PROCEEDINGS OF SCIENCE

Calibration of Flavour Tagging Algorithms at the CMS Experiment - Run 3 Results and Innovations

Svenja Diekmann[a,*] **on behalf of the CMS Collaboration**

[a] *RWTH Aachen University,*
Templergraben 55, 52062 Aachen, Germany

E-mail: svenja.diekmann@cern.ch

The efficient identification of jets originating from heavy-flavour quarks is essential for many measurements in high-energy physics, such as Higgs boson and top quark measurements as well as searches for physics beyond the standard model of particle physics. Nevertheless, differences between measured data and the simulation on which the algorithms were trained can cause performance deficits and prediction discrepancies. For precision measurements and searches for new physics, however, these flavour tagging algorithms must be well-calibrated to avoid such problems by mitigating any discrepancies arising from the application of the flavour tagging algorithms. We will present an overview of the calibration methods and first Run 3 results for data recorded by the CMS experiment. Moreover, as the flavour tagging algorithms are continuously improved, the calibrations need to be refined accordingly to ensure reliable performance. These new, more powerful flavour tagging algorithms are particularly prone to capturing mismodelling in higher-order correlations of the simulation. The impact of the state-of-the-art algorithms on the calibration and the effect of the integrated mitigation strategy based on adversarial attacks in one of the algorithms will be demonstrated and compared.

12th Edition of the Large Hadron Collider Physics Conference (LHCP2024)
3-7 June 2024
Northeastern University, Boston, USA

*Speaker

PoS(LHCP2024)263

1. Introduction to Flavour Tagging Algorithms at the CMS Experiment

Flavour tagging has become essential to many measurements in high-energy physics, ranging from precision measurements to searches for new physics. However, to enable precise predictions for these analyses, the tagging algorithms need to be well understood. Especially the prediction differences between measured data and the simulation, on which these algorithms are trained, need to be corrected to ensure a good modelling of the flavour tagging discriminant in simulation.

Currently, three different flavour tagging algorithms are used at the CMS experiment [1, 2] in the LHC Run 3, namely, **DeepJet** [3], the Run 2 state-of-the-art algorithm, **ParticleNet** [4] and **robustParticleTransformer** (robustParT) [5–7], the improved algorithms for Run 3. At the CMS experiment, five predefined working points (WP) are derived based on the misidentification efficiency, for which efficiencies in simulations and data are measured. The loose (L), medium (M), and tight (T) WPs correspond to a light-flavour (udsg) jet misidentification efficiency of 10 %, 1 %, and 0.1 %, respectively. With the start of Run 3, two new WPs were introduced for an even purer selection of b jets, namely the extra-tight and extra-extra-tight WPs, corresponding to a misidentification efficiency of 0.05 % and 0.001 %, respectively.

In these proceedings, different methods to calibrate flavour tagging algorithms are presented, including efficiency measurement for the above-mentioned WPs in data and simulation and the continuous calibration of the discriminator shape. The calibrations are presented in two data-taking periods of the year 2022. The overview of the calibrations can be found in Ref. [8].

2. B-Jet Tagging Efficiency Measurements

The efficiency of b-tagging algorithms can be estimated from different phase space regions, which can be combined if the regions are mutually exclusive. Thus, for measurements performed in similar regions, the exclusive regions can still ensure a calibration without overlap in any measurements. Phase space regions particularly suited for the efficiency measurement of b jets are the top quark pair production due to the high abundance of b jets and the quantum chromodynamics (QCD) multijet production due to its high production cross section.

2.1 B-Tagging Efficiencies from Top Quark Pair Production

The b jets produced in the top quark decay are used to determine the b-tagging efficiencies. The different methods applied at the CMS experiment exploit two decay possibilities of the top quark pair and apply different kinematic reconstructions of the top and antitop quark system in order to be exclusive and independent.

The Kinematic Fit method measures the b-tagging efficiency in the top quark pair production phase space with leptonic decays of both W bosons from the top quark decay. A template fit is performed on the prediction of a multivariate discriminator, which combines multiple kinematic observables [9].

The Tag and Probe method measures the b-tagging efficiency in the top quark pair production phase space with leptonic decays of one of the two W bosons from the top quark decay. The method is based on a tag and probe selection, in which one jet is required to pass the M WP, and the other jet is used as a probe [9].

