Eur. Phys. J. C manuscript No.
(will be inserted by the editor)

Towards an Understanding of the Correlations in Jet Substructure

Report of BOOST2013, hosted by the University of Arizona, 12^{th}-16^{th} of August 2013.

First Author[1], **Second Author**[2,3]

[1]First Address, Street, City, Country
[2]Second Address, Street, City, Country
[3]*Present Address:* Street, City, Country

Received: date / Accepted: date

Abstract Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here. Insert your abstract here.

1 Introduction

The characteristic feature of collisions at the LHC is a center-of-mass energy, 7 TeV in 2010 and 2011, of 8 TeV in 2012, and near 14 TeV with the start of the second phase of operation in 2015, that is large compared to even the heaviest of the known particles. Thus these particles (and also previously unknown ones) will often be produced at the LHC with substantial boosts. As a result, when decaying hadronically, these particles will not be observed as multiple jets in the detector, but rather as a single hadronic jet with distinctive internal substructure. This realization has led to a new era of sophistication in our understanding of both standard QCD jets and jets containing the decay of a heavy particle, with an array of new jet observables and detection techniques introduced and studies. To allow the efficient sharing of results from these jet substructure studies a series of BOOST Workshops have been held on a yearly basis: SLAC (2009, [**?**]), Oxford University (2010, [**?**]), Princeton University University (2011, [**?**]), IFIC Valencia (2012 [**?**]), University of Arizona (2013 [**?**]), and, most recently, University College London (2014 [**?**]). After each of these meetings Working Groups have functioned during the following year to generate reports highlighting the most interesting new results, including studies of ever maturing details. Previous BOOST reports can be found at [1–3].

This report from BOOST 2013 thus views the study and implementation of jet substructure techniques as a fairly mature field, and focuses on the question of the correlations between the plethora of observables that have been developed and employed, and their dependence on the underlying jet parameters, especially the jet radius R and jet p_T. Samples of quark-, gluon-, W- and Top-initiated jets are reconstructed at the particle-level using FASTJET**[REF]**, and the performance, in terms of separating signal from background, of various groomed jet masses and jet substructure observables investigated through Receiver Operating Characteristic (ROC) curves, which show the efficiency to "tag" the signal as a function of the efficiency (or rejection, being 1/efficiency) to "tag" the background. We investigate the separation of a quark signal from a gluon background (q/g tagging), a W signal from a gluon background (W-tagging) and a Top signal from a mixed quark/gluon QCD background (Top-tagging). In the case of Top-tagging, we also investigate the performance of dedicated Top-tagging algorithms, the HepTopTagger**[REF]**and John Hopkins Tagger**[REF]**. Using multivariate techniques, we study the degree to which the discriminatory information provided by the observables and taggers overlaps, by examining in particular the extent to which the signal-background separation performance increases when two or more variables/taggers are combined, via a Boosted Decision Tree (BDT), into a single discriminant.

The report is organized as follows. In Section 2 we describe the generation of the Monte Carlo event samples that we use in the studies that follow. In Section 3 we detail the jet algorithms, observables and taggers investigated in each section of the report, and in Section 4 the multivariate techniques used to combine the one or more of the observables into single discriminants. In Section 5 we describe the q/g-tagging studies, in Section 6 we describe the W-tagging studies, and in Section 7 we describe the Top-tagging studies. Finally we offer some summary of the studies and general conclusions in Section 8.

2 Monte Carlo Samples

In the below sections the Monte Carlo samples used in the q/g tagging, W tagging and Top tagging sections of this report are described. Note that in all cases the samples used contain no additional proton-proton interactions beyond the hard scatter (no pile-up), and there is no attempt to emulate the degradation in angular and p_T resolution that would result when reconstructing the jets inside a real detector.

2.1 Quark/gluon and W tagging

Samples were generated at $\sqrt{s} = 8$ TeV for QCD dijets, and for W^+W^- pairs produced in the decay of a (pseudo) scalar resonance and decaying hadronically. The QCD events were split into subsamples of gg and $q\bar{q}$ events, allowing for tests of discrimination of hadronic W bosons, quarks, and gluons.

Individual gg and $q\bar{q}$ samples were produced at leading order (LO) using MADGRAPH5, while W^+W^- samples were generated using the JHU GENERATOR to allow for separation of longitudinal and transverse polarizations. Both were generated using CTEQ6L1 PDFs**[REF]**. The samples were produced in exclusive p_T bins of width 100 GeV, with the slicing parameter chosen to be the p_T of any final state parton or W at LO. At the parton-level the p_T bins investigated were 300-400 GeV, 500-600 GeV and 1.0-1.1 TeV. Since no matching was performed, a cut on any parton was equivalent. The samples were then all showered through PYTHIA8 (version 8.176)**[REF]**using the default tune 4C**[REF]**. **ED: Need to report the size of the samples used**

2.2 Top tagging

Samples were generated at $\sqrt{s} = 14$ TeV. Standard Model dijet and top pair samples were produced with SHERPA 2.0.0**[REF]**, with matrix elements of up to two extra partons matched to the shower. The top samples included only hadronic decays and were generated in exclusive p_T bins of width 100 GeV, taking as slicing parameter the maximum of the top/anti-top p_T. The QCD samples were generated with a cut on the leading parton-level jet p_T, where parton-level jets are clustered with the anti-k_t algorithm and jet radii of $R = 0.4, 0.8, 1.2$. The matching scale is selected to be $Q_{\mathrm{cut}} = 40, 60, 80$ GeV for the $p_{T\,\mathrm{min}} = 600, 1000$, and 1500 GeV bins, respectively. For the top samples, 100k events were generated in each bin, while 200k QCD events were generated in each bin.

3 Jet Algorithms and Substructure Observables

In this section, we define the jet algorithms and observables used in our analysis. Over the course of our study, we considered a larger set of observables, but for the final analysis, we eliminated redundant observables for presentation purposes. In Sections 3.1, 3.2, 3.3 and 3.4 we first describe the various jet algorithms, groomers, taggers and other substructure variables used in these studies.

3.1 Jet Clustering Algorithms

Jet clustering: Jets were clustered using sequential jet clustering algorithms**[REF]**implemented in FASTJET 3.0.3. Final state particles i, j are assigned a mutual distance d_{ij} and a distance to the beam, $d_{i\mathrm{B}}$. The particle pair with smallest d_{ij} are recombined and the algorithm repeated until the smallest distance is instead the distance to the beam, $d_{i\mathrm{B}}$, in which case i is set aside and labelled as a jet. The distance metrics are defined as

$$d_{ij} = \min(p_{Ti}^{2\gamma}, p_{Tj}^{2\gamma})\,\frac{\Delta R_{ij}^2}{R^2}, \tag{1}$$

$$d_{i\mathrm{B}} = p_{Ti}^{2\gamma}, \tag{2}$$

where $\Delta R_{ij}^2 = (\Delta\eta)^2 + (\Delta\phi)^2$. In this analysis, we use the anti-k_t algorithm ($\gamma = -1$), the Cambridge/Aachen (C/A) algorithm ($\gamma = 0$)**[REF]**, and the k_t algorithm ($\gamma = 1$)**[REF]**, each of which has varying sensitivity to soft radiation in defining the jet.

Qjets: We also perform non-deterministic jet clustering**[REF]**. Instead of always clustering the particle pair with smallest distance d_{ij}, the pair selected for combination is chosen probabilistically according to a measure

$$P_{ij} \propto e^{-\alpha\,(d_{ij} - d_{\min})/d_{\min}}, \tag{3}$$

where $d_{\min}$ is the minimum distance for the usual jet clustering algorithm at a particular step. This leads to a different cluster sequence for the jet each time the Qjet algorithm is used, and consequently different substructure properties. The parameter α is called the rigidity and is used to control how sharply peaked the probability distribution is around the usual, deterministic value. The Qjets method uses statistical analysis of the resulting distributions to extract more information from the jet than can be found in the usual cluster sequence. We use $\alpha = 0.1$ and 25 trees per event for all the studies presented here.

3.2 Jet Grooming Algorithms

Pruning: Given a jet, re-cluster the constituents using the C/A algorithm. At each step, proceed with the merger as usual unless both

$$\frac{\min(p_{Ti}, p_{Tj})}{p_{Tij}} < z_{\mathrm{cut}} \text{ and } \Delta R_{ij} > \frac{2m_j}{p_{Tj}} R_{\mathrm{cut}}, \tag{4}$$

in which case the merger is vetoed and the softer branch discarded. The default parameters used for pruning**[REF]**in this study are $z_{\rm cut} = 0.1$ and $R_{\rm cut} = 0.5$. One advantage of pruning is that the thresholds used to veto soft, wide-angle radiation scale with the jet kinematics, and so the algorithm is expected to perform comparably over a wide range of momenta.

Trimming: Given a jet, re-cluster the constituents into subjets of radius $R_{\rm trim}$ with the k_t algorithm. Discard all subjets i with

$$p_{Ti} < f_{\rm cut}\, p_{TJ}. \tag{5}$$

The default parameters used for trimming**[REF]**in this study are $R_{\rm trim} = 0.2$ and $f_{\rm cut} = 0.03$.

Filtering:[REF] Given a jet, re-cluster the constituents into subjets of radius $R_{\rm filt}$ with the C/A algorithm. Re-define the jet to consist of only the hardest N subjets, where N is determined by the final state topology and is typically one more than the number of hard prongs in the resonance decay (to include the leading final-state gluon emission). **ED: Do we actually use filtering as described here anywhere? (BS: Yes, it is used in the HEPTopTagger.)**

Soft drop: Given a jet, re-cluster all of the constituents using the C/A algorithm. Iteratively undo the last stage of the C/A clustering from j into subjets j_1, j_2. If

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} < z_{\rm cut} \left(\frac{\Delta R_{12}}{R}\right)^{\beta}, \tag{6}$$

discard the softer subjet and repeat. Otherwise, take j to be the final soft-drop jet**[REF]**. Soft drop has two input parameters, the angular exponent β and the soft-drop scale $z_{\rm cut}$, with default value $z_{\rm cut} = 0.1$. **ED: Soft-drop actually functions as a tagger when $\beta = -1$**

3.3 Jet Tagging Algorithms

Modified Mass Drop Tagger: Given a jet, re-cluster all of the constituents using the C/A algorithm. Iteratively undo the last stage of the C/A clustering from j into subjets j_1, j_2 with $m_{j_1} > m_{j_2}$. If either

$$m_{j_1} > \mu\, m_j \ \text{or} \ \frac{\min(p_{T1}^2, p_{T2}^2)}{m_j^2} \Delta R_{12}^2 < y_{\rm cut}, \tag{7}$$

then discard the branch with the smaller transverse mass $m_T = \sqrt{m_i^2 + p_{Ti}^2}$, and re-define j as the branch with the larger transverse mass. Otherwise, the jet is tagged. If declustering continues until only one branch remains, the jet is untagged. In this study we use by default $\mu = 1.0$ and $y_{\rm cut} = 0.1$.

Johns Hopkins Tagger: Re-cluster the jet using the C/A algorithm. The jet is iteratively de-clustered, and at each step the softer prong is discarded if its $p_{\rm T}$ is less than $\delta_p\, p_{\rm T\,jet}$. This continues until both prongs are harder than the $p_{\rm T}$ threshold, both prongs are softer than the $p_{\rm T}$ threshold, or if they are too close ($|\Delta\eta_{ij}| + |\Delta\phi_{ij}| < \delta_R$); the jet is rejected if either of the latter conditions apply. If both are harder than the $p_{\rm T}$ threshold, the same procedure is applied to each: this results in 2, 3, or 4 subjets. If there exist 3 or 4 subjets, then the jet is accepted: the top candidate is the sum of the subjets, and W candidate is the pair of subjets closest to the W mass. The output of the tagger is m_t, m_W, and $\theta_{\rm h}$, a helicity angle defined as the angle, measured in the rest frame of the W candidate, between the top direction and one of the W decay products. The two free input parameters of the John Hopkins tagger in this study are δ_p and δ_R, defined above.

HEPTopTagger: Re-cluster the jet using the C/A algorithm. The jet is iteratively de-clustered, and at each step the softer prong is discarded if $m_1/m_{12} > \mu$ (there is not a significant mass drop). Otherwise, both prongs are kept. This continues until a prong has a mass $m_i < m$, at which point it is added to the list of subjets. Filter the jet using $R_{\rm filt} = \min(0.3, \Delta R_{ij})$, keeping the five hardest subjets (where ΔR_{ij} is the distance between the two hardest subjets). Select the three subjets whose invariant mass is closest to m_t. The output of the tagger is m_t, m_W, and $\theta_{\rm h}$, a helicity angle defined as the angle, measured in the rest frame of the W candidate, between the top direction and one of the W decay products. The two free input parameters of the HEPTopTagger in this study are m and μ, defined above.

Top Tagging with Pruning: For comparison with the other top taggers, we add a W reconstruction step to the trimming algorithm described above. A W candidate is found as follows: if there are two subjets, the highest-mass subjet is the W candidate (because the W prongs end up clustered in the same subjet); if there are three subjets, the two subjets with the smallest invariant mass comprise the W candidate. In the case of only one subjet, no W is reconstructed.

Top Tagging with Trimming: For comparison with the other top taggers, we add a W reconstruction step to the trimming algorithm described above. A W candidate is found as follows: if there are two subjets, the highest-mass subjet is the W candidate (because the W prongs end up clustered in the same subjet); if there are three subjets, the two subjets with the smallest invariant mass comprise the W candidate. In the case of only one subjet, no W is reconstructed.

3.4 Other Jet Substructure Observables

Qjet mass volatility: As described above, Qjet algorithms re-cluster the same jet non-deterministically to obtain a collection of interpretations of the jet. For each jet interpretation, the pruned jet mass is computed with the default pruning parameters. The mass volatility, Γ_{Qjet}, is defined as

$$\Gamma_{\mathrm{Qjet}} = \frac{\sqrt{\langle m_J^2 \rangle - \langle m_J \rangle^2}}{\langle m_J \rangle}, \tag{8}$$

where averages are computed over the Qjet interpretations.

N**-subjettiness:** N-subjettiness**[REF]**quantifies how well the radiation in the jet is aligned along N directions. To compute N-subjettiness, $\tau_N^{(\beta)}$, one must first identify N axes within the jet. Then,

$$\tau_N = \frac{1}{d_0} \sum_i p_{Ti} \min\left(\Delta R_{1i}^{\beta}, \ldots, \Delta R_{Ni}^{\beta}\right), \tag{9}$$

where distances are between particles i in the jet and the axes,

$$d_0 = \sum_i p_{Ti} R^{\beta} \tag{10}$$

and R is the jet clustering radius. The exponent β is a free parameter. There is also some choice in how the axes used to compute N-subjettiness are determined. The optimal configuration of axes is the one that minimizes N-subjettiness; recently, it was shown that the "winner-takes-all" (WTA) axes can be easily computed and have superior performance compared to other minimization techniques**[REF]**. We use both the WTA and one-pass k_t optimization axes in our analyses.

A more powerful discriminant is often the ratio,

$$\tau_{N,N-1} \equiv \frac{\tau_N}{\tau_{N-1}}. \tag{11}$$

While this is not an infrared-collinear (IRC) safe observable, it is calculable**[REF]**and can be made IRC safe with a loose lower cut on τ_{N-1}.

Energy correlation functions: The transverse momentum version of the energy correlation functions are defined as**[REF]**:

$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in j} \left(\prod_{a=1}^{N} p_{Ti_a}\right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^{N} \Delta R_{i_b i_c}\right)^{\beta}, \tag{12}$$

where i is a particle inside the jet. It is preferable to work in terms of dimensionless quantities, particularly the energy correlation function double ratio:

$$C_N^{(\beta)} = \frac{\mathrm{ECF}(N+1,\beta)\,\mathrm{ECF}(N-1,\beta)}{\mathrm{ECF}(N,\beta)^2}. \tag{13}$$

This observable measures higher-order radiation from leading-order substructure.

4 Multivariate Analysis Techniques

Multivariate techniques are used to combine variables into an optimal discriminant. In all cases variables are combined using a boosted decision tree (BDT) as implemented in the TMVA package [4]. We use the BDT implementation including gradient boost. An example of the BDT settings are as follows:

- NTrees=1000
- BoostType=Grad
- Shrinkage=0.1
- UseBaggedGrad=F
- nCuts=10000
- MaxDepth=3
- UseYesNoLeaf=F
- nEventsMin=200

Exact parameter values are chosen to best reduce the effect of overtraining. **ED: Can we describe a bit more the tests we do to ensure that we are not suffering from overtraining?**

5 Quark-Gluon Discrimination

In this section, we examine the differences between quark- and gluon-initiated jets in terms of substructure variables, and to determine to what extent these variables are correlated. Along the way, we provide some theoretical understanding of these observables and their performance. The motivation for these studies comes not only from the desire to "tag" a jet as originating from a quark or gluon, but also to improve our understanding of the quark and gluon components of the QCD backgrounds relative to boosted resonances. While recent studies have suggested that quark/gluon tagging efficiencies depend highly on the Monte Carlo generator used**[REF]**, we are more interested in understanding the scaling performance with p_T and R, and the correlations between observables, which are expected to be treated consistently within a single shower scheme.

5.1 Methodology

These studies use the qq and gg MC samples, described previously in Section 2. The showered events were clustered with FASTJET 3.03**[REF]**using the anti-k_{T} algorithm**[REF]**with jet radii of $R = 0.4$, 0.8, 1.2. In both signal (quark) and background (gluon) samples, an upper and lower cut on the leading jet p_T is applied after showering/clustering, to ensure similar p_T spectra for signal and background in each p_T bin. The bins in leading jet p_T that are considered are 300-400 GeV, 500-600 GeV, 1.0-1.1 TeV, for the 300-400 GeV, 500-600 GeV, 1.0-1.1 TeV parton p_T slices respectively. Various jet grooming approaches are applied to the jets, as described in Section 3.4. Only leading and subleading jets in each sample are used. The following observables are studied in this section:

- The ungroomed jet mass, m.
- 1-subjettiness, τ_1^β with $\beta = 1, 2$. The N-subjettiness axes are computed using one-pass k_t axis optimization.
- 1-point energy correlation functions, $C_1^{(\beta)}$ with $\beta = 1, 2$.
- The pruned Qjet mass volatility, $\Gamma_{\rm Qjet}$.
- The number of constituents ($N_{\rm constits}$).

5.2 Single Variable Discrimination

Figure 1 shows the mass of jets in the quark and gluon samples when using different groomers, and the ungroomed jet mass, for jets with R=0.8 and in the $p_T = 500 - 600$ GeV bin. Qualitatively, the application of grooming shifts the mass distributions towards lower values when compared to the ungroomed mass, as expected. No clear gain in discrimination can be seen, and for certain grooming parameters, such as the use of soft drop with $\beta = -1$ a clear loss in discrimination power is observed; this is because the soft-drop condition for $\beta = -1$ discards collinear radiation, and the differences between quarks and gluons are manifest in the collinear structure (spin, splitting functions, etc.).

The quark and gluon distributions of different substructure variables are shown in Figure 2. Among those considered, one can see by eye that $n_{\rm constits}$ provides the highest separation power, followed by $C_1^{\beta=0}$ and $C_1^{\beta=1}$, as was also found by the CMS and ATLAS Collaborations**[REF]**.

To more quantitatively study the power of each observable as a discriminator for quark/gluon tagging, ROC curves are built by scanning each distribution and plotting the background efficiency (to select gluon jets) vs. the signal efficiency (to select quark jets). Figure 3 shows these ROC curves for all of the substructure variables shown in Figure 2, along with the ungroomed mass, representing the best performing mass variable, for R=0.4, 0.8 and 1.2 jets in the $p_T = 300 - 400$ GeV bin. In addition, the ROC curve for a tagger built from a BDT combination of all the variables (see Section 4) is shown. Clearly, $n_{\rm constits}$ is the best performing variable for all Rs, even though $C_1^{\beta=0}$ is close, particularly for R=0.8. Most other variables have similar performance, except $\Gamma_{\rm Qjet}$, which shows significantly worse discrimination (this may be due to our choice of rigidity $\alpha = 0.1$, with other studies suggesting that a smaller value, such as $\alpha = 0.01$ produces better results**[REF]**). The combination of all variables shows somewhat better discrimination.

