PROCEEDINGS OF SCIENCE

Performance of Jet Reconstruction in CMS at 13 TeV

John Strologas[*]
University of Ioannina, Ioannina, Greece
E-mail: strolog@uoi.gr

We report on the performance of jet reconstruction in CMS during the LHC Run 2. The jet energy scale and resolution measurements are performed on a data sample collected from proton-proton collisions at a center-of-mass energy of 13 TeV. The calibration is extracted from data and simulated events and employs combination of several channels and methods. We also report on boosted object tagging, which is particularly relevant for searches for new physics. Finally we discuss techniques to identify and reject jets originating from pileup and to discriminate between jets originating from quarks or gluons.

38th International Conference on High Energy Physics
3-10 August 2016
Chicago, USA

[*]for the CMS collaboration

PoS(ICHEP2016)736

1. Introduction

Hadronic jets, the main manifestation of Quantum Chromodynamics (QCD) at particle colliders, are a crucial component of the CMS physics program. The precise determination of their momentum and momentum resolution is critical for QCD studies, other Standard-Model measurements and searches for new physics. At the same time, the study of jet substructure leads to identification of boosted heavy particles decaying to jets that merge into a single jet, whereas quark-gluon separation can be used for reduction of QCD background and signal enhancement. We present results from the 13-TeV LHC run, concentrating on particle-flow (PF) jets [1] reconstructed with the anti-k_T algorithm [2].

2. Jet Energy Corrections and Resolution

The jet energy corrections (JEC) are determined using detailed Monte Carlo (MC) simulations and then adjusted for data using data-driven methods applied on several samples [3]. The JEC are are extracted for jets with transverse momentum $p_T > 10$ GeV and pseudorapidity $|\eta| < 5.2$ with uncertainties about $\leq 3\%$ and are performed in stages: First, the pileup offset and noise correction are determined, then the simulation-based response is established (both stages applied to MC and data), and finally absolute and relative residual corrections are applied to data only. Optional flavor corrections can also be applied.

Pileup is the effect additional proton-proton collisions have on the processing of any triggered event. These additional collisions can occur in the same bunch crossing (in-time pileup – ITP) or a different bunch crossing (out-of-time pileup – OOTP). The result is an offset in reconstructed-jet energy. The pileup-offset corrections are determined from the simulation of dijet events processed with and without pileup. They are parametrized as a function of offset energy density ρ, jet area A, η, and $p_{T,\text{uncorrected}}$. The OOTP is mitigated with proper calorimetry signal processing (e.g., by extrapolating the tails of signal within the calorimetry processing time window), whereas ITP is reduced by 50% with the removal from jet clustering of the charged particles that come from additional ("pileup") good vertices in the event ("Charged-Hadron Subtraction" [4]). PUPPI ("Pileup Per Particle Identification" [5]) is another option: it weights charged and neutral jet constituents based on the probability that they come from pileup vertex. After the above pileup filters, the jets are corrected for the remaining pileup utilizing a hybrid jet-area method and simulation/data scale factors are extracted from zero-bias events with random (η,ϕ) cones. The offset scale factor at $|\eta| < 2.4$ is less than 5%, but increases up to 20% outside of the tracking coverage near the boundary between the endcap and forward region of the detector. Pileup-offset error is at the level of 1% at 30 GeV.

The simulation-based correction is applied to jets that have been corrected for pileup offset. Using MC simulations we determine the response $R(< p_T >, \eta) =< p_T > / < p_{T,\text{part}} >$, where p_T and $p_{T,\text{part}}$ are the transverse momentum of reconstructed jet and particle-level jet respectively, binned in bins of $p_{T,\text{part}}$ and η (Figure 1a). The results show that the response is corrected to within 0.5% with respect to the particle-level jet, for p_T from about 20 GeV to 2 TeV. Residual differences between data and MC jet momenta are determined with data-based techniques, and corrections are applied to data jets. All corrections remove initial- and final-state radiation (ISR, FSR) effects.

