on centrality is further studied. The last part describes the process of the preparation of a training data set, set up of the neural network, and the analysis of it using tools provided by the MultiLayerPerceptron library of the ROOT framework. The final result of the work is a neural network that improves the resolution of the response.

Keywords: ATLAS, heavy ions, jet energy calibration, neural networks

Název práce: Kalibrace jetů pomocí technik strojového učení

Autor: Patrik Novotný

Ústav: Ústav částicové a jaderné fyziky

Vedoucí práce: Mgr. Martin Rybář, Ph.D, Institute of Particle and Nuclear Physics

Abstrakt: Jety, kolimované spršky částic, jsou považované za ideální nástroj ke studiu hmoty, která vzniká v těžko-iontových srážkách. Práce se zabývá možnostmi využití technik strojové učení ke zlepšení celkové kalibrace odezvy jetů na experimentu ATLAS v CERN v rámci výzkumu srážek jader olova. Prvně je ukázána současná energetická škála a rozlišení jetů v závislosti na hodnotě příčné hybnosti, pseudorapidity a centrality srážek. Dále je ukázano, že současná kalibrace nezohledňuje, jsou-li jsou jety indukované kvarkem či gluonem. Následně je vytipována čtveřice proměnných, pomocí níž by mohlo být možné rozlišit mezi sebou tyto dvě skupiny jetů. Pro danou čtveřici jsou provedeny studie míry korelace s střední odezvou detektoru jakožto funkce centrality. V rámci poslední části je popsán proces přípravy souboru trénovacích dat, nastavení a analýzy neuronových sítí pomocí nástrojů, které nám poskytuje knihovna MultiLayerPerceptron frameworku ROOT. Finální výsledkem práce je nalezení sítě, která zlepšuje rozlišení odezvy.

Klíčová slova: ATLAS, těžké ionty, kalibrace jetů, neuronové sítě

Contents

Introduction

In the previous decade the existence of the quark-gluon plasma (QGP) under extreme conditions was experimentally proved [1]. Quark-gluon plasma is the unique state of matter predicted by quantum chromodynamics. The quark-gluon plasma is expected to strongly affect particles with a colored charge traversing through it, which gives us a way to investigate its properties. We are focused on the suppression of collimated sprays of particles, so-called jets. In order to accurately measure the properties of the quark-gluon plasma, it is, therefore, necessary to accurately measure the properties of jets.

In general, the reconstructed momentum or energy based on measurement in calorimeters differs from the momentum or the energy on a particle level. Currently, a correction is applied to the measured data, which ensures that the reconstructed energy corresponds on average to the particle level. However, the finite jet energy resolution is affected by both, detector performance and properties of jets. This work aims to try to find a correction using machine learning methods, which will improve the jet energy resolution using the additional information about the jet origin and structure. The data from 5 TeV proton-proton collisions will serve as the foundation of the training sample. We do not practice training on lead-lead data in order not to obtain calibration dependent on centrality. Similarly, when identifying suitable variables for correction, we must perform a correlation study to exclude variables correlated with centrality.

The inspiration for the work was the article [2], where the authors used a neural network implemented within SciKit library, which corrected jet energy in 13 TeV proton-proton collisions. As mentioned above, we try to generalize their results for jets in heavy-ion collisions. Also, instead of the general-purpose SciKit library, we worked with tools defined in the TMultiLayerPerceptron class [3], which is integrated within the ROOT framework.

1. Experimental setup

1.1 ATLAS coordinate system

The origin of the coordinate system used by the ATLAS experiment is chosen as the nominal interaction point. The z-axis is taken along the beam direction. Plane x-y is transversal to the beam direction, thus it is also referred to it as to the transverse plane.

The direction of axes is defined as follows. The positive x-axis points from the origin of the coordinate system towards the center of LHC. The positive y-axis points upward to the surface level. The direction of the positive z-axis is therefore determined from the requirement of right-handedness of the coordinate system. The azimuthal angle ϕ is measured around the z-axis. A point located on the positive x-axis has $\phi = 0$ and ϕ increases clock-wise by definition. The polar angle θ is measured around x-axis. A point located on the beam axis has $\theta = 0$ and ϕ increases up to $\theta = \frac{\pi}{2}$, which belongs to a point located on the transverse plane.

Coordinate system with example measurement of momentum p and its projection into transversal plane p_{T} is shown in Fig 1.1.

1.2 Useful kinematic quantities at ATLAS

A common quantity [4] used to describe objects in relativistic physics is rapidity y. The convenience of using rapidity in relativistic physics is the fact that rapidities are additive in a one-dimensional case, unlike velocities or momenta. Rapidity is used very often in particle physics, inasmuch as particle velocities are close to the speed of light c.

At ATLAS, we are able to determine particle momentum components p_x, p_y and p_z from the particle track under the impact of magnetic field and energy E from energy deposited in the calorimetric system. According to [4] a rapidity could be determined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{1.1}$$

However, there are problems with rapidity determination. The higher the momentum is, the less accurately we can measure it. In addition to that, for high values of the rapidity z-component of the momentum is especially large in comparison to the other two components. Due to described problems in limit case[1] $E \approx p$ it is convenient to define new quantity named pseudorapidity η:

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right). \tag{1.2}$$

It is not obvious where formula 1.2 comes from. From figure 1.1 we can see, that z-component of momentum is related to the size of momentum by the following relation:

$$p_z = p \cos(\theta). \tag{1.3}$$

[1] The equivalent condition is $m \ll p$.

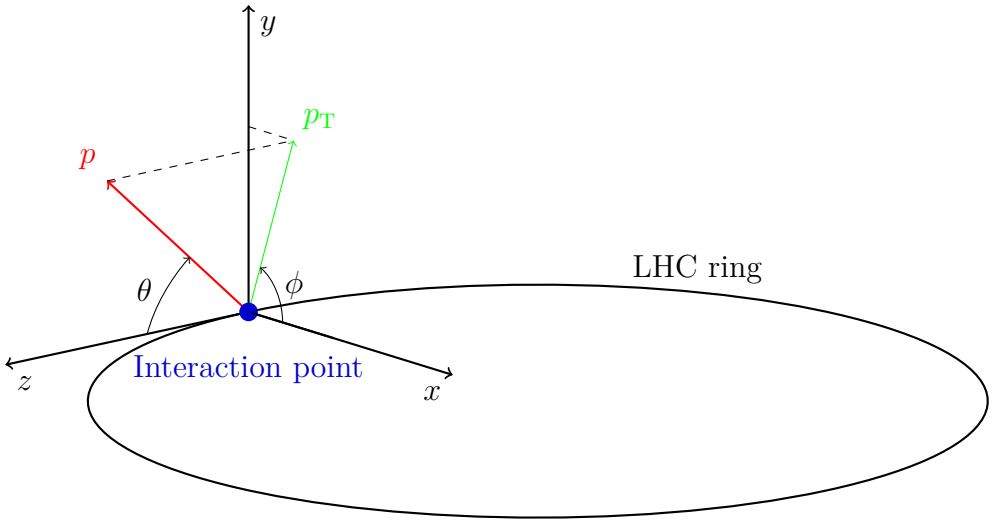

Figure 1.1: Coordinates system used at ATLAS experiment

Then formula 1.2 can be derived easily by using formulae 1.1 and 1.3 in assumed limit[2]:

$$y \approx \frac{1}{2}\ln\left(\frac{p+p_z}{p-p_z}\right) = \frac{1}{2}\ln\left(\frac{1+\cos(\theta)}{1-\cos(\theta)}\right) = \frac{1}{2}\ln\left(\frac{\cos^2(\frac{\theta}{2})}{\sin^2(\frac{\theta}{2})}\right) = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{1.4}$$

There are also the restrictions on the use of the pseudorapidity arising from used assumptions. Apart from the rapidity, the pseudorapidity is no longer additive in one-dimensional case and it can not give us any piece of information about the mass of particle.

Other useful quantities used in particle physics experiments are the transverse momentum p_T, the transverse energy E_T, which are defined as a projection of momentum p, respectively energy E into x-y plane. The transversal energy and momentum are mutually convertible by invariant mass of particle m:

$$m = \sqrt{E_T^2 - p_T^2}. \tag{1.5}$$

The invariant mass of a particle provides crucial information for particle identification. However, the invariant mass in HEP is not derived this way, inasmuch as it would be inapplicable. The reason comes from the fact that transverse momentum resolution σp_T^2 is directly proportional to its value p_T^2, which follows from the method of measurement from a sagitta curvature [5].

Last quantity often used as a veto in particle physics analyses or calibrations is the ΔR, which is the distance in $\eta - \phi$ space defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{1.6}$$

[2]The penultimate equality can be obtained by using the common-known double-angle formulae.

1.3 Collider physics quantities

A fundamental quantity used in collider physics for description of the collision of two particles is the centre of mass energy $\sqrt{s}$, which is prescribed by formula:

$$\sqrt{s} = \sqrt{(E_1 + E_2)^2 - (\vec{p_1} + \vec{p_2})^2}, \tag{1.7}$$

where lower indices denote a pertinence of quantity to a particle.

From 1.7 one can clearly see, why it is more advantageous to collide two beams of particles directly than to used fixed target. At least, in the question of achieved centre-of-mass energy $\sqrt{s}$. Nevertheless, other questions also come into play with regard to the studied phenomenon. Hence, accelerators and detectors at CERN differ strongly from each other.

The number of events per second generated in a collider is given by:

$$N_{\text{events}} = L\sigma_{\text{process}}, \tag{1.8}$$

where $\sigma_{process}$ is the cross-section of the studied process and L is the machine luminosity. Luminosity is the quantity, which can be written for a Gaussian beam distribution as:

$$L = \frac{N^2 n^2 f \gamma F}{4\pi\sigma_x\sigma_y}, \tag{1.9}$$

where N quantifies the number of particles in a bunch [3] and n is the number of colliding bunches. f describes revolution frequency, γ is Lorenz factor, and F is the geometric luminosity reduction factor. However, the description of its properties is beyond the scope of this work. More details can be found in [7]. Last two quantities requiring explanation are σ_x and σ_y, which represents physical cross-section along axes x and y. Furthermore, it is useful to define integrated luminosity:

$$\mathcal{L} = \int L dt, \tag{1.10}$$

where the integration is done with respect to time t.

1.4 Large Hadron Collider

Large Hadron Collider [8] has been the largest and highest-energy particle collider in the world since 2008. LHC was designed to be able to obtain center-of-mass energy $\sqrt{s} = 14$ TeV for protons and $\sqrt{s} = 5.5$ TeV per nucleon for lead ions. The collider was built and has been operated by the European Organization for Nuclear Research (CERN). A high number of experiments are underway at CERN. These experiments cover various topics in physics from dark matter and antimatter to quark-gluon plasma. Eight of the experiments are located on Large Hadron Collider. The group can be divided into a group of four big experiments - ALICE, ATLAS, CMS, and LHCb and four smaller experiments - TOTEM, LHCf, MoEDAL, and FASER. The rest of the experiments are associated with

[3] A bunch is a designation for a group of protons accelerated together in the accelerator, at LHC there was around 1.1 $\times 10^{11}$ protons in the bunch in year 2018. Beam parameters over time can be seen at [6]

smaller colliders - Super proton synchrotron (SPS) and Proton synchrotron (PS). The largest group of experiments associated with SPS and PS are fixed-target experiments, these experiments are located in so-called North Area. SPS and PS also gradually accelerate bunches before they are injected into the LHC.

LHC was built in the same tunnel, which was used by previous collider named Large electron-positron collider (LEP). The major motivation for upgrading to a hadron accelerator was lower energy losses through bremsstrahlung due to the significantly heavier mass of protons in comparison to electrons. In LEP, antiparticles circulated in the opposite directions due to the sign dependence of Lorentz force. This property can not be preserved at LHC. Thus, the major needed change in construction was adding a second tube for beam guidance.

All the colliders mentioned so far are called synchrotrons. The crucial issue in accelerator physics of circular accelerators is to ensure synchronisation while accelerating [9]. The trajectory of a particle is a path where the forces acting on the particle are equal. In our case, the centrifugal force and the Lorenz force. However, the centrifugal force increases with the energy of the particle. Furthermore, the circulation time is not constant. In order to ensure the same trajectory of the particle, the synchrotron increases the bending magnetic field with the energy of the particle. The energy limit delivered by synchrotron accelerator is typically dependent on the maximum strength of the magnetic field. This limit may be increased by using a synchrotron with a larger radius, which is why the LHC is as large, as well as the use of superconducting magnets, as is also the case of the LHC.

Let us describe LHC specifically [10]. There are 1232 dipole magnets for bending a particle path and 392 quadrupole magnets for focusing a beam, there are also used multipole magnets for squeezing bunches in order to obtain higher luminosity. The acceleration must also be performed in an ultra-high vacuum environment to prevent the interaction of ultra-relativistic particles with the surrounding environment. Such an interaction would blur the beam, reduction of its energy, and damage the accelerator system or detector. Acceleration itself takes place in so-called electromagnetic cavities. The frequency in them is also tuned to match the energy of the particles.

Since this work is dedicated to the topic, which belongs to heavy-ion physics, the process of acceleration of heavy ions will be explained. Both protons and ions are accelerated in LHC. However, the beginning of their journey is different. Protons are generated from hydrogen source. Hydrogen is ionised and accelerated firstly at linear collider LINAC 2 [4] to the energy of 50 MeV. The second step in the proton production chain is in PS Booster, where protons acquire kinetic energy equal to 1.4 GeV. Moreover, Booster stabilizes the beam so that more than 100 times more protons can be injected into the PS than without using it [12]. Let us denote, that not all protons are sent into the PS. Their paths split into several experimental sites, which is manifested in figure 1.2. However, we are interested only in protons used at LHC.

The Proton Synchrotron is one of the oldest apparatus in CERN, because it has been working since 1959. Protons and ions for LHC are using the same infrastructure beyond this point. The protons are accelerated there up to 25 GeV

[4]LINAC 2 will be replaced in run 3 by new linear accelerator LINAC 4, which proposed properties can be found in [11].

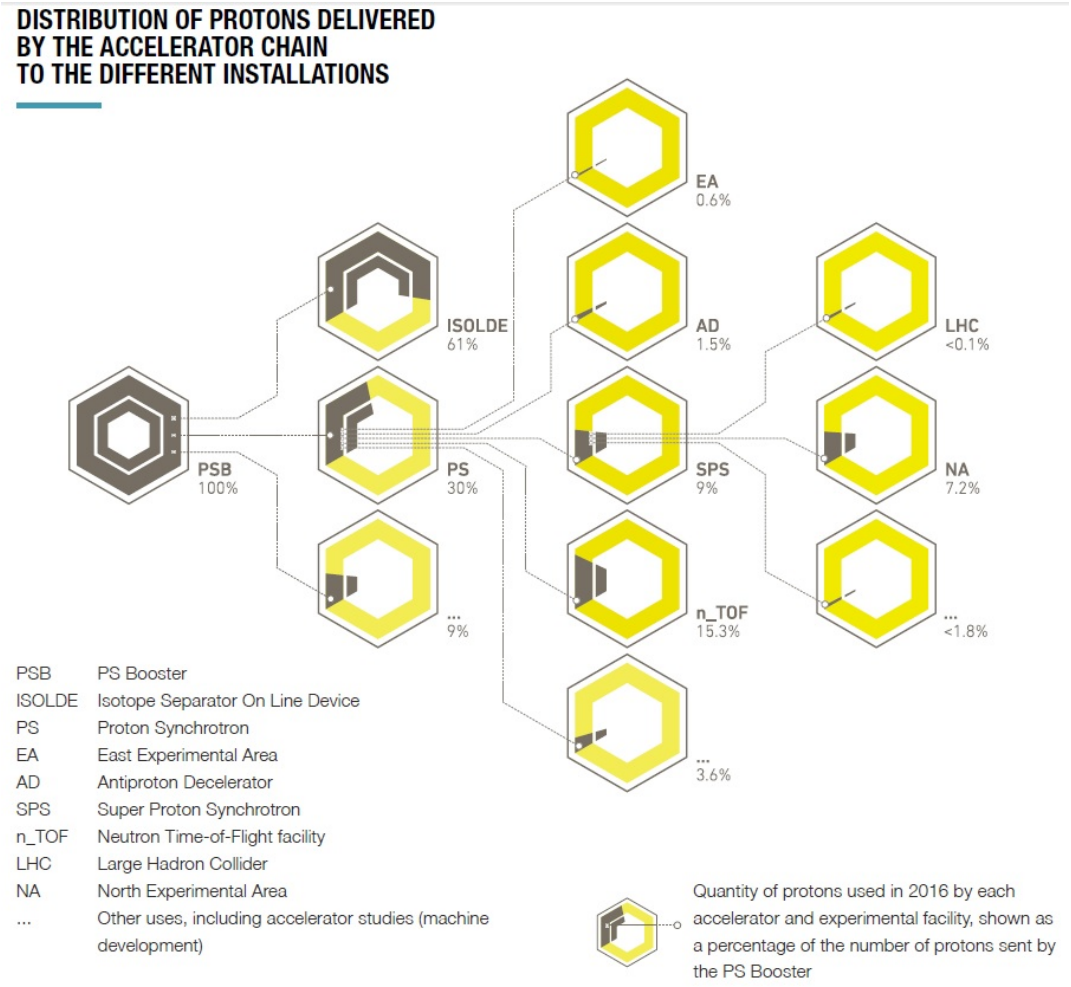

Figure 1.2: Distribution of protons delivered by the accelerator chain to the experiments. Taken from [13].

kinetic energy and injected to the SPS, which increases the energy of protons to 450 GeV [14]. For example, W and Z bosons were discovered in collisions provided by SPS in 1983. Finally, the protons injected into LHC are able to obtain the energy of 7 TeV [5]. LHC is probably the best known for the discovery of the Higgs boson in 2012 [15].

Until this point we mainly discussed protons. However, this work is dedicated to heavy-ion physics, which has only a limited time to collect data from collisions. In the longer term LHC schedule [16] heavy-ion physics has about a month every year to measure. [6] To be precise, the granted time was 28 days in 2018, which was the last run of lead-lead collisions. Four days are typically dedicated to set-up the machines and the rest of the time is used to collect data.

The process of acceleration of lead ions is described in this paragraph. The major source of information is [17], because the information from the original design report of injector chain [18] is a bit outdated due to the steady upgrades. On the other hand, the second source is more extensive, it is only necessary to

[5] 14 TeV is the energy of collision of two bodies

[6] Similar division of the measurement time in LHC for heavy-ion physics is expected to be determined after the long shutdown 2.

carefully take into account the modifications made to the production chain. The first step in accelerating the ions is the lead ions source. Solid pure isotope 208 of lead is vaporized, then the vapour is ionised and turned into the particle beam in the radiofrequency cavity. The cavity is followed by LINAC 3, which has two tasks to provide. Not only to accelerate the ions, apart from LINAC 2. Since the Lorentz force is directly proportional to the atom charge, then in order to achieve the highest center-of-mass energy in the collisions, it is reasonable to try to get rid of the atom of all electrons and leave only the nucleus. That is the second task from LINAC3. It accelerated the atoms of lead to 4.2 MeV per nucleon [7]. Moreover, during the journey inside the accelerator, the atoms of lead are so-called stripped. Stripping is a process when an accelerated atom breaks through a thin foil and loses electrons due to their interaction with the foil. At LINAC3 the atom subsequently undergoes a change from state $^{208}Pb^{27+}$ to state $^{208}Pb^{54+}$. LINAC3 is followed by the Low Energy Ion Ring (LEIR), which has the same tasks as PS Booster for protons. LEIR splits beam from LINAC3 into bunches[8] and accelerates each atom of lead up to the energy equal to 72.2 MeV per nucleus and then inject the bunches into the PS. In Proton synchrotron the energy per nucleus of lead atoms reaches 5.9 GeV. Then, during the transfer to SPS, the final stripping changing an atom from $^{208}Pb^{54+}$ state to a nucleus with no electrons is applied. Stripping as well as other manipulations with the bunches, is not lossless. However, the improvement of the properties of bunches has been developed continuously. The last pre-accelerator for ions is the Super proton synchrotron, where the ions obtain kinetic energy up to 176.4 GeV per nucleon before they are injected into the LHC. In run 2 the achieved value of the center of mass energy $\sqrt{s}$ of two colliding ions has been 5.02 TeV per nucleon [19].

To sum it up, LHC does not only provide collisions to experiments, but it also serves as a prototype of a number of vacuum and low-temperature technologies. The whole CERN experimental complex, which was presented above is shown schematically in figure 1.3.

1.5 ATLAS detector at LHC

Detector ATLAS [21], whose name is an abbreviation for A Toroidal LHC ApparatuS, is one of the four big experiments on LHC ring and along with CMS one of the two general-purpose detectors. The ATLAS detector is symmetric along the z-axis with respect to the interaction point (IP), it also covers the full 2π in azimuth angle. The ATLAS detector is composed of several various sub-detectors, which are labelled in figure 1.4. The sub-detectors are typically divided into three groups - the inner detector (ID), the calorimeter system, and the muon spectrometer.

Inner detector (ID) measured the path of charged particles and their momenta. ID is the closest detector to the interaction point. Inner detector is composed of three independent subsystems - Pixel detector, Semiconductor Tracker, and Transition Radiation Tracker. The first two detectors use very accurate silicon

[7] It is customary to use energy per nucleon rather than the energy of the whole projectile.
[8] Each bunch contains $2.2 \cdot 10^8$ ions.

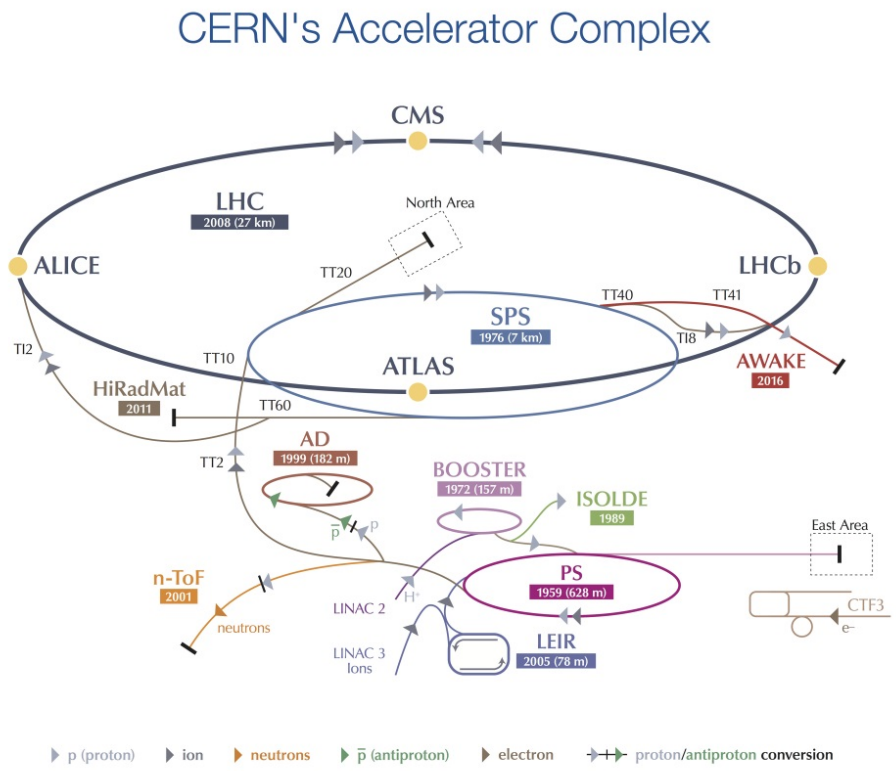

Figure 1.3: The schema of CERN site with labelled experiments and colliders. Taken from [20].

