Estimation of the b-tag efficiency using top quarks at CMS

The recently started Large Hadron Collider located at the CERN laboratory near Geneva, collides protons at unprecedented energies. This allows the experiments, like the CMS detector, located at the Large Hadron Collider to study the Standard Model of elementary particles at the TeV scale. It is expected that at this scale discoveries of new physics phenomena will be made and that the existence of the Higgs boson will be confirmed or ruled out. In many data analyses at these experiments the b quark plays an important role in discriminating between interesting signal events and the abundantly produced multi-jet background events which are omnipresent at hadron colliders. To identify jets originating from b quarks several powerful b-tagging algorithms have been developed at CMS making use of the specific properties of b quark jets. A good calibration of the performance of these algorithms is crucial for the success of the LHC physics program. An important challenge is to develop calibration methods that are independent of information obtained from simulation.

In this thesis a data-driven method is developed to calibrate b-tagging algorithms based on events where a top quark pair has been created, detected by the Compact Muon Solenoid detector. The top quark decays nearly always to a b quark resulting thus in a large sample of b quark jets in tt events. The method is designed to select semi-muonic decaying top quark events and reduce the large amount of background events coming from multi-jet, W+jets and Z+jets processes. This is done by applying dedicated selection criteria based on the topology of the semi-muonic tt events. For an integrated luminosity of 1 fb⁻¹ at a center-of-mass energy of 10 TeV, about 10000 semi-muonic tt events are expected to be selected while the expected number of background events is about 8400. In the selected events the jets are assigned within the expected decay topology of the event using a jet-quark matching algorithm making use of the top quark mass and W boson mass constraint in tt events. Based on the assignment of the jets, a jet sample is selected which is found to contain around 33% b quark jets. On this selected jet sample the estimation of the b-tag efficiency is performed.

The method developed in this thesis is designed to estimate the b-tag discriminant distribution for b quark jets by selecting a b-enriched (45%) and a b-depleted (15%) subsample in the sample of selected jets. The contribution of jets not originating from a b quark in the subsamples is estimated in a data-driven way from a control sample. In this way the b-tag efficiency can be estimated from collision data for various b-tagging algorithms. For an integrated luminosity of 1 fb⁻¹ at a center-of-mass energy of 10 TeV it is expected that an absolute (relative) statistical uncertainty of respectively 2.9% (3.8%), 2.2% (4.4%) and 1.4% (5.7%) can be reached for a b-tag efficiency of approximately 25%, 50% and 75%. The systematic uncertainty is expected to be respectively 3.2% (4.2%), 2.5% (5.0%) and 1.5% (6.2%), making rather conservative assumptions about the systematic uncertainties.
The method for the inclusive estimation of the $b$-tag efficiency is extended to perform a differential estimation of the $b$-tag efficiency as a function of the transverse momentum and the pseudo-rapidity of the jets. The estimation is performed in five bins of the transverse momentum or the pseudo-rapidity and for an average $b$-tag efficiency of 50%. For a dataset corresponding to 1 fb$^{-1}$ the total relative uncertainty on the estimation of the $b$-tag efficiency is found to range from 11% up to about 33% for the $b$-tag efficiency as a function of the transverse momentum. For the estimation of the $b$-tag efficiency as a function of the pseudo-rapidity the total relative uncertainty is found to range from 15% up to about 27%. The method developed on simulated proton collisions in this thesis is found to be able to provide an estimation of the $b$-tag efficiency. Additionally a differential estimation of the $b$-tag efficiency as a function of the transverse momentum and the pseudo-rapidity of the jets is explored. These results can be applied to calibrate the performance of the $b$-tagging algorithms in a data-driven way and can be cross-checked with the estimation of the $b$-tag efficiency obtained with other methods in the CMS experiment.