Thesis

BELFKIR Mohamed

December 2020

Contents

List of Figures

List of Tables

Jets and Tagging in ATLAS

Jet object and specifically b-quark jet is also crucial to study the $HH \rightarrow \gamma\gamma\bar{b}b$ properties. Jets in ATLAS are currently reconstructed in two ways (EMTopo and PFlow) as explained in Section **??**. This chapter details jet reconstruction, b-jet identification and calibration. I implemented a calibration method specific to b-jets which provides better reconstruction performances, thus better $m_{\bar{b}b}$ invariant mass resolution compared to generic jet corrections which are driven by light jets.

1.1 Jet Reconstruction

For both jet types at Run-2, jet reconstruction starts with a clustering algorithm based on calorimeter objects to form the topological cluster [1]. The algorithm identifies calorimeter cells with energy larger than $4\,\sigma_{noise}$ (expected noise standard deviation) as cluster seeds. Then, neighbouring cells with $|E_{cell}| > 2\sigma_{noise}$ are grouped to form the cluster as well as all adjacent cells. Clusters with two or more local energy maxima are split in two with cells being associated to each maxima according to their relative distance.
The EMTopo jets are built after using the Anti-k_t algorithm. This algorithm identifies the most energetic cluster and merge to it all neighbouring clusters in descending transverse momentum order satisfying :

$$min(k_{T,i}^{-2}, k_{T,j}^{-2})\frac{\Delta R_{ij}}{R^2} < k_{T,i}^{-2},\tag{1.1}$$

where $k_{T,i}^{-2}$ is the transverse momentum of cluster i, ΔR_{ij} is the relative distance between the clusters, and R is the jet radius fixed to 0.4 for all the jets used in this thesis.
Alternatively, the PFlow jets are built by combining the topoclusters to the ID tracking information. Firstly, well-measured tracks are selected by requesting a tight quality selection. Tracks are required to be within $|\eta| < 2.5$ and have $p_T > 0.5$ GeV. Tracks must have at least nine hits in the silicon detectors and no missing Pixel hits. Tracks with $p_T > 40$ GeV are excluded as well as tracks matched to electrons or muons with medium quality criteria and their deposited energy in the calorimeter is taken into account in the algorithm [2]. No requirement on the track-vertex is applied to the track in this stage. Then, the algorithm assigns each track to a single topological cluster to from a track/topo-cluster system. The extrapolated (to the second sampling of the EM) distance between the cluster barycenter and the track is used in addition to the $E^{cluster}/p^{track}$ in the matching. The expected energy in the calorimeter, deposited by the particle, is then computed. Additional topo-clusters are added to the track/topo-cluster system, if the particle deposits its energy in more than one topo-cluster. At this stage, the expected energy deposited by the particle in the calorimeter is subtracted cell by cell from the set of the matched topo-clusters. Finally, if the remaining energy in the system is consistent with the expected shower fluctuations, the topo-cluster remnants are removed. The PFlow algorithm combine between the calorimeter information and soft tracks which do not reach the calorimeter to improve jet resolution at low p_T compare to EMTopo jets (As Figure 1.3 demonstrates). This algorithm is sketched in Figure 1.1.
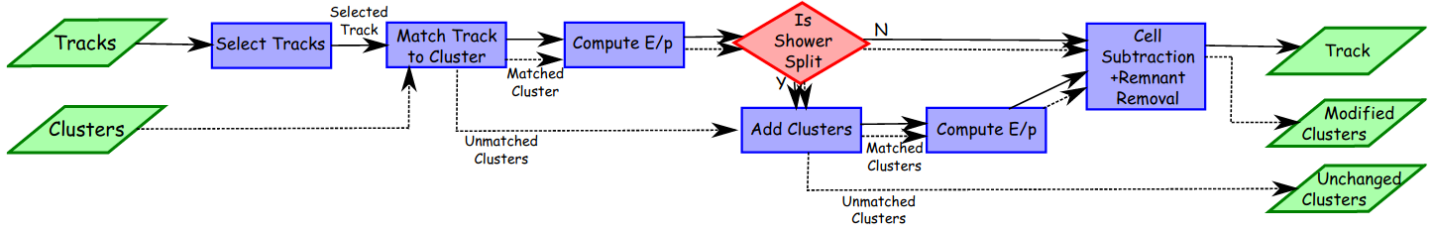
Figure 1.1: A flow chart of how the particle flow algorithm proceeds, starting with track selection and continuing until the energy associated with the selected tracks has been removed from the calorimeter [2].

Similarly to EMTopo jets, anti-k_t algorithm with R=0.4 is used to reconstruct the PFlow jets. The algorithm takes as inputs the survived topo-clusters and the associated tracks that are matched to hard scatter primary vertex. This criteria is applied to reject tracks originating from pile-up interaction [3]. The PFlow is the jet type used in this thesis, while in the rest of the chapter word "jet" means both EMTopo and PFlow.

1.2   Jet Energy Calibration

The resulting jets from Monte Carlo simulation are needed to be calibrated such that, on average, the jet energy corresponds to that observed in the ATLAS detector. A jet calibration procedure is developed to calibrate the jet energy. This calibration proceeds in several sequential stages of calibration derived from a combination of Monte Carlo based methods, which correct the jet 4-momentum to the truth one in simulation, and $in-situ$ techniques which measure the difference in jet response between real and simulated data with residual correction applied to jet in data only.

1.2.1   Jet Calibration Chain

The first step in the calibration chain aims to recalculate the jet direction to point to the hard scatter primary vertex rather than the center of the detector. This correction only affects the jet η and resulting on a better η resolution without changing the jet energy. Figure 1.2 presents an overview of the ATLAS calibration scheme for jets.

Figure 1.2: Calibration stages for jets. Other than the origin correction, each stage of the calibration is applied to the four-momentum of the jet [4].

1.2.1.1 Area-based and residual pile-up correction

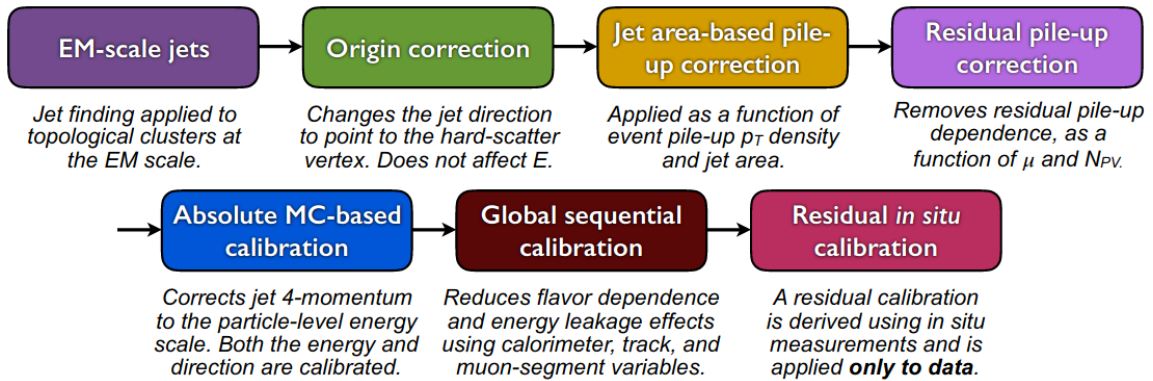This correction aims at subtracting the per-event energy excess due to in-time and out-of-time pile-up (Section **??**). The contribution from pile-up is evaluated from the median transverse energy density ρ. The correction is linked to jet area A and derived for jet within $|\eta| < 2$ given the higher calorimeter occupancy in the forward region. An additional residual correction is derived to complete correct for the pile-up rate dependence remain after the area correction [5]. The dependence is seen on the number of reconstructed primary vertices N_{PV}, sensitive to the in-time pile-up, and the number of pp interactions in the event μ, sensitive to the out-of-time pile-up. The residual correction is derived as the difference between the reconstructed jet p_T and the truth jet p_T using a geometric reconstructed-truth matching with $\Delta R = 0.3$. The area-based and pile-up correction is applied as :

