TECHNISCHE UNIVERSITÄT DRESDEN

Optimization of the Reconstruction of the Missing Transversal Momentum in the new Software of the ATLAS Experiment

Bachelor-Arbeit
zur Erlangung des Hochschulgrades
Bachelor of Science
im Bachelor-Studiengang Physik

vorgelegt von

Johannes Heinz Wünsche
geboren am 30.03.1998 in Ebersbach/Sa.

Institut für Kern- und Teilchenphysik
Fakultät Physik
Bereich Mathematik und Naturwissenschaften
Technische Universität Dresden
2019

Eingereicht am 17. Juni 2019

1. Gutachter: Prof. Dr. Michael Kobel
2. Gutachter: Dr. Frank Siegert

Abstract

Many of the processes described and hence predicted by the Standard Model of particle physics were already observed and verify that theory in this way. Some vector boson scattering processes belong to the predictions which have not been measured precisely so far. It is the aim of current efforts to use the data from the ATLAS experiment at the LHC (CERN) to succed with the verification.
Commonly appearing decay products of $W^{\pm}$ bosons involved in such scattering processes are neutrinos. These particles interact only very weakly and are therefore practically undetectable at the ATLAS experiment. To gain information about them anyhow, it is necessary to use an indirect method. The consideration of the missing transversal momentum (MET) seems to be constructive because it can be directly associated with the transversal energy of the neutrino. The MET reconstruction based on the given detector signals is performed with the ATLAS software Athena which was recently updated to version 21. That new version adds several variations of the reconstruction algorithm to the existing ones, for example by defining and calibrating new working points for the jet selection. This bachelor thesis is targeted on the investigation of the available variations in order to find the most appropriate ones for the application in the research on vector boson scattering. Therefore, the algorithm is applied to data samples from Monte Carlo simulations of the process $W^{\pm}Z \rightarrow W^{\pm}Z$. By comparing the obtained results with the truth values, the MET reconstruction performance is evaluated and conclusions are made about the differences between the variations.

Kurzzusammenfassung

Zahlreiche der vom Standard-Modell beschriebenen und damit auch vorhergesagten Prozesse sind bereits häufig beobachtet worden und bestätigen diese Theorie somit. Einige Vektorbosonen-Streuprozesse gehören jedoch zu jenen Vorraussagen, die noch nicht mit hoher Signifikanz vermessen werden konnten. Es ist das Ziel aktueller Bemühungen, sie mit Hilfe der durch das ATLAS-Experiment am LHC (CERN) gesammelten Daten zu verifizieren.
Häufig auftretende Zerfallsprodukte von $W^{\pm}$-Bosonen, die in solchen Prozessen involviert sein können, sind Neutrinos. Diese Elementarteilchen treten nun sehr schwach mit anderer Materie in Wechselwirkung und sind daher durch das ATLAS-Experiment praktisch nicht detektierbar. Um dennoch Informationen über sie zu gewinnen, sind indirekte Methoden nötig. Die betrachtung des fehlenden Transversalimpulses (MET) erscheint sehr sinnvoll, da er direkt in Verbindung mit dem Neutrinoimpuls gebracht werden kann. Die MET-Rekonstruktion erfolgt auf der Grundlage von gegebenen Detektorsignalen mit der ATLAS-Software Athena, die vor kurzem zur 21. Version aktualisiert wurde. Diese neue Version fügt mehrere Variationen

des Rekonstruktionsalgorithmus hinzu, beispielsweise durch die Definition neuer Arbeitspunkte für die Jet-Auswahl. Diese Bachelor-Arbeit zielt darauf ab, die nun verfügbaren Variationen zu untersuchen, um die geeignetste unter ihnen für die Anwendung bei der Erforschung der Vektorboson-Streuung zu finden. Dazu wird der Algorithmus auf Datensätze von Monte-Carlo-Simulationen zum Prozess $W^{\pm}Z \rightarrow W^{\pm}Z$ angewandt. Durch den Vergleich der erhaltenen Werte mit den wahren Werten wird die Funktionsweise der MET-Rekonstruktion eingeschätzt. Daraus werden schließlich Schlussfolgerungen über die Unterschiede zwischen den Variationen gezogen.

Contents

1 Introduction

At the beginning of the 20th century, the β-decay of radioactive nuclei was studied intensely for the first time. Each nucleus was supposed to emit only one electron. Conducted experiments led to the conclusion that these emitted electrons do not have a discrete energy value, as it was stated by many phisicists before. In contrast, their energy distribution formed a continuous spectrum. If, despite of the electron, no other particle was produced, the energy conservation law would not be universally valid. Because of these thoughts, Wolfgang Pauli postulated the existence of an unknown elementary particle which appears in the β-decay as an additional decay product to maintain the validity of the energy conservation. According to the observations, this particle was supposed to be elecrically neutral and almost undetectable. Finally, it was called *neutrino* following a proposal by Enrico Fermi. Its first experimental confirmation suceeded in 1956 by an experiment conducted by the research group of Clyde Cowan and Frederick Reines.

An current object of research in particle physics is the confirmation of vector boson scattering processes predicted by the Standard Model. The ATLAS experiment at the LHC is one of the most important experiments used for this purpose. In these scattering processes, $W^{\pm}$ bosons occur as products very often. If they decay leptonically, there are always neutrinos produced. Although they are not detectable, they have a certain energy and momentum and thus they carry a certain amount of information about interesting unstable states which could have existed shortly after the associated proton-proton collision. Therefore, efforts should be made to gain knowledge about them.

This can happen by considering the momentum conservation. In the transversal direction in respect to the LHC beam line, the total momentum should theoretically be zero. Practically, adding up all transversal momentum components of particles reconstructed on the basis of detector signals, this result will be achieved only in rare cases. On the one hand, there are effects of mis-calibration, mis-measurement and generally the imperfectness of the detector leading to deviations. On the other hand, there are disregarded particles, namely the neutrinos, which do not contribute in this calculation because they can not be reconstructed due to missing detector signals. But conversely, the transversal momentum of neutrinos can be estimated by regarding the amount of that *missing transversal momentum* (MET). Accordingly, it seems to be helpful to achieve an optimization of the applied reconstruction algorithm for the MET to attain more reliable conclusions about the neutrinos and thereby about the investitaged vector boson scattering processes. This bachelor thesis is focussed on this aim.

2 Theoretical Background

2.1 The Standard Model of Particle Physics

The Standard Model of particle physics is the most comprehensive theory of contemporary physics. It contains a description of the three most relevant interactions between the known elementary particles. It is based on a construction consisting of symmetries and charges explaining the characteristic phenomena of these three interactions.
Concretely, they are considered to be a consequence of the existence of so-called local gauge symmetries, i.e. the fundamental Lagrange density describing all processes in particle physics is supposed to remain invariant under certain local gauge transformations. These symmetries imply the occurrence of gauge fields which can be identified as mediating particles. These mediating particles are spin-1-particles and are therefore called gauge bosons. Furthermore, according to Noether's theorem, the symmetries lead to the existence of associated charges as conserved quantities. Every particle is characterized by the charges it carries deciding if it is sensitive to a certain interaction or not. The three concerned interactions are the electromagnetic, the weak and the strong interaction. The electromagnetic interaction is the implication of the $U(1)$ gauge symmetry. It is connected to the electromagnetic charge Z and its mediating particle is the massless photon γ. The weak interaction is a consequence of the SU(2) gauge invariance and is hence more multifarious, it has several gauge bosons called W^+, W^- and Z bosons and its associated charge is the weak Isospin $\vec{I_W}$ (for the description of weakly charged particles it suffices to regard the third component of their weak isospin: I_W^3). The strong interaction is implied by the SU(3) gauge symmetry. It is mediated by eight different gauge bosons namely gluons. The affiliated strong charge $\vec{C}$ is also called colour charge. This knowledge about these three interactions is summarized in Table 2.1. The fourth known interaction, the gravitation, is not included in the theory.

The known matter particles (which are all spin-1/2-particles and thus fermions) whose behaviour is described by the Standard Model are divided into quarks and leptons. Quarks are the only elementary particles which undergo strong interaction. Due to the fact that gluons are self-interacting mediating particles, quarks form bound states (called hadrons) which have no net colour charge. Because they are also weakly and electrically charged they are influenced by the three interactions included in the model. Considering the electric charge, quarks can be

Interaction	Gauge group	Associated charge	Gauge boson
Electromagnetism	$U(1)_Z$	Electrical charge Z	Photon γ
Weak interaction	$SU(2)_L$	Weak isospin $\vec{I_W}$	$W^{\pm}$, Z
Strong interaction	$SU(3)_C$	Colour charge $\vec{C}$	Gluon g

Table 2.1: Overview about the three interactions described by the Standard Model with their gauge bosons

Generation	Particle	Electric charge Z	I_W^3	Mass m in MeV
1	up quark u	$+2/3$	$+1/2$	≈ 2
	down quark d	$-1/3$	$-1/2$	≈ 5
	electron e	-1	$-1/2$	0.51
	electron-neutrino ν_e	0	$+1/2$	≈ 0
2	charm quark c	$+2/3$	$+1/2$	$\approx 1\,300$
	strange quark s	$-1/3$	$-1/2$	≈ 96
	muon μ	-1	$-1/2$	105.66
	muon-neutrino ν_μ	0	$+1/2$	≈ 0
3	top quark t	$+2/3$	$+1/2$	$\approx 173\,000$
	bottom quark b	$-1/3$	$-1/2$	$\approx 4\,200$
	tauon τ	-1	$-1/2$	1776.99
	tau-neutrino ν_τ	0	$+1/2$	≈ 0

Table 2.2: Overview about the known matter particles (fermions)

divided into two groups, the three up-like ones with $Z = +\frac{2}{3}$ (up, charm, top) and the three down-like ones with $Z = -\frac{1}{3}$ (down, strange, bottom). Leptons can also be divided into two groups: the three electrically charged ones (electron, muon, tauon) and the three electrically neutral ones. The latter are called neutrinos (electron neutrino, muon neutrino, tau neutrino). Another criterion to distinguish between these twelve matter particles is the mass. Considering this property, three so-called generations can be formed containing the particles from each of the just mentioned sub-groups with the lowest mass, the medium mass and the highest mass, respectively. The matter particles are summarized in Table 2.2. For each matter particle, there is also an antimatter particle with the same properties despite of opposite charges.

The last, not yet mentioned elementary particle which is known nowadays is the Higgs boson. It is a consequence of the Brout-Englert-Higgs mechanism which is included in the Standard Model to explain some properties of the weak interaction. Due to an additional field, the Higgs field, an original $SU(2)_L \times U(1)_Y$ symmetry was broken. The outcome of this symmetry breaking is a mixing of the associated gauge fields to the observable ones, belonging to the electromagnetic and the weak interaction. The photon γ and the Z boson consist partially of the B boson (the gauge boson of the more fundamental $U(1)_Y$ gauge field connected to the weak hypercharge Y_W as charge) and the W^3 boson (the third of the three gauge fields appearing as a consequence of the more fundamental $SU(2)_L$ gauge field connected to the weak isospin $\vec{I_W}$ as charge). The W^+ and W^- bosons are a superposition of the W^1 and the W^2 bosons. In this way, the masses of the weak gauge bosons and all other massive particles can be explained.

2.2 Vector Boson Scattering

As mentioned in Section 2.1, the electromagnetism is based on the U(1) gauge symmetry. Since U(1) is an abelian group, there are no terms in the Lagrange density implying multiple-photon vertices. Therefore, photons are not able to interact with each other. By contrast, U(2) is a non-abelian group causing terms in the Lagrange density which imply multiple-boson vertices. Among these scattering processes predicted by the Standard Model, there is the process $W^{\pm}Z \to W^{\pm}Z$. Because of its very small cross-section, this process has not been confirmed so far. With the help of the LHC, the confirmation could take place in the foreseeable future.
Using this collider, vector boson scattering processes can occur in a way which is shown in Figure 2.1. $W^{\pm}$ and Z bosons can be emitted by two quarks. These quarks are hence deflected theirselves causing one jet each. After that, the vector bosons interact via a WZ-vertex. After this scattering interaction, each of the bosons decays in two fermions. The two possible options are the hadronical decay, into two quarks, and the leptonical decay, into leptons. This bachelor thesis focusses on the already mentioned $W^{\pm}Z \to W^{\pm}Z$ process, limited on fully leptonic decays of the two bosons. Only cases are considered where the Z boson decays into a charged lepton and its anti-particle and the $W^{\pm}$ decays into a charged lepton and a neutrino. Cases, where tau leptons appear decay products are also excluded. Accordingly, the process should be recognizable in the detector signal data by two jets and three leptons (muons or electrons) leaving the interaction point.

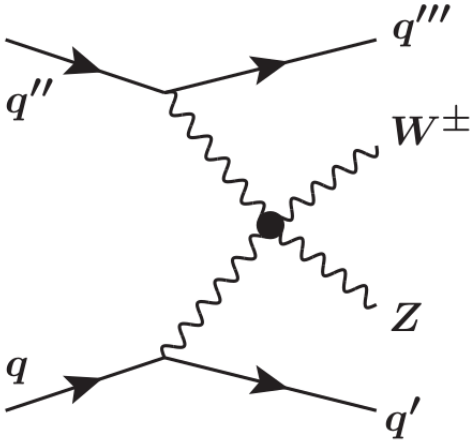

Figure 2.1: Feynman diagram of a WZ scattering process as it could occur at the LHC: one initial quark emits a W boson and another one emits a Z boson, both interact via a WZ-vertex; the outgoing bosons can decay hadronically or leptonically (figure taken from reference [1])

3 ATLAS Detector and Definition of MET

3.1 The ATLAS Detector

Currently, the most powerful particle accelerator is the LHC (Large Hadron Collider) generating pp-collisions with center-of-mass energies up to $\sqrt{s} = 13$ TeV. It is connected to several highly developed detectors which are, among others, used to test predictions of the Standard Model. One of these experiments is the ATLAS detector on which this bachelor thesis is focussed on. Its inner structure is visualized in Figure 3.1. It is useful to introduce a coordinate system and some specific angles to describe this complex construction as precise as possible. Fig. 3.2 shows how the coordinate system is usually defined. It is a right-handed system, its origin is located in the interaction point within the detector. The z-axis is aligned with the beam direction, i.e. the velocity vectors of the incoming beam protons from both sides are (anti-) parallel to the z-axis. The x-axis is directed to the center of the LHC ring. The y-axis is consequently perpendicular to the xz-plane, it is directed upwards. Commonly used angles are the azimuthal angle ϕ, which expresses how the projection of a considered vector on the xy-plane is directed, and the polar angle θ, which equals the angle enclosed by the z-axis and that vector. In the most cases, the so-called pseudorapidity $\eta = -\ln\tan(\frac{\theta}{2})$ is used instead of θ.

