CERN SUMMER SCHOOL

Jet-To-Tau ($j \rightarrow \tau_h$) Fake Rate

Zobeyer Ghafoor

zobeyer.ghafoor@stud.uni-goettingen.de

Georg-August University Göttingen, Germany

at the II. Physikalischen Institut

Supervisors: Dr. Izaak Neutelings, Dr. Arne Reimers

Abstract:

The present project concerns the misidentification of jets as hadronic τ lepton decays. In order to ascertain the frequency with which jets resemble a τ jet, the per-jet misidentification rate is determined using dijet events recorded by the CMS detector. The dependency of the jet $\rightarrow \tau_h$ fake rate on the τ identification working points, jet flavors, transverse momenta of the jets p_T^{jet} and pseudorapidity of the jet η^{jet} was investigated during this project. It has been found that the fake rate does depend on all four variables. Additionally, the fake rate exhibits disparate behaviour in different regions of p_T^{jet} and η^{jet}, particularly within the barrel region ($|\eta| < 1.3$) and end-cap ($1.3 < |\eta| < 2.5$) regions of the CMS detector.

Contents

1. Introduction

The Standard Model (SM) of particle physics represents a fundamental pillar of our understanding of the universe, elucidating the nature of the smallest building blocks of matter. For centuries, the question of the fundamental composition of our universe has been a driving force in the field of physics, leading to the development of the SM of particle physics, which remains the most profound theory in the field to date.
With the commissioning of the Large Hadron Collider (LHC) in Geneva in 2008 and the construction of multiple detectors such as the Compact Muon Solenoid (CMS), open questions in the field of particle physics are being investigated that the SM cannot answer (yet). The analysis of a vast quantity of data helps the research efforts worldwide, with a particular emphasis on open questions in the field of (particle) physics. The role of the tau lepton, an elementary particle within the SM is crucial in many analyses. In order to accurately interpret the results of such analyses, it is essential to consider the decay of the tau lepton into other particles that subsequently evolve into jets. Jets are defined as collections of particles. However, it is important to note that not only tau decays can result in the formation of jets. Other processes can also lead to the creation of jets, which may be misidentified as jets originating from tau decays but also other elementary particles can mimic a τ-jet. The objective of this project is to investigate the rate at which these objects are misidentified as τ leptons.

2. Standard Model of Particle Physics

he Standard Model (SM) of particle physics is a theoretical framework that unifies known elementary particles and their interactions.
It classifies particles into fermions (spin-1/2) and bosons (spin-1). Fermions, the matter particles, include six quarks and six leptons. The up, charm, and top quarks have an electric charge of $+2/3$, while the down, strange, and bottom quarks carry $-1/3$. These quarks are grouped into three generations, with baryons formed from three quarks and mesons from a quark-antiquark pair. Leptons, divided into three generations, include the electron, muon, and tau (charge -1), each with a corresponding electrically neutral neutrino.
The gauge bosons are the mediators for three of the four fundamental forces: the gluon g is the mediator of the strong force, the two $W^{\pm}$ bosons and the Z^0 boson function as the mediators of the weak interaction and the photon γ acts as the mediator of the electromagnetic interaction. The Higgs boson was theoretically predicted in 1964 [11–13] and experimentally discovered in 2012 [2, 5]. The Higgs mechanism has provided insight into the origin of the masses of elementary particle [16].

2.1. Tau Lepton

The tau lepton τ^- is the most massive lepton with a mass of $m_{\tau^-} = 1.78\,\text{GeV}$ [15] whereas the masses of μ^- and e^- are in the MeV- and keV-region, respectively. It is an unstable particle with a lifetime of $\tau_{\tau^-} = 0.29\,\text{ps}$ [15]. This instability gives rise to a spectrum of decay modes for the tau lepton. They can be divided into two sections: the leptonic and hadronic decay modes. In the leptonic decay modes, only leptons appear in the final state of the decay, leading two only two possible outcomes: a decay into an electron e^-, anti-electron-neutrino $\nu^-_{e^-}$ and tau-neutrino ν_{τ^-} and a decay into a muon μ^-, anti-muon neutrino $\nu^-_{\mu^-}$ and tau-neutrino ν_{τ^-}. The total branching ratio (BR) of the leptonic decay mode has a value of BR($\tau \to l^- \bar{\nu}_{l^-} \nu_\tau$)$\approx$ 35.21 % [15]. Hadronic decay modes of the tau lepton are characterised by the number of neutral π^0

and charged pions $\pi^{\pm}$ emitted. The probability for the tau lepton to decay into one negatively charged pion π^- and a neutral pion π^0 amounts to BR($\tau^- \to \pi^-\pi^0\nu_\tau$)$\approx$ 25.4 % [15]. The tau lepton τ^- can also decay into one pion: BR($\tau^- \to \pi^-\nu_\tau$)$\approx$ 10.8 % [15]. Decays into three pions also occur: BR($\tau^- \to 2\pi^-\pi^+\nu_\tau$)$\approx$ 9.3 % and BR($\tau^- \to 2\pi^0\pi^-\nu_\tau$)$\approx$ 9.3 % [15]. More decay modes with different particles in the final state are listed in [15]. The hadrons in the final state can then form jets which are collections of long-lived particles.

3. The Large Hadron Collider (LHC) and the CMS Detector

The Large Hadron Collider (LHC) at the European Organisation for Nuclear Research (CERN) is the largest particle accelerator in the world. The circular collider has a length of 27 km and is situated at a depth of around 100 m below the Earth's surface. Protons with a centre-of-mass energy of $E_{\mathrm{cm}} = 13.6\,\mathrm{TeV}$ collide in bunches at a luminosity of $10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$, and the collision products are detected by the multiple experiments positioned along the beam. An outline of CERN and the LHC with all its experiments is shown in Figure 2.

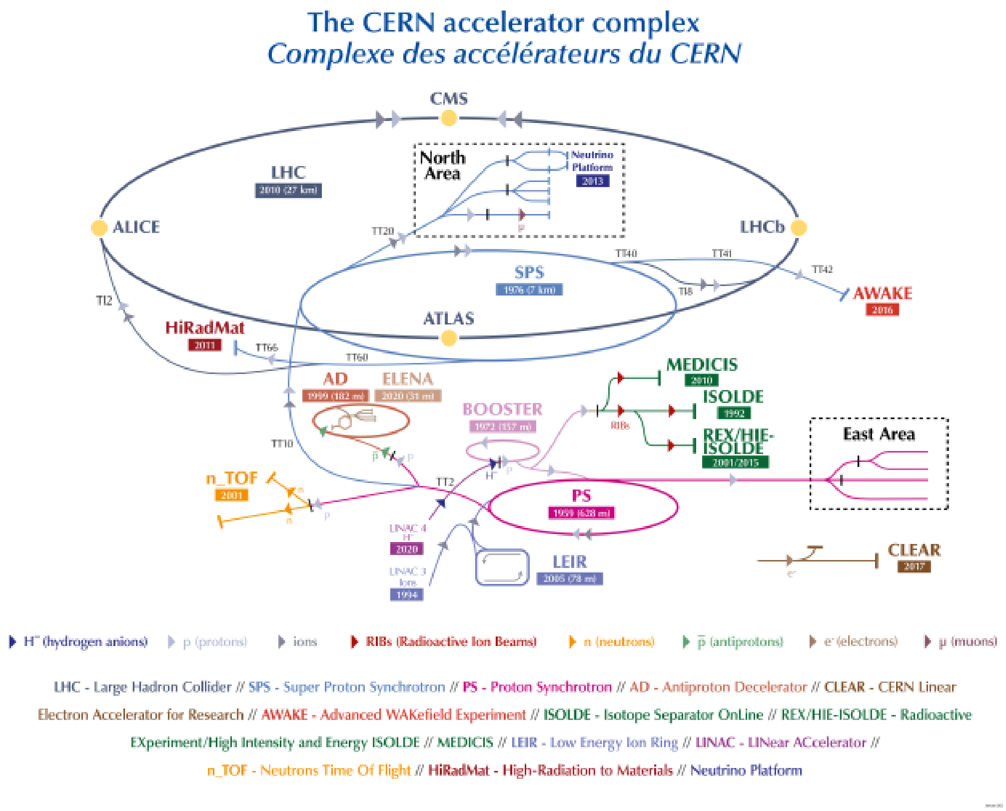


Figure 2: The CERN accelerator complex with the LHC ring and all its experiments [14].

3.1. The Compact Muon Solenoid (CMS) Detector

The Compact Muon Solenoid (CMS) detector is along with the ATLAS detector a large general purpose detector with a length of 21 m, a diameter of 15 m and it is 15 m high. The main feature of the CMS detector is the solenoid magnet that delivers a field strength of 4 T to bend the tracks

of charged particles. This bending is employed to measure the momentum of charged particles with high precision. The magnet consists of superconducting fibres that generate the magnetic field [10]. The layout of the CMS detector is shown in Figure 3.
The coordinate system of the CMS detector is defined as follows: the beam axis of the LHC defines the z-axis. The y-axis is perpendicular to the z- and x- axes and points upwards, while the x-axis points towards the centre of the LHC ring. Spherical coordinates are used at CMS besides a cartesian coordinate system. The beam angle θ is the angle between the beam axis and the x-y-plane and the azimuthal angle ϕ is around the beam axis in the x-y-plane. In many analyses, three variables play an important role: the transverse momentum $p_T = \sqrt{p_x^2 + p_y^2}$ and the rapidity $y = \frac{1}{2} \ln \left[(E + p_z) \,/\, (E - p_z) \right]$. Differences in rapidity are Lorenz-invariant under a boost in beam direction, and in the relativistic limit, the rapidity becomes the pseudorapidity $\eta = -\ln \left[\tan (\theta/2) \right]$. In addition, the distance in the pseudorapidity-azimuthal plane is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.
Inside of the magnet, the tracker and the calorimeter find their place. The inner tracker consisting of silicon pixel and silicon microstrip detectors is placed in closest proximity to the beam axis and is the first layer that the particles encounter upon occurrence of proton-proton (pp) collisions. A charged particle interacts electromagnetically with the tracker layers and leaves an electric signal [10].
The operation of calorimeters is based on the principle of energy deposition when a particle traverses the calorimeters. Ionisation and excitation of the material lead to the deposition of energy and gives information on the particles' energy. Electromagnetic and hadronic calorimeters are used to measure the energy. Lead tungstate is used as the active material for the electromagnetic and hadronic calorimeter. Both calorimeters exhibit a wide coverage, almost encompassing the beam axis. A preshower is situated in front of the end-cap region of the electromagnetic calorimeter with the objective of capturing particles that deposit their energy prior to entering the calorimeter. Additionally, the preshower serves to reduce the background from neutral pions that decay to two photons [10].
Muon detectors represent the final layer of the CMS detector as muons are minimum ionising particles which makes them distinguishable from other particles as they travel a long distance and are not stopped by the calorimeter system. Gaseous detectors (drift chambers) are used for the muon spectrometer to identify the muons and measure their momentum [10].
The trigger system employs a selective data reduction process, whereby a pre-defined subset of data is extracted for offline analysis. This approach is necessary due to the inherent limitations in storing the vast amount of data generated by the CMS detector. The Level-1 (L1) Trigger and the High-Level Trigger (HLT) are the two key components of this data reduction process, effectively reducing the data volume by a factor of millions. [10].

4. Tau Reconstruction at CMS

In order to facilitate an offline analysis, it is necessary to reconstruct the particles in question. In the case of hadronically decaying τ leptons (denoted as τ_h), this will be achieved through the application of the hadrons-plus-strip (HPS) [4, 7–9] algorithm.
The goal is to reconstruct a τ_h per seed. In the first stage of the process, the particles within a seed are reconstructed. A seed region is defined by a reconstructed hadronic jet, which will be reconstructed via the anti-k_t [3] algorithm with a distance parameter of $\Delta R = 0.4$. All particles within a cone of radius $\Delta R = 0.5$ are considered in the reconstruction of τ_h [4].
Neutral pions π^0, present in the hadronic decay modes of τ leptons, decay into two photons

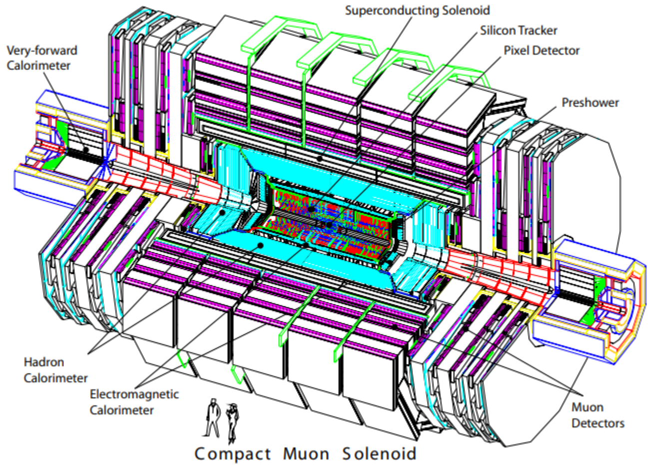
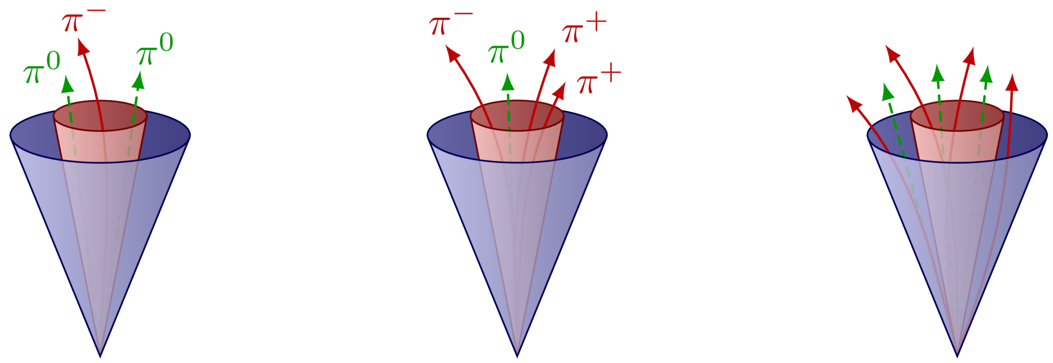


Figure 3: A sketch of the CMS detector with all its components [10].

in 98.8 % of cases [15]. The photons themselves interact with the material of the electromagnetic calorimeter and it will create a shower of secondary, lower-energetic particles (electrons, positrons, photons) due to pair production, bremsstrahlung and annihilation. The deposited energy of e^-e^+ pairs in the same $\Delta\eta\times\Delta\phi$ space are grouped together to form the strips in the HPS algorithm. Each strip has a momentum which is obtained by adding the four-momenta of the electrons, positron and photons in each strip. Hadrons in the final state of the τ_h decays are chosen by the Particle Flow algorithm [6]. The τ_h candidate is then reconstructed from the strips and reconstructed hadrons for each decay mode. The HPS algorithm uses the four momenta of the strips and hadrons to reconstruct the four momenta of the tau lepton. One τ_h is selected among all the candidates within a seed region [4].

