CMS Physics Analysis Summary

Contact: cms-pog-conveners-jetmet@cern.ch 2016/01/24

Top Tagging with New Approaches

The CMS Collaboration

Abstract

Methods for boosted hadronically decaying top quark identification (top tagging) are tested in view of the 13 TeV data-taking run of the LHC. In addition to evaluating the discrimination power of single observables, the gain from combining different substructure techniques is quantified. The interplay between different top tagging variables and b tagging is evaluated and the performance of selected observables for high transverse momentum jets is studied. Aditionally, a comparison of top tagging variables between data and simulation is performed, using collision events with a center-of-mass energy of 8 TeV.

1 Introduction

Many models of physics beyond the Standard Model (SM) predict new resonances decaying to final states containing highly energetic top quarks. Examples include Z′ bosons arising in topcolor models [1–3], and Kaluza-Klein gluons predicted by Randall-Sundrum models [4] of extra dimensions. Both the ATLAS [5–7] and the CMS [8, 9] collaborations have previously published searches for these particles based on 7 and 8 TeV data, where no significant deviations from SM predictions have been observed.

Furthermore, the center-of-mass energy of 13 TeV achieved by the LHC for the data taking period from 2015 onwards implies a large number of energetic top quarks originating from SM processes such as single top and top pair production. This makes the reconstruction and identification of boosted hadronically decaying top quarks an important technique for upcoming analyses of LHC data.

Considering a top quark with a Lorentz boost $\gamma = E/m$, the distance between its daughter W boson and b quark will be approximately $\Delta R = 2/\gamma$. If the W boson decays hadronically, the resulting jets, two from the W boson decay and one from the b quark hadronization, will often be contained in an angular region smaller than $2/\gamma$. Thus, a jet clustering algorithm with a distance parameter R larger than $2/\gamma$ will typically collect all the decay products into a single jet. This resulting object is referred to as a "top jet".

Different techniques, such as groomed jet masses, substructure observables, and combined taggers, have been considered to identify heavy objects which fragment into three prongs, i.e. top quark decays. Grooming methods attempt to remove soft and wide angle radiation inside a jet, produced by pileup interactions, underlying events, or parton shower activity. In this work, filtering [10], trimming [11], pruning [12], and softdrop [13] methods have been considered. Substructure observables are quantities, other than jet masses, used to distinguish top jets from the hadronization products of single light quarks or gluons. Examples of substructure observables include Qjets [14], which measures the jet mass stability under different randomized jet clustering histories, n-subjettiness [15], sensitive to the number of subjets inside the jet, and the k_{T} splitting scale ($\sqrt{d_{12}}$) [16]. Combined taggers attempt a complete reconstruction of the top decay kinematics, either by providing a set of variables such as the CMS Top Tagger (CMSTT) [17, 18], the HEPTopTagger [19, 20] (HTT) and HEPTopTagger Version 2 [21] (HTT V2), or by condensing this information into a single likelihood-based variable, as in shower deconstruction [22, 23].

These algorithms are briefly outlined in this document, followed by a study of their performance. The stability of different selection criteria is evaluated as a function of p_{T}, η, and number of pileup vertices. In addition, mistag rates for different parton flavors are studied to understand the flavor dependence of tagging efficiency. Finally, the performance of new techniques is commissioned with the 8 TeV LHC dataset and compared to simulated events.

This document builds on the the work documented in [24], including also the novel softdrop, HTT V2, and shower deconstruction methods. Combinations of algorithms and investigations of kinematic dependencies are used to set recommended working points for LHC Run II analyses.

2 Simulated Samples, Object Definitions and Event Selection

Top quark identification techniques are first studied using simulated events at a center of mass energy of 13 TeV. Then, the performance of these algorithms is validated using data recorded

at a center of mass energy of 8 TeV. For these purposes, two different sets of selection criteria and simulated events have been considered.

The performance of top-tagging techniques is evaluated over a broad range of jet transverse momentum (p_T), using events simulated at $\sqrt{s} = 13$ TeV. An enriched signal sample of high p_T hadronically decaying top quarks is obtained from the decay of heavy Z' resonances simulated with MADGRAPH 5 [25] and interfaced with PYTHIA 8.2 [26] for the hadronization. For the Z' production the NNPDF3.0LO [27] PDF set is used. These resonances are narrow, having intrinsic widths equal to 1% of the resonance mass. The background sample is represented by jets produced by the fragmentation of quarks and gluons in QCD multijet events, generated using PYTHIA 8.2 in exclusive p_T bins of the leading quark or gluon. For background events NNPDF2.3LO [28] PDF set is used.

For the comparison of data to simulation at 8 TeV, top quark pair events are generated with the next-to-leading-order generator POWHEG 1.0 [29–32] interfaced to PYTHIA 6.4 [33] for the showering. Alternatively, results for top quark pair production are obtained using the MC@NLO 3.4 [34] generator interfaced with HERWIG 6.5 [35] for showering and hadronization. MADGRAPH 5 and PYTHIA 6 are matched through the MLM algorithm [36] to produce W and Z bosons in association with jets, where up to 4 additional partons are simulated at matrix element level. The production of single top quarks is simulated using POWHEG. The misidentification rate is measured in QCD multijet events, where two different simulations are adopted. One sample is generated using MADGRAPH 5 for the hard scattering, which is showered and hadronized with PYTHIA 8.1 [37]. For the second one, generation and showering are both performed with HERWIG++ [38]. The MADGRAPH samples use the CTEQ6L [39] parton distribution function (PDF). The $t\bar{t}$ production simulated with POWHEG uses CT10 [40] PDF, while the single top process has been generated using the CTEQ6M [39] PDF. Finally, the QCD PYTHIA sample uses the CTEQ61L PDF set, while the alternate HERWIG++ sample adopts MRST2001 [41].

All simulated events include the activity of additional inelastic proton-proton interactions within the same bunch crossing ("in-time pileup") and the simulation of additional contributions in the signal readout from the previous and subsequent bunch crossings ("out-of-time pileup") which in 2012 were 50 ns apart from the main bunch crossing. For the 13 TeV sample a separation of 25 ns is considered. To properly compare data with simulation, simulated events have been reweighted to reproduce the pileup distribution as observed in data in studies at $\sqrt{s} = 8$ TeV.

A data sample corresponding to 19.7 fb^{-1} of integrated luminosity is considered, collected by the CMS experiment at $\sqrt{s} = 8$ TeV. The sample was recorded by a single muon trigger, selecting events with a least one reconstructed muon with $p_T > 40$ GeV without any isolation requirement. In addition, a trigger based on the scalar sum of the transverse momenta of all jets in the event is used to obtain a QCD multijet enriched sample.

Jets are clustered starting from particles reconstructed by the Particle Flow (PF) algorithm [42], where charged hadron subtraction [43] is used to mitigate in-time pileup. Then, selected leptons are removed before reconstructing jets.

The Cambridge/Aachen (CA) clustering algorithm [44] is used to find jets at low p_T while, for high p_T, the anti-k_t (AK) algorithm [45] is employed. These choices were made for consistency with respect to established techniques [24] for boosted W and Higgs boson reconstruction [46] within CMS. Two different jet cone sizes are used for jet reconstruction: $R = 1.5$ (CA15) for jets with low p_T and $R = 0.8$ (AK8) for high p_T ones.

For the 13 TeV studies, all distributions of top quark candidate quantities require a *truth-level*

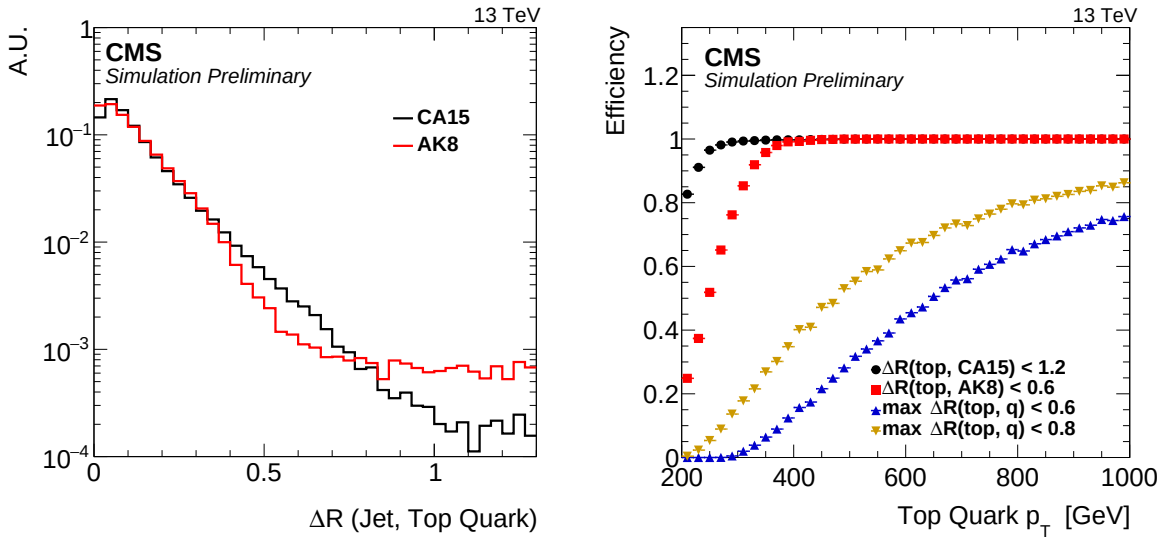

Figure 1: Distance between reconstructed jet and closest generated top quark (left). Jet reconstruction efficiency (right) as function of the generated top quark p_{T}. The efficiency is defined as the fraction of top quarks for which a reconstructed jet with $p_{\mathrm{T}} > 200$ GeV can be found within $\Delta R < 1.2$ ($\Delta R < 0.6$) for CA15 (AK8) jets. Superimposed is the fraction of merged top quarks as function of p_{T} for the two thresholds used: 0.8 (0.6) at low (high) boost. All distributions are made using hadronically decaying top quarks with $p_{\mathrm{T}} > 200$ GeV.

object close to the reconstructed jet. The truth-level matching is done using top quarks before the decay, in the case of signal jets, and partons (u, d, s, c, b and gluon) from the hard scattering, in the case of background jets.

Matching particles to the generator-level is performed by selecting the closest reconstructed jet in ΔR to a truth-level parton or top quark. The maximum distance depends on the jet size and corresponds to $\Delta R < 1.2$ ($\Delta R < 0.6$) for $R = 1.5$ ($R = 0.8$) jets. These values approximately correspond to the minima of the ΔR distribution between jets and truth-level top quarks, as shown in Fig. 1 (left).

The jet cone sizes are at typical values used in top quark reconstruction [24]. Fig. 1 (right) shows the efficiency with which a reconstructed jet with $p_{\mathrm{T}} > 200$ GeV is found within $\Delta R < 1.2$ ($\Delta R < 0.6$), for CA15 (AK8) jets, with respect to a hadronically decaying top quark as function of the quark p_{T}. The efficiency reaches a plateau close to 100% at about 280 GeV (380 GeV) for jets with $R = 1.5$ ($R = 0.8$). In fact, at low p_{T} the larger cone size allows for the collection of all decay products from the top quark. In contrast, at high p_{T} the top jet is more collimated and the smaller cone size reduces the amount of additional radiation clustered into the jet.

To evaluate the properties of top-tagging methods in an unbiased way over a large phase space, taggers are studied in exclusive regions of p_{T}, defined by the transverse momentum of the matched truth-level parton (top quark, light quark, or gluon). Then, a weight is assigned to each truth-level object such that the resulting distributions in p_{T}- and η-space are approximately flat. As the distribution of top-jet candidates becomes more central with higher p_{T} values, we tighten different selection criteria on $|\eta|$ as the p_{T} increases. Finally, to avoid the study of incompletely merged top quarks (i.e. top quarks for which not all decay products are contained in a jet), we restrict our selection to top quarks where the two quarks from the W boson decay and the b quark have a maximum distance from the top quark direction $(\max(\Delta R(\mathrm{t}, \mathrm{q})))$

Table 1: Overview of signal and background samples and selection criteria used for studies of top-tagging performance. R denotes the distance parameter used for jet reconstruction. Background samples are denoted by the p_T bin whereas signal samples are denoted by the mass of the Z′ resonance. For parton to jet matching, labeled $\Delta R(\mathrm{p,jet})$, where p stands for the top quark in the case of signal jets and light quarks, and partons from the hard scattering in the case of background ones. The merged top requirement ($\max(\Delta R(\mathrm{t,q}))$) restricts the maximal distance between the top quark (t) and the three decay products (q).

p_T range [GeV]			Max. $\lvert\eta\rvert$	QCD			$m_{Z'}$ [TeV]	R	$\Delta R(\mathrm{p,jet})$	$\max(\Delta R(\mathrm{t,q}))$
200	–	300	2.4	170	–	300	1.0	1.5	1.2	0.8
300	–	470	2.4	300	–	470	1.0	1.5	1.2	0.8
470	–	600	2.1	470	–	600	1.25	0.8	0.6	0.6
600	–	800	2.1	600	–	800	2.0	0.8	0.6	0.6
800	–	1000	1.5	800	–	1000	2.0	0.8	0.6	0.6
1000	–	1400	1.5	1000	–	1400	3.0	0.8	0.6	0.6

smaller than 0.8 (0.6) for $p_T < 470$ GeV ($p_T > 470$ GeV). The fraction of top quarks passing this requirement is shown in Fig. 1 (right). This requirement is only enforced for the signal sample and for some distributions reported in Section 5. An overview of the 13 TeV samples and selections is given in Table 1.

For the 8 TeV sample, we also define electron and muon selection criteria. Electron candidates with $p_T > 35$ GeV and $|\eta| < 2.5$ are identified via an MVA based criterion [47]. Muon candidates are selected using criteria [48] based on the track fit parameters and the number of muon, pixel, and tracking detector hits, allowing an average efficiency of at least 96% for prompt muons. Only muon candidates with $p_T > 45$ GeV and $|\eta| < 2.1$ are considered. For both electrons and muons, no isolation requirement is applied.

The missing transverse energy in the event is measured as the negative vector sum of the reconstructed PF-candidate momenta, as defined in [49]. In addition, jet energy calibrations are used to correct the PF missing energy estimate [50, 51].

3 Top Tagging Methods

3.1 Groomed Masses

The primary observable to identify top quarks in a background of light quark and gluon jets is the large invariant mass of the jet itself. This quantity is calculated as the invariant mass of the sum of constituent four vectors, which are assumed to be light compared to the top quark. Due to contributions from initial state radiation, the underlying event, and multiple hadron scattering (pileup), the reconstructed jet mass can be far higher than that of the initial parton, which should be close to zero for jets initiated by light quarks or gluons. This effect is exacerbated when using a large cone size for jet reconstruction as the amount of pileup is proportional to the jet area.

Methods which remove these additional contributions from jets, called grooming, allow for major improvements in top jet reconstruction and identification. Initial developements include filtering [10] ($m_{\mathrm{Filt.}}$) and trimming [11] ($m_{\mathrm{Tr.}}$). Both methods start by re-clustering the constituents of the initial jet with a smaller distance parameter r, which is set to 0.2 in the studies presented in this note. In the case of filtering, the resulting jets (called subjets) are sorted by descending

p_T and only the leading n subjets are considered. In the case of trimming, all resulting subjets with $p_{T,subjet} > f \cdot p_{T,jet}$ are retained, where f is a constant parameter. Finally, one reconstructs the top quark candidate as the four-vector sum of the remaining subjets.

In contrast, pruning [12] ($m_{\text{Pr.}}$) removes single constituents in multiple clustering stages. Starting with all the individual constituents of an initial jet, the clustering procedure is repeated while two conditions are evaluated for each recombination of two objects i and j into p:

$$z = \frac{\min(p_{Ti}, p_{Tj})}{p_{Tp}} < z_{cut} \text{ and } \Delta R_{ij} > r_{cut}. \tag{1}$$

Here p_{Tx} denotes the transverse momentum of object x, ΔR_{ij} the angular distance between objects i and j, while z_{cut} and r_{cut} are parameters chosen to suppress soft and wide angle radiation, respectively. If both conditions are fulfilled, the two branches are merged, otherwise, the lower-p_T jet is discarded. The clustering algorithm then continues for the remaining constituents of the jet.

Finally, softdrop [13] ($m_{\text{SD.}}$) declustering reduces the dependency of groomed variables on non-global logarithms. The clustering of a jet with distance parameter R is reverted step by step, breaking the jet j in two subjets j_1 and j_2 at each iteration. If the softdrop condition

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut} \cdot \left(\frac{\Delta R_{12}}{R}\right)^{\beta} \tag{2}$$

holds then j is considered the final jet and the procedure stops. Otherwise, the same strategy is repeated by relabelling the higher p_T subjet as j and discarding the lower p_T one. The softdrop algorithm has two free parameters: z_{cut} determines the strength of the fractional p_T selection, while β can be set to values greater than zero to soften the selection for collinear radiation.

3.2 N-Subjettiness

Another feature useful for seperating signal from background events is the energy distribution inside the jet. Jets resulting from hadronic decays of top quarks are expected to have three regions where most of the energy is deposited, which correspond to the three partons resulting from the top quark decay. In contrast, jets arising from the hadronization of light quarks or gluons are expected to only have one or two such regions.