We now examine how performance of masses and substructure observables changes with p_T and R. For jet masses, few variations are observed as the radius parameter of the jet reconstruction is increased in the two highest p_T bins; this is because the radiation is more collimated and the dependence on R is consequently smaller. However, for the $300 -$ 400 GeV bin, the use of small-R jets produces a shift in the mass distributions towards lower values, so that large-R jet masses are more stable with p_T and small-R jet masses are smaller at low-p_T as expected from the spatial constraints imposed by the R parameter. These statements are explored more quantitatively later in this section. (**BS: Do we have plots for this?**)

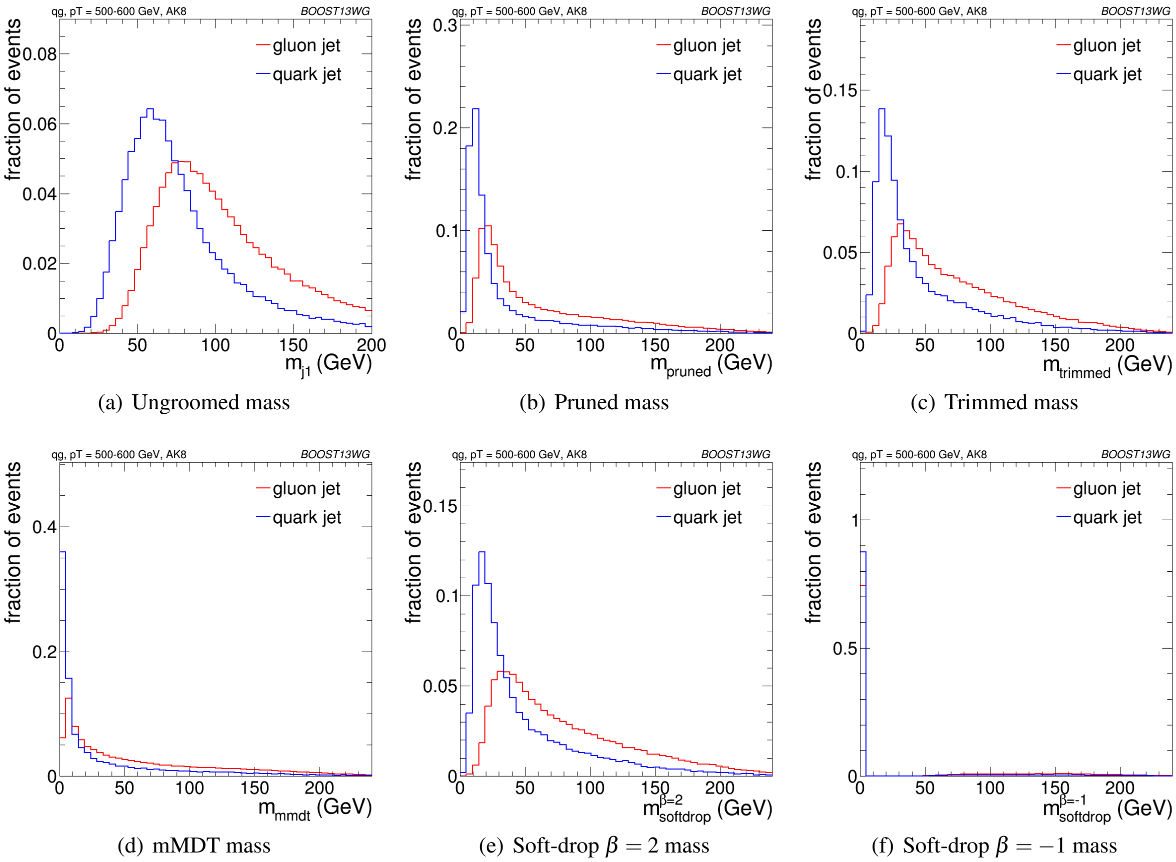
The evolution of some of the substructure variable distributions with p_T and R is less trivial than for the jet masses. In particular, changing the R parameter at high p_T changes significantly the C_a^β for $\beta > 0$ and the $n_{\rm constits}$ distributions, while leaving all other distributions qualitatively unchanged. This is illustrated in Figure 4 for $\beta = 0$ and $\beta = 1$ using $a = 1$ in both cases for jets with $p_T = 1.0 - 1.1$ TeV.

The shift towards lower values with changing R is evident for the $C_1^{\beta=1}$ distributions, while the stability of $C_1^{\beta=0}$ can also be observed. These features are present in all p_T bins studied, but are even more pronounced for lower p_T bins. The shape of the Q-jet volatility distribution shows some non-trivial shape that deserves some explanation. Two peaks are observed, one at low volatility values and one at mid-volatility. These peaks are generated by two somewhat distinct populations. The high volatility peak arises from jets that get their mass primarily from soft (and sometimes wide-angle) emissions. The removal of some of the constituents when building Q-jets thus changes the mass significantly, increasing the volatility. The lower volatility peak corresponds to jets for which mass is generated by a hard emission, which makes the fraction of Q-jets that change the mass significantly to be smaller. Since the probability of a hard emission is proportional to the colour charge (squared), the volatility peak is higher for gluon jets by about the colour factor C_A/C_F.

In summary, the overall discriminating power between quarks and gluons decreases with increasing R due to the reduction in the amount of out-of-cone radiation differences and and increased contamination from the underlying event (**BS: is this ok?**). The broad performance features discussed for this p_T bin also apply to the higher p_T bins. These is further quantified in the next section.

5.3 Combined Performance and Correlations

The quark/gluon tagging performance can be further improved over cuts on single observables by combining multiple observables in a BDT; due to the challenging nature of q/g-tagging, any improvement in performance with multivariable techniques could be critical for certain analyses, and the improvement could be more substantial in data than the marginal benefit found in MC and shown in Fig. 3. Fur-



(a) Ungroomed mass (b) Pruned mass (c) Trimmed mass

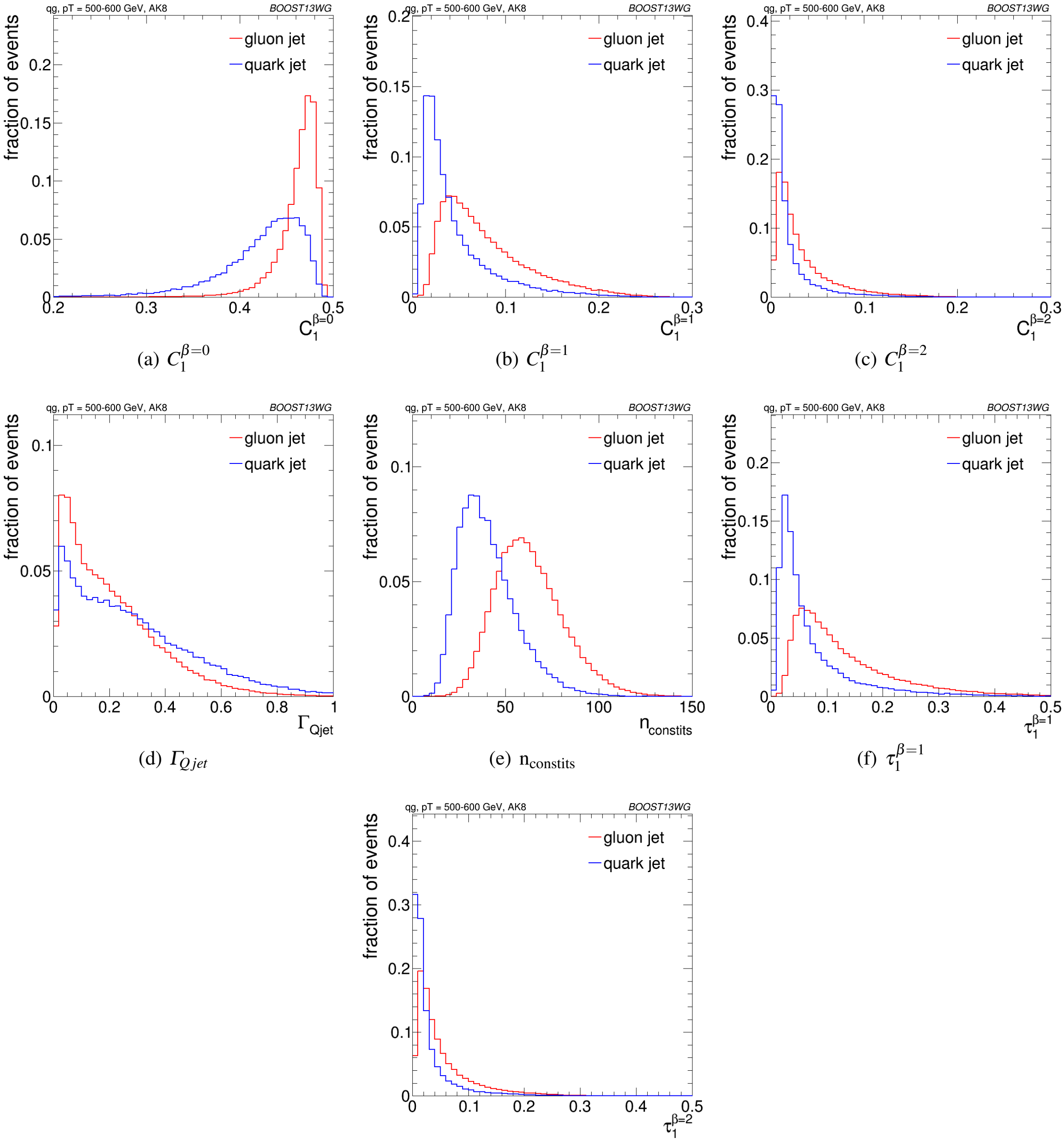
(d) mMDT mass (e) Soft-drop $\beta = 2$ mass (f) Soft-drop $\beta = -1$ mass

Fig. 1 Comparisons of ungroomed and groomed quark and gluon mass distributions for leading jets in the $p_T = 500-600$ GeV bin using the anti-k_T R=0.8 algorithm.

thermore, insight can be gained into the features allowing for quark/gluon discrimination if the origin of the improvement is understood. To quantitatively study this improvement, we build quark/gluon taggers from every pair-wise combination of variables studied in the previous section for comparison with the all-variable combination.

In order to quantitatively study the value of each variable for quark/gluon tagging, we study the gluon rejection, defined as $1/\varepsilon_{\text{gluon}}$, at a fixed quark selection efficiency of 50%, using jets with $p_T = 1-1.1$ TeV and for different R parameters. Figure 5 shows the gluon rejection for each pair-wise combination. The pair-wise gluon rejection at 50% quark efficiency can be compared to the single-variable values shown along the diagonal. The gluon rejection for the BDT all-variable combination is also shown on the bottom right of each plot. As already observed in the previous section, n_{constits} is the most powerful single variable and $C_1^{(\beta=0)}$ follows closely. However, the gains are largely correlated; the combined performance of n_{constits} and $C_1^{(\beta=0)}$ is generally poorer than combinations of n_{constits} with other jet substructure observables such as τ_1. Interestingly, in spite of the high correlation between n_{constits} and $C_1^{(\beta=0)}$, the two-variable combinations of n_{constits} generally fare worse than two-variable combinations with $C_1^{(\beta=0)}$. In particular, the combinations of $\tau_1^{\beta=1}$ or $C_1^{(\beta=1)}$ with n_{constits} are capable of getting very close to the rejection achievable through the use of all variables for $R = 0.4$ and $R = 0.8$.

Tagger performance is generally better at small R. The overall loss in performance with increasing R can be seen in most single variables we study; this is expected, since more of the parton radiation is captured in the jet and more contamination from underlying event occurs, suppressing the differences between q/g jets. The principal exceptions are $C_1^{(\beta=0)}$ and the Q-jet mass volatility, which are both quite resilient to increasing R. For $C_1^{(\beta=0)}$, this is due to the fact that the exponent on ΔR is zero, and so soft radiation at the periphery of the jet does not substantially change the distribution; as a result, the performance is largely independent of R. Similarly, the soft radiation distant from the jet centre will be vetoed during pruning regardless of the cluster sequence, and so the R-dependence of Γ_{Qjet} is not significant. (**BS: Check my logic?**) Their combination, however, does perform slightly worse at larger R. (**BS: I don't understand this, but it is a $\sim 10\%$ effect, so maybe not too**



(a) $C_1^{\beta=0}$ (b) $C_1^{\beta=1}$ (c) $C_1^{\beta=2}$

(d) Γ_{Qjet} (e) $\mathrm{n}_{\mathrm{constits}}$ (f) $\tau_1^{\beta=1}$

(g) $\tau_1^{\beta=2}$

Fig. 2 Comparisons of quark and gluon distributions of different substructure variables for leading jets in the $p_T = 500-600$ GeV bin using the anti-k_{T} R=0.8 algorithm.

significant?). By contrast, $\tau_1^{(\beta=2)}$ and $C_1^{(\beta=2)}$ are particularly sensitive to increasing R since, for $\beta = 2$, large-angle emissions are given a larger weight.

These observations are qualitatively similar across all ranges of p_T. Quantitatively, however, there is a loss of rejection power for the taggers made of a combination of variables as the p_T decreases. This can be observed in Fig. 6 for anti-k_{T} R=0.4 jets of different p_Ts. Clearly, most single variables retain their gluon rejection potential at lower p_T. However, when combined with other variables, the highest performing pairwise combinations lose ground with respect to other pairwise combinations. This is also reflected in the rejection of the tagger that uses a combination of all variables, which is lower at lower p_Ts. **[do we understand this?] (BS: This is a bit of a guess, but could it be that there is typically less radiation for low p_T, and so you're more sensitive to fluctuations; since you have less access**

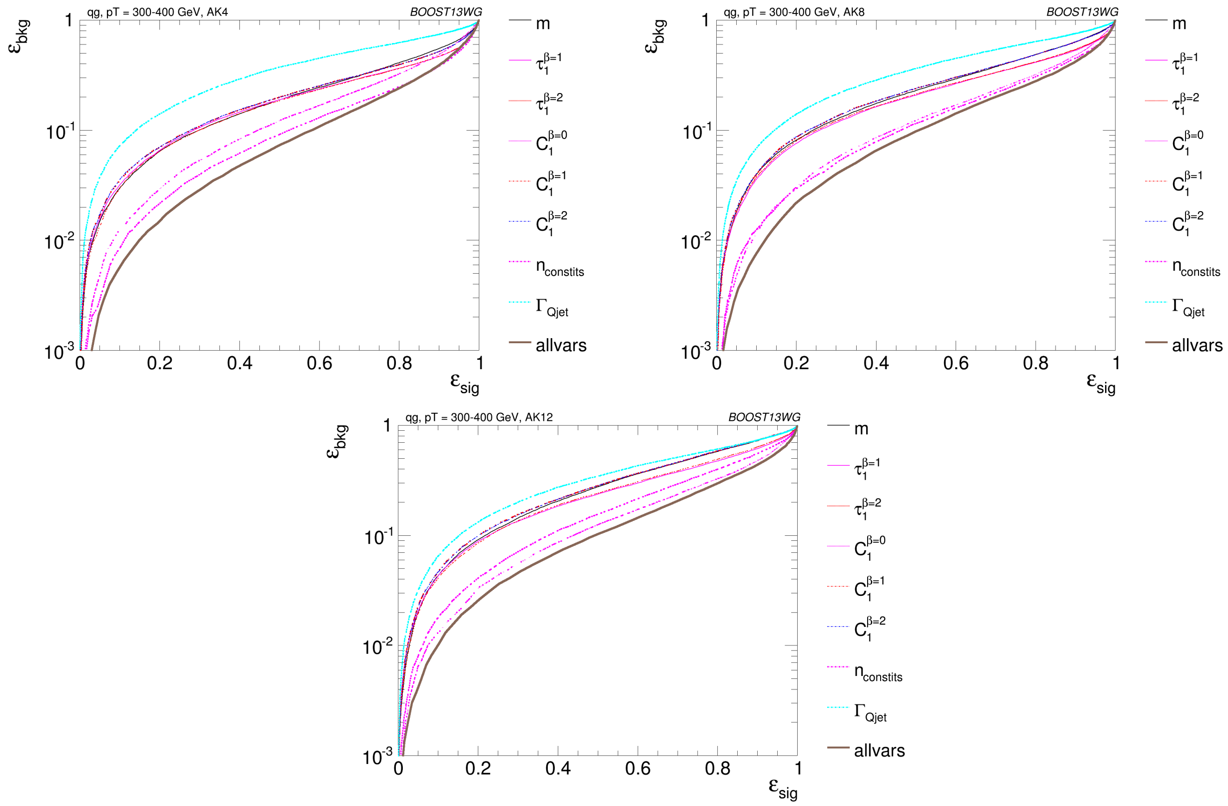


Fig. 3 The ROC curve for all single variables considered for quark-gluon discrimination in the p_T 300-400 GeV bin using the anti-k_T R=0.4, 0.8 and 1.2 algorithm.**ED: Hard to tell the lines on the plots apart**

to information, combinations of observables perform less well than at high p_T.)

6 Boosted W-Tagging

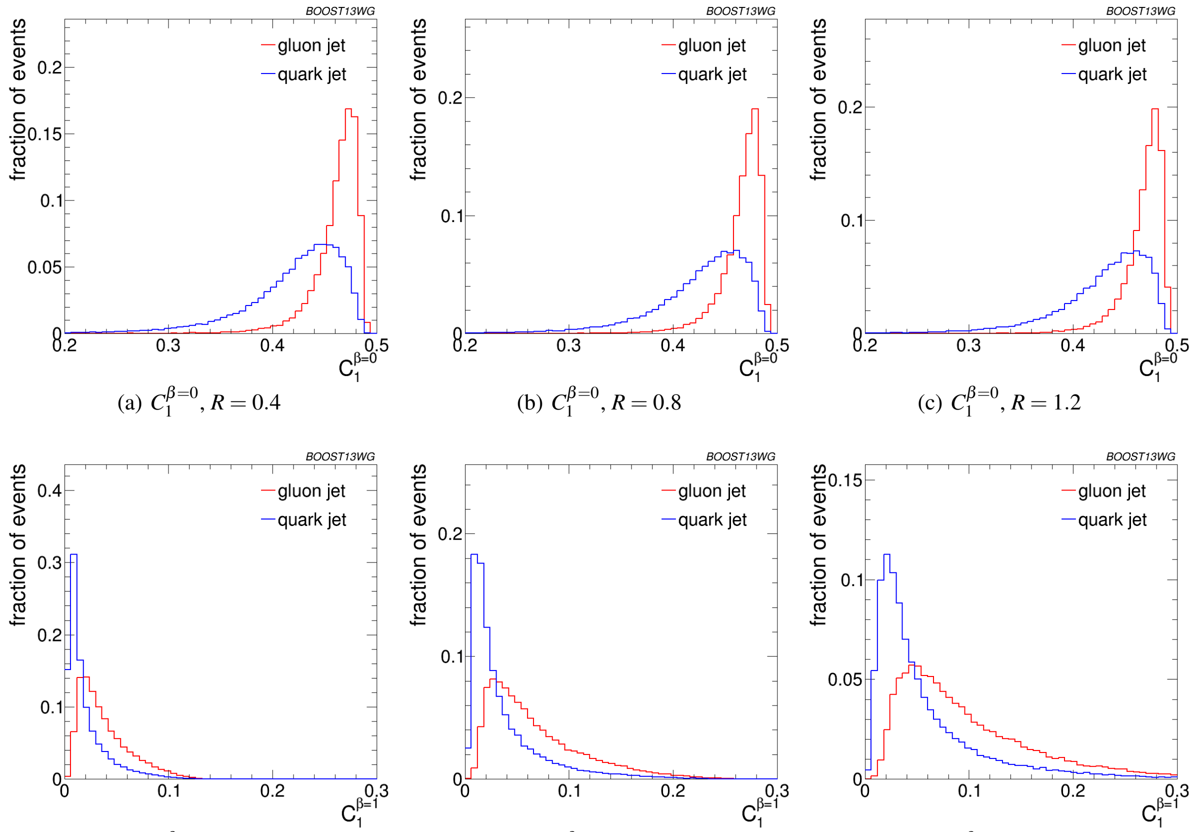
In this section, we study the discrimination of a boosted hadronically decaying W signal against a gluon background, comparing the performance of various groomed jet masses, substructure variables, and BDT combinations of groomed mass and substructure. We produce ROC curves that elucidate the performance of the various groomed mass and substructure variables. A range of different distance parameters R for the anti-k_T jet algorithm are explored, as well as a variety of kinematic regimes (lead jet p_T 300-400 GeV, 500-600 GeV, 1.0-1.1 TeV). This allows us to determine the performance of observables as a function of jet radius and jet boost, and to see where different approaches may break down. The groomed mass and substructure variables are then combined in a BDT as described in Section 4, and the performance of the resulting BDT discriminant explored through ROC curves to understand the degree to which variables are correlated, and how this changes with jet boost and jet radius.

6.1 Methodology

These studies use the WW samples as signal and the dijet gg as background, described previously in Section 2. Whilst only gluonic backgrounds are explored here, the conclusions as to the dependence of the performance and correlations on the jet boost and radius have been verified to hold also for qq backgrounds. **ED: To be checked!**

As in the q/g tagging studies, the showered events were clustered with FASTJET 3.03 using the anti-k_T algorithm with jet radii of $R = 0.4, 0.8, 1.2$. In both signal and background samples, an upper and lower cut on the leading jet p_T is applied after showering/clustering, to ensure similar p_T spectra for signal and background in each p_T bin. The bins in leading jet p_T that are considered are 300-400 GeV, 500-600 GeV, 1.0-1.1 TeV, for the 300-400 GeV, 500-600 GeV, 1.0-1.1 TeV parton p_T slices respectively. The jets then have various grooming approaches applied and substructure observables reconstructed as described in Section 3.4. The substructure observables studied in this section are:

- The ungroomed, trimmed (m_trim), and pruned (m_prun) jet masses.
- The mass output from the modified mass drop tagger (m_mmdt).
- The soft drop mass with $\beta = -1, 2$ (m_sd).



(a) $C_1^{\beta=0}$, $R = 0.4$ (b) $C_1^{\beta=0}$, $R = 0.8$ (c) $C_1^{\beta=0}$, $R = 1.2$

(d) $C_1^{\beta=1}$, $R = 0.4$ (e) $C_1^{\beta=1}$, $R = 0.8$ (f) $C_1^{\beta=1}$, $R = 1.2$

Fig. 4 Comparisons of quark and gluon distributions of $C_1^{\beta=0}$ (top) and $C_1^{\beta=1}$ (bottom) for leading jets in the $p_T = 1-1.1$ TeV bin using the anti-k_T algorithm with $R = 0.4$, 0.8 and 1.2.

- 2-point energy correlation function ratio $C_2^{\beta=1}$ (we also studied $\beta = 2$ but do not show its results because it showed poor discrimination power).
- N-subjettiness ratio τ_2/τ_1 with $\beta = 1$ ($\tau_{21}^{\beta=1}$) and with axes computed using one-pass k_t axis optimization (we also studied $\beta = 2$ but did not show its results because it showed poor discrimination power).
- The pruned Qjet mass volatility, Γ_Qjet.

6.2 Single Variable Performance

In this section we will explore the performance of the various groomed jet mass and substructure variables in terms of discriminating signal and background, and how this performance changes depending on the kinematic bin and jet radius considered.

Figure 7 the compares the signal and background in terms of the different groomed masses explored for the anti-k_T R=0.8 algorithm in the p_T 500-600 bin. One can clearly see that in terms of separating signal and background the groomed masses will be significantly more performant than the ungroomed anti-k_T R=0.8 mass. Figure 8 compares signal and background in the different substructure variables explored for the same jet radius and kinematic bin.

