Table of Contents

Treatment of the Correlations in b-Tagging Systematics in ATLAS and CMS

Introduction

- Top physics at LHC has entered the realm of precision physics for both experiments ATLAS and CMS
    - ♦ gain in precision by combining the results of both experiments
- Correct treatment of the uncertainties important
- Flavor tagging is one of the dominant systematics uncertainties, therefore compare for ATLAS and CMS
    - ♦ the correlations between flavor tagging algorithms and calibration techniques,
    - ♦ the sources of uncertainty and provide procedures for the combination
- b-jet identification (tagging) is a key ingredient of many analyses
    - ♦ so far no correlation has been considered
- the two collaborations use different approaches regarding every aspect of b-jet identification:
    - ♦ b-tagging algorithms and working point definition
    - ♦ calibrations samples and methods
    - ♦ combination strategy
    - ♦ source of systematics considered and their treatment
- we compared the different approaches, and identified a list of common sources of uncertainty:
    - ♦ treatments of each uncertainty compared to understand how it's effect is correlated in the flavour tagging
    - ♦ size of the uncertainties has been found to be in reasonable agreement across the whole pT spectrum of jets from top decays
- a proposal is advanced for the treatment of b-tagging correlations for future top physics combinations at LHC

Contact Persons

- Contact Person form ATLAS Liza Mjovic, Bonn, Martin zur Nedden, HU Berlin
- Contact Person from CMS Luca Scodellaro, IFCA

Documentation

- ATLAS Flavour Tagging Public Results
- CMS Flavour Tagging Public Results

Flavour Tagging Algorithms at ATLAS / CMS

- b-jets can be identified by the distinctive signatures of the B hadrons from the b-quark hadronization
    - ♦ large life times lead to:
        - ◊ tracks with large impact paramenters (wrt primary vertex)
        - ◊ displaced secondary vertices
    - ♦ high semi-leptonic decay branching ratios

        - presence of soft leptons inside the jet cone
        - charateristical signatures from muons and electrons
- flavour tagging algorithms
    - provide a discriminator value for each jet
    - most performing algorithms using multivariate approaches to combine information from displaces tracks and secondary vertices
- best performing algorithms for ATLAS
    - MV1: combination of the above information with a neural network to build a discriminator
    - operation points: fixed b-jet efficiencies (usually 60, 70 and 80%)
- best performing algorithm for CMS
    - CSV: combination of secondary vertices and track-based lifetime information to build a likelihood-based discriminator
    - operation points: fixed mis-identification rate for light jets at 0.1, 1 and 10%
- top physics analyses in both experiments are using mainly the **medium** operation point (MV1 at 70% (ATLAS) / CSV at 1% (CMS))
    - other tagging algorithm are available and calibrated

Flavour Tagging Calibrations

- Discriminator values from the b-tagging algorithms not perfectly described by Monte Carlo simulations
    - calibration of the MC performance needed
    - done as a function of kinematic variables (e.g. pT or eta)
- measure b-jet content of a b-quark enriched sample before and after flavour tagging requirement
- data-to-MC scale factors for b-jet tagging efficiencies and for light jet mis-identification rates
    - b-tag efficiency SFs provided as a function of jet pT
    - eta dependence also studied and found to be flat in both experiments
    - c-jet calibration: independent analysis at ATLAS, b-jet calibration with inflated uncertainties for CMS
- Measurement of the systematic uncertainty introduced by flavour tagging
    - as important as the measurement of the scale factors
    - central ingredient for the total systematic uncertainty of a measurement using flavour tagging
- Calibration with **di-jet events**: jets with a soft muon
    - exploiting the high semileptonic decay branching ratio of B-hadrons
    - often looking at the away jet to enrich the sample in b-bbar content
    - independent data sample from top selection
    - muon requirement can have a bias to the impact parameter based flavour tagging algorithms
- Calibration with **top-pair events**: jets from top quark decays
    - BR(t Wb) ~100%
    - isolated leptons from W decays to reduce the background contamination

    - single lepton and dilepton decay channels providing two orthogonal samples
    - very pure and well known b-jet sample
    - using the same data set as for top-physics measurement
        - systematic uncertainties must be treated carefully

Calibration methods

- Di-jet based calibrations:
    - ATLAS:
        - pTrel: template fit of the muon pT w.r.t the jet axis (only for 7 TeV data)
        - System8: equations with 8 unknowns, using two samples (with different purities) and two weakly correlated taggers (tagger under study and the muon pTrel)
    - CMS:
        - pTrel and system8 method as well
        - extending pTrel method up to 800 GeV looking at muon IP3D
        - Lifetime tagger (LT) method: template fit of a reference discriminator (usually JP which is calibrated in data)
- top-pair based calibration:
    - ATLAS:
        - tag counting: fit b-jet efficiency on tagged jets multiplicity (only for 7 TeV data)
        - Kinematic selection and tag&probe: based on sample composition estimates
        - combinatorial likelihood PDF approach (only 8 TeV data), **most precise measurement**
        - kinematic fit: use a kinematic fit to reconstruct the final states and increase the purity of the sample
    - CMS:
        - Tag counting techniques (inclusive in pT)
        - bSample method (kinematic based, inclusive in pT)
        - LT method on dilepton events (only for 8 TeV data)
- Final usage: combination of calibrations
    - ATLAS:
        - various combinations of the calibrations provided among di-jets (pTrel, S8) and ttbar calibraions
        - default calibration is a combination of di-jet (S8) and ttbar (PDF-method)
        - the choice of the combination is up to the individual analysis teams
        - global fit with all systematic uncertainties as nuisance parameters which can shift the mean data/MC SF
        - systematic uncertainties can be fully correlated or uncorrelated in each single kinematic bin or across kinematic bins
    - CMS:
        - all measurements on jets with a soft muon and top based LT analysis
        - combination also provided without top-pair based calibrations
        - using the least squared BLUE method