2.2 B-Tagging Efficiencies from Multijet Production

In the multijet production most jets are light-flavoured, but requiring a low momentum (soft) muon within the jet, originating from the decay of heavy mesons, results in an enrichment of b jets.

The P_T Rel. method measures the b-tagging efficiency in multijet events, which contain jets with a soft muon within the jet radius. The efficiency measurement is based on a fit of the transverse momentum spectrum of the muon within the jet relative to the jet axis to evaluate the number of jets passing or failing the WP studied [9].

The System8 method measures the b-tagging efficiency in b-enriched multijet events by solving a system of eight equations with eight unknown parameters, which include the b-tagging efficiency of the WP studied, in eight different regions in the selected data events [9].

2.3 Combination of B-Tagging Efficiency Measurements

The individual calibration measurements are combined using the Best Linear Unbiased Estimate (BLUE) method [10], in which the weighted average is determined, taking into account the correlation between the systematic uncertainties. In Figure 1, the fit to these weighted averages differential in the jet transverse momentum is displayed in addition to the individual measurements for the robustParticleTransformer algorithm and its tight WP.

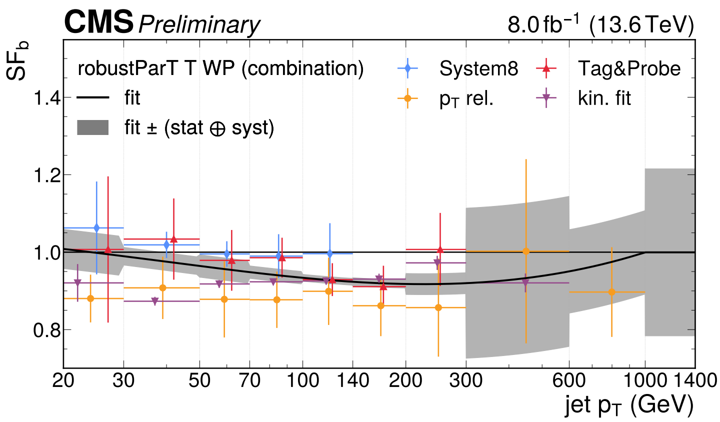

Figure 1: Measurements of b-tagging scale factors (SF_b) in different phase space regions differential in the jet transverse momentum (jet p_T). The individual calibrations are displayed by different coloured points with uncertainties, and the combination is displayed by the black line and the grey uncertainty band [8].

3. Light-Flavour Jet Tagging Efficiency Measurements

The misidentification efficiency of light-flavoured jets is measured in the multijet production by the negative tag method. In contrast to the b-jet efficiency measurements, the presence of a low momentum muon within the studied jet is not required.

The Negative Tag method measures the light-flavour jet mistagging efficiency in an inclusive multijet phase space. A negative tagging algorithm is defined by reversing the sign of the impact parameters of the tracks inside the jet. After its evaluation the WP criterion can be applied to measure the efficiency, similar as described in Ref. [9].

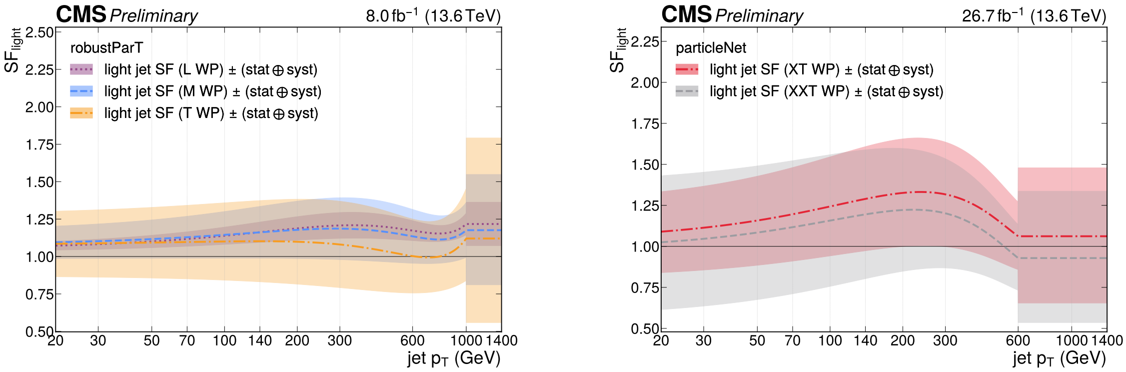

Figure 2: Measurements of light-flavour jet mistagging calibrations (SF_{light}) in the inclusive multijet production differential in the jet transverse momentum (jet p_T) for the robustParticleTransformer (left) and the ParticleNet (right) algorithms. The correction factors are displayed for the L, M, and T WPs displayed by the dotted, dashed, and dashed-dotted lines (left) and for the XT and XXT WPs by the dashed and dashed-dotted lines (right) including the corresponding uncertainty bands [8].