Figures 9, 10 and 11 show the single variable ROC curves compared to the ROC curve for a BDT combination of all the variables (labelled "allvars"), for each of the anti-k_T distance parameters considered in each of the kinematic bins. One can see that, in all cases, the "allvars" option is considerably better performant than any of the individual single variables considered, indicating that there is considerable complementarity between the variables, and this will be explored further in the next section.

Although the ROC curves give all the relevant information, it is hard to compare performance quantitatively. In Figures 12, 13 and 14 are shown matrices which give the background rejection for a signal efficiency of 70% when two variables (that on the x-axis and that on the y-axis) are combined in a BDT. These are shown separately for each p_T bin and jet radius considered. In the final column of these plots are shown the background rejection performance for three-variable BDT combinations of $m_{sd}^{\beta=2} + C_2^{\beta=1} + X$. These results will be discussed later in Section 6.3.3. The diagonal of these plots correspond to the background rejec-

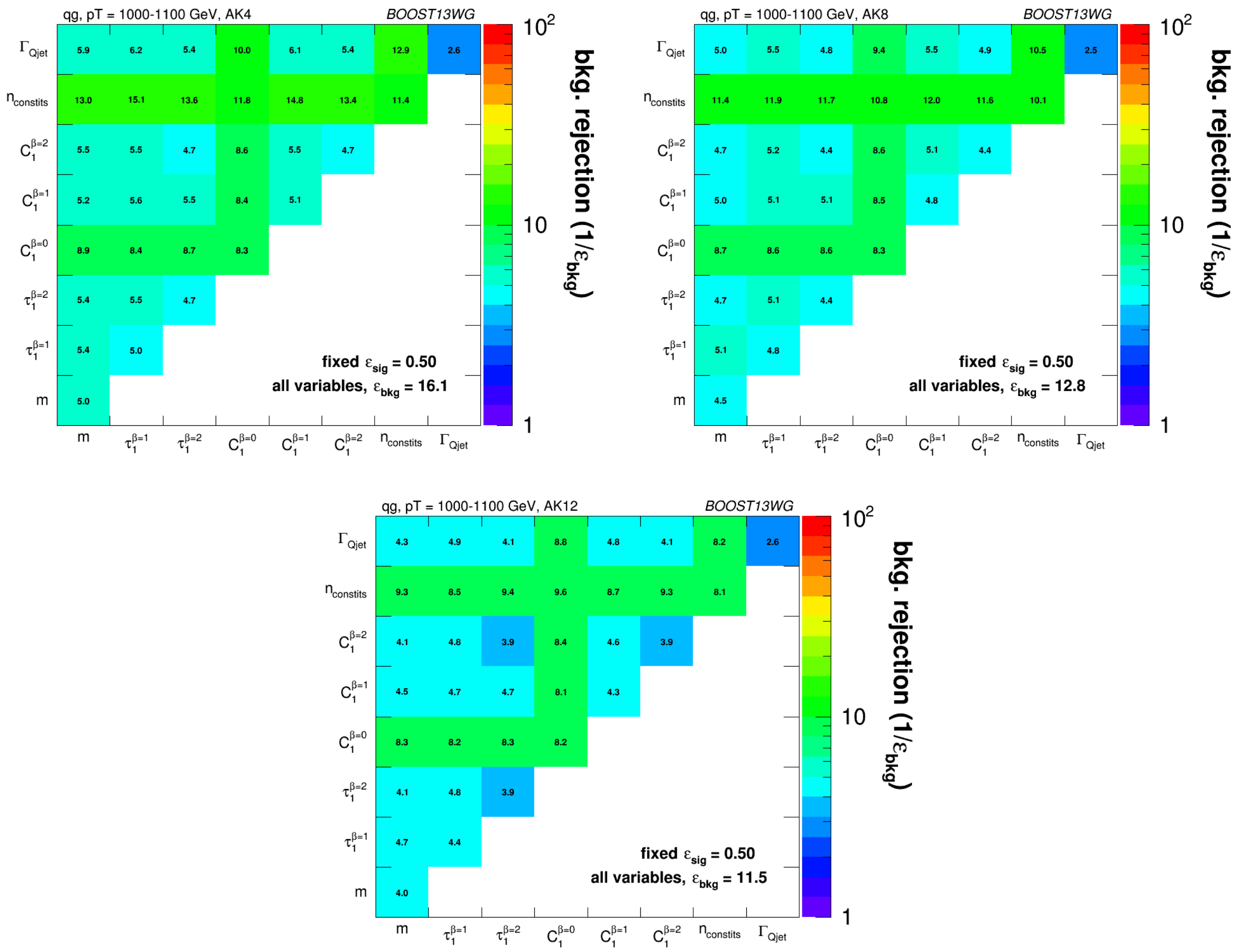


Fig. 5 Gluon rejection defined as $1/\varepsilon_{\text{gluon}}$ when using each 2-variable combination as a tagger with 50% acceptance for quark jets. Results are shown for jets with $p_T = 1-1.1$ TeV and for (left) $R = 0.4$; (centre) $R = 0.8$; (right) $R = 1.2$. The rejection obtained with a tagger that uses all variables is also shown in the plots.

tions for a single variable BDT, and can thus be examined to get a quantitative measure of the individual single variable performance, and to study how this changes with jet radius and momenta.

One can see that in general the most performant single variables are the groomed masses. However, in certain kinematic bins and for certain jet radii, $C_2^{\beta=1}$ has a background rejection that is comparable to or better than the groomed masses.

By comparing Figures 12(a), 13(a) and 14(b), we can see how the background rejection performance evolves as we increase momenta whilst keeping the jet radius fixed to R=0.8. Similarly, by comparing Figures 12(b), 13(b) and 14(c) we can see how performance evolves with p_T for R=1.2. For both R=0.8 and R=1.2 the background rejection power of the groomed masses increases with increasing p_T, with a factor 1.5-2.5 increase in rejection in going from the 300-400 GeV to 1.0-1.1 TeV bins. **ED: Add some of the 1-D plots comparing signal and bkgd in the different masses and pT bins here?** However, the $C_2^{\beta=1}$, Γ_{Qjet} and $\tau_{21}^{\beta=1}$ substructure variables behave somewhat differently. The background rejection power of the Γ_{Qjet} and $\tau_{21}^{\beta=1}$ variables both decrease with increasing p_T, by up to a factor two in going from the 300-400 GeV to 1.0-1.1 TeV bins. Conversely the rejection power of $C_2^{\beta=1}$ dramatically increases with increasing p_T for R=0.8, but does not improve with p_T for the larger jet radius R=1.2. **ED: Can we explain this? Again, should we add some of the 1-D plots?**

By comparing the individual sub-figures of Figures 12, 13 and 14 we can see how the background rejection performance depends on jet radius within the same p_T bin. To within $\sim 25\%$, the background rejection power of the groomed masses remains constant with respect to the jet radius. However, we again see rather different behaviour for the substructure variables. In all p_T bins considered the most performant substructure variable, $C_2^{\beta=1}$, performs best for an anti-k_T distance parameter of R=0.8. The performance of this variable is dramatically worse for the larger jet radius

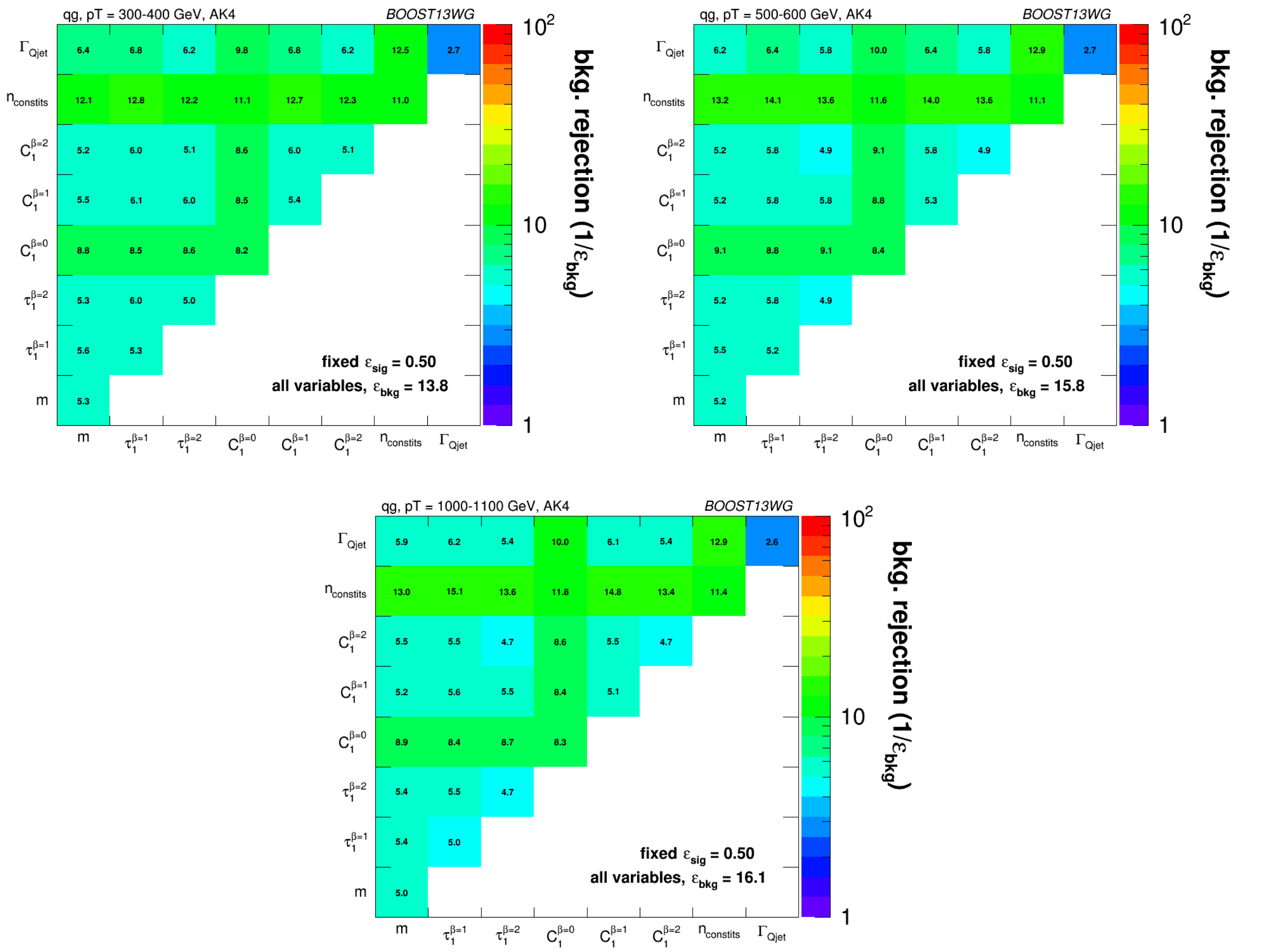


Fig. 6 Gluon rejection defined as $1/\varepsilon_{\mathrm{gluon}}$ when using each 2-variable combination as a tagger with 50% acceptance for quark jets. Results are shown for R=0.4 jets with $p_T = 300-400$ GeV, $p_T = 500-600$ GeV and $p_T = 1-1.1$ TeV. The rejection obtained with a tagger that uses all variables is also shown in the plots.

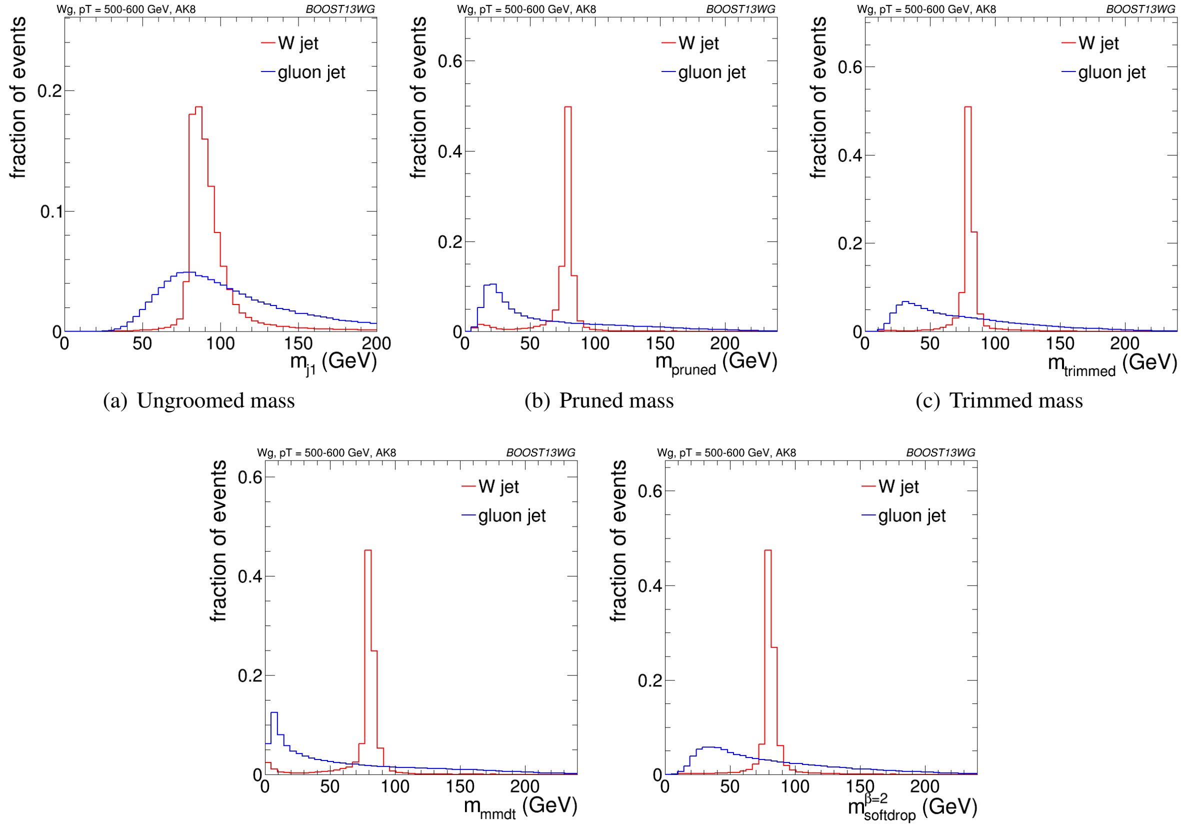
of R=1.2 (a factor seven worse background rejection in the 1.0-1.1 TeV bin), and substantially worse for R=0.4. For the other jet substructure variables considered, Γ_{Qjet} and $\tau_{21}^{\beta=1}$ their background rejection power also reduces for larger jet radius, but not to the same extent. **ED: Insert some nice discussion/explanation of why jet substructure power generally gets worse as we go to large jet radius, but groomed mass performance does not. Probably need the 1-D figures for this.**

6.3 Combined Performance

The off-diagonal entries in Figures 12, 13 and 14 can be used to compare the performance of different BDT two-variable combinations, and see how this varies as a function of p_T and R. By comparing the background rejection achieved for the two-variable combinations to the background rejection of the "all variables" BDT, one can understand how much more discrimination is possible by adding further variables to the two-variable BDTs.

One can see that in general the most powerful two-variable combinations involve a groomed mass and a non-mass substructure variable ($C_2^{\beta=1}$, Γ_{Qjet} or $\tau_{21}^{\beta=1}$). Two-variable combinations of the substructure variables are not powerful in comparison. Which particular mass + substructure variable combination is the most powerful depends strongly on the p_T and R of the jet, as discussed in the sections that follow.

There is also modest improvement in the background rejection when different groomed masses are combined, compared to the single variable groomed mass performance, indicating that there is complementary information between the different groomed masses. In addition, there is an improvement in the background rejection when the groomed masses are combined with the ungroomed mass, indicating that grooming removes some useful discriminatory information from the jet. These observations are explored further in the section below.



(a) Ungroomed mass (b) Pruned mass (c) Trimmed mass

(d) mMDT mass (e) Soft-drop $\beta = 2$ mass

Fig. 7 Comparisons of the QCD background to the WW signal in the p_T 500-600 GeV bin using the anti-k_{T} R=0.8 algorithm: leading jet mass distributions.

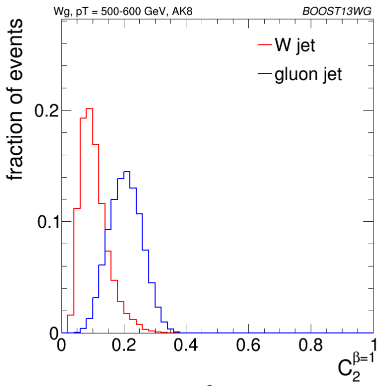
Generally one can see that the R=0.8 jets offer the best two-variable combined performance in all p_T bins explored here. This is despite the fact that in the highest 1.0-1.1 GeV p_T bin the average separation of the quarks from the W decay is much smaller than 0.8, and well within 0.4. This conclusion could of course be susceptible to pile-up, which is not considered in this study.

6.3.1 Mass + Substructure Performance

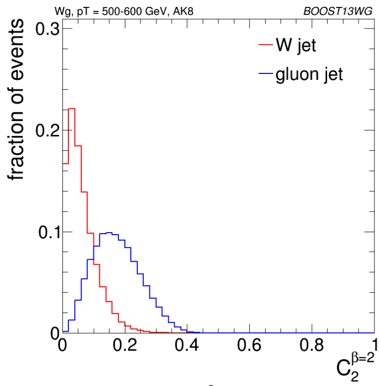
As already noted, the largest background rejection at 70% signal efficiency are in general achieved using those two variable BDT combinations which involve a groomed mass and a non-mass substructure variable. For both R=0.8 and R=1.2 jets, the rejection power of these two variable combinations increases substantially with increasing p_T , at least within the p_T range considered here.

For a jet radius of R=0.8, across the full p_T range considered, the groomed mass + substructure variable combinations with the largest background rejection are those which involve $C_2^{\beta=1}$. For example, in combination with $m_{sd}^{\beta=2}$, this produces a five-, eight- and fifteen-fold increase in background rejection compared to using the groomed mass alone. In Figure 15 the low degree of correlation between $m_{sd}^{\beta=2}$ versus $C_2^{\beta=1}$ that leads to these large improvements in background rejection can be seen. One can also see that what little correlation exists is rather non-linear in nature, changing from a negative to a positive correlation as a function of the groomed mass, something which helps to improve the background rejection in the region of the W mass peak.

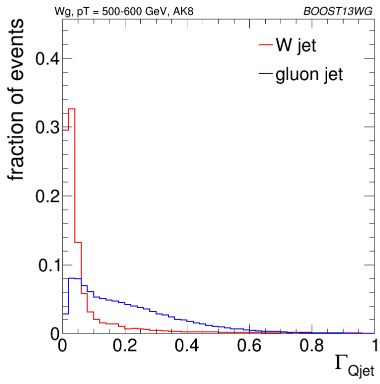
However, when we switch to a jet radius of R=1.2 the picture for $C_2^{\beta=1}$ combinations changes dramatically. These become significantly less powerful, and the most powerful variable in groomed mass combinations becomes $\tau_{21}^{\beta=1}$ for all jet p_T considered. Figure 16 shows the correlation between $m_{sd}^{\beta=2}$ and $C_2^{\beta=1}$ in the p_T 1.0 - 1.2 TeV bin for the various jet radii considered. Figure 17 is the equivalent set of distributions for $m_{sd}^{\beta=2}$ and $\tau_{21}^{\beta=1}$. One can see from Figure 16 that, due to the sensitivity of the observable to to soft, wide-angle radiation, as the jet radius increases $C_2^{\beta=1}$ increases and becomes more and more smeared out for both signal and background, leading to worse discrimination power. This does not happen to the same extent for $\tau_{21}^{\beta=1}$. We can see from Figure 17 that the negative correlation between $m_{sd}^{\beta=2}$ and $\tau_{21}^{\beta=1}$ that is clearly visible for R=0.4 decreases for larger



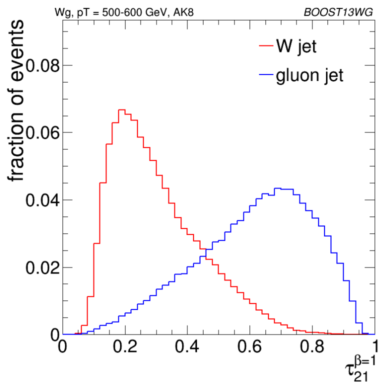
(a) $C_2^{\beta=1}$



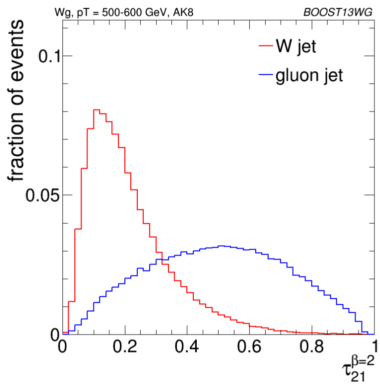
(b) $C_2^{\beta=2}$



(c) Γ_{Qjet}

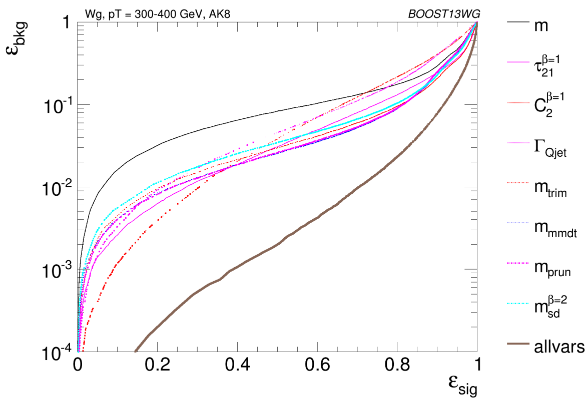


(d) $\tau_{21}^{\beta=1}$

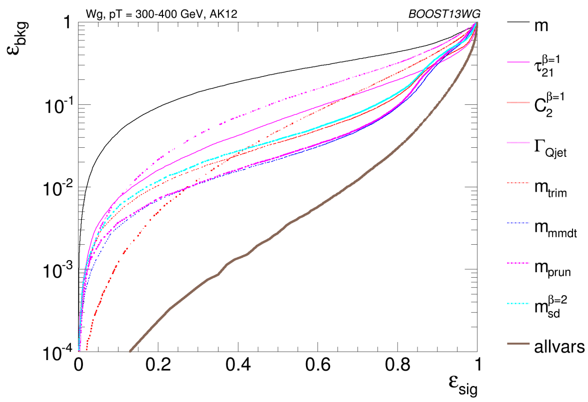


(e) $\tau_{21}^{\beta=2}$

Fig. 8 Comparisons of the QCD background to the WW signal in the p_T 500-600 GeV bin using the anti-k_{T} R=0.8 algorithm: substructure variables.