PoS(ICHEP2016)736

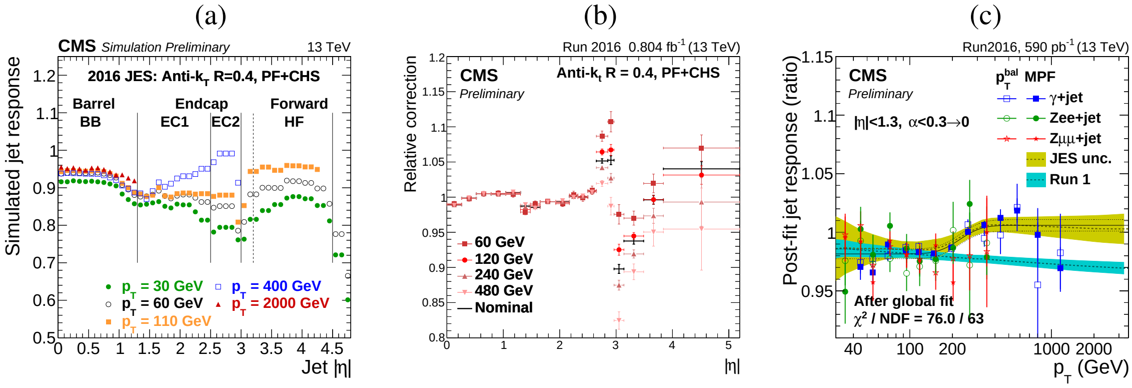

Figure 1: Simulation-based JEC response (a), relative (b), and absolute (c) residual corrections.

The residual corrections are performed in two stages: relative and absolute residual corrections. The relative residual correction is an η-dependent response determined with use of dijet data, with reference jet in the detector barrel ($|\eta| < 1.3$) and by applying the missing-transverse-energy projection fraction method, and it is cross-checked with the common p_T-balance method. Process repeats for data and MC, and the ratio of response gives relative correction seen in Figure 1b. The uncertainty is $\sim 0.5-2.5\%$, depending on η, p_T. For the absolute residual correction, first, a p_T-independent correction is determined with $Z(\rightarrow \mu\mu)$+jet events. The p_T-dependent response within $|\eta| < 1.3$ and $30 < p_T < 800$ GeV is determined with use of Z+jet and γ+jet and multijet events. Above 800 GeV it is constrained with multijet events. All datasets are part of the same global fit, and ISR/FSR biases are removed. Procedure is repeated for data and MC, and the ratio of response gives absolute correction shown in Figure 1c. Uncertainty is at the level of $\sim 1\%$.

Heavy-flavor jets (originated from b and c quarks) have lower response, compared to light-flavor jets, due to their semileptonic decays to softer quarks and leptons. The jet-flavor-specific corrections are derived from MC simulation and checked in data with $Z+b$-jet events.

The jet energy resolution is defined as $\sigma(< p_T > / < p_{T,\text{part}} >)$ after the application of JEC [3]. It is determined with p_T asymmetry in dijet data and p_T-balancing in Z/γ+jets, and it is parametrized as a function of particle-level jet $p_{T,\text{part}}$ and average number μ of pileup interactions in bins of jet η. The resolution is stable against pileup above p_T= 100 GeV and measures about 10% (5%) above 100 GeV (1 TeV).

3. Jet Substructure

Jets that are decay products of a boosted ($p_T/M > 1$) parent particle (top, W, etc) usually merge into a single jet. Several algorithms are developed to identify the substructure within reconstructed wider jets and associate subjets to decay products of the parent particle [6]. Jet-grooming algorithms used include the “pruning” algorithm [7], which removes soft and wide-angle contributions from jets after reclustering, and the “soft drop” algorithm [8] which removes soft jet constituent (less collinear); both combined with pileup removal algorithms. The “N-subjetiness” algorithm [9] uses the distribution of jet constituents relative to jet axis to determine how well the jet can be divided to N subjets. The ratios of the subjetiness is used to specify particular jet substructure (τ_3/τ_2 for top, τ_2/τ_1 for W/Z). Figure 2a shows the W-tagging efficiency as a function of extra (pileup) vertices, whereas Figure 2b shows the application of softdrop-based tagging on boosted top jets.