$$p_{\mathrm{T}}^{\mathrm{corr}} = p_{\mathrm{T}}^{\mathrm{reco}} - \rho A - \alpha \left(N_{\mathrm{PV}} - 1\right) - \beta\mu, \tag{1.2}$$

where α and β are the slopes of the linear dependency of jet response respect to N_{PV} and μ.

1.2.1.2 Absolute MC-based calibration

The aim of this correction is to correct, through a Monte Carlo based procedure, the reconstructed jet energy scale (JES) to the particle-level scale as well as the direction biases caused by the different calorimeter technologies and granularity. The average JES response is defined as the mean of a Gaussian fit to the core of the $R = E^{reco}/E^{truth}$ distribution in (E^{truth}, η) bins [6, 7]. A numerical inversion is used to parameterize the correction as a function of E^{reco}, to correction is applied as :

$$E_{JES}^{\mathrm{jet}} = \frac{E^{\mathrm{jet}}}{F(E^{\mathrm{jet}}, \eta)}, \tag{1.3}$$

where E_{JES}^{jet} and E^{jet} are jet energy after and before the JES calibration, the $F(E^{\mathrm{jet}}, \eta)$ is the inverted jet response for given (E,η) bin. A second correction is derived to account for a bias seen in the reconstructed jet η. The η correction is computed as the difference between the reconstructed η^{reco} and the truth η^{truth} as a function of E^{truth} and η. Similarly to JES correction, a numerical inversion is used to derive the correction in E^{reco}. This correction is evaluated on multi-jet events sample. At this step EMTopo (PFlow) jet is denoted EM+JES (PFlow+JES). Comparisons of the jet energy resolution measurement for PFlow+JES and EM+JES jets as a function of jet p_T is shown if Figure 1.3.

1.2.1.3 Global Sequential Calibration (GSC)

In this correction, the aim is to correct effects due to intrinsic properties of the jets not properly taken into account and the hadron composition yielded by the fragmentation process. The average particle composition and shower shape of a jet varies between initiating particles, most notably between quark- and gluon-initiated jets. The GSC uses five observables to improve the resolution of the JES jets. For each observable, an independent jet 4-momentum correction is derived as a function of p_T^{truth} and $|\eta|$ by inverting the reconstructed jet response in Monte Carlo events. The effect of each correction is therefore to remove the dependence of the jet response on each observable while conserving the overall energy scale at the JES. No correlation between observables is take into account. The five observables are :

1. f_{Tile0}, the fraction of jet energy measured in the first layer of the hadronic Tile calorimeter.

2. f_{LAr3}, the fraction of jet energy measured in the third layer of the electromagnetic LAr calorimeter.
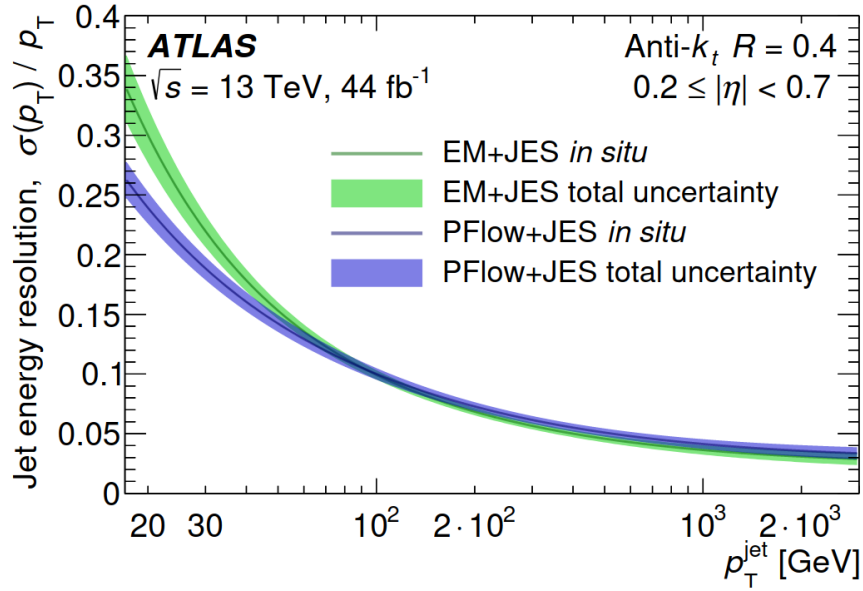
Figure 1.3: The relative jet energy resolution for fully calibrated PFlow+JES jets (green curve) and EM+JES jets (blue curve) as a function of p_T jet.

3. n_{trk}, the number of tracks with $p_T > 1$ GeV associated with the jet.

4. $\mathcal{W}_{trk}$, the average p_T-weighted transverse distance between the jet axis and all tracks of $p_T > 1$ GeV associated to the jet.

5. $n_{segments}$, the number of muon track segments associated with the jet.

1.2.1.4   Residual $in-situ$ calibration

The last step of the jet calibration chain recovers the disagreement between real data a simulated Monte Carlo in the jet response coming from the imperfect detector simulation and mismodelling. The correction is extracted by balancing the jet p_T against that of well-measured reference objects from Z+jets, γ+jets and dijet events. Figure 1.4 shows the final *insitu* combination as a function of jet p_T. To complete the calibration, the inverse of the curve (R_{MC}/R_{data}) is taken as the scaling factor and applied to data. Additional details on the each calibration steps can be found at [4].

Figure 1.3 demonstrates the performance of PFlow at low jet p_T coming from combining calorimeter and track information as well as the low threshold on track p_T, for high jet p_T the performances of both EMTopo and PFlow are comparable.