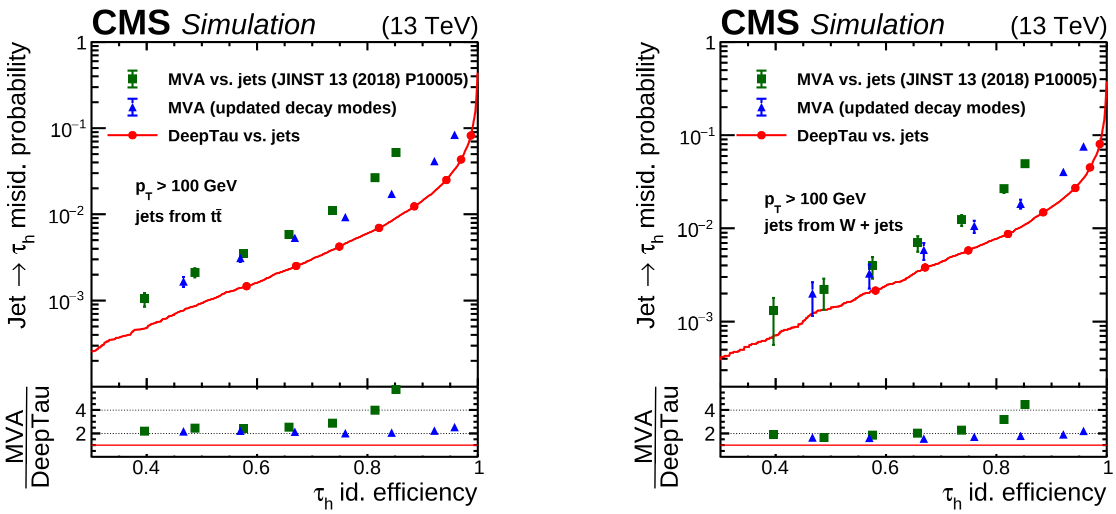
5. Misidentification of τ_h jets

As stated in Chapter 2.1, the tau lepton is observed to decay into a hadronic final state with a probability of two-thirds. It should be noted that hadronic tau decays are not the sole source of jets. Similarly, quarks and gluons, which are involved in other processes (W+jets, $t\bar{t}$, Drell-Yann (DY) and in QCD processes), create jets likewise. These jets must be distinguished from jets originating from tau decays. One feature of hadronic tau decays are that the hadrons are collimated and isolated, meaning the decay products are mainly inside the signal cone and with small contamination from other particles around them whereas jets from the background processes - especially quark and gluon jets - are wider spread and have a lot of activity around them which distinguishes them from jets emerging from τ decays. This is visualised in Figure 4. The DeepTau algorithm [1, 4] was developed to differentiate between these two categories of jets and electrons and muons misidentified as τ_h jets. It is a deep neural network (DNN) that uses the information of the tracker and calorimeter which have been segmented in rectangular $\eta\times\phi$-grids. The output of the DNN is combined with "high-level" variables, including the BR of the τ_h decay mode, its lifetime, and information of the tracker and deposited energy in the calorimeters, in another neural network. The efficacy of multivariate analysis classifiers has been evaluated in comparison with the DeepTau algorithm, as illustrated in Figure 5. The misidentification of jets as τ_h jets as a function of the τ_h candidates ($p_T > 100$ GeV) is shown there. This assessment demonstrates that the DeepTau algorithm exhibits the greatest capacity to differentiate between jets and τ_h decays.



(a) One-prong τ decay **(b)** Three-prong τ decay **(c)** Jets from background processes

Figure 4: (a) (b) Jets emerging from τ_h decays and (c) from background processes



(a) W+jets **(b)** $t\bar{t}$

Figure 5: The misidentification rate of jets from W+jets and $t\bar{t}$ (simulation) as a function of the efficiency of τ_h with $p_T > 100\,\text{GeV}$ [4]. Two multivariate analysis classifier and the DeepTau algorithm are used to determine the misidentification rate jets as τ_h jets. It shows that the DeepTau algorithm performs better at all τ_h efficiencies.

5.1. Jet $\to \tau_h$ Fake Rate

Nevertheless, as the τ_h efficiency of the process increases, so too does the misidentification rate. Accordingly, a jet $\to \tau_h$ fake rate is employed and calculated to serve as a scale factor for simulation. Up until this point, the fake rate has been calculated in control regions (CRs), which have been defined through data and then applied to the signal region (SR). This method has several disadvantages. One significant issue is the high-$p_T^{\tau_h}$ region, which has limited statistics, resulting in considerable uncertainty. Furthermore, the definition of CRs is complex, making the calculation of the jet $\to \tau_h$ fake rate intricate and it is also dependent on the specific analysis. It must be calculated individually for each analysis. Therefore, the objective is to establish a fake rate that mitigates these disadvantages. For this project, a Dijet selection was applied: the leading jets must have $p_T > 200\,\text{GeV}$ and the difference between the azimuthal angles of the two leading jets $\Delta\phi$ must be larger than 2.4 (see event selection in Section A). Additionally, there must be minimal additional activity. Furthermore, the triggers are based on the leading jet p_T or H_T ($= \sum_i p_T^i$, i runs over anti-k_t jets) and the triggers therefore require $H_T > 1200\,\text{GeV}$. This results in selected events being almost exclusively QCD dijet (see Figure 6). The fake rate is measured in the QCD CR it is then calculated as follows:

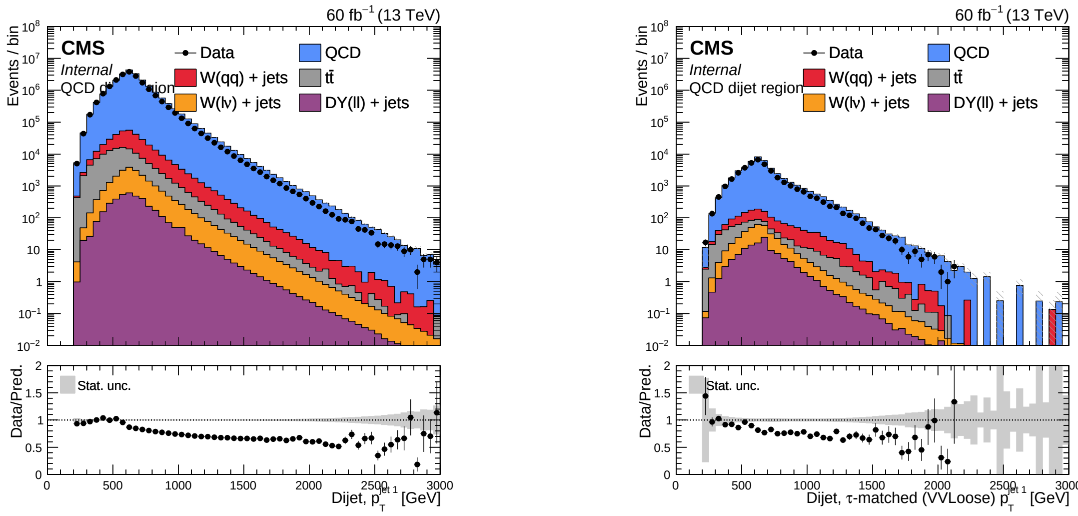
$$\text{jet} \to \tau_h \text{ fake rate} = \frac{\text{taus matched to one of the two leading jets}}{\text{total number of jets (two per event)}}. \tag{5.1}$$

The question that must be addressed is which variables influence the fake rate. In the following section, the dependency on the τ_h identification (ID) working point (WP), jet flavor, p_T^{jet} and η^{jet} will be investigated.

6. Analysis

6.1. p_T Distribution

To ensure that enough statistics are in the high-p_T region, the corresponding distributions before and after the matching are plotted and it is shown for the τ_h ID WP `vvloose` in Figure 6. It is clearly visible that high-momenta jets are evident and thus, a jet $\to \tau_h$ fake rate in the high-p_T region can be calculated.



(a) before matching to τ_h ID WP

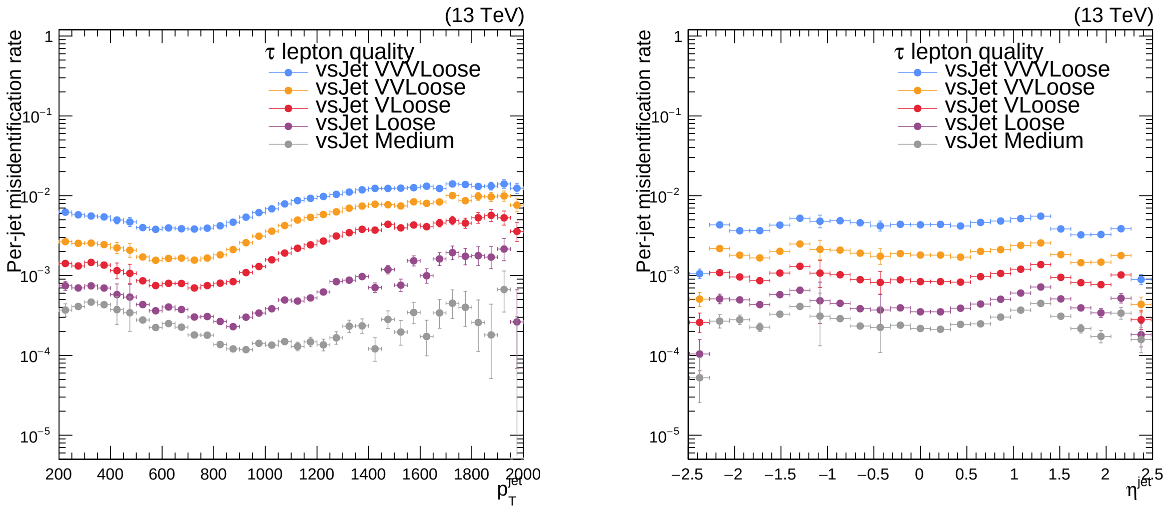
(b) after matching to τ_h ID WP `vvloose`

Figure 6: The p_T distributions before and after matching with the τ_h ID WP `vvloose` of the first jet.

6.2. Fake Rate vs. τ_h ID WP

The initial step involved an investigation into the dependency of the fake rate on the ID WP for the τ_h. Figure 7 shows the fake rate as a function of the transverse momenta and pseudorapidity of the jet for different working points. The WPs employed in this investigation and the associated matching of the τ_h are as follows: `very-very-very loose (vvvloose), very-very loose (vvloose), very loose (vloose), loose, medium` and `tight`. The `tight` WP is not shown here as it reduces the number of events drastically and statistical fluctuations dominate.

It is evident that the value of the jet $\to \tau_h$ fake rate declines with a more stringent WP. The tighter the WP is the lower the misidentification rate is, as expected due to Figure 5. The discrepancy between the fake rates between the WPs is limited to a maximum of two dimensions. Furthermore, the fake rate is observed to vary across the p_T and η range. With regard to the misidentification rate as a function of p_T, the fake rate is seen to decline more rapidly for tighter WPs within the low-p_T region. It takes on a minimum between $p_T = 600$ GeV and $p_T = 1200$ GeV. Furthermore, it can be stated that the misidentification rate displays a larger gradient in cases of tighter WPs. Additionally, it is observed that the statistical uncertainty is greater in instances where the transverse momentum of the jet is higher. It can also be stated that the fake rate as a function of η^{jet} takes on a different shape at around $|\eta^{\text{jet}}| = 1.3$. At this value of $|\eta|$, the end-cap region of the CMS detector begins. The rate of jets that fake a hadronic decay of a tau lepton also decreases in the end-cap region compared to the fake rate in the barrel region.



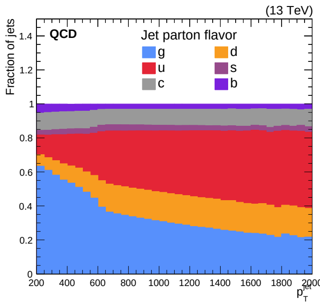
(a) p_T^{jet} **(b)** η^{jet}

Figure 7: The fake rate as a function of the jet p_T and η for different τ_h ID WPs.