The structure of the ATLAS detector can be divided into four parts: the inner detector, the electromagnetic calorimeter, the hadronic calorimeter and the muon chambers. The inner detector contains a tracking system which consists mainly of silicon pixel detectors and microstrip detectors. These trackers enable a reconstruction of trajectories with a very high precision. This inner detector part is included in a solenoid magnet leading to a deflection of electrically charged particles which allows a determination of their specific momenta. The tracking system covers only the region with $|\eta| < 2.5$. Closer to the beam pipe, the needed space is already filled with other detector parts and the beam pipe itself. The neighbouring electromagnetic calorimeter consists of liquid argon sampling calorimeters with a high granularity. It is extended up to $|\eta| = 3.2$. The hadronic calorimeter is based on scintillator tile calorimeters. It covers a much shorter range in comparison to the electromagnetic calorimeter ($|\eta| < 1.7$). The so-called end-caps contain further liquid-argon calorimeters with a coverage up to $|\eta| = 4.9$. The muon chambers are located around the calorimeters. Using different tracking systems with a coverage up to $|\eta| = 2.7$, it is able to measure the trajectories of muons which commonly pass the calorimeter layers of the detector.

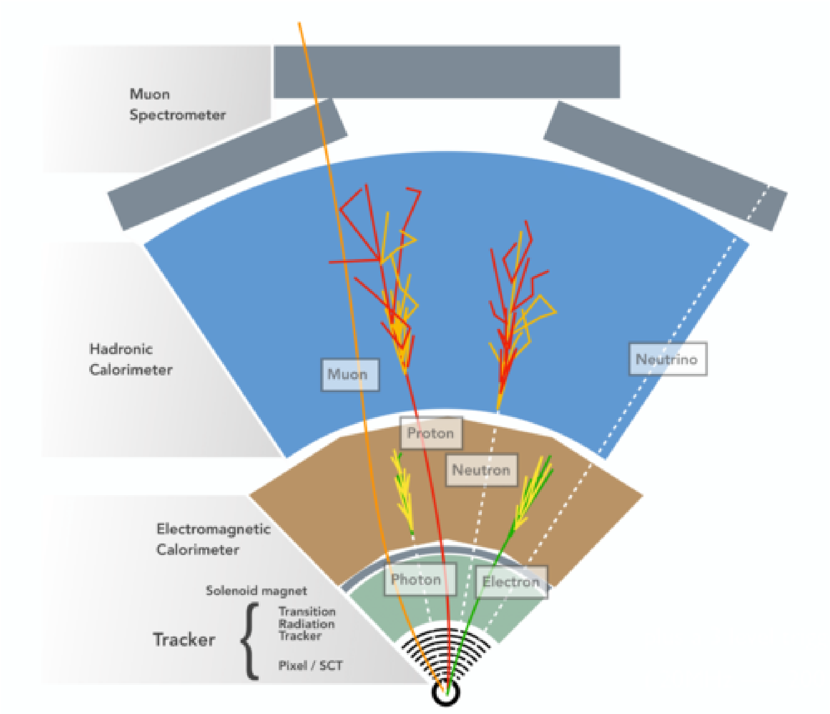

Figure 3.1: Schematic image showing the inner structure of the ATLAS detector (in xy-plane); charged particles are detectable in the inner detector where their tracks are measured and in the calorimeters where they are decelerated and deposit their energy due to interactions with detector elements, muons usually overcome the calorimeters and reach the muon chamber where their tracks can be measured again; photons do not cause signals in the inner detector due to their missing electrical charge, but they are detectable in the calorimeter where they also deposit all their energy; neutrinos are the only undetectable (known) particles (figure taken from https://cds.cern.ch/record/1096081)

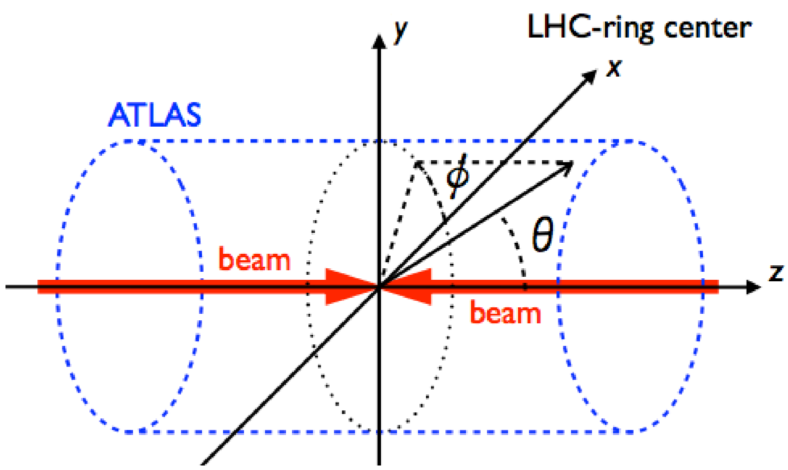

Figure 3.2: Visualization of the coordinate system (figure taken from Reference [2])

3.2 MET Definition

The ATLAS detector is able to detect a big part of the stable and long-living product particles which are produced near the interaction point during bunch crossings. Among the known elementary particles, only neutrinos are undetectable. They do not carry an electromagnetic or colour charge, thus they are not able to generate measurable signals in any part of the detector mediated by the electromagnetic or the strong interaction. Merely effects caused by the weak interaction may occur, but due to the extremely low cross sections of the corresponding processes they can be neglected practically. Although direct evidence of their appearance is excluded, conclusions about their existence and their kinematic properties can be made indirectly.

The key to these indirect conclusions is the momentum conservation. As we know, the velocity vectors of the incoming protons as collision partners are aligned parallelly to the z-axis, resp. antiparallelly (velocity components in the directions perpendicular to the z-axis are negligible because of the high-relativistic movement of the bunches). Momentum conservation implies that the sum of the momenta of the outgoing particles has to be equal to the sum of the momenta of the two incoming protons. While the sum of the momenta in z-direction is not exactly known due to differing momentum fractions of the interacting partons in the colliding protons, the sum of the transversal momenta has to be zero. Practically, adding up all reconstructed momenta of the outgoing particles will lead only in rare cases to this result. The main reason is neutrinos whose momenta are not respected because they can not be perceived by the detector. Other reasons are fake detection caused by defects of detector parts, a miscalibration or wrong reconstruction of collision products as well as potentially unknown elementary particles. The difference between zero and the reconstructed transversal momentum is called *missing transversal momentum* $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$, or also *missing transversal energy* (MET). In other words, the missing transversal momentum is defined as the negative sum of the transversal momentum vectors of all objects which are regarded as products of the collision event.

The missing transversal momentum can be divided into several terms containing different contributions:

$$\vec{p}_{\mathrm{T}}^{\mathrm{miss}} = \vec{p}_{\mathrm{T,\,hard}}^{\mathrm{miss}} + \vec{p}_{\mathrm{T,\,soft}}^{\mathrm{miss}} = \vec{p}_{\mathrm{T},\,e}^{\mathrm{miss}} + \vec{p}_{\mathrm{T},\,\mu}^{\mathrm{miss}} + \vec{p}_{\mathrm{T},\,\tau}^{\mathrm{miss}} + \vec{p}_{\mathrm{T},\,\gamma}^{\mathrm{miss}} + \vec{p}_{\mathrm{T,\,jet}}^{\mathrm{miss}} + \vec{p}_{\mathrm{T,\,soft}}^{\mathrm{miss}} \quad (1)$$

The first division leads to two terms which are refered to as the *hard term* $\vec{p}_{\mathrm{T,\,hard}}^{\mathrm{miss}}$ and the *soft term* $\vec{p}_{\mathrm{T,\,soft}}^{\mathrm{miss}}$. The hard term includes all contributions to the MET which are associated with reconstructed objects ("hard" objects) and it can be divided into several terms itself, each for a specific type of particles: electrons, muons, hadronically decaying tau-leptons, photons and jets. The soft term contains all other contributions from detector signals not associated with hard objects. The algorithm used for this bachelor thesis modifies this usual division by including the contributions of tau leptons and photons in the soft term. They are not reconstructed as hard objects.

4 MET Reconstruction Algorithm

4.1 Challenges of the MET Reconstruction

Practically, the MET reconstruction is done by a certain algorithm which should include all the signal data from the detector including particle tracks and energy deposits. It is complicated by some disturbing effects whose influence should be logically reduced.

One of these effects are pileup interactions. All hard-scatter vertices which appear additionally to considered one are called pile-up vertices. Due to the large number of protons per bunch in the accelerator, the number of *pp*-collision events per bunch crossing is accordingly high (in the two-digit range). The much more numerous outgoing collision products superpose hence strongly. Since it is the aim to determine the MET for a single hard-scatter process to gain knowledge about its undetectable products, there is the difficulty that relevant detector signals have to be separated from irrelevant ones. Consequently, the reconstruction algorithm has to be able to reduce this pile-up influence by the help of appropriate methods.

Another mentionable problem is the imperfectness of the detector. As described in Chapter 3.1, all detector parts have a specific coverage up to a certain value of $|\eta|$. Particularly the inner detector with its tracking system has only a coverage up to $|\eta| = 2.5$. In contrast, the calorimeters are able to measure signals from particles in the range $|\eta| < 4.9$. Hence, for particles with $2.5 < |\eta| < 4.9$, no track information can be acquired, only the data for the energy deposits in the calorimeters are usable. The MET reconstruction algorithm should be able to handle this difficulty, not neglecting potentially important contributions from that detector region. In general, it is an open question if the consideration of tracking information improves the result of the reconstruction in a significant way.

4.2 MET Reconstruction in Release 21.2

The used MET reconstruction algorithm is part of Athena, the analysis framework used for the ATLAS experiment. On the basis of the detector data about signals which were caused by particles, it reconstructs the properties of these physical objects. The software is continuously improved in order to adapt the algorithms to current detector conditions and collision data. This work leads to frequent actualizations. Since recently, release 21 of Athena is available. It offers several new variations of the MET reconstruction algorithm whose performance shall be investigated here.

These variations concern the way in which the reconstruction of the terms is done and also quality and isolation criteria. This subsection gives an overview of their characteristics. Jets, electrons and muons are reconstructed as hard objects, all other included signals including tau leptons and photons are part of the soft term.

4.2.1 Reconstruction of Jet Contributions

In Release 21, there are two different types of jet reconstruction algorithms usable: EMTopo and EMPFlow. The EMTopo jet reconstruction disregards track information from the inner detector. It uses only data about the topology of energy deposit clusters from the calorimeters. On these data, the anti-k_t algorithm (with $R = 0.4$) is applied. The anti-k_t reconstruction algorithm is described in detail in Reference [3]. The jets are calibrated following the EM+JES Scheme (see References [4] and [5]).
After the reconstruction and calibration, the so-called jet-vertex-tagging algorithm is applied. It calculates a discriminant value for each jet which allows an appreciation of the probability for the jet to come from a certain hard-scatter vertex. By comparing the values for different vertices, the jet can be matched to the one which is most likely its origin. Higher values indicate a higher probability. Conversely, if the focus already lies on a specific hard-scatter vertex, this discriminant can be used to appreciate if the considered jet is originated there. Therefore, it is utilized to filter pile-up jets. For the performance of the algorithm, track data from the inner detector are also included. Because of this fact, it can only be applied to non-forward jets (with $|\eta| < 2.4$).
For former releases, the discriminant was defined for a jet i and a regarded primary vertex j in the following way:

$$\mathrm{JVT}_i(\mathrm{jet}_i, \mathrm{PV}_j) = \frac{\sum_m p_T(\mathrm{track}_m^{\mathrm{jet}_i}, \mathrm{PV}_j)}{\sum_n \sum_l p_T(\mathrm{track}_l^{\mathrm{jet}_i}, \mathrm{PV}_n)} \tag{2}$$

where the index m runs over all tracks associated with the jet i matched to vertex j, n over all primary vertices in the collision event and l over all tracks associated with jet i matched to the primary vertex n. More easily expressed, it is the ratio of the sum of the p_T of the tracks matched to the regarded vertex j and the sum of the p_T of the tracks matched to other vertices for jet i. Accordingly, it is also called *jet vertex fraction*. Details can be found in [5]. Currently, the JVT discriminant calculation was changed, now it includes multivariate methods (see [6]). But it is still similar and based on the same idea. Using different JVT values as thresholds for the filtering, four jet selection working points were defined: Loose, Tight, Tighter and Tenacious. Their concrete definitions are given in table 4.1. Next to different JVT thresholds, they distinguish in respect to the p_T-cuts.
Additionally, there is the possibility of using a so-called forward-jet-vertex tagger (fJVT). This algorithm was created as an attempt to reduce disturbing influences of jets in the forward region (with $|\eta| > 2.5$) with transversal momenta 20 GeV $< p_T <$ 50 GeV which have their origin in pile-up vertices. At first, for each pile-up vertex i, the MET $\vec{p}_{\mathrm{T,i}}^{\mathrm{miss}}$ is reconstructed on the basis of detector signals from the central region. There should only be one central jet matched to that vertex and this one should definitely not be originated in the interesting hard-scatter vertex. Therefore, it has to satisfy a condition concerning the JVT discriminant: JVT < 0.14. Next,

Working point	p_T (all jets)	p_T (forward jets)	JVT (for jets with $\|\eta\| < 2.4$)
Loose	> 20 GeV	> 20GeV	> 0.59 (for $p_T < 60$GeV)
Tight	> 20GeV	> 30GeV	> 0.59 (for $p_T < 60$GeV)
Tighter	> 20GeV	> 35GeV	> 0.59 (for $p_T < 60$GeV)
Tenacious	> 20GeV	> 35GeV	> 0.91 (for $p_T < 40$GeV), > 0.59 (for $p_T < 60$GeV), > 0.11 (for $p_T < 120$GeV)

Table 4.1: Overwiew about the thresholds for the jet selection working points for the EMTopo jet reconstruction

the criterion $\Delta R_{\mathrm{pT}} > 0.2$ is applied to the considered forward jet. It has the task to ensure that the forward het belongs to the class of QCD pile-up jets (jets coming almost exclusively from a single QCD process in a single pile-up vertex) having generally a higher ΔR_{pT} than stochastic pile-up jets (combining particles from different hard-scatter events) which should be removed. The exact definition of ΔR_{pT} and other details can be read in Reference [6]. Finally, a fJVT discriminant value for every pile-up vertex i can be defined in the following way:

$$\mathrm{fJVT}_i = \frac{\vec{p}_{\mathrm{T,i}}^{\,\mathrm{miss}} \cdot \vec{p}_{\mathrm{T}}^{\,\mathrm{fj}}}{|\vec{p}_{\mathrm{T}}^{\,\mathrm{fj}}|^2} \tag{3}$$

It is the normalized projection of the missing transversal momentum $\vec{p}_{\mathrm{T,i}}^{\,\mathrm{miss}}$ on the transversal momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{fj}}$ of the forward jet. The algorithm follows the idea that a forward jet is very likely to be originated in a pile-up vertex if one of them can be identified whose transversal momentum constructed by the usage of detector data from the central region is directed back-to-back in the azimuthal angle ϕ with respect to the transversal momentum of this forward jet. In the case of a forward pile-up jet from vertex i, $\vec{p}_{\mathrm{T,i}}$ in the jet direction is therefore supposed to be proportional to $\vec{p}_{\mathrm{T}}^{\,\mathrm{fj}}$. Accordingly, a high value means that the jet has its origin in this pile-up interaction with a high probability. A certain value of this discriminant can be used as a threshold. If a pile-up vertex with an fJVT discriminant above that value is found for the considered forward jet, it is tagged as forward pile-up jet. Different thresholds were used to define two fJVT working points: Medium (with fJVT $= 0.4$ as cut) and Tight (with fJVT $= 0.5$ as cut). Medium can be used with the jet selection working points Tight and Tighter, whereas Tigth can be used with Tighter and Tenacious as jet selection working points.