The n-subjettiness [15] τ_i measures the compatibility of a jet with the hypothesis that it is composed of i subjets. It is defined as:

$$\tau_i = \frac{1}{\sum_k p_{T,k} R} \sum_k p_{T,k} \min\left(\Delta R_{1k}, \Delta R_{2k}, \cdots, \Delta R_{ik}\right), \tag{3}$$

where the index k enumerates the constituents of the input jet, $p_{T,k}$ is the transverse momentum of the k-th constituent, R denotes the distance parameter used for the initial jet reconstruction, and ΔR_{ik} is the angular distance between the i-th subjet and the k-th constituent. The subjets are reconstructed using the one-pass k_t axes approach.

The value of τ_i tends to zero in the case of exactly i subjets and to one if there are less than i subjets. For top quark decays, a topology with three subjets is expected while jets from light quarks or gluons will have fewer subjets. This makes the ratio τ_3/τ_2 a robust variable for top quark identification.

3.3 Qjet volatility

The methods described so far are either sensitive to the mass of the particle that intiated the jet (groomed masses) or to the number of centers of radiation inside the jet (n-subjettiness). Qjet volatility [14] is a complementary approach which evaluates the stability of the jet mass when the input jet is reclustered multiple times (in this study 50 re-clusterings are calculated per jet) while introducing a random element to reweight the clustering distance at each step.

The weight function is defined as:

$$\omega_{ij}(\alpha) = \exp\left(-\alpha \frac{(d_{ij} - d^{min})}{d^{min}}\right). \tag{4}$$

Here d_{ij} is the standard CA or k_T jet clustering distance between the objects i and j, d^{min} is the minimum of all pair-wise distances at the current clustering step, and α denotes a free parameter, called rigidity. The probability to choose a specific clustering is then set to $\Omega_{ij} = \frac{\omega_{ij}}{\sum_{ij}\omega_{ij}}$ and a random number is used to select which clustering to perform according to this probability distribution.

The degree of randomness is controlled by the rigidity parameter, where $\alpha = 0$ corresponds to a completely random choice, while $\alpha \to \infty$ corresponds to a deterministic jet clustering. This study uses a value of $\alpha = 0.1$.

The observable used for discriminating top jets from QCD ones is the Qjet volatility, defined as:

$$V = \frac{\sqrt{\mathrm{Var}(m)}}{\langle m \rangle} \tag{5}$$

where $\langle m \rangle$ denotes the mean and $\mathrm{Var}(m)$ the variance of the mass distribution.

3.4 b-quark Tagging

The presence of B-hadron decay products inside a jet is a powerful indicator that distinguishes top jets from the fragmentation of light quarks or gluons. To this end, the impact parameter significance of charged-particle tracks as well as the presence and properties of reconstructed decay vertices are used. In particular, a multivariate discriminator called Combined Secondary Vertex (CSV) [52] is employed. Unless otherwise noted, softdropped subjets are used as inputs to b-tagging algorithms.

3.5 CMS Top Tagger

The CMS top tagging algorithm (CMSTT) [17, 18, 24] operates in two steps: first the initial jet is split into two subjets (primary decomposition), then an attempt is made to further split each subjet (secondary decomposition).

The splitting is done by going through the pair-wise clustering sequence used to find the initial jet in reverse order. Selection criteria on the transverse momenta and angular distances of the subjets are used to veto too close and too low-p_{T} unclusterings.

The discriminating variables provided by the CMS Top Tagger are the number of subjets $N_{subjets}$, the mass of the whole jet m_{jet}, and the minimum pairwise mass m_{min}, defined as the minimum invariant mass of all the pairwise combinations of the three leading subjets.

3.6 HEPTopTagger

The HTT [19–21] version 2 (HTT V2) is another multi-stage top-tagging algorithm. In its first step, the clustering history of a CA15 jet is undone step-by-step. If, after the unclustering of jet j into subjets j_1 and j_2 (where $m_{j1} > m_{j2}$), the masses of the jet fulfill the condition $m_{j1} < 0.8m_j$, both subjets are analysed further, otherwise j_2 is discarded. The unclustering continues until all subjets have a mass below 30 GeV. If less than three subjets are found the candidate is discarded, otherwise the triplet of subjets with highest p_{T} is selected for further considerations.[1] In fact, the constituents of this triplet are then passed to a filtering step with $R_{filt} = \min(0.3, \Delta R_{jk}/2)$ and the leading five filtered subjets are kept. Finally, their consistuents are re-clustered into exactly three subjets, discarding candidates where less than three subjets with $p_{\mathrm{T}} > 30$ GeV are obtained. These subjets correspond to the partons from the top quark decay.

Finally, the discriminating observables are represented by the top quark mass m_{123}, defined as the invariant mass calculated from the four-vector sum of the filtered subjets, and the ratio f_{Rec} between the reconstructed W boson and top quark masses. This variable is defined as:

$$f_{Rec} = \min_{ij} \left| \frac{\frac{m_{ij}}{m_{123}}}{\frac{m_W}{m_t}} - 1 \right|. \tag{6}$$

Here m_{ij} denotes the invariant mass of the pair of subjets i and j, m_W the mass of the W boson, and m_t the mass of the top quark.

In addition, HTT V2 introduces the optimal-R approach, implementing the idea that there is an optimal cone size for reconstructing top quarks as a function of p_{T}. Starting from $R = 1.5$, subsequently smaller values of the jet size, down to $R = 0.5$, are tested by applying the procedure described above. The final candidate is reconstructed with the smallest possible R choice (named R_{opt}) so that the reconstructed mass is less than 20 % away from the value found when using $R = 1.5$.

Thus, another discrimination variable $\Delta R_{opt} = R_{opt} - R_{opt}^{calc}$ is introduced as the difference between the optimal radius R_{opt} and its expected estimation R_{opt}^{calc}. The value of R_{opt}^{calc} is determined using a sample of pure top-jets as a function of the jet p_{T}, obtained after applying a filtering step with $R_{filt} = 0.2$ and $N_{max} = 10$.

3.7 Shower Deconstruction

Shower deconstruction [22, 23] is a full information approach to top tagging, defined as the ratio of the likelihood that the jet originates from a hadronic top quark decay to the likelihood that the jet is produced by a single parton.

In the first step, the initial jet is reclustered using the k_T algorithm with a smaller distance parameter, *MJ R*, into up to nine so-called microjets. We use *MJ R* = 0.2 at low p_{T} and *MJ R* = 0.1 at high p_{T}. The resulting ensemble of four-momenta is denoted by $\{p\}_N = \{p_1, p_2, \ldots, p_N\}$. Then, the signal $P(\{p\}_N|S)$ and the background likelihood $P(\{p\}_N|B)$ are estimated by calculating the probabilities that a simplified parton shower Monte Carlo would generate $\{p\}_N$ according to the signal or background hypothesis. This takes into account both the color flow

[1] In the initial version of the HTT the resulting triplet with a mass closest to true top mass was selected but HTT V2 uses the transverse momentum based criterion to avoid a bias in the mass distribution for background samples.

information in the event and the Sudakov factors. This variable is defined as:

$$\chi = \frac{P(\{p\}_N|S)}{P(\{p\}_N|B)} \tag{7}$$

By construction, high values of χ correspond to jets likely originating from top quarks, whereas low values of χ are observed for light quark and gluon initiated jets.

4 Distribution of Top Tagging Discriminating Variables

In this section, the fiducial p_T and η selection criteria for truth-level top quarks and background partons, described in Section 2, are applied to all distributions. Furthermore, the weights leading to flat p_T and η distributions are applied. In every plot, the number of entries of each histogram are reported inside the legend, evaluated with respect to the fiducial selection. This efficiency is the product of truth-level matching efficiency, jet reconstruction efficiency, boosted algorithm efficiency, preselection efficiency, and the probability of finding a value of the discrimination variable within the chosen axis range.

4.1 Masses

The mass of the CA (AK) jet after pruning is presented in Fig. 2 (top row). In all the explored p_T bins, except the lowest one, a clear peak is visible around the top mass value. The tail towards lower values and the second peak at around the W boson mass are predominantly caused by not-fully-merged tops, where one parton is outside the jet. All the background distributions are peaking below 30 GeV. Both signal and background distributions are stable across the different high p_T bins.

Overall, the effect of softdrop grooming on the mass distributions, presented in Fig. 2 (center and bottom), mimics what is observed for pruned jets. The mass is pushed to even lower values for background jets while a clear top mass peak is observed for the high p_T signal samples. Signal jets show an additional peak near the W boson mass, which occurs when the top quark's decay products are not fully contained in the jet. It vanishes when fully merged top-quarks are required. At low p_T the fraction of jets that passes the softdrop procedure can be increased by simulatenously increasing the z and β parameters. However, using a β-value larger than zero has the side-effect of shifting the background jet distribution to higher values, as well as altering the shape. The groomed mass has an intrinsic efficiency equal to one, thus the differences from one reported in Fig. 2 are due to the jet reconstruction efficiency and the mass lying outside the shown axis range.

4.2 Shape Observables

The ungroomed τ_3/τ_2 distribution is reported in Fig. 3 (top). As expected, the distribution is centered at lower values for top jets than for background ones. At low p_T, where CA15 jets are considered, the discrimination power is lower than at high p_T as the value of the variable for top jets is shifted upwards, while the background shape shifts downwards. However, once the jet cone size is fixed, the n-subjettiness is relatively stable with respect to p_T. The fraction of not-fully-merged top candidates inside the AK8 cone, which typically exhibit a two prong structure, introduces a peak in the τ_3/τ_2 distribution close to the one expected for background jets, that can be reduced by a top mass selection ($110 < m_{SD.} < 210$ GeV). The recovery of these partially merged candidates should be investigated further.

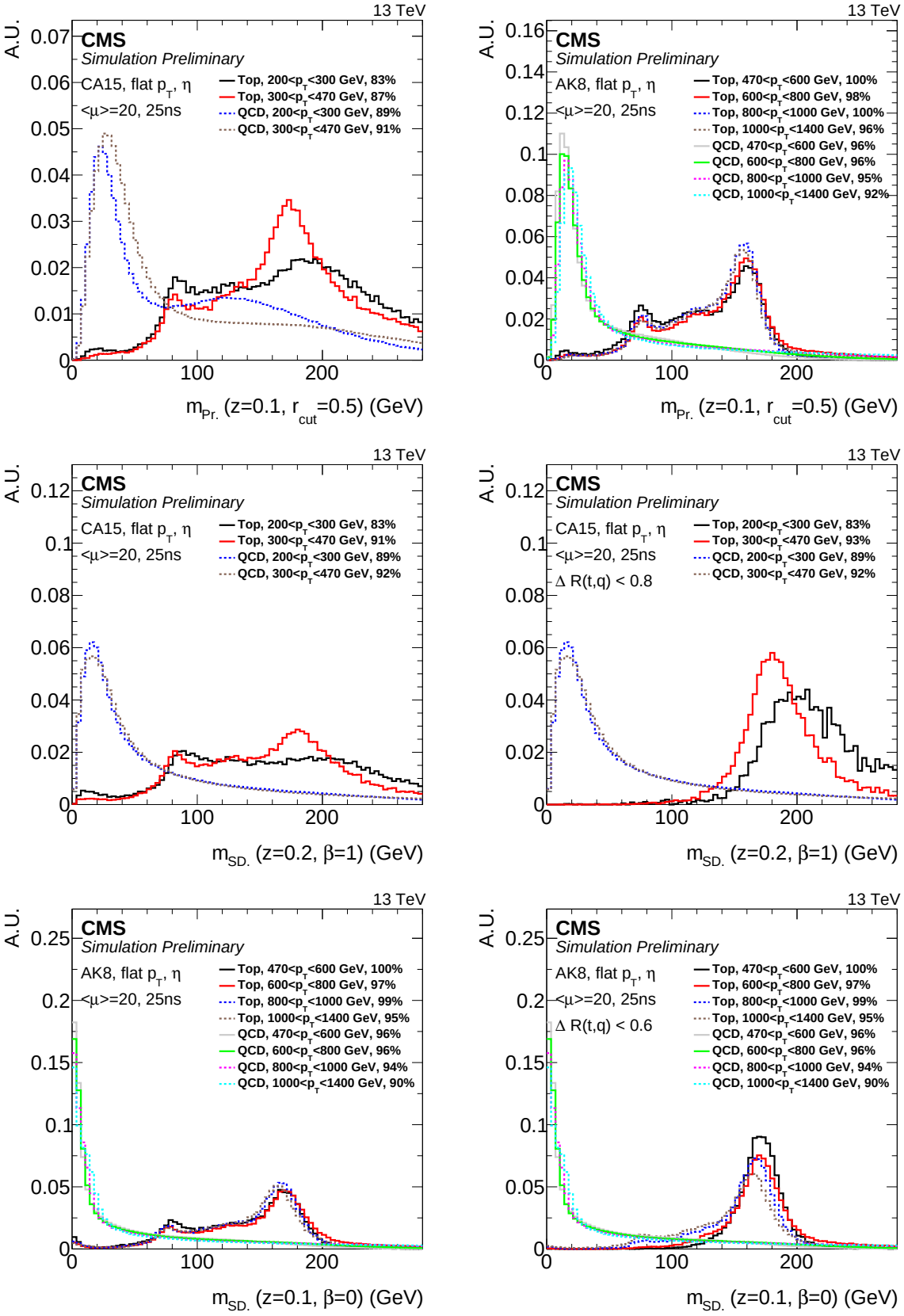

Figure 2: Distribution of pruned mass (top), softdrop ($z = 0.2$, $\beta = 1$) mass for low p_{T} jets reconstructed using CA15 (center), and softdrop ($z = 0.1$, $\beta = 0$) mass (bottom) for high p_{T} jets reconstructed using AK8. The softdrop mass distributions are shown for a fiducial selection without (left) and with (right) the merged-top requirement. The percentage in the legend indicates the fraction of entries shown in the plot with respect to the fiducial selection. Events correspond to an average number of $\langle\mu\rangle = 20$ pileup interactions and a bunch spacing of 25 ns.

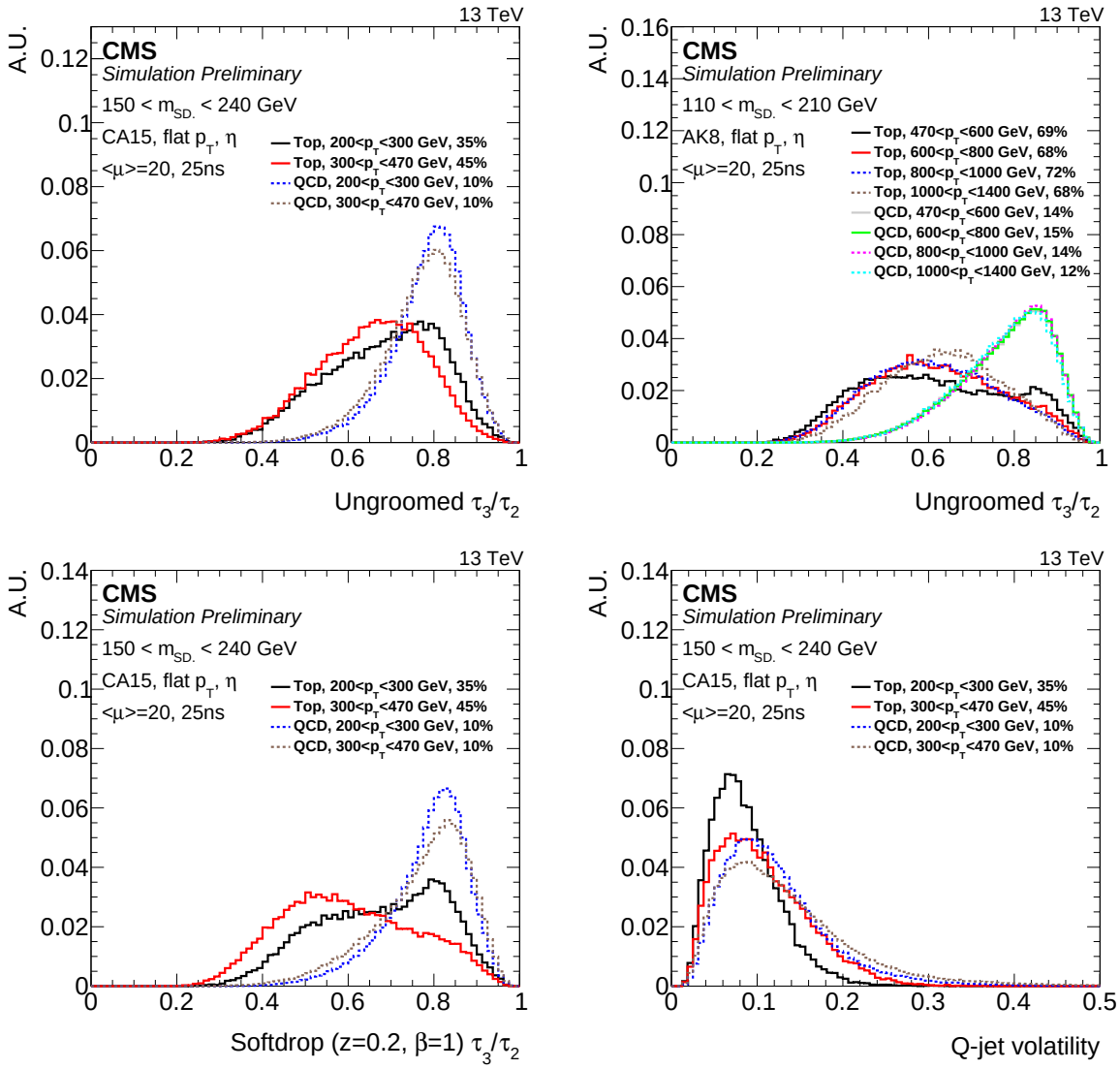

Figure 3: Distribution of ungroomed n-subjettiness (top) at low p_T (left) and high p_T (right). In addition, the softdrop n-subjettiness (bottom left) and the Qjet volatility (bottom right) are shown for low p_T jets clustered using CA15 jets. All distributions are shown after selecting on the jet mass. The percentage in the legend indicates the fraction of entries shown in the plot with respect to the fiducial selection. Events correspond to an average number of $\langle\mu\rangle = 20$ pileup interactions and a bunch spacing of 25 ns.