1.3   Jet Tagging : Pile-up and flavour tagging

Discuss the jet tagging algorithms JVT and b-tagging

Jets coming from pile-up (PU) vertices can have a disastrous impact on physics analysis, by leading to fake jets and therefore biasing the event selection. PU jets are coming either from hard-scatter interaction from a primary vertex that is not the vertex from which the process of interest is arising, defining the most frequent type of jet QCD-PU, or from interactions that are not related to the initial one defining the stochastic PU. Removing PU jets has a critical importance [8, 9]. *Jet Vertex Tagger* (JVT) and
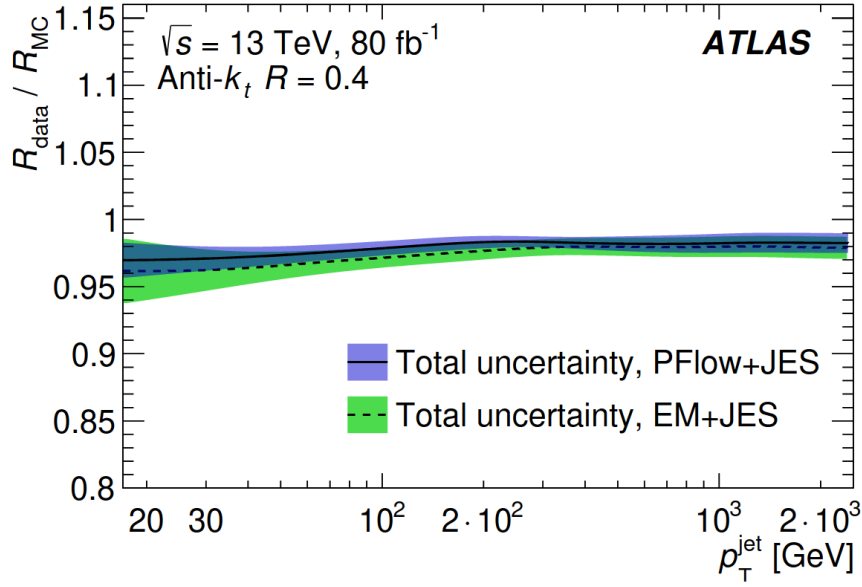
Figure 1.4: Ratio of the PFlow+JES (blue curve) and EM+JES (green curve) jet response in data to that in the nominal Monte Carlo event generators as a function of p_T jet.

forward Jet Vertex Tagger (fJVT) are two algorithms used to identify hard-scatter (HS) and PU jets [10, 11]. In addition to PU jets, identifying the initial nature of the jets is needed for specific analysis such as the $HH \rightarrow \gamma\gamma b\bar{b}$ which require their jets to be originating from the b-quark fragmentation instead of being from gluons or other quarks. For that specific b-tagging algorithm designed to discriminate between b-jets and other jets [12]. This section includes details on each tagging algorithm.

1.3.1 Jet Vertex Tagging (JVT) algorithms

1.3.1.1 JVT

This section will focus on the jet vertex tagger used in the analysis.
Jet can be matched its vertex origin within $|\eta| < 2.5$ (the ID acceptance) where the tracking information is available. Information from the this matching the main ingredient of the JVT algorithm [13]. A global JVT discriminant is constructed from two track-based variables using k-nearest-neighbour algorithm [14]. The two variables are defined as :

$$corrJVF = \frac{\sum_k p_T^{track_k} (\mathrm{PV}_0)}{\sum_l p_T^{track_l} (\mathrm{PV}_0) + \frac{p_T^{\mathrm{PU}}}{k.n_{tracks}^{\mathrm{PV}}}}, \tag{1.4}$$

$$R_{p_T} = \frac{\sum_k p_T^{track_k} (\mathrm{PV}_0)}{p_T^{jet}}, \tag{1.5}$$

The $\sum_k p_T^{track_k} (\mathrm{PV}_0)$ term, appearing in both formulas, represents the scalar sum of tracks p_T originating from the hard scattered vertex and associated to a jet, corrJVF also uses a similar term for PU vertices $p_T^{\mathrm{PU}} = \sum_{n \geq 1} \sum_l p_T^{track_l} (\mathrm{PV}_n)$ where PU vertices are noted as PV_n, $\frac{1}{k.n_{tracks}^{\mathrm{PU}}}$ where k=0.01 is a correction factor used to correct the behaviour from linear increase with the number of tracks. For both EMTopo and PFlow the JVT is built similarly, while WPs are defined differently for the two types of jets given the fact that PFlow has built-in pileup jet suppression in it reconstruction algorithm. For EMTopo three WPs

are defined (Loose, Medium and Tight) and only two WPs (Medium and Tight) are defined for PFlow jets. Figure 1.5 shows the JVT reconstructed variable as well as the fake rate efficiency for JVF, R_{p_T}, corrected JVF and the combined JVT discriminate variables.

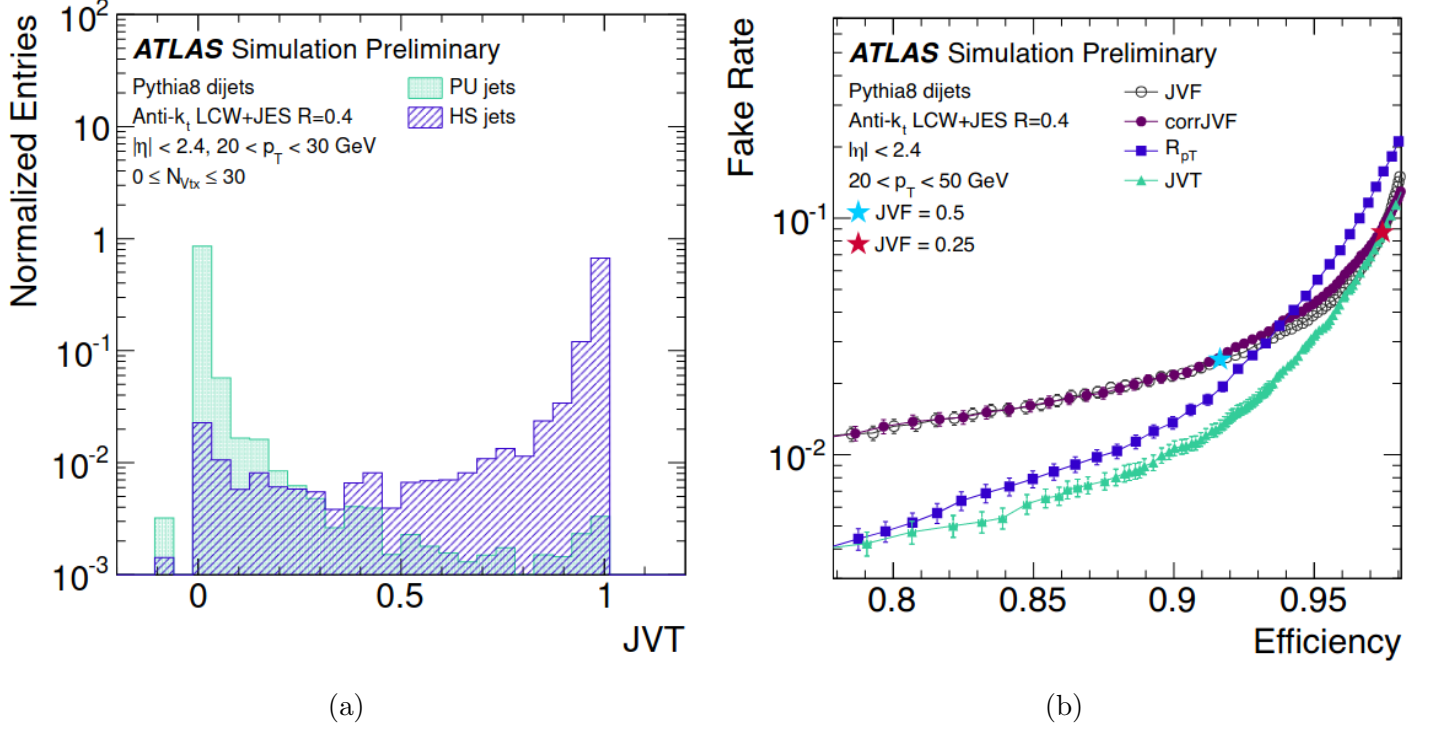(a)   (b)

Figure 1.5: (a) Comparative distribution, scaled to unity, of JVT for truth-matched HS and PU. (b) PU-jet fake rate as a function of the HS-jet efficiency for cuts on the JVT discriminant and JVT input variables. The performance is compared to that of the JVT inputs, corrJVF and R_{p_T}, as well as that of the no corrected JVF.