The EMPFlow reconstruction works differently, it includes also track information from the inner detector. It is extensively described in Reference [7]. Analogously to the proceeding with EMTopo, pile-up jets are intended to be filtered by applying JVT thresholds. For EMTopo, they are chosen differently for the existing working points, see Table 4.2. Also the fJVT algorithm can be applied, but in the case of Tenacious it is not usable.

Working point	p_T (all jets)	p_T (forward jets)	JVT (for jets with $\|\eta\| < 2.4$)
Loose	> 20 GeV	> 20GeV	> 0.2 (for $p_T < 60$GeV)
Tight	> 20GeV	> 30GeV	> 0.2 (for $p_T < 60$GeV)
Tighter	> 20GeV	> 35GeV	> 0.2 (for $p_T < 60$GeV)
Tenacious	> 20GeV	> 35GeV	> 0.91 (for $p_T < 40$GeV), > 0.59 (for $p_T < 60$GeV), > 0.11 (for $p_T < 120$GeV)

Table 4.2: Overwiew about the thresholds for the jet selection working points for the EMPFlow jet reconstruction

4.2.2 Other Hard Object Contributions

Electrons

Electrons are reconstructed by the consideration of detector signals from the inner detector and the electromagnetic calorimeter. According to the covered range of the calorimeter, they require to have $|\eta| < 2.47$ (with exclusion of the range $1.37 < |\eta| < 1.52$). They are neglected in the MET reconstruction if they do not satisfy $p_T > 5$ GeV. Additionally, the ID cut with working point Loose_BL and the isolation cut with LooseTrackOnly are applied.

Muons

Muons are reconstructed by the use of signals from the muon chambers in combination with data from the inner detector and in some cases also with signals from the calorimeters. Due to the construction of the muon chambers (coverage up to $|\eta| = 2.7$), only muons with $|\eta| > 2.7$ can be considered. Details about the muon reconstruction can be found in Reference [8]. In the MET reconstruction, muons are also rejected if they do not fulfill $p_T > 10$ GeV. Additionally, the ID cut with working point Loose and the isolation cut with FixedCutLoose are applied.

4.2.3 Determination of the Soft Term Contribution

For the soft term reconstruction, the used algorithm should be able to decide reasonably if a specific track signal in the detector can be associated with a hard-scatter vertex (in the MET reconstruction, a track-based soft term is used). In the case of the usage of EMTopo, this succeeds by the use of the following selection conditions (for tracks associated with hard-scatter vertices): the transversal momentum shall be above 0.4 GeV, the track should not be aligned in forward direction ($|\eta| < 2.5$) and

$$\left|\frac{d_0}{\sigma(d_0)}\right| < 2$$

as well as

$$|z_0 \sin\theta| < 3 \text{ mm}$$

should be satisfied. The last to conditions are geometrical conditions considering z_0 as the z-coordinate of the track point which is located the closest to the beam line and d_0 as the shortest distance from the track to the beam line in transversal direction (and $\sigma(d_0)$ its resolution). A certain list of classes is excluded from the soft term to assure that no particle is counted several times, although such signals would fulfill the selection criteria (see Reference [9]):

- tracks in the range $\Delta\eta < 0.2$ and $\Delta\phi < 0.05$ around photons and electrons
- tracks with a cone size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.2$ around hadronically decaying τ-leptons
- tracks from the inner detector associated with identified muons
- tracks matched to jets (as described in Section 4.2.1)
- isolated tracks with $p_T > 120$ GeV (> 200 GeV for $|\eta| < 1.5$) whose estimated relative resolution on their p_T is larger than 40 % or who have no energy deposit in the calorimeters with a p_T larger than 65 % of the track p_T, in order to reduce the influence of mismeasured tracks

For the usage of EMPFlow, the tracks not associated with reconstructed hard objects have to fulfill similar requirement, only minor differences appear in the definition: p_T has to be greater than 0.5 GeV and $|z_0 \sin\theta| < 3$ mm must be satisfied. Additionally, there is no constraint concerning d_0. Detailed information about the soft term are findable in Reference [10].

4.3 Conclusion: usable Variations

According to the already mentioned facts, there are several choices to be made for the application of the MET reconstruction algorithm. The first choice is the decision between EMTopo and EMPFlow. EMPFlow is expected to give better results because it includes additionally the track information from the inner detector. The second one concerns the selection of the working point. Four different working points are selectable (Loose, Tight, Tighter, Tenacious), they distinguish in respect to the "tightness" (strength of the p_T threshold and minimal JVT discriminant). The third choice is the decision if the fJVT algorithm shall be used or not. It is also expected to improve the result by removing pileup jets in the forward direction. By the combination of these three options and taking into account that fJVT cannot be used for Tenacious in the case of the usage of EMPFlow, 15 different variations exist which could be chosen. This bachelor thesis is targeted on the comparison of the results generated by their use in the MET reconstruction. In the optimal case, an appropriate choice would be found.

5 Performance of MET Studies

5.1 Used Samples

To investigate the MET reconstruction performance, the reconstruction algorithm with its variations is applied to Monte Carlo simulations of the process $W^{\pm}Z \rightarrow l\nu lljj$. In this way, the results can be compared with the truth values of the MET which would be obtained in the optimal case. Therefore, the sample group corresponding to that process (DSID 364284) was used. The considered samples are part of the campaigns MC15 and MC16. From MC16 the subcampaigns MC16a and MC16d were included in the MET studies. The MC16a simulation bases on data taking in 2015 and 2016, whereas MC16d was generated based on the data taking in 2017. There are two major differences between them. At first, the number of contained events differs significantly, MC16a includes about 60.000 events in comparison to MC16d with only about 6.000. Accordingly, the MET reconstruction applied to MC16d implies higher relative uncertainties. Despite this disadvantage, it is reasonable to use that sample in the testing phase of the studies because of the shorter computing time. The second difference is the higher proton density of the bunches leading to significantly higher numbers of interactions per bunch crossing in MC16d. To compare these two samples with respect to the pile-up, the Plots 5.1 and 5.2 are given. They show the distribution of the average number of pile-up vertices per bunch crossing N_{Pileup} for the MC16a and MC16d simulation data. This number is determined by averaging the number of pile-up vertices over all bunch crossings occurring within one lumi block which lasts for about one minute. For all collisions during one lumi block, the conditions are relatively homogeneous. If a certain amount of protons is accelerated initially, the bunches have a high proton density. Accordingly, N_{Pileup} is large in the beginning. With the proceeding of the time, the proton density decreases due to the collisions. This implies a reduction of N_{Pileup}. In that way, a continuous distribution of N_{Pileup} is produced.
MC15 is also based on the detector conditions and the data taking in 2015 and 2016. In contrast to MC16, it is an analysis reconstructed with release 20.7 of Athena. The results obtained by the usage of this campaign can therefore serve as reference for the reconstruction with release 21. Logically, it is the aim to proof which of the existing working points leads to results which are similar or even better than the ones generated with the older version.
The regarded data can be limited on certain phase space regions. Throughout this bachelor thesis the (fiducial) inclusive phase space region is used. Its definition can be found in the appendix.

Details about the Monte Carlo samples can be found in Reference [1].

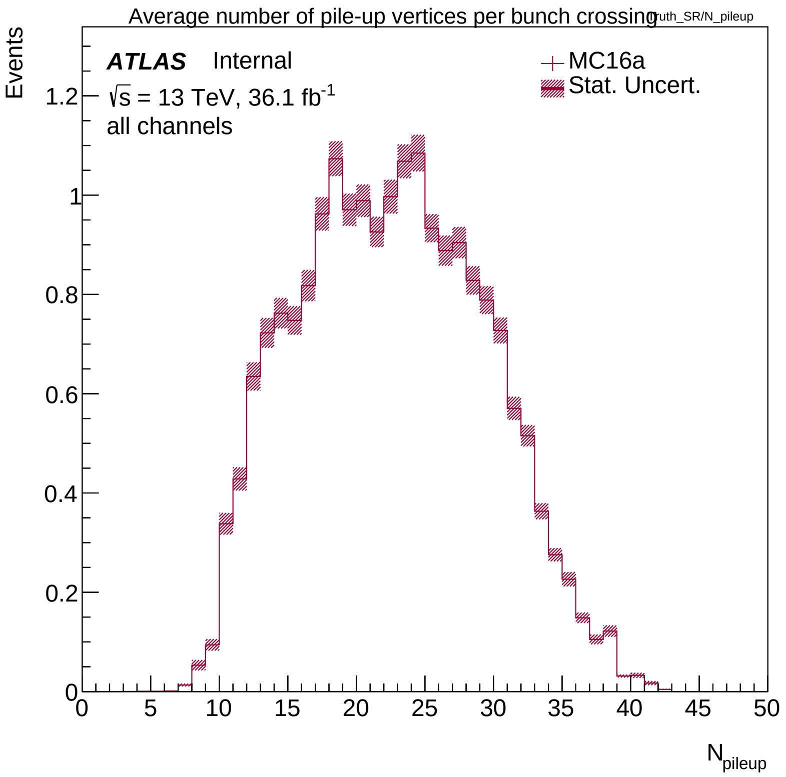

Figure 5.1: Occurrence of different average numbers of pileup vertices per bunch crossing for the MC16a data, according to the 2015+2016 data taking (in arbitrary units)

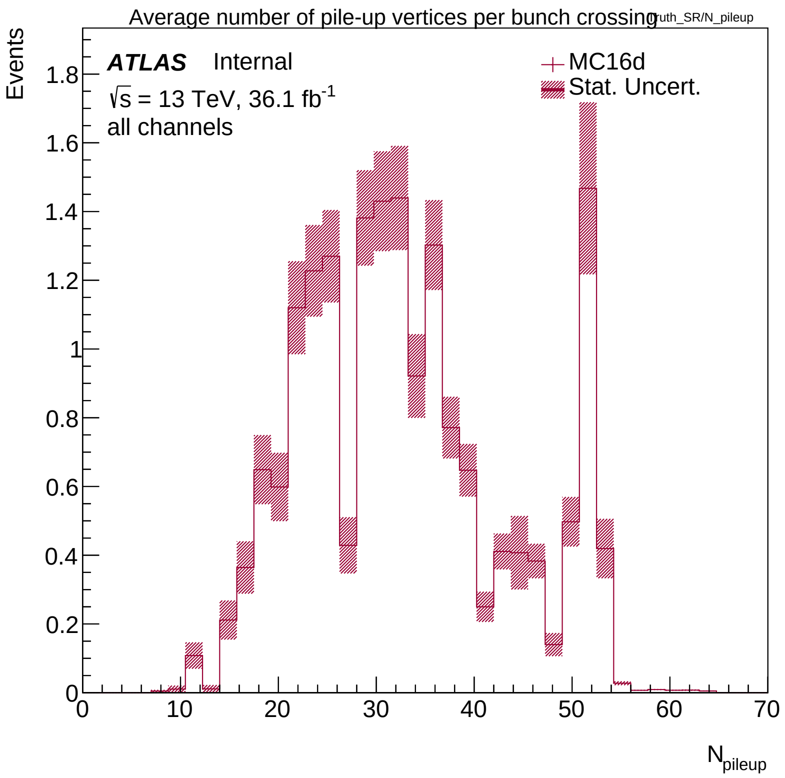

Figure 5.2: Occurrence of different average numbers of pileup vertices per bunch crossing for the MC16d data, according to the 2017 data taking (in arbitrary units)

5.2 Fitting

In this section, the two function types used for the fitting are described.

5.2.1 Gaussian Fit

The Gaussian distribution is a symmetric, continuous propbability density function which is very commonly applied. It depends on only two parameters: the mean value x_0 and the standard deviation σ. Its definition is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{(x-x_0)^2}{2\sigma^2}} \tag{4}$$

The values for the difference between the truth and the reconstructed values of the MET are supposed to be approximately Gauss-distributed (see [9]). The mean value and the standard deviation obtained from Gaussian fittings are thus commonly used to characterize the distrubution and the MET reconstruction itself. In this context, the standard deviation is called *MET resolution* within this bachelor thesis because it describes how strongly the values for the difference are scattered and therefore how precisely the reconstruction algorithm works. Similarly, the mean value is called *shift*. It describes which differences between the truth and the reconstructed value of the MET are more likely and which of them are less likely. In the optimal case, the MET resolution as well as the shift would reach values close to zero. In general, the focus lies on the resolution value because it allows conclusions about the ability of the reconstruction algorithm to achieve results in a narrow range around the most probable value. For example, in a combination of a large shift value with a small resolution value, the latter one would have a higher weight in an evaluation of the result. The large shift value could easily be corrected by adding a constant to the reconstructed MET.

5.2.2 Crystal Ball Function

The second function used for the fitting within this bachelor thesis is a type of the so-called Crystal Ball function. It is a function related to the Gaussian function. Around the most probable value it is indeed Gauss-shaped, the difference concerns the region beyond the central part. One or both tails of the function are defined by an exponential decline. By choosing the transition points between Gaussian part and exponential part and also by adapting other parameters (depending on the definition), a possible asymmetry of the distribution can be taken into account. Within this bachelor thesis, a less common definition is used:

$$\begin{aligned}
f(x) &= C \cdot e^{-\frac{t^2}{2}} && \text{for } \alpha_{\text{low}} \leq x \leq \alpha_{\text{high}}\ ;\\
f(x) &= C \cdot e^{\frac{\alpha_{\text{low}}^2}{2} - \alpha_{\text{low}} t} && \text{for } x \leq \alpha_{\text{low}}\ ;\\
f(x) &= C \cdot e^{\frac{\alpha_{\text{high}}^2}{2} - \alpha_{\text{high}} t} && \text{for } \alpha_{\text{high}} \leq x
\end{aligned} \tag{5}$$

with the abrevation $t = \frac{x-x_0}{\sigma}$. Due to this definition, the function is continuous and continuously differentiable in the transition points. Obviously, there are five parameters: the mean value x_0, the standard deviation σ, the transition points α_{low} and α_{high} (calculated as the distance to the mean value in multiples of σ) and the normalization constant C. Similarly to the terms chosen for the Gaussian fitting, the standard deviation is also called MET resolution and the mean value is referred to as shift. These two values obtained from the fitting with both functions are apparently not directly comparable.