The resulting corrections for the ParticleNet and robustParticleTransformer algorithms are presented in Figure 2 and show a similar behaviour across the jet momentum spectrum and for the different WPs.

4. B-Tagging Discriminator Shape Calibration and Performance

If the full shape of the b-tagging discriminant is calibrated, it becomes accessible to machine learning algorithms to improve the sensitivity of physics analyses. However, if no fixed cut-off threshold is defined, the b-tagging efficiency cannot be estimated directly. Instead, the distribution is calibrated by a simultaneous measurement of the b and light-flavoured jet contributions in the heavy-flavour-enriched top quark pair production and the light-flavour-enriched Z boson production in association with jets. The fit is performed iteratively to deal with the contamination from other flavoured jets. These contaminations contributions are estimated from simulation, which is updated with the preliminary calibrations throughout the fitting procedure [9].

The application of the continuous calibration to top pair production phase space, in which the W bosons from the top quark decay into an electron and a muon and their corresponding neutrinos, is displayed in Figure 3. The calibrations for both presented algorithms show a good closure compared to the pre-calibration agreement between data and simulation.

The calibration of the full discriminator shape allows for the correction of the receiver operating characteristic (ROC) curve to estimate the tagging performance in measured data, as it can be seen in Figure 4. The newly established algorithms for Run 3 show a comparable performance over a large range of the b-jet identification efficiency. Only for a low light-flavour jet misidentification rate, the robustParticleTransformer algorithm achieves a higher b-jet identification efficiency.

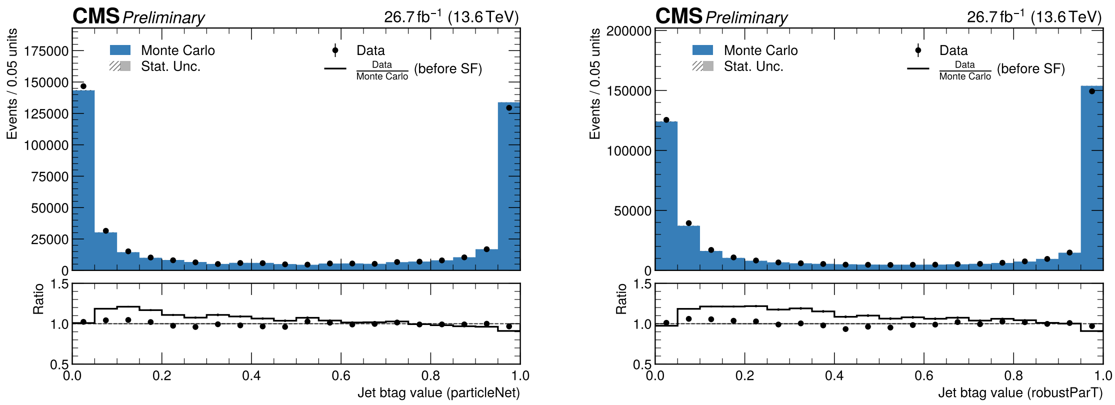

Figure 3: Distribution of the b-tagging discriminator score for the ParticleNet (left) and the robustParticle-Transformer (right) algorithms showing the agreement between data and simulation after the application of the continuous light and b-jet calibrations. In the lower panel, the ratio of data over simulation is shown by the black points and the line before and after the application, respectively [8].