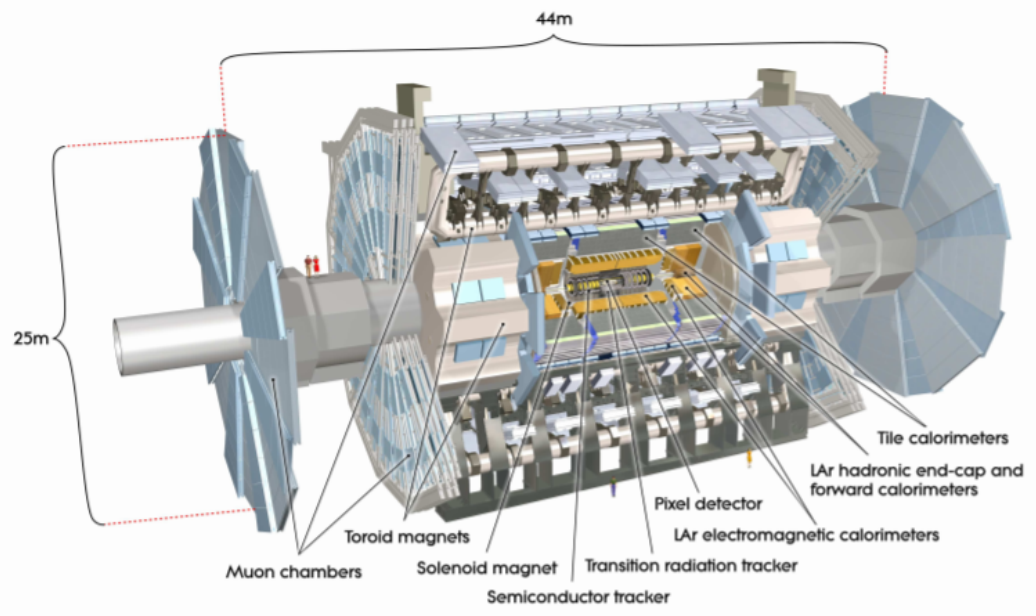

Figure 1.4: Layout of ATLAS. Taken from [21].

detectors, pixel, respectively strip silicon detectors. The last part of ID uses xenon-filled straws. A charged particle emits transition radiation as it passes between the straws. The spatial resolution of this detector is not good in comparison to the silicon detectors. However, it is a suitable poor man solution, which

provides us a lot of information about charged particle tracks. TRT is also very radiation resistant. The cut-away view of the ATLAS ID is shown in figure 1.5. The ID active area is placed into a solenoid magnetic field of intensity $B = 2$ T. The curvature of the particle path is caused by the Lorentz force. It helps us to determine the momentum of particles. The ID covers the pseudorapidity in range $\eta \in$ (-2.5,2.5). This value is significant for data selection used in this work, which is mentioned in chapter 4.

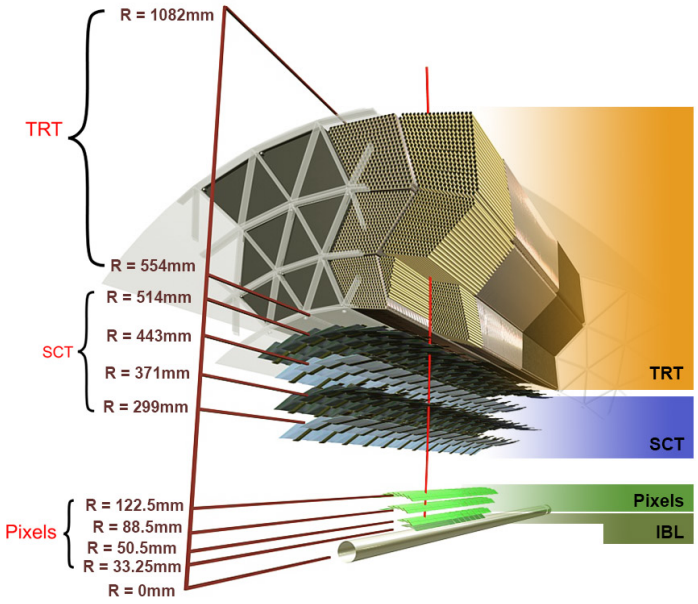

Figure 1.5: The cut-away view of the ATLAS ID. Taken from [22].

The apparatus with a major impact on jet physics study is a calorimeter detection system. The principle of energy measurement is divided into two steps. Interaction of passing particle with the detector mass leads to energy deposition. It is necessary to measure all the energy of the particle, hence the calorimeters are vast instruments. The subsequent task is to measure the energy and convert its amount to a signal. Unfortunately, it is not that easy. Various particles passing through material behave very differently. Hence, the ATLAS detector has two main sub-systems of calorimeters - electromagnetic calorimeter and hadronic calorimeter.

Both calorimeters are sampled, which means that there are layers of an absorber, a passive material, which is cheap and within which the mean free path of the passing particles is as short as possible. The absorber is interleaved with an active material, which measures the energy deposited in active volume due to the interactions in an absorber. The advantage of this approach is the price, but the sampling method results in degradation of resolution. Therefore, for example, the CMS detector has electromagnetic calorimeter made fully of scintillator wedges.

The calorimeter system covers the pseudorapidity range $\eta \in (-4.9, 4.9)$. The elmag. calorimeter gather information for pseudorapidity in range $|\eta| < 1.475$ in barrel part and in range $1.375 < |\eta| < 3.2$ in two end-caps. The electromagnetic calorimeter at ATLAS uses lead as the absorber and liquid Argon scintillator (LAr) as an active material. The big advantage is the absence of an atomic lattice,

so there is no radiation damage of that type. Phenomena such as electron-positron pair formation, bremsstrahlung, Compton scattering, and ionization occur in the electromagnetic calorimeter. Overall, we can say that the number of electrons, positrons, and photons increases, but their average energy decreases in the shower that is developed when an electron or photon is passing through a material. Until the so-called electromagnetic shower stops completely in the material. The hadronic calorimeter covers the regions with pseudorapidity $|\eta| < 1.0$ with barrel part, $0.8 < |\eta| < 1.7$ with extended barrel part, and $1.5 < |\eta| < 3.2$ with endcaps. Barrel parts use steel as the absorber and the measurement is done by plastic scintillator tiles. Endcaps use copper as the absorber and LAr as active material. The coverage of the forward regions, where pseudorapidity meets the condition $3.1 < |\eta| < 4.9$, is provided with forward electromagnetic calorimeter module made of copper absorber and LAr as active material and hadronic calorimeter module made of tungsten absorber and LAr as active material. Copper, same as lead and steel or tungsten, is a suitable choice because it cannot be magnetized, otherwise we would interrupt the bending magnetic field. A hadron shower occurs in a hadron calorimeter. The phenomena taking place here are hadron production, nuclear deexcitation, and pion and muon decays. Moreover, neutral pion almost immediately decays into a pair of photons, the energy of the neutral pions creates a new electromagnetic shower, which propagates as stated above. Hadronic calorimeters are bigger in general than electromagnetic ones, inasmuch as the electromagnetic showers are absorbed over a shorter distance than the hadron ones. Therefore, the hadron calorimeter is placed behind the electromagnetic one. One can see that in the cut-away view of the ATLAS calorimeters is displayed in figure 1.6

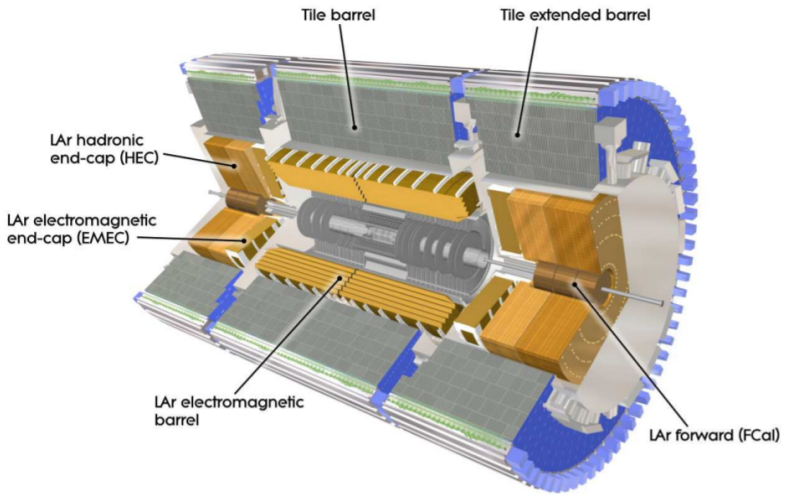

Figure 1.6: The cut-away view of the ATLAS calorimeters. Taken from [21].

The outermost part of the detector is the muon spectrometer (MS). Only the muons have sufficient energy to reach the spectrometer [9] The whole detector is exposed to a magnetic field by superconducting air-core toroids. The magnetic induction of the field ranges between 2 and 6 T. The field is used to deflect the trajectories of muons. Precise measurement is realised within the region of

[9]Neutrinos reach the spectrometer as well but are not detectable in the spectrometer.

$|\eta| < 2.7$ with three layers of monitored drift tubes. Moreover, in the forward region in order to suppress higher background signals cathode-strip chambers are located. The last part of the muon trigger system serving as a fast trigger are resistive plate chambers in the barrel, and thin-gap chambers in the endcap regions within the region of $|\eta| < 2.4$.

Thanks to the small cross-section of the interaction with baryon matter [10], muons reach the spectrometer quickly and the signal from them can be identified easily compared to other particles in other sub-detectors. The overview of the muon spectrometer is displayed in 1.7.

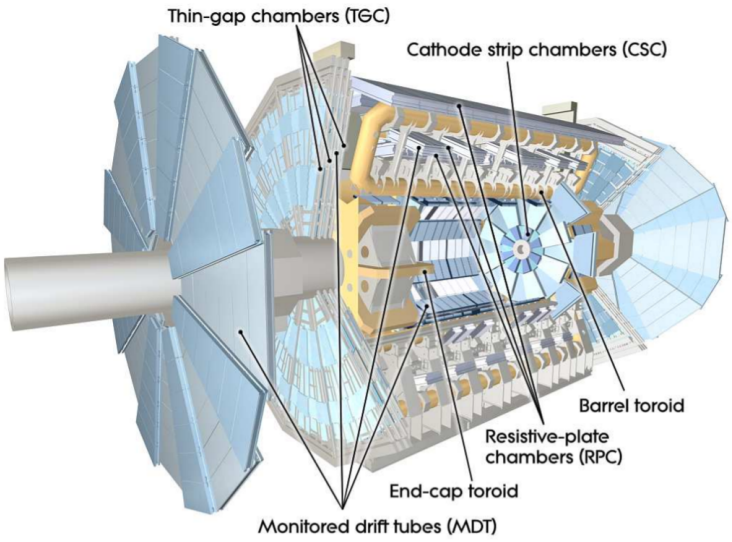

Figure 1.7: The cut-away view of the ATLAS muon spectrometer. Taken from [21].

Figure 1.8 shows how various particles appear inside the ATLAS detector. The shown particles are those that are stable enough to leave a detectable trace. Less stable exotic particles can be identified by connecting the pathways to vertices. It is also impossible to measure a neutrino in the data at ATLAS, as it interacts only very weakly. Its presence is manifested by the lack of transverse energy and momentum in data. The topic of particle detection and identification is discussed in depth in [5].

Delivered integrated luminosity to ATLAS from accomplished runs, which characterised both - the detector as well as the collider, are displayed for proton-proton collisions annually as well as overall in figure 1.9. For lead-lead collisions in figure 1.10.

The given graphs show the increasing amount of data taken from the LHC machine over time, which allows us to study rarer processes.

[10] For given energies of muons in the order of tens and hundreds of GeV.

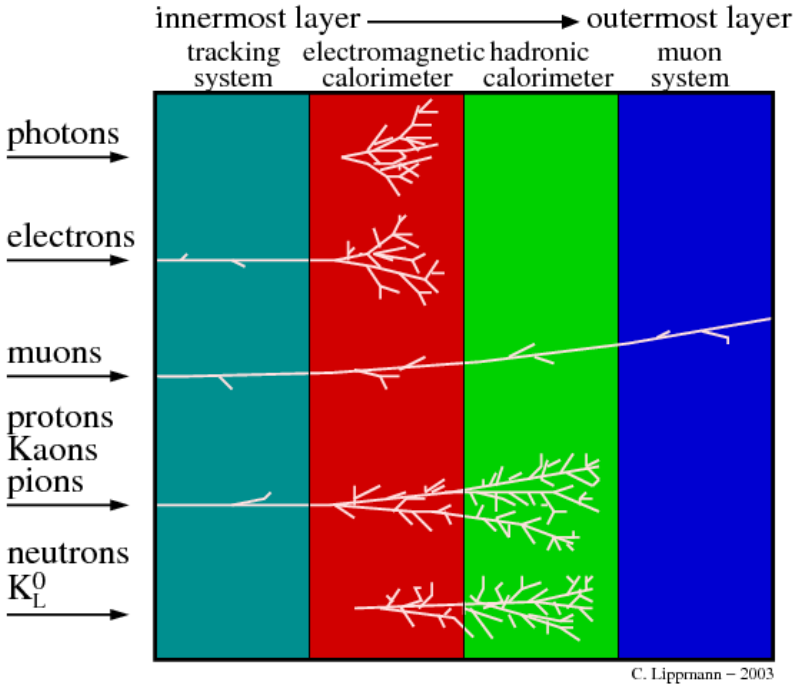

Figure 1.8: Various particle signatures in ATLAS. Taken from [23].

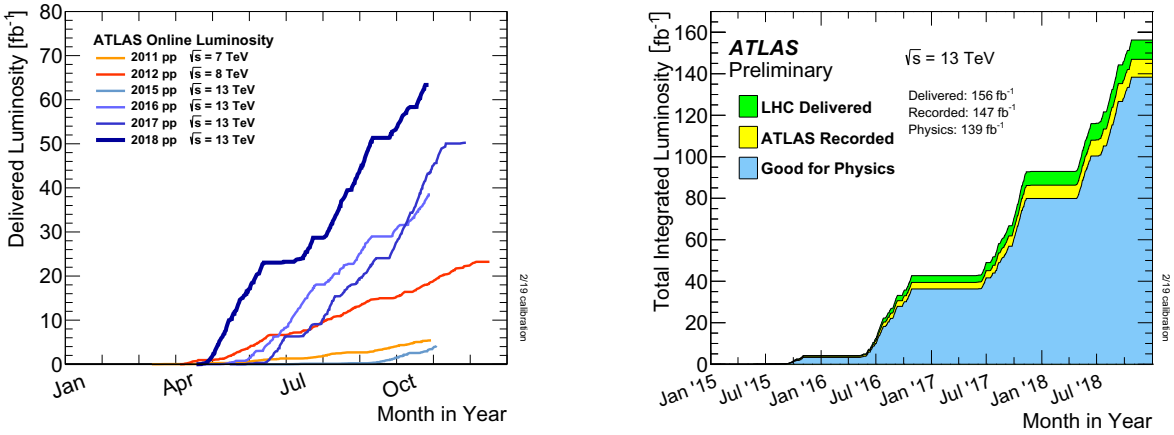

Figure 1.9: Integrated luminosity collected in proton-proton collisions at ATLAS at center of mass energy $\sqrt{s}$. Annual values in the left picture, overall value in the right picture. Taken from [24].

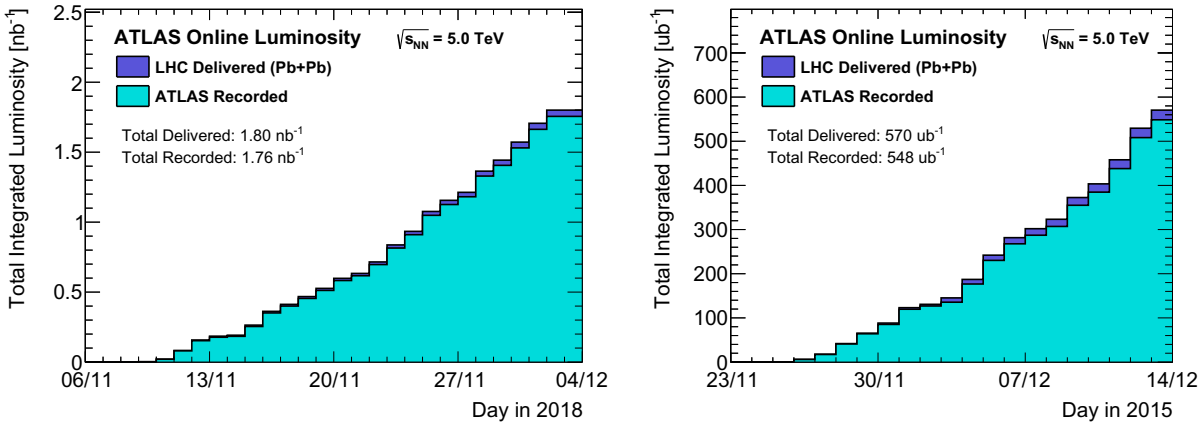

Figure 1.10: Integrated luminosity collected in lead-lead collisions at ATLAS at center of mass energy $\sqrt{s} = 5.02$ TeV per nucleon. In the left picture for 2015 run, in the right picture for 2018 run. Taken from [24].

2. Physics background

The Standard Model of particle physics (SM) is the set of theories [1], which altogether provides the description of three fundamental interactions - electromagnetic, weak, and strong except for gravitational force. Any theory of gravity has not been successfully implemented into the model yet. Considering energy and mass scales in current experimental particle physics one can see that we can without prejudice to generality omit phenomena associated with gravitation force. Thus, the SM is a cardinal theory for every experimental particle physicist.

2.1 Quark gluon plasma in Standard model

At first, we need to outline a brief introduction to the principles, which follow from the theory of strong interaction called Quantum Chromodynamics (QCD). A correct derivation of QCD principles is far beyond the scope as well as the intention of this work. The following paragraphs are based on [25], which does not reflect the changes in the SM in last the thirty years, but it provides us a sufficient theory of quarks and QGP.

Six quarks are one of the seventeen building blocks of the universe according to the SM. These building stones called particles are indivisible. Each particle has intrinsic quantities related to the interaction, which the particle undergoes. Quarks interact via all four know interactions. Thus, they have quantities such as mass related to gravitational force, electric charge related to electromagnetic force and color charge related to strong force, and a few more, which describes inner states of freedom such as spin.

The matter we encounter in everyday life, is composed of particles with an inner structure called baryons and further indivisible electrons. There are only two stable baryons, under some circumstances [2], proton and neutron. In general, baryon is a particle made up of three quarks, each with different color charge. Color charges are three - blue, green, and red. Quarks can also make up composite particles called mesons, which consist of one quark with arbitrary color charge and one antiquark[3] with corresponding anti-color, an example of a meson is a pion. There also exist extremely rare groups of particles made up of four or five quarks, named tetraquarks [27] and pentaquarks [28], respectively. These groups were discovered relatively recently. Overall, all these groups of particles formed by quarks are called hadrons. Quarks inside hadrons are bound through strong force carrier particles, gluons. There are eight types of different gluons, in the language of quantum mechanics it is more accurate to say eight states - octet state. Gluons number among bosons, which means that the wave function of a multi-gluon system is symmetric. As a result, gluons are not subject to

[1] Various elements of this theory set has been verified and expanded for more than last fifty year all over the world.

[2] A free neutron is unstable. However, the neutron bound in a nucleus is stable. Proton stability is still an open problem. According to measurement done by [26] mean lifetime of proton is more than $3.6 \cdot 10^{29}$ years at 90% Bayesian credibility level

[3] Antiquark is an antiparticle to quark. In general, an antiparticle has the same mass, but the rest quantities have the opposite values

the Pauli exclusion principle. Thus, the occurrence of gluons in a region with respect to other gluons is not limited in any way. Let us note that indifference to the Pauli exclusion principle is the reason why bosons are carriers of interactions. Altogether gluons and quarks, which are the only groups of particles participating in strong interactions, are called partons.

The QCD theory has two main properties. The first one explains why we are unable to observe bare quarks outside hadrons. This phenomenon is called color confinement [29], which is explained later in section 2.3. It results in behavior such that further quarks are, the stronger they attract each other. The second crucial QCD property is called asymptotic freedom. It is manifested that strong interaction between particles becomes asymptotically weaker with increasing energy scale or decreasing characteristic dimension scale of the system. The dependence of a coupling constant on a characteristic scale comes from the action of the field on virtual color-charged particles. Quarks screens the field, on the other hand, gluons anti-screens the field. Thus, the existence of asymptotic freedom of QCD is a consequence of the fact that anti-screening caused the gluons to overcome the screening caused by the quarks due to the number of existing gluons and quark flavors. More details about the phenomenon of asymptotic freedom are given in a relatively easy-to-understand way in [30]. As the result of asymptotic freedom, when the matter is extremely squeezed, in other words, it has high density and temperature, then quarks can be liberated. A new state of matter called quark-gluon plasma is formed.

According to the current cosmological models and measurements done in existing heavy-ion experiments, the primeval state of matter in the universe was QGP [31]. The more we understand the properties of the QGP, the more accurate cosmological model we are able to obtain. QGP physical parameters might also provide fundamental information for string theory and the theory of quantum gravity [32].

2.2 Heavy-ion collisions

In the previous section we have noticed that there exists state of matter called QGP, which occurs in extremely hot or dense hadronic matter. The phase transition between hadronic matter and quark-gluon plasma described in classical terms of thermodynamics [33] is displayed in figure 2.1.

In figure 2.1 we can see both feasible approaches in producing QGP. The first approach how to prepare QGP is to collide heavy-ions such as nuclei of gold, lead, or xenon. This approach increases the temperature (energy) of hadrons so much that they undergo a phase transition and turn into QGP. There is also an area on the left from the critical point, where according to the theory smooth transition between QGP and hadronic matter should be realized. The second possible approach is to squeeze hadrons as happens in neutron stars. By squeezing them we enlarge the value of chemical potential μ_B, which is a quantity describing the internal energy per mole of substance [4]. The detailed form of the phase diagram is the subject of research.

When one takes arbitrary heavy-ion, at rest it has temperature $T = 0$ MeV

[4] It can be also interpreted as energy that can be absorbed due to the phase transition.