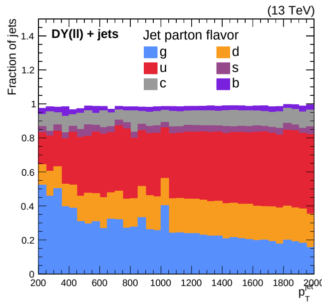
6.3. Jet Flavor Composition

In order to investigate the dependence of the fake rate on the jet flavor, it is necessary to extract the jet flavors from the simulation. The normalised jet flavor composition is presented in Figure 8 for p_T^{jet} and in Figure 9 for η^{jet}. In the case of the $t\bar{t}$ sample, the contribution of b-jets to the jet flavor composition is greater than that of other flavors, due to the exclusive decay of the top quark to a b-quark and a W boson. The contribution of s- and b-quark jets is negligibly small in all plots, with the exception of the $t\bar{t}$ plots, which pertain to the fraction of b-jets. Additionally, in certain instances, the jet fractions do not total to one. This is due to the fact that for some jets, it was not feasible to assign them a flavor. Figure 8 illustrates that gluon jets are dominant in the low-p_T region, whereas jets originating from quarks are more common in the high-p_T region. Figure 9 illustrates the dependency of jet flavor on η^{jet}, which allows us to infer the favoured emission direction of jets depending on their flavor. Gluon initiated jets tend to be emitted in the direction perpendicular to the beam axis as the corresponding distribution peaks at $\left|\eta^{\text{jet}}\right| = 0$.

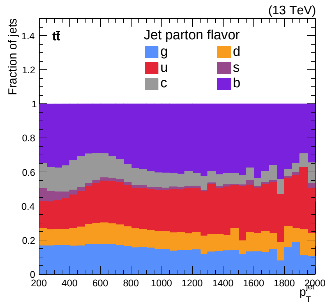
However, the normalisation of the jet flavor fractions obscures the actual behaviour of the fractions. To make predictions on the direction of emission of jets per flavor, the absolute values of the fraction of the jets must be looked at. Figure 10 and 11 shows them. It is evident that all jet flavors exhibit a peak in the low-p_T^{jet} region, rather than solely gluon-initiated jets, and are predominantly emitted around $|\eta^{\text{jet}}| = 0$.



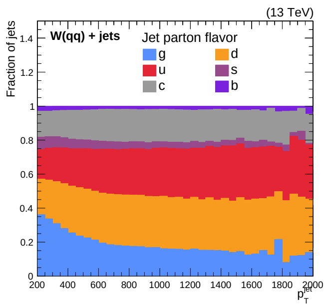
(a) QCD



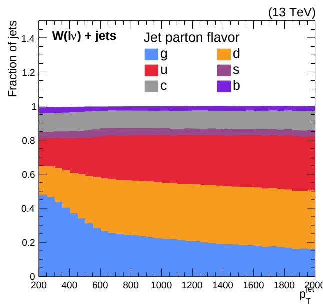
(b) DY($l\bar{l}$)



(c) $t\bar{t}$

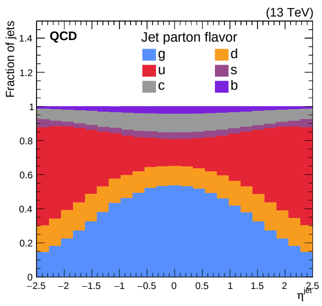


(d) $W(qq)$+jets

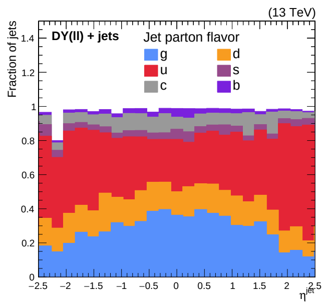


(e) $W(l\nu)$+jets

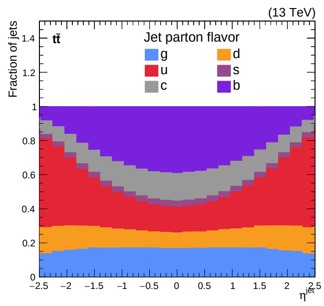
Figure 8: The normalised jet flavor composition as a function of p_T^{jet} in the QCD/Dijet CR for the different samples.



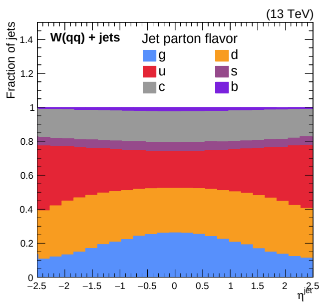
(a) QCD



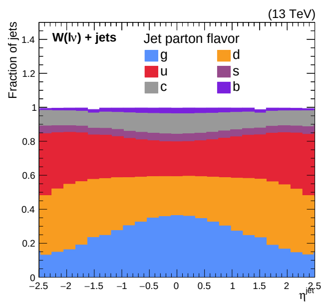
(b) DY($l\bar{l}$)



(c) $t\bar{t}$

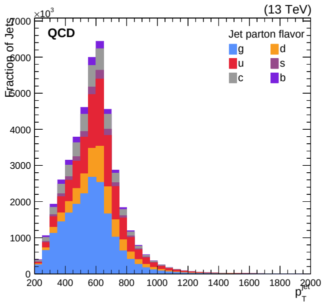


(d) $W(qq)$+jets

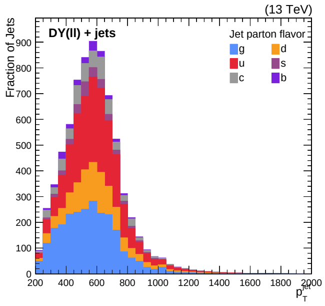


(e) $W(l\nu)$+jets

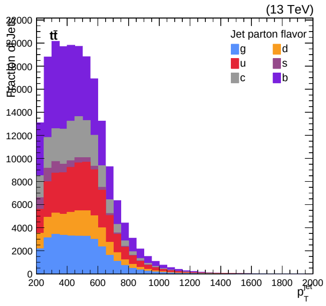
Figure 9: The normalised jet flavor composition as a function of η^{jet} in the QCD/Dijet CR for the different samples.



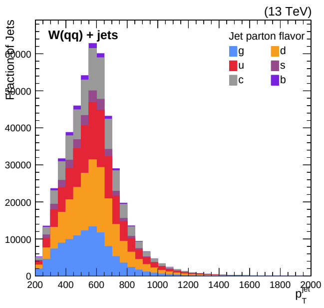
(a) QCD



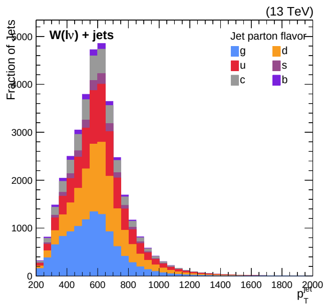
(b) DY($l\bar{l}$)



(c) $t\bar{t}$

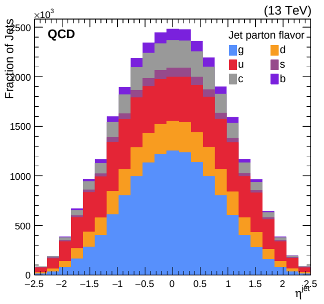


(d) $W(qq)$+jets

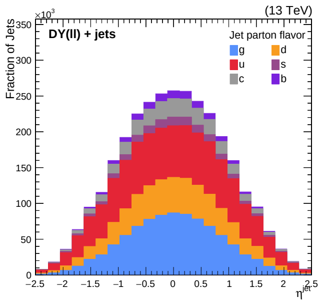


(e) $W(l\nu)$+jets

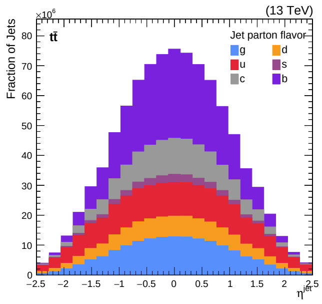
Figure 10: The jet flavor composition as a function of p_T^{jet} in the QCD/Dijet CR for the different samples.



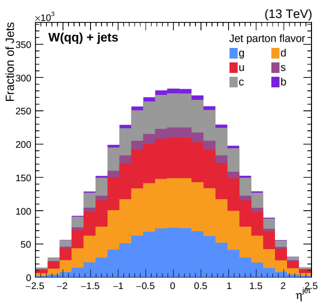
(a) QCD



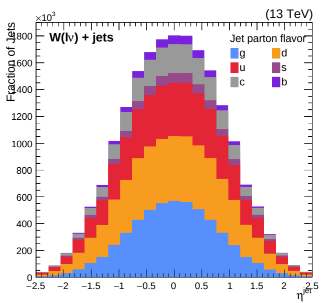
(b) DY($l\bar{l}$)



(c) $t\bar{t}$



(d) $W(qq)$+jets

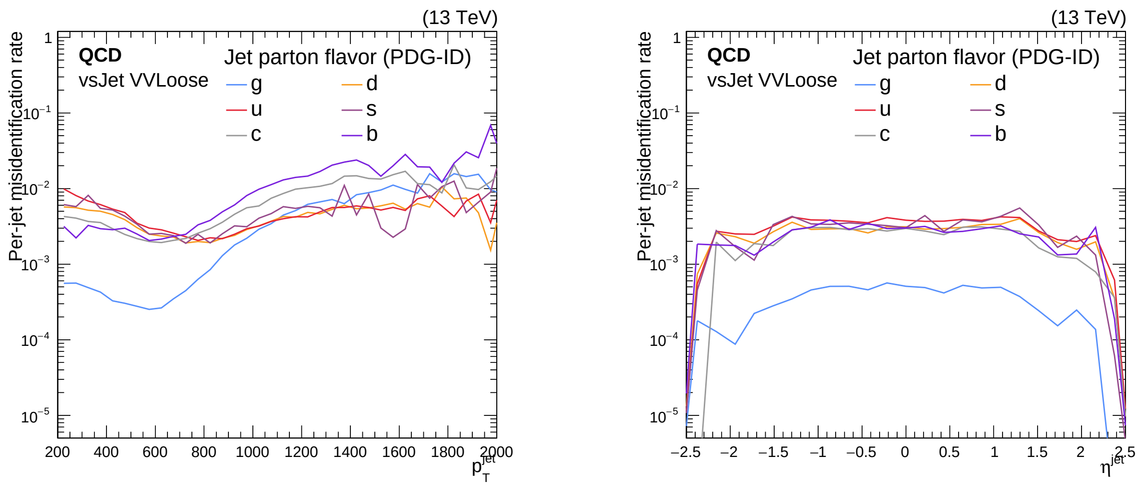


(e) $W(l\nu)$+jets

Figure 11: The jet flavor composition as a function of η^{jet} in the QCD/Dijet CR for the different samples.

6.4. Fake Rate vs. Jet Flavor

Figure 12 depicts the jet $\to \tau_h$ fake rate as a function of the jet-p_T and jet-η for each jet flavor. Only the plots for the `vvloose` τ_h ID WP are shown, as the higher the WP, the higher the statistical fluctuations are (the other distributions are listed in Section B). Additionally, the fake rate has different values for for each jet flavor but is relatively similar in the case of quark-jets. Notably, the fake rate for gluon-initiated jets is almost two orders of magnitude lower for $p_T^{\text{jet}} < 1000\,\text{GeV}$ and approximately 10 times smaller over the full η-range than the fake rate for jets originating from quarks.
The misidentification rate varies across the full range of p_T and η, demonstrating a correlation between these two variables. The behaviour of the fake rate is similar for all jet flavors in both plots in Figure 12. Of particular note is the decrease of the fake rate distribution for very low values of p_T^{jet}, which is followed by an increase from $p_T^{\text{jet}} = 600\,\text{GeV}$ onwards until $p_T^{\text{jet}} \approx 1300\,\text{GeV}$. Beyond this point, statistical fluctuations become the dominant factor in particular for this WP. At $|\eta^{\text{jet}}| \approx 1.3$, the fake rate decreases for all jet flavor, while in the $|\eta^{\text{jet}}|$ region of -1.3 to 1.3, it is approximately constant.



(a) p_T^{jet} **(b)** η^{jet}

Figure 12: Jet $\to \tau_h$ fake rate as a function of p_T^{jet} and η^{jet}.

The disparate behaviours observed in different p_T^{jet} and η^{jet} regions point to the necessity of investigating the fake rate in these regions individually. For this, different p_T- and η-bins where defined fake rate as a function of one variable in the bins of the other variable. The pseudorapidity of a jet η^{jet} was segmented in the three following bins:

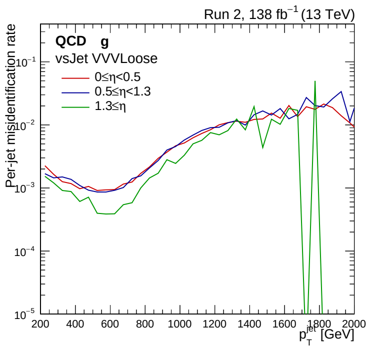
1. $0 \leq |\eta^{\text{jet}}| < 0.5$,
2. $0.5 \leq |\eta^{\text{jet}}| < 1.3$ and
3. $1.3 \leq |\eta^{\text{jet}}|$.

The transverse momenta p_T^{jet} is separated in four bins:

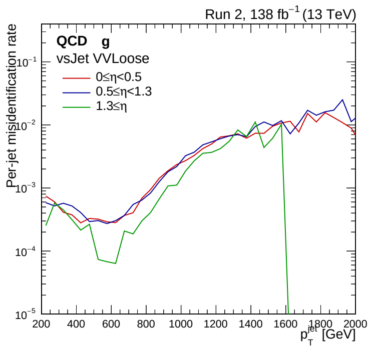
1. $0 \leq p_T < 300\,\text{GeV}$,
2. $300 \leq p_T < 600\,\text{GeV}$,

3. $600 \leq p_T < 1000\,\text{GeV}$ and
4. $1000\,\text{GeV} \leq p_T$.