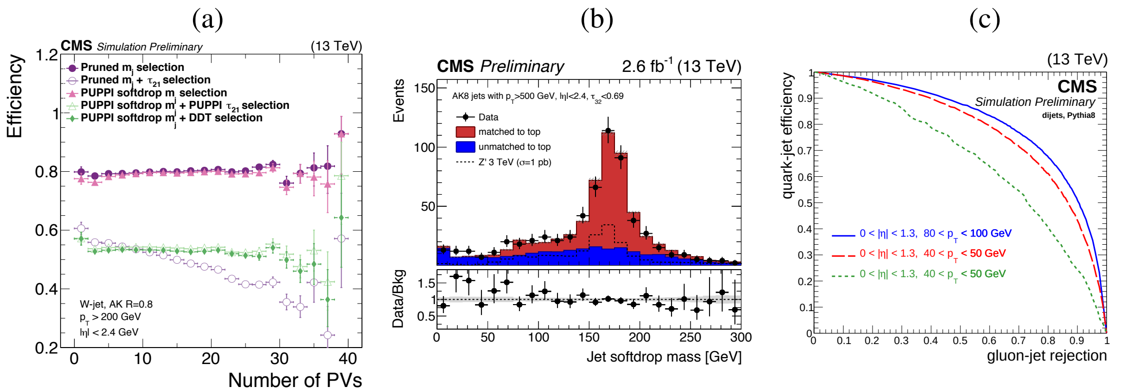

Figure 2: W-tagging efficiency (a), boosted-top performance (b) and quark-gluon ROC curves (c).

4. Quark versus gluon jets

A likelihood that a jet is originated from a quark or a gluon [10] is constructed and trained with QCD multijet simulation (binned in η, p_T, and ρ). Such likelihood can be used to reduce QCD backgrounds in selected analyses, improve the momentum and mass resolution of particles that decay to quarks, and facilitate searches for new physics. Training variables are the jet PF-constituent multiplicity, the minor axis of the ellipse (defined by p_T^2-weighted PF-constituents in (η,ϕ) space), and the jet fragmentation distribution. Figure 2c shows the quark-jet tagging efficiency versus gluon-jet rejection for selected $|\eta|$ and p_T regions.

5. Pileup Jet ID

Pileup jets [11] originate from overlapping lower-energy particles from pileup interactions. A boosted decision tree is utilized to identify them. It uses track-based variables, most sensitive of which is the sum of PF-candidates p_T from primary vertex over sum of all PF-candidates p_T. It also uses shape-based variables, since pileup jets are wider, most sensitive of which is the p_T^2-weighted distance of PF candidates from the jet axis in the (η,ϕ) space.

References

[1] CMS Collaboration, CMS-PAS-PFT-10-001 (2010).

[2] M. Cacciari, G. P. Salam, and G. Soyez, JHEP 04 (2008) 063.

[3] CMS Collaboration, CMS-DP-2016-020 (2016), and CMS-JME-13-004, arXiv:1607.03663 (2016).

[4] CMS Collaboration, CMS-PAS-JME-14-001 (2014).

[5] D. Bertolini, P. Harris, M. Low, N. Tran, JHEP 1410 (2014) 59, and CMS-DP-2015-034 (2015).

[6] J. High Energy Phys. 12 (2014) 017, CMS-DP-2016-039 (2016), and CMS-JME-15-002 (2016).

[7] S. D. Ellis, C. K. Vermilion, and J. R.Walsh, Phys.Rev. D **81**, 094023 (2010).

[8] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, JHEP 1405 (2014) 146.

[9] J. Thaler and K. Van Tilburg, JHEP 1202 (2012) 093.

[10] CMS Collaboration, CMS-DP-2016-070 (2016).

[11] CMS Collaboration, CMS-PAS-JME-13-005 (2013).