Introduction

The identification of jets originated from b quarks is crucial for a broad range of physics analyses. Various b tagging algorithms exist at CMS and are further developed with the use of the machine learning techniques. Constant monitoring of the basic quantities provided to the high-level taggers is fundamental to ensure a good tagging performance and to spot potential issues in the data taking. We present a comparison between the proton-proton collision data collected by the CMS detector in 2016 and simulation. The comparison is between input variables used by the heavy flavour tagging algorithms and the taggers distributions in several event topologies.

Event Topologies

- Inclusive multijet sample: enriched in light jets
- Muon-enriched jet sample: dominated by jets containing heavy-flavor hadrons
- Dilepton ttbar sample: enriched in b jets from top quark decays
- Single-lepton ttbar sample: higher fraction of c jets

Heavy flavour jet identification algorithms

Jet Probability Taggers (JP): combines the probabilities from all tracks to originate from the Primary Vertex (PV).

Combined Secondary Vertex Taggers

- CSVv2 tagger: combines the information of the displaced tracks with the information of the secondary vertices associated with the jet using multivariate techniques.
- DeepCSV: compared to CSVv2, uses a deep neural network with more hidden layers, more nodes per layer, and a simultaneous training in all vertex categories and for all jet flavours.

Soft Lepton (SL) Combined Taggers:

- rely on the presence of a muon or electron from semileptonic b-decays.
- can be used as input for a combined tagger.

Combined tagger cMVAv2: takes input from JP, SL and CSVv2 to perform a MVA training.

Heavy Flavour Jet Discriminating variables

Heavy-flavour jet identification algorithms use variables associated with the properties of heavy-flavour hadrons that are present in jets.

- The lifetime of hadrons containing b quarks is of the order of 1.5 ps. This results in a displacement of the tracks from which a secondary vertex (SV) may be reconstructed.
- Impact Parameter (IP): the distance between the primary vertex and the tracks at their points of closest approach.
- Exploits the large (~15%) semileptonic branching ratio of heavy hadrons. The presence of a muon or electron permits the selection of a pure sample of heavy-flavour jets

Identification of b jets in boosted topologies

- particles decaying to b quarks can be produced with large Lorentz boost
- jets are reconstructed with a cone of R=0.8 (AK8 jets)
- CSVv2 applied either to an AK8 jet or its subjets
- dedicated double-b tagger for boosted jets with double b content

Two samples for validation purposes:

- muon-enriched boosted subjets sample
- double-muon tagged boosted sample

Reference:

1: Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV
The CMS Collaboration JINST 13 (2018) no.05, P05011

Fig I: Data-MC comparison for input variables used for heavy flavour tagging for several topologies. The total number of entries is normalized to the observed entries in data.

Fig II: Data-MC comparison of discriminator distributions for for several topologies. The total number of entries is normalized to the observed entries in data.