Applying softdrop grooming before calculating the n-subjettiness clearly improves the discrimination power for lower p_T jets, especially for top quarks with a p_T around 400 GeV, as shown in Fig. 3 (bottom left). At the same time, for AK8 jets, the groomed n-subjettiness shows a more stable performance as a function of the jet p_T with respect to ungroomed one.

Finally, the Qjet volatility exhibits lower values for true top quarks, where the decay of a heavy particle is responsible for the jet mass, than for backgrounds, where the clustering of radiation into the jet dominates. However, after requiring a soft drop mass between 150 and 240 GeV (Fig. 3 bottom right), most of the separation power disappears. The deviation from one in the efficiency reported in Fig. 3 is dominated by the applied softdrop mass selection.

4.3 Combined tagger variables

The three most important HTT decision variables are shown in Fig. 4. The candidate mass peaks around 150 GeV for top jets while at about 40 GeV for background ones. This method removes automatically not-fully-merged top candidates. In fact, only a small bump around the W boson mass is present after applying the algorithm, while background jets show a smoothly falling tail towards higher values with a bump at around 110 GeV for the lowest p_{T} bin. The top candidate mass is very stable with respect to p_{T}, as is shown in Fig. 4 (top). The efficiency of the HTT algorithm is far from one, driven by intermediate stopping conditions such as rejecting jets where less than three subjets can be reconstructed. This is strongly related to the fraction of not fully merged top quarks, as shown in Fig. 1 (left).

The main difference in the f_{Rec} (center) distribution between signal and background events is the sharpness of the peak around zero. Signal jets fall off more rapidly towards higher values. Similarly, the peak of the ΔR_{opt} distribution (bottom) at zero sharpens with increasing p_{T} both for signal and background samples. Among the HTT decision variables, ΔR_{opt} is the least discriminating.

Finally, Fig. 5 shows the shower deconstruction (χ) distribution for both top and background jets, where signal jets peak at a value of $\log(\chi)$ around 5, while backgrounds are centered at $\log(\chi)$ around 0. The efficiency is lower than one, caused by jets for which the discriminator cannot be calculated. This occurs when the algorithm fails to find a shower history matching the signal or background hypothesis within the given constraints, e.g. the top quark and W boson mass windows.

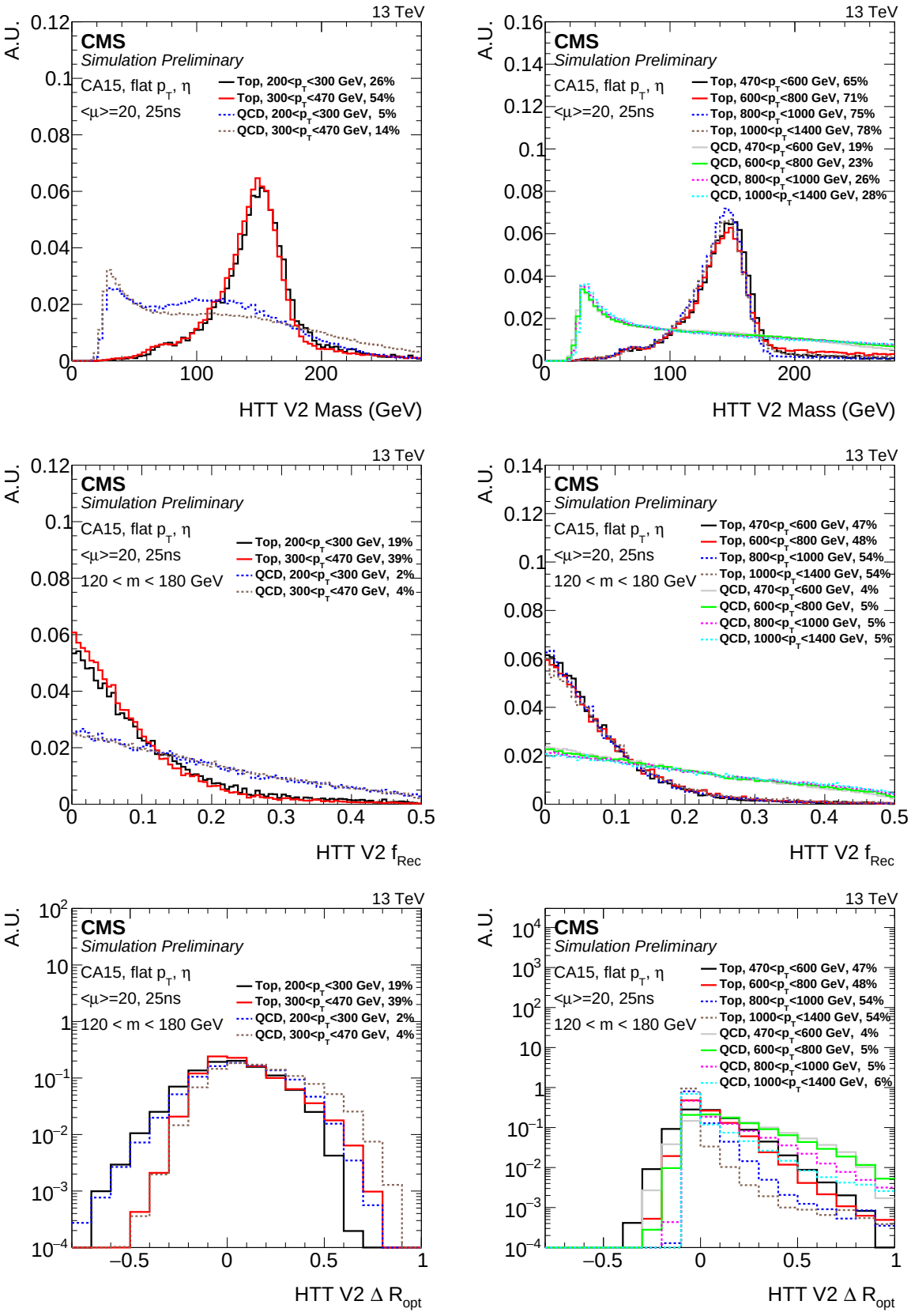

Figure 4: Distribution of the HTT V2 candidate mass (top), f_{Rec} (center) and ΔR_{opt} (bottom) for low p_{T} jets (left) and high p_{T} jets (right) reconstructed using CA15 jets. The percentage in the legend indicates the fraction of entries shown in the plot with respect to the fiducial selection. Events correspond to an average number of $\langle\mu\rangle = 20$ pileup interactions and a bunch spacing of 25 ns.

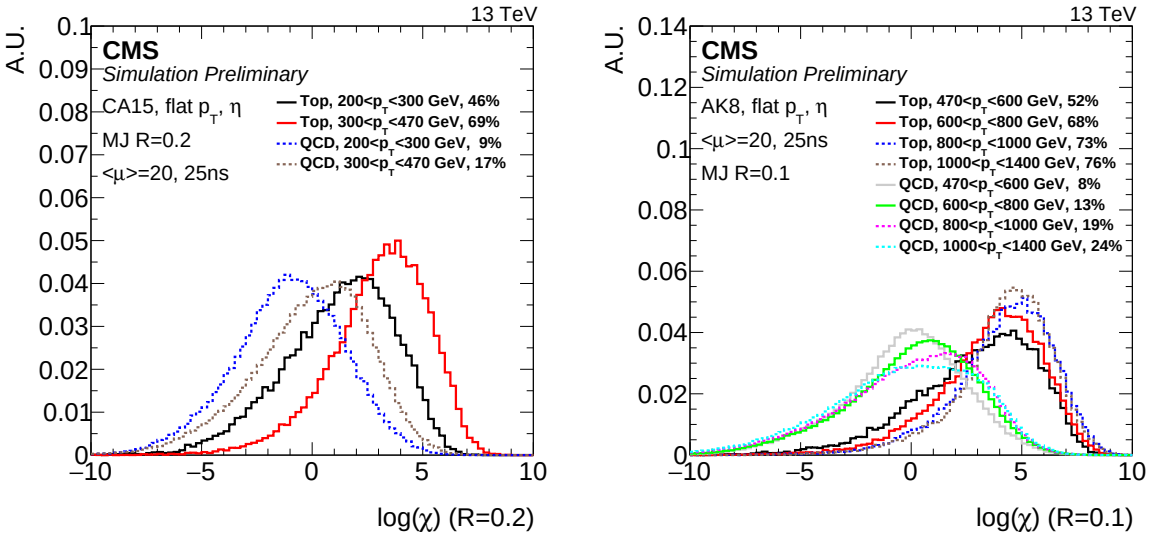

Figure 5: Distribution of the shower deconstruction discriminator for low (high) p_{T} jets reconstructed using CA15 (AK8) on the left (right). For the low (high) p_{T} jets a microjet distance parameter of 0.2 (0.1) is used. No selection on the jet mass is applied. The percentage in the legend indicates the fraction of entries shown in the plot with respect to the fiducial selection. Events correspond to an average number of $\langle\mu\rangle = 20$ pileup interactions and a bunch spacing of 25 ns.

5 Top-tagger Performance

To gain an initial understanding of the tagging performance of different approaches, we analyze the single variable receiver operating characteristic (ROC) curves shown in Fig. 6 (left), evaluated through the TMVA [53] package, where each point on the curve corresponds to a single selection window for the considered observable. The signal efficiency ε_S is shown on the horizontal axis and the background mistag rate ε_B on the vertical one. The efficiency is defined with respect to the total number, after reweighting, of truth-level top quarks and background partons passing the fiducial p_{T}, η, and merged top requirements described in Section 2. Results are shown for a representative highly boosted region, where partons have a transverse momentum between 800 and 1000 GeV.

Shower deconstruction provides the strongest discrimination power, outperforming the next strongest variable by at least a factor of three over large parts of the phase space. In both cases, the ungroomed n-subjettiness ratio comes second at low signal efficiency and the mass calculated using the HTT V2 algorithm for higher ones. Both pruning and softdrop mass variations perform similarly.

To investigate the improved discrimination power from pairs of variables a z-score diagram is prepared. The z-score is defined as the inverse of the background mistag rate ($1/\varepsilon_B$) at a signal efficiency of 30%. For the diagonal entries, corresponding to the single variable performance, a projective likelihood estimator is used. For the off-diagonal ones, corresponding to pairs of variables, a Boosted Decision Tree is trained.

The diagonal of the z-score diagram Fig. 6 (right) can be compared to the ROC curves in Fig. 6 (left). At a signal efficiency of 30%, shower deconstruction (χ) shows a background rejection of approximately 0.25% (a z-score of 385), followed by the mass reconstructed by HTT V2 and CMSTT as well as τ_3/τ_2. When pairs of variables are considered, combining χ with groomed masses improves the z-score by approximately 50%, while combining it with the subjet b tag yields an improvement by a factor two. Finally, other strong combinations include τ_3/τ_2 and a mass variable.

The ROC curves for combinations of tagging methods are shown in Fig. 7, where each point on the curve corresponds to a single selection window for each of the considered variables. At low boost (Fig. 7 left) and considering only fully merged top jets, a combination of the HTT V2 variables, softdropped τ_3/τ_2 and b tagging as well as a combination of the softdrop mass, n-subjettiness ratio, shower deconstruction and b tagging achieve the best overall performance. The distributions at high boost (Fig. 7 right) show a similar behaviour with the difference that groomed τ_3/τ_2 is not necessary.

We compare the susceptibility of different jet flavours to being misidentified as a top jet. This measurement is performed on simulated mulitijet events, where the jet flavor is assigned by matching the reconstructed jet with generator level partons. Considering a sample of illustrative working points for top tagging, defined in Table 2 (Table 3) for low (high) p_{T} jets respectively, the mistag rates for different parton flavors are compared. In particular, Fig. 8 (left) shows the mistag rate for high p_{T} AK8 jets when a softdrop mass in the range between 110 and 210 GeV and ungroomed $\tau_3/\tau_2 < 0.5$ are required. In contrast, Fig. 8 (right) shows the rate obtained for low p_{T} CA15 jets when selecting a softdrop mass window of 150 to 240 GeV and n-subjettiness $\tau_3/\tau_2 < 0.6$.

We do not apply any b tagging requirements as these trivially increase the relative mistagging rate of b-jets. Even without this requirement, jets from b-partons are most likely to pass the top tagging requirement, followed by charm and then gluon jets.

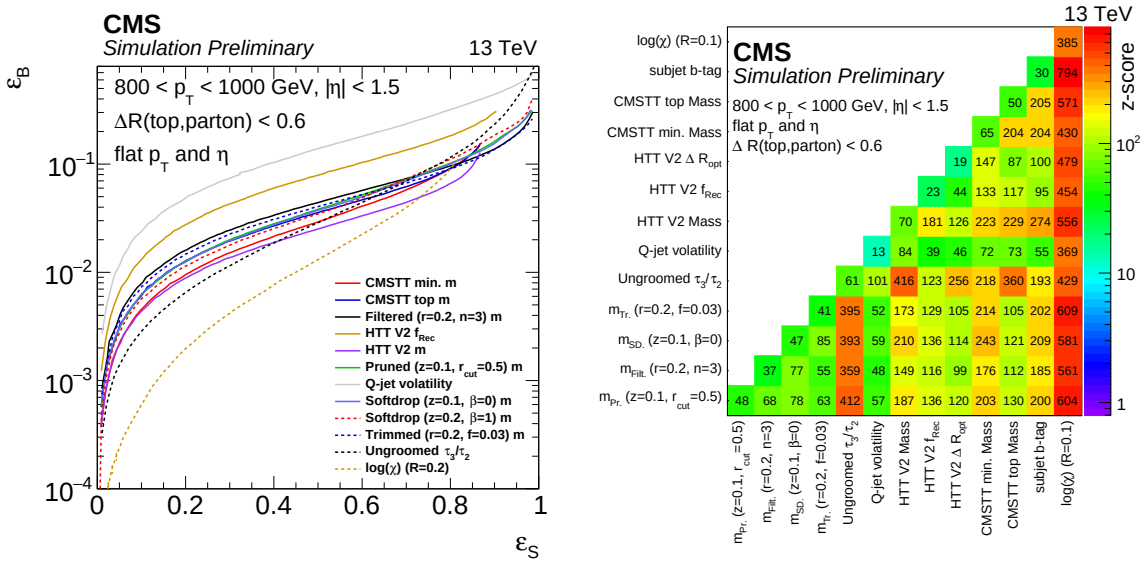

Figure 6: Single variable ROC curves (left) and z-score, defined as $1/\varepsilon_B$ at a signal efficiency of 30% (right) calculated per-parton for objects passing the fiducial selection criteria for a high p_{T} sample. Each point on the ROC curve corresponds to a simple selection window using the tagging variable. The z-score is determined using a likelihood estimator for the diagonal and a BDT for the off-diagonal elements.

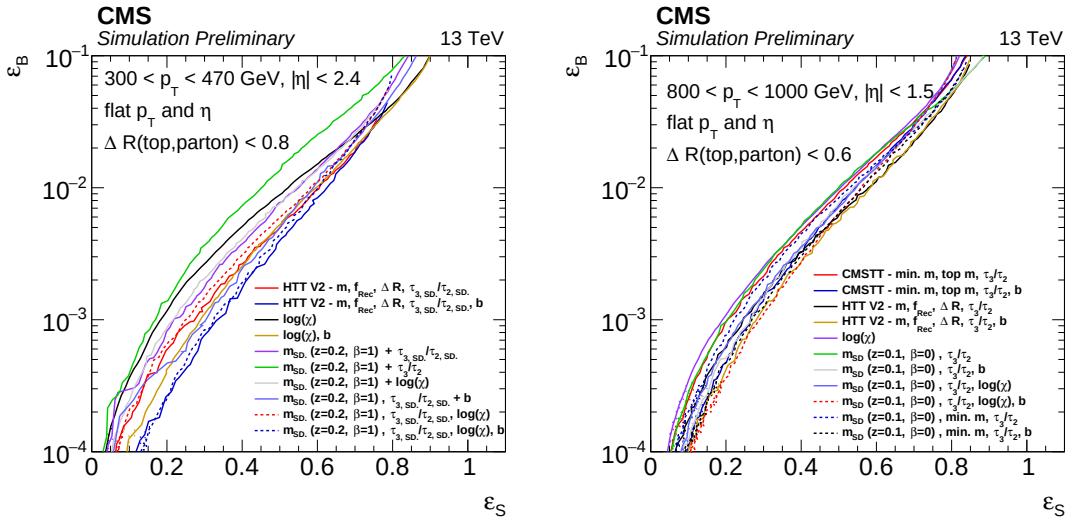

Figure 7: ROC curves for calculated per-parton for objects passing the fiducial selection criteria for merged top quarks at low boost (left) and high boost (right). Each point on the curve corresponds to a set of simple selection windows on the given variables.