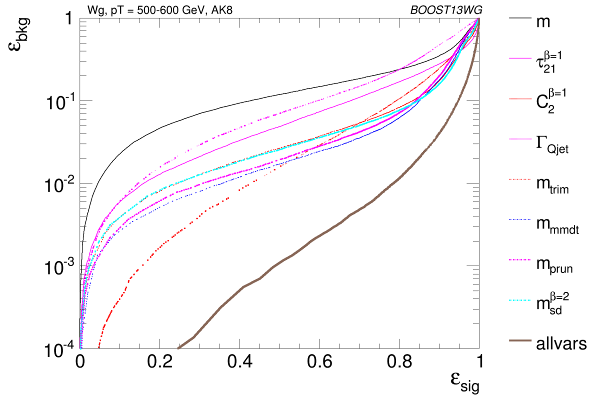


(a) anti-k_{T} R=0.8, p_T 300-400 GeV bin

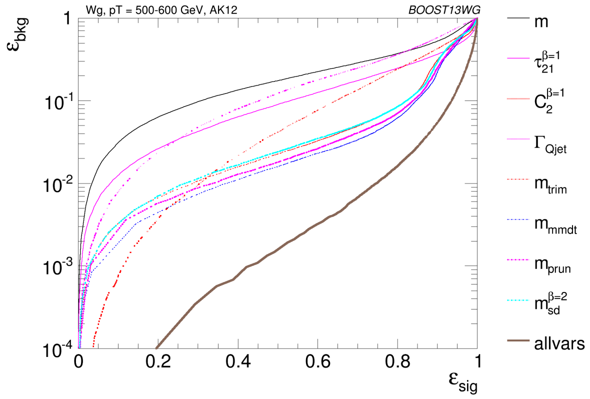


(b) anti-k_{T} R=1.2, p_T 300-400 GeV bin

Fig. 9 The ROC curve for all single variables considered for W tagging in the p_T 300-400 GeV bin using the anti-k_{T} R=0.8 algorithm and R=1.2 algorithm.

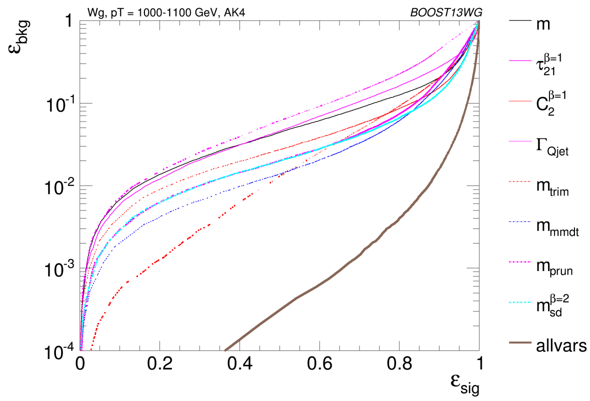


(a) anti-k_{T} R=0.8, p_T 500-600 GeV bin

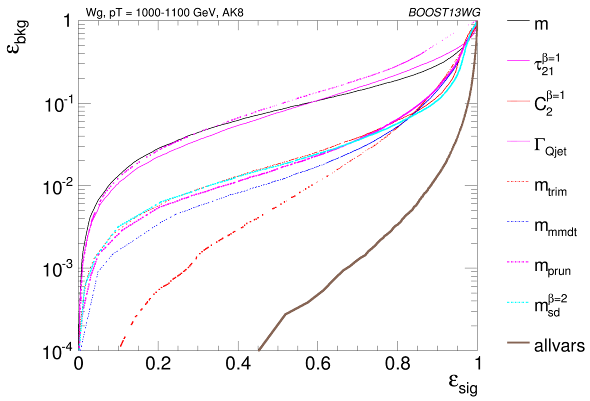


(b) anti-k_{T} R=1.2, p_T 500-600 GeV bin

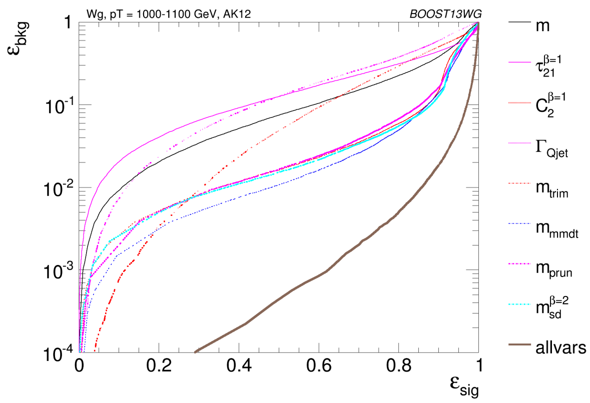
Fig. 10 The ROC curve for all single variables considered for W tagging in the p_T 500-600 GeV bin using the anti-k_{T} R=0.8 algorithm and R=1.2 algorithm.



(a) anti-k_{T} R=0.4, p_T 1.0-1.1 TeV bin

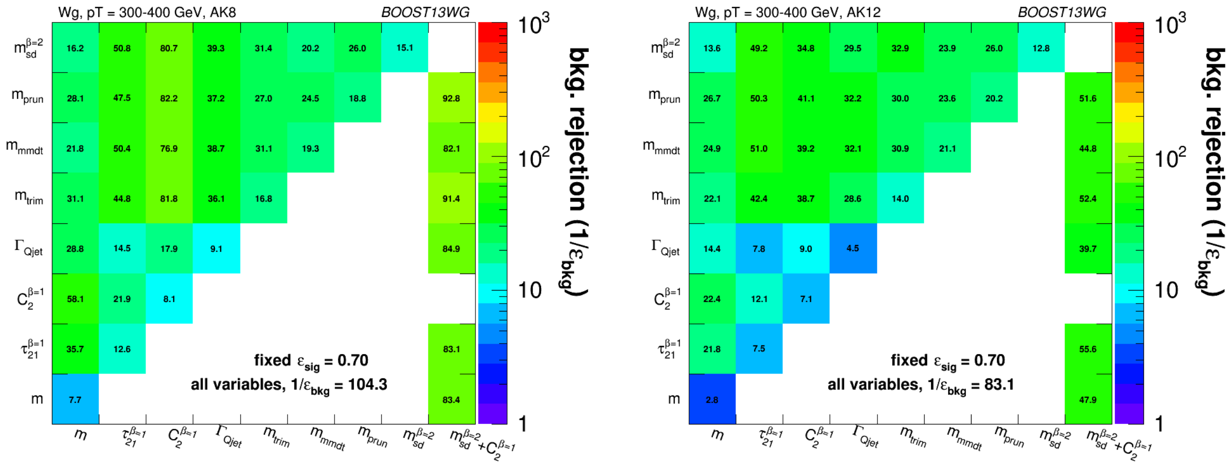


(b) anti-k_{T} R=0.8, p_T 1.0-1.1 TeV bin



(c) anti-k_{T} R=1.2, p_T 1.0-1.1 TeV bin

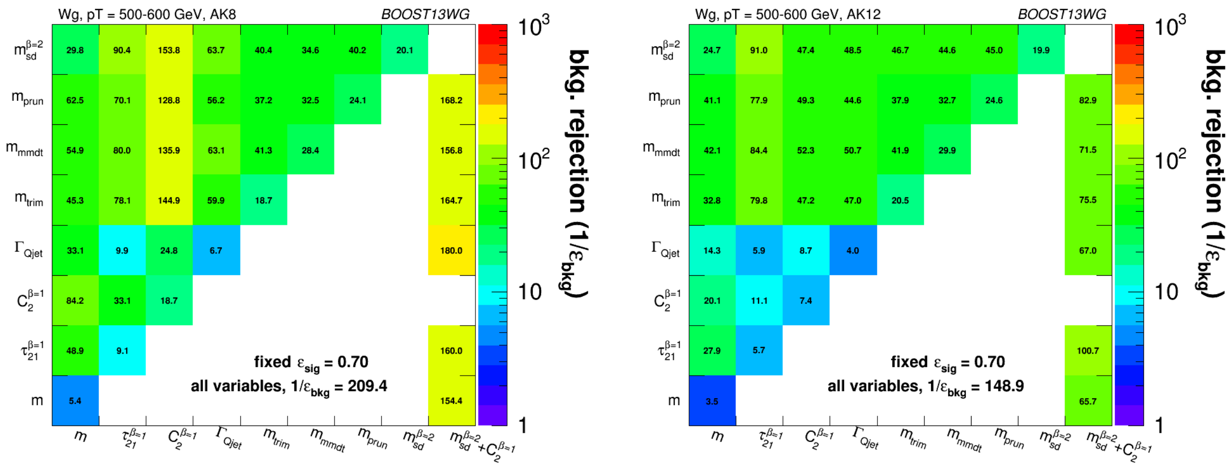
Fig. 11 The ROC curve for all single variables considered for W tagging in the p_T 1.0-1.1 TeV bin using the anti-k_{T} R=0.4 algorithm, anti-k_{T} R=0.8 algorithm and R=1.2 algorithm.



(a) anti-k_{T} R=0.8, p_T 300-400 GeV bin

(b) anti-k_{T} R=1.2, p_T 300-400 GeV bin

Fig. 12 The background rejection for a fixed signal efficiency (70%) of each BDT combination of each pair of variables considered, in the p_T 300-400 GeV bin using the anti-k_{T} R=0.8 algorithm and R=1.2 algorithm. Also shown is the background rejection for a BDT combination of all of the variables considered.



(a) anti-k_{T} R=0.8, p_T 500-600 GeV bin

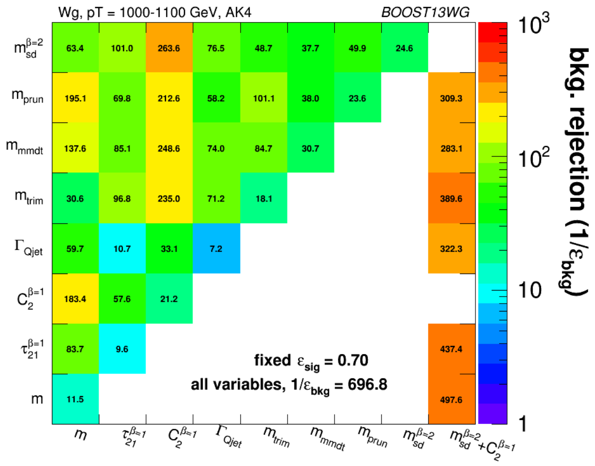
(b) anti-k_{T} R=1.2, p_T 500-600 GeV bin

Fig. 13 The background rejection for a fixed signal efficiency (70%) of each BDT combination of each pair of variables considered, in the p_T 500-600 GeV bin using the anti-k_{T} R=0.8 algorithm and R=1.2 algorithm. Also shown is the background rejection for a BDT combination of all of the variables considered.

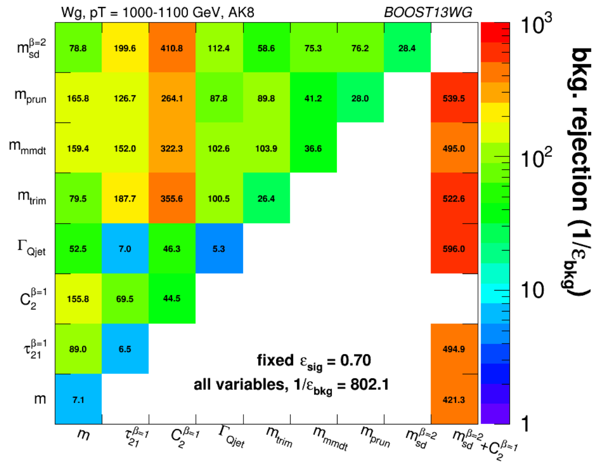
jet radius, such that the groomed mass and substructure variable are far less correlated and $\tau_{21}^{\beta=1}$ offers improved discrimination within a $m_{sd}^{\beta=2}$ mass window.

6.3.2 Mass + Mass Performance

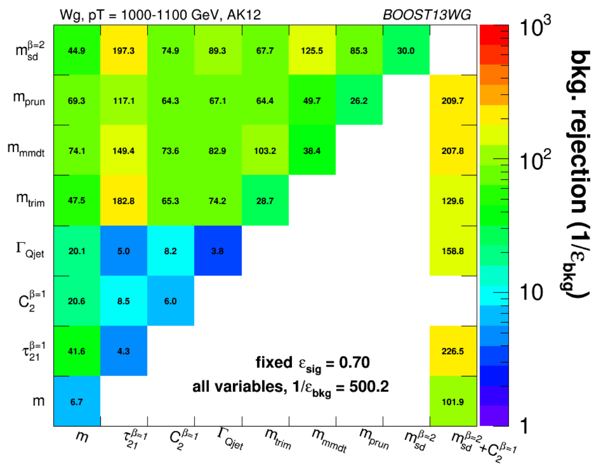
The different groomed masses and the ungroomed mass are of course not fully correlated, and thus one can always see some kind of improvement in the background rejection (relative to the single mass performance) when two different mass variables are combined in the BDT. However, in some cases the improvement can be dramatic, particularly at higher p_T , and particularly for combinations with the ungroomed mass. For example, in Figure 14 we can see that in the p_T 1.0-1.1 TeV bin the combination of pruned mass with ungroomed mass produces a greater than eight-fold improvement in the background rejection for R=0.4 jets, a greater than five-fold improvement for R=0.8 jets, and a factor ∼two improvement for R=1.2 jets. A similar behaviour can be seen for mMDT mass. In Figures 18, 19 and 20 is shown the 2-D correlation plots of the pruned mass versus the ungroomed mass separately for the WW signal and gg background samples in the p_T 1.0-1.1 TeV bin, for the various jet radii considered. For comparison, the correlation of the trimmed



(a) anti-k_{T} R=0.4, p_T 1.0-1.1 TeV bin



(b) anti-k_{T} R=0.8, p_T 1.0-1.1 TeV bin



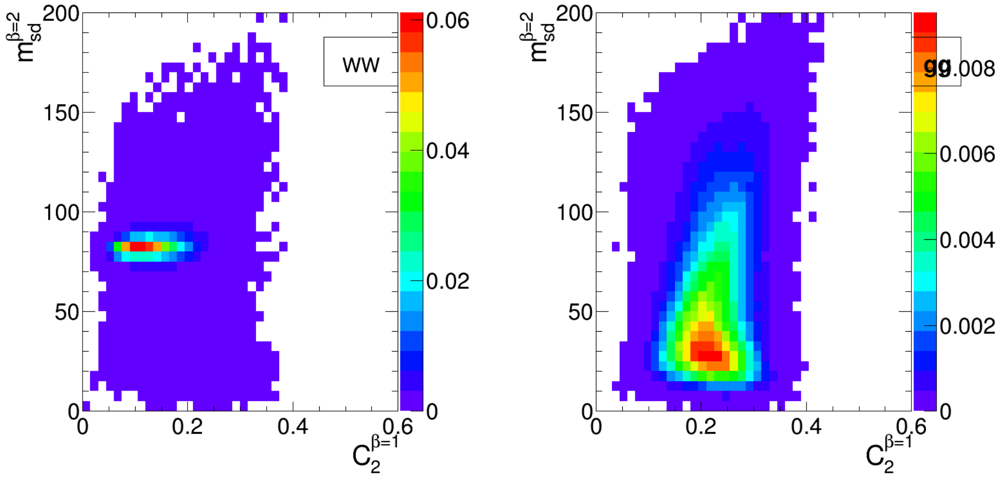
(c) anti-k_{T} R=1.2, p_T 1.0-1.1 TeV bin

Fig. 14 The background rejection for a fixed signal efficiency (70%) of each BDT combination of each pair of variables considered, in the p_T 1.0-1.1 TeV bin using the anti-k_{T} R=0.4, R=0.8 and R=1.2 algorithm. Also shown is the background rejection for a BDT combination of all of the variables considered.

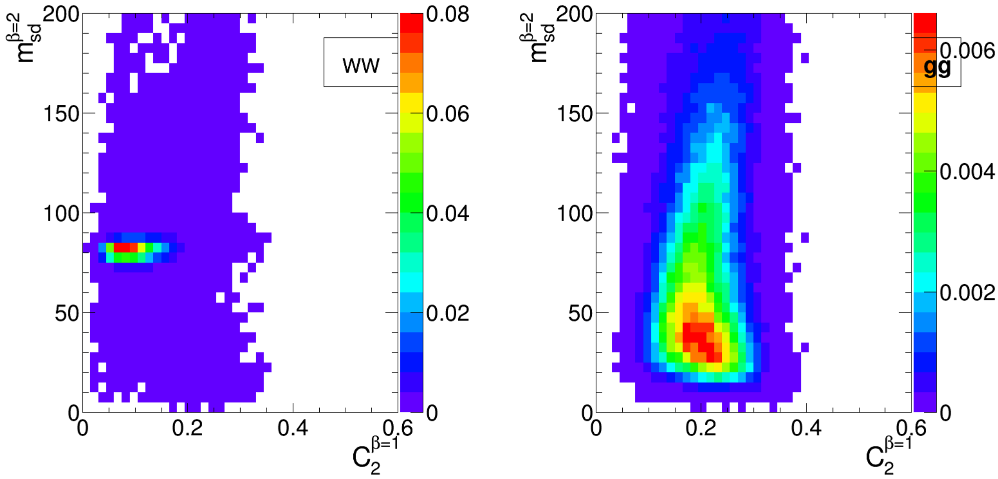
mass with the ungroomed mass, a combination that does not improve on the single mass as dramatically, is shown. In all cases one can see that there is a much smaller degree of correlation between the pruned mass and the ungroomed mass in the backgrounds sample than for the trimmed mass and the ungroomed mass. This is most obvious in Figure 18, where the high degree of correlation between the trimmed and ungroomed mass is expected, since with the parameters used (in particular $R_{trim} = 0.2$) we cannot expect trimming to have a significant impact on an R=0.4 jet. The reduced correlation with ungroomed mass for pruning in the background means that, once we have made the requirement that the pruned mass is consistent with a W (i.e. ∼80 GeV), a relatively large difference between signal and background in the ungroomed mass still remains, and can be exploited to improve the background rejection further. In other words, many of the background events which pass the pruned mass requirement do so because they are shifted to lower mass (to be within a signal mass window) by the grooming, but these events still have the property that they look very much like background events before the grooming. A single requirement on the groomed mass only does not exploit this. Of course, the impact of pile-up, not considered in this study, could significantly limit the degree to which the ungroomed mass could be used to improve discrimination in this way.