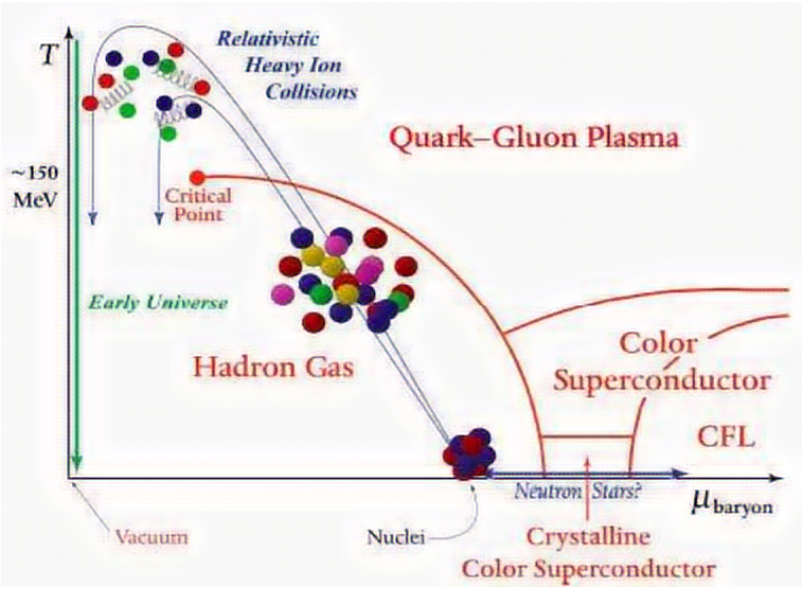

Figure 2.1: The phase diagram of strongly interacting matter in temperature-chemical potential $T - \mu_{\mathrm{B}}$ space. Taken from [34].

and chemical potential close to nucleon mass. In simplified description, ion goes through hotter and hotter equilibrium states even after it becomes QGP.

LHC apparatus described in the first chapter is used for accelerating heavy-ions. Most of the time, experiments on LHC are done with bunches of protons instead of heavy-ions. However, for studies of QGP it is preferable to use heavy-ions since the multiplicity of parton collisions is typically greater by two orders of magnitude. The high multiplicity allows us to study the collective behavior of QGP. A truly not negligible volume of QGP is created for a moment. On the other hand, requirements for simultaneous detection are significantly higher. Historically, the first observation of QGP was achieved with detectors STAR and PHOENIX at collider RHIC in 2003 [1].

Apart from protons one must also take into account that nuclei can no longer be considered as points in collision geometry. Final state properties [5] in an ion collision strongly depend on collision impact parameter b. The impact parameter is a quantity, which quantifies the distance between centers of projectiles in the transversal plane and it is directly related to the number of participating nucleons, N_{part}, the number of spectators, and the number of single nucleon-nucleon collisions, N_{coll}. Unfortunately, these quantities can not be measured experimentally. The fundamental quantity used in this study is called centrality and it is generally defined as [35]:

$$c = \frac{\int_0^b \frac{d\sigma}{db'} db'}{\int_0^\infty \frac{d\sigma}{db'} db'}, \tag{2.1}$$

where $\frac{d\sigma}{db'}$ is differential cross-section with respect to impact parameter. Experimentally, centrality is measured at ATLAS using the total E_{T} deposited in forward calorimeters. Its distribution is divided into intervals to include given fractions of the total cross-section. This is further parametrised using simulations and model to extract N_{part} and N_{coll} if needed. Numerical values of centrality,

[5] This dependence is many times highlighted in the discussion of results in chapter 4.

N_{part}, N_{coll} for certain values of b are typically determined in Monte Carlo simulations based on the used physical model of collision. The leading model used in heavy-ion physics is probably the Glauber model, whose approaches and results are briefly summarized in [36]. The centrality interval 0–10%[6] refers to the 10% of the most central collisions while the 80–100% to the most peripheral collisions. Centrality, impact parameter b, and their relation in heavy-ion collision are shown in figure 2.2.

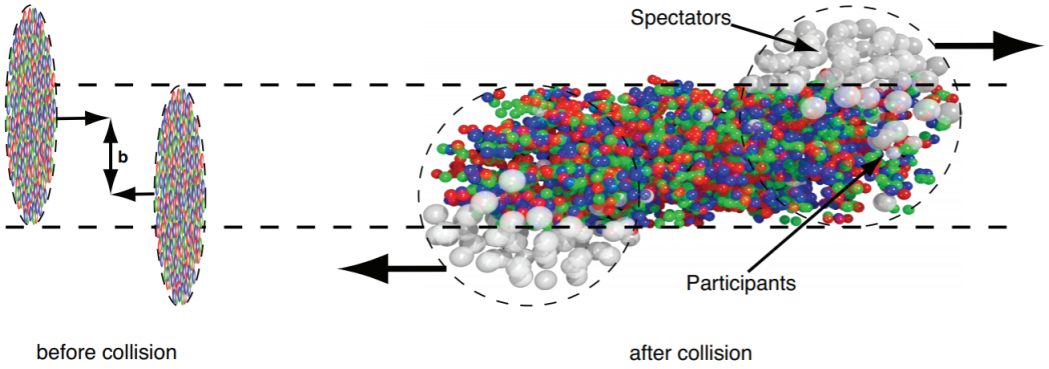

Figure 2.2: A schema of hadronic projectiles collision with impact parameter b and nucleons divided into groups of so called participants (involved in collisions) and spectators (not involved in collisions). Taken from [37].

2.3 Jet physics, jets as probe of QGP

In section 2.1 it was mentioned that one of the two crucial QCD theory phenomena is called color confinement. The phenomenon is named with respect to the finding that partons can not exist as bare particles due to their non-zero color charge. In the first section of this chapter, there were given examples of hadron types. Altogether these types have a total color charge equal to zero. In a macroscopic world, we can imagine a situation where we overcome the attractive force by moving the attracted particles away from each other with sufficient energy. The sufficient energy must exceed the potential associated with attractive force. However, the strange nature of color confinement does not allow this approach. Let us describe the action of the strong interaction on the analogy with a spring. As the imaginary spring between quarks stretches, the stored energy in the spring increases until a certain level of energy. The critical energy level is the energy sufficient to form a quark-antiquark pair. The newly born partons from vacuum together with previously existing ones form hadrons. This phenomenon is called hadronisation. The described behavior is shown in the figure 2.3.

Consequently, after a hard scattering process and following hadronisation, in the first approximation there are two collimated sets of hadrons moving back-to-back. These sets of hadrons are called jets. As a jet evolves, the paths of the jet hadrons form a cone. If one identifies the hadrons that make up the jet and measures their properties, one can reconstruct the properties of the original

[6]It is customary to give the value of the centrality as a percentage, not as a decimal number.

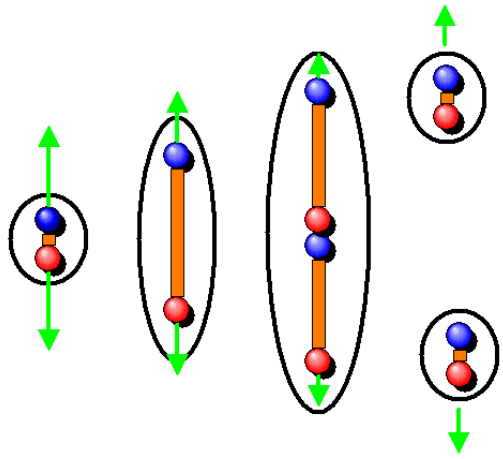

Figure 2.3: Confined quarks cooled below the deconfinement phase transition temperature undergoing hadronization. Taken from [38].

parton. However, in heavy-ion physics, there is another use of jets. The jets can serve as a useful probe into the properties of the QGP.

The jets, which form in heavy-ion collisions interact during propagating through surrounding QGP. Depending on the properties of the QGP along the jet path, the jet suffers differential energy loss. The jet may even disappear completely. This phenomenon is called jet quenching. Naturally, we are unable to measure the initial properties of the jet and study its change over the path. However, we can for example compare the asymmetry in the properties of a pair of jets initiated at the same collision. The jet quenching process is schematically shown in figure 2.4.

The jet quenching mechanism and its use to study QGP properties was first predicted, calculated, and published in [40]. In the paper, it is derived that the differential energy loss $\frac{dE}{dx}$ is roughly proportional to the square of the plasma temperature. There is also made an assumption of the type of events, which could be easily observed. The type is called mono jet signature and it represents the category of events when the hard scattering takes place near the edge of QGP. One jet escapes the area through hadronic gas with negligible energy loss. The other one is fully absorbed in the plasma. One can imagine the event as an even more extreme case of the situation shown in the figure 2.4. Indeed, such events were observed in 2010 at ATLAS [41]. ATLAS also was the first to directly observe jet quenching in the measurement of dijet balance [41]. An event display from the paper showing the Inner Detector and calorimeter systems is shown in figure 2.5. One can notice that in the Inner Detector, there are two back-to-back collimated particle showers. However, only one of that two back-to-back sets of particles deposited a significantly stronger signal in hadronic calorimeter. This event can be interpreted as an imbalanced dijet event where one jet was quenched.

2.4 Jet reconstruction at ATLAS

In the previous section, we presented arguments for the importance of jet reconstruction in high-energy physics, especially in ultra-relativistic heavy-ion physics.

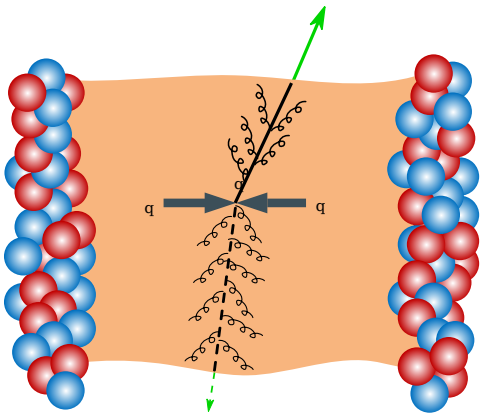

Figure 2.4: Schematic view of two partons created in hard scattering process propagating through QGP. Taken from [39].

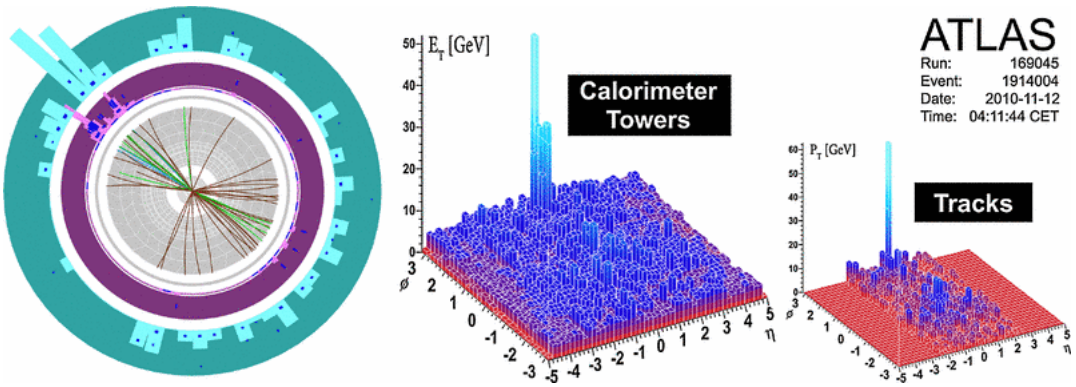

Figure 2.5: Event display of an asymmetric dijet event, recorded by ATLAS in 2010 in lead-lead collisions. Taken from [41].

In this section, we describe how specifically the reconstruction is done.

At first, let us introduce what the signal from jets looks like in the ATLAS detector. The smallest unit of the calorimeter is called a cell. In general, cells do not have the same spatial resolution. Each cell is defined by η, ϕ, and the layer to which it belongs. If one sums the signal from cells overall layers such that $\Delta\eta$ x $\Delta\phi = 0.1$ x 0.1, one obtains an object called a calorimeter tower.

Subsequently, we provide the explanation of the reconstruction algorithm, which reconstructs jets from calorimetric tower signals. Algorithms responsible for jet clustering are important tool for analysing data from hadron collisions. There is a plenty of algorithms. A specific choice of an algorithm is based on studied phenomenon. The most commonly used algorithms are sequential recom-

bination and cone jet algorithm. The group of sequential algorithms includes the widely used anti-k_t algorithm, which was used for reconstruction of jets in our training data set. Anti-k_t algorithm [42] is a soft-resilient jet algorithm, which means that a jet shape and a jet area is not influenced by soft radiation. In order to explain algorithm itself, let us define distance between particles d_{ij} and distance between particle and the beam d_{iB} by following relations:

$$d_{ij} = min\left(\frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2}\right)\left(\frac{\Delta R_{ij}}{R}\right)^2, \tag{2.2}$$

$$d_{iB} = \frac{1}{k_{ti}^2}, \tag{2.3}$$

where $\Delta\ R_{ij}$, R, and k_{ti} are the distance between the i-th and the j-th particle in $\eta - \phi$ space defined by (1.6), radial parameter and momentum of the i-th particle respectively.

In each iteration, the algorithm calculates the distances and identifies the smallest distance. If the smallest distance is d_{ij}, then the algorithm recombines entities i and j into one. While if the smallest distance is d_{iB}, the algorithm tag i as a jet and removes it from the list of entities. After that, the distances are recalculated and the process repeats until there are no remaining entities to be clustered.

Other commonly used reconstruction algorithms may be obtained just by using different value of the power of momentum $k_{ti/j}$ in definitions (2.2) and (2.3). In k_t algorithm quantities d_{ij} and d_{iB} are functions of $k_{ti/j}^2$ instead of $\frac{1}{k_{ti}^2}$. Cambridge/Aachen algorithm works independently to momenta, i. e. with power index of zero.

The performance of the anti-k_t algorithm with given radial parameter R can be compared to the other algorithms with other choices of the value of the power of momentum $k_{ti/j}$ under considering of an event, which contains a few well spatially separated hard particles and many soft particles. Transverse momenta of the i-th hard particle is k_{it}.

Let us analyse the behaviour of the anti-k_t algorithm. The distance d_{1i}, where 1 represents a hard particle, is significantly suppressed by the high value of the transverse momentum compared to the momenta of the soft particles despite similar spatial distribution. Thus, soft particles are likely to cluster with hard ones rather than with other soft particles. If there is no other hard particle in the distance of $2R$ from the hard particle, all the soft particles within a radius R from the hard particle are clustered with it.

If there are two hard particles [7] closer than $2R$, there are two options. If $\Delta R_{12} \in (R,2R)$, there are reconstructed two jets. However, apart from the previous case, reconstructed jets cannot be perfect cones. If the momentum of the first hard particle significantly exceeds the momentum of the second hard particle, then the jet associated with the first hard particle is a cone and the second jet is just partly conical. Such situational will be performed graphically below. [8]

[7] Second hard particle is represented with number 2.

[8] One can see such jet in figure 2.6 in graph for anti-k_t algorithm in area around coordinates $y = 2$, $\phi = 5$.

If momenta of both hard particles are approximately the same, then both cones are cut off by common boundary b, which is determined by condition [9]:

$$\frac{\Delta R_{1b}}{k_{t1}} = \frac{\Delta R_{2b}}{k_{t2}}. \tag{2.4}$$

Remaining case to study is the situation for a pair of hard particles with $\Delta R_{12} < R$. Logically, only a single jet is reconstructed. If the momentum of the first hard particle significantly exceeds the momentum of the second hard particle, then the conical jet is centred on the place, where the first hard particle occurs. In case the momenta of both hard particles are approximately the same, then the shape of the reconstructed cone is a more complex union of conical shapes.

The described behavior of anti-k_t algorithm for a pair of hard particles can be naturally generalized on a higher number of hard particles.

Jets reconstructed by different jet algorithms from a particle-level event with a high count of random soft noise particles are illustrated in figure 2.6. It is clear that the clusters delivered by anti-k_t algorithm are resilient to a specific occurrence of random soft particles, apart from k_t and Cambridge/Aachen algorithms.

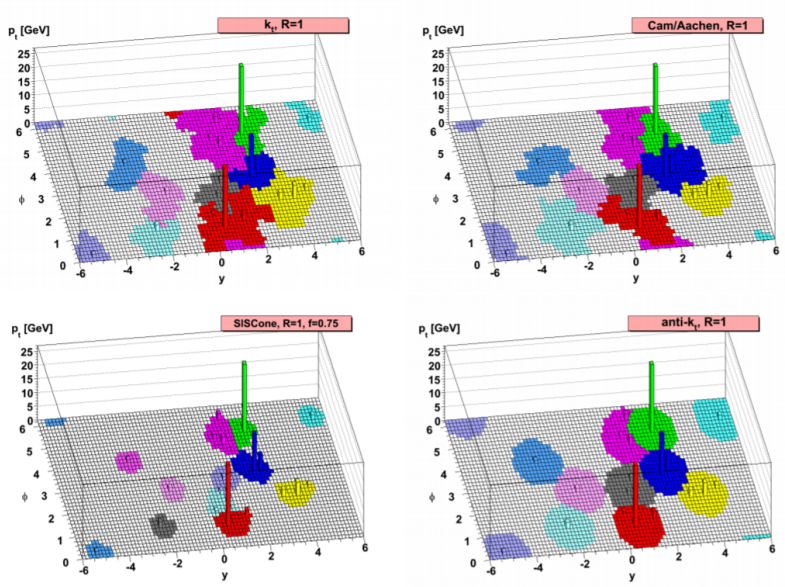

Figure 2.6: A comparison of results provided by four different jets algorithms processing a parton-level event containing many random noise soft particles. Taken from [42].

In figure 2.6 one can notice important behavior of anti-k_t algorithm. Hard jets clustered by anti-k_t algorithm are represented by circles in $\eta - \phi$ space with strictly given radius R. On the other hand, softer jets could be more variable in the representation of their shape.

The collisions of two Pb nuclei has a significantly higher level of the underlying event in comparison to proton-proton collisions. Hence, specific algorithms

[9] The condition follows from (2.2) providing a set of points j with the same distance from both hard particles

[43] have to be used event-by-event to correct underlying event background energy deposited in the jet region. Underlying event energy is subtracted from all calorimeter towers.

Additional corrections are also applied. If an algorithm that has already finished the subtraction procedure cannot decide which jets have an origin in a realised hard process and which are fake signals, the decision is made according to the ratio of maximum and average transverse energy in the towers of the jet. The described procedure might be improved with data from tracking jets or smaller jets with $R = 0.2$. The second correction is called a $\eta - \phi$ correction, which fixes non-uniformities of the reconstructed jet in the ϕ direction due to the presence of dead modules and non-uniformities in the calorimeters.

3. Jet calibration

3.1 Data sets

The data sets utilised in this study are MC16 Pythia 8 Event Generator jet events at $\sqrt{s} = 5.02$ TeV with the A14 ATLAS tune [1] and parton distribution function [2] NNPDF23LO grids [44]. The definitions of used proton-proton MC samples can be found in tables 3.1 and 3.2.

JZ	Dataset name
2	mc16_5TeV.420012.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2R04.recon.AOD.e*_s3238_r11199
3	mc16_5TeV.420013.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3R04.recon.AOD.e*_s3238_r11199
4	mc16_5TeV.420014.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4R04.recon.AOD.e*_s3238_r11199
5	mc16_5TeV.420015.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5R04.recon.AOD.e*_s3238_r11199

Table 3.1: A list of 5.02 TeV Pythia 8 proton-proton MC samples.

JZ	$p_{\mathrm{T}}\,\vert_{R=0.4}$ [GeV]	$\sigma \cdot \epsilon$ [nb]	Number of events
2	60–160	$(6.4 \cdot 10^{5}) \cdot (4.29 \cdot 10^{-3})$	8 M
3	160–400	$(4.7 \cdot 10^{3}) \cdot (5.30 \cdot 10^{-3})$	8 M
4	400–800	$(2.7 \cdot 10^{1}) \cdot (4.59 \cdot 10^{-3})$	8 M
5	800–1600	$(2.2 \cdot 10^{-}1) \cdot (2.18 \cdot 10^{-3})$	8 M

Table 3.2: 5.02 TeV Pythia 8 proton-proton MC samples with cross-setions and filtering efficiency.

The lead-lead MC sample uses the same event generator, tune, and particle distribution functions as in proton-proton case. The lead-lead MC sample is created by overlaying Pythia8 dijet MC events on top of Pb+Pb minimal bias events. The corresponding samples are recorded in the tables 3.3 and 3.4.

Many subsystems of ATLAS have a sensitivity window longer than the time interval between arrival of two bunches, which is approximately 25 ns. Hence, the following or the previous collision may affect the studied one. Moreover, in each bunch collision, there occur many parton interactions within the detector sensitivity window. Thus, each event is affected by background signal from other interactions. Such a phenomenon is called a pile-up. Necessarily, the proton-proton MC is generated with the pile-up contamination in accordance with the one in the measured data.

3.2 Current approach to calibration

Due to the difference of elmag. and strong interactions, calorimeters measure different signals for hadronic and electromagnetic components of jets of the same

[1] Each MC generator is specially tuned in order to provide the best performance for used apparatus.

[2] Parton distribution function is a relation giving the probability to find a parton with given flavor in a hadron based on the fraction of the hadron total momentum carried by the parton.

JZ	Dataset name
2	mc16_5TeV.420012.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2R04.merge.AOD.e4108_d1516_r11439_r11217
3	mc16_5TeV.420013.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3R04.merge.AOD.e4108_d1516_r11439_r11217
4	mc16_5TeV.420014.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4R04.merge.AOD.e4108_d1516_r11439_r11217
5	mc16_5TeV.420015.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5R04.merge.AOD.e4108_d1516_r11439_r11217

Table 3.3: A list of 5.02 TeV Pythia 8 lead-lead MC samples.

JZ	$p_{\mathrm{T}}\mid_{R=0.4}$ [GeV]	$\sigma \cdot \epsilon$ [nb]	Number of events
2	60–160	$(6.4 \cdot 10^{5}) \cdot (4.29 \cdot 10^{-3})$	7.7 M
3	160–400	$(4.7 \cdot 10^{3}) \cdot (5.30 \cdot 10^{-3})$	7.7 M
4	400–800	$(2.7 \cdot 10^{1}) \cdot (4.59 \cdot 10^{-3})$	7.5 M
5	800–1600	$(2.2 \cdot 10^{-}1) \cdot (2.18 \cdot 10^{-3})$	7.7 M

Table 3.4: 5.02 TeV Pythia 8 lead-lead MC samples with cross-setions and filtering efficiency.

energy. The quality of calorimeter calibration is quantified by response $\Re$, which is defined as:

$$\Re = \frac{p_{\mathrm{T,reco}}^{\mathrm{EM}}}{p_{\mathrm{T}}}, \tag{3.1}$$

where p_{T} is transverse momentum generated by Monte Carlo simulation on the particle level. Transverse momentum of the reconstructed particle is $p_{\mathrm{T,reco}}^{\mathrm{EM}}$. In a dedicated model, we simulate the interaction of generated particles with a model of ATLAS detector and obtain an output signal with the same structure as the signal from the real data. Such a signal is processed in order to reconstruct an initial particle state. In general, the response of a fully calibrated jet approaches unity. The calibration for jets reconstructed with the heavy-ion algorithm is derived in proton-proton MC collisions using numerical inversion method [43].

Jet reconstruction performance can be characterized by two quantities, jet energy scale (JES) and its resolution (JER) defined as

$$\mathrm{JES} = \langle \frac{p_{\mathrm{T}}^{\mathrm{reco}}}{p_{\mathrm{T}}} \rangle, \tag{3.2}$$

$$\mathrm{JER} = \sigma \left(\frac{p_{\mathrm{T}}^{\mathrm{reco}}}{p_{\mathrm{T}}} \right). \tag{3.3}$$

The jet energy resolution can be described by empirical formula [45]:

$$\sigma(\Re) = \frac{a}{\sqrt{p_{\mathrm{T}}}} \oplus \frac{b}{p_{\mathrm{T}}} \oplus c, \tag{3.4}$$

where parameters a and c depends on the response of detector. Contrariwise, parameter b has two components - the centrality dependent based on fluctuations of underlying event (or a contribution from the pile-up decays in case of proton-proton collisions), and the second one independent of centrality caused by electronic noise. Therefore, in an effort to improve calibration, we vary parameters a and c, not b.