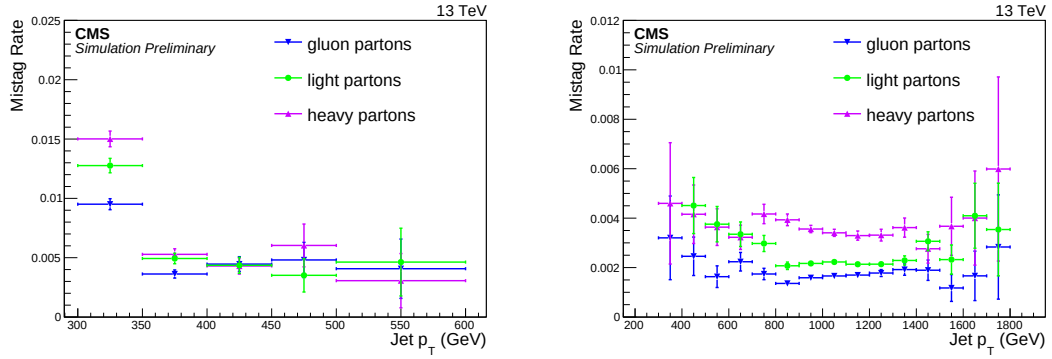

Figure 8: Comparisons of efficiencies for the low p_{T} working points (left): softdrop ($z = 0.2$, $\beta = 1$) mass with $150 < m_{SD} < 240\,\mathrm{GeV}$ and ungroomed $\tau_3/\tau_2 < 0.58$ and high p_{T} working points (right): softdrop ($z = 0.1$, $\beta = 0$) mass with $110 < m_{SD} < 210\,\mathrm{GeV}$ and ungroomed $\tau_3/\tau_2 < 0.5$.

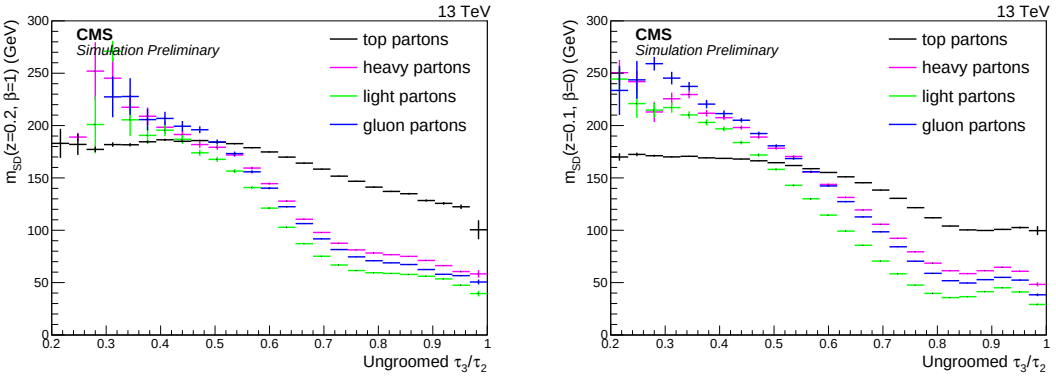

Figure 9: Profile plots showing the average values of m_{SD} as a function of ungroomed τ_3/τ_2 for (left) CA15 jets and (right) AK8 jets. For CA15 (AK8) jets the softdrop parameters $z = 0.2$, $\beta = 1$ ($z = 0.1$, $\beta = 0$) are used.

This behavior can be understood by considering the correlation between the n-subjetiness ratio and the softdrop mass. The profile of the scatter plot between softdrop mass and τ_3/τ_2 is reported in Fig. 9, where the bulk of QCD jets are characterized by low softdrop mass and a n-subjettiness ratio close to one. As published in [46], the lowest groomed mass is observed for light quark jets, followed by gluon jets and heavy flavor (c, b) quark jets.

Table 2: Summary of working points considered for studying their dependence on top quark p_T, η, and number of pileup vertices. The working points here are for the low-p_T region, $R = 1.5$ jets, and correspond to a background efficiency of 0.3%. The quantity χ_2 refers to the shower deconstruction output using a microjet size of $R = 0.2$.

Substructure Algorithms Used for Low p_T Working Points

$150 < M_{SD}(z = 0.2, \beta = 1) < 240$ GeV			$\tau_3/\tau_2 < 0.6$	
$150 < M_{SD}(z = 0.2, \beta = 1) < 240$ GeV			Groomed $\tau_3/\tau_2 < 0.5$	
$150 < M_{SD}(z = 0.2, \beta = 1) < 240$ GeV			Groomed $\tau_3/\tau_2 < 0.6$	b tag > 0.7
$120 <$ HTT V2 Mass < 180 GeV	$f_{\text{Rec}} < 0.2$		Groomed $\tau_3/\tau_2 < 0.6$	
$120 <$ HTT V2 Mass < 180 GeV	$f_{\text{Rec}} < 0.2$	$-0.5 < \Delta R_{opt} < 0.2$	Groomed $\tau_3/\tau_2 < 0.6$	
$120 <$ HTT V2 Mass < 180 GeV	$f_{\text{Rec}} < 0.3$	$-0.7 < \Delta R_{opt} < 0.3$	Groomed $\tau_3/\tau_2 < 0.7$	b tag > 0.8
	SD $\log(\chi_2) > 4.5$			
$130 < M_{SD}(z = 0.2, \beta = 1) < 230$ GeV	SD $\log(\chi_2) > 3.8$		Groomed $\tau_3/\tau_2 < 0.7$	
$130 < M_{SD}(z = 0.2, \beta = 1) < 300$ GeV	SD $\log(\chi_2) > 2.8$		Groomed $\tau_3/\tau_2 < 0.9$	b tag > 0.8

Table 3: Summary of working points considered for studying their dependence on top quark p_T, η, and number of pileup vertices. The working points here are for the high-p_T region, $R = 0.8$ jets, and correspond to a background efficiency of 0.3%. The quantity χ_1 refers to the shower deconstruction output using a microjet size of $R = 0.1$.

Substructure Algorithms Used for High p_T Working Points

$110 < M_{SD}(z = 0.1, \beta = 0) < 210$ GeV			$\tau_3/\tau_2 < 0.5$	
$110 < M_{SD}(z = 0.1, \beta = 0) < 210$ GeV			$\tau_3/\tau_2 < 0.6$	b tag > 0.8
$140 <$ CMSTT Mass < 250 GeV			$\tau_3/\tau_2 < 0.5$	
$140 <$ CMSTT Mass < 250 GeV			$\tau_3/\tau_2 < 0.6$	b tag > 0.8
$120 <$ HTT V2 Mass < 180 GeV	$f_{\text{Rec}} < 0.2$	$-0.1 < \Delta R_{opt} < 0.1$	$\tau_3/\tau_2 < 0.6$	
$90 <$ HTT V2 Mass < 170 GeV	$f_{\text{Rec}} < 0.2$	$-0.2 < \Delta R_{opt} < 0.2$	$\tau_3/\tau_2 < 0.7$	b tag > 0.8
	SD $\log(\chi_1) > 5.4$			
$90 < M_{SD}(z = 0.1, \beta = 0) < 180$ GeV	SD $\log(\chi_1) > 4.6$		$\tau_3/\tau_2 < 0.7$	
$70 < M_{SD}(z = 0.1, \beta = 0) < 190$ GeV	SD $\log(\chi_1) > 3.9$		$\tau_3/\tau_2 < 0.7$	b tag > 0.6

6 Tagger Efficiency Kinematic Dependence

In this section, we examine the observables previously described, comparing the performance of top-tagging working points defined in Table 2 and Table 3. An optimal working point should provide high signal efficiency and background rejection across a wide range of jet p_T and $|\eta|$ values, as well as maintain stable performance for large amounts of pileup activity in the event. First, a working point is chosen to correspond to a background efficiency of 0.3%, using the ROC curves previously shown in Fig. 7, without reweighting. The values come from a single bin in p_T, but the working point is then tested for stability across p_T, $|\eta|$, and number of pileup vertices. Table 2 summarizes the working points considered for the low p_T regime, while Table 3 refers to those for high p_T jets.

Signal efficiency and mistag rate are shown in Fig. 10 as a function of the generator truth matched parton p_T and the number of pileup vertices for jets clustered with CA15. In contrast, Fig. 11 shows the result obtained for high p_T jets clustered via AK8. The error bars represent the statistical uncertainty in each specific bin, due to a limited amount of simulated events. In all cases, the working points are chosen to correspond to a background tagging effiency of 0.3% in the inclusive sample. No merged top requirement is enforced in the signal sample to avoid an artificial increase of the top tagging efficiency.

In the low p_T regime, the efficiency of all taggers increases as a function of p_T, since the top

quark decay products become more collimated. The combination of softdrop mass with shower deconstruction provides the highest efficiency but has a mistag rate with a large dependence on the parton p_T. The large p_T dependence of the shower deconstruction variable is shown in Fig. 5 (left) and is likely due to microjets not calibrated through jet energy corrections. Other algorithms, such as the HTT mass and softdrop mass, show mistag rates with little to no dependence on the jet momentum. Additionally, the efficiency decreases as a function of number of pileup vertices for all taggers. The addition of the n-subjettiness ratio τ_3/τ_2 and subjet b tagging tends to stabilize the efficiency and mistag rates as a function of the number of pileup vertices.

For the high-p_T regime, with the smaller jet cone size, a more visible turn-on behavior is observed as a function of the top quark transverse momentum, from 400 to 600 GeV. This is due to the decrease in angular distance between the top quark decay products approaching the threshold of the jet cone size. Above this transition, the efficiencies for the various algorithms are relatively flat as a function of jet p_T, with the exception of shower deconstruction, which has a much slower turn-on behavior. Similarly to the low p_T case, jets tagged using shower deconstruction have a mistag rate that increases dramatically with p_T, which may not be suitable for physics analyses. The algorithm with the largest efficiency for the high-p_T region is again found to be shower deconstruction, when combined with softdrop mass, subjet b tagging, and τ_3/τ_2.

Based on the stable performance shown, we recommend softdrop combined with n-subjettiness to be used as the primary strategy for top tagging at high p_T. At low p_T, the CMS default prescription uses softdrop combined with groomed n-subjettiness. These working points will now be referred to as CMSTT version two for high p_T (CMSTT V2H) and for low p_T (CMSTT V2L), respectively. In addition, analyses interested in a tagging method valid over a wide p_T range should consider the HTT V2 strategy. To increase background rejection and signal efficiency stability, subjet b tagging should be added to each chosen working point.

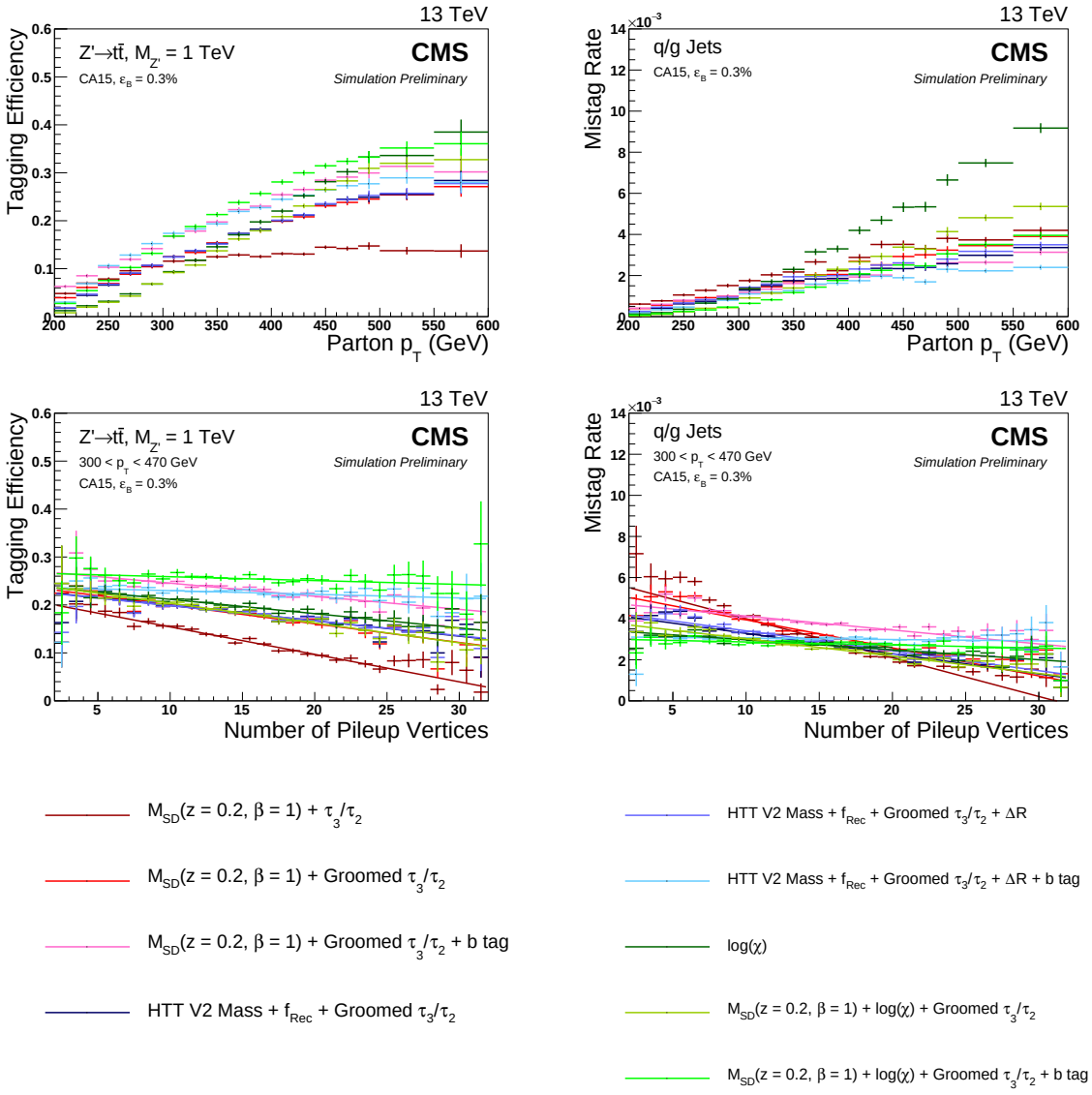

Figure 10: Top-tagging efficiency for the low p_{T} working points listed in Table 2. The top plots show efficiency as a function of parton p_{T}, while the bottom ones show efficiency as a function of the number of pileup vertices. The plots in the first column are based on a $\mathrm{Z}' \rightarrow t\bar{t}$ sample with $M_{Z'} = 1\,\mathrm{TeV}$, while those in the second column refer to QCD multijet production. The top right plot uses a flat parton p_{T} distribution whereas a p_{T}=300–470 GeV QCD background sample is used for the bottom right one.

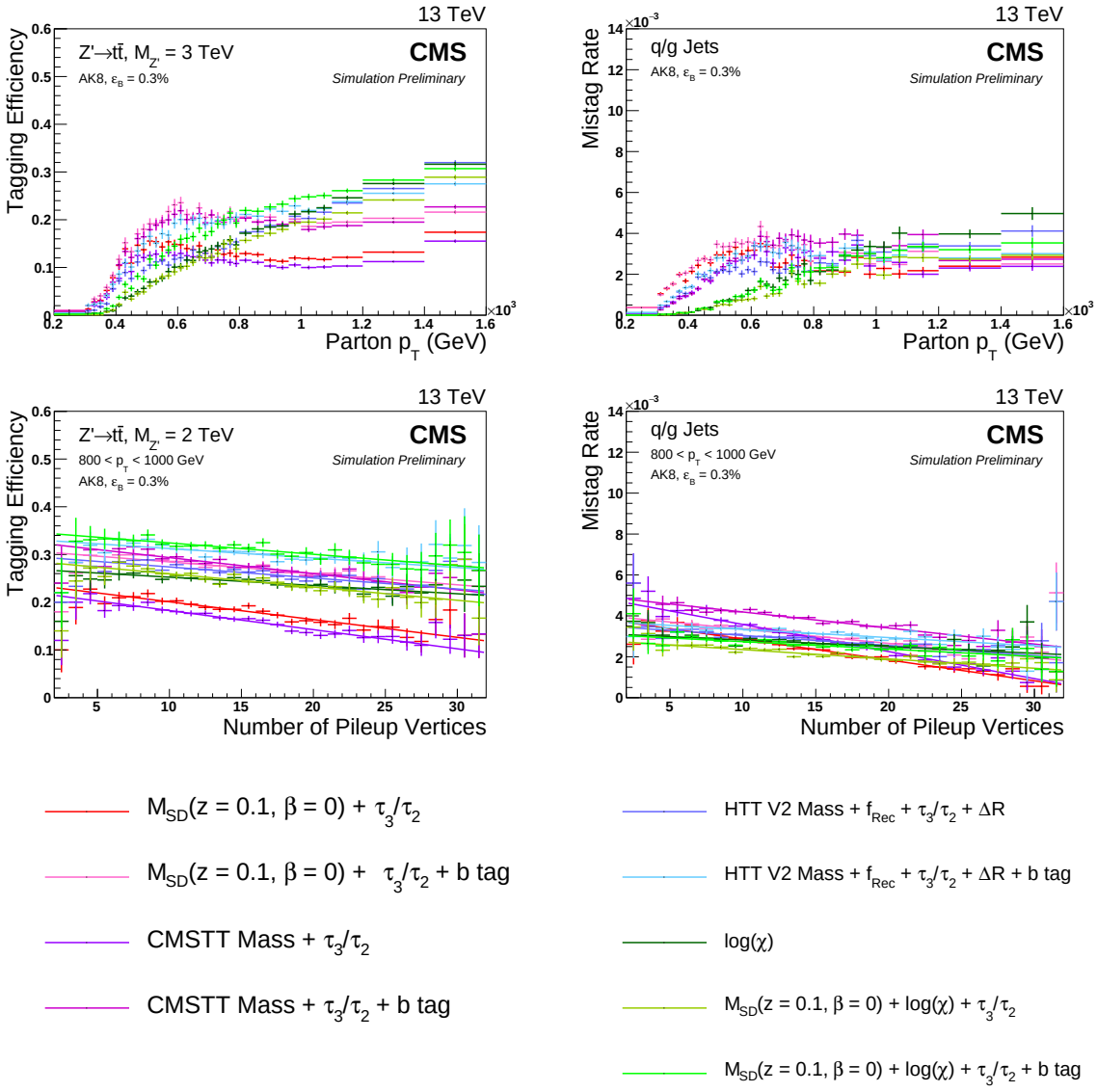

Figure 11: Top-tagging efficiency for the high p_{T} working points listed in Table 3. The top plots show efficiency as a function of parton p_{T}, while the bottom ones show efficiency as a function of the number of pileup vertices. The plots in the first column are based on a $\mathrm{Z}' \to t\bar{t}$ sample with $M_{\mathrm{Z}'} = 3\,\mathrm{TeV}$ or 2 TeV, while those in the second column refer to QCD multijet production. The top right plot uses a flat parton p_{T} distribution whereas a p_{T}=300–470 GeV QCD background sample is used for the bottom right one.