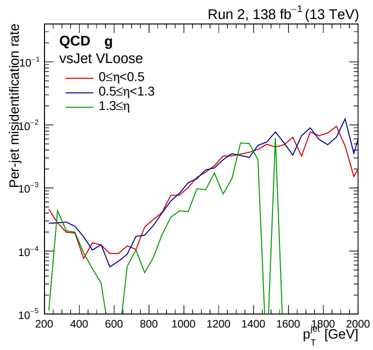
The corresponding plots can be seen in Figure 13 for the three loosest τ_h ID WP but only for QCD events and gluon-jets. The fake rate plots for the other τ_h ID WPs, for other samples and for different jet-flavors are shown in Section C. The general shape of the fake rate across the p_T and η ranges remains consistent in the different bins and is similar to the total fake rate distributions in Figure 12. This is the primary conclusion to be drawn from these distributions. Furthermore, the fake rate decreases with tighter τ_h ID WP and statistical fluctuations start to get dominant at the `loose` working point, especially in the first η- and p_T-bin. What also stands out is that in the p_T^{jet}-fake factor plots, the fake rate value seems to only change slightly for high transverse momenta and is at around 0.01. Additionally, the fake rate in the bin $1.3 \leq |\eta^{\text{jet}}|$ is generally lower than the fake rate in the other two p_T-bins. Especially around a transverse momentum of $p_T^{\text{jet}} = 600\,\text{GeV}$, the fake rate takes on a minimum. The misidentification rate in the η^{jet} plots in Figure 13 decreases similarly with tighter WP. The fake rate for $p_T^{\text{jet}} \geq 1000\,\text{GeV}$ in the η^{jet} plots is significantly larger than the other three fake rate distributions that have similar values over the whole η range. Thus, this indicates an area that needs to be studied separately in the fake rate calculation.



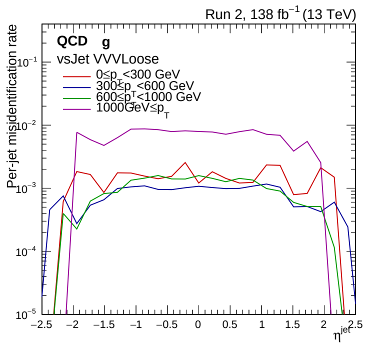
(a) `vvvloose`



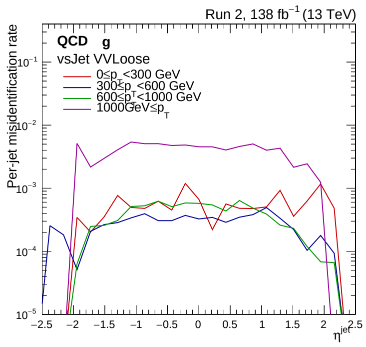
(b) `vvloose`



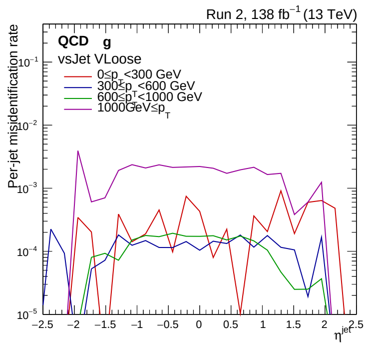
(c) `vloose`



(d) `vvvloose`



(e) `vvloose`



(f) `vloose`

Figure 13: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} and η^{jet} for gluon-initiated jets and for the three loosest τ_h ID WPs in three $|\eta|$ bins, in four p_T bins respectively.

7. Conclusion and Outlook

The project addressed the issue of misidentification of jets from background processes as jets originating from tau decays. In many analyses, this problem must be addressed in order to reconstruct the actual τ_h jets as τ_h jets and not misidentify other jets, electrons or muons as τ_h jets. Previously, various techniques have been devised to eliminate such jets and particles that have been erroneously identified as τ_h jets. One notable approach is the DeepTau algorithm, which is a deep neural network (DNN). However, these methods are not without limitations. This project aimed to develop a methodology for estimating the fake rate, which can then be used as a scaling factor on simulation.
The objective of this project was to investigate the dependency of the jet $\to \tau_h$ fake rate on the τ_h ID WPs, jet flavor, transverse momentum of the jet p_T^{jet} and pseudorapidity of the jet η^{jet} in the QCD/Dijet CR. It has been seen that the fake rate changes with all four variables. In order to calculate the fake rate, it is necessary to include these variables, as they exert a significant influence on the value of the fake rate. The significance of the jet flavor has been consistently evident throughout the entirety of the analysis. Gluon jets appear to be less indicative of a τ_h jet, whereas quark-initiated jets demonstrate a higher propensity to be misidentified as a τ_h jet. Also, the fake rate values observed in different region of p_T^{jet} and η^{jet} exhibit notable discrepancies. At a p_T^{jet} of approximately 600 GeV, the fake rate attains a minimum for all jet flavors and all WPs, and is evident that the misidentification rate for jets with pseudorapidity greater than 1.3 differs from that of jets with $\left|\eta^{\text{jet}}\right| < 1.3$. For $\left|\eta^{\text{jet}}\right| \geq 1.3$, the barrel region of the CMS detector ceases to be operational and the end-cap region becomes active. This suggests that the precision of the data-taking in this area may be subject to change and may differ from that observed in the barrel region.
The subsequent phase of the investigation must be the examination of additional dependencies on other variables. Also, it must be shown that global event variables such as the total number of jets in the whole event or ΔR(closet jet but not τ_h) do not impact the fake rate. Furthermore, the fake rate must be calculated in other CRs to focus on other background processes. To improve the fake rate calculation, the tightening of the CRs is crucial to guarantee that only the intended background process is enriched in its CR. It is similarly crucial to ensure the extraction of jet flavor data, and subsequently applied to the simulation. In order to calculate the fake rate in data, it is necessary to extract the jet flavor. One potential approach is to define at least a second CR with a different jet flavor composition, and to use the matrix method to extract the jet flavors in data.

This field encompasses a vast research area. This is a significant challenge that researchers must address in their analysis. With the knowledge of a fake rate, which can be used to classify the problem of misidentified jets as tau jets, it will facilitate the analysis of processes such as Higgs boson decaying to two τ's. This will allow for more accurate results to be achieved, which will in turn lead to new insights.

References

[1] ANDREJKOVIC, J. W. *Data-Driven Background Modeling for Precision Studies of the Higgs Boson and Searches for New Physics with the CMS Experiment.* Presented 06-07-2022 (TU Vienna, 2022). `https://cds.cern.ch/record/2823368`.

[2] ATLAS COLLABORATION. *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.* Phys. Lett. B **716,** 1–29 (2012).

[3] CACCIARI, M.; SALAM, G. P.; SOYEZ, G. *The anti-k_t jet clustering algorithm.* JHEP **04,** 063 (2008).

[4] CMS COLLABORATION. *Identification of hadronic tau lepton decays using a deep neural network.* JINST **17,** P07023 (2022).

[5] CMS COLLABORATION. *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC.* Phys. Lett. B **716,** 30–61 (2012).

[6] CMS COLLABORATION. *Particle-flow reconstruction and global event description with the CMS detector.* JINST **12,** P10003 (2017).

[7] CMS COLLABORATION. *Performance of τ-lepton reconstruction and identification in CMS.* JINST **7,** P01001–P01001 (2012).

[8] CMS COLLABORATION. *Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13\,\text{TeV}$.* JINST **13,** P10005 (2018).

[9] CMS COLLABORATION. *Reconstruction and identification of τ lepton decays to hadrons and ν_τ at CMS.* JINST **11,** P01019 (2016).

[10] CMS COLLABORATION. *The CMS Experiment at the CERN LHC.* JINST **3,** S08004 (2008).

[11] ENGLERT, F.; BROUT, R. *Broken Symmetry and the Mass of Gauge Vector Mesons.* Phys. Rev. Lett. **13** (ed TAYLOR, J. C.) 321–323 (1964).

[12] HIGGS, P. W. *Broken Symmetries and the Masses of Gauge Bosons.* Phys. Rev. Lett. **13** (ed TAYLOR, J. C.) 508–509 (1964).

[13] HIGGS, P. W. *Broken symmetries, massless particles and gauge fields.* Phys. Lett. **12,** 132–133 (1964).

[14] LOPIENSKA, E. *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022.* General Photo. `https://cds.cern.ch/record/2800984` (2022).

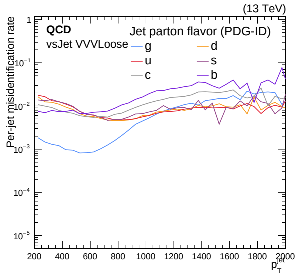
[15] NAVAS, S. et al. *Review of Particle Physics.* Phys. Rev. D **110,** 030001 (3 2024).

[16] THOMSON, M. *Modern Particle Physics* (Cambridge University Press, 2013).

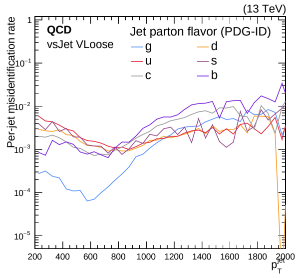
A. Event Selection

- 0 loose e^- and μ^- (cut based ID)
- consider AK4 CHS jets cleaned from loose e^- and μ^-
  - $p_T > 30\,\text{GeV}$
  - $|\eta| < 2.5$
- Trigger: HLT_PFHT1050 || HLT_PFJet550 || HLT_PFJet500
- Dijet Selection:
  - leading jets must have $p_T > 200\,\text{GeV}$, $\Delta\phi > 2.4$
- MET$< 100\,\text{GeV}$
- offline HT (computed with AK4 CHS jets)$> 1200\,\text{GeV}$

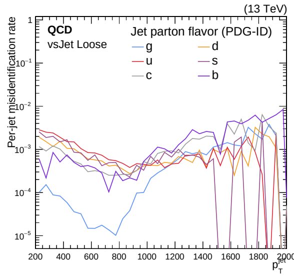
B. Fake Rate vs. Jet Flavor



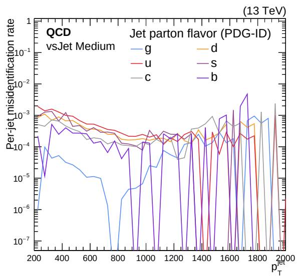
(a) `vvvloose`



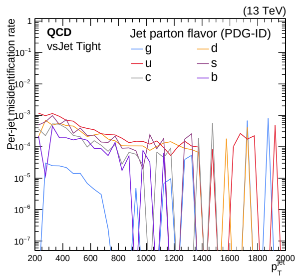
(b) `vloose`



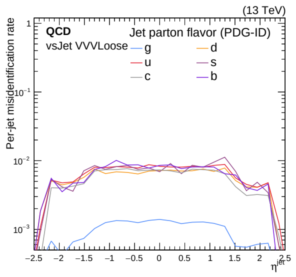
(c) `loose`



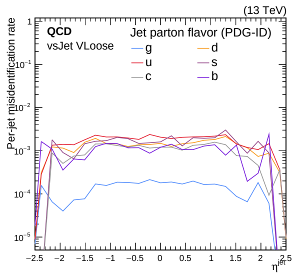
(d) `medium`



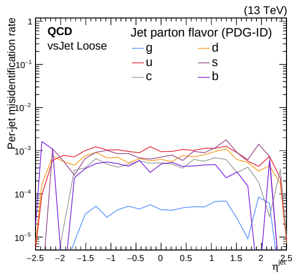
(e) `tight`



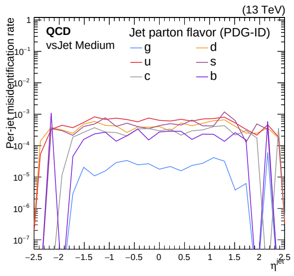
(f) `vvvloose`



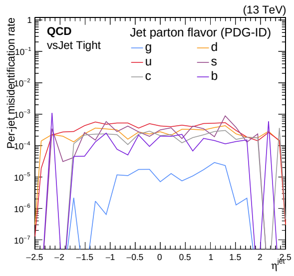
(g) `vloose`



(h) `loose`

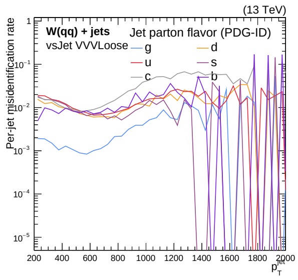


(i) `medium`

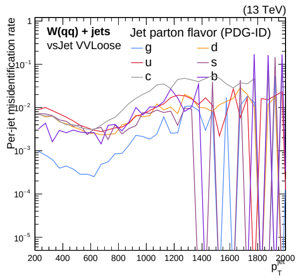


(j) `tight`

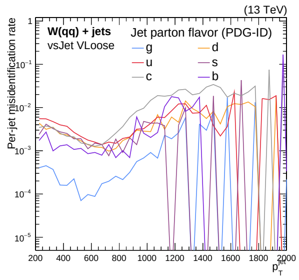
Figure 14: Jet $\to \tau_h$ fake rate as a function of p_T^{jet} and η^{jet} for the QCD sample.