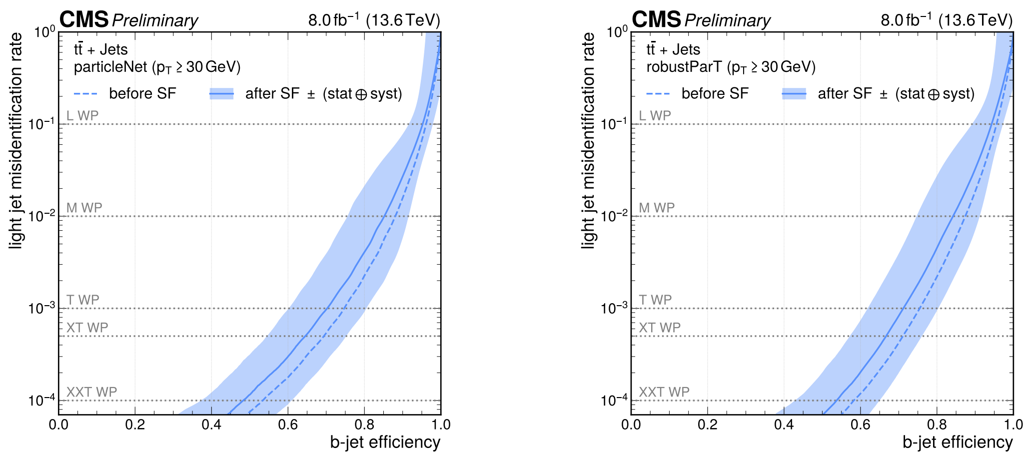

Figure 4: ROC curve of the ParticleNet (left) and the robustParticleTransformer (right) algorithms evaluated on top quark pair production simulation before and after the application of the full discriminator shape calibration displayed as the dashed and solid lines in combination with the blue uncertainty band [8].

5. Innovations in the Flavour Tagging Description and Calibrations

The recent developments in machine learning can not only be exploited in the flavour tagging algorithms themselves but also in their calibrations. In an adversarial approach, the calibration of the discriminator shape can be improved by an unbinned treatment of the discriminator values and the kinematics, in which the b-tagging corrections are differential. Additionally, it allows for higher-dimensional dependencies [11]. Furthermore, graph neural networks can be exploited to determine efficiency maps, which are applied instead of direct tagging [12].

6. Conclusion

In these proceedings, the first b-tagging calibrations for the data measured in the LHC Run 3 in 2022 are presented. The efficiency measurements are stable across different working points and tagging algorithms and perform well for the newly established working points, which are defined by lower light-flavour jet misidentification rates and thus correspond to higher b-jet identification efficiencies. The calibration of the full discriminator shape shows a good closure in its application, with which the b-jet identification performance on data can be estimated in order to perform a complete comparison of the algorithms.

References

[1] The CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004

[2] The CMS Collaboration, *Development of the CMS detector for the CERN LHC Run 3*, *JINST* **19** (2024) P05064

[3] E. Bols, J Kieseler, M. Verzetti, M. Stoye, and A. Stakia, *Jet Flavour Classification Using DeepJet*, *JINST* **15** (2020) P12012

[4] H. Qu, L. Gouskos, *ParticleNet: Jet Tagging via Particle Clouds*, *Phys. Rev. D* **101** (2020) 056019

[5] H. Qu, C. Li, and S. Qian, *Particle Transformer for Jet Tagging*, arXiv:2202.03772 (2022)

[6] The CMS Collaboration, *Transformer models for heavy flavor jet identification*, CMS Detector Performance Summary 2022

[7] A. Stein, X. Coubez, S. Mondal, A. Novak, and A. Schmidt, *Improving Robustness of Jet Tagging Algorithms with Adversarial Training*, *Comput. Softw. Big Sci.* **6** (2022) 15

[8] The CMS Collaboration, *Performance summary of AK4 jet b tagging with data from 2022 proton-proton collisions at 13.6 TeV with the CMS detector*, CMS Detector Performance Summary 2024

[9] The CMS Collaboration, *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV*, *JINST* **13** (2018) P05011

[10] L. Lyons, D. Gibaut, and P. Clifford, *How to combine correlated estimates of a single physical quantity*, *Nucl. Instrum. Meth. A* **270** (1988) 110

[11] The CMS Collaboration, *Adversarial Neural Network-based data-simulation corrections for heavy-flavour jet-tagging*, CMS Detector Performance Summary 2019

[12] The CMS Collaboration, *Efficiency parametrization of b-tagging classifier using Graph Neural Networks*, CMS Detector Performance Summary 2022

PoS(LHCP2024)263