PROCEEDINGS OF SCIENCE

Two New Developments on the Statistical Treatment of Flavour Tagging Uncertainties in ATLAS

Yan Ke,[a] **Ilaria Luise,**[a,*] **Giacinto Piacquadio**[a] **and Quentin Buat**[b]

[a] *Stony Brook University, Stony Brook, NY - USA*
[b] *University of Washington, Seattle, WA - USA*

E-mail: ilaria.luise@cern.ch

The document introduces two new methods on the implementation of flavour tagging uncertainties in ATLAS physics analyses. In order to reduce the number of flavor-tagging calibration uncertainties, the physics analyses use an eigenvector decomposition approach. However, the resulting flavour tagging eigenvectors are in general not the same across flavour tagging selections, so the uncertainties can not be directly correlated in combination analyses. A new method, called *eigenvector recomposition*, has been designed to overcome this problem. This proceeding describes the method and gives practical examples about its usage in physics analyses, focusing on the $VH, H \to b\bar{b}$ analysis.

The second development involves the flavour tagging uncertainties in analyses with high-transverse momentum jets. The in-situ calibration of the flavour tagging uncertainties is computed using events with jet-p_T spectra up to 140-250 GeV and used through all the jet-p_T spectrum. Therefore, at higher transverse momenta the calibration needs dedicated extrapolation uncertainties in order to account for possible deviations from the central value. The second part of the document describes the method used to extract these extrapolation uncertainties starting from Z' simulated events.

41st International Conference on High Energy physics - ICHEP2022
6-13 July, 2022
Bologna, Italy

*Speaker

1. Eigenvector recomposition

About ~100 separate uncertainty sources (η) are accounted for in the b-jet calibration analysis. As described in [1], an *eigenvector decomposition* method is used to reduce the number of flavor-tagging uncertainty components in the analyses to $O(10)$ θ parameters while correctly preserving correlations across jet kinematic bins. The eigenvectors are closely related to the diagonalization of a specific covariance matrix, so they can not be directly correlated across analyses. This study introduces the concept of *eigenvector recomposition* to overcome this problem.

Most physics analyses using b-tagging in ATLAS rely on a maximum-likelihood fit to extract their parameters of interest, with the diagonalized b-tagging uncertainties encoded in the nuisance parameters θ. The corresponding likelihood function is:

$$\mathcal{L} = L(\vec{\mu}, \theta_1, \theta_2, ..., \theta_L)P(\theta_1)P(\theta_2)...P(\theta_L), \tag{1}$$

where $\vec{\mu}$ is the vector of the parameters of interest, for example the Higgs boson production signal strengths, and the $P(\theta_k)$ represent the Gaussian priors centered at zero and with variance set to unity for the flavour tagging eigenvectors.

The aim of the eigenvector recomposition approach is to transform the likelihood function in Eq. 1, back into the corresponding likelihood function where the b-tagging uncertainty components are encoded into the nuisance parameters η of the original sources of systematic uncertainties:

$$\mathcal{L} = L(\vec{\mu}, \eta_1, \eta_2, ..., \eta_M)P(\eta_1)P(\eta_2)...P(\eta_M), \tag{2}$$

where the $P(\eta_k)$ again represent Gaussian priors. The nuisance parameters θ (after diagonalization) and the nuisance parameters η (before diagonalization) represent the exact same uncertainties in the space of the b-tagging scale factors, i.e. the same central values and same covariance matrix.

In order to modify Eq. 1 to make it identical to Eq. 2 the the variables $\vec{\theta}$ are expressed as a function of $\vec{\eta}$, using a translation matrix described in Ref. [1].
The validation has been performed modifying the $VH, H \rightarrow b\bar{b}$ analysis workspace (WS) [2]. The recomposed uncertainties in the modified workspace are equal to the original source uncertainty within 0.2%. The ful mathematical description of the method and its validation are further described in Ref. [1].

2. High-p_T extrapolation uncertainties

The second development introduced in this note involves the treatment of the calibration uncertainties for high-p_T jets, in particular in the extrapolation regime above the highest p_T of the in-situ flavour tagging calibration.

2.1 Derivation of the extrapolation uncertainties

The p_T value that determines the division between high p_T region and the standard in-situ calibration region is called the p_T *reference point*:

$$SF_{p_T} = SF_{p_T,ref} \cdot R(p_T) \tag{3}$$

where R is a factor depending on p_T. The value of R is assumed to be equal to one since the SFs do not have a strong dependence on p_T, as reported in Ref. [3]. R has associated uncertainties that need to be evaluated and propagated as additional extrapolation uncertainties to cover possible deviations from unity. The relative uncertainty of the SF in the high-p_T extrapolation region can be expressed as $\Delta_{rel}(SF_{p_T})$. It depends on the relative SF uncertainty at the reference point $\Delta_{rel}(SF_{p_T,ref})$ as determined by the in-situ calibrations and on the relative uncertainty of R, defined as $\Delta_{rel}(R)$, as follows:

$$\Delta_{rel}(SF_{p_T}) = \frac{\Delta(SF_{p_T} \cdot R)}{SF_{p_T} \cdot R} = \frac{\Delta(SF_{p_T,ref}) \cdot R + SF_{p_T,ref} \cdot \Delta(R)}{SF_{p_T} \cdot R} \tag{4}$$