6.3.3 "All Variables" Performance

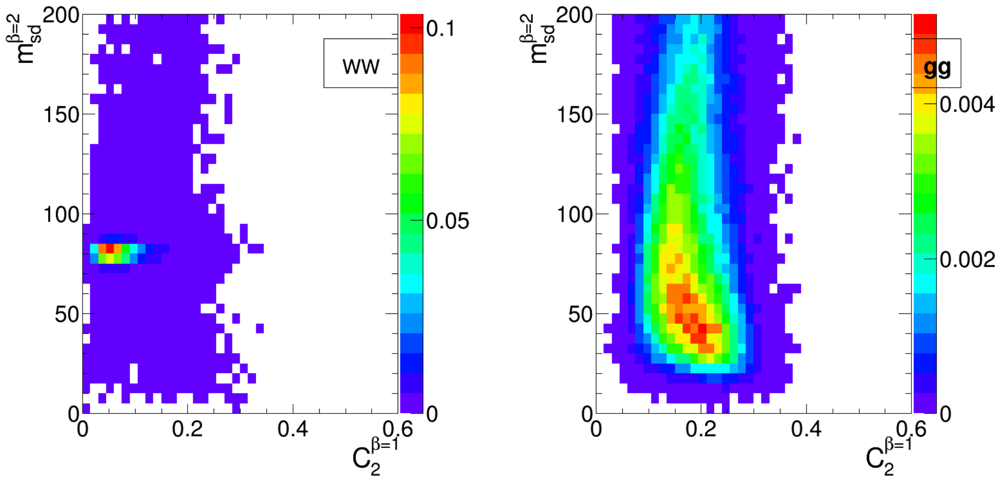
As well as the background rejection at a fixed 70% signal efficiency for two-variable combinations, Figures 12, 13 and 14 also report the background rejection achieved by a combination of all the variables considered into a single



(a) p_T 300-400 GeV

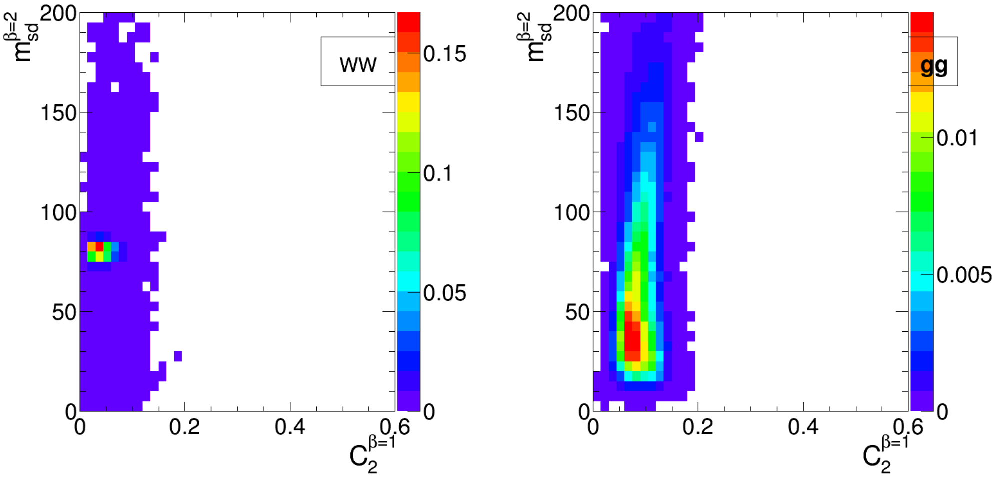


(b) p_T 500-600 GeV

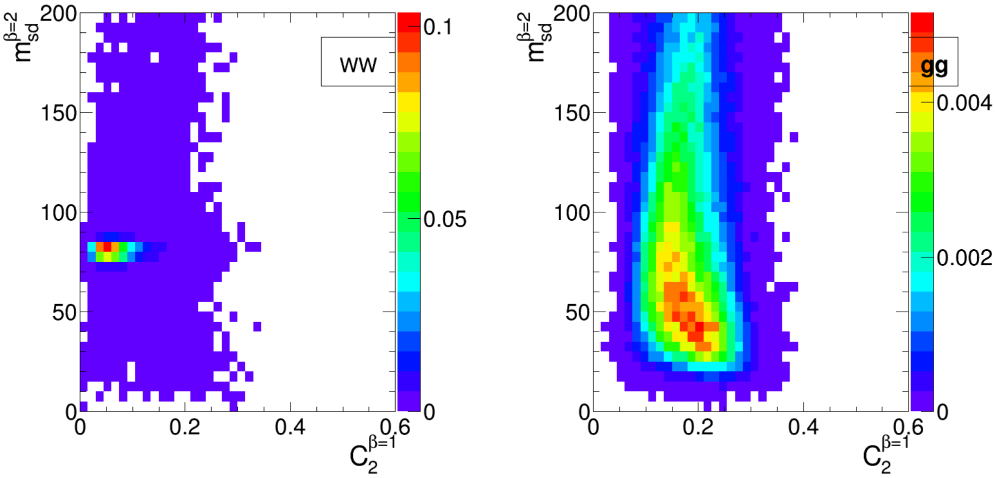


(c) p_T 1.0-1.1 TeV

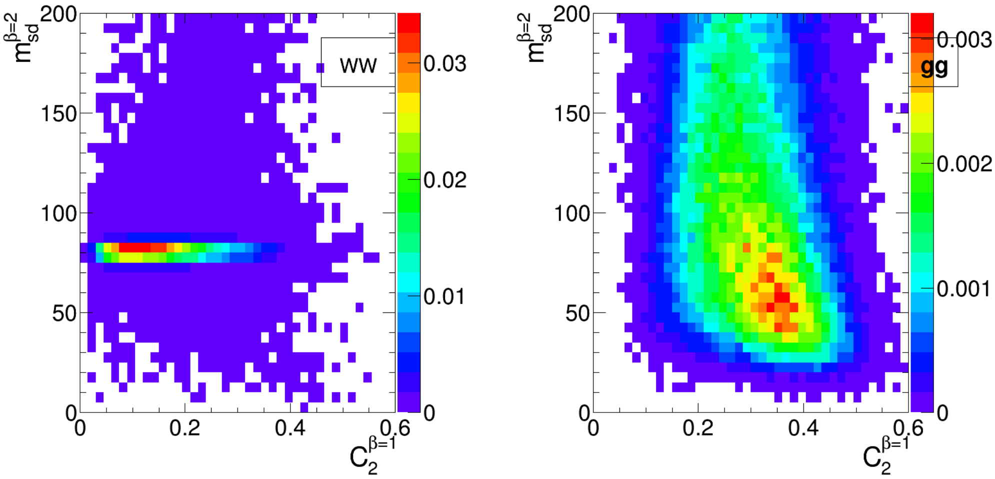
Fig. 15 2-D plots showing $m_{sd}^{\beta=2}$ versus $C_2^{\beta=1}$ for R=0.8 jets in the various p_T bins considered.



(a) p_T R=0.4

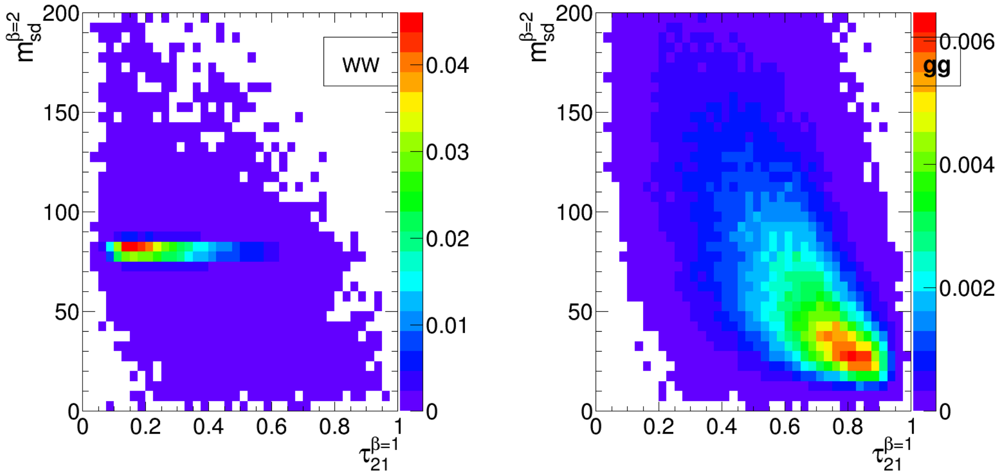


(b) p_T R=0.8

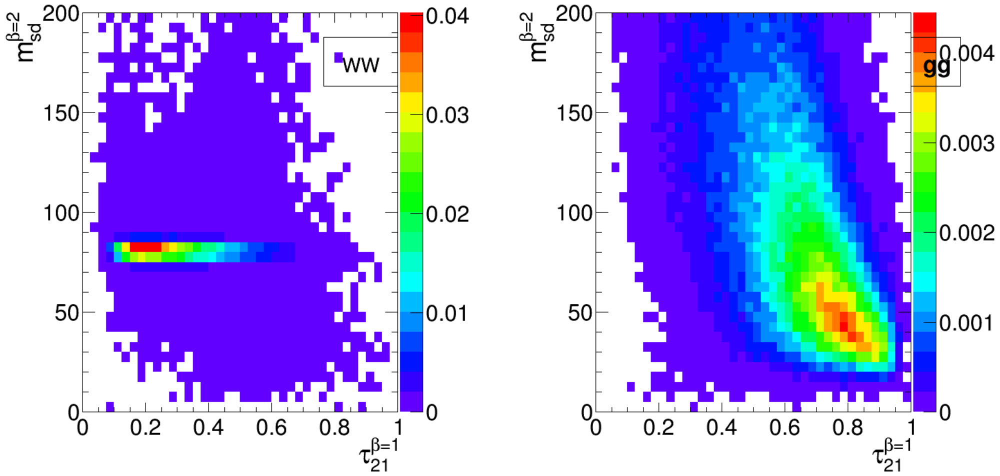


(c) p_T R=1.2

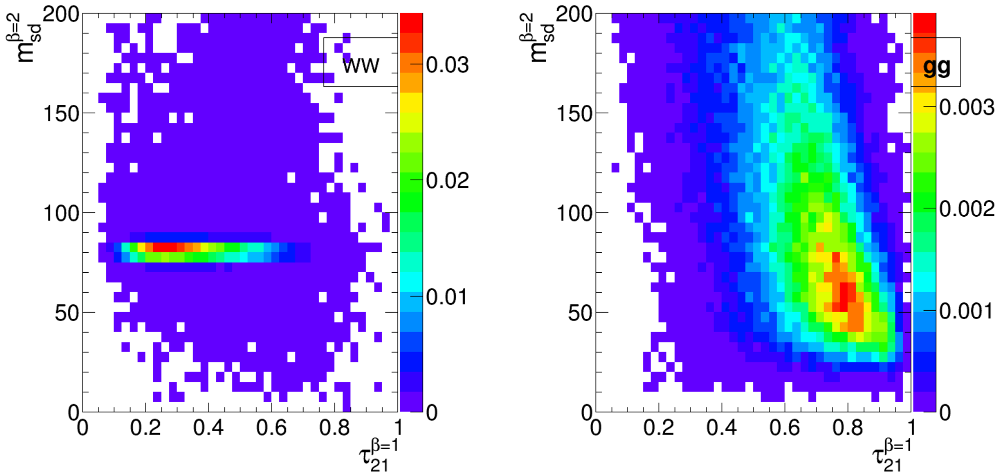
Fig. 16 2-D plots showing $m_{sd}^{\beta=2}$ versus $C_2^{\beta=1}$ for R=0.4, 0.8 and 1.2 jets in the p_T 1.0-1.1 TeV bin.



(a) p_T R=0.4

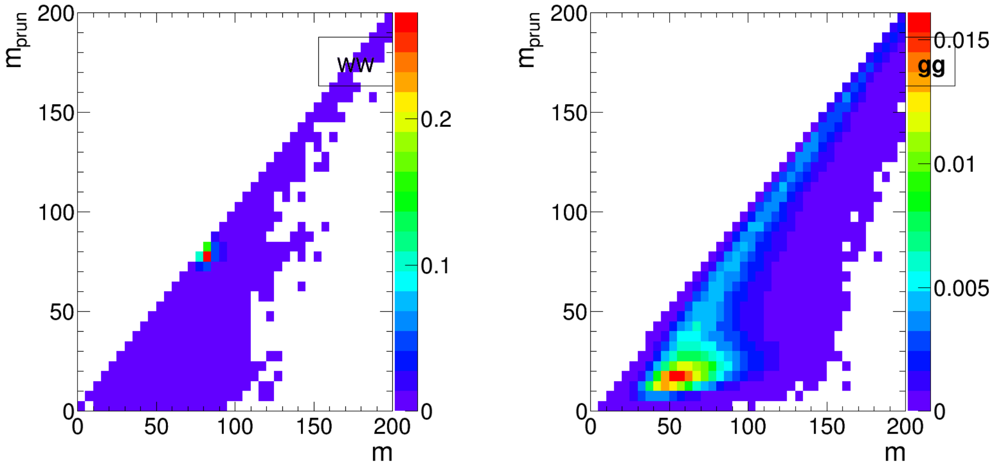


(b) p_T R=0.8

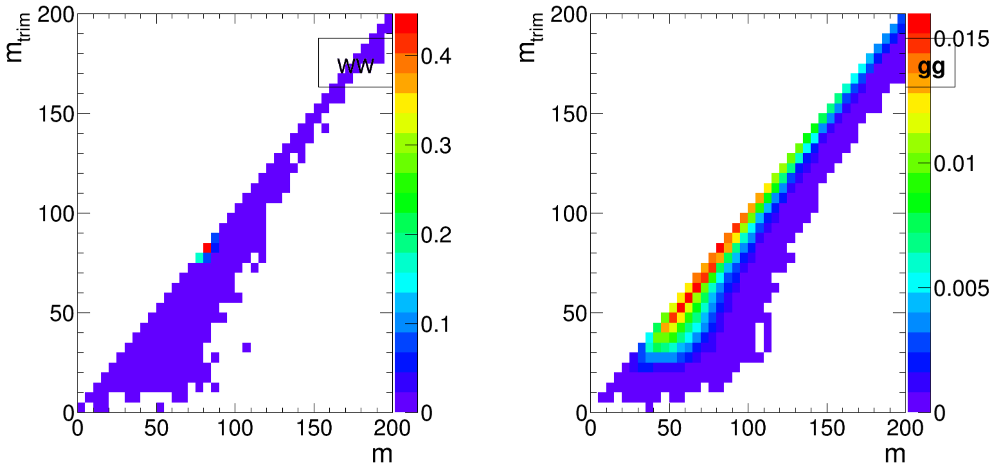


(c) p_T R=1.2

Fig. 17 2-D plots showing $m_{sd}^{\beta=2}$ versus $\tau_{21}^{\beta=1}$ for R=0.4, 0.8 and 1.2 jets in the p_T 1.0-1.1 TeV bin.



(a) Pruned mass vs ungroomed mass

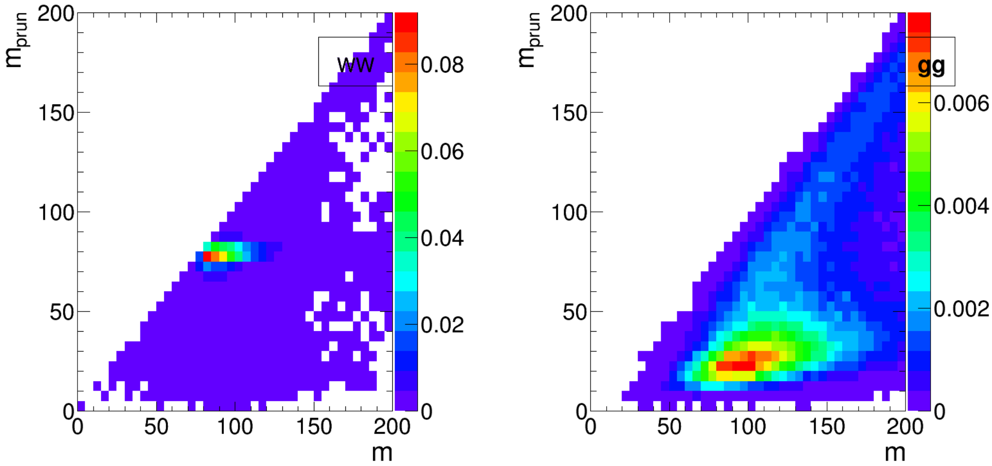


(b) Trimmed mass vs ungroomed mass

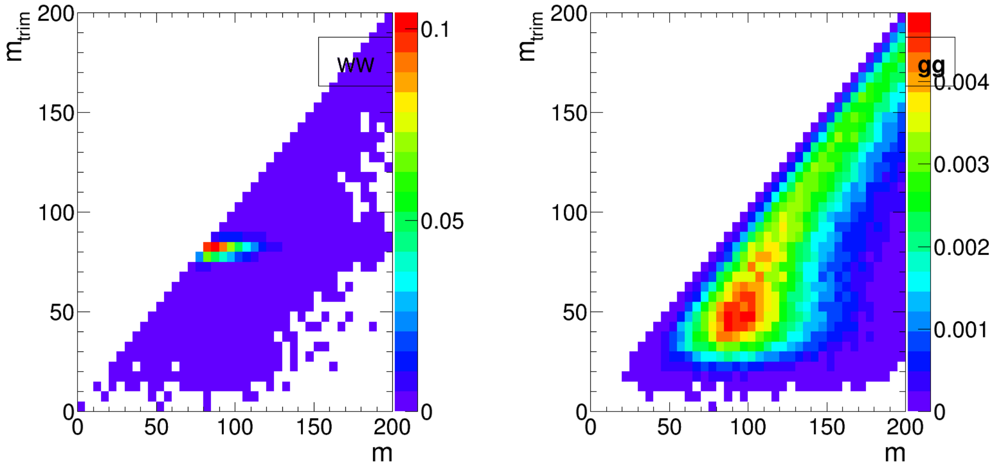
Fig. 18 2-D plots showing the correlation between groomed and ungroomed mass for WW and gg events in the p_T 1.0-1.1 TeV bin using the anti-k_{T} R=0.4 algorithm.

BDT discriminant. One can see that, in all cases, the rejection power of this "all variables" BDT is significantly larger than the best two-variable combination. This indicates that beyond the best two-variable combination there is still significant complementary information availiable in the remaining variables in order to improve the discrimination of signal and background. How much complementary information is available appears to be p_T dependent. In the lower p_T 300-400 and 500-600 GeV bins the background rejection of the "all variables" combination is a factor ~ 1.5 greater than the best two-variable combination, but in the highest p_T bin it is a factor ~ 2.5 greater.

The final column in Figures 12, 13 and 14 allows us to explore the all variables performance a little further. It shows the background rejection for three variable BDT combinations of $m_{sd}^{\beta=2}+C_2^{\beta=1}+X$, where X is the variable on the y-axis. For jets with R=0.4 and R=0.8, the combination $m_{sd}^{\beta=2}+C_2^{\beta=1}$ is the best performant (or very close to the best performant) two-variable combination in every p_T bin considered. For R=1.2 this is not the case, as $C_2^{\beta=1}$ is superceded by $\tau_{21}^{\beta=1}$ in performance, as discussed earlier. Thus, in considering the three-variable combination results it is best to focus on the R=0.4 and R=0.8 cases. Here we see that, for the lower p_T 300-400 and 500-600 GeV bins, adding the third variable to the best two-variable combination brings us to within $\sim 15\%$ of the "all variables" background rejection. However, in the highest p_T 1.0-1.1 TeV bin, whilst adding the third variable does improve the performance considerably, we are still $\sim 40\%$ from the observed "all variables" background rejection, and clearly adding a fourth or maybe even fifth variable would bring considerable gains. In terms of which variable offers the best improvement when added



(a) Pruned mass vs ungroomed mass



(b) Trimmed mass vs ungroomed mass

Fig. 19 2-D plots showing the correlation between groomed and ungroomed mass for WW and gg events in the p_T 1.0-1.1 TeV bin using the anti-k_{T} R=0.8 algorithm.

to the $m_{sd}^{\beta=2}+C_2^{\beta=1}$ combination, it is hard to see an obvious pattern; the best third variable changes depending on the p_T and R considered.

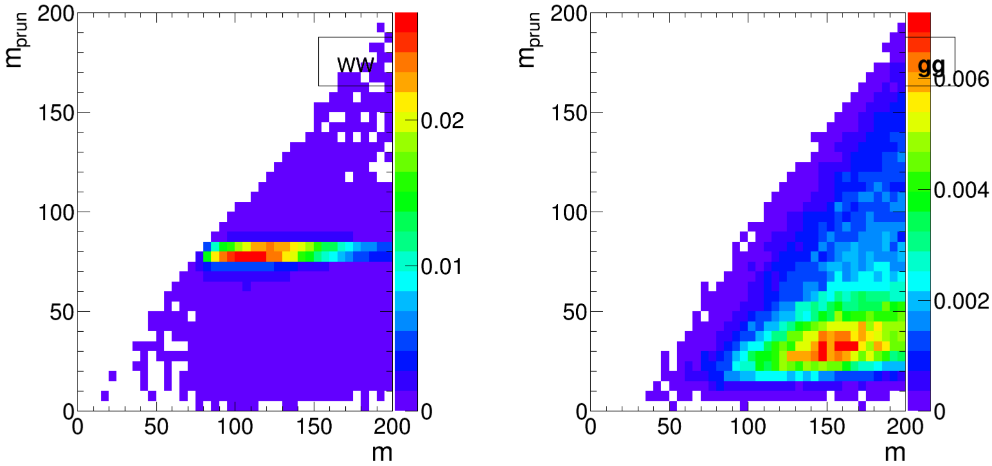
In conclusion, it appears that there is a rich and complex structure in terms of the degree to which the discriminatory information provided by the set of variables considered overlaps, with the degree of overlap apparently decreasing at higher p_T . This suggests that in all p_T ranges, but especially at higher p_T , there are substantial performance gains to be made by designing a more complex multivariate W tagger.

6.4 Conclusions

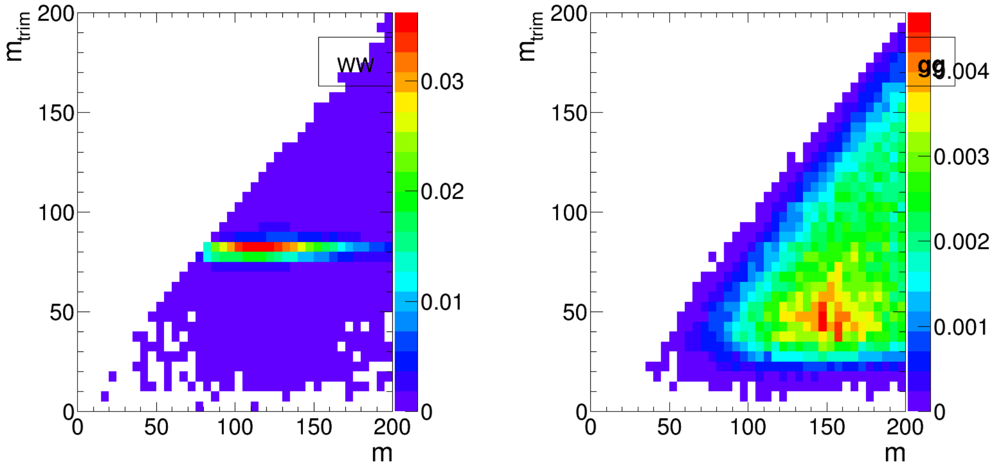
We have studied the performance, in terms of the degree to which a hadronically decaying W boson can be separated from a gluonic background, of a number of groomed jet masses, substructure variables, and BDT combinations of the above. We have used this to build a picture of how the discriminatory information contained in the variables overlaps, and how this complementarity between the variables changes with p_T and anti-k_{T} distance parameter R.

In terms of the performance of individual variables, we find that, in agreement with other studies**[REF]**, in general the groomed masses perform best, with a background rejection power that increases with increasing p_T , but which is more constant with respect to changes in R. Conversely, the performance of other substructure variables, such as $C_2^{\beta=1}$ and $\tau_{21}^{\beta=1}$ is more susceptible to changes in radius, with background rejection power decreasing with increasing R.