1.3.1.2 fJVT

For jets outside the ID acceptance ($|\eta| > 2.5$), the JVT can not be used due to the absence of tracking information. The fJVT algorithm procedure is quit different from the JVT since it starts by selecting central QCD PU jets and associate them with the corresponding vertex. Then for each vertex i the missing transvers momentum $-\vec{p}_{T,i}$ is computed, averaging the jet and track component, as :

$$-\vec{p}_{T,i} = \frac{1}{2}\left(k \cdot \sum_{tracks \in PV_i} \vec{p}_T^{track} + \sum_{jets \in PV_i} \vec{p}_T^{jet} \right) \tag{1.6}$$

where k=2.5 is a factor to account for intrinsic differences between the jet and tracks terms. The fJVT discriminant is therefore for a given jet is defined is :

$$fJVT = max(fJVT_i) = max(\frac{-\vec{p}_{T,i} \cdot \vec{p}_T^{jet}}{\vec{p}_T^{jet} \cdot \vec{p}_T^{jet}}) \tag{1.7}$$

where $fJVT_i$ is the jet fJVT with respect to the vertex i. Figure 1.6 shows the fJVT discriminant tends to have large values for QCD pileup jets, while the distribution for hard-scatter jets falls steeply.
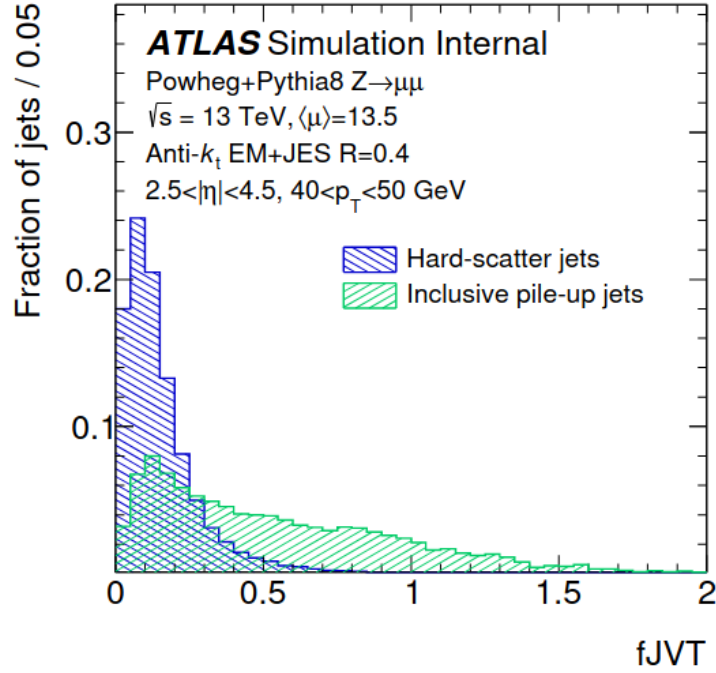
Figure 1.6: The fJVT distribution for hard-scatter (blue) and pileup (green) forward jets.

It is important to note that the fJVT discriminant is only clearly defined for QCD PU jets, as opposed to stochastic pileup jets. To be able to select stochastic jets, fJVT is accompanied with an additional requirement on the jet timing. Two fJVT working point are defined Loose WPs which correspond to fJVT< 0.5 and a Tight one with fJVT< 0.4, their efficiency is shown in Figure 1.7.
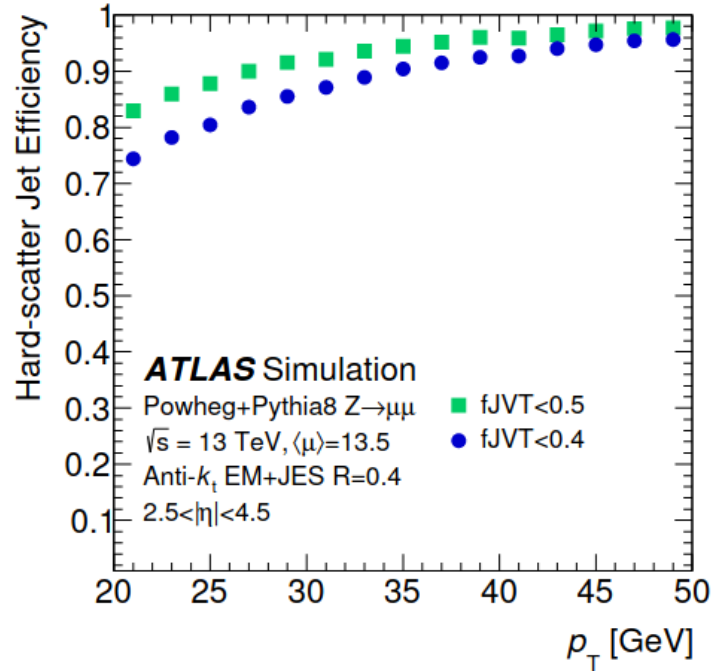
Figure 1.7: Efficiency for hard-scatter jets as a function of the forward jet p_T.

An additional improvement in the fJVT performances is achieved by combining fJVT, timing and jets

shapes and kinematics information using of machine learning techniques [15].

1.3.2 b-jet tagging

This section will discuss the DLr1 and MV2c10 tagger to identify the b-jet.

The identification of jets originating from b-quark decay, called b-tagging, is important for all analyses involving b-quark in their final state such as the analysis presented in this thesis $HH \rightarrow \gamma\gamma b\bar{b}$. The b-jet identification relies of identifying the secondary displaced vertex given the long lifetime of b-hadrons $\sim 491.1\ \mu m$ compare to c-jet or gluon- (light) jets. B-tagging algorithms are designed to discriminate between b-jets and other types. They are based on several low-level algorithms that exploit information from reconstructed tracks in the ID and MS which is combined by a high-level b-tagging algorithm into a single discriminate used at the analysis level. Two high-level b-tagging are developed MV2 which is based on the BDT and the DL1 based on neural networks (NNs), both algorithms use almost the same input from low-level algorithms except for the DL1 which include additional inputs related to the semi-leptonic decay of heavy-flavoured hadrons topology. Each low-level flavour tagging algorithm focuses on different flavour-sensitive information, deriving different sets of variables based on the impact parameter algorithms of tracks in jets (IPs), the presence and properties of an inclusive secondary vertex (SV) and the reconstruction of the b- and c-hadron multi-vertices chain topology (JetFitter):

- Impact Parameter based algorithms : use a log-likelihood ratio (LLR) to separate tracks associated to jets according to whether or not they are compatible to the primary vertex hypothesis. The LLR is computed as the sum of per-track contributions $\sum_{i=1}^{N} log(p_b/p_u)$ where N is the number of tracks and p_b, p_u are the probability density functions (PDFs) for the b- and light-jet hypotheses respectively. The algorithms do not take into account tracks correlation [16, 17].