7 Validation on 8 TeV Data

In this section, simulated events are compared to collision data with the aim of validating the top tagging strategies previously defined, as well as the simulation of substructure variables. If the distribution of an observable is not properly described by the simulation, but is used as a decision variable, it can lead to differences in the performance of the algorithm between data and simulation. The studies help to point out both differences and shortcomings of the simulations used. Results can be used to both improve our simulation and to obtain correction factors for the efficiency and mistag rates for a given top tagging algorithm.

Distributions of substructure variables are compared in a sample dominated by top-jets in data as well as in an enriched QCD multijet background region. Measurements of tagging efficiency and mistag rate are presented for various taggers, performed using data collected at $\sqrt{s} = 8\,\text{TeV}$ for a total integrated luminosity of 19.7 fb^{-1}.

The measurement of the efficiency and mistag rate in data is more challenging than the calculation in simulation, since the generator information is not availible. In addition, it is not possible to use individual $\mathrm{t\bar{t}}$ and QCD multijet samples, as the dataset collected by a certain trigger contains a mixture of various SM processes. It has to be ensured that jets produced by a hadronically decaying top quark are used to measure the efficiency, while only jets originating from QCD interactions are considered for the mistag rate measurement. For this reason, a top-jet enriched sample is isolated in data from a region of the phase space dominated by semi-leptonic $\mathrm{t\bar{t}}$ production, while an almost pure sample of high p_T QCD jets is selected in di-jet events. In both cases, a tag and probe method is used to measure tagging efficiency as well as the mistag rate. In the signal case, the top jet candidate is required to be reconstructed in the opposite hemisphere with respect to a leptonically decaying W boson. In contrast, for background studies, the di-jet topology is considered where one of the two jets is randomly selected as a tag object, while the other one as a probe jet.

Since the top tagging algorithms are applied on jets with different distance parameters, as described in Section 3, the event selection is applied independently for two different CA jet collections, reconstructed with $R = 0.8$ (CA8) or $R = 1.5$ (CA15).

7.1 Signal selection

The efficiency of top tagging algorithms is measured in a semi-leptonic $\mathrm{t\bar{t}}$ control region, as the leptonic leg of the event provides a clean signature. The measurement is performed in the muon channel, because of the excellent detector performance for the reconstruction and identification of muons, which results in a sample with a negligible contamination from QCD multijet events.

Events are recorded by a single muon trigger, which requires the presence of one reconstructed and identified muon in the event with p_T larger than 40 GeV. To reject backgrounds, mainly coming from leptonically decaying W bosons produced in association with jets, the following selection criteria are applied:

- exactly one muon with $p_\mathrm{T} > 45\,\text{GeV}$ and $|\eta| < 2.1$,
- no electron with $p_\mathrm{T} > 35\,\text{GeV}$ and $|\eta| < 2.5$,
- missing transverse energy $E_T^{\text{miss}} > 20\,\text{GeV}$,
- $H_T^{\text{lep}} > 150\,\text{GeV}$, where $H_T^{\text{lep}} = E_T^{\text{miss}} + p_\mathrm{T}^{\mu}$,
- at least one b-tagged AK5 jet according to the CSV algorithm [54] considering a working point with a nominal background rejection of 99%,

Table 4: Sample composition after the signal selection including the statistical uncertainty.

Process	fraction of events [%]	
	CA8 selection	CA15 selection
$t\bar{t}$ unmatched	13.6 ± 0.5	19.7 ± 0.5
$t\bar{t}$	55.5 ± 1.0	55.0 ± 0.8
Single top	9.5 ± 0.4	7.7 ± 0.3
W+jets	18.6 ± 0.6	14.9 ± 0.4
Other	3.2 ± 0.2	2.8 ± 0.2

- at least one reconstructed CA15 or CA8 jet with $p_T > 200\,\text{GeV}$ and $|\eta| < 2.4$.

The selection is performed twice, once for CA8 jets, and once for CA15 jets. The CA jet is used as a candidate for the probe jet. The differences between the two selections are mostly due to kinematics, where for the CA15 selection the jet transverse momenta will be higher on average. After this selection, the sample is enriched with $t\bar{t}$ events, making up more than 70% of the total. The event composition after this selection is summarized in Table 4. About 13–20% of jets in $t\bar{t}$ events passing the signal selection are not matched to hadronically-decaying top quarks within $\Delta R < 1.2$ (0.6) for CA15 (CA8) selection according to MC simulation studies, but no attempt is made to subtract them in the efficiency measurement presented in this document.

The reconstruction of the leptonic leg of the top-quark decay is used to obtain a top-quark tag independent of the jet substructure of the CA jet assigned to the hadronic decay leg. This reconstruction uses a χ^2 method [55, 56] to solve the system for the two possible neutrino four-momenta obtained from the W boson mass constraint and multiple jet assignments. Once the leptonically decaying top quark candidate (tag object) is obtained, the leading CA jet, placed in the opposite hemisphere ($\Delta\phi > 2.1$), is selected as probe jet.

After the event selection the signal sample consists of about 15–19% of events from W production in association with jets and about 8% of single top events. After the selection, top quarks from single top production decay preferentially into leptons (66% of cases), and are therefore considered as background events. Thus, the efficiency is defined as:

$$\epsilon = \frac{N_{\text{tagged jets}} - N_{\text{tagged jets (background events)}}}{N_{\text{all jets}} - N_{\text{all jets (background events)}}}, \tag{8}$$

where the denominator consists of all probe jets, while the numerator counts all probe jets passing the top-tagging requirements. Jets from background events are substracted from the denominator as well as the numerator.

In the efficiency measurements statistical and systematic uncertainties are shown, where the systematic uncertainties come from jet energy correction, jet energy resolution, top p_T reweighting and parton shower models.

In Fig. 12, the p_T distribution of probe jets for the CA8 and CA15 selection are shown. The data are well described by the simulation after scaling the $t\bar{t}$ normalization with a factor of 0.9. This effect has already been observed previously [24].

7.2 Background selection

The main background to top tagging comes from jets produced by the fragmentation of gluons or light quarks. Therefore, the misidentification rate is measured in a sample enriched

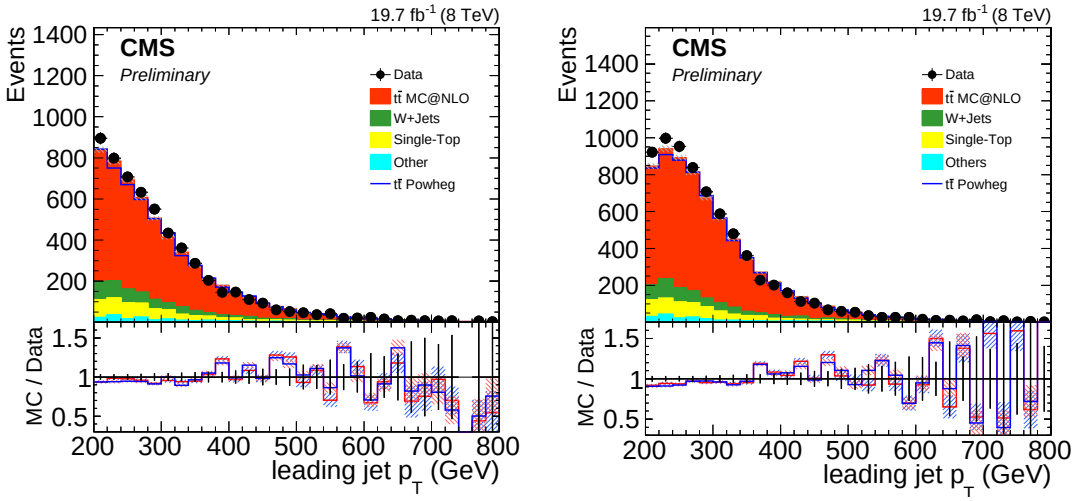

Figure 12: Distribution of p_{T} for selected CA8 (left) and CA15 jets (right) after the signal selection. Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom, where the ratio to POWHEG is reported in blue while the one to MC@NLO is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

in light-flavour jets from QCD multijet production, obtained by selecting di-jet-like topologies. Contaminations from $\mathrm{t\bar{t}}$ production in the all-hadronic channel are suppressed by the following selection criteria.

Di-jet events have been recorded by a trigger based on the scalar sum of the transverse momenta of all jets, defined as:

$$H_{\mathrm{T}} = \sum_{jets} p_{\mathrm{T}}\,, \tag{9}$$

where the sum runs over all anti-k_T jets in the event reconstructed with a distance parameter of $R = 0.5$. The trigger selects events with $H_{\mathrm{T}} > 750$ GeV.

Then, the following selection is applied:

- events with a reconstructed electron or muon are discarded,
- $H_{\mathrm{T}} > 1000$ GeV is required to ensure a fully efficient trigger,
- at least two CA jets with $p_{\mathrm{T}} > 400$ GeV and $|\eta| < 2.4$,
- the difference in azimuth $\Delta\phi$ between the CA jets has to be larger than 2.1 units.

Similar to the signal case, the event selection is performed twice, once for CA8 and once for CA15 jets. To measure the misidentification rate, one of the two jets in the event is chosen randomly as a tag jet. The selected tag jet has to fulfill the following criteria:

- the jet mass has to be in a window $140 < m_{\mathrm{jet}} < 250$ GeV around the top quark one, to achieve comparable kinematics to events from $\mathrm{t\bar{t}}$ production,
- the minimum pairwise mass, calculated from the subjets found by the CMSTT, has to be smaller than 50 GeV to further reduce the contamination by hadronically decaying top quarks.

If the criteria are fulfilled, the jet on the opposite hemisphere to the tag one is considered as the probe jet and is used to measure the misidentification rate through Eq. (8). After the back-

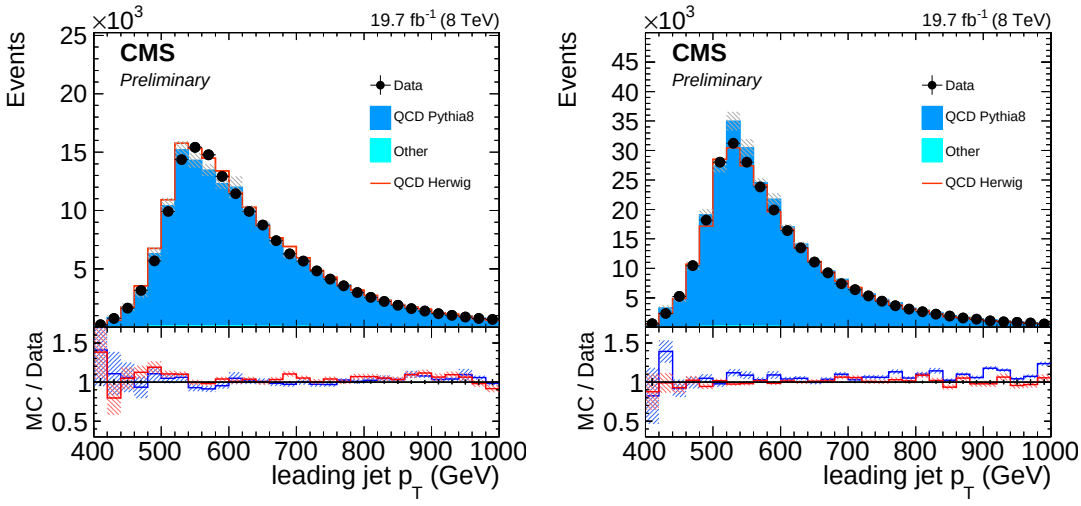

Figure 13: p_T distribution for CA8 (left) and CA15 (right) jets after the background selection. Data is compared to simulated events by PYTHIA 8 and HERWIG++. Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom, where the ratio to PYTHIA 8 is reported in blue while the one to HERWIG++ is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

ground selection, the selected event sample is dominated by multi-jet production, with less than 1% of contamination from $t\bar{t}$. For this reason, no background subtraction is performed for the mistag rate evaluation.

Good agreement between data and simulated events is achieved by reweighting the p_T distribution of the leading jet. After this procedure, also the probe jet kinematic is well described. As an example, the p_T distribution of the leading CA8 and CA15 jets in the event is shown in Fig. 13.

7.3 Results of data to simulation comparison

In this section, the top tagging efficiency and the misidentifiaction rate are measured and compared to data. Differences between data and simulation will lead to correction factors or further improvement of the simulation. For all tested top tagging strategies important substructure variables are shown and analyzed.

7.3.1 CMS Top Tagger

In Fig. 14, the ungroomed jet mass for the signal and background selection is shown. Since the CMSTT applies a jet mass selection, a good description of this variable is necessary for a correct simulation of the signal efficiency and mistag rate. In the top-jet enriched sample, the jet mass distribution shows two peaks, one in the region of the W boson mass, the other one close to the top quark one. The agreement between data and simulation is good over the whole mass range. In di-jet events, HERWIG++ describes the data as well as PYTHIA 8 does, although below 50 GeV the agreement becomse worse for both generators.

The number of subjets found by the CMSTT is shown in Fig. 15 (top). There is fair agreement between data and MC in the signal selection, albeit a slight slope in the ratio of data to simulation is observed. The description in the background region by HERWIG++ and PYTHIA is good, only high subjet multiplicities are not well modeled by PYTHIA 8. Comparing the signal

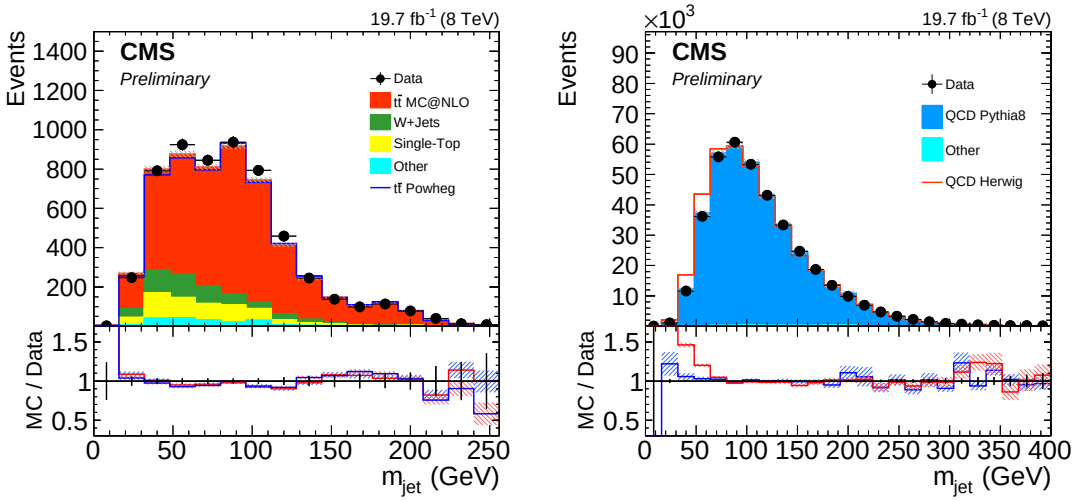

Figure 14: Jet mass for the signal selection (left) and the background one (right). Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom of each panel. In the case of the signal (background) selection, the ratio to POWHEG (PYTHIA 8) is shown in blue while the one to MC@NLO (HERWIG++) is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

with the background regions, it is confirmed that the number of subjets is a powerful variable to discriminate between jets originating from collimated top-quark decays and light flavor or gluon jets.

The minimum pairwise mass distribution is shown in Fig. 15 (bottom). This variable can be only calculated if the input jet has at least three subjets. It is well described for the signal selection for values above 60 GeV, although a large disagreement is visible in the turn-on region, where m_{min} is below 50 GeV. This has already been observed in earlier studies [24]. At around 80 GeV, a peak from the decay of the W boson is observed. In contrast, the peak region in the background sample is well described by both PYTHIA 8 and HERWIG++ generators, while the agreement becomes worse towards low m_{min} values.

The CMSTT observables are used to define a top tagging strategy, requiring the jet mass to be in the range from 140 to 250 GeV, a minimum pairwise mass above 50 GeV and at least two subjets. Fig. 16 shows the tagging efficiency and the mistag rate for this working point as a function of the top-jet candidate transverse momentum.