The best two-variable performance is obtained by combining a groomed mass with a substructure variable. Which particular substructure variable works best in combination



(a) Pruned mass vs ungroomed mass



(b) Trimmed mass vs ungroomed mass

Fig. 20 2-D plots showing the correlation between groomed and ungroomed mass for WW and gg events in the p_T 1.0-1.1 TeV bin using the anti-k_{T} R=1.2 algorithm.

is strongly dependent on p_T and R. $C_2^{\beta=1}$ offers significant complimentarity to groomed mass at smaller R, owing to the small degree of correlation between the variables. However, the sensitivity of $C_2^{\beta=1}$ to soft, wide-angle radiation leads to worse discrimination power at large R, where $\tau_{21}^{\beta=1}$ performs better in combination. Our studies also demonstrate the potential for enhanced discrmination by combining groomed and ungroomed mass information, although the use of ungroomed mass in this may in practice be limited by the presence of pile-up that is not considered in these studies.

By examining the performance of a BDT combination of all the variables considered, it is clear that there are potentially substantial performance gains to be made by designing a more complex multivariate W tagger, especially at higher p_T .

7 Top Tagging

In this section, we study the identification of boosted top quarks at Run II of the LHC. Boosted top quarks result in large-radius jets with complex substructure, containing a b-subjet and a boosted W. The additional kinematic handles coming from the reconstruction of the W mass and b-tagging allow a very high degree of discrimination of top quark jets from QCD backgrounds.

We consider top quarks with moderate boost (600-1000 GeV), and perhaps most interestingly, at high boost ($\gtrsim$ 1500 GeV). Top tagging faces several challenges in the high-p_T regime. For such high-p_T jets, the b-tagging efficiencies are no longer reliably known. Also, the top jet can also accompanied by additional radiation with $p_T \sim m_t$, leading to combinatoric ambiguities of reconstructing the top and W, and the possibility that existing taggers or observables shape the background by looking for subjet combinations that reconstruct m_t/m_W. To study this, we examine the performance of both mass-reconstruction variables, as well as shape observables that probe the three-pronged nature of the top jet and the accompanying radiation pattern.

We use the top quark MC samples for each bin described in Section 2.2. The analysis relies on FASTJET 3.0.3 for jet clustering and calculation of jet substructure observables. Jets are clustered using the anti-k_t algorithm. An upper and lower p_T cut are applied after jet clustering to each sample to ensure similar p_T spectra in each bin. The bins in leading jet p_T that are investigated for top tagging are 600-700 GeV, 1-1.1 TeV, and 1.5-1.6 TeV. Jets are clustered with radii $R = 0.4$, 0.8, and 1.2; $R = 0.4$ jets are only studied in the 1.5-1.6 TeV bin because for top quarks with this boost, the top decay products are all contained within an $R = 0.4$ jet.

7.1 Methodology

We study a number of top-tagging strategies, in particular:

1. HEPTopTagger
2. Johns Hopkins Tagger (JH)
3. Trimming
4. Pruning

The top taggers have criteria for reconstructing a top and W candidate, and a corresponding top and W mass, as described in Section 3.3, while the grooming algorithms (trimming and pruning) do not incorporate a W-identification step. For a level playing field, where grooming is used we construct a W candidate mass, m_W, from the three leading subjets by taking the mass of the pair of subjets with the smallest invariant mass; in the case that only two subjets are reconstructed, we take the mass of the leading subjet. The top mass, m_t, is the mass of the groomed jet. All of the above taggers and groomers incorporate a step to remove pile-up and other soft radiation.

We also consider the performance of the following jet shape observables:

– The ungroomed jet mass.
– N-subjettiness ratios τ_2/τ_1 and τ_3/τ_2 with $\beta = 1$ and the "winner-takes-all" axes.
– 2-point energy correlation function ratios $C_2^{\beta=1}$ and $C_3^{\beta=1}$.
– The pruned Qjet mass volatility, $\Gamma_{\rm Qjet}$.

In addition to the jet shape performance, we combine the jet shapes with the mass-reconstruction methods described above to determine the optimal combined performance.

For determining the performance of multiple variables, we combine the relevant tagger output observables and/or jet shapes into a boosted decision tree (BDT), which determines the optimal cut. Additionally, because each tagger has two input parameters, as described in Section 3.3, we scan over reasonable values of the parameters to determine the optimal value that gives the largest background rejection for each top tagging signal efficiency. This allows a direct comparison of the optimized version of each tagger. The input values scanned for the various algorithms are:

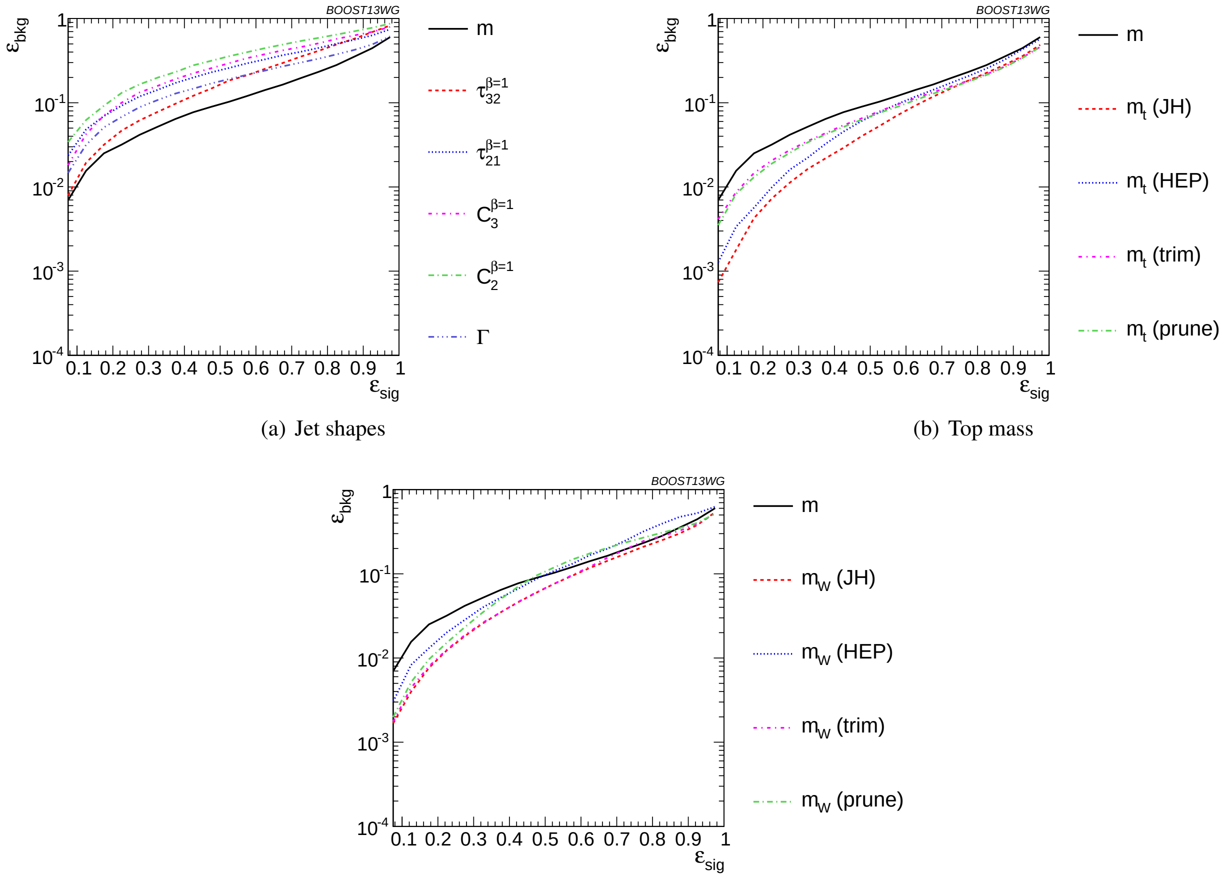
– **HEPTopTagger:** $m \in [30, 100]$ GeV, $\mu \in [0.5, 1]$
– **JH Tagger:** $\delta_p \in [0.02, 0.15]$, $\delta_R \in [0.07, 0.2]$
– **Trimming:** $f_{\rm cut} \in [0.02, 0.14]$, $R_{\rm trim} \in [0.1, 0.5]$
– **Pruning:** $z_{\rm cut} \in [0.02, 0.14]$, $R_{\rm cut} \in [0.1, 0.6]$

7.2 Single-observable performance

We start by investigating the behaviour of individual jet substructure observables. Because of the rich, three-pronged structure of the top decay, it is expected that combinations of masses and jet shapes will far outperform single observables in identifying boosted tops. However, a study of the top-tagging performance of single variables facilitates a direct comparison with the W tagging results in Section 6, and also allows a straightforward examination of the performance of each observable for different p_T and jet radius.

Fig. 21 shows the ROC curves for each of the top-tagging observables, with the bare (ungroomed) jet mass also plotted for comparison. The jet shape observables all perform substantially worse than jet mass, unlike W tagging for which several observables are competitive with or perform better than jet mass (see, for example, Fig. 7). To understand why this is the case, consider N-subjettiness. The W is two-pronged and the top is three-pronged; therefore, we expect τ_{21} and τ_{32} to be the best-performant N-subjettiness ratio, respectively. However, τ_{21} also contains an implicit cut on the denominator, τ_1, which is strongly correlated with jet mass. Therefore, τ_{21} combines both mass and shape information to some extent. By contrast, and as is clear in Fig.21(a), the best shape for top tagging is τ_{32}, which contains no information on the mass. Therefore, it is unsurprising that the shapes most useful for top tagging are less sensitive to the jet mass, and under-perform relative to the corresponding observables for W tagging.

Of the two top tagging algorithms, we can see from Figure 21 that the Johns Hopkins (JH) tagger out-performs the HEPTopTagger in terms of its signal-to-background separation power in both the top and W candidate masses. In Figure 22 we show the histograms for the top mass output from the JH and HEPTopTagger for different R in the p_T 1.5-1.6 TeV bin, and in Figure 23 for different p_T at at R =0.8, optimized at a signal efficiency of 30%. One can see from these figures that the likely reason for the better performance of the JH tagger is that, in the HEPTopTagger algorithm, the jet is filtered to select the five hardest subjets, and then three subjets are chosen which reconstruct the top mass. This re-



(a) Jet shapes

(b) Top mass

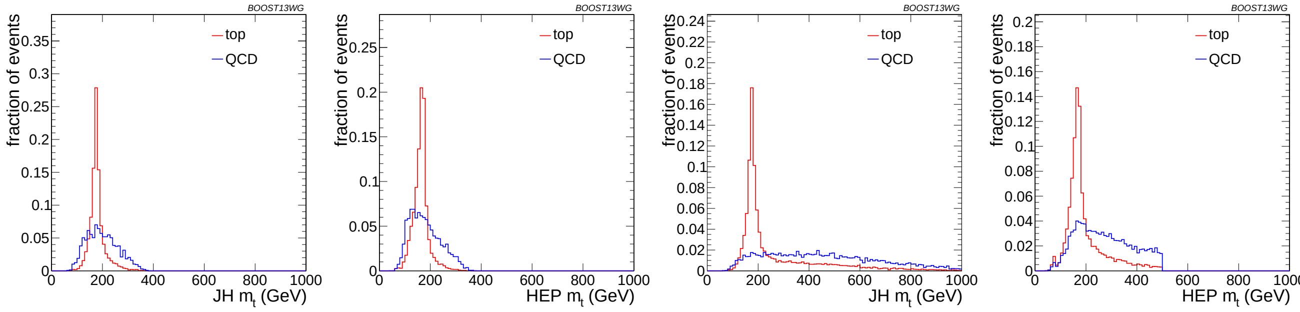
(c) W mass

Fig. 21 Comparison of single-variable top-tagging performance in the $p_T = 1 - 1.1$ GeV bin using the anti-k_{T}, R=0.8 algorithm.

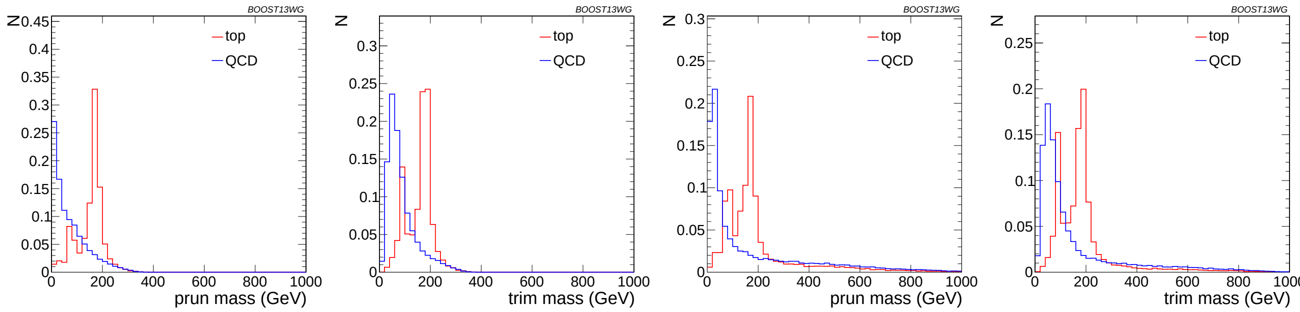
quirement tends to shape a peak in the QCD background around m_t for the HEPTopTagger, while the JH tagger has no such requirement. It has been suggested by Anders *et al.* [5] that performance in the HEPTopTagger may be improved by selecting the three subjets reconstructing the top only among those that pass the W mass constraints, which somewhat reduces the shaping of the background. The discrepancy between the JH and HEPTopTaggers is more pronounced at higher p_T and larger jet radius (see Figs. 26 and 29).

We also see in Figure 21(b) that the top mass from the JH tagger and the HEPTopTagger has superior performance relative to either of the grooming algorithms; this is because the pruning and trimming algorithms do not have inherent W-identification steps and are not optimized for this purpose. Indeed, because of the lack of a W-identification step, grooming algorithms are forced to strike a balance between under-grooming the jet, which broadens the signal peak due to UE contamination and features a larger background rate, and over-grooming the jet, which occasionally throws out the b-jet and preserves only the W components inside the jet. We demonstrate this effect in Figures 22 and 23, showing that with $\varepsilon_{\mathrm{sig}} = 0.3 - 0.35$, the optimal performance of the tagger over-grooms a substantial fraction of the jets ($\sim 20 - 30\%$), leading to a spurious second peak at the W mass. This effect is more pronounced at large R and p_T, since more aggressive grooming is required in these limits to combat the increased contamination from UE and QCD radiation.

In Figures 24 and 26 we directly compare ROC curves for jet shape observable performance and top mass performance respectively in the three different p_T bins considered whilst keeping the jet radius fixed at R=0.8. The input parameters of the taggers, groomers and shape variables are separately optimized in each p_T bin. One can see from Figure 24 that the tagging performance of jet shapes do not change substantially with p_T. The observables $\tau_{32}^{(\beta=1)}$ and Qjet volatility Γ have the most variation and tend to degrade with higher p_T, as can be seen in Figure 25. This makes sense, as higher-p_T QCD jets have more, harder emissions within the jet, giving rise to substructure that fakes the signal. By contrast, from Figure 26 we can see that most of the top mass observables have superior performance at higher p_T due to the radiation from the top quark becoming more collimated. The notable exception is the HEPTopTagger, which

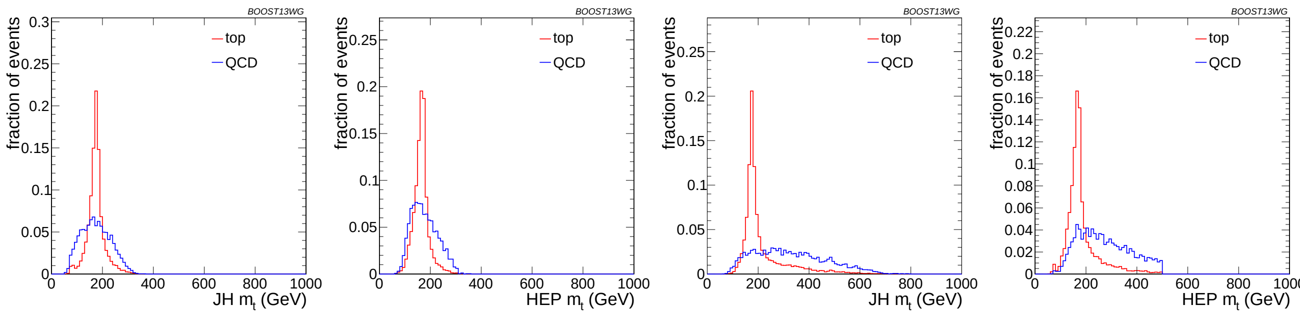


(a) JH, $R = 0.4$ (b) HEP, $R = 0.4$ (c) JH, $R = 1.2$ (d) HEP, $R = 1.2$

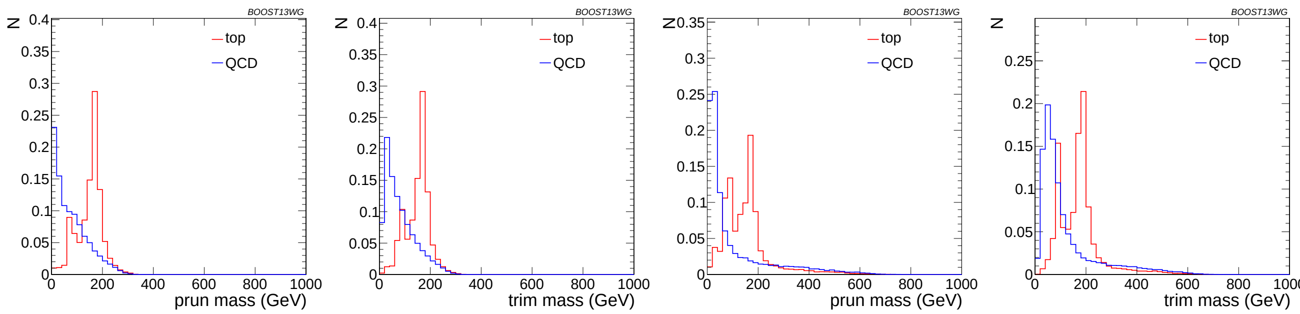


(e) prune, $R = 0.4$ (f) trim, $R = 0.4$ (g) prune, $R = 1.2$ (h) trim, $R = 1.2$

Fig. 22 Comparison of top mass reconstruction with the Johns Hopkins (JH), HEPTopTaggers (HEP), pruning, and trimming at different R using the anti-k_{T} algorithm, $p_{\mathrm{T}} = 1.5-1.6$ TeV. Each histogram is shown for the working point optimized for best performance with m_t in the $0.3-0.35$ signal efficiency bin, and is normalized to the fraction of events passing the tagger. In this and subsequent plots, the HEPTopTagger distribution cuts off at 500 GeV because the tagger fails to tag jets with a larger mass.



(a) JH, $p_T = 600-700$ GeV (b) HEP, $p_T = 600-700$ GeV (c) JH, $p_T = 1.5-1.6$ TeV (d) HEP, $p_T = 1.5-1.6$ TeV



(e) prune, $p_T = 600-700$ GeV (f) trim, $p_T = 600-700$ GeV (g) prune, $p_T = 1.5-1.6$ TeV (h) trim, $p_T = 1.5-1.6$ TeV

Fig. 23 Comparison of top mass reconstruction with the Johns Hopkins (JH), HEPTopTaggers (HEP), pruning, and trimming at different p_T using the anti-k_{T} algorithm, $R = 0.8$. Each histogram is shown for the working point optimized for best performance with m_t in the $0.3-0.35$ signal efficiency bin, and is normalized to the fraction of events passing the tagger.

degrades at higher p_T, likely in part due to the background-shaping effects discussed earlier.

In Figures 27 and 29 we directly compare ROC curves for jet shape observable performance and top mass performance respectively for the three different jet radii considered within the p_T 1.5-1.6 TeV bin. Again, the input parameters of the taggers, groomers and shape variables are separately optimized for each jet radius. We can see from these figures that most of the top tagging variables, both shape and reconstructed top mass, perform best for smaller radius. This is likely because, at such high p_T, most of the radiation from the top quark is confined within $R = 0.4$, and having a larger jet radius makes the observable more susceptible to contamination from the underlying event and other uncorrelated radiation. In Figure 28, we compare the individual top signal and QCD background distributions for each shape variable considered in the p_T 1.5-1.6 TeV bin for the various jet radii. One can see that the distributions for both signal and background broaden with increasing R, degrading the discriminating power. For $C_2^{(\beta=1)}$ and $C_3^{(\beta=1)}$, the background distributions are shifted upward as well. Therefore, the discriminating power generally gets worse with increasing R. The main exception is for $C_3^{(\beta=1)}$, which performs optimally at $R = 0.8$; in this case, the signal and background coincidentally happen to have the same distribution around $R = 0.4$, and so $R = 0.8$ gives better discrimination.