- Secondary Vertex finding : explicitly reconstructs an inclusive displaced secondary vertex within the jet. All track pairs within a jet are tested for a two-track vertex hypothesis using a χ^2. Two-track vertices are discarded if they were likely to originate from the decay of a long-lived particle, photon conversions or hadronic interactions with the detector material [18].

- Decay Chain Multi-Vertex algorithm (JetFitter) : exploits the topological structure of weak b- and c-hadron decays inside the jet and tries to reconstruct the full b-hadron decay chain. A Kalman filter is used to find a common line on which the primary vertex and the bottom and charm vertices lie, approximating the b-hadron flight path, as well as their positions [19].

Outputs from these algorithms construct input to MV and DL1. Many variations of DL1 exist with different NN configurations and additional inputs. The DL1r adopted in the work of this thesis as well as ATLAS publications also includes input from a Recurrent Neural Network (RNN) based IP algorithm to consider the correlation among tracks [20]. The final discriminant for b-tagging efficiency is calculated as:

$$\text{DL1} = \ln\left(\frac{\text{p}_b}{\text{f}_{c-\text{jets}} \cdot \text{p}_c + (1 - \text{f}_{c-\text{jets}}) \cdot \text{p}_{light-flavour}}\right), \tag{1.8}$$

where p_b, p_c and $p_{light-flavour}$ are the NN outputs probabilities that the jet is a b-jet, c-jet or light-jet respectively. f_{c-jets} is the c-jet fraction taken into account, which can be tuned separately according to physics analysis needs.

For both MV2 and DL1 4 WPs are computed for different b-tagging efficiencies. The WPs are defined as a fixed cut or p_T dependent cut on the final discriminant variable (Eq. 1.8) and labelled with the corresponding integrated efficiency on b-jets. The working points are : 85%, 77%, 70% and 60%. Figure

1.8 shows a comparison of performances for 2018 recommended versions of MV2 and DL1, and 2019 DL1r optimization. The 2019 DL1r shows a significance improvement over other algorithms. Figure 1.9 shows the efficiency of identifying b-jets as a function of jet p_T for 77% WP requirement, for 2018 recommended versions of MV2 and DL1, and 2019 DL1r optimization [21].
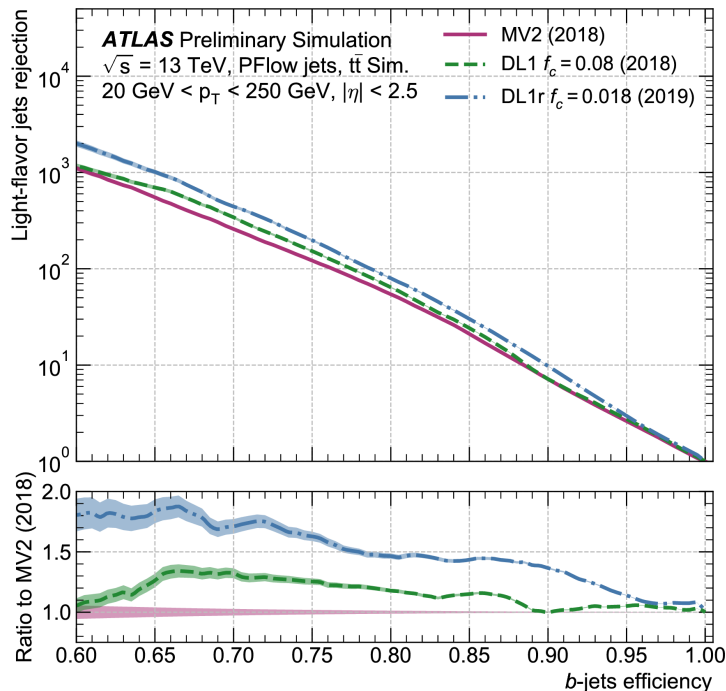

Figure 1.8: ROC curves for 2018 recommended versions of MV2 and DL1, and 2019 DL1r optimization. The x-axis corresponds to the b-jet efficiency, while the y-axis corresponds to light-flavor jets rejection. The shaded bands represent the statistical uncertainty.

1.4 b-jet energy calibration

The b-jet calibration work will fit here.

The default JES calibration (Section 1.2.1.2) applied to all jets is a correction optimised on multi-jet events which are mainly light-quark jets. This correction is not optimal for b-jets due to:

- large b-quark mass which reduces the boost of the jet and, as consequence, a larger fraction of the jet particles lay outside the usual jet cone leading to an out-of-cone radiation.

- larger semi-leptonic branching ratio of b-hadrons which produce elusive neutrino.

Figure 1.11 illustrates these statements through the schematic of a b-jet. To overcome this issue, a step-by-step approach, firstly developed in VH($H \rightarrow b\bar{b}$) analysis demonstrates a potential improvements [22], the first correction called "μ-in-jet" includes soft muon in the jet to address the semi-leptonic decays of b-hadrons inside b-jet. The second correction called "p_TReco" corrects b-jets for the neutrino presence in the semi-leptonic decay and out-of-cone radiation effect by introducing a global correction factor the jet transverse momenta. The developed approach in VH(bb) is computed only for EMTopo b-jets from the signal VH(bb) simulated events. To generalise this correction to any channel with b-jets in the final state its amplitude is computed from b-jets produced in $t\bar{t}$ simulated event for both EMTopo and PFlow jets.
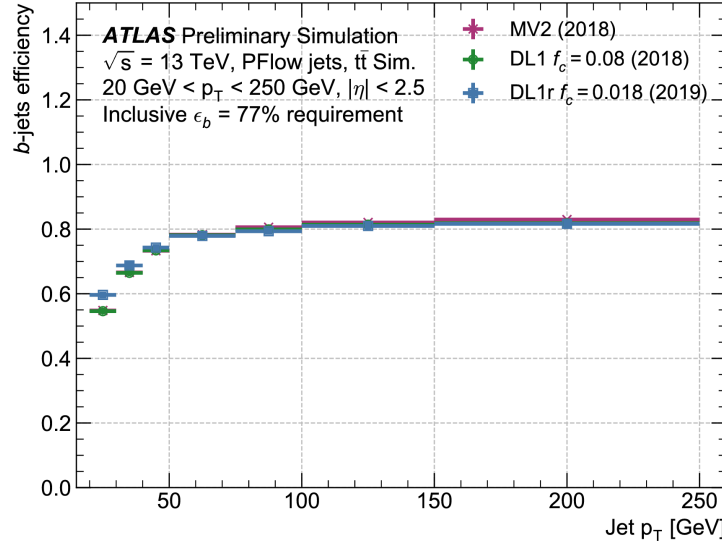
Figure 1.9: Efficiency of identifying b-jets as a function of jet pT for an inclusive 77% efficiency requirement, for 2018 recommended versions of MV2 and DL1, and 2019 DL1r optimization.