The efficiency of the CMSTT as measured in data agrees well with the one in simulation within the statistical and systematic uncertainties. The systematic uncertainties come from the uncertainties on jet energy scale and resolution measurements. The efficiency increases with increasing p_{T} up to 60%. The integrated efficiency for $p_{\text{T}} > 350$ GeV is 28%. The efficiency for jets with a p_{T} smaller than 400 GeV is very small. This is mostly due to the distance parameter used for jet clustering ($R = 0.8$), within a single jet and therefore the jet mass might be biased to lower values failing, as a consequence, the m_{jet} selection. As expected, the mistag rate in data agrees well with HERWIG++ and PYTHIA, since the substructure decision variables are well modeled by both event generators. The mistag rate for the CMSTT increases up to 6% at values of p_{T} around 800 GeV, and drops with increasing p_{T} similarly to the efficiency. The integrated mistag rate in HERWIG++ for $p_{\text{T}} > 350$ GeV is 5.4%.

In Fig. 17, efficiency and mistag rate are shown as a function of the number of reconstructed

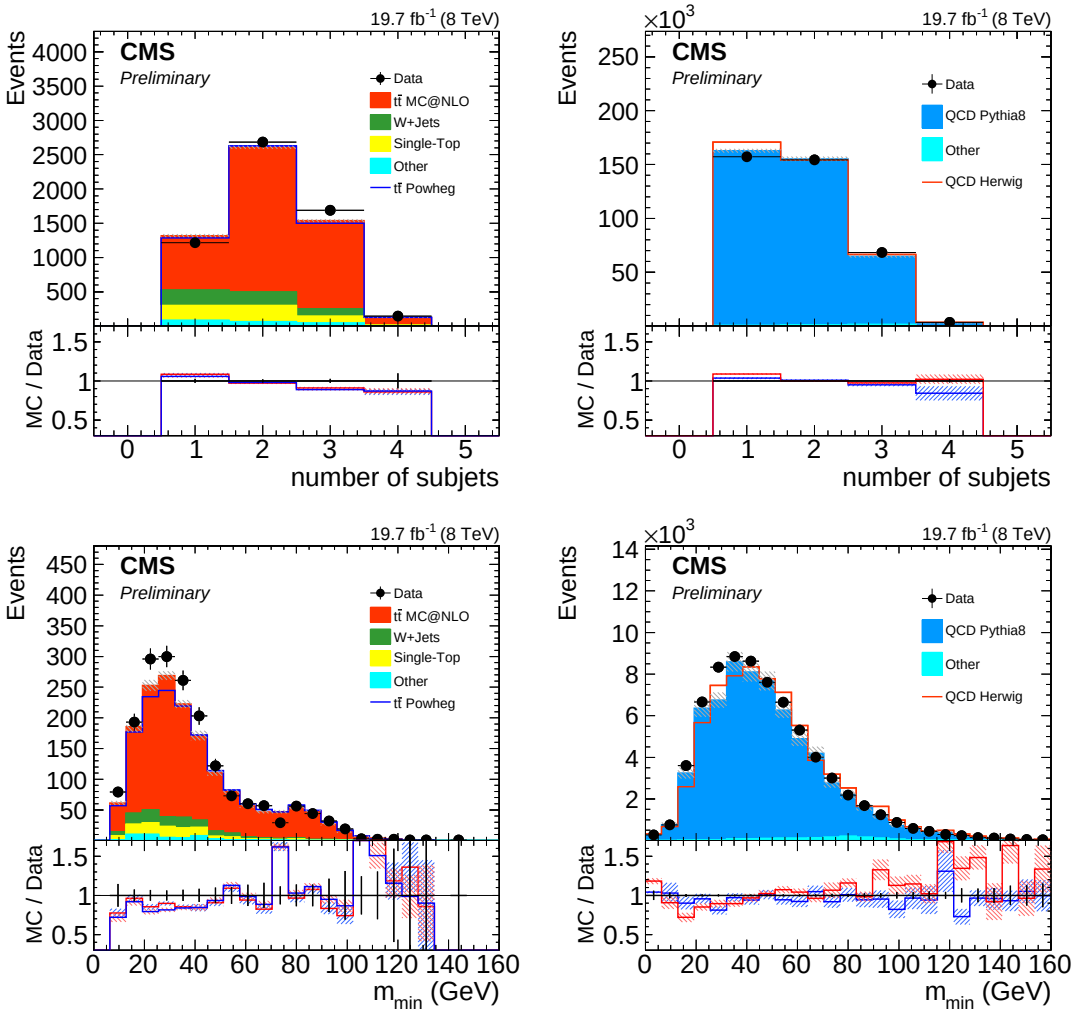

Figure 15: Distribution of the minimum pairwise mass (bottom) and the number of subjets found by the CMSTT (top) for the signal sample (left) and for the background one (right). Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom of each panel. In the case of the signal (background) selection, the ratio to POWHEG (PYTHIA 8) is shown in blue while the one to MC@NLO (HERWIG++) is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

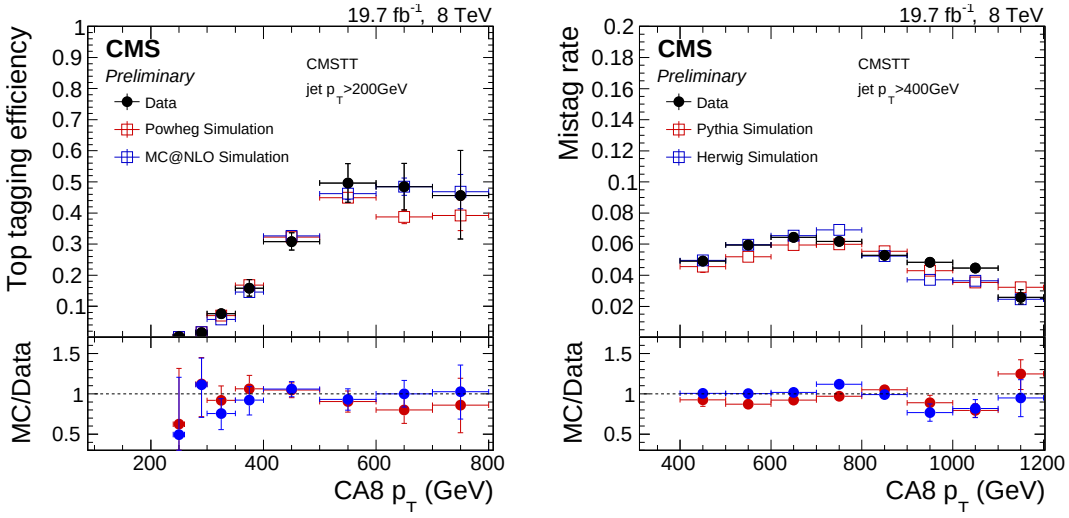

Figure 16: Efficiency (left) and mistag rate (right) of the CMSTT as function of p_{T}. The error bars show the combined statistical and systematic uncertainties. At the bottom of each panel the ratio of simulation to data is shown.

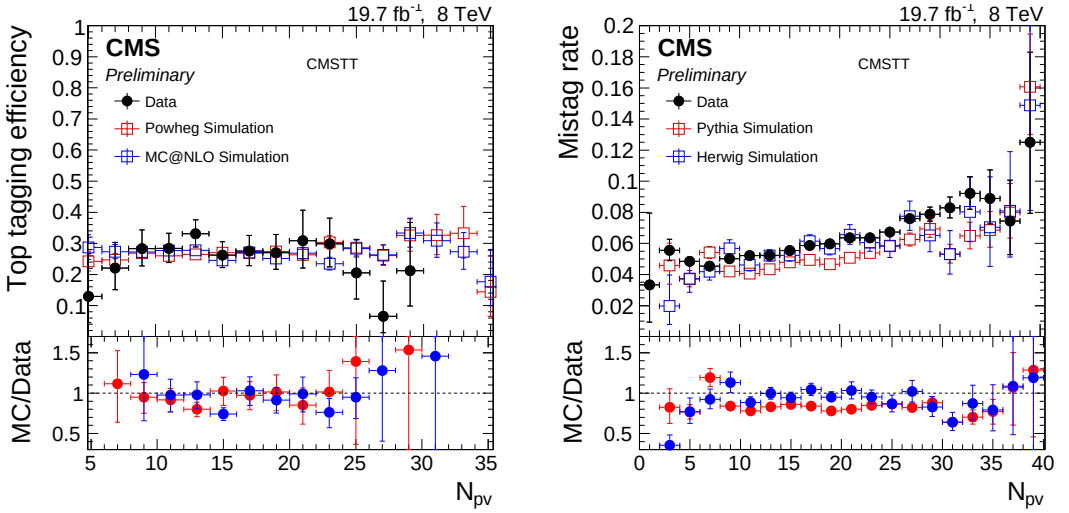

Figure 17: Efficiency (left) and mistag rate (right) of the CMSTT as function of the number of reconstructed primary vertices. The error bars show the combined statistical and systematic uncertainties. At the bottom of each panel the ratio of simulation to data is shown.

primary vertices in the event. Since pileup particles clustered inside the jet tend to shift the mass to higher values, a slope is observed in both cases. Since the upper limit on the jet mass window is rather high, with a value of (250 GeV), signal jets are not as much affected as background ones. A slope towards higher values of the mistag rate with an increasing number of primary vertices is visible in both data and simulation.

7.3.2 HEPTopTagger V2

The HTT V2 is an improvement over the original HEPTopTagger [19, 20]. The distribution of the optimal distance parameter R_{opt}, as found by the algorithm, is shown in Fig. 18 (top). In the signal selection the distribution exhibits a broad peak around one, which indicates that without any additional p_{T} selection the optimal jet radius is around one. With increasing jet p_{T},

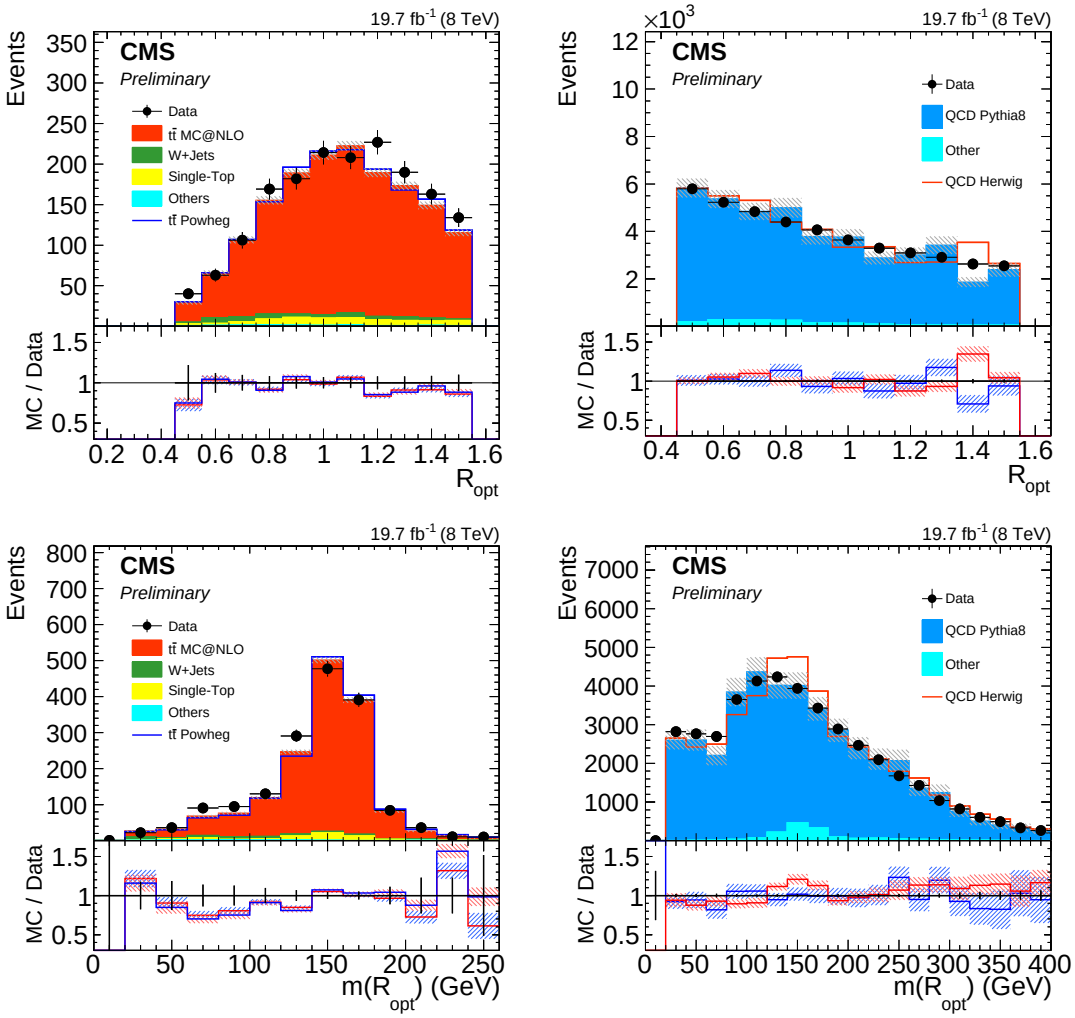

Figure 18: Distribution of the optimal radius R_{opt} (top) and the mass m_{123} at the optimal radius R_{opt} (bottom) of the HTT v2 for the signal selection (left) and for the background one (right). Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom of each panel. In the case of the signal (background) selection, the ratio to POWHEG (PYTHIA 8) is shown in blue while the one to MC@NLO (HERWIG++) is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

the peak shifts to lower values as the decay products become more collimated. In contrast, for background jets no resonant peak is visible. Overall, the agreement between data and simulation is good for both selections. The top quark mass m_{123} at the minimum radius R_{opt} is shown in Fig. 18 (bottom). For background events a smoothly falling distribution is observed with an additional peak around 30 GeV, caused by limiting the mass of the subjets to 30 GeV. While the signal region is well described by the POWHEG + PYTHIA 6 $\text{t}\bar{\text{t}}$ simulation, the distribution for background jets is better modeled by PYTHIA 8 than HERWIG++. Finally, the new decision variable $R_{\text{opt}} - R_{\text{opt}}^{\text{calc}}$ is shown in Fig. 19 and it is well described by all the different event generators and behaves as shown Section 4.

The efficiency and mistag rate as function of the jet p_{T} is reported in Fig. 20 (top), where a tight working point on HTT decision observables is considered. In particular, the top quark mass evaluated using the optimal R_{min} is required to be in the range between 140 and 250 GeV, then

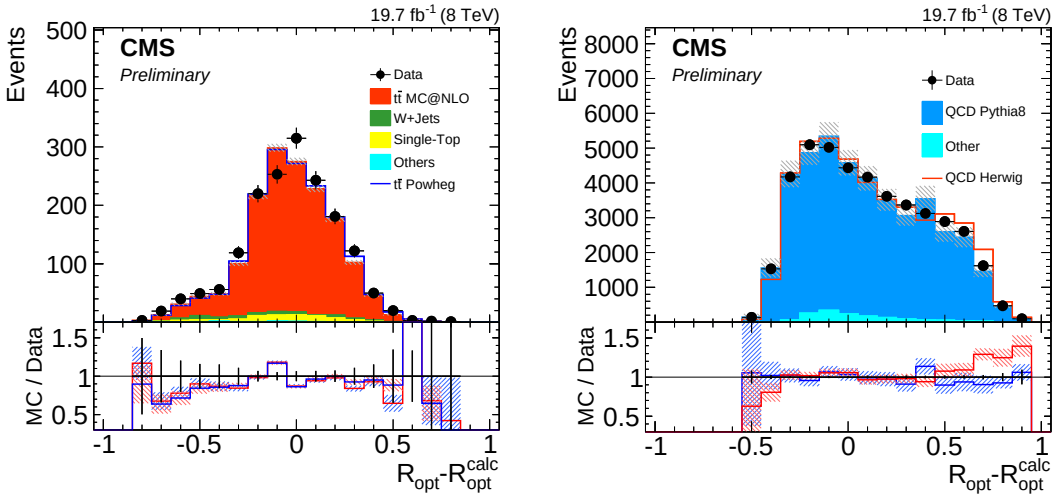

Figure 19: Distribution of the decision variable $R_{\text{opt}} - R^{\text{calc}}_{\text{opt}}$ for the signal selection (left) and background one (right). Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom of each panel. In the case of the signal (background) selection, the ratio to POWHEG (PYTHIA 8) is shown in blue and while the one to MC@NLO (HERWIG++) is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

f_{Rec} and $R_{\text{opt}} - R^{\text{calc}}_{\text{opt}}$ should be lower than 0.15 and 0, respectively.

Similar to the taggers discussed above, the signal efficiency is well described by the simulation. The integrated efficiency for $p_{\text{T}} > 350$ GeV is 28% and therefore comparable to the one of the CMSTT shown in Fig. 16. The measured mistag rate is about 2% for $p_{\text{T}} > 500$ GeV and 4% for lower values of p_{T}. The average mistag rate is 2.8% for both HERWIG++ and PYTHIA 8 and is a factor two smaller than the one of CMSTT at roughly the same total signal efficiency. Finally, the measurement of the efficiency and mistag rate as function of the number of primary vertices is reported in Fig. 20 (bottom), where HTT V2 shows to be fairly independent of pileup. This behavior is well described by simulation.