7.3 Performance of multivariable combinations

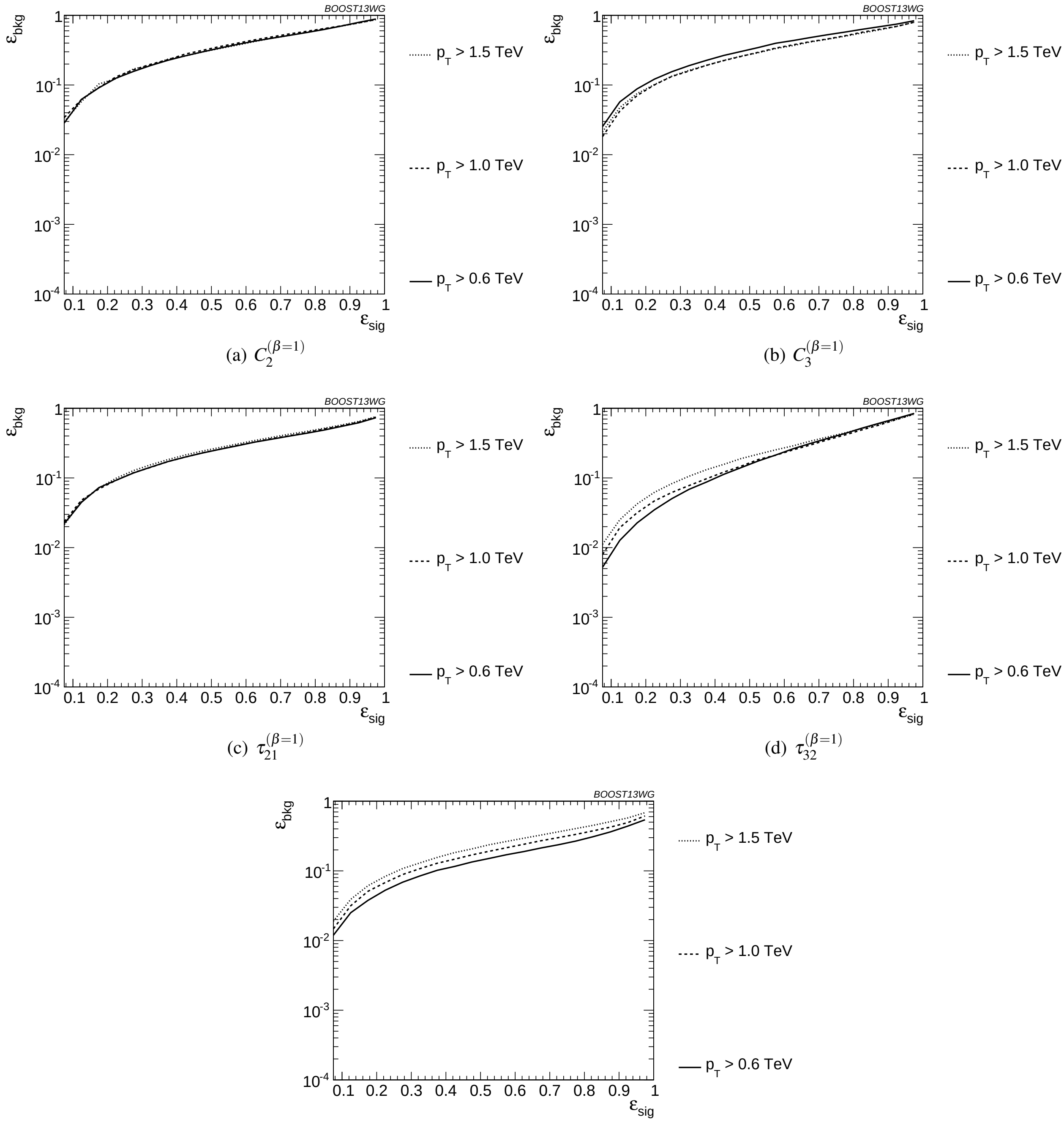
We now consider various BDT combinations of the observables from Section 7.2, using the techniques described in Section 4. In particular, we consider the performance of individual taggers such as the JH tagger and HEPTopTagger, which output information about the top and W candidate masses and the helicity angle; groomers, such as trimming and pruning, which remove soft, uncorrelated radiation from the top candidate to improve mass reconstruction, and to which we have added a W reconstruction step; and the combination of the outputs of the above taggers/groomers, both with each other, and with shape variables such as N-subjettiness ratios and energy correlation ratios. For all observables with tuneable input parameters, we scan and optimize over realistic values of such parameters, as described in Section 7.1.

In Figure 30, we directly compare the performance of the HEPTopTagger, the JH tagger, trimming, and pruning, in the $p_T = 1-1.1$ TeV bin using jet radius R=0.8, where both m_t and m_W are used in the groomers. Generally, we find that pruning, which does not naturally incorporate subjets into the algorithm, does not perform as well as the others. Interestingly, trimming, which does include a subjet-identification step, performs comparably to the HEPTopTagger over much of the range, possibly due to the background-shaping observed in Section 7.2. By contrast, the JH tagger outperforms the other algorithms. To determine whether there is complementary information in the mass outputs from different top taggers, we also consider in Figure 30 a multivariable combination of all of the JH and HEPTopTagger outputs. The maximum efficiency of the combined JH and HEPTopTaggers is limited, as some fraction of signal events inevitably fails either one or other of the taggers. We do see a 20-50% improvement in performance when combining all outputs, which suggests that the different algorithms used to identify the top and W for different taggers contains complementary information.

In Figure 31 we present the results for multivariable combinations of the top tagger outputs with and without shape variables. We see that, for both the HEPTopTagger and the JH tagger, the shape observables contain additional information uncorrelated with the masses and helicity angle, and give on average a factor 2-3 improvement in signal discrimination. We see that, when combined with the tagger outputs, both the energy correlation functions $C_2 + C_3$ and the N-subjettiness ratios $\tau_{21} + \tau_{32}$ give comparable performance, while the Qjet mass volatility is slightly worse; this is unsurprising, as Qjets accesses shape information in a more indirect way from other shape observables. Combining all shape observables with a single top tagger provides even greater enhancement in discrimination power. We directly compare the performance of the JH and HEPTopTaggers in Figure 31(c). Combining the taggers with shape information nearly erases the difference between the tagging methods observed in Figure 30; this indicates that combining the shape information with the HEPTopTagger identifies the differences between signal and background missed by the tagger alone. This also suggests that further improvement to discriminating power may be minimal, as various multivariable combinations are converging to within a factor of 20% or so.

In Figure 32 we present the results for multivariable combinations of groomer outputs with and without shape variables. As with the tagging algorithms, combinations of groomers with shape observables improves their discriminating power; combinations with $\tau_{32} + \tau_{21}$ perform comparably to those with $C_3 + C_2$, and both of these are superior to combinations with the mass volatility, Γ. Substantial improvement is further possible by combining the groomers with all shape observables. Not surprisingly, the taggers that lag behind in performance enjoy the largest gain in signal-background discrimination with the addition of shape observables. Once again, in Figure 32(c), we find that the differences between pruning and trimming are erased when combined with shape information.

Finally, in Figure 33, we compare the performance of each of the tagger/groomers when their outputs are combined with all of the shape observables considered. One can

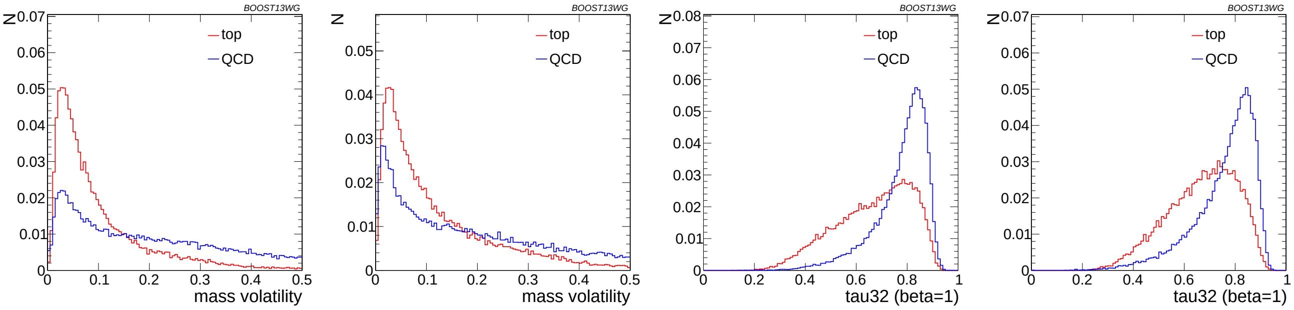


(a) $C_2^{(\beta=1)}$ (b) $C_3^{(\beta=1)}$ (c) $\tau_{21}^{(\beta=1)}$ (d) $\tau_{32}^{(\beta=1)}$ (e) Qjet mass volatility

Fig. 24 Comparison of individual jet shape performance at different p_T using the anti-k_T R=0.8 algorithm.

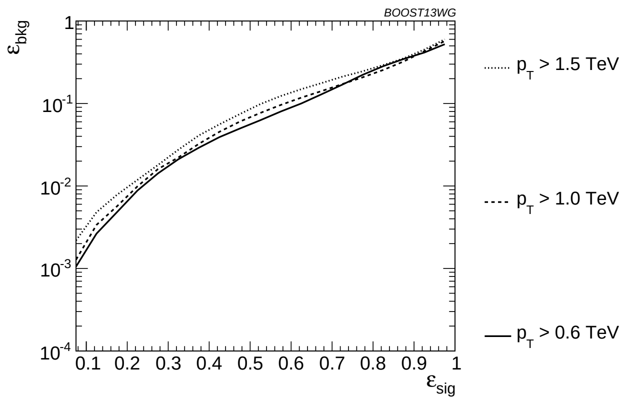
see that the discrepancies between the performance of the different taggers/groomers all but vanishes, suggesting perhaps that we are here utilising all available signal-background discrmination information, and that this is the optimal top-tagging performance that could be achieved in these conditions.

Up to this point we have just considered the combined multivariable performance in the p_T 1.0-1.1 TeV bin with jet radius R=0.8. We now compare the BDT combinations of tagger outputs, with and without shape variables, at different p_T. The taggers are optimized over all input parameters for each choice of p_T and signal efficiency. As with the single-variable study, we consider anti-k_T jets clustered with $R = 0.8$ and compare the outcomes in the $p_T = 500-600$ GeV, $p_T = 1-1.1$ TeV, and $p_T = 1.5-1.6$ TeV bins. The comparison of the taggers/groomers is shown in Figure 34.

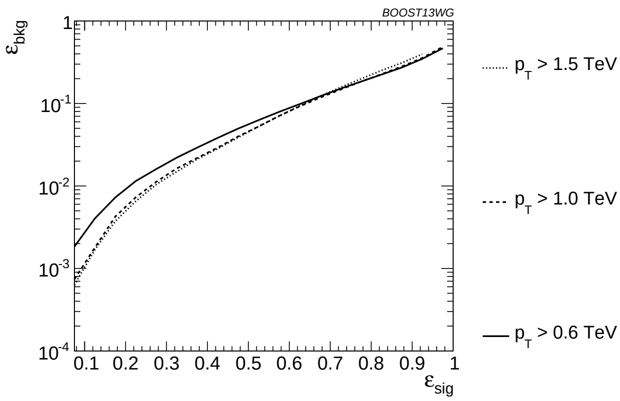


(a) $\Gamma_{\rm Qjet}$, $p_T = 600-700$ GeV (b) $\Gamma_{\rm Qjet}$, $p_T = 1.5-1.6$ TeV (c) $\tau_{32}^{(\beta=1)}$, $p_T = 600-700$ GeV (d) $\tau_{32}^{(\beta=1)}$, $p_T = 1.5-1.6$ TeV

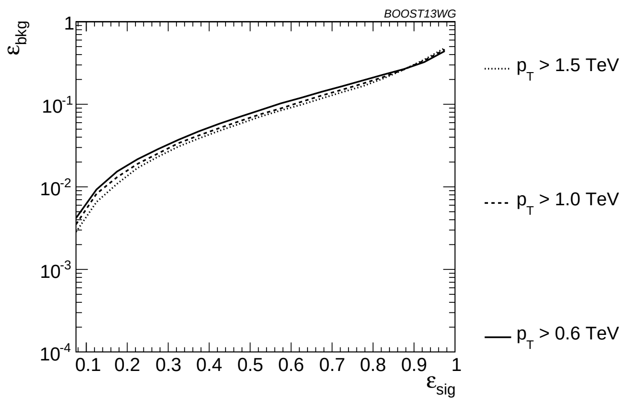
Fig. 25 Comparison of $\Gamma_{\rm Qjet}$ and $\tau_{32}^{\beta=1}$ at $R = 0.8$ and different values of the p_T. These shape observables are the most sensitive to varying p_T.



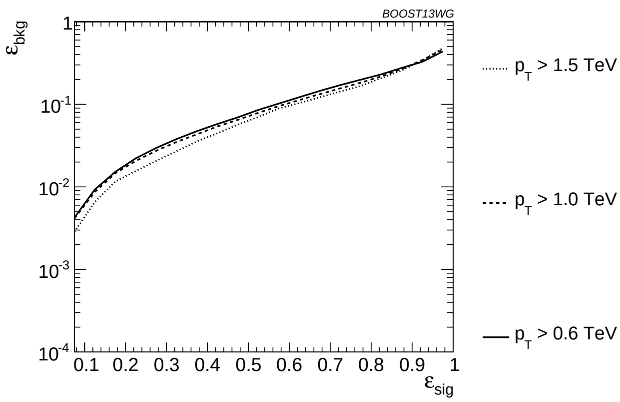
(a) HEPTopTagger m_t



(b) Johns Hopkins Tagger m_t



(c) Pruning m_t

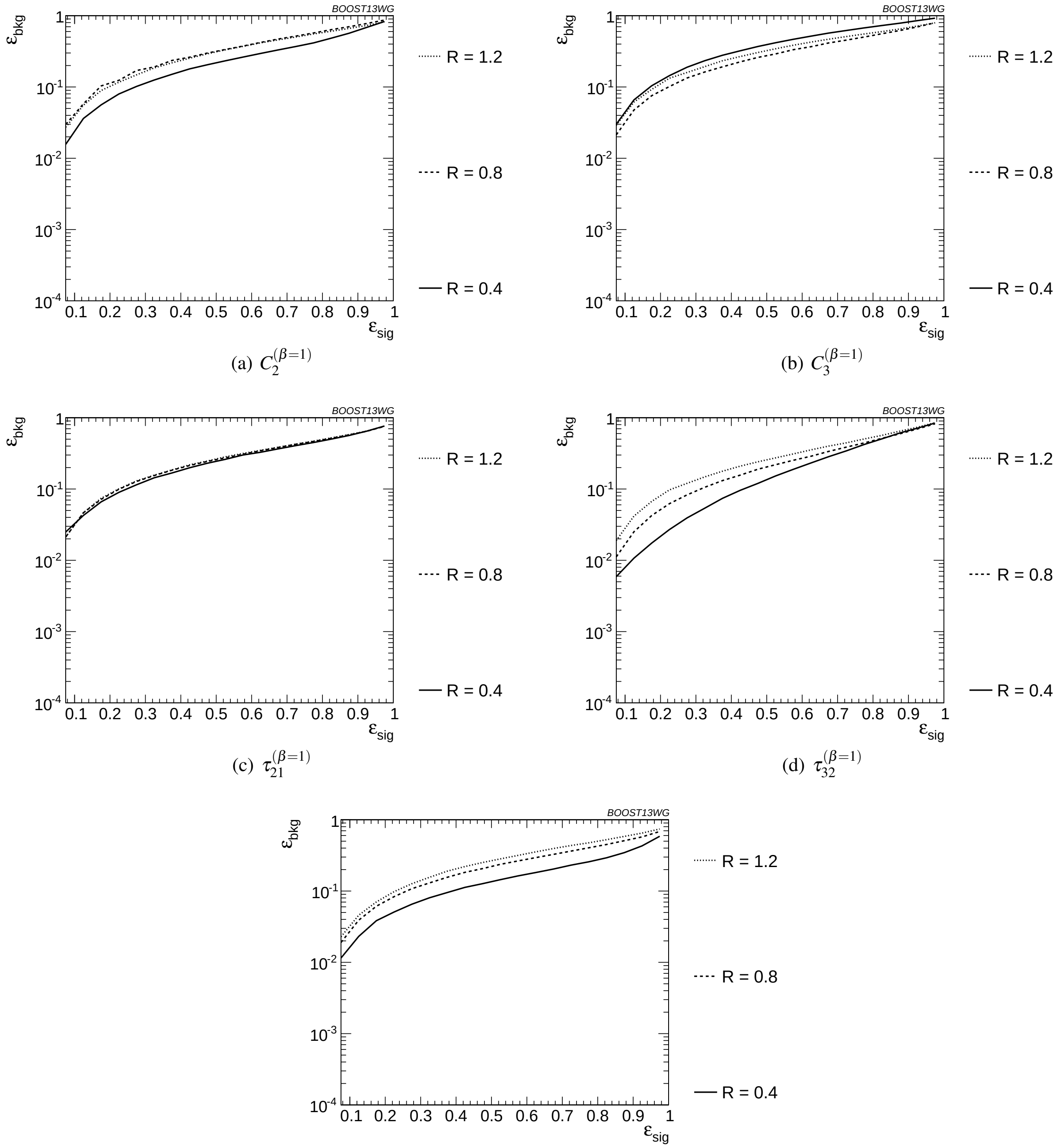


(d) Trimming m_t

Fig. 26 Comparison of top mass performance of different taggers at different p_T using the anti-$k_{\rm T}$ R=0.8 algorithm.

The behaviour with p_T is qualitatively similar to the behaviour of the m_t observable for each tagger/groomer shown in Figure 26; this suggests that the p_T behaviour of the taggers is dominated by the top mass reconstruction. As before, the HEPTopTagger performance degrades slightly with increased p_T due to the background shaping effect, while the JH tagger and groomers modestly improve in performance.

In Figure 35, we show the p_T dependence of BDT combinations of the JH tagger output combined with shape observables. We find that the curves look nearly identical: the p_T dependence is dominated by the top mass reconstruction, and combining the tagger outputs with different shape observables does not substantially change this behaviour. The same holds true for trimming and pruning. By contrast, HEPTopTagger ROC curves, shown in Figure 36, do change somewhat when combined with different shape observables; due to the suboptimal performance of the HEPTopTagger at high p_T , we find that combining the HEPTopTagger with

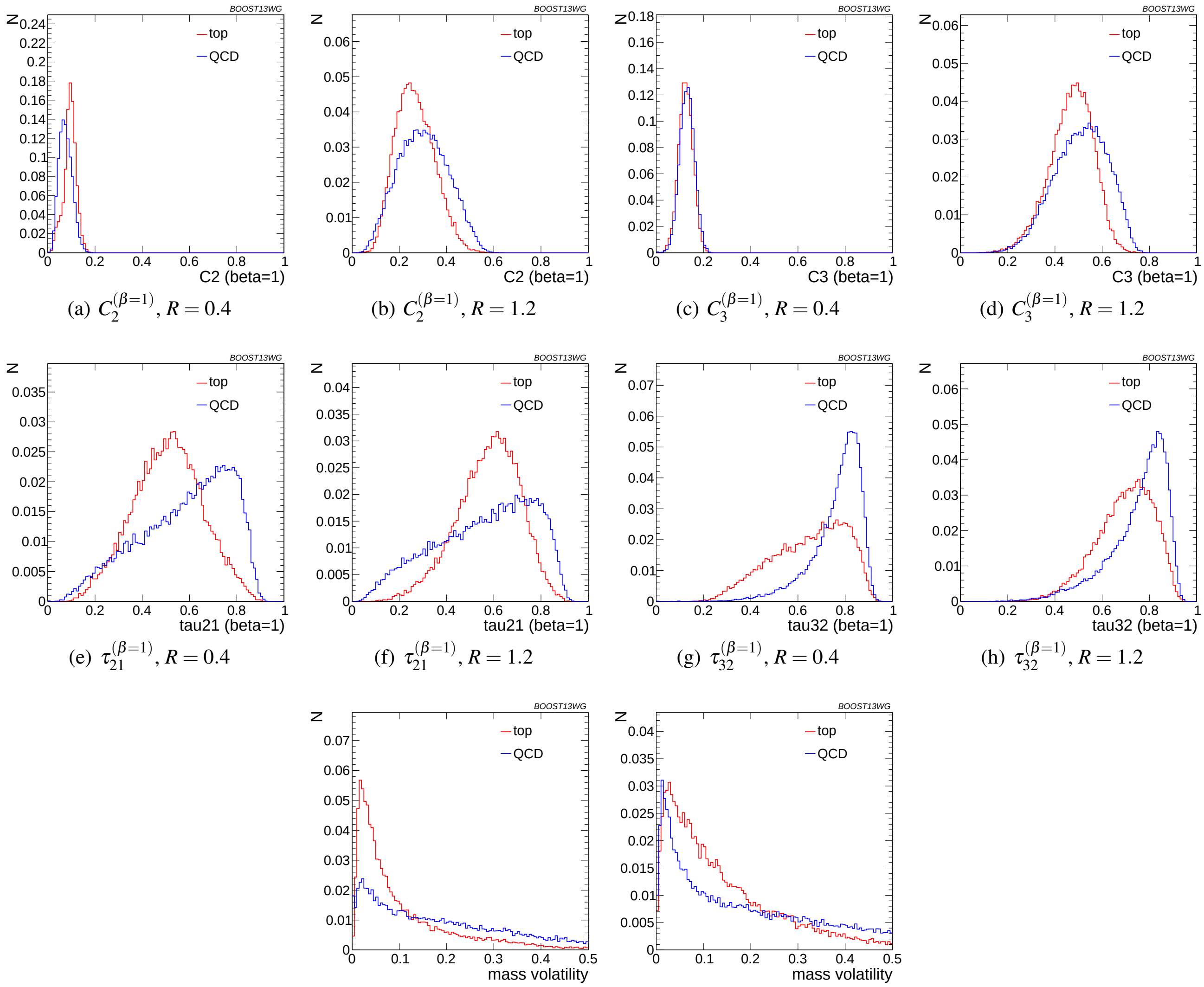


(a) $C_2^{(\beta=1)}$ (b) $C_3^{(\beta=1)}$ (c) $\tau_{21}^{(\beta=1)}$ (d) $\tau_{32}^{(\beta=1)}$ (e) Qjet mass volatility

Fig. 27 Comparison of individual jet shape performance at different R in the $p_T = 1.5-1.6$ TeV bin.

$C_3^{(\beta=1)}$, which in Figure 24(b) is seen to have some modest improvement at high p_T , can improve its performance. Combining the HEPTopTagger with multiple shape observables gives the maximum improvement in performance at high p_T relative to at low p_T .