3.3 Neural networks - multi layer perceptron

We would like to remark that we do not try to replace the calibration mentioned in the previous section. We attempt to improve the properties of that calibration even more so that we can get more reliable physical data from heavy ions experiments running on ATLAS.

At first, let us introduce briefly the fundamental principles of machine learning [46] used in this study. TMultiLayerPerceptron class [3], which is implemented in the ROOT framework, was used. The class provides the user with tools for the development, training, testing, and analysis of a multi layer perceptron (MLP) neural network (NN). Artificial neural networks (ANN) are programs that process and categorize inspirations according to the pattern of structures in the human brain. The main utilization of multi layer perceptrons is classification and regression.

The fundamental building block of a neural network is a mathematical function called a neuron, which is displayed in figure 3.1. A neuron has a number I of input values and one output value y. Each input x_i is weighted by corresponding weight w_i. There also may be an additional parameter w_0. A weighted average, called activity a, is computed from inputs. A significant parameter of each neuron network is the so-called activation function. This is a function f such that $f(a) = \mathrm{y}$. Typical choices are the one-sided linear function or the Heaviside function. Activation functions may be also a sigmoid - hyperbolic tangent or logistic function, which is the approach implemented in TMultiLayerPerceptron class. Choosing the suitable activation function is a theoretical matter beyond the scope of this work.

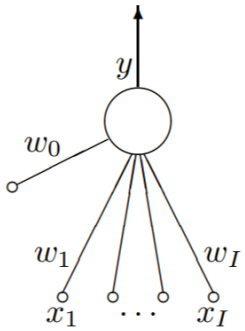

Figure 3.1: A neuron diagram. Taken from [46]

The multi layer perceptron is a neural network in which neurons are arranged in layers as one can see in figure 3.2. The input quantities are represented by the input layer. The output layer represents the variables to be corrected. The values of output quantities are called, in the theory of neural network, targets t_i. The number of targets can generally be different than the number of inputs. The layers between the input and output layers are called hidden layers.

We try to find a map between the input and output quantities. The approach of finding the map is called error back-propagation. First, we define the error function:

$$E(\vec{w}) = \frac{1}{2}(\sum_{i=1}^{n}(y_i - t_i)^2), \tag{3.5}$$

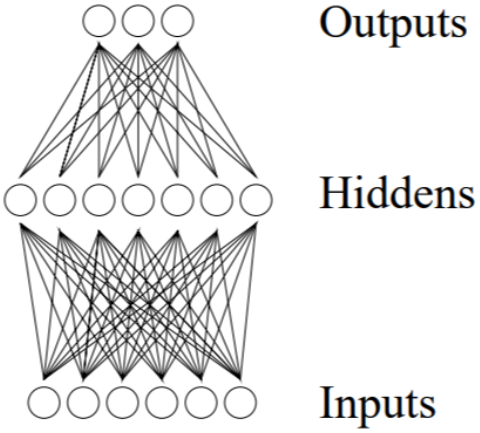

Figure 3.2: A diagram of multi layer perceptron neural network. Taken from [46]

where y_i denotes the outputs of the last hidden layer. The multiplicative constant before the expression makes it easier to work with derivatives.

Finding the map represented by weights w_i is equivalent to the task of minimizing error function $E(\vec{w})$. Each event is represented by the pair of vectors $\vec{x}, \vec{t}$, from them we calculate the error function $E(\vec{w})$. The gradient of the error function is calculated. The calculation proceeds backward from the output layer to the input layer. The chain rule is used to find the derivatives. The weights in the network are changed as in the gradient descent method. There of course exist problems related to such an approach, for example, we can find only the local minimum of $E(\vec{w})$. Hence, the used error function is a bit more sophisticated. However, the principle is the same. The second possible approach is stochastic learning, where the neural network is built by introducing random variations into the network. This approach allows the random fluctuations to help the neural network escape from the local minima of the error function. This type of learning is suitable for tasks that are described by a complex dependency.

The second possible use of MLP is classification. The network used for classification is very similar to the one used for regression. In the case of classification, we choose t_i to be equal to one, which represents the correct class, in our case quark/gluon induced jet. The rest of the elements t_j are set to zero. If we normalize the output from the last layers y_i, then these coefficients can be interpreted as the probability that the object characterized by the given inputs $\vec{x}$ belongs to the class represented by t_i. Back-propagation is the same as for a neural network used for regression. However, in the case of classification the error function is replaced by the negative natural logarithm of the likelihood function:

$$G(\vec{w}) = -\sum_{i=1}^{n}(t_i \cdot \ln y_i) \tag{3.6}$$

It is also worth mentioning a few words about the training sample. The training data set should be as balanced and broad as possible from the phenomenon it represents. For example, events with a given pseudorapidity and its negative value are the same, because the physical processes and the detector are symmetric with respect to the z-axis. Let ϵ be a small change of pseudorapidity. However,

when varying values of $\vec{w}$ with respect to ϵ, a neural network does not have the information that $\eta + \epsilon$ and $-\eta$ are close to each other. For this reason, it is necessary to know the physical background of the studied processes and prepare the training data which are not ambiguous to the neural network.

4. Results

4.1 Study of current calibration

Currently used calibration was described in section 3.2. Accordingly, we determine the response $\Re$ between the reconstructed and the truth jets. Determination was done separately for various intervals of centrality and pseudorapidity since detector response and performance varies with pseudorapidity and the best performance is expected for peripheral collision due to low underlying event contribution. Inside each interval of p_{T}, η, and centrality, we fitted response $\Re$ distribution by Gaussian function to extract mean and sigma in order to evaluate JES and JER. Let us remark that p_{T} binning is logarithmic, since the yield of jets strongly decreases with p_{T}.

Histograms with fewer than twenty entries are disqualified for further evaluation, due to potential low quality of fits. Response histograms containing enough entries are processed in the following way. At first, the mean μ and the root square mean σ are calculated for the distribution of the values stored in a histogram. Then, the values are fitted with Gaussian within boundaries $\mu \pm 2\sigma$.

Figures 4.1 and 4.2 are given as a representative example of performed fitting process with respect to the truth jet transverse momentum p_{T}, pseudorapidity η, and the centrality of collision.

The value of the mean and the root mean square of each response fit represent the jet energy scale and its resolution, respectively. Yielded graphs for two diverse values of pseudorapidity are shown in figure 4.3. The rest of the graphs describing the jet energy scale and its resolution are attached in appendix A.1 and A.2.

Jets shown in figure 4.3 seems to be well-calibrated as response is close to unity. From the bottom graphs placed in 4.3, we can see that JER gets better as collisions are more peripheral or the jet carries higher momentum. Specially, for the low p_{T} jets in the most central collisions the resolution of response $\Re$ overcomes the value of 30 % of the mean response. However, there is still a place for a better performance. We can sort jets based on the leading parton type. In figures 4.4 and 4.5 there is a comparison of jet energy scale and jet energy resolution in Pb+Pb collisions for lower values of η between jets induced by a gluon or a quark hadronisation.

One may notice that response to gluon-induced jets is underrated after current calibration. Similarly, the response to quark-induced jets is overrated. Motivation for the following section is to identify the foundations of shown discrepancies and later to correct them in order to get better overall calibration that would lead to better JER.

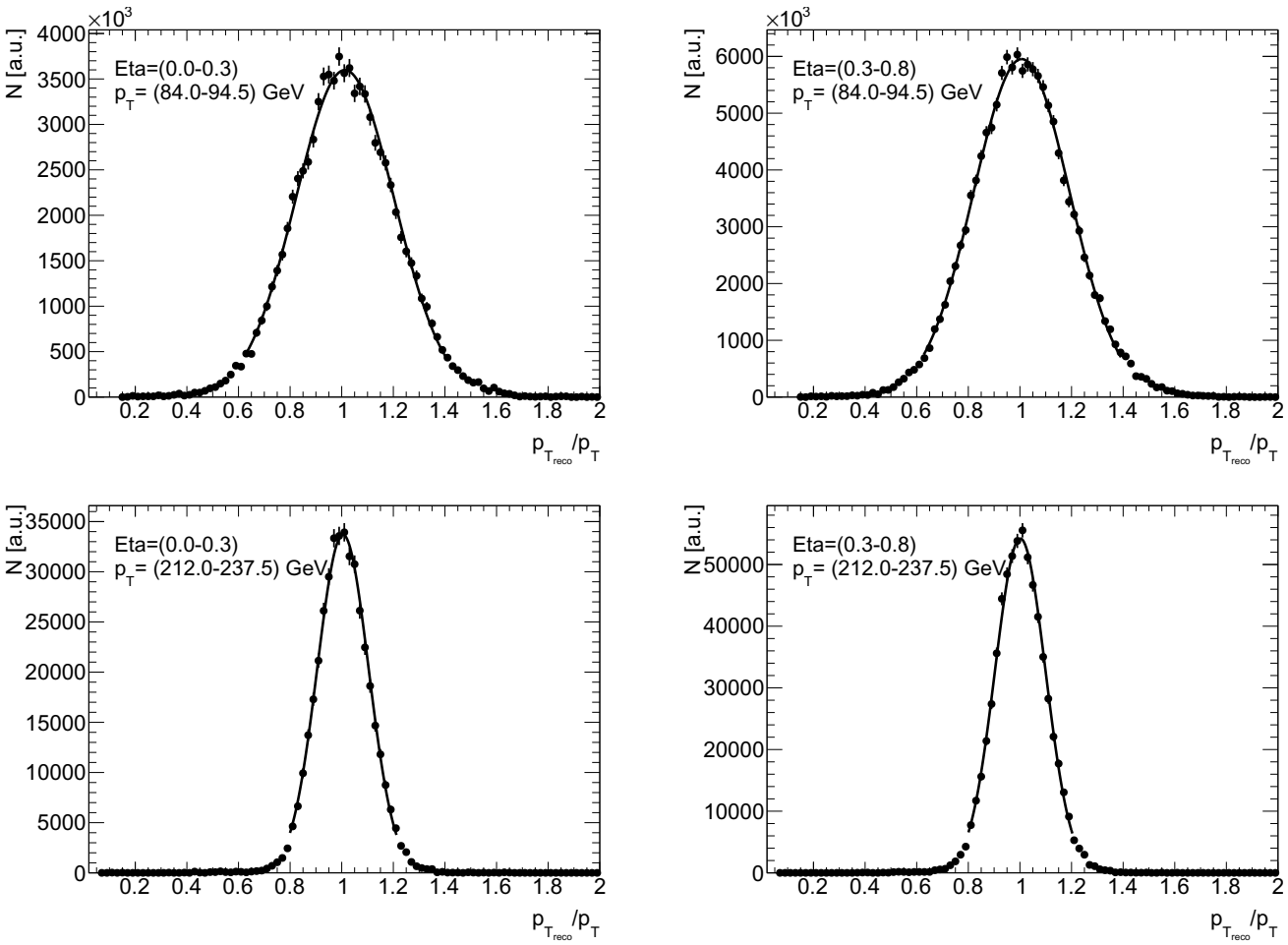

Figure 4.1: Response fits for various values of pseudorapidity η and transverse momentum p_{T} for centrality $\in (0-10)\%$.

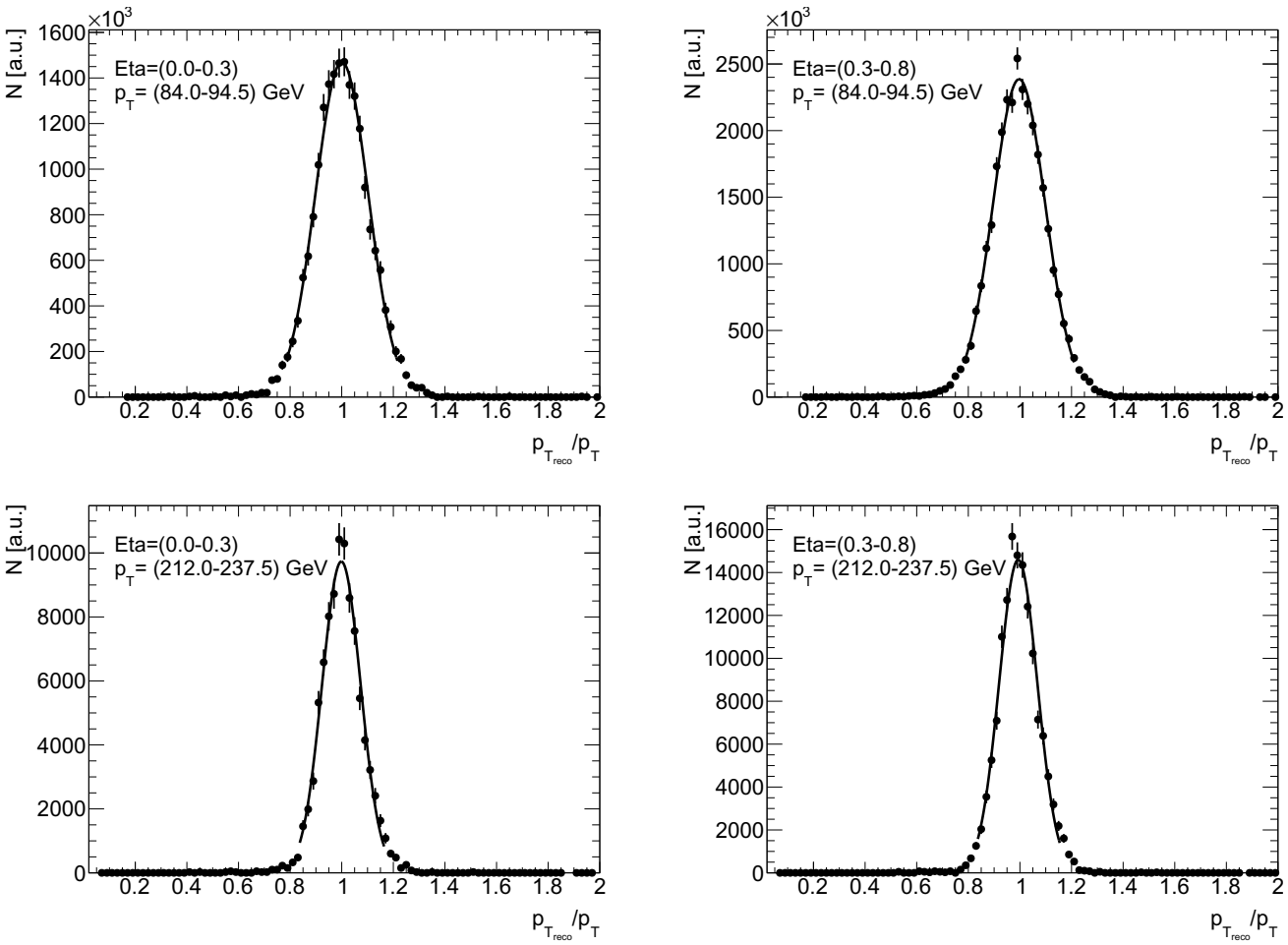

Figure 4.2: Response fits for various values of pseudorapidity η and transverse momentum p_{T} for centrality $\in (70-80)\%$.

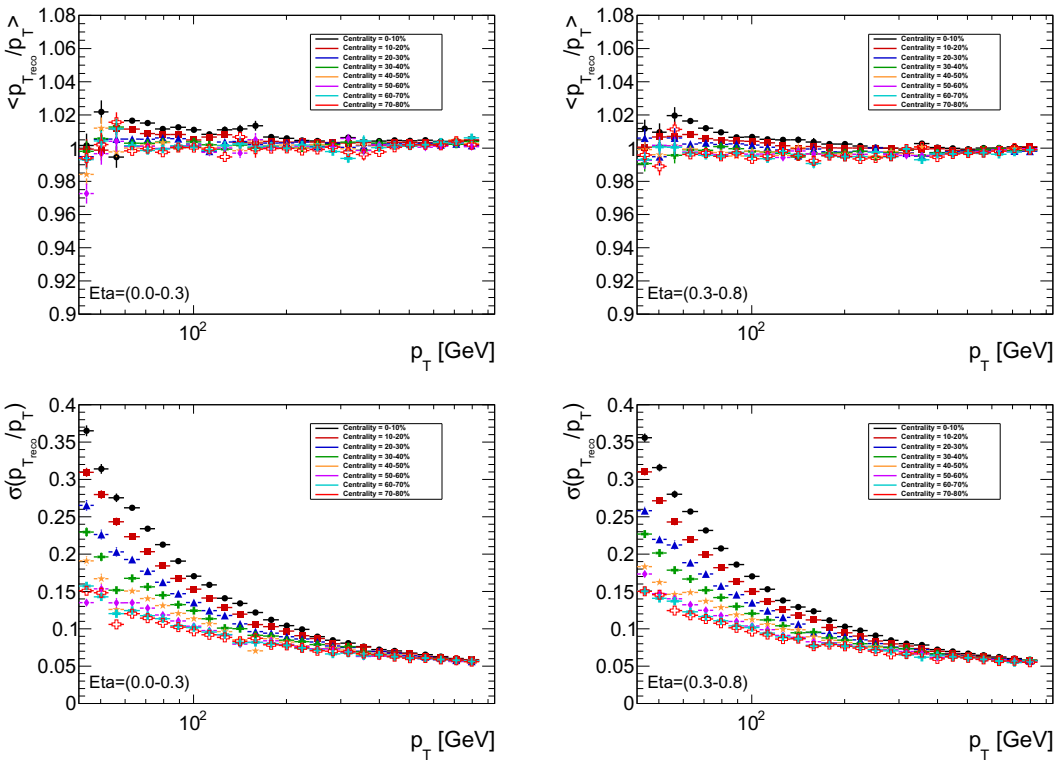

Figure 4.3: Jet energy scale and jet energy resolution in Pb+Pb collisions for central pseudorapidity η.

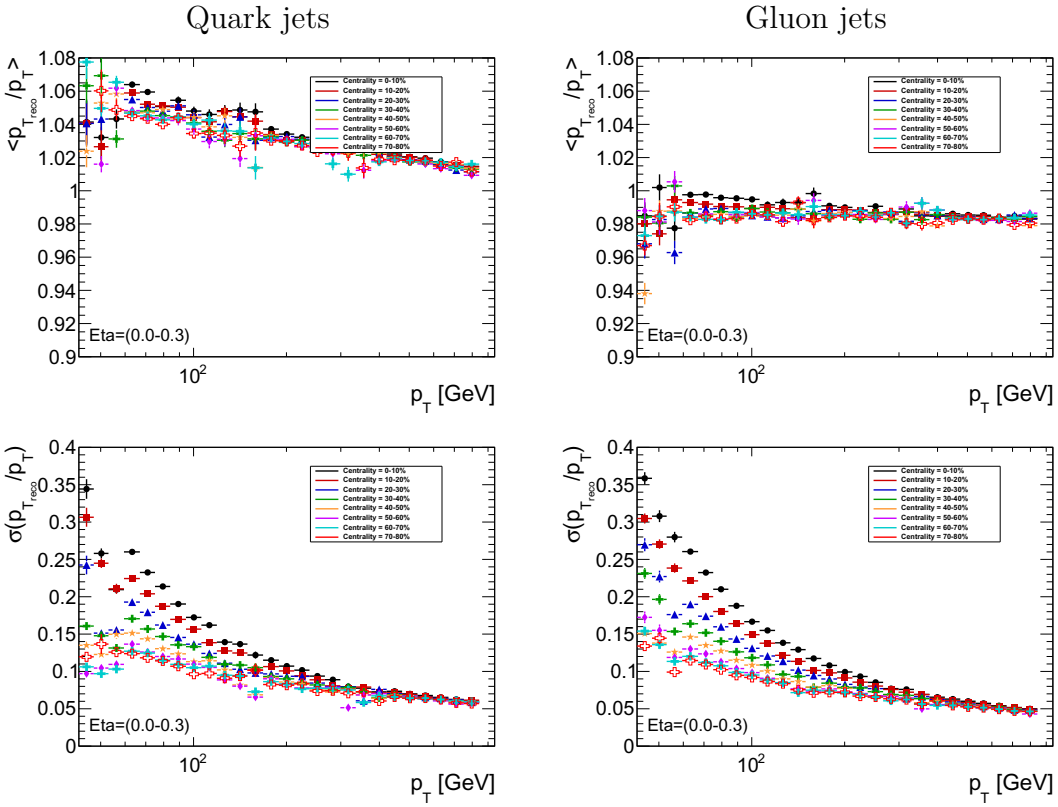

Figure 4.4: Comparison between jet energy scale and jet energy resolution in Pb+Pb collisions for quarks and gluons, $\eta \in (0.0,0.3)$.

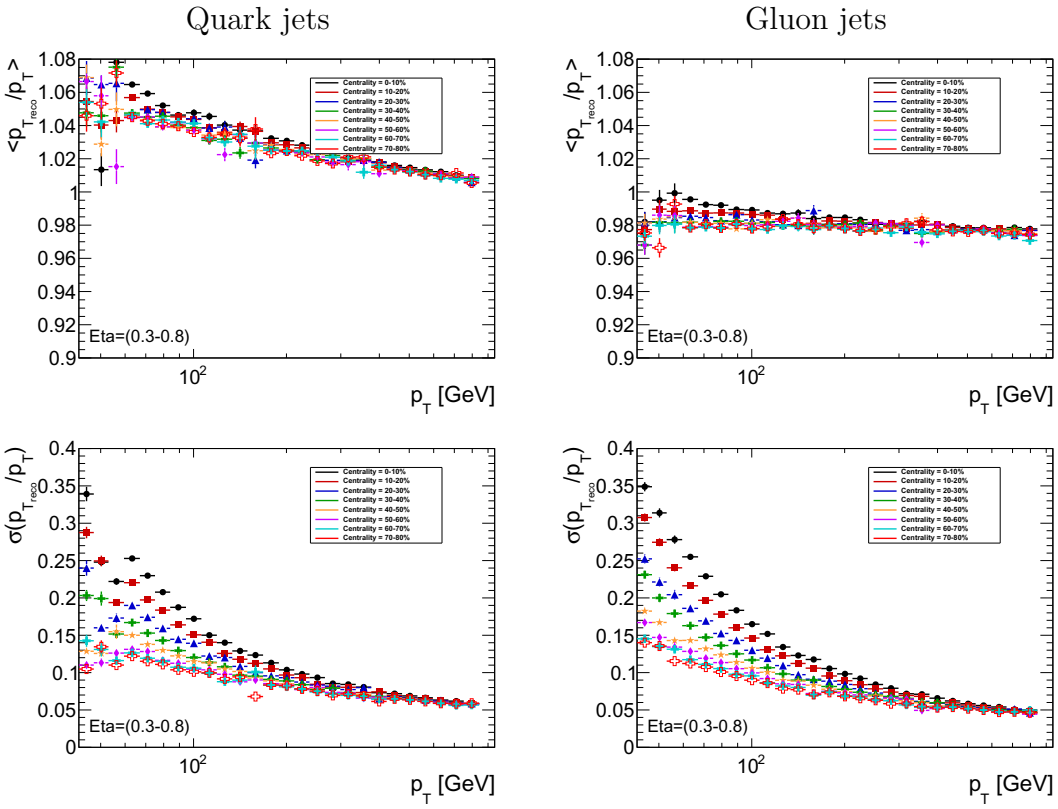

Figure 4.5: Comparison between jet energy scale and jet energy resolution in Pb+Pb collisions for quarks and gluons, $\eta \in (0.3,0.8)$.