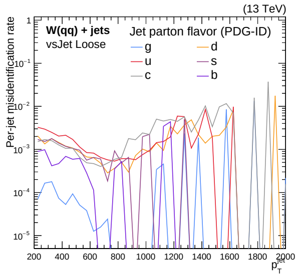
(a) `vvvloose`



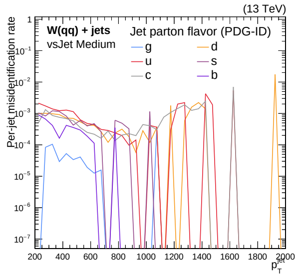
(b) `vvloose`



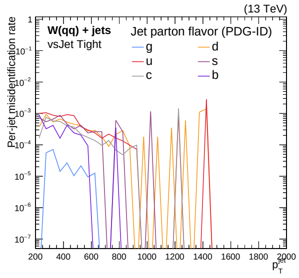
(c) `vloose`



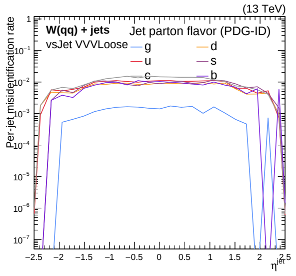
(d) `loose`



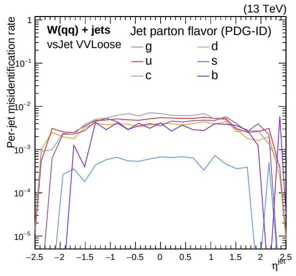
(e) `medium`



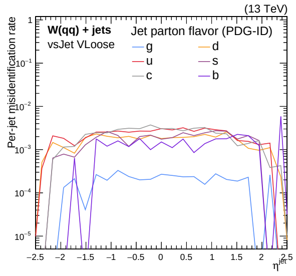
(f) `tight`



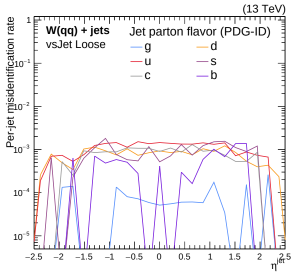
(g) `vvvloose`



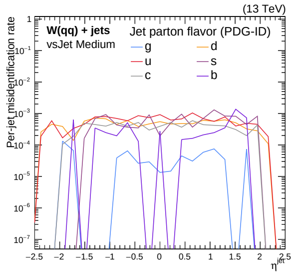
(h) `vvloose`



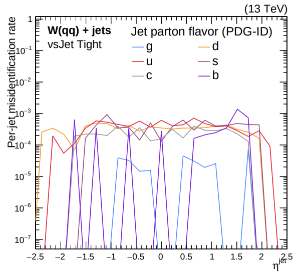
(i) `vloose`



(j) `loose`



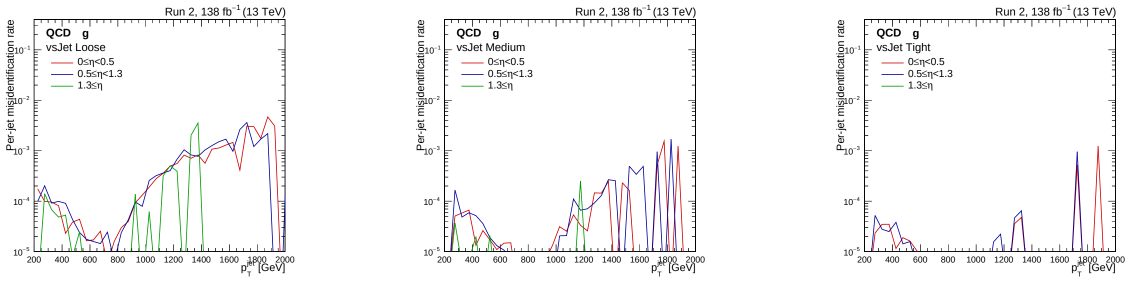
(k) `medium`



(l) `tight`

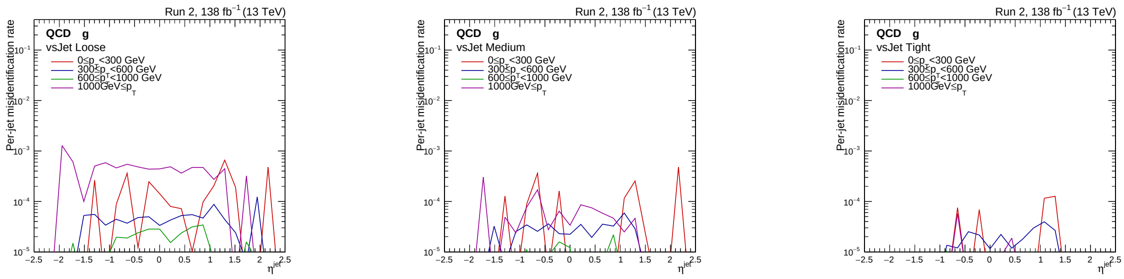
Figure 15: Jet $\to \tau_h$ fake rate as a function of p_T^{jet} and η^{jet} for the $W(qq)$+jets sample.

C. Fake Rate in Bins of p_T and η



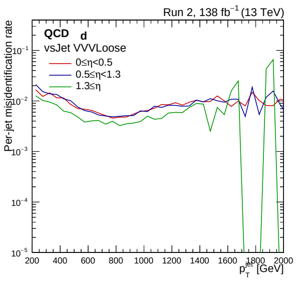
(a) `loose` **(b)** `medium` **(c)** `tight`

Figure 16: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for gluon-initiated jets in three $|\eta|$ bins (QCD sample).

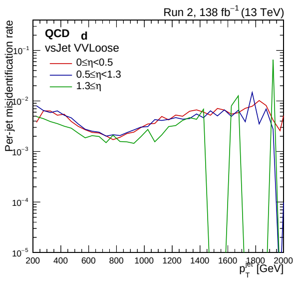


(a) `loose` **(b)** `medium` **(c)** `tight`

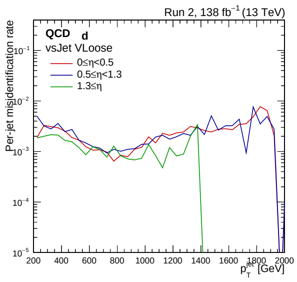
Figure 17: The jet $\to \tau_h$ fake rate as a function of η^{jet} for gluon-initiated jets in four $p_T|$ bins (QCD sample).



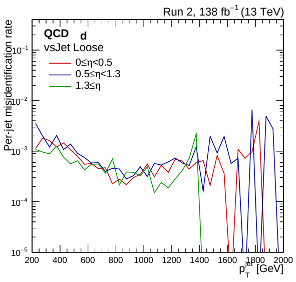
(a) `vvvloose`



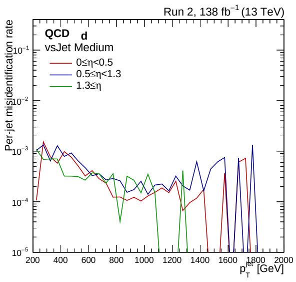
(b) `vvloose`



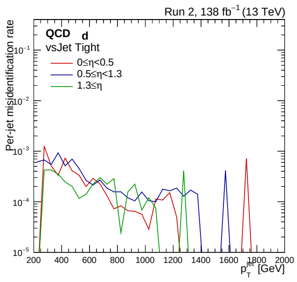
(c) `vloose`



(d) `loose`

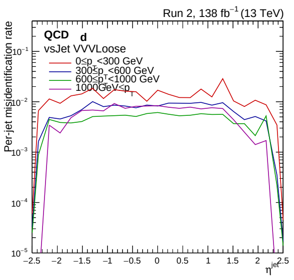


(e) `medium`

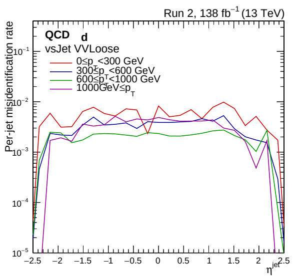


(f) `tight`

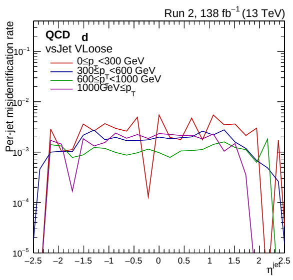
Figure 18: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for down-quark initiated jets in three $|\eta|$ bins (QCD sample).



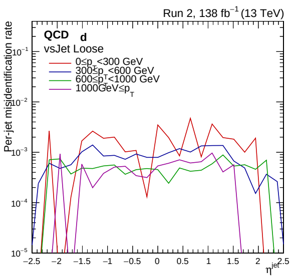
(a) `vvvloose`



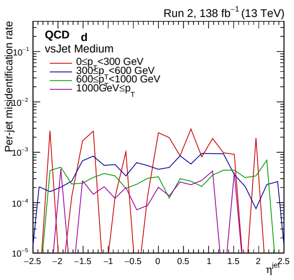
(b) `vvloose`



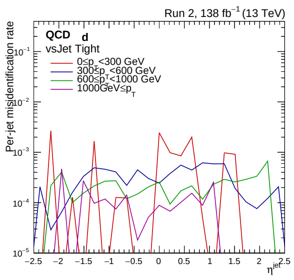
(c) `vloose`



(d) `loose`

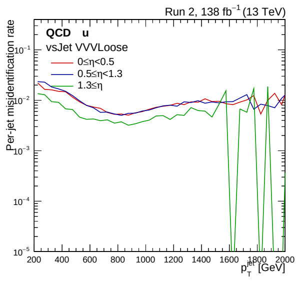


(e) `medium`

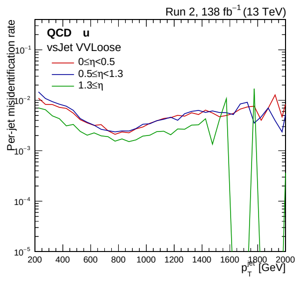


(f) `tight`

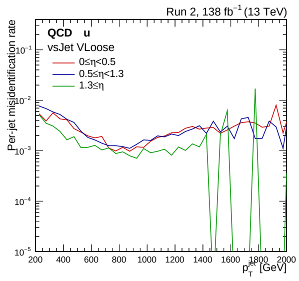
Figure 19: The jet $\to \tau_h$ fake rate as a function of η^{jet} for down-quark initiated jets in four p_T bins (QCD sample).



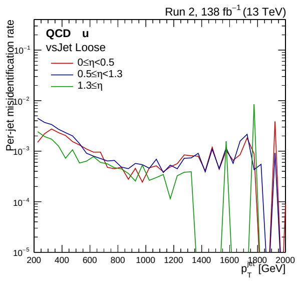
(a) `vvvloose`



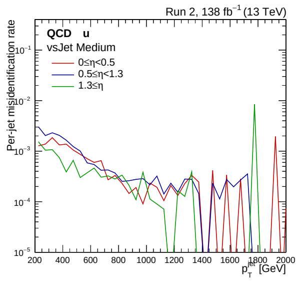
(b) `vvloose`



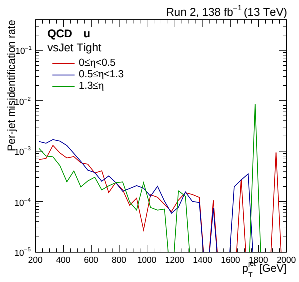
(c) `vloose`



(d) `loose`

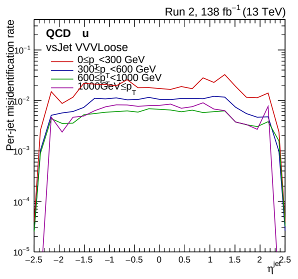


(e) `medium`

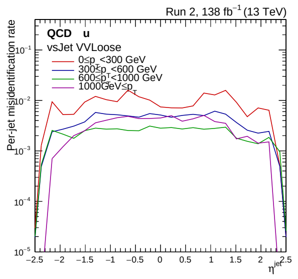


(f) `tight`

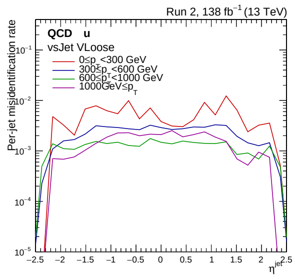
Figure 20: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for up-quark initiated jets in three $|\eta|$ bins (QCD sample).



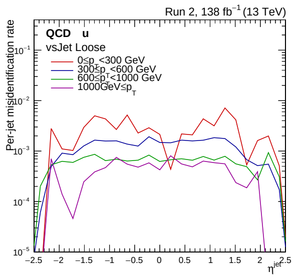
(a) `vvvloose`



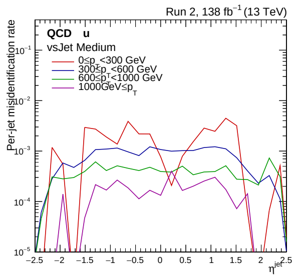
(b) `vvloose`



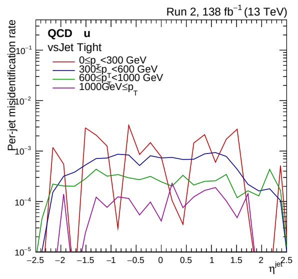
(c) `vloose`



(d) `loose`

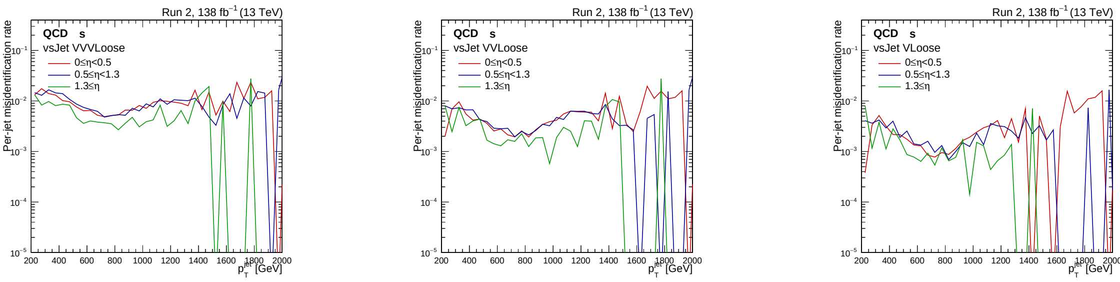


(e) `medium`

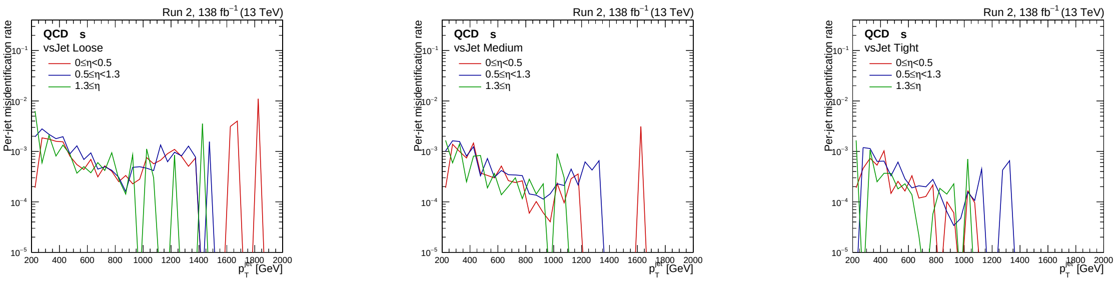


(f) `tight`

Figure 21: The jet $\to \tau_h$ fake rate as a function of η^{jet} for up-quark initiated jets in four p_T bins (QCD sample).