This correction is added as an addition step to the calibration chain already mentioned above (Figure 1.10).
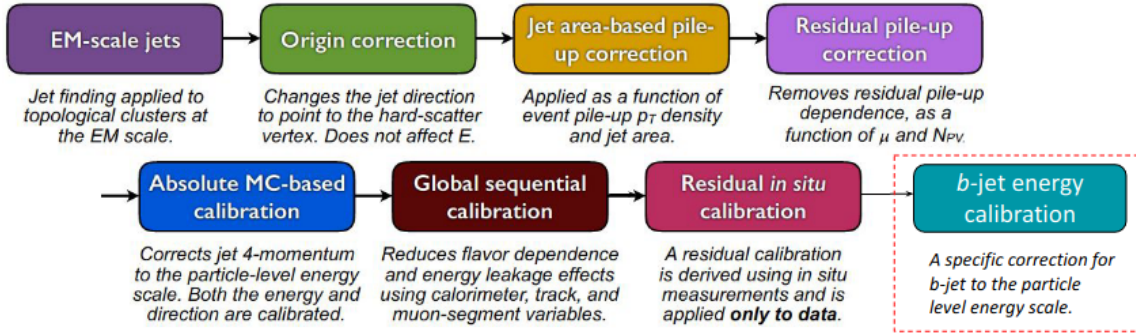

Figure 1.10: Calibration stages for jets including b-jet energy correction.

1.4.1 μ-in-jet correction

The decays of B and D hadrons produce each three type of leptons with a branching ratio of 11%. These leptons are usually mixed with other particles within jets. For electrons and taus, their identification efficiency is low and, as consequence, will not be flagged as semi-leptonic and their energies are fully included in the jet reconstruction as mentioned in Section 1.1. The energy coming from their associated neutrinos will be mixed with p_TReco. Instead, the muon is a minimum ionising particle, depositing a very little energy in the calorimeter. To correct for the semi-lepton decay, the four vector momenta of a muon found inside the jet is added to the the four vector momenta of the jet. To select the associate muon to the jet, selection criteria, listed in Table 1.1, were defined ΔR represents distance between jet and muon axis. VR is computed with using the formula :

$$\min(0.4, 0.04 + 10/p_T) \tag{1.9}$$
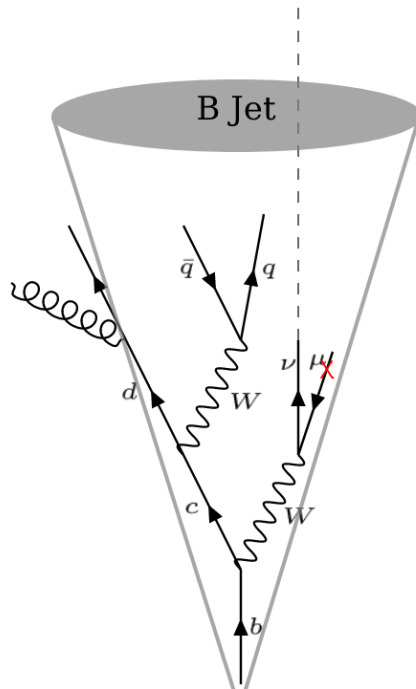
Figure 1.11: Example of a b-jet.

Criteria	Selection
p_T	> 4 GeV
ID	Medium
ΔR	<VR

Table 1.1: Muon selection requested for μ-in-jet correction

with muon p_T in GeV. Instead of a 0.4 fixed cut on ΔR, a variable cut is applied which takes into account the fact that jet boost reduces the angular distance of the muon. If more than one muon are identified, only the one with highest transverse momentum (closest) is used. Such muons are found for 12% of the b-tagged jets. The 4-vector momenta associated to muon passing the selection is added back to the jet one after having subtracted the estimated energy deposited in the calorimeter. The associated jets with a muon are labeled as "semi-leptonic" while others are labeled "hadronic".

1.4.2 p_TReco correction

On top of μ-in-jet correction, a global p_T dependent scale factor is applied to jet 4-vector to correct for the presence of neutrino in semi-leptonic decays and out-of-cone radiation. The relevant target is the particle-level `TruthWZ` jets container, which contains reconstructed jets with the same reconstruction algorithm ($AntiK_t$), using all stable hadrons and non-isolated muons and neutrinos. To generalize the correction and make it signal independent and used by any analysis with b-jet in final state, the correction factor is derived from a $t\bar{t}$ sample. It is binned in $ln(p_T)$ of reconstructed jet up to 400 GeV. The log scale is used because of the exponential shape of p_T spectrum and to provide a smooth distribution for interpolation. The scale factor is computed separately for semi-leptonic and hadronic jets and evaluated as the mean of the distribution of the ratio between reconstructed b-jet p_T and the corresponding `TruthWZ` jet p_T ,

matched within $\Delta R < 0.3$ for the given p_T bin. For each b-tagged jet, the 4-vector is scaled by the scale factor computed using a linear interpolation by `TH1::Interpolate(ln(`p_T` ))` function. The correction is derived for each jet reconstruction algorithms EMTopo or PFlow and for each b-tagging algorithms MV2 or DL1r. Figure 1.12 shows the correction factor distribution for different MV2 jet labels and reconstruction algorithms. Figure 1.12 shows that :
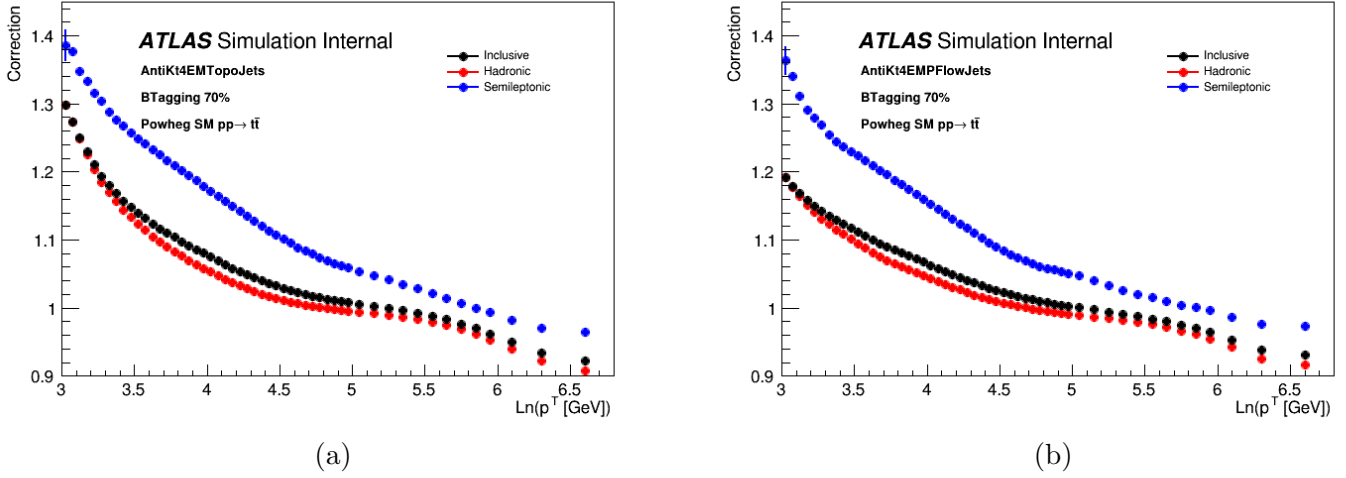

(a)   (b)

Figure 1.12: p_TReco correction distribution for (a) topological cluster jets (b) particle flow jets, b-tagging 70% WP is applied with MV2 algorithm.