4.2 Comparison of current approach with neural network calibration

4.2.1 Correlation study – sensitivity of jet properties to response

In the previous section, we have discussed the necessity of being able to distinguish quark and gluon-induced jets in order to obtain more precisely jet energy scale and its resolution. At first, we need to identify jet properties correlated with response. However, these properties must not be correlated with centrality. One may wonder why. The reason is that centrality truly affects response through the underlying event, as we can see for example in fig. 4.3. Nevertheless, centrality describes the collision itself, it is not a jet property.

We have put to the correlation test four following properties:

- n_{trk} := the number of tracks in a jet.
- $r_{\mathrm{trk}} := \frac{\sum_{\mathrm{trk}\in\mathrm{jet}} p_{\mathrm{T}}^{\mathrm{trk}}}{p_{\mathrm{T}}}$, where $p_{\mathrm{T}}^{\mathrm{trk}}$ represents the transverse momentum of a particle measured in the tracking subdetector. Since, p_{T} represents the transverse momentum measured in the calorimeters, it also contains the momentum of neutral particles apart from $p_{\mathrm{T}}^{\mathrm{trk}}$.
- $\mathcal{C} := \frac{p_{\mathrm{T}}^{\mathrm{reco}}|_{R=0.2}}{p_{\mathrm{T}}^{\mathrm{reco}}|_{R=0.4}}$, where $p_{\mathrm{T}}{}^{\mathrm{reco}}\,|_{R=0.2}$ is a p_{T} of $R = 0.2$ jet and $p_{\mathrm{T}}{}^{\mathrm{reco}}\,|_{R=0.2}$ is a p_{T} of $R = 0.4$ jet.

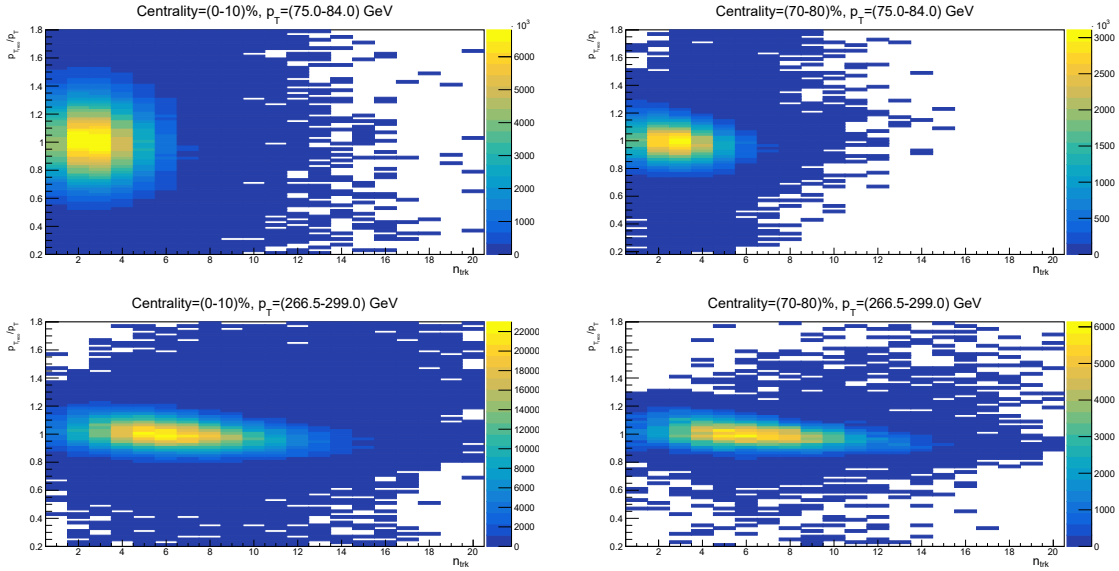

Figure 4.6: Dependence of response on n_{trk} in two centrality and p_{T} intervals.

- $\mathcal{W} := \frac{1}{n_{\mathrm{trk}}+1} \frac{\sum_{\mathrm{trk}\in\mathrm{jet}} p_{\mathrm{T}}^{\mathrm{trk}} \Delta R_{\mathrm{trk,\ jet}}}{\sum_{\mathrm{trk}'\in\mathrm{jet}} p_{\mathrm{T}}^{\mathrm{trk}'}}$; $\mathcal{W}$ is in our analysis defined only for jets with more then one track. The choice of $\mathcal{W}$ was based on [2].

.

As mentioned in section 1.5, the ID covers the pseudorapidity in range $\eta \in (-2.5, 2.5)$. Therefore, if we construct jets with a radial parameter $R = 0.4$, we must select jets with the pseudorapidity in range $\eta \in$ (-2.1,2.1). We also excluded reconstructed jets without corresponding truth jet and vice versa. The results of performed correlation study are presented in figures 4.6, 4.8, 4.10, and 4.12 in form of 2-D histogram of response $\Re$ versus the studied variable. The mean response $\Re$ as a function of a given variable is shown in figures 4.7, 4.9, 4.11, and 4.13.

From fig. 4.6 and fig. 4.7 we can see that response $\Re$ depends on n_{trk}. Moreover, the dependence of response $\Re$ on n_{trk} varies only weakly with collision centrality and p_{T}. Hence, n_{trk} is interesting for our exploration of new calibration.

Similarly, from fig. 4.8 and fig. 4.9 we can see that response $\Re$ depends on r_{trk}. Moreover, the dependence of response $\Re$ also on r_{trk} varies only weakly with collision centrality and p_{T}. Hence, r_{trk} is as well as n_{trk} interesting for our exploration of new calibration.

From fig. 4.10 and fig. 4.11 we can see that $\mathcal{C}$ is strongly correlated with centrality. Thus, we have to exclude $\mathcal{C}$ from neural network training.

From fig. 4.12 and fig. 4.13 we can see that response $\Re$ depends on the variable $\mathcal{W}$. Moreover, the dependence of response $\Re$ on $\mathcal{W}$ varies only weakly with collision centrality and p_{T}.

One may notice, in correlation studies for all four jet properties, that the distribution of response for centrality 0-10% is wider than for centrality 70-80%. These results are in agreement with our study of inclusive jet energy resolution because of higher underlying event contribution in more central events.

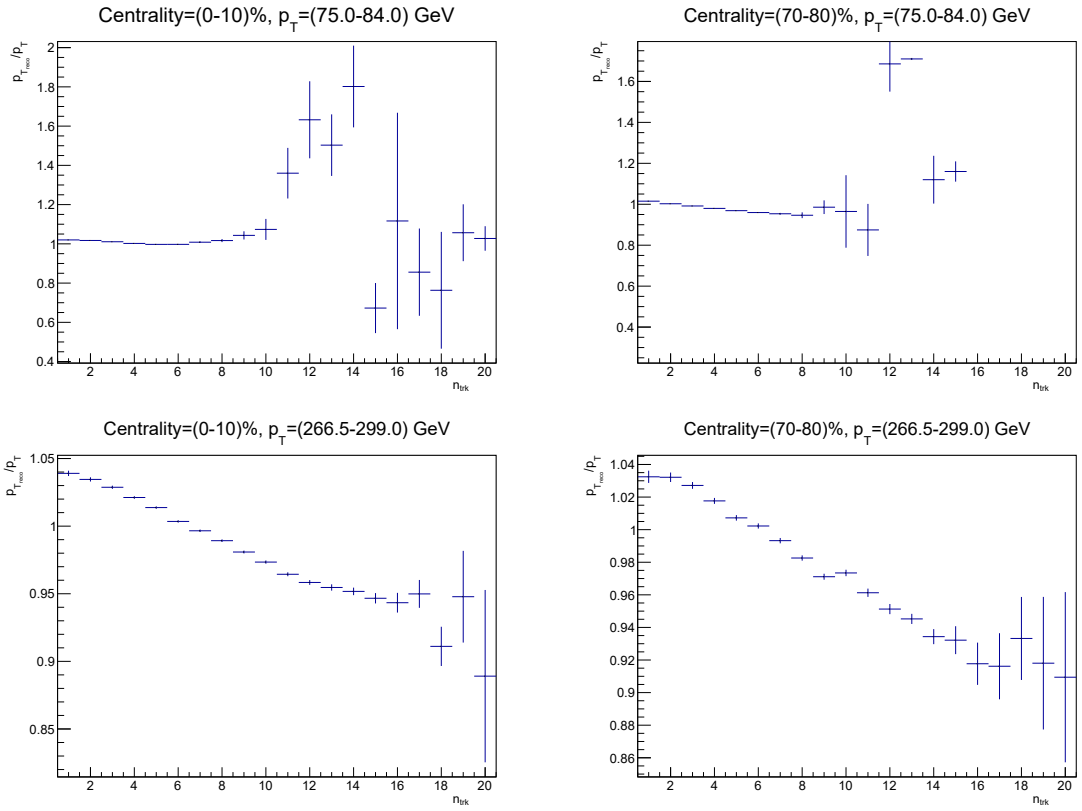

Figure 4.7: Dependence of response on n_{trk} in two centrality and p_{T} intervals as profile.

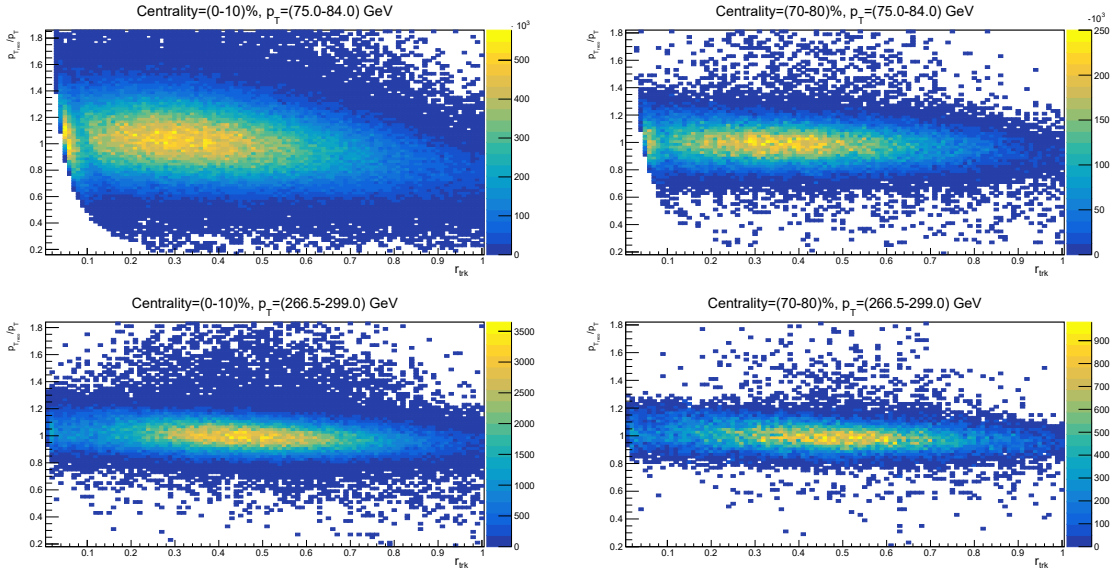

Figure 4.8: Dependence of response on r_{trk} in two centrality and p_{T} intervals.

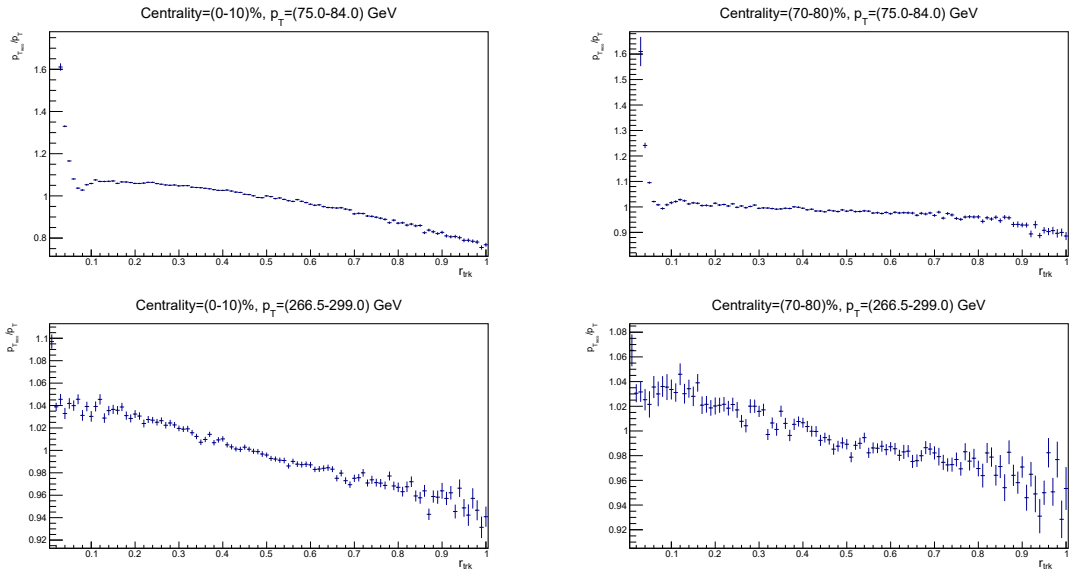

Figure 4.9: Dependence of response on r_{trk} in two centrality and p_{T} intervals as profile.

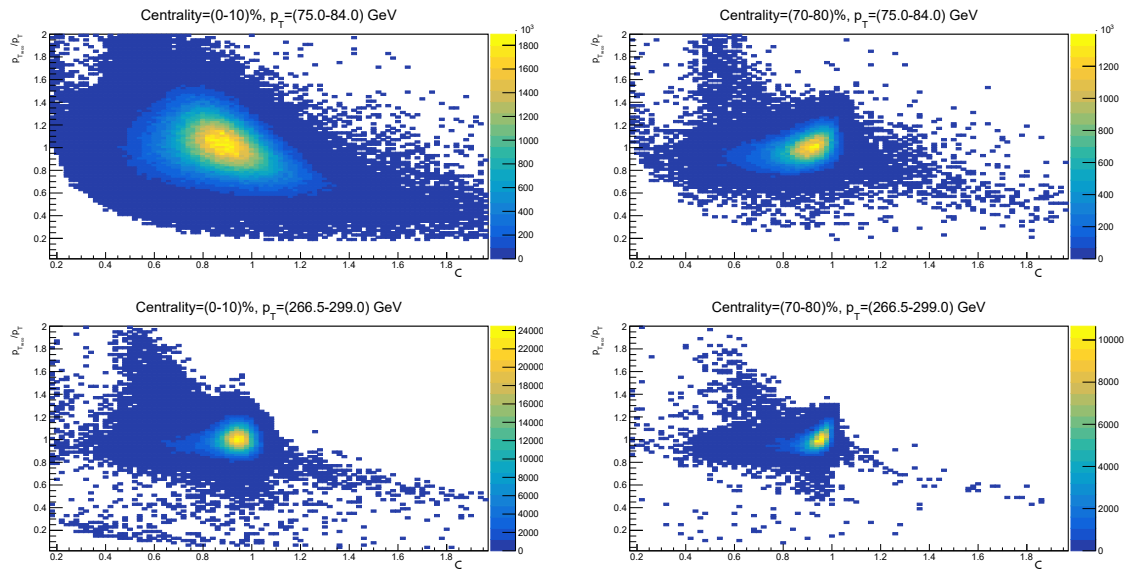

Figure 4.10: Dependence of response on $\mathcal{C}$ in two centrality and p_{T} intervals.

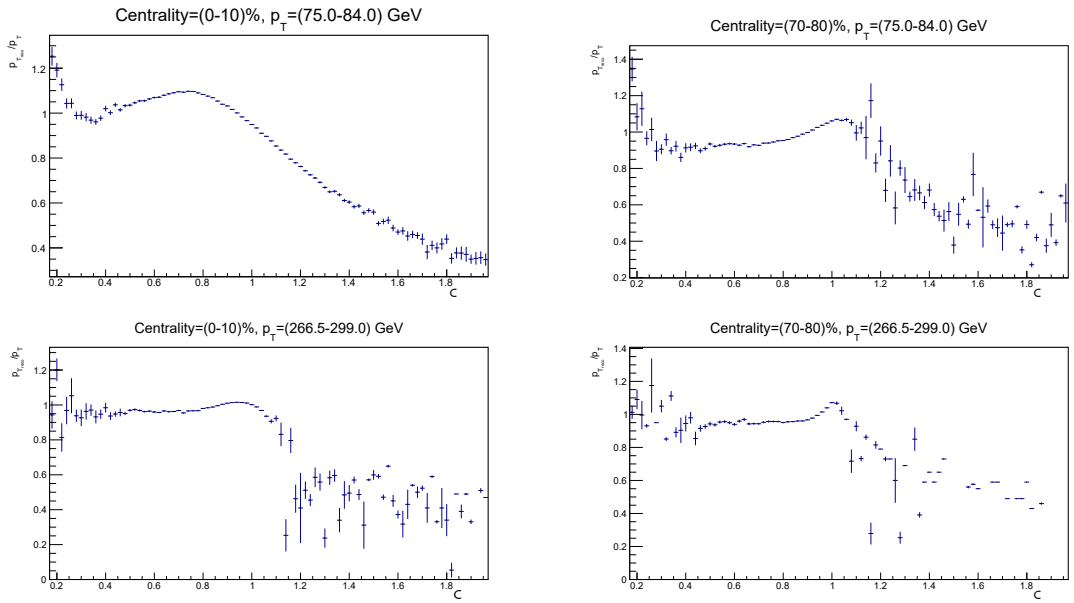

Figure 4.11: Dependence of response on $\mathcal{C}$ in two centrality and p_{T} intervals as profile.

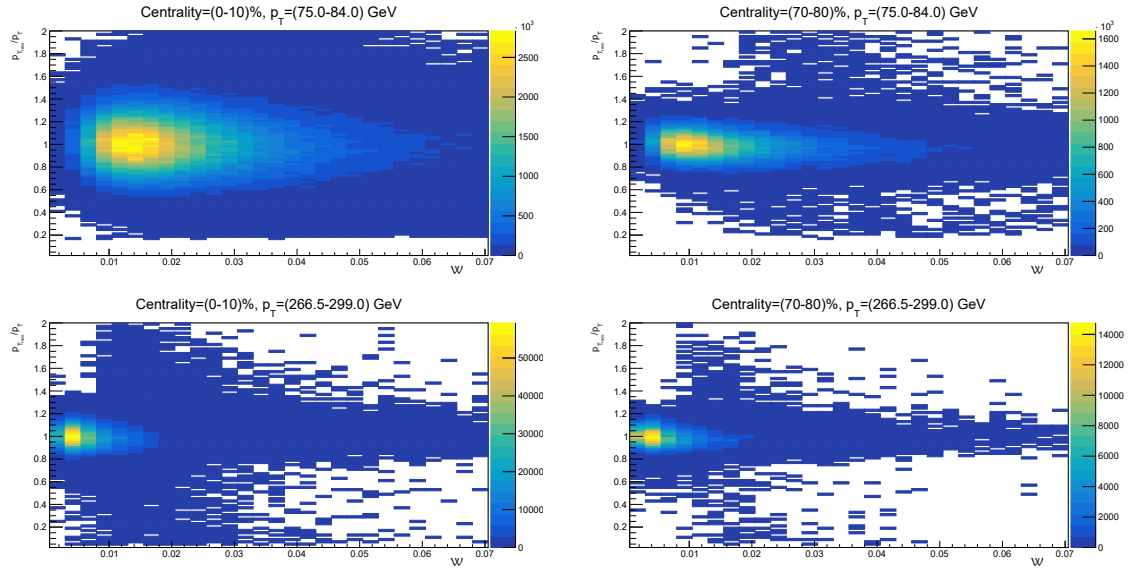

Figure 4.12: Dependence of response on $\mathcal{W}$ in two centrality and p_{T} intervals.

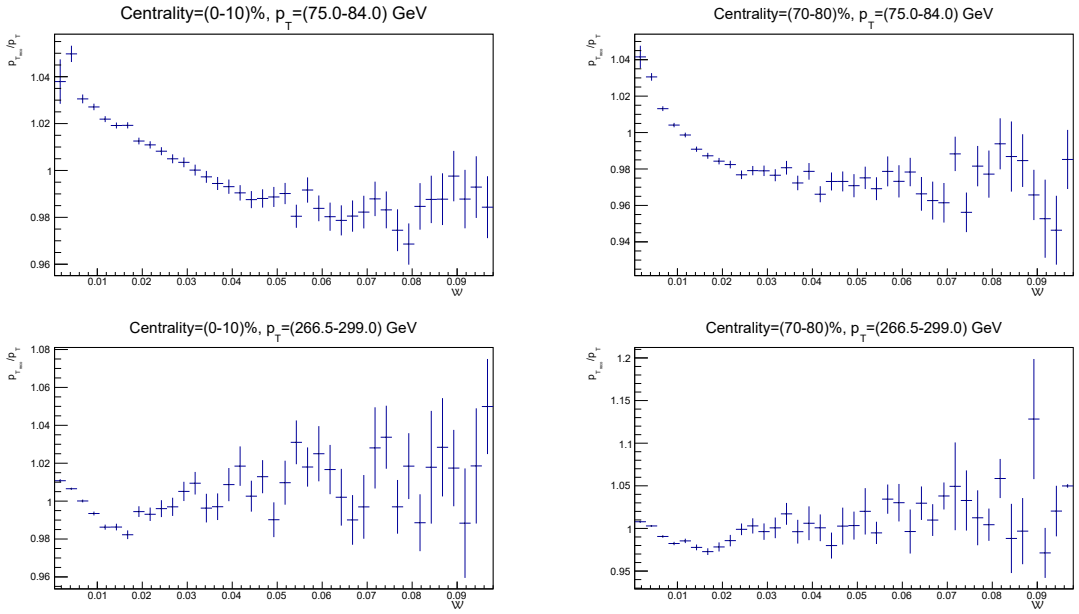

Figure 4.13: Dependence of response on $\mathcal{W}$ in two centrality and p_{T} intervals as profile.

4.2.2 The study of neural network

The studies of the behavior of neural networks were performed in The Regional Computing Center for Particle Physics in Prague - the Golias farm. Details about the machines used at Golias farm properties can be found in [47]. During a power outage at the Golias farm it ran on the IPNP twelve i7-980 cores server.

Root MLP class provides six different back-propagation algorithms, which definitions can be found in [3]. The first task was to test them and decide which one is the most suitable for our research goals. For the purpose of choosing an appropriate algorithm, we have prepared three types of samples containing 50 thousand events, 250 thousand events, and 2.5 million events. For each sample, we studied the performance of error $E(w)$ as the number of so-called epochs increases. An epoch is a period when an entire dataset passes once through the neural network. The initial value of the error function varies according to the geometry of the network. Therefore, it makes sense to compare individual networks in different setups at first only qualitatively. We have found out that the time of convergence rises with the number of neurons and the amount of training data. However, as the amount of data increased, the resulting value of the error function decreased. All these observations met our expectations and the general characteristics of neural networks.