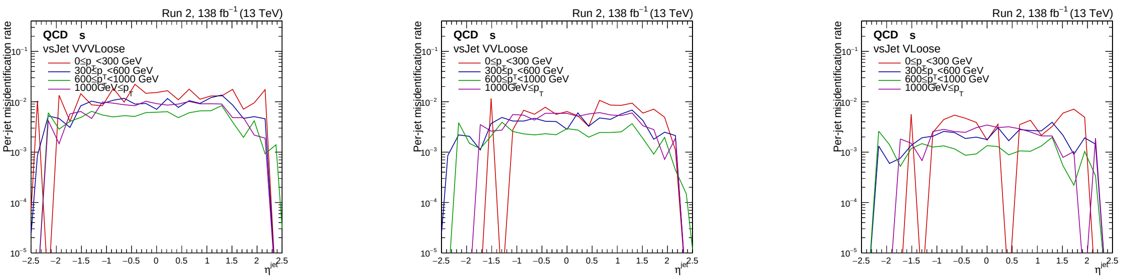


(a) `vvvloose` **(b)** `vvloose` **(c)** `vloose`

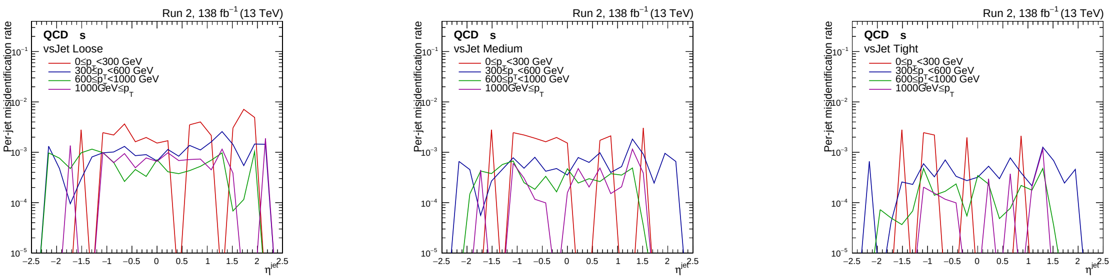


(d) `loose` **(e)** `medium` **(f)** `tight`

Figure 22: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for strange-quark initiated jets in three $|\eta|$ bins (QCD sample).

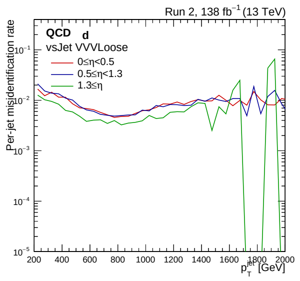


(a) `vvvloose` **(b)** `vvloose` **(c)** `vloose`

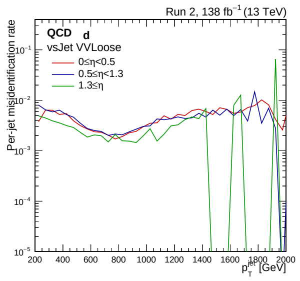


(d) `loose` **(e)** `medium` **(f)** `tight`

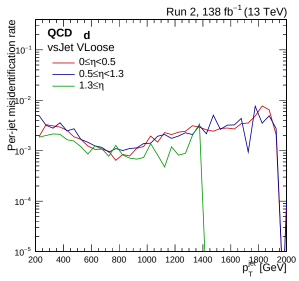
Figure 23: The jet $\to \tau_h$ fake rate as a function of η^{jet} for strange-quark initiated jets in four p_T bins (QCD sample).



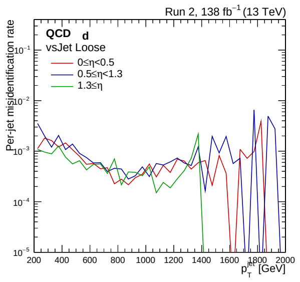
(a) `vvvloose`



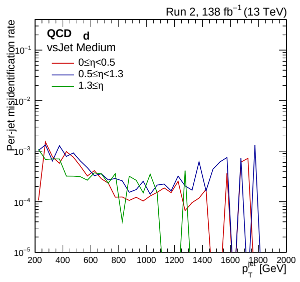
(b) `vvloose`



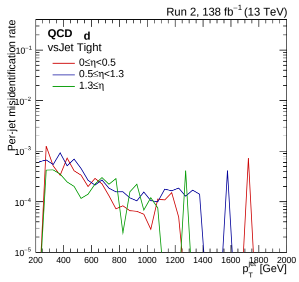
(c) `vloose`



(d) `loose`

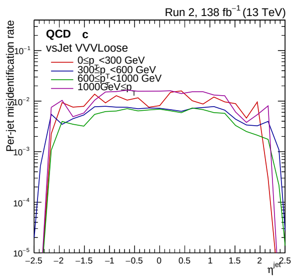


(e) `medium`

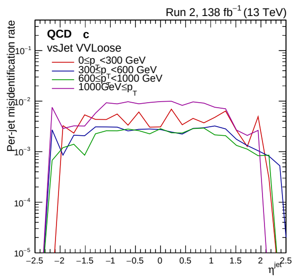


(f) `tight`

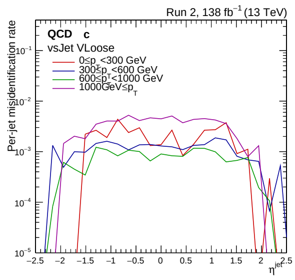
Figure 24: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for charm-quark initiated jets in three $|\eta|$ bins (QCD sample).



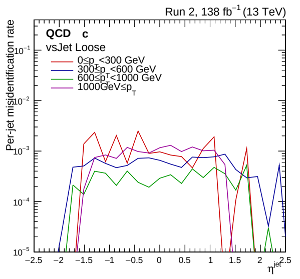
(a) `vvvloose`



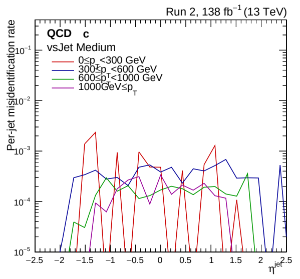
(b) `vvloose`



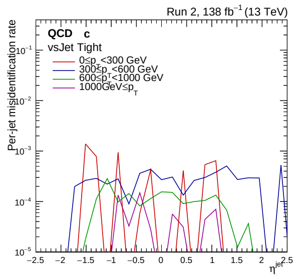
(c) `vloose`



(d) `loose`

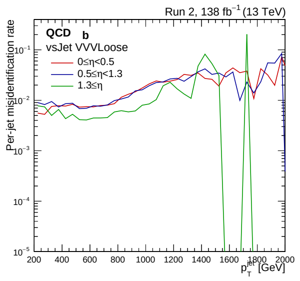


(e) `medium`

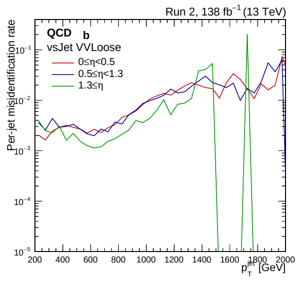


(f) `tight`

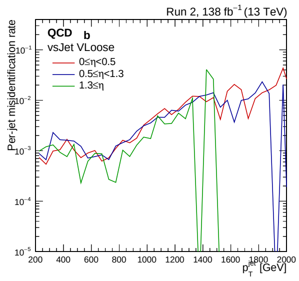
Figure 25: The jet $\to \tau_h$ fake rate as a function of η^{jet} for charm-quark initiated jets in four p_T bins (QCD sample).



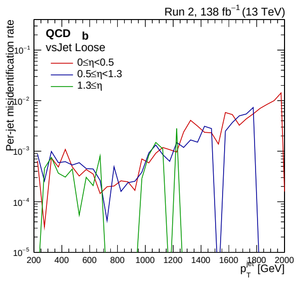
(a) `vvvloose`



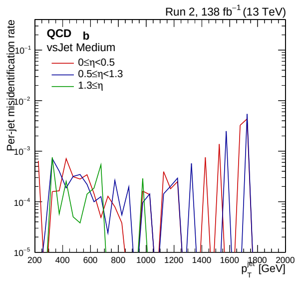
(b) `vvloose`



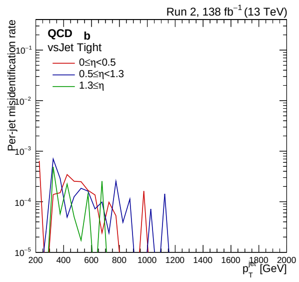
(c) `vloose`



(d) `loose`

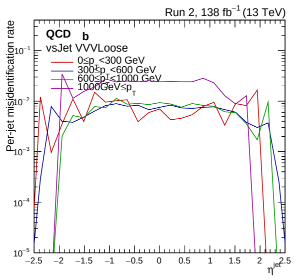


(e) `medium`

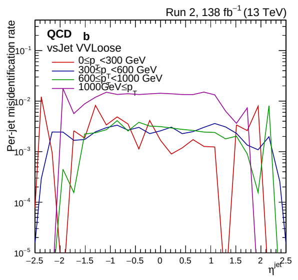


(f) `tight`

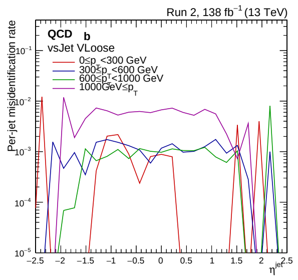
Figure 26: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for bottom-quark initiated jets in three $|\eta|$ bins (QCD sample).



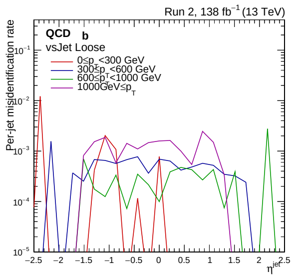
(a) `vvvloose`



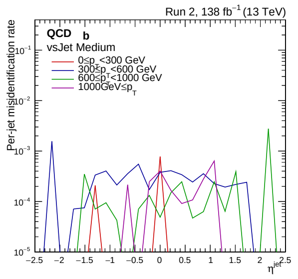
(b) `vvloose`



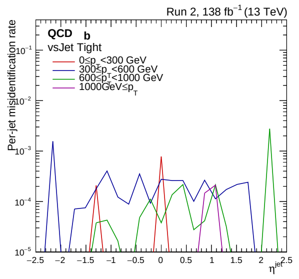
(c) `vloose`



(d) `loose`

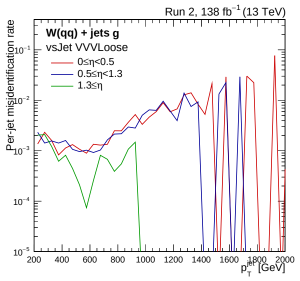


(e) `medium`

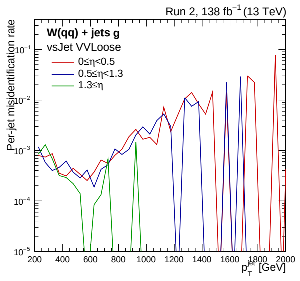


(f) `tight`

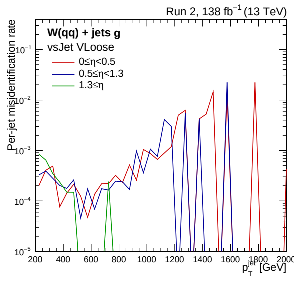
Figure 27: The jet $\to \tau_h$ fake rate as a function of η^{jet} for bottom-quark initiated jets in four p_T bins (QCD sample).