- EMTopo hadronic factor is much higher than the one for PFlow at low jet p_T ($\sim 10\%$). This is explained by the fact that PFlow combines tracking, soft activity and calorimetric information to reconstruct the jet, so it includes low p_T ($p_T > 0.5$ GeV) tracks that do not reach the calorimeter, which helps to reduce the out-of-cone effects, as mentioned in Section (1.1).

- The presence of neutrino in semi-leptonic jets explains the higher correction compared to hadronic jets (up to $\sim 10\%$ for low p_T jet).

- At high jet p_T the distribution on the ratio between the reconstructed p_T and the truth one has a large tail below 1 which leads to a correction factor below 1 indicating that the reconstructed jet p_T is higher than the truth. This can be explained by the fact that at high p_T the b-quark fragmentation at truth level gives closest jets in which only one is identified as b-jet, but both jets are reconstructed as a single b-jet.

1.4.3 b-jet energy correction improvement

Figure 1.13, shows the maximum possible gain in b-jet energy resolution in $t\bar{t}$ sample (used to compute the p_TReco correction). The energy resolution is improved by 40% (20%) for high (low) energy for semi-leptonic jet, and about 20% (5%) for high (low) energy for hadronic jet. Almost similar improvement is observed for both jet types (EMTopo and PFlow).

1.4.3.1   Improvement on $H \to \bar{b}b$

The power of $HH \to \gamma\gamma\bar{b}b$ is coming from the $H \to \gamma\gamma$ part wish has a very narrow peak helping in rejection a huge amount of the continuum background by defining a narrow windows on $m_{\gamma\gamma}$ spectrum
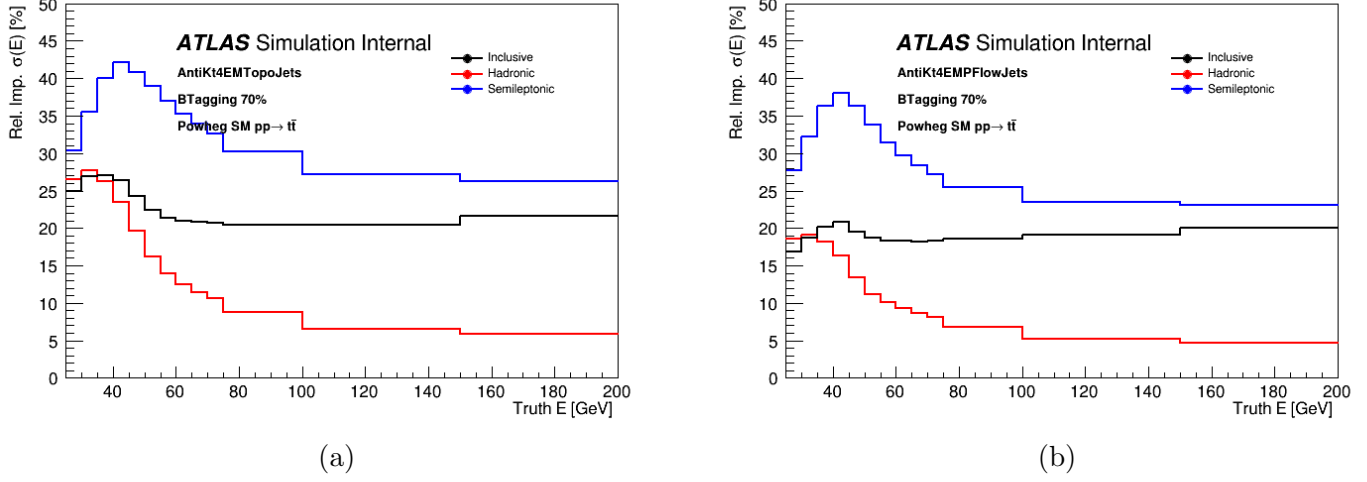
Figure 1.13: Relative energy resolution improvement after b-jet energy calibration in bin of truth energy for $t\bar{t}$ using (a) topological cluster jets (b) particle flow jets.

to select the Higgs signal. The $H \to \bar{b}b$ is suffering from the b-jet mis-calibration and has a large m_{bb} resolution compared to $m_{\gamma\gamma}$, which explain why the b-jet energy calibration is mandatory in such analysis. The gain on m_{bb} resolution from the decoupled method is quantifies on $HH \to \gamma\gamma\bar{b}b$. The gain in b-jet energy resolution reported before is translated into $\sim 22\%$ improvement in the m_{bb} invariant mass resolution as Figure 1.14 shows. In addition, the m_{bb} peak position get closer to the Higgs mass. Similar improvement in both topological and particle flow jets. The two steps μ-in-jet and p_TReco are combined in a decoupled
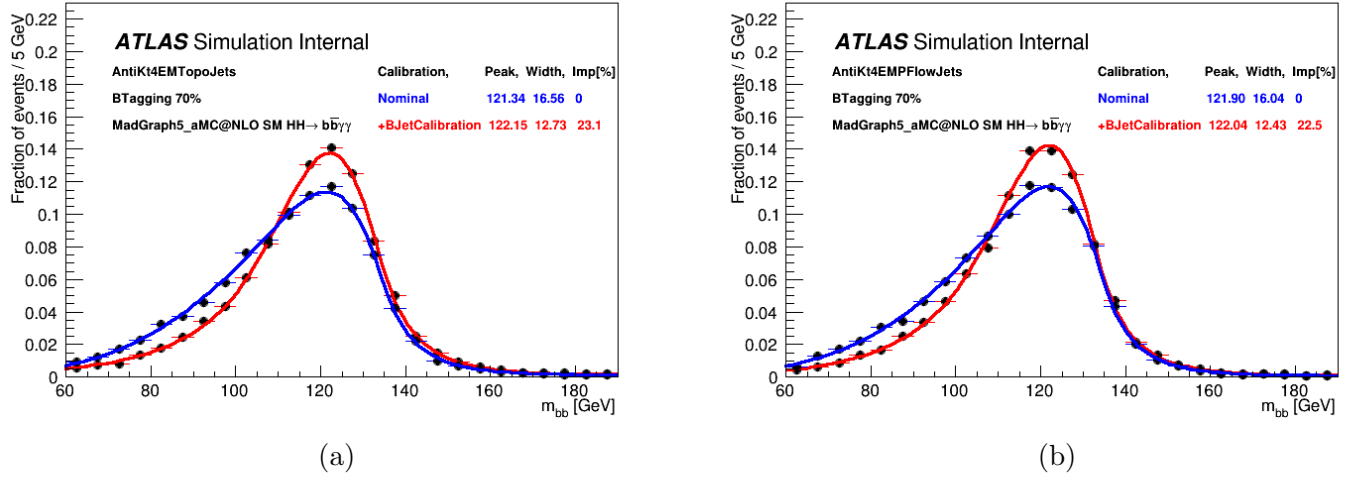
Figure 1.14: Distribution of m_{bb} before calibration (blue) and after calibration (red) in $HH \to bb\gamma\gamma$ for (a) topological cluster jets (b) particle flow jets.