5.3 Test Reconstruction with MC16d Sample

The first step of the practical work was getting an overview of the reconstruction performance in dependence on the used working point. For this part, the MC16d sample was used in order to profit from the reduction of the computing time. According to the aim, it was not necessary to regard larger amounts of data initially. Consequently, the reconstruction algorithm could be applied to the data without the need to limit the consideration on a subset of the usable working points.
The MET reconstruction in the practical part of this bachelor thesis is done by the usage of the analysis framework ELCore which is maintained at TU Dresden. It is based on the ATLAS software Athena. It accomplishes all the reconstruction work including the MET reconstruction. Its output files contain all information about the objects included in the data. For different properties (like the number of jets and the MET reconstructed with both EMTopo and EMPFlow) it yields the reconstructed as well as the truth values.
As already indicated, the evaluation of the reconstruction results occurs by regarding the distribution of the values for the difference between the truth and the reconstruction values of the MET. For the search for qualitative differences between the distributions belonging to the various working points, it seems reasonable to plot them. Here, the plotting of the results is realized by the usage of the so-called common analysis framework (CAF), another framework maintained at TU Dresden specifically for this purpose. In Figure 5.3 and 5.4, the plotted distributions for the EMTopo and the EMPFlow reconstruction and the jet selection working point Loose without the application of the fJVT algorithm are shown exemplarily. The plotting results for all the other applicable variations can be found in the appendix. Apparently, there are no significant qualitative differences between them. Indeed, the supposed Gaussian shape can be observed. A closer look reveals a certain asymmetry. The decline from the maximum to the right tail is slightly steeper, the left side, on the contrary, drops off more slowly. The left tail seems to be more elongated than the right one. However, the columns are generally afflicted with comparatively large relative uncertainties, some details could potentially not be resolved.

The second step of the evaluation is the consideration of the quantitative results for the MET resolution and the shift, as they were defined in Section 5.2. Therefore, a script was

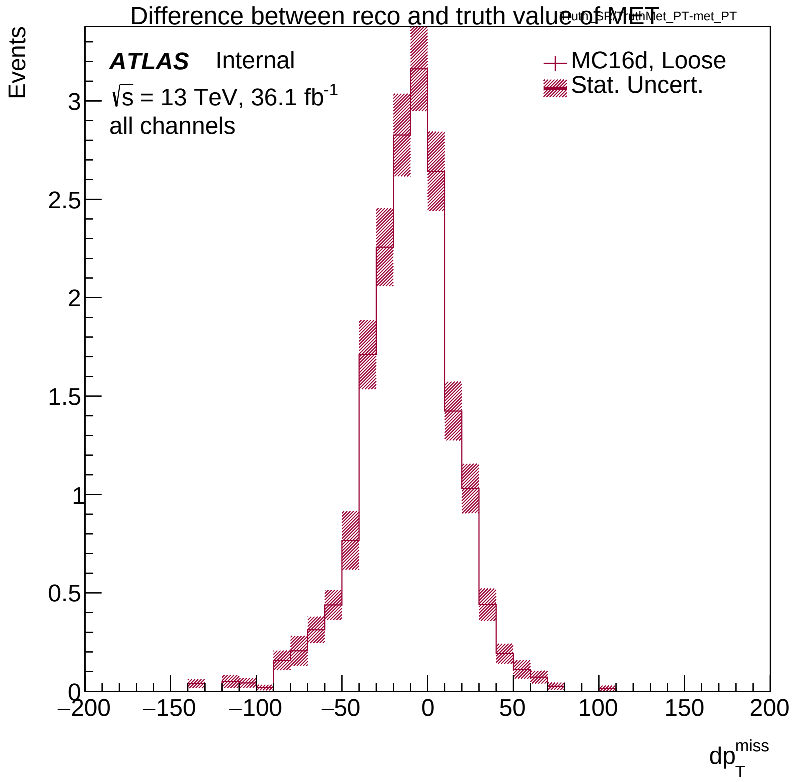

Figure 5.3: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Loose without fJVT, used with EMTopo reconstruction

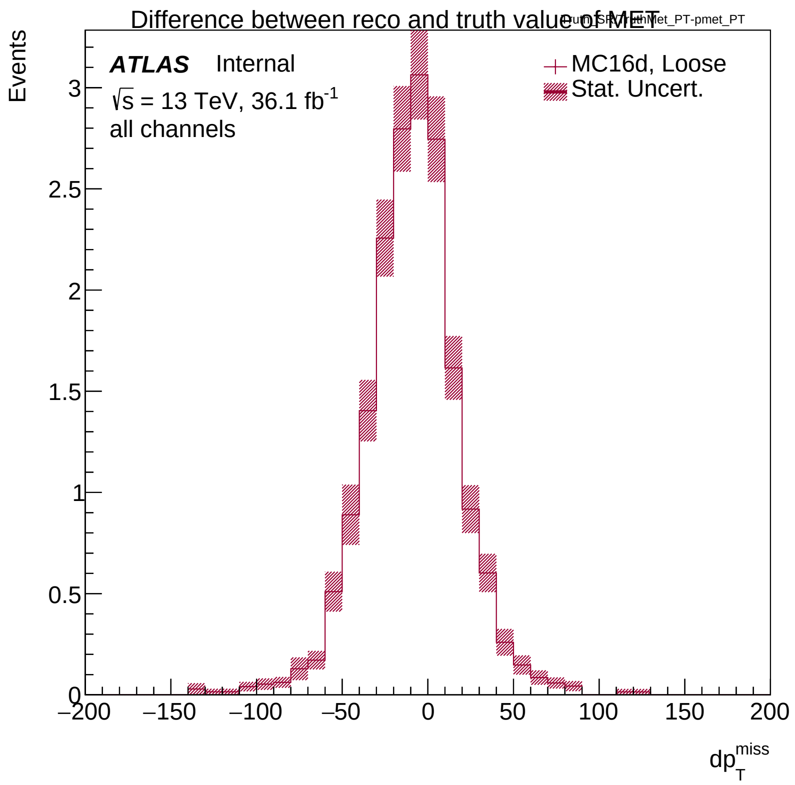

Figure 5.4: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Loose without fJVT, used with EMPFlow reconstruction

prepared which could be executed with CAF applying a Gaussian fit on the chosen distribution obtained from the ELCore output file. Figure 5.5 shows the result for the jet selection working point Loose (without fJVT, EMTopo). There are some significant deviations, particularly the maximum and the left tail of the distribution differ strongly from the fitted function. The ratio of χ^2 calculated by CAF for the fitting and the number of degrees of freedom indicates also major discrepancies: $\chi^2/N_{\mathrm{Df}} = 5.349$. Obviously, another more suitable function should be used to achieve meaningful values for the MET resolution and the shift. Considering the purpose, getting an overview of the results for existing variations and determining if there are significant differences between them, the Gaussian fit can be regarded as sufficient.
The following table contains the resulting values of the fitting parameters:

Working points		EMTopo		EMPFlow	
Jet selection	fJVT	MET res. / GeV	Shift / GeV	MET res. / GeV	Shift / GeV
Loose	-	22.5 ± 0.3	-9.3 ± 0.4	22.8 ± 0.4	-7.9 ± 0.4
Tight	-	22.2 ± 0.4	-8.6 ± 0.4	21.6 ± 0.4	-7.0 ± 0.4
Tighter	-	22.5 ± 0.4	-9.1 ± 0.4	21.6 ± 0.4	-7.1 ± 0.4
Tenacious	-	22.5 ± 0.4	-9.1 ± 0.4	21.5 ± 0.4	-7.1 ± 0.4
Tight	Tight	22.1 ± 0.4	-8.4 ± 0.4	21.0 ± 0.4	-6.6 ± 0.4
Tighter	Tight	22.2 ± 0.4	-8.6 ± 0.4	21.6 ± 0.4	-7.0 ± 0.4
Tighter	Medium	22.5 ± 0.4	-9.0 ± 0.4	21.4 ± 0.4	-7.0 ± 0.4
Tenacious	Medium	22.2 ± 0.4	-8.7 ± 0.4	-	-

Primarily, it is noticeable that the shift values are negative. They are located in the interval between -10 GeV and -6 GeV. Accordingly, the reconstruction algorithm is more likely to yield MET values which are smaller than the truth value instead of being larger. Furthermore, the usage of EMPFlow instead of EMTopo for a certain variation leads to a significant reduction of the shift value. The difference (in each case around 2 GeV) is about five times larger than one standard deviation. This observation supports the assumption that EMPFlow causes an improvement of the MET reconstruction. But at the same time, it is also important to regard the MET resolution values because they have a higher expressiveness concerning the evaluation of the results. They show a less significant tendency of an improvement by the usage of EMPFlow. They differ not more than three standard deviations from each other. But nevertheless, this tendency is perceptible. Comparing the variations, there are no major differences observable despite the outstanding results for the jet selection working point Tight used with EMPFlow. There, the MET resolution value is even higher than the one obtained with EMTopo.
It was the aim of this part of the practical work to decide which variations are worth being further investaged by their application on the data of sample MC16a. Unfortunately, the differences do not suffice to state clearly which of them should be chosen. Even the results

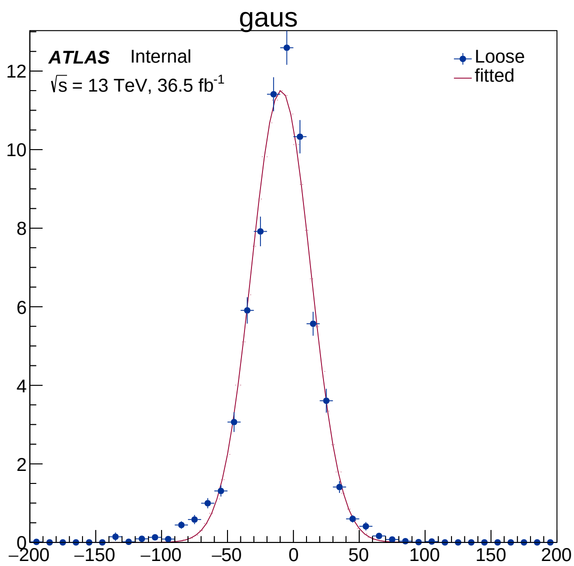

Figure 5.5: Gaussian fit applied to distribution of difference between truth and reconstructed value of MET for jet selection working point Loose without fJVT (sample MC16d)

for the jet selection working point Loose do not deviate that strongly that they can be easily neglected. Because of the absent indications, the decision has to be made differently. It appears reasonable to involve the default choice of working points (Tight without fJVT) into the consideration. Additionally, one version with a tighter and one with a less tight jet selection should be chosen to test if the different thresholds change the result in a positive way. Due to these intention, the jet selection working points Loose and Tenacious (each without fJVT) are included in the following considerations. Finally, a variation including the fJVT algorithm should also be used. Therefore, the jet selection working point Tight was selected combined with fJVT working point Tight.

5.4 Reconstruction and Comparison of MC15 and MC16a Sample

The proceeding with the samples MC15 and MC16a happened in analogy to the consideration of the sample MC16d. The important difference was the limitation of the used variations. According to the considerations in the last section, the following four variations were used for MC16a: Loose without fJVT, Tight without fJVT, Tight with Tight fJVT and Tenacious without fJVT. So far, the combination of the jet selection working point Tight without fJVT is used as default. The MC15 sample was only used with that default setting because it serves merely as reference.

By the application of the chosen working points, the corresponding output files containing the reconstruction results were generated with ELCore. The plotting was again done by the help of CAF. The results of the plots can also be found in the appendix, the example for the Loose working point is again shown here in the main part (for EMTopo and EMPFlow, see Figures 5.6 and 5.7). Obviously, the characteristics of the distribution shape which were already visible in the distributions for the sample MC16d are confirmed. The drop at the right side of the maximum is steeper than the one at the left side, the left tail is more elongated in comparison to the other one. Due to the much higher number of events included in MC16a, the relative uncertainties are significantly reduced. Consequently, the conclusions about the qualitative properties of the distributions are now more reliable.

It is now more profitable than before to determine which fitting function should be used to obtain meaningful and comparable results for the values of the MET resolution and the shift. The results for MC16d have already indicated that the Gaussian fit is not the optimal choice because there were significant deviations, especially around the maximum and in the region of the left tail. Figure 5.8 shows the result for the Gaussian fitting applied to the distribution belonging to the jet selection working point Tight without fJVT. The same specific deviations can be found again. Particularly the discrepancy concerning the maximum is problematic and it is to be feared that this fact has a strong influence on the values. Instead of the Gaussian distribution, a fitting function should be used which takes into account that the distributions are slightly asymmetric and the tails decline more slowly. Some research led to the assumption that the Crystal Ball function could fit significantly better. This function is divided into three parts: a Gauss-shaped center and two exponential tails (in the definition which is used here, see Section 5.2).

Therefore, the Crystal Ball function fitting was implemented in a script executable with CAF. The application of this fit leads to a result that can be examined in Figure 5.9. By naked eye, it already appears obvious that the attained fitted function deviates clearly less from the distribution. In the region of the tails, there are no differences visible anymore. Also around the maximum, the remaining deviations are hardly recognizable. By comparing the values of χ^2/N_{Df} for the mentioned example, the assumption can be confirmed impressively: for the Gaussian fit it yields 48.40, while for the Crystal Ball function fitting it yields 2.61. For

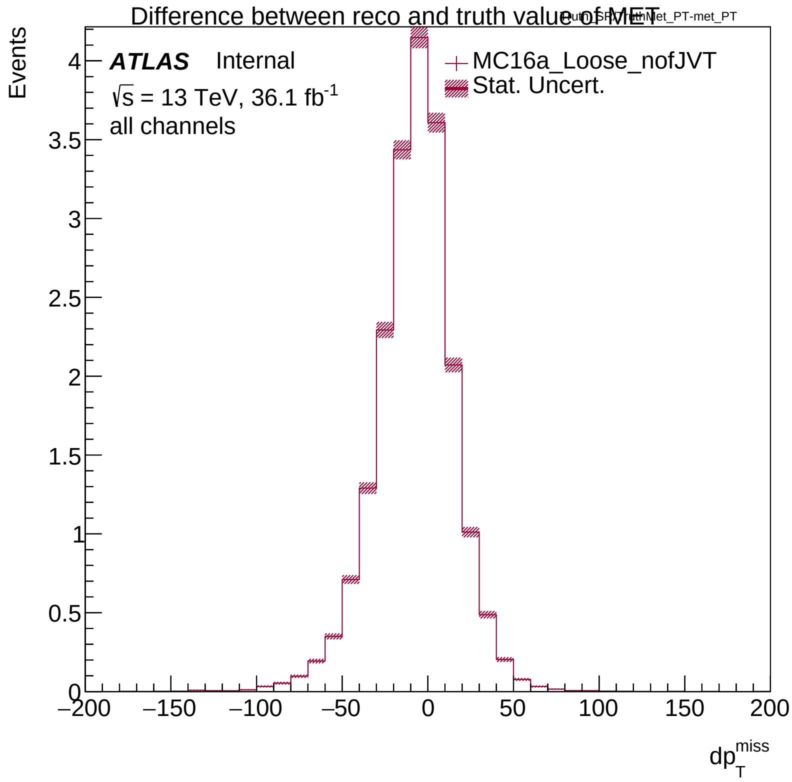

Figure 5.6: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Loose without fJVT, used with EMTopo reconstruction