The six teaching methods in the studied library can be divided into two groups, stochastic and steepest descent algorithms. Both groups were introduced in section 3.3. Used algorithms may be further tuned by a change of internal parameters. However, a few tests performed with a different parameters setup did not show significant changes of error function convergence. Thus, we kept the default parameters as they were initialised in the library definition. It would certainly be possible to achieve even better results by changing these parameters, but detailed research of their influence would require considerable computational power and

time.

Significantly better results were given by gradient descent methods compared to stochastic ones. For example, the region where $p_{\mathrm{T}} \in$ 180-360 GeV for training 1.1 million events, as well as 1.1 million testing events. Inputs were $p_{\mathrm{T}}^{\mathrm{reco}}, \mathcal{W}$, the target was p_{T}. The first hidden layer contained 40 neurons, the second one 80 neurons. All variables were normalised. After 20 hours of real-time run, gradient descent algorithms gave about 15 % lower value of error function in comparison to stochastic algorithms.

In accordance with the previous paragraph the fastest convergence for given training time and the lowest error function value was provided by Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS). BFGS, like many other algorithms, improves the gradient descent method by replacing equidistant steps with steps affected by the information of curvature obtained from a Hessian matrix. Moreover, it replaces the Hessian matrix with gradual approximation using secants instead of matrix equations. More details about the BFGS algorithm can be found in [48]. The computational complexity of matrix equations is $\mathcal{O}(n^3)$, contrary to the approximations used in BFGS, which require only complexity $\mathcal{O}(n^2)$. Thus, in such case the computational complexity played an extremely important role.

Let us also mention the importance of normalization as a last general note on the neural network setup optimisation. During the network testing, we confirmed that normalization of data is necessary. Without it, stochastic methods did not work, and those associated with the gradient descent method converged much more slowly and to a higher value of the error function. No modification has improved the behavior of the network more than the introduction of the normalization.

In this paragraph we present the properties of the used training sample. The training sample was created from simulated proton-proton collisions. We were limited by computing power, thus we explored and provided proof of concept for $p_{\mathrm{T}}^{\mathrm{reco}} \in (0, 422.5)$ GeV, which is the region used in heavy-ion analyses. The limitations of the training sample are the same as those applied in the correlation study. Furthermore, the restriction on p_{T} to be less than 360 GeV. A slightly different setup in the parameters was used for the n_{trk} study, which will be pointed out at the appropriate place. Originally, the selection was applied on $p_{\mathrm{T}}^{\mathrm{reco}}$, because that is the only information available in the experimental data. However, the outputs from the neural network had a strongly biased mean response as shown in figure 4.14.

Such a selection would devalue the training sample. So another approach was used where a threshold on p_{T} and subsequent two times higher threshold on $p_{\mathrm{T}}^{\mathrm{reco}}$ are applied to remove outliers and keep the ability to perform normalisation. This selection criterion is justified because the fitting of uncorrected data in order to obtain JES and JER does not take place in a region exceeding the value of response approximately 1.8. We also divide both variables $p_{\mathrm{T}}^{\mathrm{reco}}$ and p_{T} by the value of this second threshold in order to normalize the data. Then, we obtain a response with a mean close to unity, which is exactly what we demand from the training sample.

Even if we had enough computation power, we would not be able to train the network over the entire range of momentum p_{T}. In general, it can be assumed that

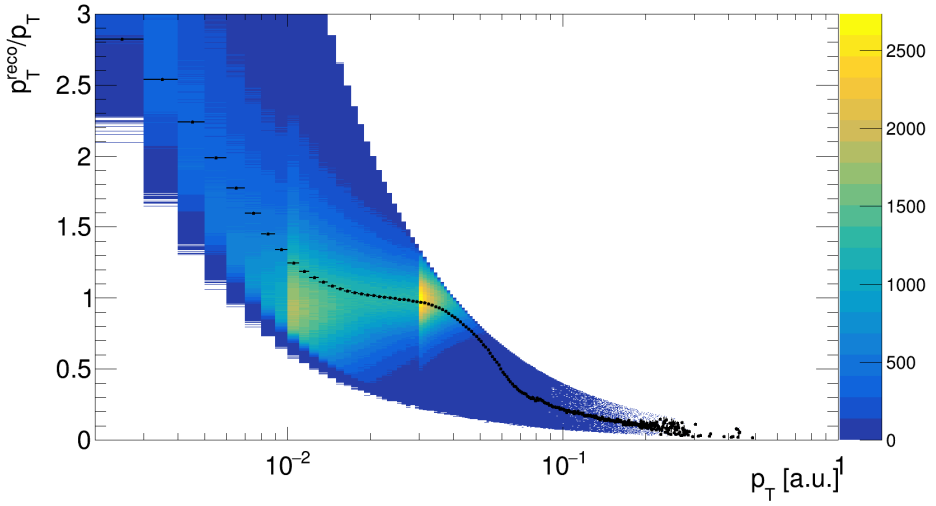

Figure 4.14: Response $\Re$ as function of $p_{\mathrm{T}}^{\mathrm{reco}}$. The mean of the response is visualised by points with a corresponding error.

different physical processes are dominant on different sections of the p_{T} spectrum. If we divide the p_{T} scale in the right way, we might get a more accurate calibration.

Once the error function was only decreasing negligibly we stopped the training and plotted the graphs characterising the performance of the NN. It is good to realize that we are not even trying to squeeze the error function to zero as there are free parameters, which we do not study. There exist two potential issues with machine learning: 1) The NN can be undertrained. This case is demonstrated in the right panel of figure 4.15; 2) The NN can be overtrained. This case is demonstrated in the left panel of figure 4.15.

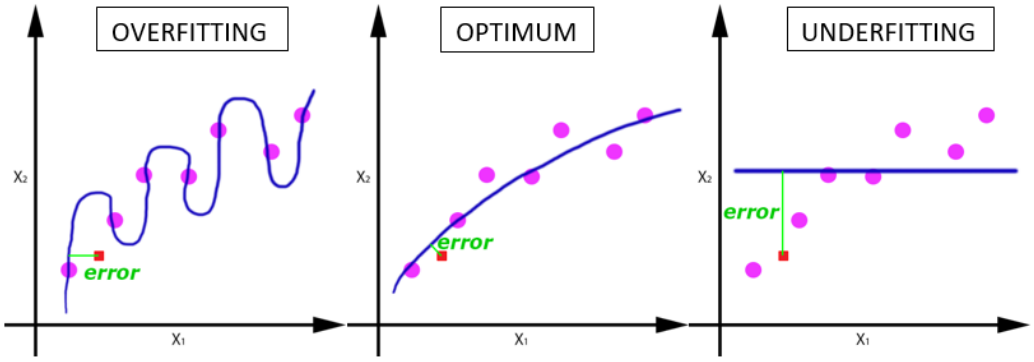

Figure 4.15: A demonstration of pathological phenomena in neural network training on the problem of simple fit. Taken from [49].

In order to avoid described issues with training the NN, we analysed the performance of each trained network. The MLPAnalyzer class [50] provides a set of graphs after the training process that allows to characterise the performance of the NN. An example of graphs to analyse the performance of the NN is given in figure 4.16.

The upper left panel shows how much the neural network changes when one

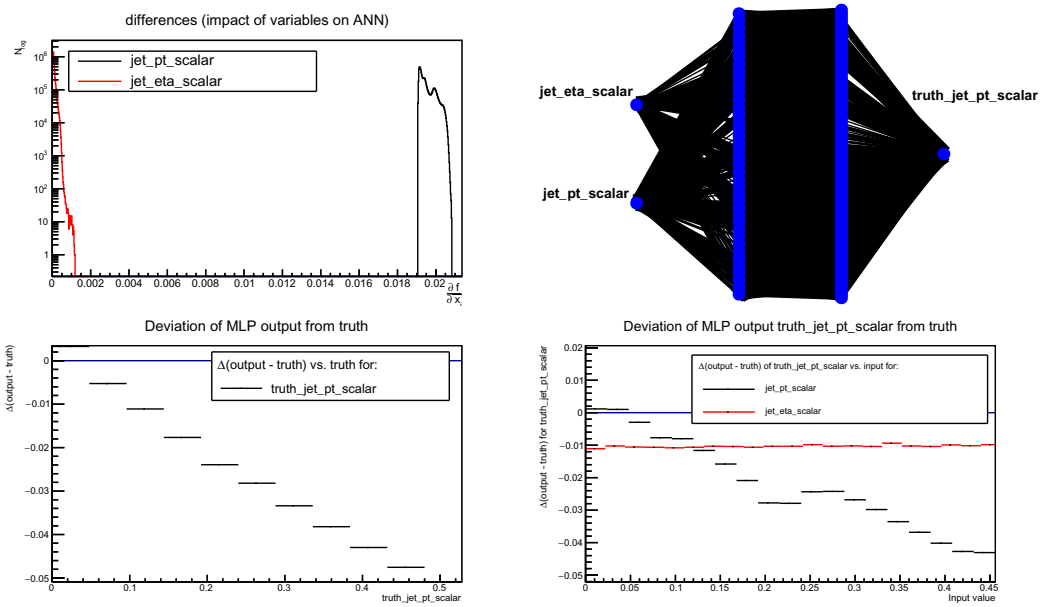

Figure 4.16: Analysis chart, taken from $|\eta|$ based neural network study.

of the neural network inputs is varied around its nominal value. Such an operation is done for each entry, as a result, distribution is produced. The produced distribution approximates the derivative of the NN distribution with respect to each input. The upper right panel shows the structure of a trained neural network and the found connections with appropriate weights. It can be used for a simple diagnosis. For example, it is suspicious if the network has only a few connections. Then, probably a sufficient degree of convergence was not reached or the training data set is not suitable. The network shown in the picture does not exhibit any obvious pathologies. The bottom left panel displays the average change of the target after the use of the neural network. The bottom right panel shows the change of the target as the function of each input. The same analysis was performed for each neural network investigated. The most of the setups of the trained network were rejected as inapplicable at this point of research.

The graphs on the figure 4.16 can be used to discuss, whether pseudorapidity is a suitable variable for correcting the response. The response dependence on pseudorapidity due to the design of the detector is significant and non-trivial for uncalibrated jets. However, the default calibrations were used before testing the neural network. Thus, the results can be also interpreted as the check, whether the response dependence on pseudorapidity is corrected sufficiently in default calibration. From the top left and the bottom right picture in 4.16 we can see that η^{reco} affects response $\Re$ negligibly.

If the neural network passed the analysis successfully, we used the network to correct the lead ion collision data. To be precise, we ran the correction obtained from the neural network on a subset of jets that met the same criteria as those in the proton-proton training set. Initially, we trained networks with only two inputs. Then we moved on to the networks with three input variables. We obtained two neural networks worth mentioning. The first one was trained on 8.06 million proton-proton events for training as well as for testing such that $p_{\text{T}} < 360$ GeV and $p_{\text{T}}^{\text{reco}} < 720$ GeV and $|\eta_{\text{reco}}| < 2.1$. Jets where $\mathcal{W}$ = -1

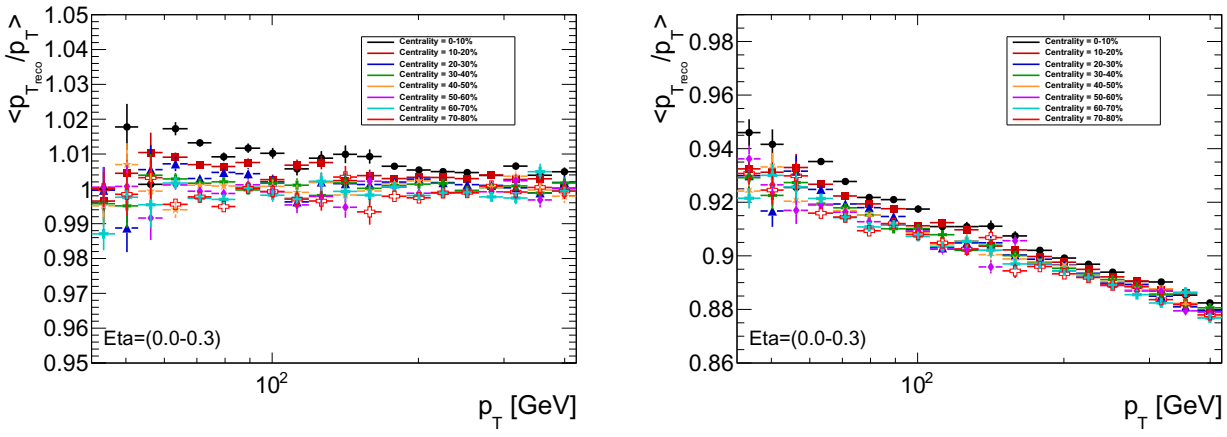

Figure 4.17: The comparison of JES for uncorrected jets (left) and jets corrected (right) by neural network, which depends on r_{trk} and $\mathcal{W}$. The jets are restricted to $|\eta| < 0.3$.

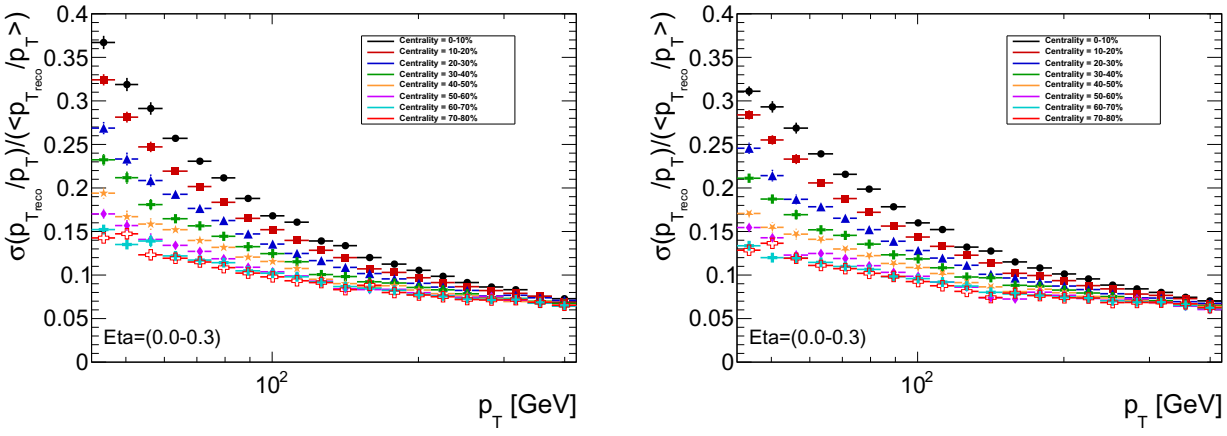

Figure 4.18: The comparison of JER scaled by JES for uncorrected jets (left) and jets corrected (right) by neural network, which depends on r_{trk} and $\mathcal{W}$. The jets are restricted to $|\eta| < 0.3$.

were also excluded from the training sample. Inputs were $p_{\mathrm{T}}^{\mathrm{reco}}$, r_{trk}, and $\mathcal{W}$. The target was p_{T}. The training was performed by the BFGS algorithm for 20 epochs. Uncorrected JES is placed on the left, corrected JES on the right in figure 4.17.

As one can see from figure 4.17 the used NN changes the average response and additional a numerical inversion need to be applied after the correction. The numerical inversion transforms the $p_{\mathrm{T}}^{\mathrm{reco}}$ in the way such that the response to be equal to one. It is beyond the scope of this work to prepare such a numerical inversion. However, in order to create a clear comparison between the uncorrected data and data on which the correction obtained from a neural network was applied, the value of resolution was scaled by the value of mean response in each p_{T} bin. Obtained results of JER are displayed in figure 4.18. JER with original calibration is placed on the left, corrected JER on the right side of fig. 4.18.

Obtained JES and JER for the absolute intervals of pseudorapidity in range of 0.3 to 2.1 can be found in the appendix A.3 and A.4, respectively.

Fig. 4.19 presents examples of the response distribution for completeness.

Three p_{T} slices representing low, medium, and higher p_{T} jets with $|\eta| < 0.3$ in centrality 0-10 % interval are shown. The jets with default calibration are placed on the right, the corrected jets on the left.

Quantitative comparison between jets with and without NN-based calibration with $|\eta| < 0.3$ in centrality 0-10 % interval is given in table 4.1.

p_{T} [GeV]	default JER	cor. JER	cor./default JER [%]
42.5-47.5	0.3671±0.008	0.311±0.006	84.7±1.5
47.5-53.0	0.3188±0.008	0.293±0.007	91.9±1.7
53.0-59.5	0.2912±0.007	0.269±0.006	92.3±1.7
59.5-67.0	0.257±0.002	0.239±0.001	93.1±0.4
67.0-75.0	0.231±0.001	0.216±0.001	93.5±0.3
75.0-84.0	0.217±0.001	0.199±0.001	93.9±0.4
84.0-94.5	0.188±0.001	0.178±0.001	94.9±0.4
94.5-106.0	0.168±0.002	0.160±0.001	95.2±0.5
106.0-119.0	0.161±0.002	0.152±0.002	94.7±0.7
119.0-133.5	0.139±0.002	0.132±0.002	94.9±0.7
133.5-149.5	0.134±0.003	0.128±0.002	95.4±0.9
149.5-168.0	0.120±0.002	0.115±0.002	95.8±0.8
168.0-189.0	0.1126±0.0003	0.1081±0.0003	96.1±0.1
189.0-212.0	0.1055±0.0004	0.1013±0.0003	96.0±0.2
212.0-237.5	0.0986±0.0005	0.0956±0.0005	97.0±0.2
237.5-266.5	0.0914±0.0006	0.0885±0.0005	96.8±0.3
266.5-299.0	0.0863±0.0007	0.0841±0.0007	97.5±0.3
299.0-335.5	0.083±0.001	0.0801±0.0009	96.3±0.5
335.5-376.5	0.075±0.001	0.074±0.001	98.9±0.6
376.5-422.5	0.073±0.001	0.0703±0.0009	96.5±0.6

Table 4.1: JER (scaled by JES) comparison in various p_{T} slices between default jets and jets corrected by the NN using $\mathcal{W}$ and r_{trk}. The jets are restricted to $|\eta| < 0.3$ and centrality 0-10 %.

The second analysed neural network was trained on 5.46 million proton-proton events for training as well as for testing such that $p_{\mathrm{T}} < 400$ GeV, $\Re < 2$ and $|\eta_{\mathrm{reco}}| < 2.1$. Inputs were $p_{\mathrm{T}}^{\mathrm{reco}}$, r_{trk}, and n_{trk}. The target was p_{T}.The training was performed by BFGS algorithm for 30 epochs. This second approach differs in the input and in the jet selection. Training and validation samples are now contain data where $n_{\mathrm{trk}} = 0$. Contrary, such a case had to be excluded for neural network training using $\mathcal{W}$, inasmuch as it behaves nonlinearly in the studied region.

Uncorrected data are placed on the left panel, corrected on the right panel of figure 4.20.

One can see that JES in figure 4.20 is similar to the JES resulting from the first neural network. Thus, a numerical inversion would need to be applied after the correction as well. Obtained results of JER scaled by JES are displayed in figure 4.21 The jets with the default calibration are placed on the left panel, the jets corrected by the NN are placed on the right panel of fig. 4.21.

Obtained JES and JER for the absolute intervals of pseudorapidity in the range of 0.3 to 2.1 for the second neural network can be found in the appendix A.5, and A.6, respectively.

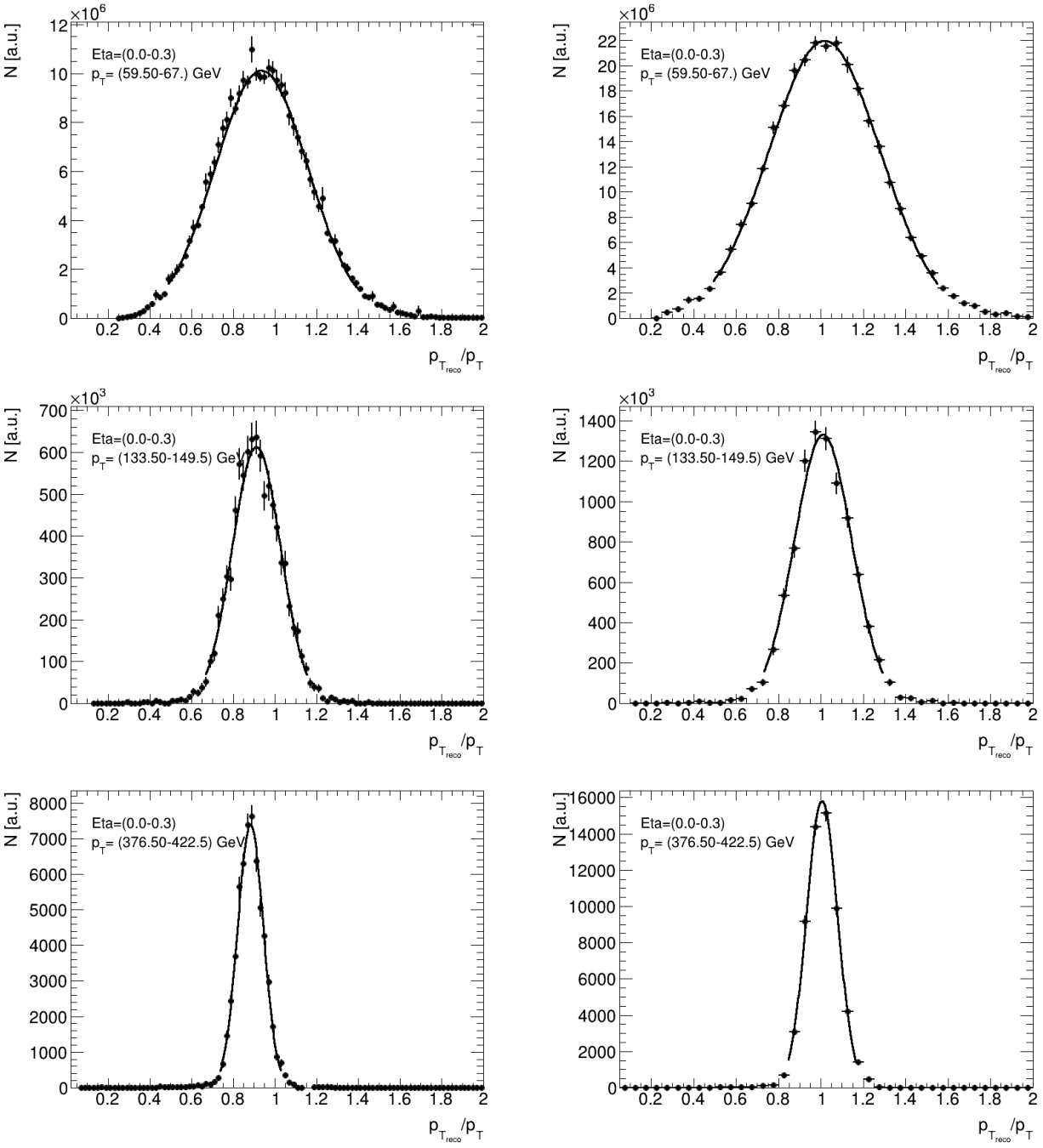

Figure 4.19: The comparison of the response $\Re$ distributions for the jets with default calibration (right) and jets corrected (left) by neural network, which depends on r_{trk} and $\mathcal{W}$. The jets are restricted to $|\eta| < 0.3$ and centrality 0-10 %.