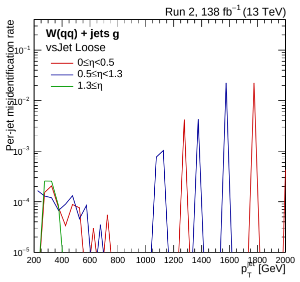
(a) `vvvloose`



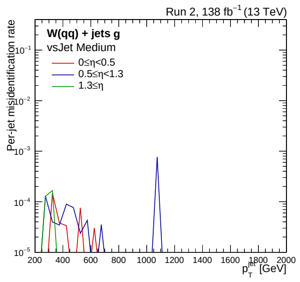
(b) `vvloose`



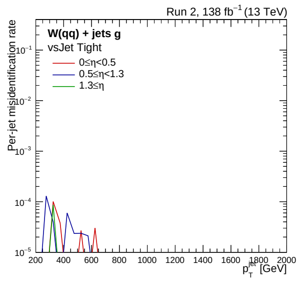
(c) `vloose`



(d) `loose`

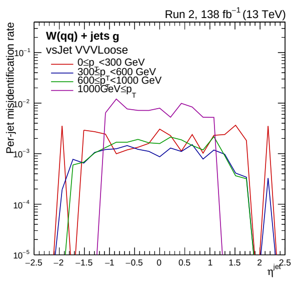


(e) `medium`

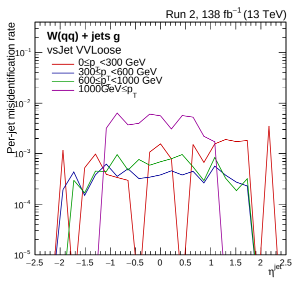


(f) `tight`

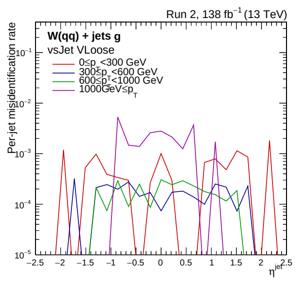
Figure 28: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for down-quark initiated jets in three $|\eta|$ bins ($W(qq)$+jets sample).



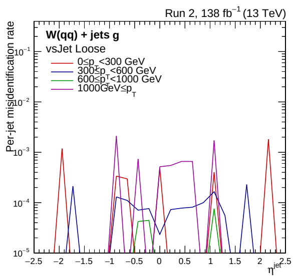
(a) `vvvloose`



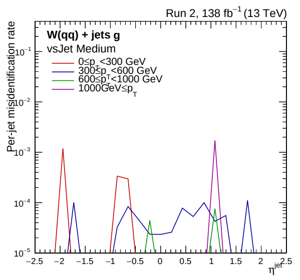
(b) `vvloose`



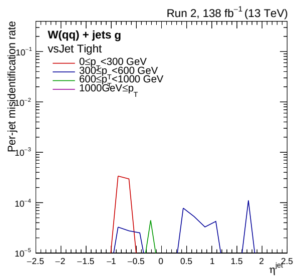
(c) `vloose`



(d) `loose`

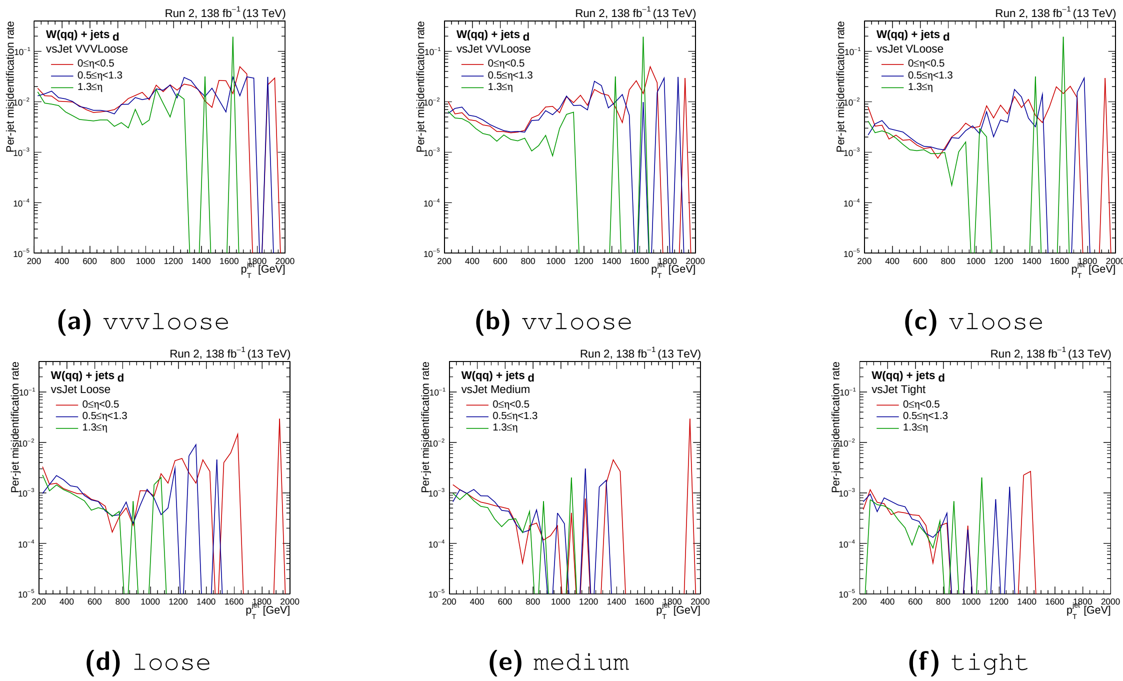


(e) `medium`



(f) `tight`

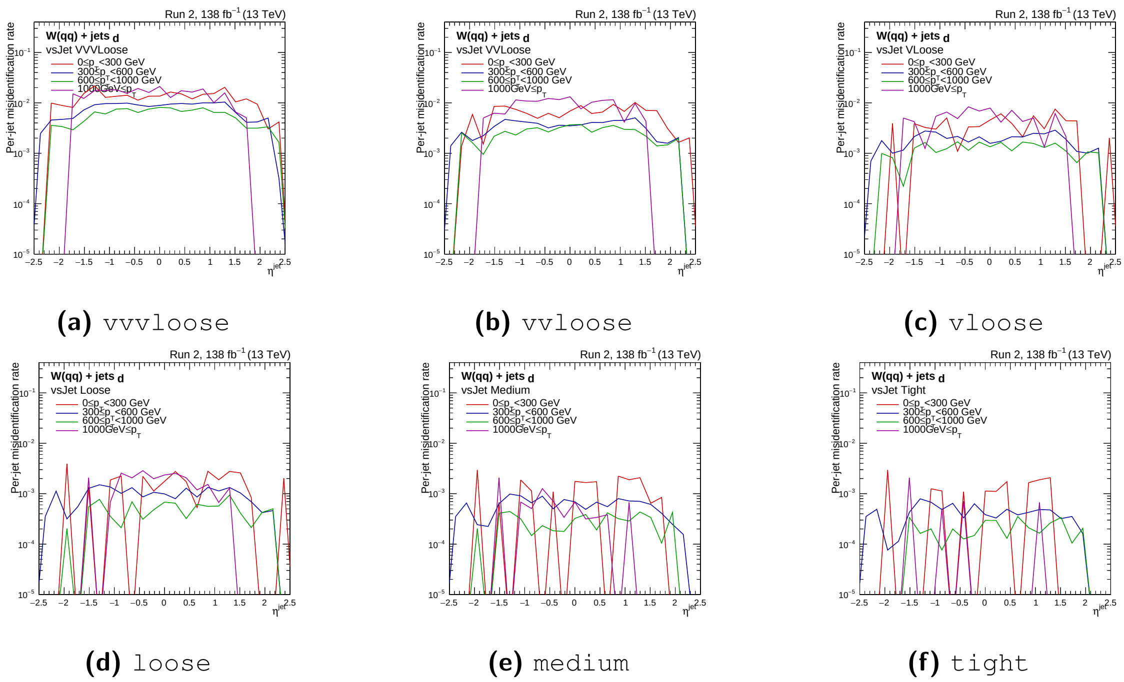
Figure 29: The jet $\to \tau_h$ fake rate as a function of η^{jet} for gluon-initiated jets in four p_T bins ($W(qq)$+jets sample).



(a) vvvloose **(b)** vvloose **(c)** vloose

(d) loose **(e)** medium **(f)** tight

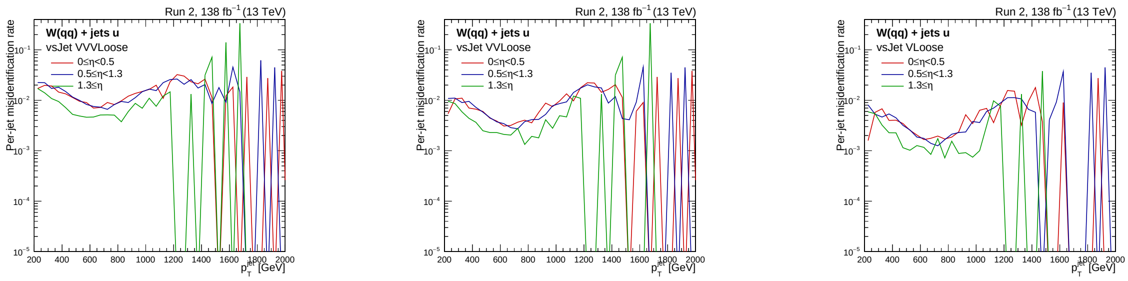
Figure 30: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for down-quark initiated jets in three $|\eta|$ bins ($W(qq)$+jets sample).



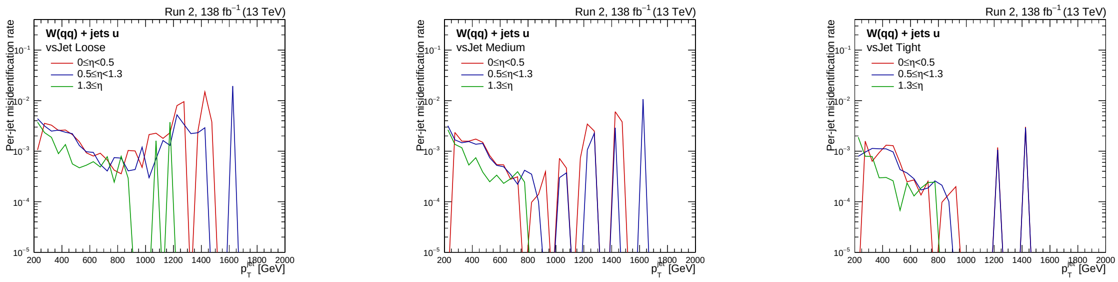
(a) vvvloose **(b)** vvloose **(c)** vloose

(d) loose **(e)** medium **(f)** tight

Figure 31: The jet $\to \tau_h$ fake rate as a function of η^{jet} for down-quark initiated jets in four p_T bins ($W(qq)$+jets sample).

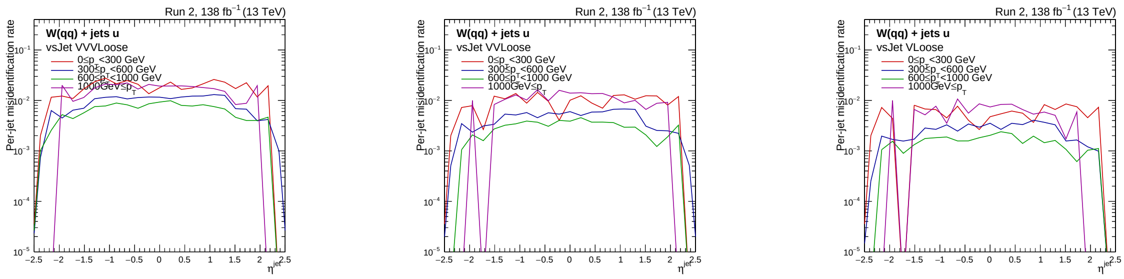


(a) `vvvloose` **(b)** `vvloose` **(c)** `vloose`

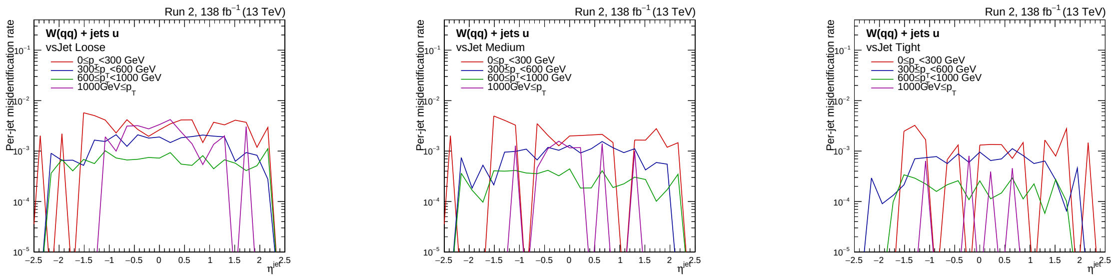


(d) `loose` **(e)** `medium` **(f)** `tight`

Figure 32: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for up-quark initiated jets in three $|\eta|$ bins ($W(qq)$+jets sample).

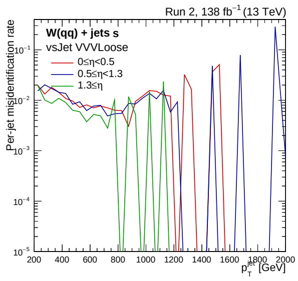


(a) `vvvloose` **(b)** `vvloose` **(c)** `vloose`

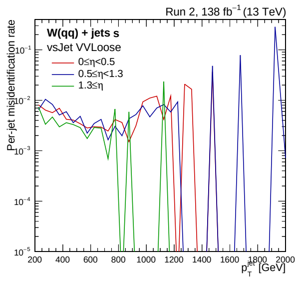


(d) `loose` **(e)** `medium` **(f)** `tight`

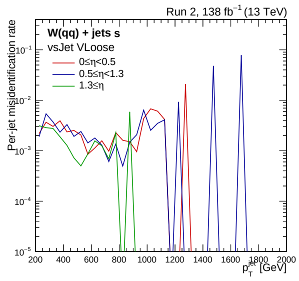
Figure 33: The jet $\to \tau_h$ fake rate as a function of η^{jet} for up-quark initiated jets in four p_T bins ($W(qq)$+jets sample).