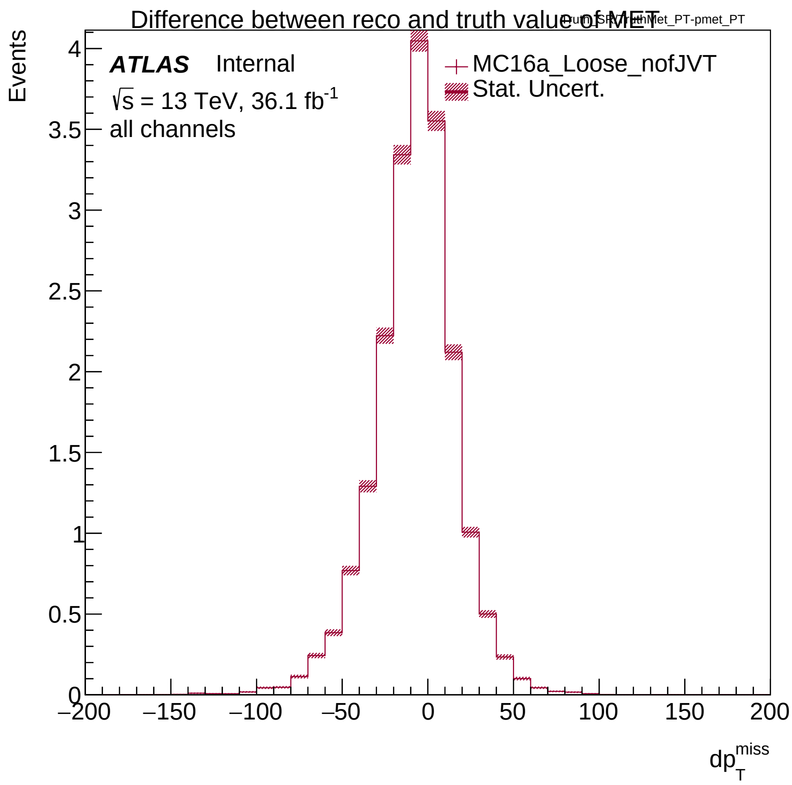

Figure 5.7: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Loose without fJVT, used with EMPFlow reconstruction

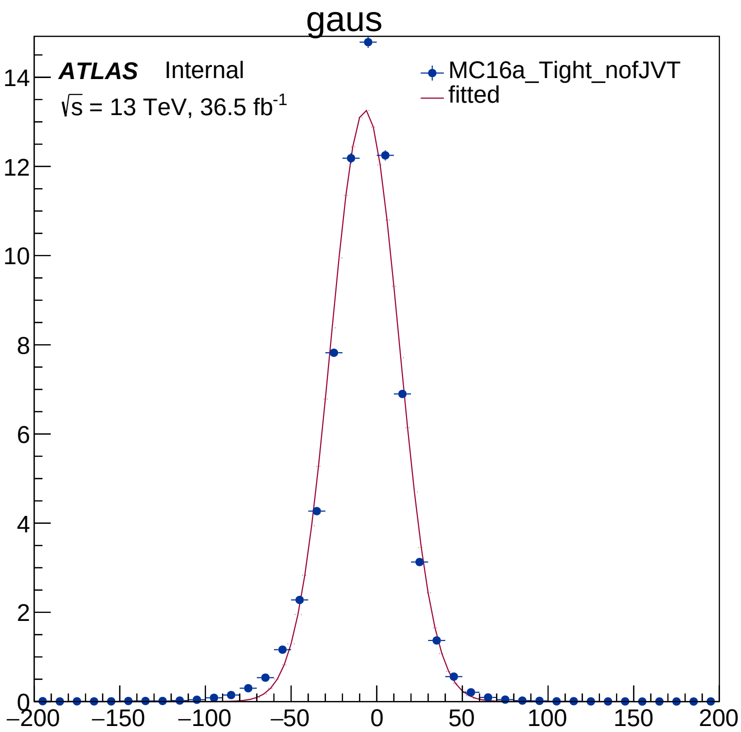

Figure 5.8: Gaussian fit applied to distribution of difference between truth and reconstructed value of MET for jet selection working point Tight without fJVT (sample MC16a)

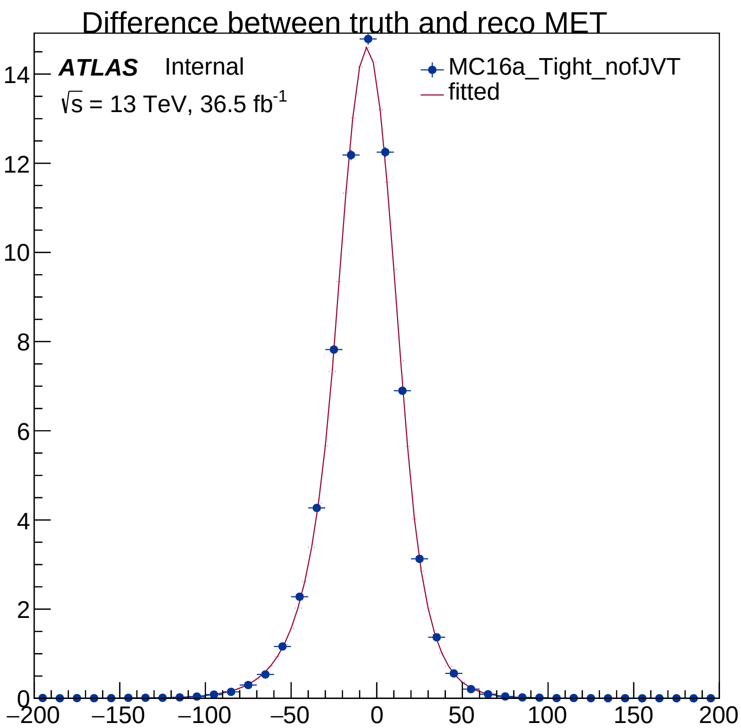

Figure 5.9: Crystal Ball function fit applied to distribution of difference between truth and reconstructed value of MET for jet selection working point Tight without fJVT (sample MC16a)

all other considered variations applied to the MC16a sample there are no major qualitative differences concerning the Crystal Ball function fitting, also the χ^2/N_{Df} values differ only negligibly from each other. The same applies to the comparison between the MC15 and the MC16a sample as well. For this reason, that function was chosen for the further proceeding.
After the decision, the MET resolution and shift value determination for all variations was carried out. The mentioned fitting script for CAF was executed for all considered MET reconstruction output files. The following table contains the obtained fitting parameters for each variation:

Sample	Jet selection	fJVT	Shift / GeV	MET res. / GeV	α_{low}	α_{high}	χ^2/N_{Df}
MC15	Tight (EMTopo)	-	-5.58 ± 0.10	16.60 ± 0.17	-1.098 ± 0.018	1.413 ± 0.028	2.68
MC16a	Loose (EMTopo)	-	-5.81 ± 0.09	17.12 ± 0.15	-1.102 ± 0.016	1.460 ± 0.029	3.06
MC16a	Loose (EMPFlow)	-	-5.63 ± 0.10	17.11 ± 0.15	-1.050 ± 0.015	1.378 ± 0.026	2.20
MC16a	Tight (EMTopo)	-	-5.73 ± 0.10	17.11 ± 0.16	-1.089 ± 0.016	1.441 ± 0.030	2.61
MC16a	Tight (EMPFlow)	-	-5.62 ± 0.10	17.03 ± 0.15	-1.035 ± 0.015	1.351 ± 0.025	2.31
MC16a	Tenacious (EMTopo)	-	-5.91 ± 0.10	17.46 ± 0.16	-1.097 ± 0.016	1.447 ± 0.030	2.79
MC16a	Tenacious (EMPFlow)	-	-5.86 ± 0.10	17.40 ± 0.16	-1.032 ± 0.015	1.357 ± 0.025	2.33
MC16a	Tight (EMTopo)	Tight	-5.71 ± 0.10	17.10 ± 0.16	-1.089 ± 0.016	1.444 ± 0.029	2.21
MC16a	Tight (EMPFlow)	Tight	-5.60 ± 0.10	16.98 ± 0.16	-1.037 ± 0.016	1.346 ± 0.025	2.34

The normalization constant as the fifth parameter is neglected in theistable because it is neither needed for the consideration of the MET resolution and the shift nor does it help to understand the interesting characteristics of the fit (i.e. in which range it is Gaussian and where the exponential tails begin). As already mentioned, the χ^2/N_{Df} values indicate that the Crystal Ball function fits for all variations are almost optimal. The differences between them are small, therefore the yielded values for the MET resolution and the shift can be regarded as well comparable. Considering the α_{low} and α_{high} value, there are also significant similarities. The transition from the Gaussian center to the right exponential tail is situated always about one MET resolution in front of the maximum. In contrast, the transition to the right tail happens around 1.4 MET resolutions behind the maximum. This fact reflects the uniformity of the distributions and the obvious asymmetry of each of them, particularly the elongated left tail. It is observable that the Gaussian-shaped center is wider for the EMTopo results.
As also recognized in Section 5.3 with the results for sample MC16d, the values for the shift are negative again. Hence the MET is more frequently overestimated than underestimated. The shift values fluctuate in a relatively narrow range between -6.0 GeV and -5.5 GeV. Accordingly, there is no outstanding value which is notably low or high. Indeed, the result obtained by the usage of the sample MC15 with the jet selection working point Tight without fJVT yields the lowest shift value. From the MC16a results, the application of working point Tight with Tight fJVT and EMPFlow leads to the result with the lowest shift value, it is only

insignificantly higher than the first mentioned one. The highest values were obtained by the usage of Tenacious (without fJVT) for EMTopo and EMPFlow as well.
Similarly, the MET resolution values are not strongly scattered. They are all situated in a small range around 17 GeV. Nevertheless, some of them stand out: the result for the MC15 sample and for the MC16a sample with the application of Tenacious. The value determined for the reconstruction with sample MC15 is notably lower than the lowest MET resolution from the MC16a results (two standard deviations). In contrast, the results for sample MC16a with Tenacious reach about 17.4 GeV and thus they are two standard deviations above the other MC16a results. These are all located within one standard deviation mutually near 17.1 GeV. The deviation of the MC15 results can be explained by the fact that version 20 of Athena has been corrected several times in order to increase the quality of the results. Release 21 will be continuously improved by corrections, a decrease of the MET resolution values can therefore be expected. The difference between the values for the sample MC16a with Tenacious and the values for the other variations applied to the same sample indicates that the MET reconstruction algorithm leads to slightly less reliable results if it is used with Tenacious. Most probably, the higher p_T-cut causes this tendency because it implies that more jets are rejected although many among these higher energetical ones lead to non-negligible contributions to the MET. This difference is well perceivable but cannot be regarded as highly significant. Accordingly, the conclusion that the usage of Tenacious implies definitively worse results and should thus not be applied is inappropriate in this case.
Further observations concerning the MET resolution are that the values obtained by the application of EMPFlow are smaller than the ones obtained with EMTopo but this effect has only a minimal extent. The differences are generally small in comparison to one standard deviation. Only for the jet selection working point Tight with Tight fJVT, the values differ by almost one standard deviation. Obviously, the usage of EMPFlow does not influence the result in a significant way although a perceptible improvement was expected due to the additional inclusion of track information from the inner detector. Also, the usage of the fJVT algorithm does not lead to remarkable changes. In comparison to the results for Loose and Tight without fJVT, the MET resolution value is only slightly reduced, the difference is small in comparison to one standard deviation. The indications from the outcome in section 5.3 can thus not be confirmed. One have to assume that the observable differences were a consequence of the suitability of the fits which distinguished between EMTopo and EMPFlow. As the Crystal Ball function fitting results pointed out, the Gaussian-shaped center for example has always been smaller for the usage of EMPFlow.

5.5 Dependence on the Pile-up

In this last section of the main part of the bachelor thesis, the pile-up dependence of the MET reconstruction basing on the different variations is regarded. The consideration of this topic was also motivated by the results described in the last section where no significant difference between the different working points and between EMTopo and EMPFlow was observed.

The pile-up distributions for the sample MC16a was already visible in the Figure 5.1. The distribution for MC15 is can be found in the appendix, there are no major qualitative or quantitative differences observable. The idea for the investigation of the pile-up dependence is dividing the samples into subsets with respect to the average number of pile-up interactions per bunch crossing N_{Pileup}. On each of these data subsets, which are called *pile-up bins* here, the Crystal Ball function fitting is applied to obtain values for the MET resolution and the shift. Finally, the changes of these values in dependence on the pile-up bin can be considered. In order to minimize the statistical uncertainties, the pile-up should be chosen in a way that assures that every bin contains almost the same numbers of events. Additionally, their number should not be too high to have still an appropriate number of events in each of them for a higher significance. Finally, the following division into four bins was made:

Index of pile-up bin	Range
1	$0\text{ GeV} \leq N_{\text{Pileup}} \leq 17\text{ GeV}$
2	$18\text{ GeV} \leq N_{\text{Pileup}} \leq 24\text{ GeV}$
3	$25\text{ GeV} \leq N_{\text{Pileup}} \leq 31\text{ GeV}$
4	$32\text{ GeV} \leq N_{\text{Pileup}}$

The division is visualized in Figure 5.10.

On each of the created subsets of data, the Crystal Ball function fitting was applied analogously to the proceeding with the whole data in the last section. In the appendix, a table containing the fitting results for each pile-up bin for all the selected variations can be found, including the MET resolution and the shift with their uncertainties as well as the χ^2/N_{Df} value to appreciate the suitability of the fits. The latter ones are very similar. They are scattered in a narrow range between 0.76 and 2.40. Accordingly, specific observations concerning the tendency of the pile-up dependence for a certain variation or concerning the comparison of several among are very likely no consequence of less suitable fittings. Instead, it can be regarded as an implication of the different functioning of the applied variations.