        - source of uncertainties common between two or more methods are taken as (anti-)correlated
        - systematics considered correlated across the bins

Correlations between Systematics among ATLAS and CMS

- Two types of correlations of the systematic uncertainties of flavour tagging calibrations
    - correlations with other parts of **top analyses**
        - sources of uncertainties on flavour tagging performance measurements also considered in the physics analysis
        - example: Jet Energy Scale in a top quark mass measurement
    - correlations between the two **experiments**
        - common sources of uncertainty
        - related to general physics modeling of the calibration sample
- correlations affecting different categories of systematics
    - General physics modelling
        - e.g. ISR/FSR, parton showering
        - correlated to analysis and between experiments
    - Specific physics modelling
        - e.g. pT spectrum of soft muons, light/charm ratio
        - uncorrelated to analysis but correlated between experiments
    - Detector description
        - e.g. JES, pile up
        - correlated to analysis but uncorrelated between experiments
    - Method specific
        - uncorrelated to analysis and between experiments
- **Proposed treatment of systematics:**
    - Systematic uncertainties uncorrelated between experiments will be added together in an uncorrelated systematics category
    - for the time being, the correlations between different part of top physics analyses will not be considered
        - experiments must decide internally how the deal with them
    - the main sources of systematic uncertainties correlated between the experiments will be quoted separately in the systematic breakdown
    - the systematics which come from the same source will be treated with a correlation factor of one in the combination
    - the very small correlated systematics will be merged to an existing similar category

Physics Modelling Systematics

- The sources of systematic uncertainties correlated between the experiment are related to the general physics modelling in the simulated samples
- Physics modelling of muon jets
    - production of b- and c-quarks (fraction of gluon splitting)
    - decays of B-hadron

    - b-quark fragmentation
    - ratio of charm-to-light jets
    - simulation of the pT spectrum of the muon
- Physics modelling of top-pair events
    - top-pair MC generator
    - description of the parton shower
    - ISR / FSR
    - underlying events

Source	treatment in ATLAS	Treatment in CMS	Corr.
b/c production	b,c --> gg scale by 50%	b,c --> gg scale by 50%	YES
B decay	reweight acc. to BR	neglected	NO
b-quark frag.	av. B hadron energy fraction varied by +/- 5%	av. B hadron energy fraction varied by +/- 5%	YES
c/l ratio	c/l ratio scaled by factor 2	l/c ratio scaled by 20%	YES
muon pT	pT spectrum reweighting	vary cut on muon pT	YES
top generator	compare POWHEG to MC@NLO	compare fit to templates for QCD	NO
parton shower	compare HERWIG to PYTHIA	comoare HERWIG to PYTHIA	YES
ISR / FSR	use AcerMC+Pythia	varying Q2 scale and ME-PS threshold	YES
underl. event	negligible	varing parameters	NO

Split of Systematics

- Six major correlated sources identified
    - b/c production, muon pT, charm-to-light ratio, b-fragmentatin, PS, ISR/FSR
    - contributing at the level of 0.2 - 1.3%
    - to be provided as separate uncertainties
    - charm-to-light systematics equally small for ATLAS and CMS (up to 0.2%)
        - very small uncertainty
        - **add to b-fragmentation, not consider separately**
- Analyses should propagate in their analyses all the six uncertainties as a function of pT
- Analyses should merge the charm-to-light systematics with the b-fragmentation systematics into one systematics
    - we have then only five systematics left for the combination: b/c production, muon pT, charm-to-light ratio + b-fragmentatin, PS, ISR/FSR
- Resulting five uncertainties of final result can be combined taking into account the correlations among the experiments
- Summary table (for all six systematics, 8 TeV data):

source	**size at ATLAS**	**size at CMS**
b/c prod.	low pT: 0.1% - 0.2%, high pT for b-prod.: 1.2% - 2.0%	low pT: 0.1% - 0.3%, high pT: 0.5% - 1.3%
mu pT		

	first pT bin: 2.5%, 0.2% - 0.9% elsewhere	low pT: 0.1% - 1.1%, high pT: 0.1 - 0.9%
c/l ratio	<0.1% - 0.2%	<0.1% - 0.2%
b-frag	0.2% - 2.7%	0.2% - 0.8%
PS	0.1% - 1.5%	0.3% - 0.6%
IFSR	0.3% - 1.4%	0.3% - 0.6%

Systematic Breakdown

- The breakdown of the systematic is provided for the tagger mostly used in top analyses, and for different data taking and calibration periods
    - please, contact us know if your analysis is using
- ATLAS breakdown of systematics provided for MV1 and MV1c in the CalibrationDataInterface for the 8 TeV data :
    - link to internal ATLAS twiki
- CMS breakdown for CSVL and CSVM are provided for the PromptReco and Winter13 reprocessing of 8 TeV data:
    - link to internal CMS twiki

-- RobertoChierici - 13 Jun 2014

This topic: LHCPhysics > BTaggingSystematics
Topic revision: r12 - 2015-11-12 - LizaMjovic