method denoted b-jet energy calibration under a simple `BJetCalibration` tool. The tool is documented and public to be used by analysis with b-jet in their final state. It is been the standard b-jet calibration across all DiHiggs channels with b-jet ($HH \to \gamma\gamma\bar{b}b$, $HH \to b\bar{b}b\bar{b}$, $HH \to b\bar{b}\tau\tau$, $HH \to b\bar{b}ll + MET$ and $HH \to b\bar{b}VV$).

1.5   Conclusion

The related features to reconstruct and identify jets and b-jets mandatory to reconstruct the $HH \to \gamma\gamma b\bar{b}$ events such as reconstruction algorithms, energy calibration, b-tagging and the specific b-jet calibration are introduced in this chapter. A significance improvement is achieved on the m_{bb} resolution leads to a better signal versus background separation. The developed method shows similar performance to an alternative approach which use a simplified NNs regression to calibrate the b-jet energy trained on $t\bar{t}$. The simplified NNs regression uses similar inputs as the decoupled, while the full NNs regression includes additional information about the tracks and secondary vertices ($\sim$90 variables) and achieves a resolution of $\sim$10 GeV in the m_{bb}. The huge number of inputs needed by the full NNS regression makes its implementation at analysis level not trivial and kept for the next analysis round at Run-3.

Bibliography

[1] L Asquith et al. *Performance of Jet Algorithms in the ATLAS Detector*. Tech. rep. ATL-PHYS-INT-2010-129. Geneva: CERN, 2010. URL: https://cds.cern.ch/record/1311867.

[2] M. Aaboud et al. "Jet reconstruction and performance using particle flow with the ATLAS Detector". In: *The European Physical Journal C* 77.7 (2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5031-2. URL: http://dx.doi.org/10.1140/epjc/s10052-017-5031-2.

[3] G. Aad et al. "Charged-particle multiplicities in pp interactions at s = 900 GeV measured with the ATLAS detector at the LHC". In: *Physics Letters B* 688.1 (Apr. 2010), pp. 21–42. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2010.03.064. URL: http://dx.doi.org/10.1016/j.physletb.2010.03.064.

[4] M. Aaboud et al. "Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at s=13 TeV with the ATLAS detector". In: *Physical Review D* 96.7 (2017). ISSN: 2470-0029. DOI: 10.1103/physrevd.96.072002. URL: http://dx.doi.org/10.1103/PhysRevD.96.072002.

[5] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Physics Letters B* 659 (2008), pp. 119–126. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2007.09.077. URL: http://dx.doi.org/10.1016/j.physletb.2007.09.077.

[6] G. Aad et al. "Jet energy measurement with the ATLAS detector in proton-proton collisions at $\sqrt{s}$ = 7 TeV". In: *The European Physical Journal C* 73.3 (2013). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-013-2304-2. URL: http://dx.doi.org/10.1140/epjc/s10052-013-2304-2.

[7] G. Aad et al. "Jet energy measurement and its systematic uncertainty in proton–proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector". In: *The European Physical Journal C* 75.1 (Jan. 2015). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-014-3190-y. URL: http://dx.doi.org/10.1140/epjc/s10052-014-3190-y.

[8] *Pile-up subtraction and suppression for jets in ATLAS*. Tech. rep. ATLAS-CONF-2013-083. Geneva: CERN, Aug. 2013. URL: https://cds.cern.ch/record/1570994.

[9] *Tagging and suppression of pileup jets*. Tech. rep. ATL-PHYS-PUB-2014-001. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1643929.

[10] *Tagging and suppression of pileup jets with the ATLAS detector*. Tech. rep. ATLAS-CONF-2014-018. Geneva: CERN, May 2014. URL: https://cds.cern.ch/record/1700870.

[11] Ariel Schwartzman, Matthew Henry Klein, and Francesco Rubbo. *Identification and rejection of pileup jets in the forward region with the ATLAS detector*. Tech. rep. ATL-COM-PHYS-2016-710. Geneva: CERN, July 2016. URL: https://cds.cern.ch/record/2158916.

[12]   G. Aad et al. "Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV TeV with the ATLAS detector". In: *The European Physical Journal C* 74.8 (2014). ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-014-3023-z`. URL: `http://dx.doi.org/10.1140/epjc/s10052-014-3023-z`.

[13]   G. Aad et al. "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector". In: *The European Physical Journal C* 76.11 (2016). ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-016-4395-z`. URL: `http://dx.doi.org/10.1140/epjc/s10052-016-4395-z`.

[14]   N. S. Altman. "An Introduction to Kernel and Nearest-Neighbor Non parametric Regression". In: *The American Statistician* 46.3 (1992), pp. 175–185. DOI: `10.1080/00031305.1992.10475879`. eprint: `https://www.tandfonline.com/doi/pdf/10.1080/00031305.1992.10475879`. URL: `https://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879`.

[15]   Louis Portales. "Observation of electroweak $WZjj$ production, and studies on pile-up mitigation with the ATLAS detector. Observation de la production $WZjj$ électrofaible, et études sur la suppression des jets d'empilement avec le détecteur ATLAS". Presented 01 Oct 2020. 2020. URL: `https://cds.cern.ch/record/2743491`.

[16]   *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. Tech. rep. ATL-PHYS-PUB-2017-003. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2255226`.

[17]   *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. Tech. rep. ATL-PHYS-PUB-2017-013. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2273281`.

[18]   *Secondary vertex finding for jet flavour identification with the ATLAS detector*. Tech. rep. ATL-PHYS-PUB-2017-011. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2270366`.

[19]   *Topological b-hadron decay reconstruction and identification of b-jets with the JetFitter package in the ATLAS experiment at the LHC*. Tech. rep. ATL-PHYS-PUB-2018-025. Geneva: CERN, 2018. URL: `https://cds.cern.ch/record/2645405`.

[20]   Marie Christine Lanfermann, Tobias Golling, and Andrea Coccaro. *Deep Neural Network based higher level flavour tagging algorithm at the ATLAS experiment*. Tech. rep. ATL-COM-PHYS-2017-1596. Geneva: CERN, 2017. URL: `https://cds.cern.ch/record/2290144`.

[21]   ATLAS Collaboration. *Expected performance of the 2019 ATLAS b-taggers*. Geneva. URL: `http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/`.

[22]   *Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a $b\bar{b}$ pair in pp collisions at 13 TeV using the ATLAS detector*. Tech. rep. ATLAS-CONF-2016-091. Geneva: CERN, Aug. 2016. URL: `http://cds.cern.ch/record/2206813`.