7.3.3 Shower deconstruction

The shower deconstruction algorithm is based on a single likelihood-based observable (χ), as described in Section 3, which is computed from a given set of reconstructed microjets inside the input jet. In this section, only results for CA15 jets are shown.

Fig. 21 (top) shows the number of microjets as found by re-clustering the particle flow candidates associated to the top-jet with the k_T-algorithm, with a distance parameter $R = 0.2$. Good agreement is observed for the signal selection. It is observed that limiting the maximum number of microjets to nine, in order to reduce computational time, does not lead to a loss of signal since almost no signal jets contain more than nine subjets. HERWIG++ does not provide a good description of the microjet multiplicity for background jets, while PYTHIA 8 shows a good agreement within the statistical uncertainties. The invariant mass of all microjets is shown in Fig. 21 (bottom). The agreement between data and simulation is good for both selections. Only the very low mass region for background jets is not well modeled by either of the two event generators.

The decision variable χ for jets with $p_{\text{T}} > 350$ GeV is shown in Fig. 22. The distribution is well modeled in the signal region, confirming the good description of microjets in $\bar{\text{t}}\text{t}$ production.

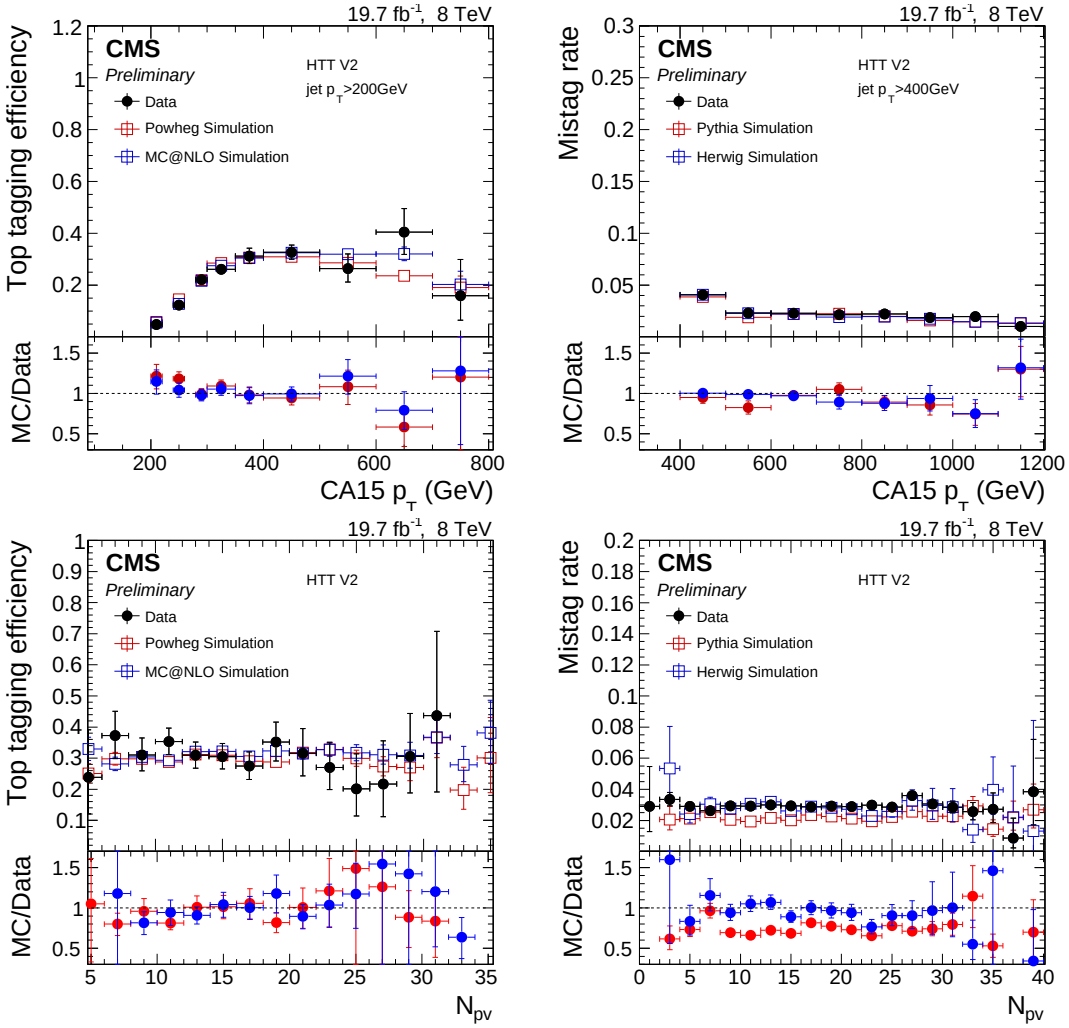

Figure 20: Efficiency (left) and mistag rate (right) of the HTT V2 as function of p_{T} (top) and the number of reconstructed primary vertices (bottom). The error bars show the combined statistical and systematic uncertainties. At the bottom of each panel the ratio of simulation to data is shown.

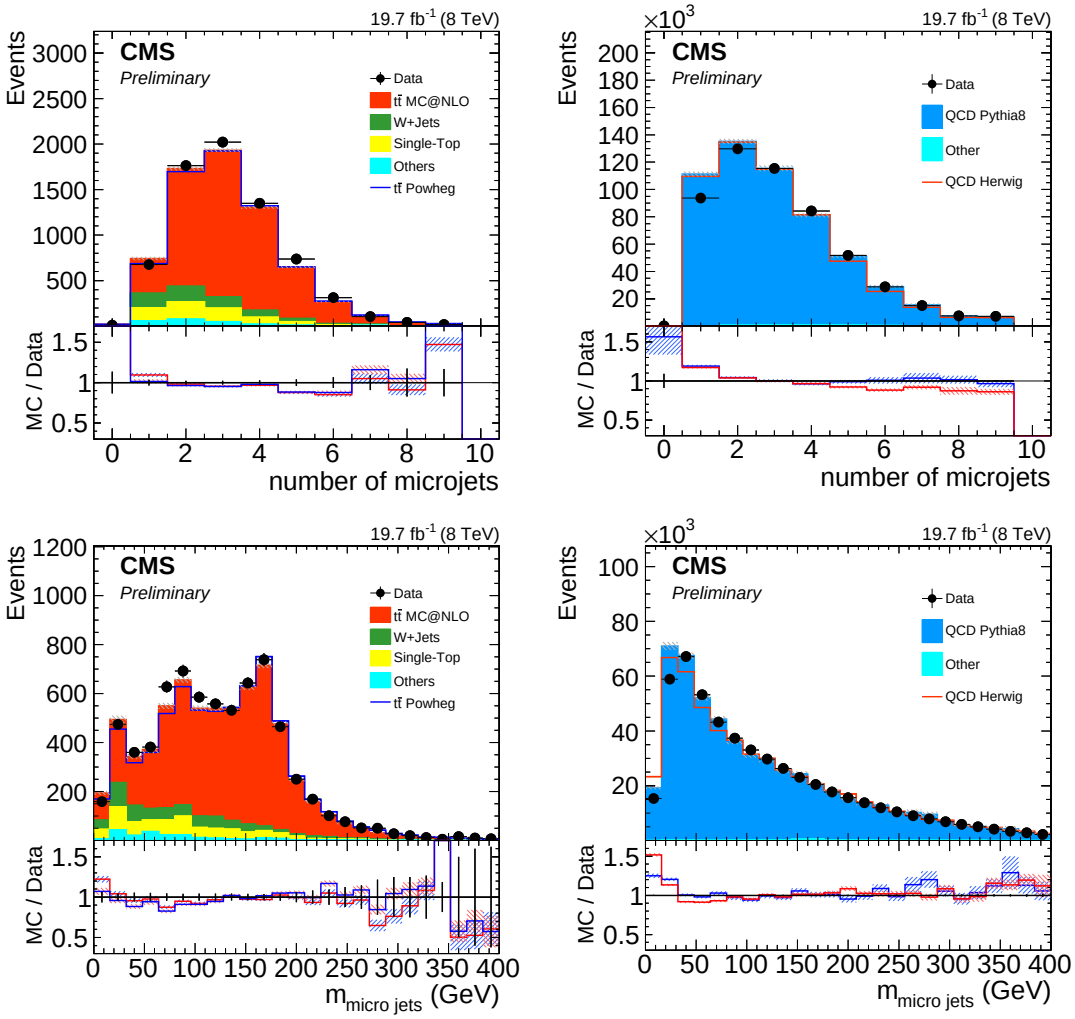

Figure 21: Distribution of the microjet multiplicity (top) and invariant mass of all microjets (bottom) for the signal sample (left) and background one (right). Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom of each panel. In the case of the signal (background) selection, the ratio to POWHEG (PYTHIA 8) is shown in blue while the one to MC@NLO (HERWIG++) is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

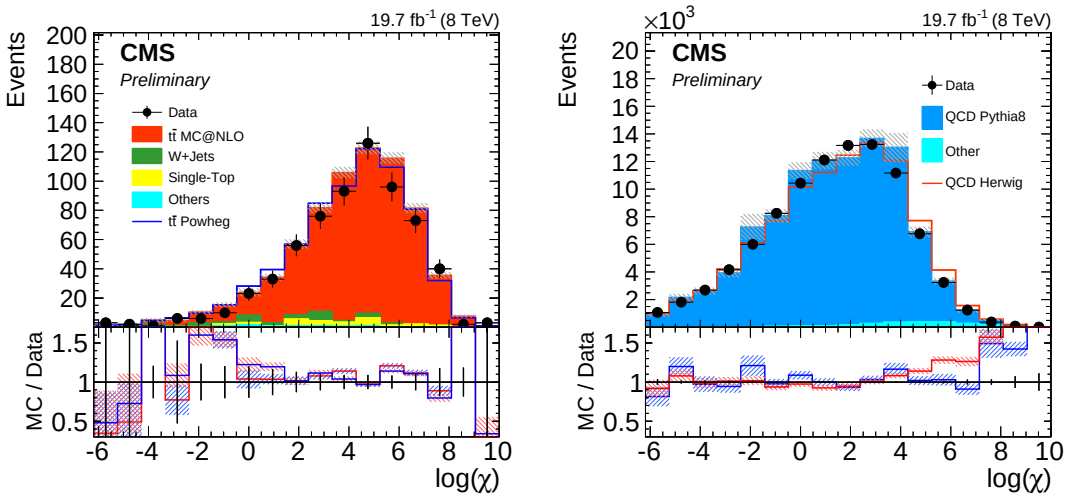

Figure 22: Distribution of the decision variable χ for jets with $p_T > 350$ GeV for signal jets (left) and background ones (right). Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom of each panel. In the case of the signal (background) selection, the ratio to POWHEG (PYTHIA 8) is shown in blue while the one to MC@NLO (HERWIG++) is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

In the background one, some discrepancy between data and the simulation for HERWIG++ is observed. This can be traced back to the mismodeling of the microjet multiplicity.

Finally, the measurement of the efficiency and the mistag rate for $\log(\chi) > 3.5$ is shown in Fig. 23. The efficiency increases up to 45% at values of p_T of around 600 GeV. The shower deconstruction tagger is not very efficient in the low p_T region, and should not be used in analyses sensitive to the transition region between the fully-merged and resolved decay of the top quarks. The integrated efficiency for $p_T > 350$ GeV is 23.1%. Data and simulation are in good agreement for the efficiency. The mistag rate is almost flat and below 5%. It is slightly better described by the PYTHIA event generator. Fig. 23 (bottom) shows the efficiency as function of the number of primary vertices N_{pv}. The agreement between data and simulation is good for the signal selection and no dependence on N_{pv} is observed. Also the mistag rate shows no significant dependence on the number of primary vertices. Again, the agreement between data and simulation is slightly better for the PYTHIA multijet sample than for HERWIG++.

7.3.4 Softdrop mass and n-subjettiness

The studies described in Section 5 show softdrop as a very promising new technique for top tagging. In this section, we explore how well the softdrop mass is modeled and if it is feasible to use it in future analyses, considering as soft threshold $z_{cut} = 0.2\,(0.1)$ and an angular parameter $\beta = 1\,(0)$ for CA15 (CA8) jets.

In Fig. 24 the jet mass after applying softdrop is shown, where a clear peak at the top quark mass is visible for the signal selection. In background events, the mass is shifted to smaller values and good separation power between signal and background is clearly visible. The softdrop mass distribution is well described by both HERWIG++ and PYTHIA 8.

As recommended in Section 5, the efficiencies and mistag rates for a mass selection on the softdropped jet of 110 GeV (150 GeV)$< m_{SD} <$210 GeV (240 GeV) combined with a one on n-

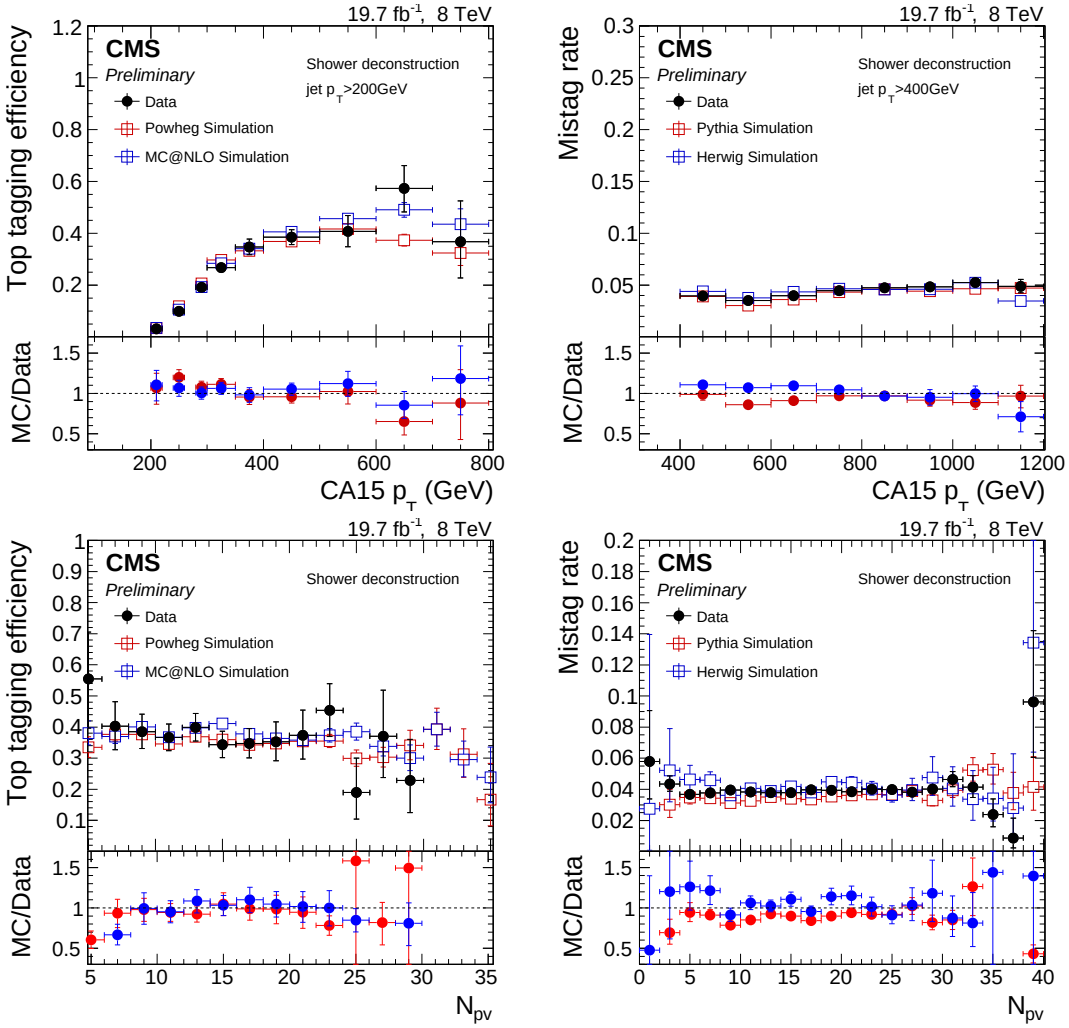

Figure 23: Efficiency (left) and mistag rate (right) of the shower deconstruction tagger requiring $\log(\chi) > 3.5$ as function of p_{T} (top) and the number of reconstructed primary vertices (bottom). The error bars show the combined statistical and systematic uncertainties. At the bottom of each panel the ratio of simulation to data is shown.

