Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run

Introduction to b-tagging
- The identification of jets containing b-hadrons (b-tagging) is an important ingredient for several physics analyses (Higgs, Top physics, Beyond Standard Model searches)
- The long lifetime of hadrons with b-quarks (1.5 ps) results in a typical decay topology with at least one vertex displaced from the primary vertex from the hard-scattering collision

- Identification of the b-hadron jets is based on distinct strategies encoded in three basic algorithms:
  - An impact parameter based algorithm (IP), an inclusive secondary vertex reconstruction algorithm (SV) and a decay chain multi-vertex reconstruction algorithm (JetFilter)
  - The output of these algorithms are combined in a multivariate discriminant (MV2) which provides the best separation between the different jet flavours

Monte Carlo samples
- ttbar events generated at 13 TeV
  - simulated with Powheg+Pythia6
- Z–ttbar simulated with Pythia8 at a Z resonance mass of 4 TeV

IP2D and IP3D: The Impact Parameter-based tagging Algorithms
- Input of IP2D/IP3D: transverse and longitudinal impact parameter significance of each track associated to the jet to form a per-track 2D template that takes correlations into account
- Log likelihood ratio (LLR) to separate b from light-jets is computed as the sum of per-track contributions
- Reference histograms are separated in categories depending on the hit pattern of a given track:
  - By knowing whether the hit is expected or not in the various pixel (IBL, b-layer, L1 and L2) or SCT layers or if the hit is split or shared, a quality criteria is defined

Secondary Vertex Finding Algorithm (SV)
- Reconstructing an inclusive displaced secondary vertex within the jet
- Single secondary vertex built by combining all track pairs except when compatible with conversion, V0 decays or material interactions

1. Track multiplicity criteria (use only first 25 tracks ordered in p<sub>T</sub>) for highly energetic jets (p<sub>T</sub> > 300 GeV)
   - reduce the impact of sub-leading tracks from jet fragmentation
2. Additional track cleaning for tracks associated to jets in the end-caps -- mitigate the hadronic interactions with the detector material
3. Tracks with low d<sub>0</sub> significance and high z<sub>0</sub> significance removed -- reduce the impact of pileup and fake vertices

Decay Chain Multi-Vertex Algorithm (Jet Fitter)
- Decay chain multi-vertex reconstruction algorithm exploiting the topology of b-to-hadron decays inside a jet
- Properties of the decay topology and secondary vertices reconstructed by the algorithm

Multivariate MV2 Algorithm and flavour-tagging performance
- Combining output of the three basic taggers (IP, SV, TF) with a BDT algorithm
  - Training of the classifier performed on b-, c- and light-flavour jets from ttbar events
  - Kinematic properties of the jets (p<sub>T</sub>/eta) included among the input variables → b- and light-flavour jets are reweighted to the same p<sub>T</sub> and eta spectrum
- MV2 training procedure and BDT parameters significantly optimized with respect to the 2015 configuration (MV2c20):
  - MV2c20 algorithm chosen (7% c-fraction in the training)
  - +40% c-rejection, +44% light-rejection @ 77% c-efficiency working point

- MV2 algorithmic improvements for the 2016 dataset:
  1. Training parameters employed in the BDT (number of trees, depth, minimal node size) have been optimised to account for the new conditions
  2. Modified training procedure (downgrading) to account for invalid input variables in the training (e.g. whenever IP, SV or TF fail to reconstruct a secondary vertex)
  3. Modification of the c-jet fraction in the training to enhance the c-jet rejection

Wrapping-up and conclusions
- Optimisation of the ATLAS flavour-tagging algorithms ahead of 2016 data-taking
- Development of a new MV2 training resulting in a gain of approximately 40% c- and 4% light-flavour rejection gauged at 77% b-jet efficiency working point compared to the 2015 configuration