Some of the dependencies are depicted in the Figures 5.11 to 5.13. For every variation, there is the same general tendency observable: higher N_{pileup} imply larger MET resolution values. This observation corresponds to the expectation, as more hard-scatter vertices lead logically to more pile-up vertices and thus more disturbing effects influence the MET reconstruction. However, for many variations, there is a slight drop recognizable at the transition from pile-up bin three

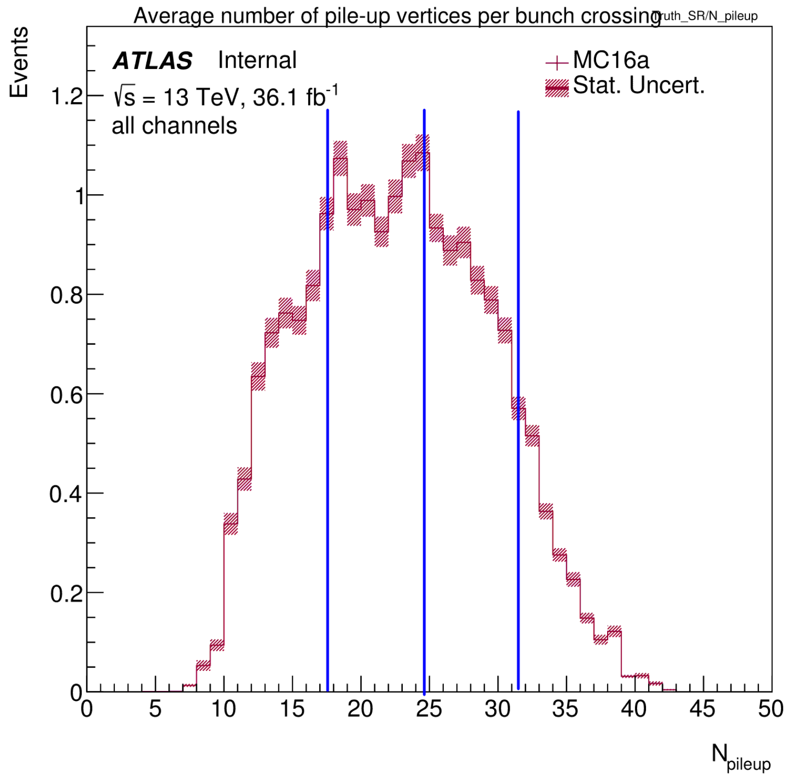

Figure 5.10: Visualization of the pile-up bin division of the data contained in sample MC16a: the three vertical blue lines symbolize the limits of the bins

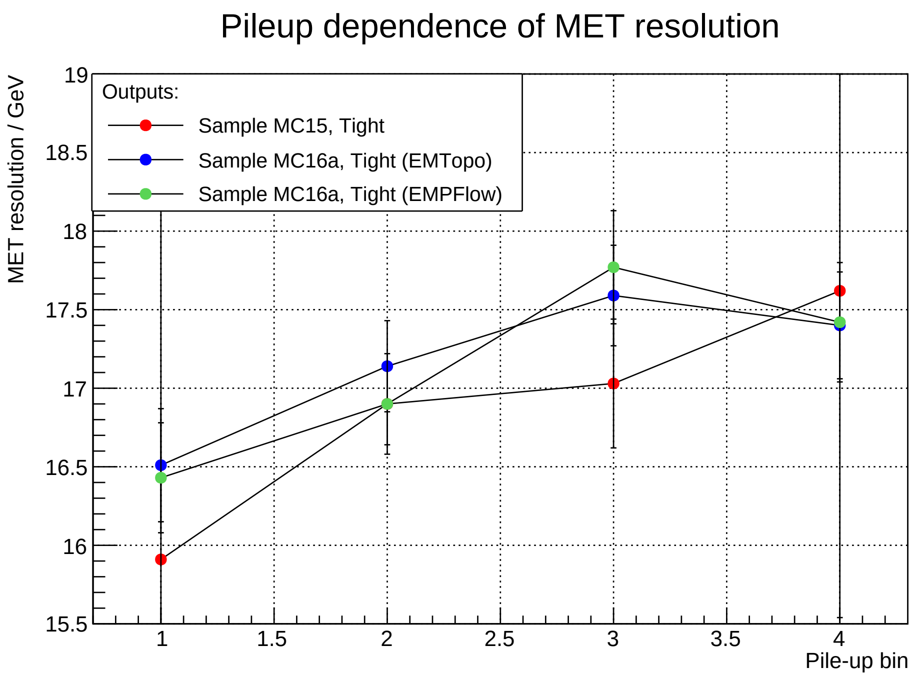

Figure 5.11: MET resolution in dependence on the average number of pile-up interactions per bunch crossing for jet selection working point Tight (without fJVT) for sample MC15 and sample MC16a used with and EMTopo and EMPFlow

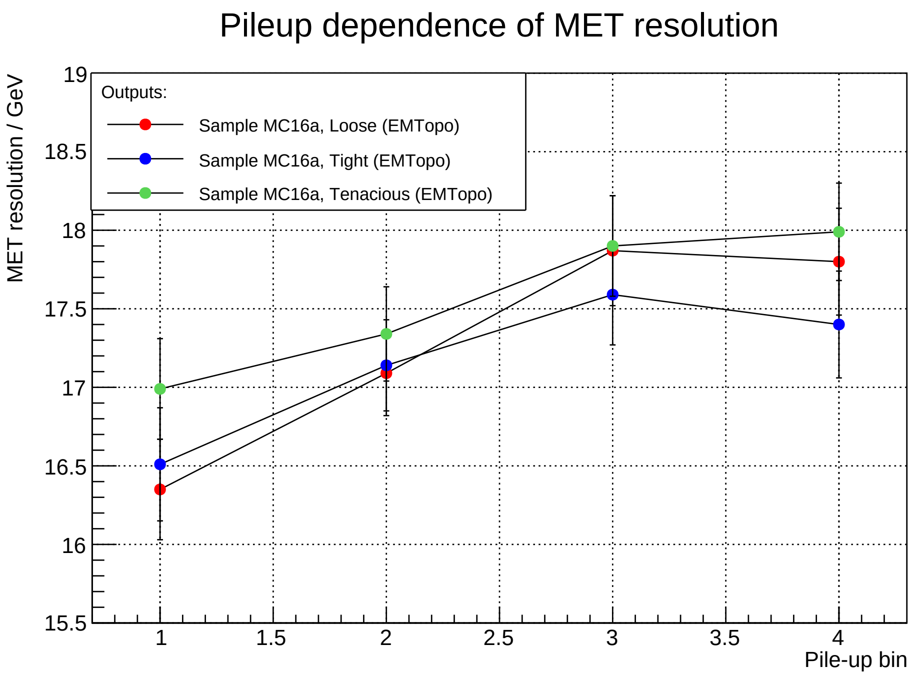

Figure 5.12: MET resolution in dependence on the average number of pile-up interactions per bunch crossing for jet selection working points Loose, Tight and Tenacious (without fJVT) used with EMTopo

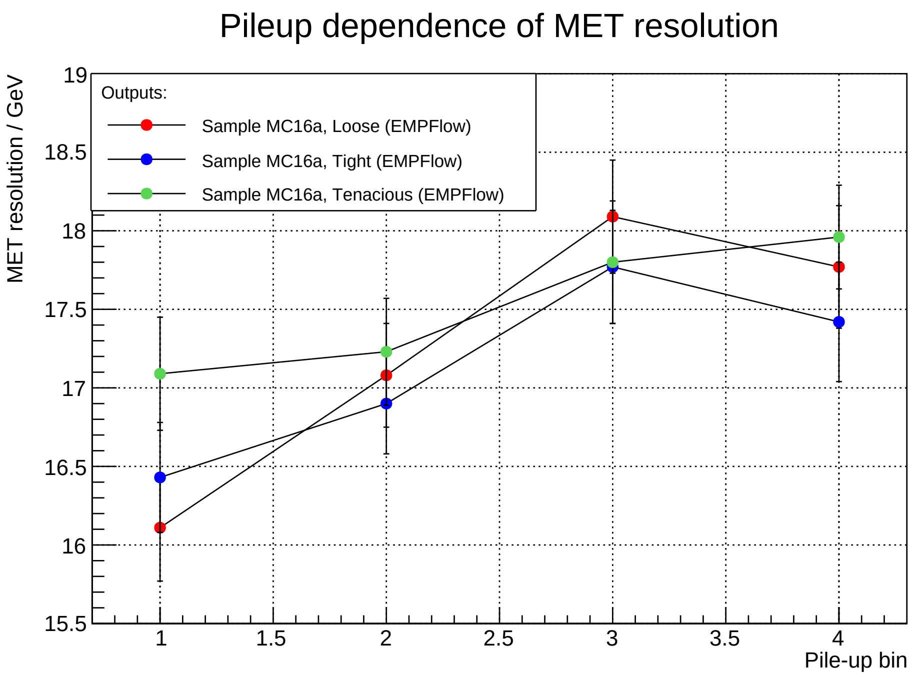

Figure 5.13: MET resolution in dependence on the average number of pile-up interactions per bunch crossing for jet selection working points Loose, Tight and Tenacious (without fJVT) used with EMPFlow

to bin four which contradicts the expectation. Only the MET resolutions achieved with sample MC15 and with the jet selection working point Tenacious show a different behaviour. On the one hand, the effect is observable in numerous cases. On the other hand, the drop cannot be regarded as significant, with respect to the uncertainties. Therefore, it would be questionable to state that this behaviour is caused by the functioning of the MET reconstruction algorithm or that it appears generally if the MET reconstruction is applied to data which is influenced by strong pile-up effects. So far, no reason could be found which explains this observation in a plausible way. Potentially, it is caused by a particularity of the used Monte Carlo sample instead of being a general effect.

Another observation confirms an assumption which was made regarding the results in Section 5.4. The pile-up dependences indicate again that the decision between the usage of EMTopo and EMPFlow has only a minimal influence on the result. The Plots 5.11 to 5.13 demonstrate that the application of EMPFlow leads to a slight improvement of the MET resolution in the pile-up bins one and two. But at the same time, it causes larger resolution values in the bins three and four so that the improvement effect in the first two bins is almost compensated if the whole data is included in the MET reconstruction.

Furthermore, the Figures 5.12 and 5.13 show that especially for low pile-up influence it is profitable to use a less tight jet selection working point. For EMPFlow, the MET resolution value for Tenacious is about 1 GeV larger than the values obtained with Loose and Tight (Loose leads in each case to a better resolution). Obviously, the usage of a tighter working point effects the negligence of some important contributions to the MET. The tendency of less tight working points to cause smaller MET resolution values disappears for higher pile-up influence.

Considering which variation leads to the most precise results for all bins, there is no clear conclusion possible. The usage of the jet selection working point Tight with EMPFlow and the application of fJVT (working point Tight) yields the smallest MET resolution value for three of the four bins (each despite of the first one) but the advantage is also not significant. Accordingly, the investigation of the pile-up dependence does also not succeed to reveal significant differences between the variations.

6 Summary and Outlook

The achieved results described in the last sections are approximately as expected if the variations are considered individually. The obtained distributions of the difference between the truth and the reconstructed value of the MET are very similar to a Gaussian distribution. The Crystal Ball function was found as a more suitable fitting function, therefore it was used for the application of the fitting on the MET reconstruction output files.
The comparison of the results for the different variations, in contrast ,revealed a behaviour which was not expected. The EMPFlow jet recosntruction, which can be used since recently, includes track information from the inner detector. It was thus thought to lead to a perceptible improvement of the MET reconstruction expressed by smaller shift and MET resolution values. A certain difference exists but with respect to the uncertainties it is only a minimal effect which cannot be regarded as significant. Also the fJVT algorithm did not effect the MET reconstruction in the desired way. By rejecting the contribution of forward jets which are probably originated in pile-up vertices, it was supposed to achieve a more precise MET reconstruction. The MET resolution values obtained by the usage of Tight fJVT (see Section 5.4) are the lowest ones with respect to the other MC16a results but also this difference is insignificant. Furthermore, the investigation of the pile-up dependence was similarly not able to reveal any major differences between the variations. It was found that the expected increase of the MET resolution value for a growing pile-up influence appears indeed although in many cases there is a still unexplained slight decline at the transition from pile-up bin three to four. Additionally, the investigation of the pile-up dependence indicates that it is more reasonable to use a less tight jet selection working point to achieve a more precise MET reconstruction in the region where less pile-up influence occurs.
Regarding this outcome, one has to make the conclusion that no definitive recommendation is possible asserting which variation is the most appropriate one for the described aim. The observed differences are far too weak to make a such decision. All in all, the outcome contradicts the expectation of a clear improvement by the usage of the recently available variations of the MET reconstruction algorithm. For future investigations of the MET reconstruction algorithm and its variations, it could be helpful to use Monte Carlo samples with a higher number of included events in order to reduce the statistical uncertainties further and increase the significance of occurring differences. Additionally, one could include background signals leading to the similar signatures in the detector and investigate their influence on the result.

Appendix

A Phase space definitions

Definition of phase space regions (extracted from reference [1]):

Variable	Total	Fiducial inclusive	Fiducial VBS
Lepton $\|\eta\|$	—	< 2.5	< 2.5
p_{T} of ℓ_Z, p_{T} of ℓ_W [GeV]	—	$> 15, > 20$	$> 15, > 20$
m_Z range [GeV]	$66 - 116$	$\|m_Z - m_Z^{\mathrm{PDG}}\| < 10$	$\|m_Z - m_Z^{\mathrm{PDG}}\| < 10$
m_{T}^{W} [GeV]	—	> 30	> 30
$\Delta R(\ell_Z^-, \ell_Z^+)$, $\Delta R(\ell_Z, \ell_W)$	—	$> 0.2, > 0.3$	$> 0.2, > 0.3$
p_{T} two leading jets [GeV]	—	—	> 40
$\|\eta_j\|$ two leading jets	—	—	< 4.5
Jet multiplicity	—	—	≥ 2
$\eta_{j1} \cdot \eta_{j1}$	—	—	< 0
m_{jj} [GeV]	—	—	> 500
$\Delta R(j, \ell)$	—	—	> 0.3
$N_{b-\mathrm{quark}}$	—	—	$= 0$

B Plots for MET reconstruction with sample MC16d

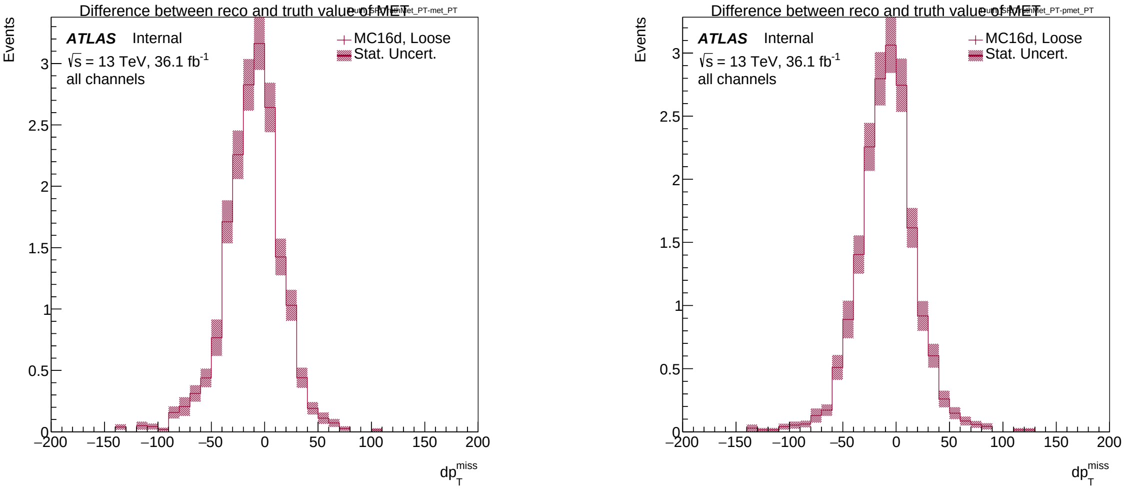

Figure B.1: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Loose without fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

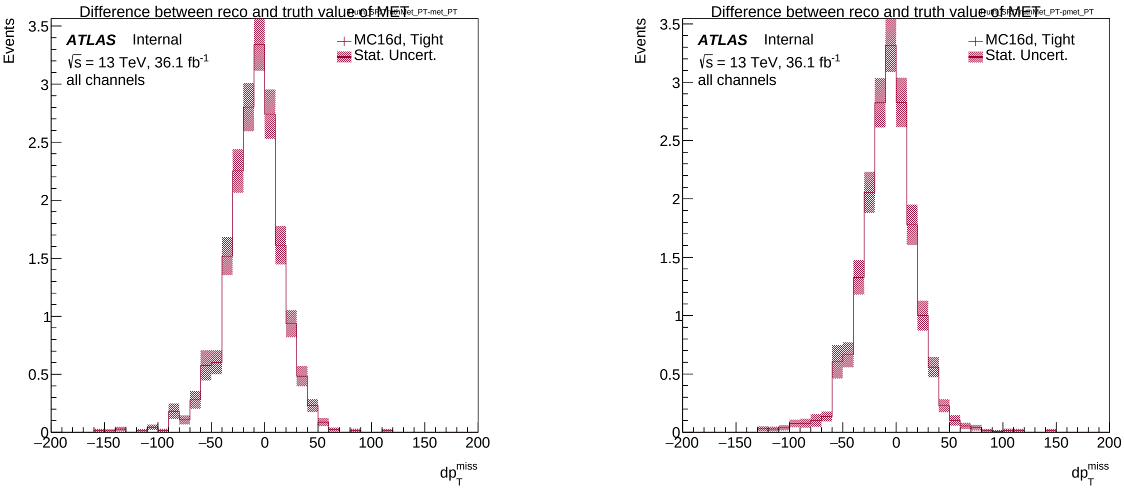

Figure B.2: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tight without fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