$$= \Delta_{rel}(SF_{p_T,ref}) + \Delta_{rel}(R) \tag{5}$$

The relative variation of R is evaluated from Monte-Carlo samples as described below[1]. Potential variations of the efficiency in data with respect to MC are estimated based on the systematic variations applied to simulation as $\Delta(\epsilon_{data}) = SF \cdot \Delta(\epsilon_{MC}) = \frac{\epsilon_{data}}{\epsilon_{MC}}\Delta(\epsilon_{MC})$. Therefore, given the variation of the MC b-tagging efficiency, corresponding to a specific source of uncertainty the scale factor is varied correspondingly by $SF' = \frac{\epsilon_{data}+\Delta(\epsilon_{data})}{\epsilon_{MC}} = \frac{\epsilon_{data}(1+\Delta_{rel}(\epsilon_{MC}))}{\epsilon_{MC}} = SF(1+\Delta_{rel}(\epsilon_{MC}))$. Based on the result above, the variation on R becomes:

$$R' = \frac{SF_{p_T}(1 + \Delta_{rel}(\epsilon_{MC,p_T}))}{SF_{p_T,ref}(1 + \Delta_{rel}(\epsilon_{MC,p_T,ref}))} \approx R \cdot (1 + \Delta_{rel}(\epsilon_{MC,p_T}) - \Delta_{rel}(\epsilon_{MC,p_T,ref})) \tag{6}$$

Therefore, the relative uncertainty on R is $\Delta_{rel}(R) \approx \Delta_{rel}(\epsilon_{MC,p_T}) - \Delta_{rel}(\epsilon_{MC,p_T,ref})$, assuming R=1. Combining the equations above, the SF extrapolation uncertainty becomes:

$$\Delta_{rel}(SF_{p_T}) = \Delta_{rel}(SF_{p_{T,ref}}) + \Delta_{rel}(\epsilon_{MC,p_T}) - \Delta_{rel}(\epsilon_{MC,p_{T,ref}}) \tag{7}$$

2.2 Results

The extrapolation uncertainties are computed using simulated $Z' \to q\bar{q}$ ($q = uds, c, b$) events with a Z' mass of 3 TeV to enhance the high-p_T region. The jets in the sample are reconstructed as Variable-Radius (VR) track-jets [4]. The extrapolation uncertainties are evaluated separately for truth b-, c- and *light*-jets. The uncertainties considered in the extrapolation are tracking and modelling uncertainties. In particular, the modelling uncertainties include final state radiation, parton shower modelling and quasi-stable simulation[2].

Figure 1 shows the total uncertainty together with the modelling and tracking components for b-jets in the 60-70% working point interval of the pseudo-continuous b-tagging calibration [5]. The total uncertainty is the quadrature sum of the individual uncertainties. The extrapolation uncertainties have been integrated into the eigenvector decomposition method together with the in-situ calibration uncertainties. The outcome of this procedure is a set of eigenvectors with mixed contributions from both regimes, ready to be used in physics analyses.

[1]Please note that the statistical uncertainties in the high-p_T region are inherited from the statistical uncertainties of the highest p_T bin in the standard calibration region. Therefore in the current procedure, only the systematic variations are evaluated in the high p_T region.

[2]The quasi-stable uncertainty is introduced since the b hadron at high p_T might decay after traveling through the first few layers (30 ~ 100 mm from the collision point) of the pixel inner tracker, and the current simulation does not consider the interactions of charged B and D hadrons with the detector. The quasi-stable uncertainties only affect the b- and c-jets due to the B- and D-hadron's longer lifetime that allow them to decay also after the first layer

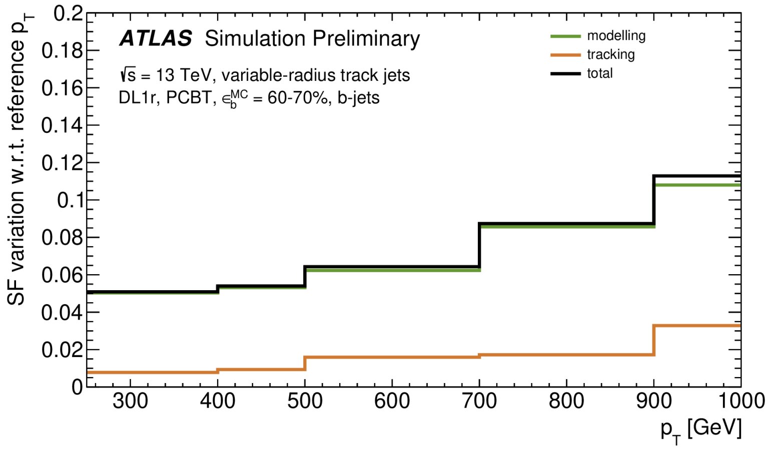

Figure 1: Monte Carlo based extrapolation uncertainties for the DL1r tagger applied to Variable Radius track jets as a function of the reconstructed b-jet pT (black curve). The contribution from experimental uncertainties affecting the reconstruction of tracks is shown in orange. Theoretical uncertainties on the modelling of the jet fragmentation are drawn in green. The total variation is obtained summing in quadrature the single contributions. More details are available in Ref. [5].

References

[1] ATLAS Collaboration. Eigenvector recomposition: a new method to correlate flavor-tagging systematic uncertainties across analyses. Technical report, CERN, Geneva, 2022.

[2] ATLAS Collaboration. Measurements of WH and ZH production in the $H \to b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector. *Eur. Phys. J. C*, 81(2):178, 2021.

[3] ATLAS Collaboration. Simulation-based extrapolation of b-tagging calibrations towards high transverse momenta in the ATLAS experiment. Technical report, CERN, Geneva, Mar 2021.

[4] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jets with Variable R. *JHEP*, 06:059, 2009.

[5] Yan Ke, Ilaria Luise, and Giacinto Piacquadio. Montecarlo-based extrapolation uncertainties for variable radius track jets with multiple b-tagging working points. Technical report, CERN, Geneva, 2022.