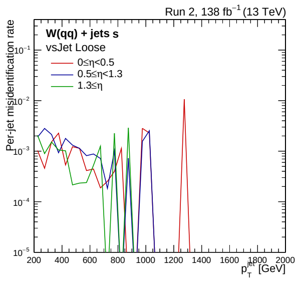
(a) vvvloose



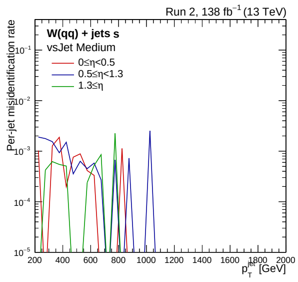
(b) vvloose



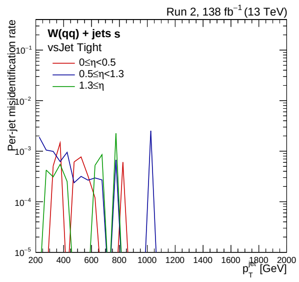
(c) vloose



(d) loose

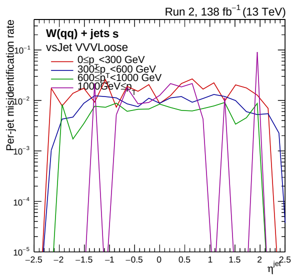


(e) medium

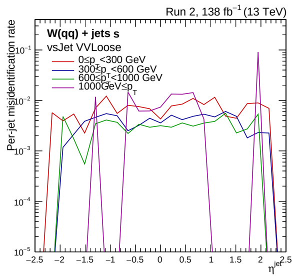


(f) tight

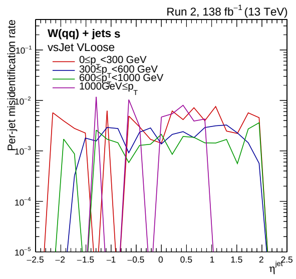
Figure 34: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for strange-quark initiated jets in three $|\eta|$ bins ($W(qq)$+jets sample).



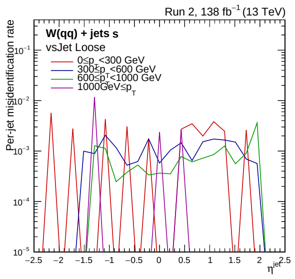
(a) vvvloose



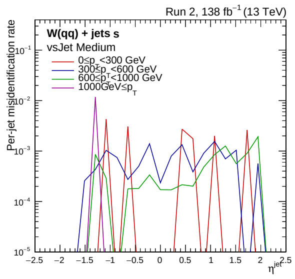
(b) vvloose



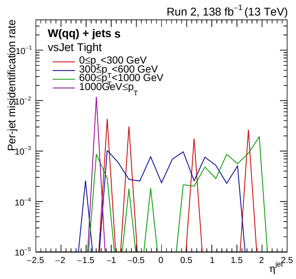
(c) vloose



(d) loose

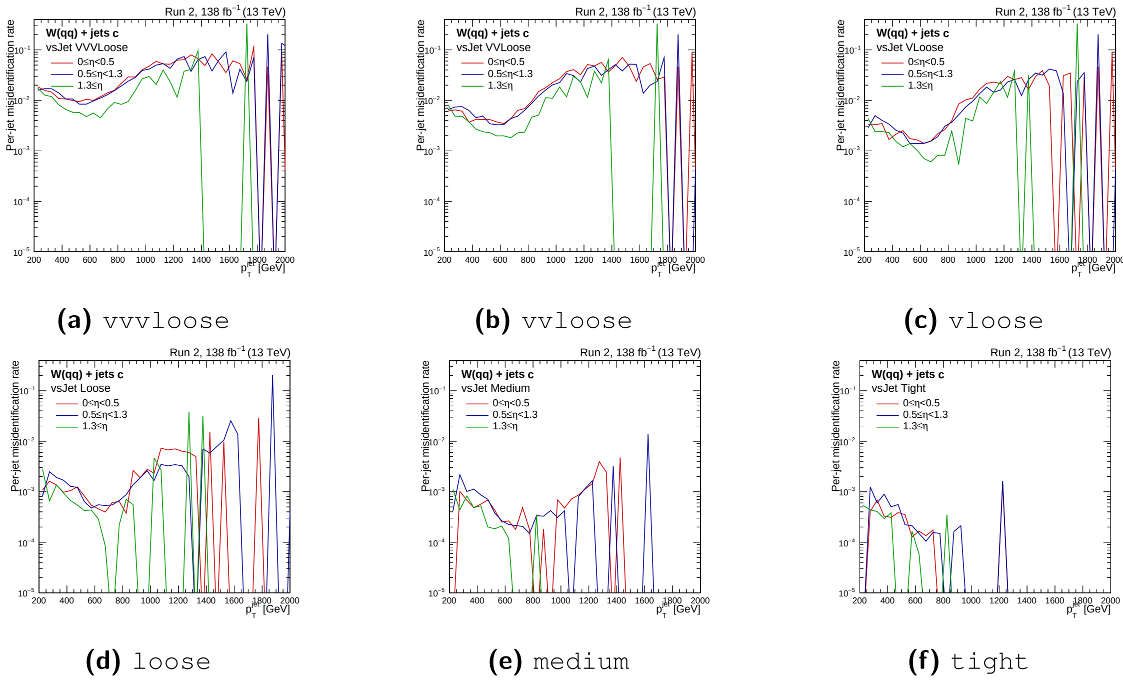


(e) medium



(f) tight

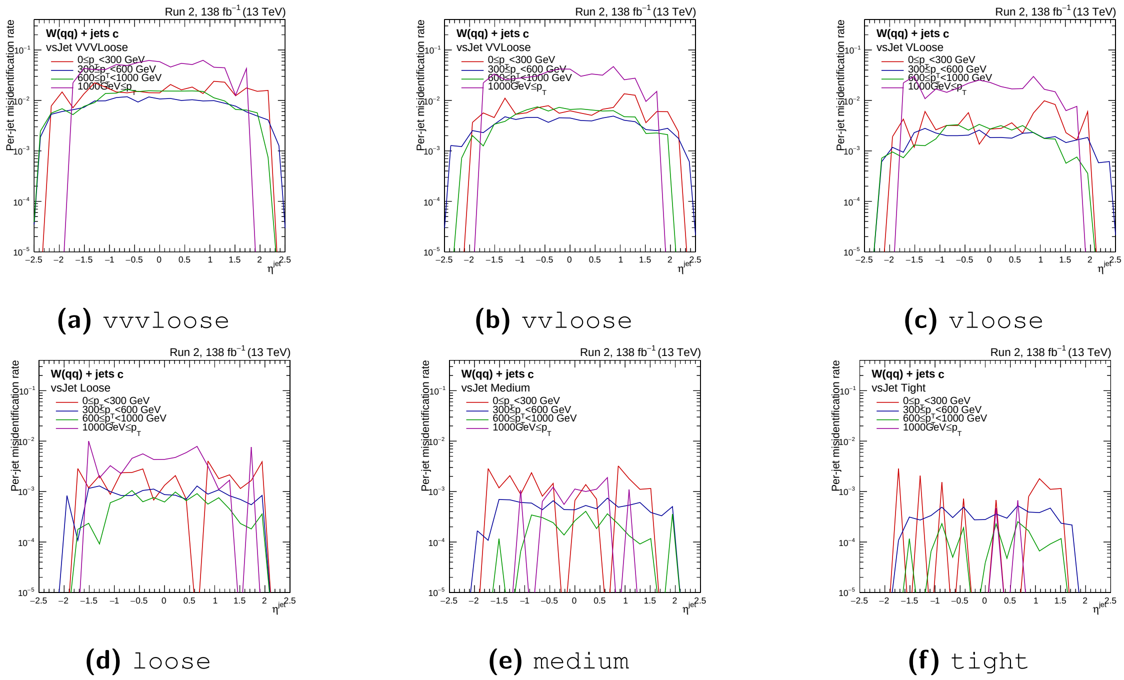
Figure 35: The jet $\to \tau_h$ fake rate as a function of η^{jet} for strange-quark initiated jets in four p_T bins ($W(qq)$+jets sample).



(a) vvvloose **(b)** vvloose **(c)** vloose

(d) loose **(e)** medium **(f)** tight

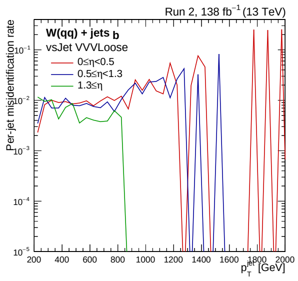
Figure 36: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for charm-quark initiated jets in three $|\eta|$ bins ($W(qq)$+jets sample).



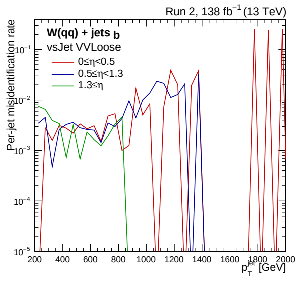
(a) vvvloose **(b)** vvloose **(c)** vloose

(d) loose **(e)** medium **(f)** tight

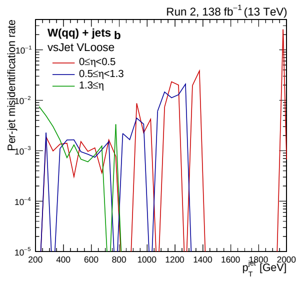
Figure 37: The jet $\to \tau_h$ fake rate as a function of η^{jet} for charm-quark initiated jets in four p_T bins ($W(qq)$+jets sample).



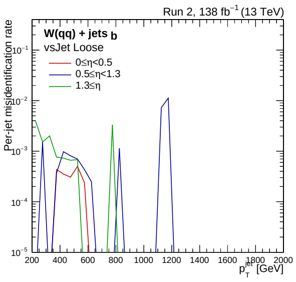
(a) `vvvloose`



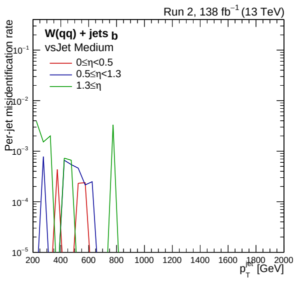
(b) `vvloose`



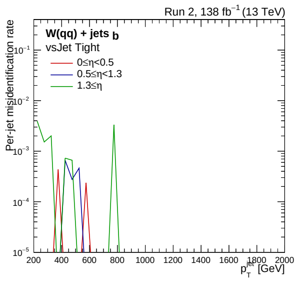
(c) `vloose`



(d) `loose`

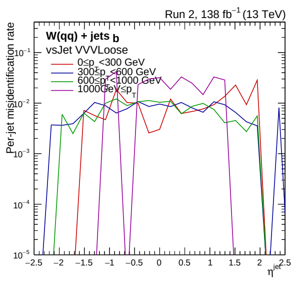


(e) `medium`

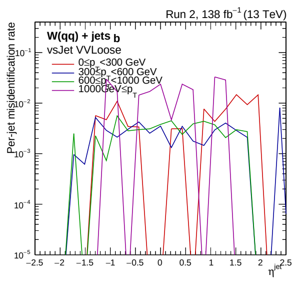


(f) `tight`

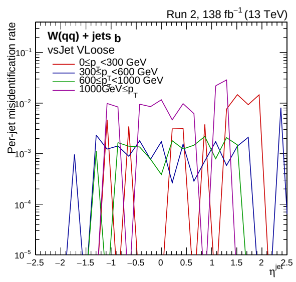
Figure 38: The jet $\to \tau_h$ fake rate as a function of p_T^{jet} for bottom-quark initiated jets in three $|\eta|$ bins ($W(qq)$+jets sample).



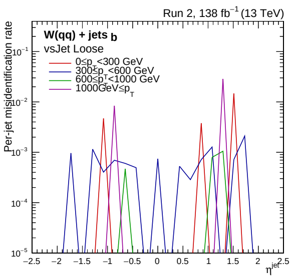
(a) `vvvloose`



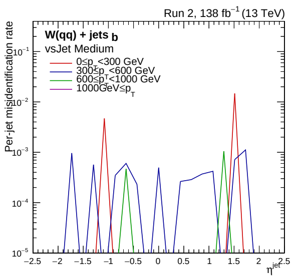
(b) `vvloose`



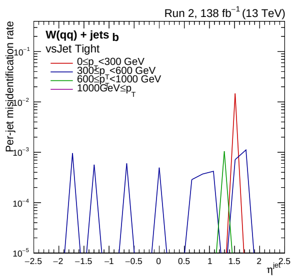
(c) `vloose`



(d) `loose`



(e) `medium`



(f) `tight`

Figure 39: The jet $\to \tau_h$ fake rate as a function of η^{jet} for bottom-quark initiated jets in four p_T bins ($W(qq)$+jets sample).

Acknowledgement

I would like to start by thanking my supervisors, Izaak and Arne, for all their help, support and encouragement. It was a real pleasure working with you both. I could not have asked for better supervisors! You both showed me what a supervisor should be. You both helped me get the best possible outcome from this project. I really appreciate your advice and knowledge during these eleven weeks and I learned a lot from your expertise in physics and in this topic. Thank you for choosing me as your summer student. It was an experience I will never forget. You are not just great at physics and research, you are also fantastic people. I felt safe and comfortable in your group from day one. Thank you for this amazing time!

I would like to also thank the friends I made along the way: Christine, Zsófia, Fabi, Alice, Moritz, Laura, Luisa, Katrin, Janek, Robin, Naza, Houyem and Niklas! You all have my heart! The eleven weeks we spend at CERN were the best eleven weeks I ever had until now. The trips we made, the talks we had, the things we experienced, the evenings and nights in R1, the movie nights - it was an incredible time that I will never forget. It was special. Thank you. It is not a goodbye, it is a see you later! Much love to you all.