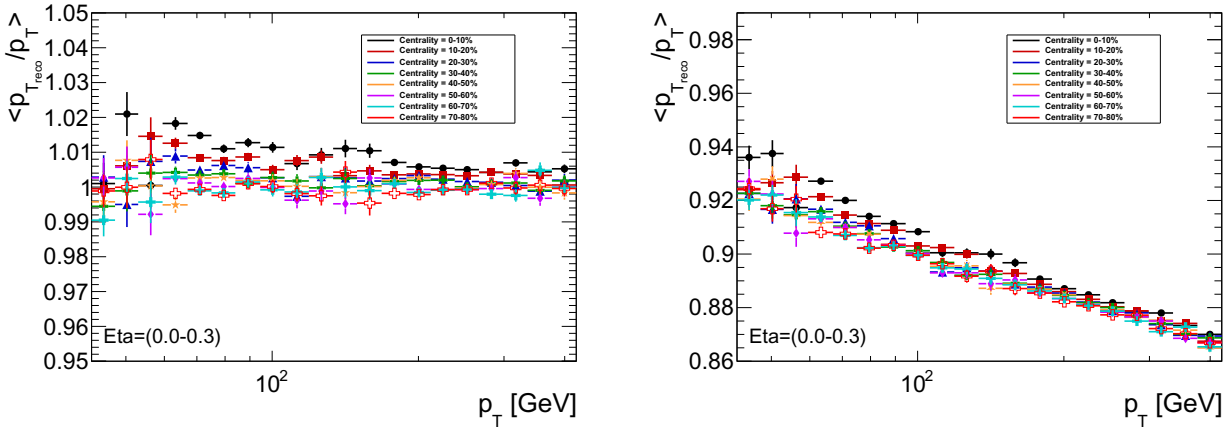

Figure 4.20: The comparison of JES for uncorrected jets (left) and jets corrected (right) by neural network, which depends on r_{trk} and n_{trk}. The jets are restricted to $|\eta| < 0.3$.

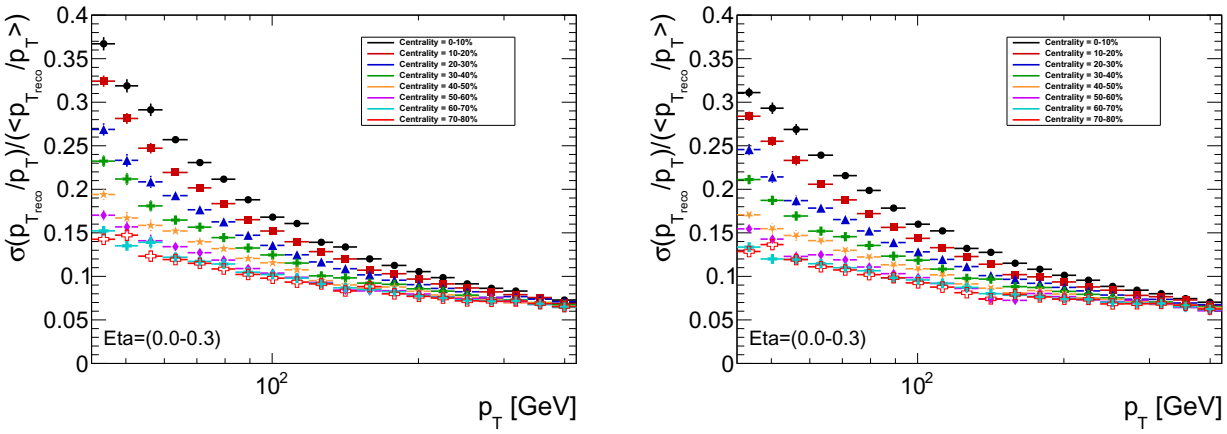

Figure 4.21: The comparison of JER scaled by JES for uncorrected jets (left) and jets corrected (right) by neural network, which depends on r_{trk} and n_{trk}. The jets are restricted to $|\eta| < 0.3$.

Fig. 4.22 presents examples of the response distribution for completeness. Three p_{T} slices representing low, medium, and higher p_{T} jets with $|\eta| < 0.3$ in centrality 0-10 % interval are shown. The jets with default calibration are placed on the right, the corrected jets on the left.

Table 4.2 presents qualitative comparison of improved resolution for centrality 0-10 % and pseudorapidity up to 0.3.

p_{T} [GeV]	default JER	cor. JER	cor./default JER [%]
42.5-47.5	0.366±0.007	0.299±0.005	81.9±1.3
47.5-53.0	0.319±0.007	0.281±0.006	88.0±1.6
53.0-59.5	0.289±0.007	0.258±0.006	89.5±1.6
59.5-67.0	0.256±0.002	0.232±0.001	90.3±0.4
67.0-75.0	0.231±0.001	0.2093±0.0009	90.8±0.3
75.0-84.0	0.212±0.001	0.192±0.001	90.9±0.4
84.0-94.5	0.188±0.001	0.173±0.001	92.0±0.4
94.5-106.0	0.168±0.002	0.155±0.001	92.5±0.5
106.0-119.0	0.161±0.002	0.148±0.002	92.1±0.6
119.0-133.5	0.139±0.002	0.127±0.002	91.5±0.7
133.5-149.5	0.134±0.003	0.121±0.002	90.8±0.9
149.5-168.0	0.120±0.002	0.110±0.002	91.5±0.8
168.0-189.0	0.1126±0.0003	0.1036±0.0003	92.0±0.1
189.0-212.0	0.1056±0.0004	0.0965±0.0003	91.4±0.2
212.0-237.5	0.0987±0.0005	0.0908±0.0004	92.0±0.2
237.5-266.5	0.0915±0.0006	0.0840±0.0005	91.8±0.2
266.5-299.0	0.0865±0.0007	0.0790±0.0006	91.4±0.3
299.0-335.5	0.083±0.001	0.0757±0.0008	90.9±0.4
335.5-376.5	0.075±0.001	0.0702±0.0009	93.3±0.6
376.5-422.5	0.073±0.001	0.0662±0.0009	90.9±0.5

Table 4.2: JER (scaled by JES) comparison in various p_{T} slices between default jets and jets corrected by the NN using n_{trk} and r_{trk}. The jets are restricted to $|\eta| < 0.3$ and centrality 0-10 %.

If one compares data in table 4.1 and 4.2, then one can see that response $\Re$ is sensitive to $\mathcal{W}$ at low p_{T}, while response $\Re$ sensitivity to n_{trk} holds over all p_{T} scale. The dependencies of response $\Re$ can be also seen in figure 4.23, which compares the parts of an analysis of both neural networks. One can there note also that $\mathcal{W}$ ranges only from 0.0 to 0.2, better normalization in future research of this variable could also improve the overall behavior of the first presented neural network.

Let us also point out the general validity of obtained correction as the neural network was trained on proton-proton collisions data, but used to correct lead-lead collision data. Therefore, there is a low risk of bias resulting from excessive similarity of the training data set and verification data. However, it is clear that it is necessary to verify the found calibration through as much lead-lead data as possible.

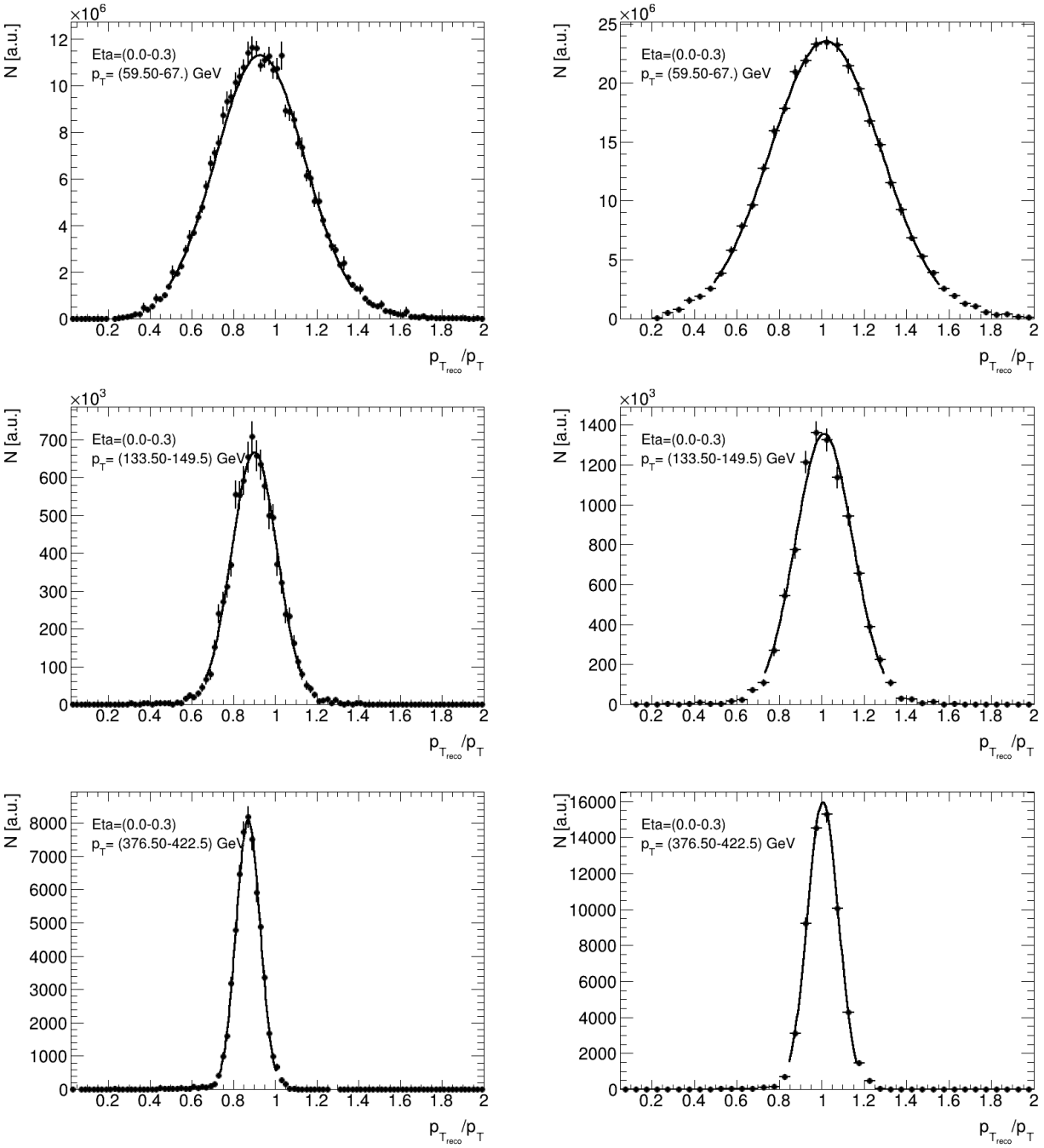

Figure 4.22: The comparison of the response $\Re$ distributions for the jets with default calibration (right) and jets corrected (left) by neural network, which depends on r_{trk} and n_{trk}. The jets are restricted to $|\eta| < 0.3$ and centrality 0-10 %.

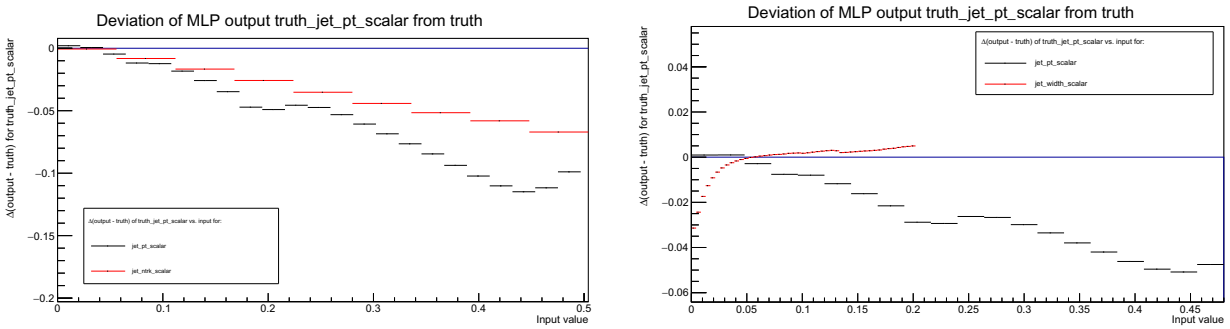

Figure 4.23: The comparison of the change of p_{T} based on the values of n_{trk} (left) and $\mathcal{W}$ (right). All values are normalised.

Conclusion

The jets used in physics analysis, both in proton-proton and in Pb+Pb collisions at the ATLAS experiment, utilised numerical inversion as the default calibration method to account for the different response of particle showers of various types and for the detector non-uniformities. We confirmed that the currently used calibration leads to the mean response of the detector being approximately equal to unity. Furthermore, we proved that currently used calibration sufficiently compensates the dependence of response $\Re$ on pseudorapidity η. We also showed that the mean response is different for quark-induced and gluon-induced jets. This information opens up space for finding a better correction if we can distinguish these groups of physical states. This work tries to identify jet properties for training the neural network which is sensitive to calorimeter response.

The major result of the study was a pair of neural networks that significantly improve the calibration leading to overall improvement of the performance of jet reconstruction. The first neural network depends on $p_{\mathrm{T}}^{\mathrm{reco}}$, $\mathcal{W}$ and r_{trk}. This network improved the resolution by approximately 10 % for low values of p_{T} and by 5 % for higher values of p_{T}. The second neural network depends on $p_{\mathrm{T}}^{\mathrm{reco}}$, n_{trk} and r_{trk}. It improved the resolution by approximately 10 % on the whole p_{T} scale, with even higher improvement for the lower values of p_{T}.

We succeeded in the search for a correction that can significantly improve the current calibration. We did not exclude some problematic regions from our training data, which leaves unnecessary degrees of freedom in problem-free areas. For example, there is no lower threshold on $p_{\mathrm{T}}^{\mathrm{reco}}$ nor p_{T}. The jet properties using tracking information in this low p_{T} region are more affected by the presence of fake tracks. Removing these jets from the training sample could lead to further improvements in neural network calibration.

Another direction of the future research offers the opportunity to explore the possibility of jet classification, which should be implemented very similarly. We believe that n_{trk} and $\mathcal{W}$ could be suitable quantities for this purpose with regard to the physical properties of jets and the ability to improve the response resolution over the entire studied range of momentum $p_{\mathrm{T}}^{\mathrm{reco}} \in (0, 422.5)$ GeV.

The next logical step in the study of the usage of neural networks in search for improved jet calibration is using SciKit Python library instead of ROOT MLP class as SciKit provided better documentation, variability in the setup of the neural network as well as a vast developer community. This library also provides networks with the ReLU activation function, which has been used more often in recent years than sigmoid, because its derivatives are either 0 or 1, which facilitates calculations. Another important step will be to get in touch with neural network training using GPUs provided by the cluster LXPLUS at CERN or tensor processing units (TPU), which are integrated circuits developed specifically for neural network machine learning. This improvement in hardware utilization could speed up the learning process by 2-3 orders of magnitude. It can also be perceived that we could take a significantly larger training data set at the same time.

Bibliography

[1] Miklos Gyulassy and Larry McLerran. New forms of QCD matter discovered at RHIC. *Nucl. Phys. A*, 750:30–63, 2005.

[2] ATLAS collaboration. Generalized Numerical Inversion: A Neural Network Approach to Jet Calibration. Technical report, CERN, Geneva, Jul 2018.

[3] Rene Brun and Fons Rademakers. ROOT - An Object Oriented Data Analysis Framework: TMultiLayerPerceptron Class Reference. `https://root.cern.ch/doc/master/classTMultiLayerPerceptron.html`. 25.04.2021.

[4] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

[5] Claus Grupen and Boris Shwartz. *Particle Detectors*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2 edition, 2008.

[6] R. Bruce, N. Fuster-Martinez, A. Mereghetti, D. Mirarchi, and S. Redaelli. Review of LHC Run 2 Machine Configurations. In *9th LHC Operations Evian Workshop*, Geneva, Switzerland, 2019.

[7] Werner Herr and B Muratori. Concept of luminosity. 2006.

[8] LHC Machine. *JINST*, 3:S08001, 2008.

[9] Edmund J N Wilson. An introduction to particle accelerators. `https://cds.cern.ch/record/513326`, 2001.

[10] CERN. The large hadron collider. `https://home.cern/science/accelerators/large-hadron-collider`. 25.04.2021.

[11] Maurizio Vretenar, J Vollaire, R Scrivens, C Rossi, F Roncarolo, S Ramberger, U Raich, B Puccio, D Nisbet, R Mompo, S Mathot, C Martin, L A Lopez-Hernandez, A Lombardi, J Lettry, J B Lallement, I Kozsar, J Hansen, F Gerigk, A Funken, J F Fuchs, N Dos Santos, M Calviani, M Buzio, O Brunner, Y Body, P Baudrenghien, J Bauche, and T Zickler. *Linac4 design report*, volume 6 of *CERN Yellow Reports: Monographs*. CERN, Geneva, 2020.

[12] CERN. The Proton Synchrotron Booster. `https://home.cern/science/accelerators/proton-synchrotron-booster`. 25.04.2021.

[13] CERN Annual report 2016. Technical report, CERN, Geneva, 2017.

[14] CERN. The super proton synchrotron. `https://home.cern/science/accelerators/super-proton-synchrotron`. 25.04.2021.

[15] CERN. The higgs boson: A landmark discovery. `https://atlas.cern/higgs-boson-landmark-discovery`. 26.04.2021.

[16] CERN. Longer term lhc schedule. `https://lhc-commissioning.web.cern.ch/schedule/LHC-long-term.htm`. 26.04.2021.

[17] John Jowett. Colliding Heavy Ions in the LHC. page TUXGBD2. 6 p, 2018.

[18] Michael Benedikt, Paul Collier, V Mertens, John Poole, and Karlheinz Schindl. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

[19] CERN. Time for lead collisions in the lhc. `https://home.cern/news/news/accelerators/time-lead-collisions-lhc`. 26.04.2021.

[20] CERN. Facts and figures about the lhc. `https://home.cern/resources/faqs/facts-and-figures-about-lhc`. 25.04.2021.

[21] ATLAS collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[22] ATLAS Collarboration. Alignment of the ATLAS Inner Detector in Run-2. *Eur. Phys. J. C*, 80:1194. 41 p, Jul 2020.

[23] Christian Lippmann. Particle identification. *Nucl. Instrum. Meth. A*, 666:148–172, 2012.

[24] ATLAS Collaboration. Luminosity public results run2. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`. 23.04.2021.

[25] F. Halzen and Alan D. Martin. Quarks and leptons: An introductory course in modern particle physics, 1984.

[26] M. Anderson et al. Search for invisible modes of nucleon decay in water with the SNO+ detector. *Phys. Rev. D*, 99(3):032008, 2019.

[27] Kazuo Abe et al. Observation of a new narrow charmonium state in exclusive B+- —> K+- pi+ pi- J / psi decays. In *21st International Symposium on Lepton and Photon Interactions at High Energies (LP 03)*, 8 2003.

[28] Roel Aaij et al. Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \to J/\psi K^- p$ Decays. *Phys. Rev. Lett.*, 115:072001, 2015.

[29] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys. B*, 126:298–318, 1977.

[30] David J. Gross. Twenty five years of asymptotic freedom. *Nucl. Phys. B Proc. Suppl.*, 74:426–446, 1999.

[31] S. Acharya et al. Anisotropic flow in Xe-Xe collisions at $\sqrt{\mathbf{s}_{\mathrm{NN}}} = \mathbf{5.44}$ TeV. *Phys. Lett. B*, 784:82–95, 2018.

[32] Oliver DeWolfe, Steven S. Gubser, Christopher Rosen, and Derek Teaney. Heavy ions and string theory. *Prog. Part. Nucl. Phys.*, 75:86–132, 2014.

[33] Herbert B Callen. *Thermodynamics and an introduction to thermostatistics; 2nd ed.* Wiley, New York, NY, 1985.

[34] Terry Awes. Status of the quark gluon plasma search. *Pramana*, 67:915–925, 11 2006.

[35] Betty Abelev et al. Centrality determination of Pb-Pb collisions at $\sqrt{s_{NN}} =$ 2.76 TeV with ALICE. *Phys. Rev. C*, 88(4):044909, 2013.

[36] Michael L. Miller, Klaus Reygers, Stephen J. Sanders, and Peter Steinberg. Glauber modeling in high energy nuclear collisions. *Ann. Rev. Nucl. Part. Sci.*, 57:205–243, 2007.

[37] Raimond Snellings. Elliptic Flow: A Brief Review. *New J. Phys.*, 13:055008, 2011.

[38] Brett Joseph Teeple. *Deconfinement and Duality of (super) Yang-Mills on Toroidially-compactified Spacetimes for all Gauge Groups*. PhD dissertation, Toronto U., 2015.

[39] Rybář Martin. *Study of jets in p+p and Pb+Pb collisions at LHC*. PhD dissertation, Univerzita Karlova, Matematicko-fyzikální fakulta, Ústav částicové a jaderné fyziky, 2015.

[40] J. D. Bjorken. *Energy Loss of Energetic Partons in Quark - Gluon Plasma: Possible Extinction of High p(t) Jets in Hadron - Hadron Collisions*. 8 1982.

[41] Georges Aad et al. Observation of a Centrality-Dependent Dijet Asymmetry in Lead-Lead Collisions at $\sqrt{s_{NN}} = 2.77$ TeV with the ATLAS Detector at the LHC. *Phys. Rev. Lett.*, 105:252303, 2010.

[42] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[43] ATLAS Collaboration. Numerical inversion for calibration of Heavy-ion jets. 2014. ATL-COM-PHYS-2014-795.

[44] ATLAS Collaboration. ATLAS Pythia 8 tunes to 7 TeV data. *ATL-PHYS-PUB-2014-021*, 2014.

[45] Morad Aaboud et al. Measurement of the nuclear modification factor for inclusive jets in Pb+Pb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV with the ATLAS detector. *Phys. Lett. B*, 790:108–128, 2019.

[46] David J. C. MacKay. *Information Theory, Inference & Learning Algorithms*. Cambridge University Press, USA, 2002.

[47] Petr Chaloupka, Pavel Jakl, Jan Kapitan, Michal Zerola, and Jerome Lauret. Setting up a STAR Tier 2 Site at Golias/Prague Farm. *J. Phys. Conf. Ser.*, 219:072031, 2010.

[48] Quasi-newton methods. `https://doi.org/10.1007/978-0-387-40065-5_6`, 2006.

[49] Sagar Sharma. Epoch vs batch size vs iterations, towards data science. `https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9`. 20.05.2021.

[50] Rene Brun and Fons Rademakers. ROOT - An Object Oriented Data Analysis Framework: TMLPAnalyzer Class Reference. https://root.cern.ch/doc/master/classTMLPAnalyzer.html. 26.05.2021.