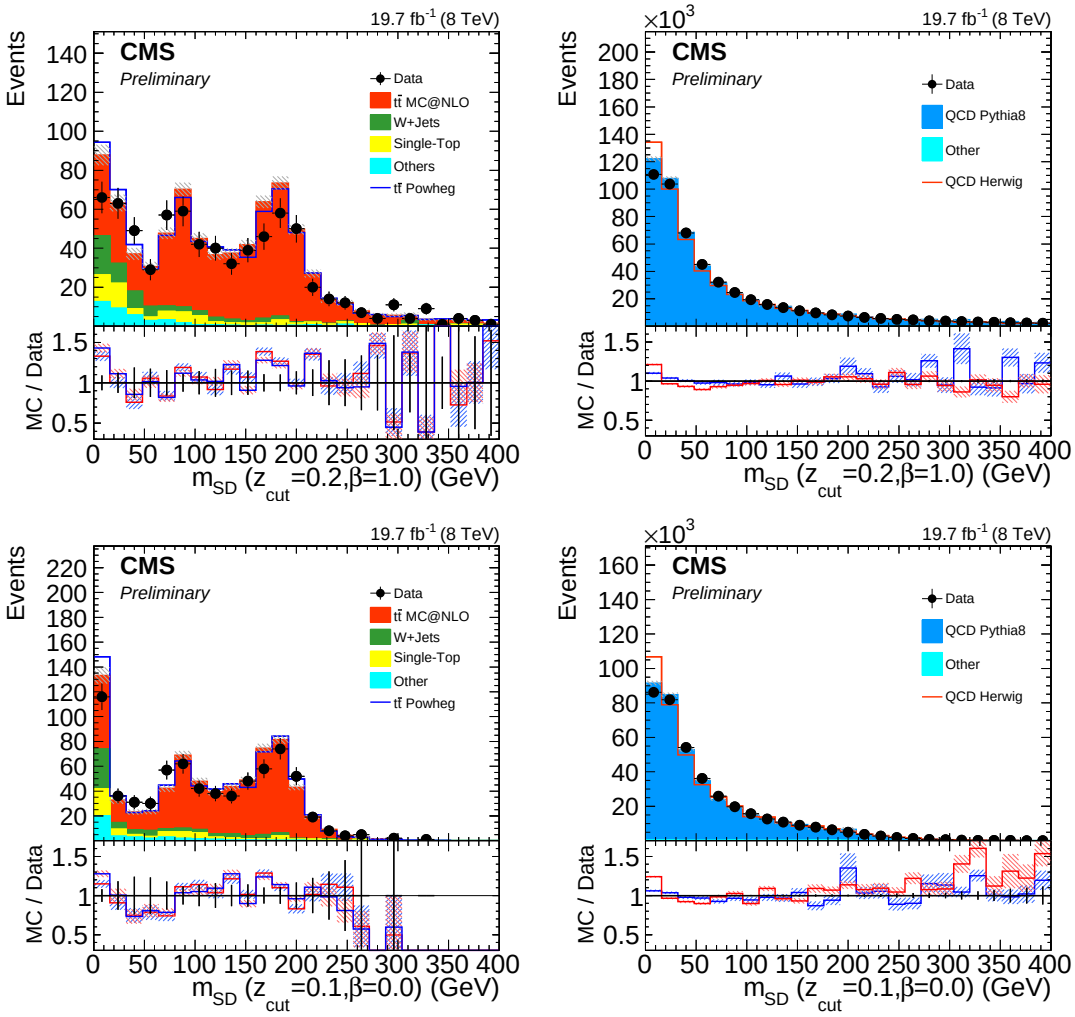

Figure 24: Distribution the jet mass after applying softdrop for the CA15 selection with $z_{\text{cut}} = 0.2$ and $\beta = 1$ (top) and the CA8 selection with $z_{\text{cut}} = 0.1$ and $\beta = 0$ (bottom) for signal jets (left) and background ones (right). Only statistical uncertainties are shown. The ratio of simulation to data is shown at the bottom of each panel. In the case of the signal (background) selection, the ratio to POWHEG (PYTHIA 8) is shown in blue while the one to MC@NLO (HERWIG++) is shown in red. The hashed bands depict the statistical uncertainty of the simulated samples, whereas the vertical bars show the statistical uncertainties of data.

subjettiness ($\tau_3/\tau_2 < 0.5\,(0.6)$ for CA8 (CA15) jets are shown in Fig. 25. For the CA8 selection (top row), the efficiency rises up to 20% and the obtained mistag rate is between 1–2%. The efficiency is well modeled by POWHEG + PYTHIA and is in good agreement with the data within the statistical and systematic uncertainties. The mistag rate is better modeled with PYTHIA, and HERWIG++ overestimates the mistag rate by 20%.

The efficiency and mistag rate for CA15 jets are shown in Fig. 25 (bottom row). The efficiency increases up to 20% and the mistag rate is in the whole p_{T} range below 1%. Within the statistical and systematic uncertainties the simulated signal events provide a fair description of the data. The mistag rate is described better by PYTHIA, since in the p_{T} region below 800 GeV HERWIG++ overestimates the mistag rate.

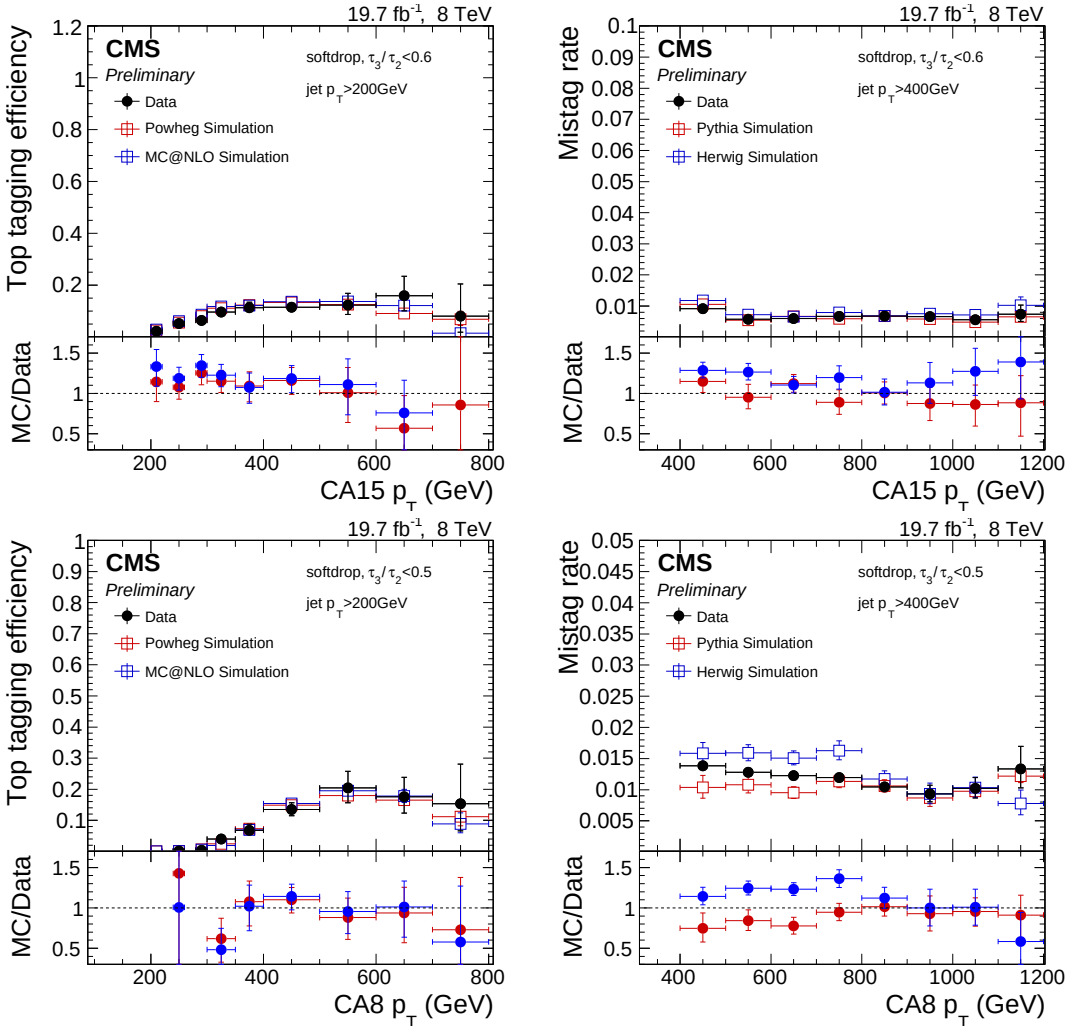

Figure 25: Efficiency (left) and mistag rate (right) of the softdrop based tagging criteria as function of p_{T}, for CA15 jets, $z_{\mathrm{cut}} = 0.2$ and $\beta = 1.0$ (top) and CA8 jets, $z_{\mathrm{cut}} = 0.1$ and $\beta = 0$ (bottom). The error bars show the combined statistical and systematic uncertainties. At the bottom of each panel the ratio of simulation to data is shown.

8 Conclusions

A review of top-tagging techniques aimed at Run II of the LHC is presented. By analyzing the performance with respect to generator truth-level, inside a fiducial region with flat jet transverse momentum and η distributions, it is possible to disentangle kinematic correlations from the discriminating power of individual variables.

Softdrop, pruning and HTT V2 algorithms provide a stable reconstructed mass as a function of top-quark p_{T}. The HTT V2 method also supplies a stable f_{Rec} discriminant. Shower deconstruction provides a single variable offering strong discrimination between signal and background jets. Improvements can be obtained after combination with a groomed mass variable.

When multiple observables are considered to select top-jet candidates, a combination of the HTT V2 decision variables, the n-subjettiness, and b tagging; or alternatively a combination of shower deconstruction, the softdrop mass, the n-subjettiness, and b tagging offer superior performance. A simplified combination of the softdrop mass, the n-subjettiness, and b tagging only has at most 15% lower signal efficiency at the same background rejection, therefore it is recommended as the default technique.

A set of different working points is tested for their stability with respect to p_{T}, η, and the amount of pileup present in the event. In general, a turn-on behavior with a stable plateau at high p_{T} is observed, with the main exception of shower deconstruction where the efficiency and mistag rate rise linearly with p_{T}. This can already be inferred from the one-dimensional distributions where a shift in the shower deconstruction discriminator with p_{T} is observed.

A dependence of the expected background mistag rate on the parton flavor is observed, with bottom quarks showing the highest and gluons showing the lowest mistag rate. This is consistent with the different shapes of these variables in the two-dimensional n-subjettiness versus groomed mass plane.

In order to measure top-tagging efficiencies in an analysis setting, we compare the simulations with data collected at a center-of-mass energy of 8 TeV. To measure the efficiency and mistag rate in data, two selections were designed to enrich a sample with semi-leptonic $\mathrm{t\bar{t}}$ events and the other with di-jet ones. After reweighting the p_{T} distribution of the leading jet for the background selection, we examine the decision variables of the tagging algorithms using data events.

Overall the agreement between simulation and data is good. The discriminating variables as well as the efficiencies and mistag rates are well modeled by MC@NLO and POWHEG for the signal and by HERWIG++ and PYTHIA 8 for the backgrounds.

References

[1] C. T. Hill, "Topcolor: Top quark condensation in a gauge extension of the standard model", *Phys. Lett.* **B266** (1991) 419–424, doi:10.1016/0370-2693(91)91061-Y.

[2] C. T. Hill, "Topcolor assisted technicolor", *Phys. Lett.* **B345** (1995) 483–489, doi:10.1016/0370-2693(94)01660-5, arXiv:hep-ph/9411426.

[3] R. M. Harris, C. T. Hill, and S. J. Parke, "Cross-section for topcolor Z'_t decaying to $t\bar{t}$", arXiv:hep-ph/9911288.

[4] L. Randall and R. Sundrum, "An Alternative to compactification", *Phys. Rev. Lett.* **83** (1999) 4690–4693, doi:10.1103/PhysRevLett.83.4690, arXiv:hep-th/9906064.

[5] ATLAS Collaboration, "A search for $t\bar{t}$ resonances in lepton+jets events with highly boosted top quarks collected in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector", *JHEP* **1209** (2012) 041, doi:10.1007/JHEP09(2012)041, arXiv:1207.2409.

[6] ATLAS Collaboration, "A search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 14 fb^1 of pp collisions at $\sqrt{s} = 8$ TeV",.

[7] ATLAS Collaboration, "A search for $t\bar{t}$ resonances using lepton-plus-jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP* **08** (2015) 148, doi:10.1007/JHEP08(2015)148, arXiv:1505.07018.

[8] CMS Collaboration, "Search for anomalous $t\bar{t}$ production in the highly-boosted all-hadronic final state", *JHEP* **1209** (2012) 029, doi:10.1007/JHEP09(2012)029, arXiv:1204.2488.

[9] CMS Collaboration, "Search for Anomalous Top Quark Pair Production in the Boosted All-Hadronic Final State using pp Collisions at $\sqrt{s} = 8$ TeV",.

[10] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new Higgs search channel at the LHC", *Phys. Rev. Lett.* **100** (2008) 242001, doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.

[11] D. Krohn, J. Thaler, and L.-T. Wang, "Jet Trimming", *JHEP* **1002** (2010) 084, doi:10.1007/JHEP02(2010)084, arXiv:0912.1342.

[12] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches", *Phys. Rev.* **D81** (2010) 094023, doi:10.1103/PhysRevD.81.094023, arXiv:0912.0033.

[13] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", *JHEP* **1405** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.

[14] S. D. Ellis et al., "Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure", *Phys. Rev. Lett.* **108** (2012) 182003, doi:10.1103/PhysRevLett.108.182003, arXiv:1201.1914.

[15] J. Thaler and K. Van Tilburg, "Identifying Boosted Objects with N-subjettiness", *JHEP* **1103** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.

[16] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions", *Nucl. Phys.* **B406** (1993) 187–224, doi:10.1016/0550-3213(93)90166-M.

[17] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, "Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks", *Phys. Rev. Lett.* **101** (2008) 142001, doi:10.1103/PhysRevLett.101.142001, arXiv:0806.0848.

[18] CMS Collaboration, "A Cambridge-Aachen (C-A) based Jet Algorithm for boosted to-jet tagging", CMS Physics Analysis Summary CMS-PAS-JME-09-001, 2009.

[19] T. Plehn, G. P. Salam, and M. Spannowsky, "Fat Jets for a Light Higgs", *Phys. Rev. Lett.* **104** (2010) 111801, doi:10.1103/PhysRevLett.104.111801, arXiv:0910.5472.

[20] T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, "Stop Reconstruction with Tagged Tops", *JHEP* **1010** (2010) 078, doi:10.1007/JHEP10(2010)078, arXiv:1006.2833.

[21] G. Kasieczka et al., "Resonance Searches with an Updated Top Tagger", arXiv:1503.05921.

[22] D. E. Soper and M. Spannowsky, "Finding physics signals with shower deconstruction", *Phys. Rev.* **D84** (2011) 074002, doi:10.1103/PhysRevD.84.074002, arXiv:1102.3480.

[23] D. E. Soper and M. Spannowsky, "Finding top quarks with shower deconstruction", *Phys. Rev.* **D87** (2013) 054012, doi:10.1103/PhysRevD.87.054012, arXiv:1211.3140.

[24] CMS Collaboration, "Boosted Top Jet Tagging at CMS", CMS Physics Analysis Summary CMS-PAS-JME-13-007, 2014.

[25] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[26] T. Sjstrand et al., "An Introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159–177, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[27] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[28] R. D. Ball et al., "Parton distributions with LHC data", *Nucl. Phys.* **B867** (2013) 244–289, doi:10.1016/j.nuclphysb.2012.10.003, arXiv:1207.1303.

[29] S. Frixione, P. Nason, and G. Ridolfi, "A Positive-Weight Next-to-Leading-Order Monte Carlo for Heavy Flavour Hadroproduction", *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.

[30] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[31] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[32] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[33] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.

[34] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", *JHEP* **06** (2002) 029, doi:10.1088/1126-6708/2002/06/029, arXiv:hep-ph/0204244.

[35] G. Corcella et al., "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)", *JHEP* **01** (2001) 010, doi:10.1088/1126-6708/2001/01/010, arXiv:hep-ph/0011363.

[36] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching Matrix Elements and Shower Evolution for Top-Quark Production in Hadronic Collisions", *JHEP* **01** (2007) 013.

[37] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.

[38] M. Bahr et al., "Herwig++ Physics and Manual", *Eur. Phys. J.* **C58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.

[39] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables", *Phys. Rev.* **D78** (2008) 013004, doi:10.1103/PhysRevD.78.013004, arXiv:0802.0007.

[40] H.-L. Lai et al., "New parton distributions for collider physics", *Phys. Rev.* **D82** (2010) 074024, doi:10.1103/PhysRevD.82.074024, arXiv:1007.2241.

[41] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "MRST2001: Partons and α_s from precise deep inelastic scattering and Tevatron jet data", *Eur. Phys. J.* **C23** (2002) 73–87, doi:10.1007/s100520100842, arXiv:hep-ph/0110215.

[42] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.

[43] CMS Collaboration, "Jet performance in CMS", *PoS* **EPS-HEP2013** (2013) 433.

[44] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.

[45] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[46] CMS Collaboration, "V Tagging Observables and Correlations", CMS Physics Analysis Summary CMS-PAS-JME-14-002, 2014.

[47] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV", *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[48] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

[49] CMS Collaboration, "Pileup Jet Identification", CMS Physics Analysis Summary CMS-PAS-JME-13-005, 2013.

[50] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s}$ = 8 TeV", *JINST* **10** (2015), no. 02, P02006, doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.

[51] CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** (2011), no. 09, P09001.

[52] CMS Collaboration, "Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, $\mathrm{t\bar{t}}$ and boosted topology events", CMS Physics Analysis Summary CMS-PAS-BTV-13-001, 2013.

[53] A. Hoecker et al., "TMVA: Toolkit for Multivariate Data Analysis", *PoS* **ACAT** (2007) 040, arXiv:physics/0703039.

[54] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", CMS Physics Analysis Summary CMS-PAS-BTV-11-001, 2001.

[55] CMS Collaboration, "Searches for new physics using the $\mathrm{t\bar{t}}$ invariant mass distribution in pp collisions at $\sqrt{s} = 8$ TeV", *Phys. Rev. Lett.* **111** (2013), no. 21, 211804, doi:10.1103/PhysRevLett.111.211804, arXiv:1309.2030.

[56] CMS Collaboration, "Search for Resonant $\mathrm{t\bar{t}}$ Production in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV", arXiv:1506.03062.