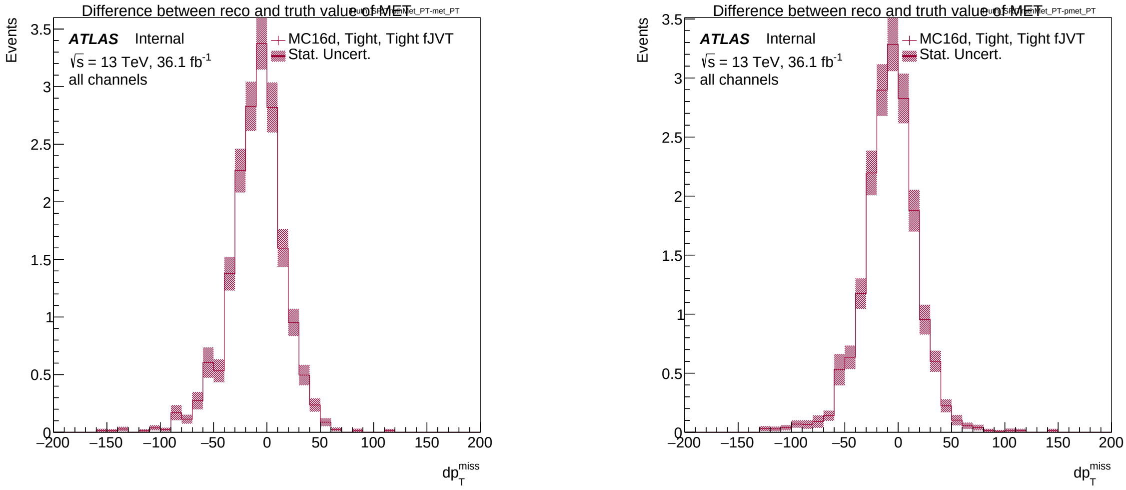

Figure B.3: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tight with Tight fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

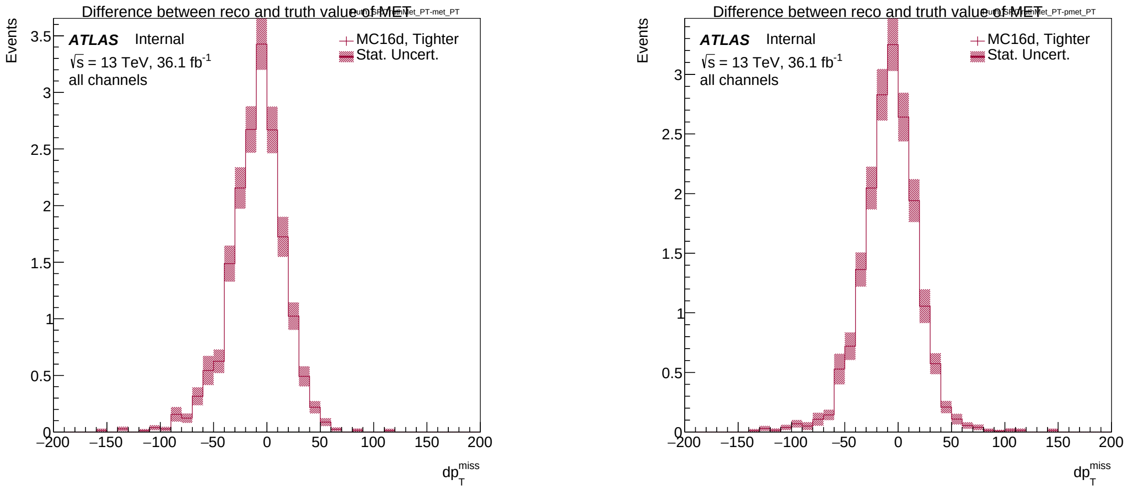

Figure B.4: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tighter without fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

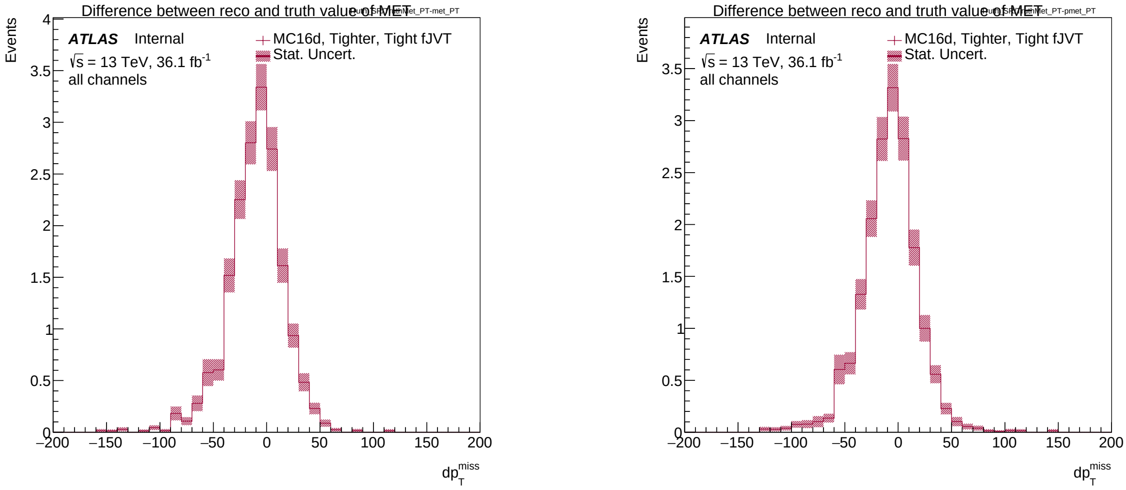

Figure B.5: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tighter with Tight fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

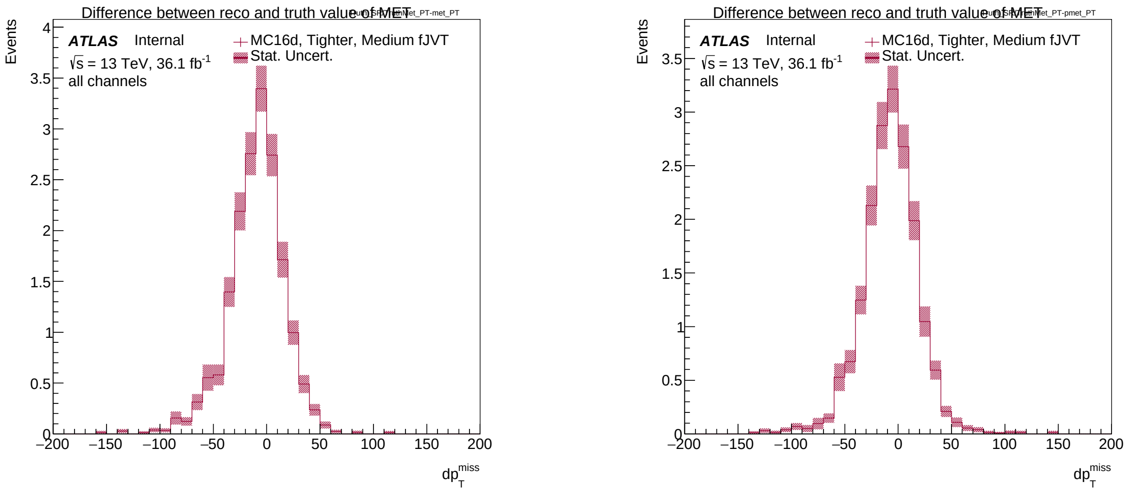

Figure B.6: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tighter with Medium fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

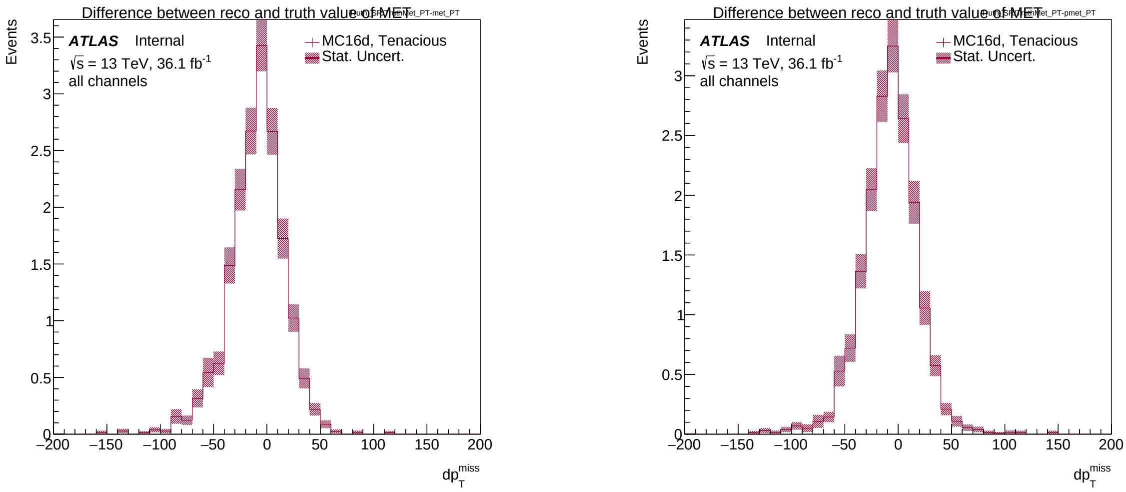

Figure B.7: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tenacious without fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

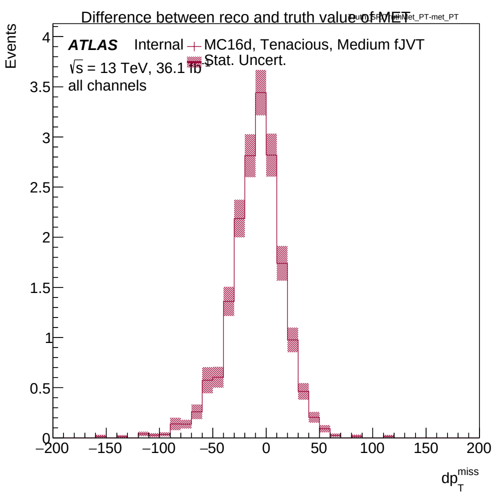

Figure B.8: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tenacious with Medium fJVT, used with EMTopo reconstruction

C Plots for MET reconstruction with sample MC16a

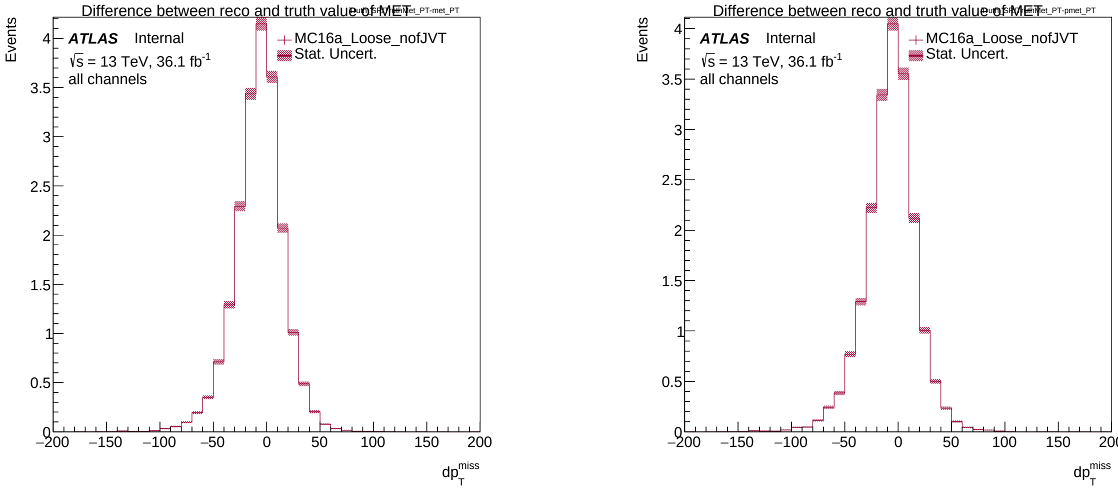

Figure C.1: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Loose without fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

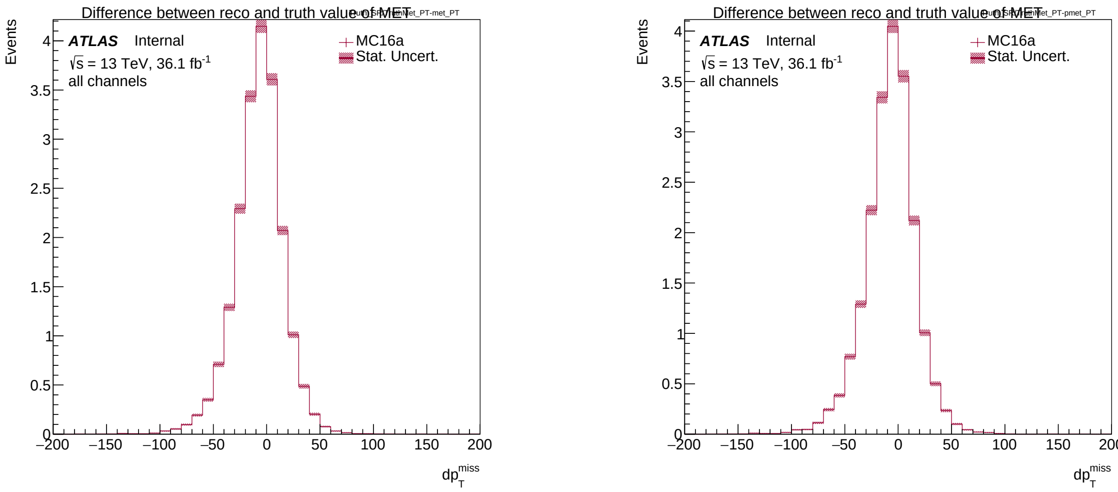

Figure C.2: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tight without fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