List of Figures

List of Tables

A. Attachments

A.1 Additional graphs - Jet energy scale, jet energy resolution

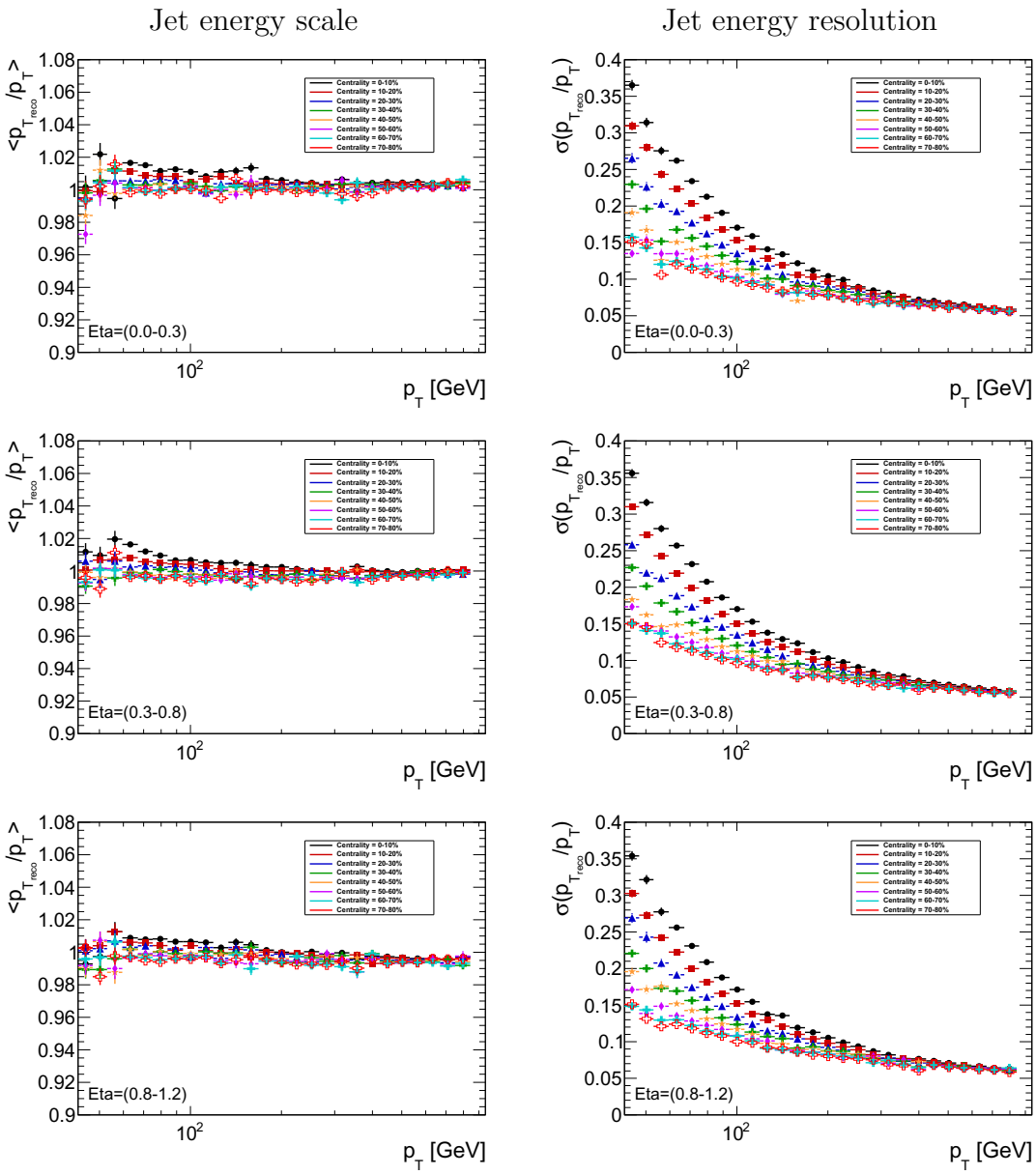

Figure A.1: Jet energy scale and jet energy resolution in Pb+Pb collisions for pseudorapidity $\eta \in (0.0, 1.2)$

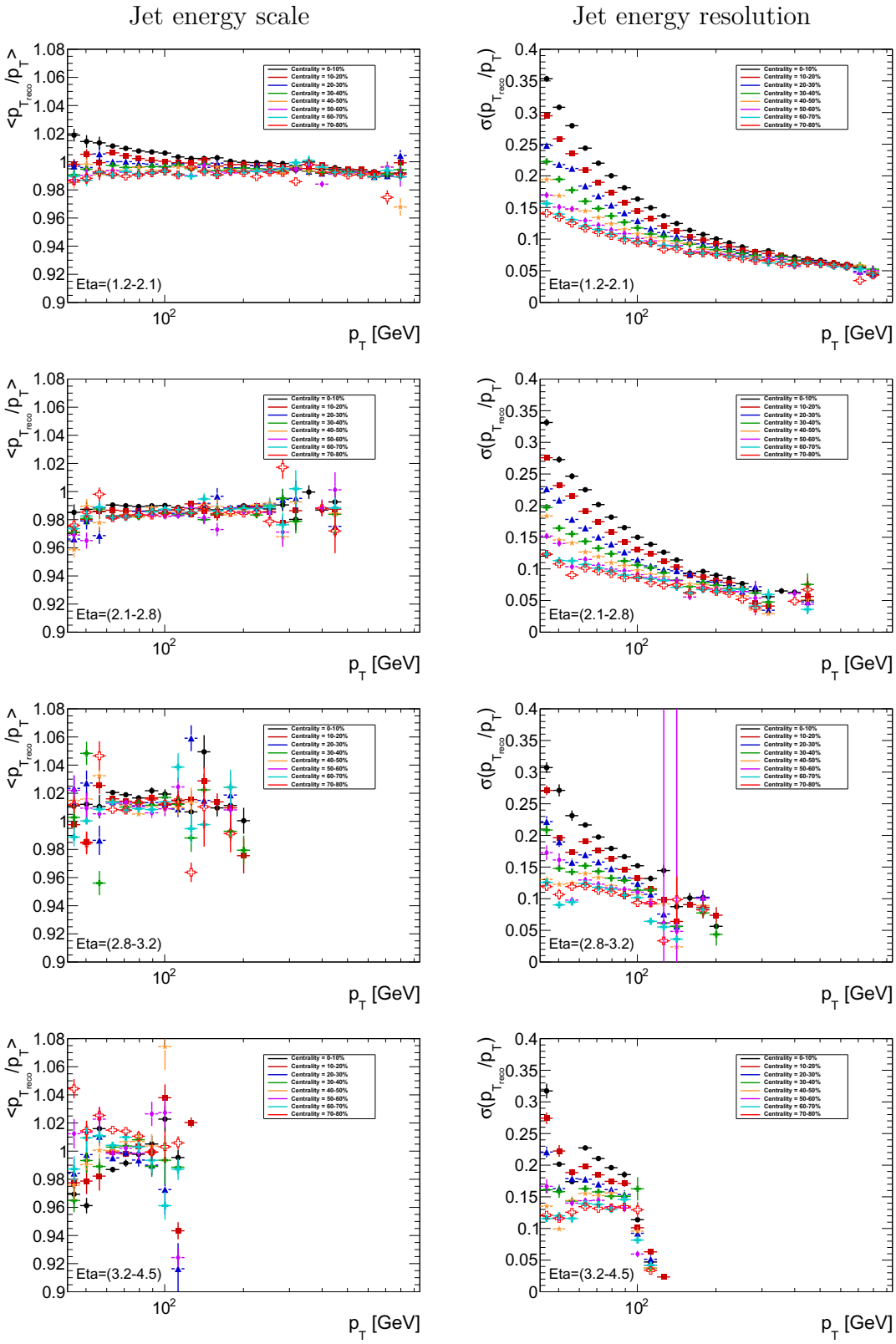

Figure A.2: Jet energy scale and jet energy resolution in Pb+Pb collisions for pseudorapidity $\eta \in (1.2, 4.5)$

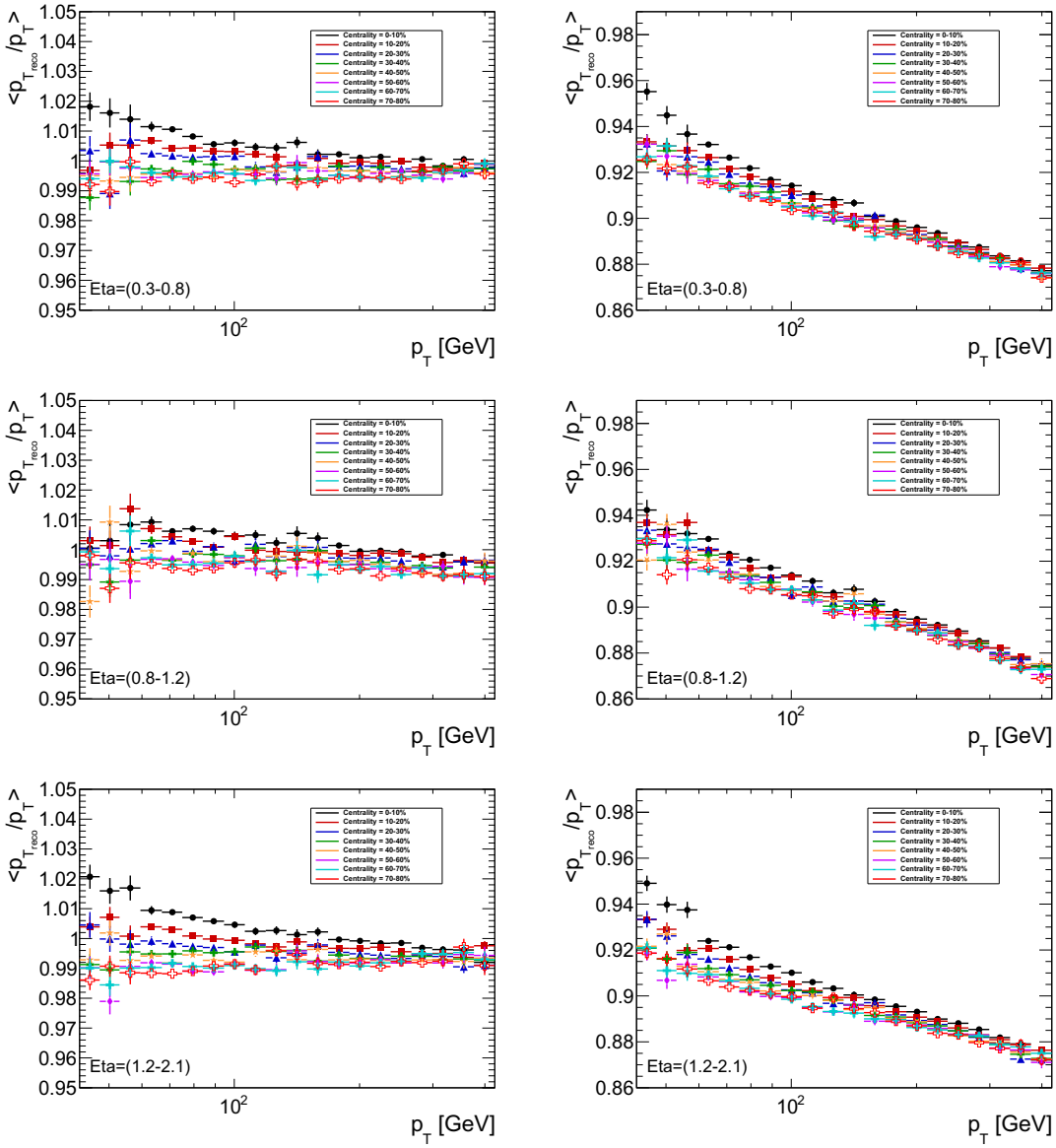

Figure A.3: The comparison of JES for uncorrected jets (left) and jets corrected (right) by neural network, which depends on r_{trk} and $\mathcal{W}$. The jets are restricted to $0.3 < |\eta| < 2.1$.

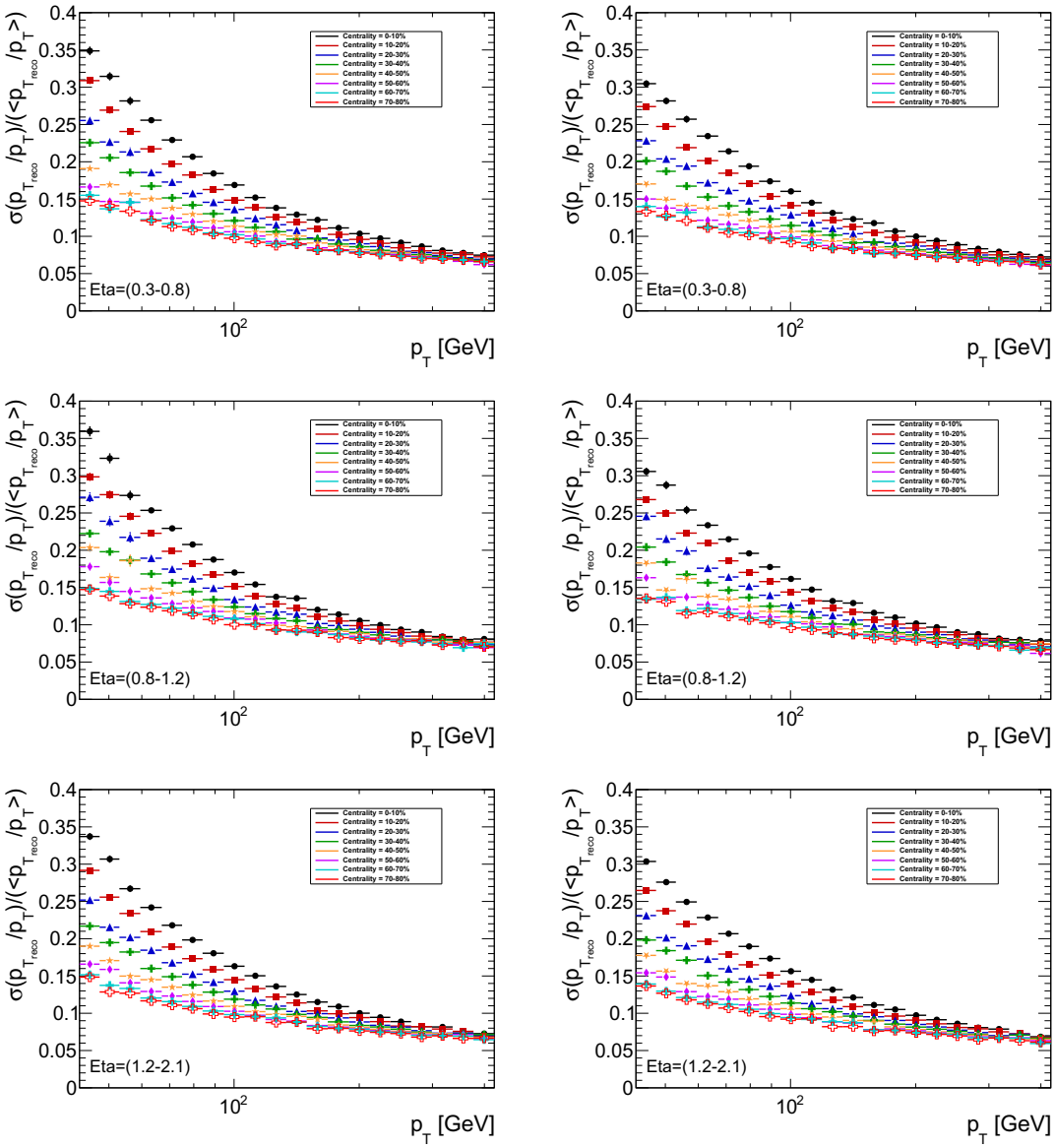

Figure A.4: The comparison of JER scaled by JES for uncorrected jets (left) and jets corrected (right) by neural network, which depends on r_{trk} and $\mathcal{W}$. The jets are restricted to $0.3 < |\eta| < 2.1$.

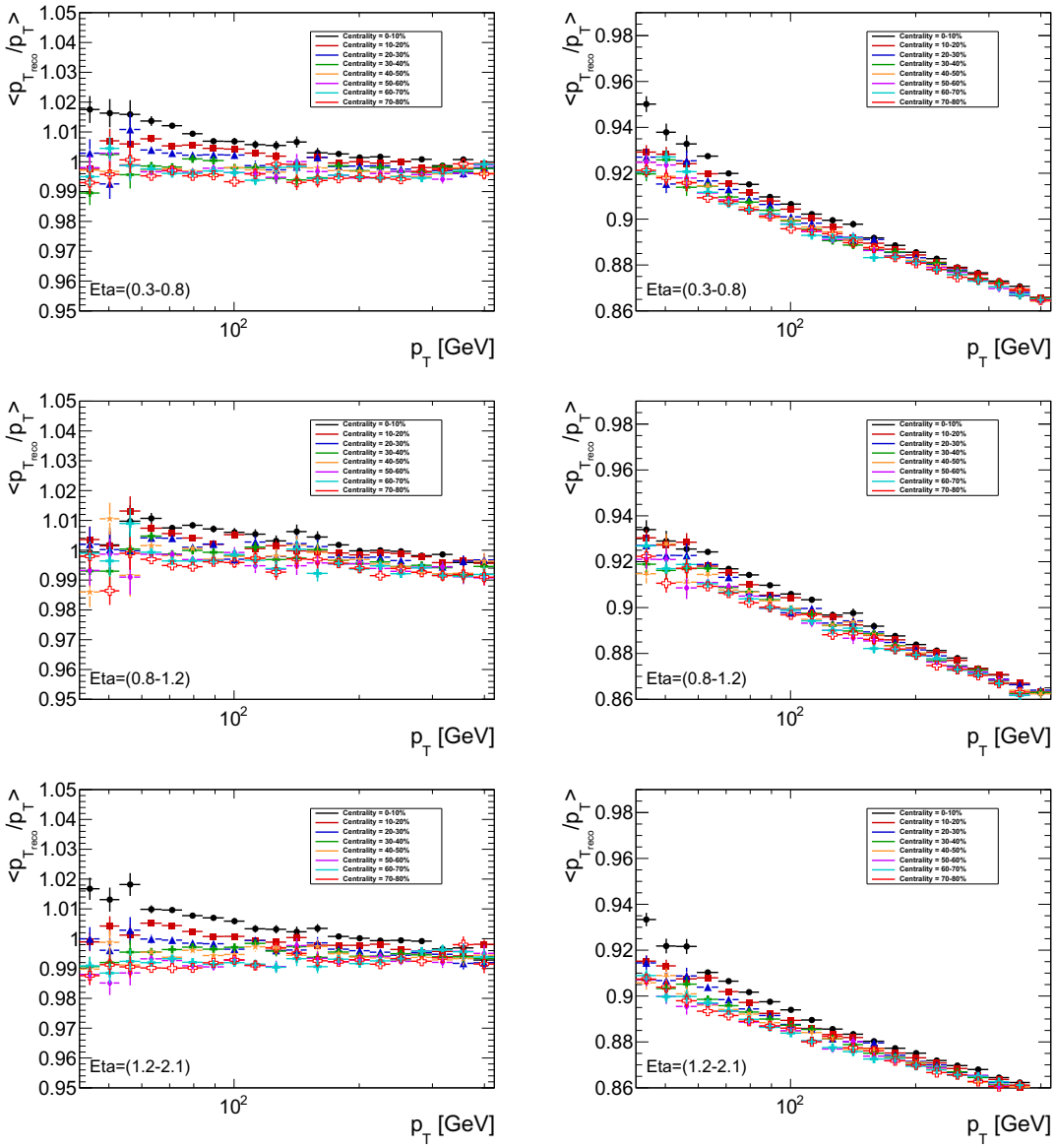

Figure A.5: The comparison of JES for uncorrected jets (left) and jets corrected (right) by neural network, which depends on r_{trk} and n_{trk}. The jets are restricted to $0.3 < |\eta| < 2.1$.

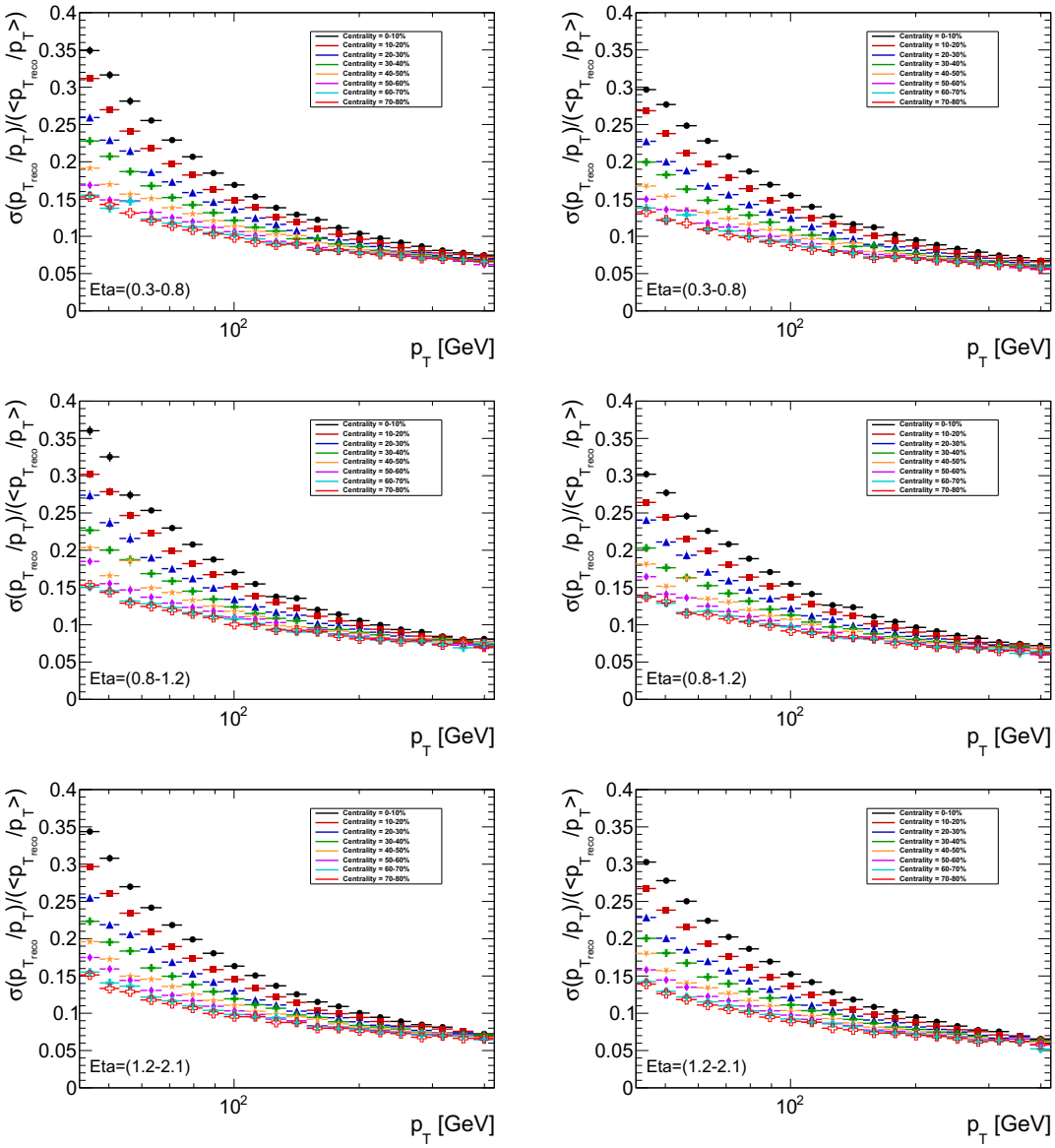

Figure A.6: The comparison of JER scaled by JES for uncorrected jets (left) and jets corrected (right) by neural network, which depends on r_{trk} and n_{trk}. The jets are restricted to $0.3 < |\eta| < 2.1$.