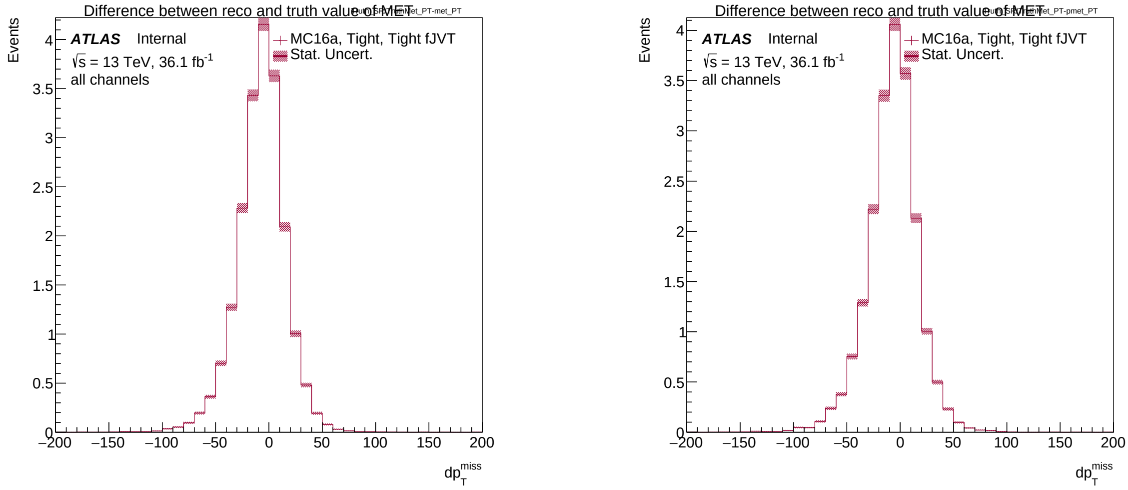

Figure C.3: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tight with Tight fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

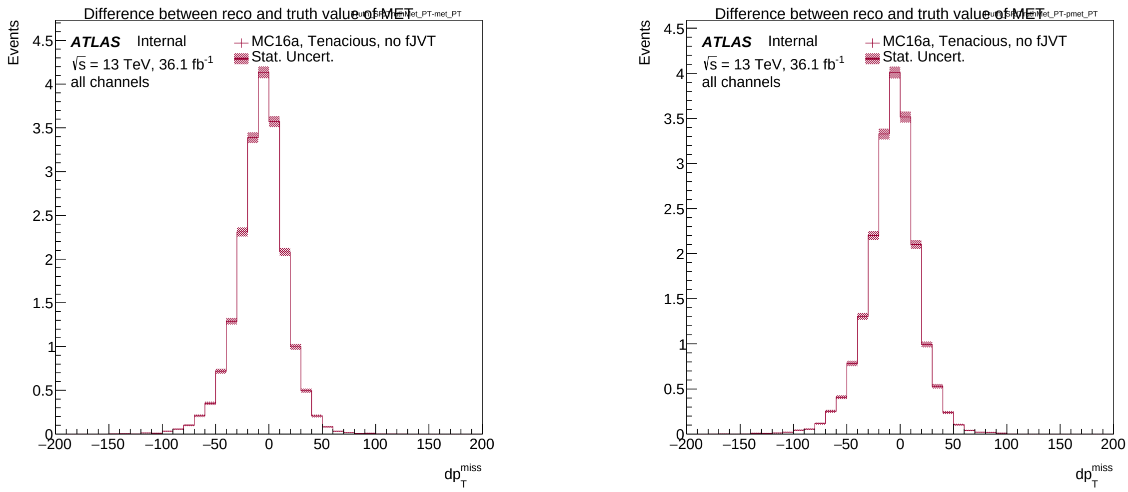

Figure C.4: Distribution of the difference between the truth and the reconstructed value of the MET for the jet selction working point Tenacious without fJVT, used with EMTopo (left side) and EMPFlow (right side) reconstruction

D Fitting results for selected variations (samples MC15 and MC16a)

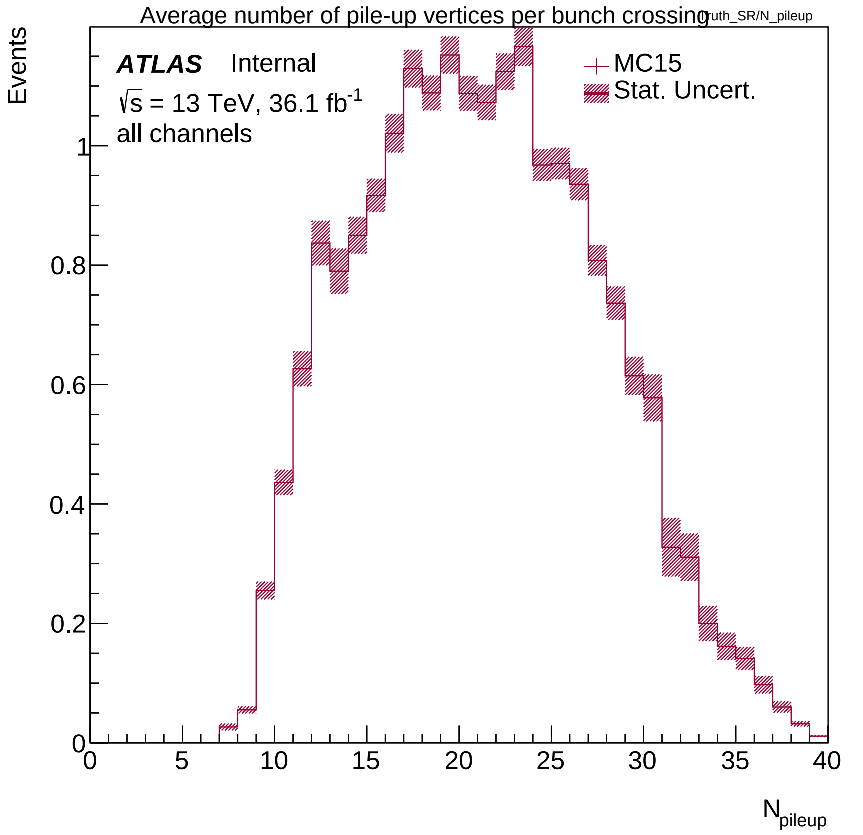

Figure D.1: Pile-up distribution for sample MC15

E Pile-up distribution for Sample MC15

Sample	Jet selection	fJVT	Shift / GeV	MET res. / GeV	α_{low}	α_{high}	χ^2/N_{Df}
MC15	Tight (EMTopo)	-	-5.58 ± 0.10	16.60 ± 0.17	-1.098 ± 0.018	1.413 ± 0.028	2.68
MC16a	Loose (EMTopo)	-	-5.81 ± 0.09	17.12 ± 0.15	-1.102 ± 0.016	1.460 ± 0.029	3.06
MC16a	Loose (EMPFlow)	-	-5.63 ± 0.10	17.11 ± 0.15	-1.050 ± 0.015	1.378 ± 0.026	2.20
MC16a	Tight (EMTopo)	-	-5.73 ± 0.10	17.11 ± 0.16	-1.089 ± 0.016	1.441 ± 0.030	2.61
MC16a	Tight (EMPFlow)	-	-5.62 ± 0.10	17.03 ± 0.15	-1.035 ± 0.015	1.351 ± 0.025	2.31
MC16a	Tenacious (EMTopo)	-	-5.91 ± 0.10	17.46 ± 0.16	-1.097 ± 0.016	1.447 ± 0.030	2.79
MC16a	Tenacious (EMPFlow)	-	-5.86 ± 0.10	17.40 ± 0.16	-1.032 ± 0.015	1.357 ± 0.025	2.33
MC16a	Tight (EMTopo)	Tight	-5.71 ± 0.10	17.10 ± 0.16	-1.089 ± 0.016	1.444 ± 0.029	2.21
MC16a	Tight (EMPFlow)	Tight	-5.60 ± 0.10	16.98 ± 0.16	-1.037 ± 0.016	1.346 ± 0.025	2.34

F Fitting results for the pile-up dependence

Sample	Jet selection	fJVT	Pile-up bin	Shift / GeV	MET resolution / GeV	χ^2/N_{Df}
MC15	Tight	-	1	-5.09 ± 4.60	15.91 ± 6.11	1.54
MC15	Tight	-	2	-5.67 ± 0.13	16.90 ± 0.26	1.25
MC15	Tight	-	3	-6.17 ± 0.23	17.03 ± 0.41	0.96
MC15	Tight	-	4	-5.88 ± 0.99	17.62 ± 2.08	2.40
MC16a	Loose (EMTopo)	-	1	-5.20 ± 0.18	16.35 ± 0.32	1.19
MC16a	Loose (EMTopo)	-	2	-5.92 ± 0.27	17.09 ± 0.27	1.84
MC16a	Loose (EMTopo)	-	3	-6.22 ± 0.18	17.87 ± 0.35	1.19
MC16a	Loose (EMTopo)	-	4	-6.46 ± 0.37	17.80 ± 0.34	1.18
MC16a	Loose (EMPFlow)	-	1	-4.85 ± 0.20	16.11 ± 0.34	1.04
MC16a	Loose (EMPFlow)	-	2	-5.76 ± 0.17	17.08 ± 0.33	0.98
MC16a	Loose (EMPFlow)	-	3	-6.09 ± 0.18	18.09 ± 0.36	1.10
MC16a	Loose (EMPFlow)	-	4	-6.43 ± 0.39	17.77 ± 0.39	0.76
MC16a	Tight (EMTopo)	-	1	-5.35 ± 0.19	16.51 ± 0.36	1.25
MC16a	Tight (EMTopo)	-	2	-5.95 ± 0.16	17.14 ± 0.29	1.40
MC16a	Tight (EMTopo)	-	3	-5.83 ± 0.17	17.59 ± 0.32	1.30
MC16a	Tight (EMTopo)	-	4	-5.86 ± 0.39	17.40 ± 0.34	1.28
MC16a	Tight (EMPFlow)	-	1	-5.14 ± 0.20	16.43 ± 0.35	0.86
MC16a	Tight (EMPFlow)	-	2	-5.82 ± 0.17	16.90 ± 0.32	1.25
MC16a	Tight (EMPFlow)	-	3	-5.78 ± 0.18	17.77 ± 0.36	1.36
MC16a	Tight (EMPFlow)	-	4	-6.02 ± 0.39	17.42 ± 0.38	0.83
MC16a	Tenacious (EMTopo)	-	1	-5.68 ± 0.18	16.99 ± 0.32	1.56
MC16a	Tenacious (EMTopo)	-	2	-6.03 ± 0.16	17.34 ± 0.30	1.34
MC16a	Tenacious (EMTopo)	-	3	-5.94 ± 0.18	17.90 ± 0.32	1.00
MC16a	Tenacious (EMTopo)	-	4	-6.12 ± 0.38	17.99 ± 0.31	2.16
MC16a	Tenacious (EMPFlow)	-	1	-5.54 ± 0.20	17.09 ± 0.36	1.02
MC16a	Tenacious (EMPFlow)	-	2	-6.01 ± 0.18	17.23 ± 0.34	1.45
MC16a	Tenacious (EMPFlow)	-	3	-5.86 ± 0.18	17.80 ± 0.39	1.18
MC16a	Tenacious (EMPFlow)	-	4	-6.34 ± 0.35	17.96 ± 0.33	0.93
MC16a	Tight (EMTopo)	Tight	1	-5.38 ± 0.19	16.53 ± 0.36	1.08
MC16a	Tight (EMTopo)	Tight	2	-5.92 ± 0.16	17.06 ± 0.29	1.22
MC16a	Tight (EMTopo)	Tight	3	-5.77 ± 0.17	17.61 ± 0.33	1.02
MC16a	Tight (EMTopo)	Tight	4	-5.86 ± 0.37	17.55 ± 0.30	1.34
MC16a	Tight (EMPFlow)	Tight	1	-5.14 ± 0.20	16.37 ± 0.35	0.85
MC16a	Tight (EMPFlow)	Tight	2	-5.79 ± 0.17	16.81 ± 0.32	1.32
MC16a	Tight (EMPFlow)	Tight	3	-5.73 ± 0.18	17.44 ± 0.35	1.10
MC16a	Tight (EMPFlow)	Tight	4	-5.96 ± 0.39	17.40 ± 0.38	0.88

Bibliography

[1] ATLAS Collaboration, *Study of electroweak $W^{\pm}Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*,.

[2] J. M. Ramos, *WZ diboson measurements with the ATLAS experiment at the LHC and Performance of resistive Micromegas in view of HL-LHC applications*. PhD thesis, Université Paris Diderot, 2013.

[3] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, `arXiv:0802.1189 [hep-ph]`.

[4] ATLAS Collaboration, M. Aaboud et al., *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. **D96** no. 7, (2017) 072002, `arXiv:1703.09665 [hep-ex]`.

[5] ATLAS Collaboration, G. Aad et al., *Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. **C76** no. 11, (2016) 581, `arXiv:1510.03823 [hep-ex]`.

[6] ATLAS Collaboration, M. Aaboud et al., *Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector*, Eur. Phys. J. **C77** no. 9, (2017) 580, `arXiv:1705.02211 [hep-ex]`. [Erratum: Eur. Phys. J.C77,no.10,712(2017)].

[7] ATLAS Collaboration, M. Aaboud et al., *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. **C77** no. 7, (2017) 466, `arXiv:1703.10485 [hep-ex]`.

[8] ATLAS Collaboration, G. Aad et al., *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC proton–proton collision data*, Eur. Phys. J. **C74** no. 11, (2014) 3130, `arXiv:1407.3935 [hep-ex]`.

[9] ATLAS Collaboration, *E_T^{miss} performance in the ATLAS detector using 2015-2016 LHC pp collisions*,.

[10] ATLAS Collaboration, G. Aad et al., *Performance of algorithms that reconstruct missing transverse momentum in $\sqrt{s}$ = 8 TeV proton-proton collisions in the ATLAS detector*, Eur. Phys. J. **C77** no. 4, (2017) 241, `arXiv:1609.09324 [hep-ex]`.

[11] K. Zuber, *Neutrino Physics*. CRC Press, Taylor Francis Group, 2012.

[12] R. Mann, *An Introduction to Particle Physics and the Standard Model*. CRC Press, Taylor Francis Group, 2010.

[13] D. Griffiths, *Introduction to Elementary Particles*. Viley-VCH Verlag GmbH Co. KGaA.

[14] "Quark." https://en.wikipedia.org/wiki/Quark. Accessed: 2019-06-13.

[15] C. Bittrich, *Study of Polarization Fractions in the Scattering of Massive Gauge Bosons $WZ \rightarrow WZ$ with the ATLAS Detector at the Large Hadron Collider*, 2015.

[16] K. Brendlinger, *Physics with Electrons in the ATLAS Detector*,. `https://cds.cern.ch/record/2228644/files/CERN-THESIS-2016-144.pdf`.

[17] "Introduction: General production and performance public results of the atlas data preparation group." https://twiki.cern.ch/twiki/bin/view/AtlasPublic/DataPrepGenPublicResults. Accessed: 2019-06-04.

[18] ATLAS Collaboration, M. Aaboud et al., *Observation of electroweak $W^{\pm}Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. **B793** (2019) 469–492, `arXiv:1812.09740 [hep-ex]`.

[19] "Normal distribution." https://en.wikipedia.org/wiki/Normal_distribution. Accessed: 2019-06-05.

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit im Rahmen der Betreuung am Institut für Kern- und Teilchenphysik ohne unzulässige Hilfe Dritter verfasst und alle Quellen als solche gekennzeichnet habe.

Johannes Wünsche
Dresden, Juni 2019