In Figure 37 we compare the BDT combinations of tagger outputs, with and without shape variables, at different jet radius R in the $p_T = 1.5-1.6$ TeV bin. The taggers are optimized over all input parameters for each choice of R and signal efficiency. We find that, for all taggers and groomers, the performance is always best at small R; the choice of R is sufficiently large to admit the full top quark decay at such high p_T , but is small enough to suppress contamination from additional radiation. This is not altered when the taggers are combined with shape observable. For example, in Figure 38



(a) $C_2^{(\beta=1)}$, $R = 0.4$ (b) $C_2^{(\beta=1)}$, $R = 1.2$ (c) $C_3^{(\beta=1)}$, $R = 0.4$ (d) $C_3^{(\beta=1)}$, $R = 1.2$

(e) $\tau_{21}^{(\beta=1)}$, $R = 0.4$ (f) $\tau_{21}^{(\beta=1)}$, $R = 1.2$ (g) $\tau_{32}^{(\beta=1)}$, $R = 0.4$ (h) $\tau_{32}^{(\beta=1)}$, $R = 1.2$

(i) Γ_{Qjet}, $R = 0.4$ (j) Γ_{Qjet}, $R = 1.2$

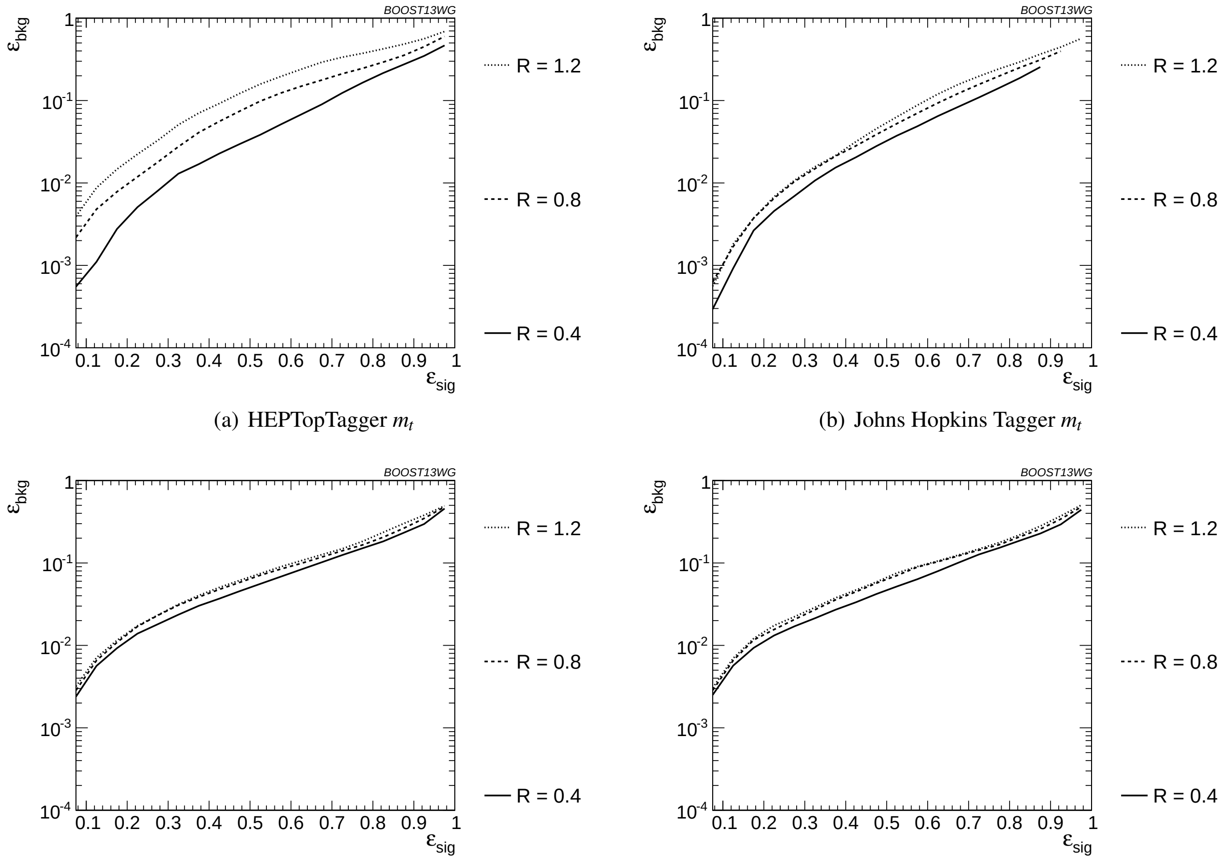
Fig. 28 Comparison of various shape observables in the $p_T = 1.5 - 1.6$ TeV bin and different values of the anti-k_{T} radius R.

is shown the dependence on R of the JH tagger when combined with shape observables, where one can see that the R-dependence is identical for all combinations. The same holds true for the HEPTopTagger, trimming, and pruning.

7.4 Performance at Sub-Optimal Working Points

Up until now, we have re-optimized our tagger and groomer parameters for each p_T, R, and signal efficiency working point. In reality, experiments will choose a finite set of working points to use. How do our results hold up when this is taken into account? To address this concern, we replicate our analyses, but only optimize the top taggers for a particular p_T/R/efficiency and apply the same parameters to other scenarios. This allows us to determine the extent to which re-optimization is necessary to maintain the high signal-background discrimination power seen in the top tagging algorithms we study. The shape observables typically do not have any input parameters to optimize. Therefore, we focus on the taggers and groomers, and their combination with shape observables, in this section.

Optimizing at a single p_T : We show in Figure 39 the performance of the top taggers, using just the reconstructed top mass as the discriminating variable, with all input parameters optimized to the $p_T = 1.5 - 1.6$ TeV bin, relative to the performance optimized at each p_T. We see that while the performance degrades by about 50% when the high-p_T optimized points are used at other momenta, this is only an order-one adjustment of the tagger performance, with trimming and the Johns Hopkins tagger degrading the most. The jagged behaviour of the points is due to the finite resolution of the scan. We also observe a particular effect asso-



(a) HEPTopTagger m_t

(b) Johns Hopkins Tagger m_t

(c) Pruning m_t

(d) Trimming m_t

Fig. 29 Comparison of top mass performance of different taggers at different R in the $p_T = 1.5 - 1.6$ TeV bin.

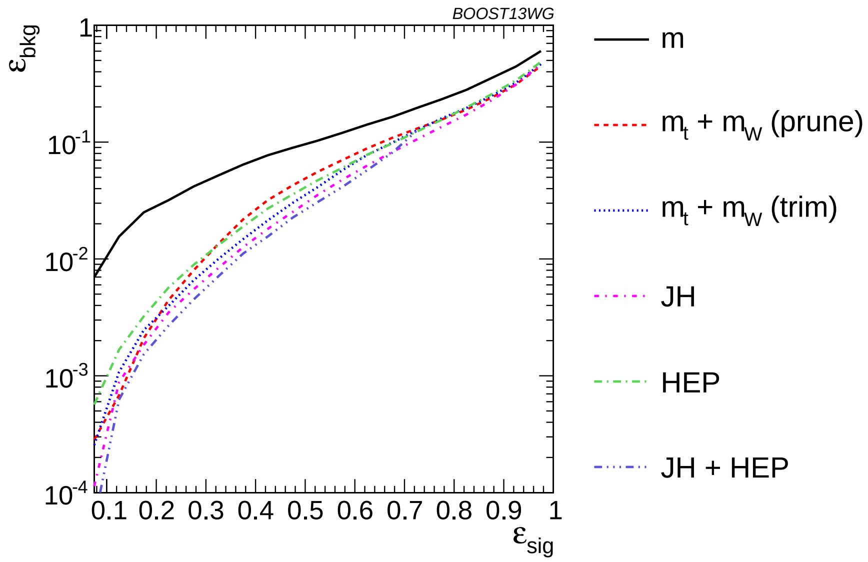
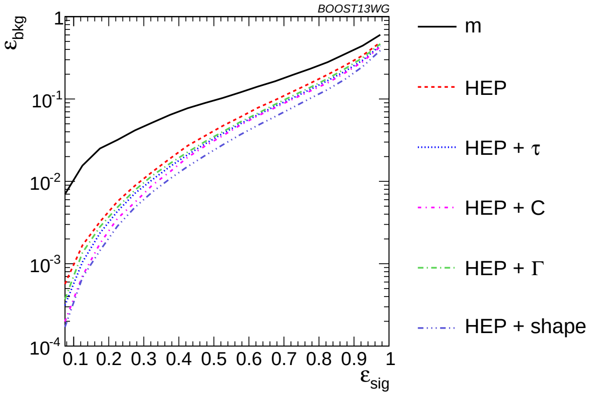
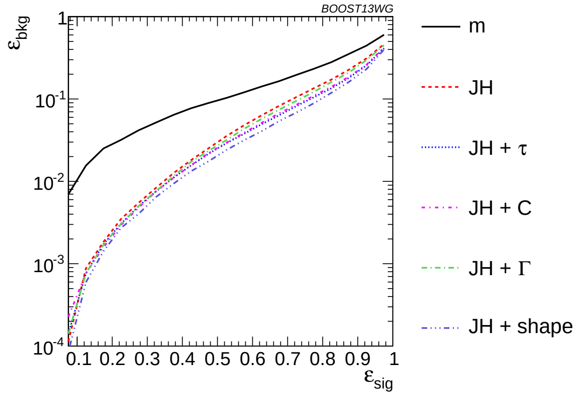


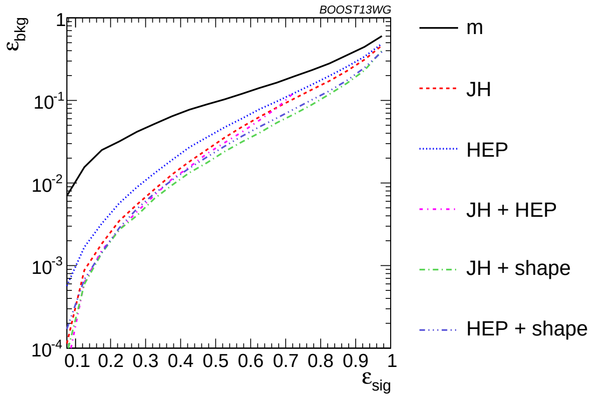
Fig. 30 The performance of the various taggers in the $p_T = 1 - 1.1$ TeV bin using the anti-k_T R=0.8 algorithm. For the groomers a BDT combination of the reconstructed $m_t$and $m_W$are used. Also shown is a multivariable combination of all of the JH and HEPTopTagger outputs. The ungroomed mass performance is shown for comparison.



(a) HEPTopTagger + Shape



(b) Johns Hopkins Tagger + shape



(c) HEP vs. JH comparison (incl. shape)

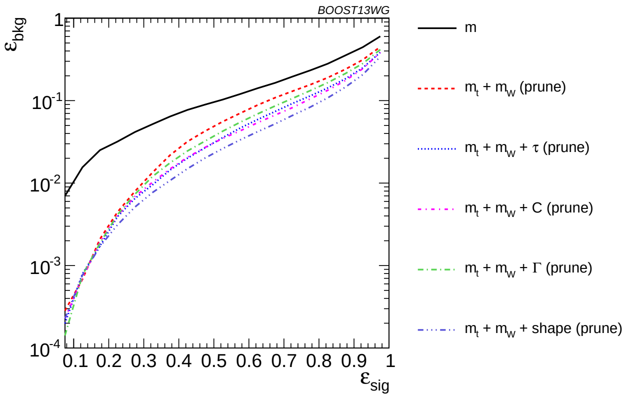
Fig. 31 The performance of BDT combinations of the JH and HepTopTagger outputs with various shape observables in the $p_T = 1-1.1$ TeV bin using the anti-k_T R=0.8 algorithm. Taggers are combined with the following shape observables: $\tau_{21}^{(\beta=1)} + \tau_{32}^{(\beta=1)}$, $C_2^{(\beta=1)} + C_3^{(\beta=1)}$, Γ_{Qjet}, and all of the above (denoted "shape").

ciated with using suboptimal taggers: since taggers sometimes fail to return a top candidate, parameters optimized for a particular efficiency ε_S at $p_T = 1.5-1.6$ TeV may not return enough signal candidates to reach the same efficiency at a different p_T. Consequently, no point appears for that p_T value. This is not often a practical concern, as the largest gains in signal discrimination and significance are for smaller values of ε_S, but it is something that must be considered when selecting benchmark tagger parameters and signal efficiencies.

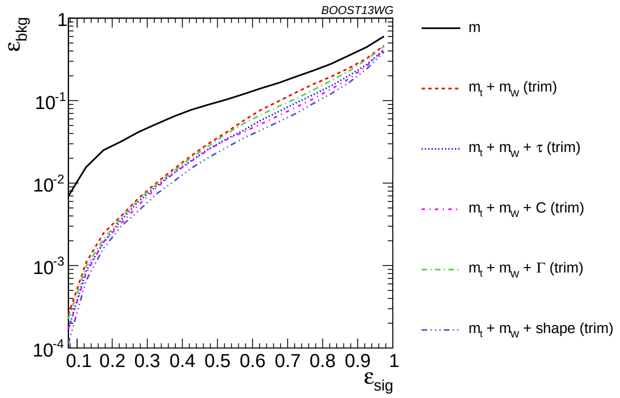
The degradation in performance is more pronounced for the BDT combinations of the full tagger outputs, shown in Figure 40), particularly at very low signal efficiency where the optimization picks out a cut on the tail of some distribution that depends precisely on the p_T/R of the jet. Once again, trimming and the Johns Hopkins tagger degrade more markedly. Similar behaviour holds for the BDT combinations of tagger outputs plus all shape observables.

Optimizing at a single R: We perform a similar analysis, optimizing tagger parameters for each signal efficiency at $R = 1.2$, and then use the same parameters for smaller R, in the p_T 1.5-1.6 TeV bin. In Figure 41 we show the ratio of the performance of the top taggers, using just the reconstructed top mass as the discriminating variable, with all input parameters optimized to the $R = 1.2$ values compared to input parameters optimized separately at each radius. While the performance of each observable degrades at small ε_{sig} compared to the optimized search, the HEPTopTagger fares the worst as the observed is quite sensitive to the selected value of R. It is not surprising that a tagger whose top mass reconstruction is susceptible to background-shaping at large R and p_T would require a more careful optimization of parameters to obtain the best performance.

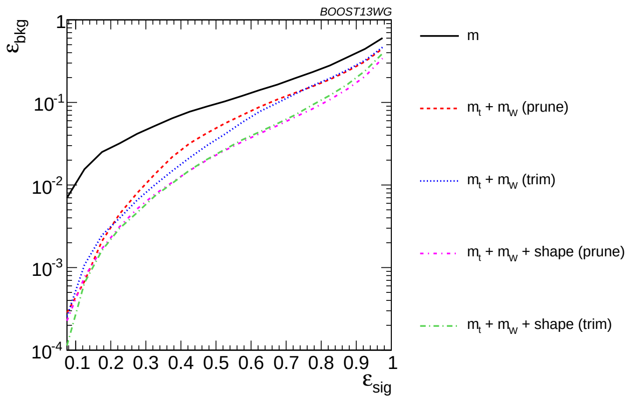
The same holds true for the BDT combinations of the full tagger outputs, shown in Figure 42). The performance for the sub-optimal taggers is still within an $O(1)$ factor of the optimized performance, and the HEPTopTagger performs better with the combination of all of its outputs relative to the performance with just m_t. The same behaviour holds for the BDT combinations of tagger outputs and shape



(a) Pruning + Shape



(b) Trimming + Shape



(c) Trim vs. Prune comparison (incl. shape)

Fig. 32 The performance of the BDT combinations of the trimming and pruning outputs with various shape observables in the $p_T = 1-1.1$ TeV bin using the anti-k_T R=0.8 algorithm. Groomer mass outputs are combined with the following shape observables: $\tau_{21}^{(\beta=1)} + \tau_{32}^{(\beta=1)}$, $C_2^{(\beta=1)} + C_3^{(\beta=1)}$, $\Gamma_{\rm Qjet}$, and all of the above (denoted "shape").

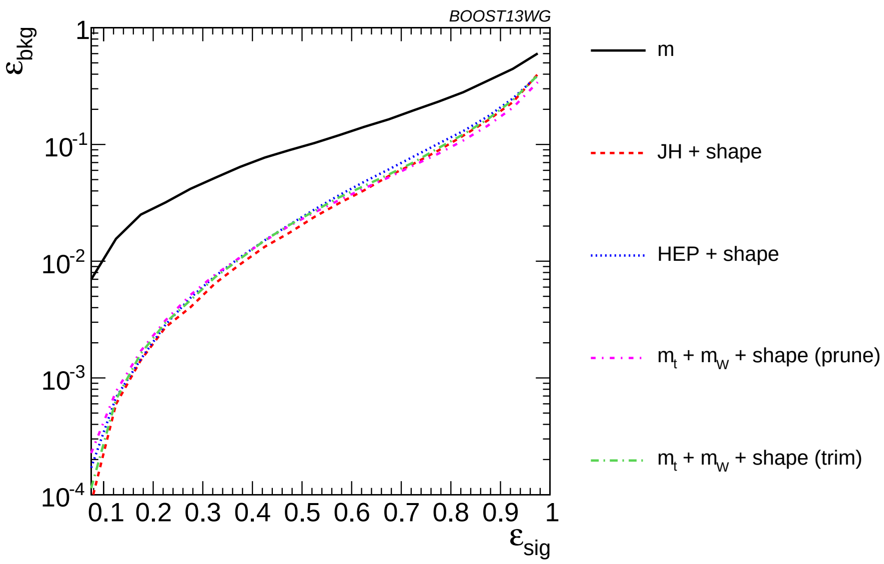
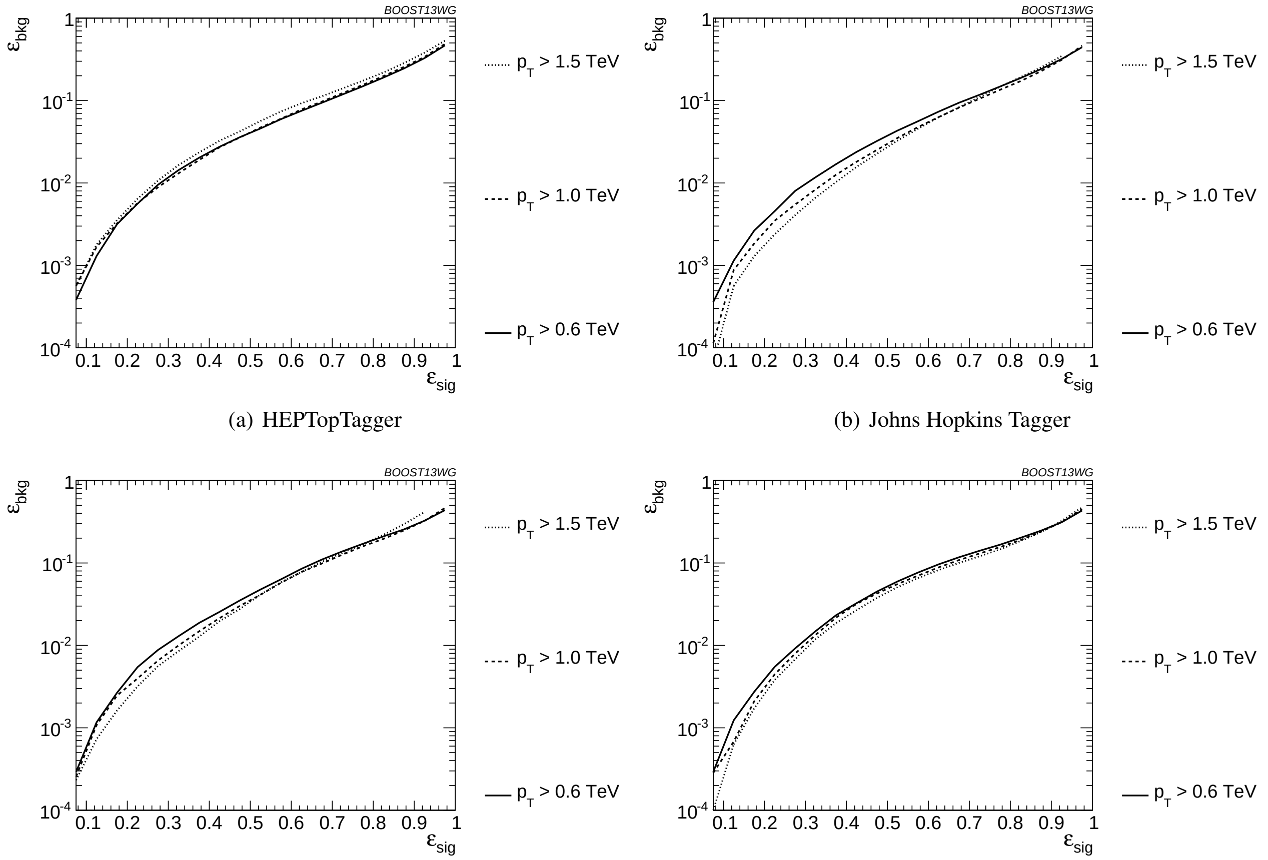


Fig. 33 Comparison of the performance of the BDT combinations of all the groomer/tagger outputs with all the available shape observables in the $p_T = 1-1.1$ TeV bin using the anti-k_T R=0.8 algorithm. Tagger/groomer outputs are combined with all of the following shape observables: $\tau_{21}^{(\beta=1)} + \tau_{32}^{(\beta=1)}$, $C_2^{(\beta=1)} + C_3^{(\beta=1)}$, $\Gamma_{\rm Qjet}$.



(a) HEPTopTagger (b) Johns Hopkins Tagger

(c) Trimming (d) Pruning

Fig. 34 Comparison of BDT combination of tagger performance at different p_T using the anti-k_{T} R=0.8 algorithm.

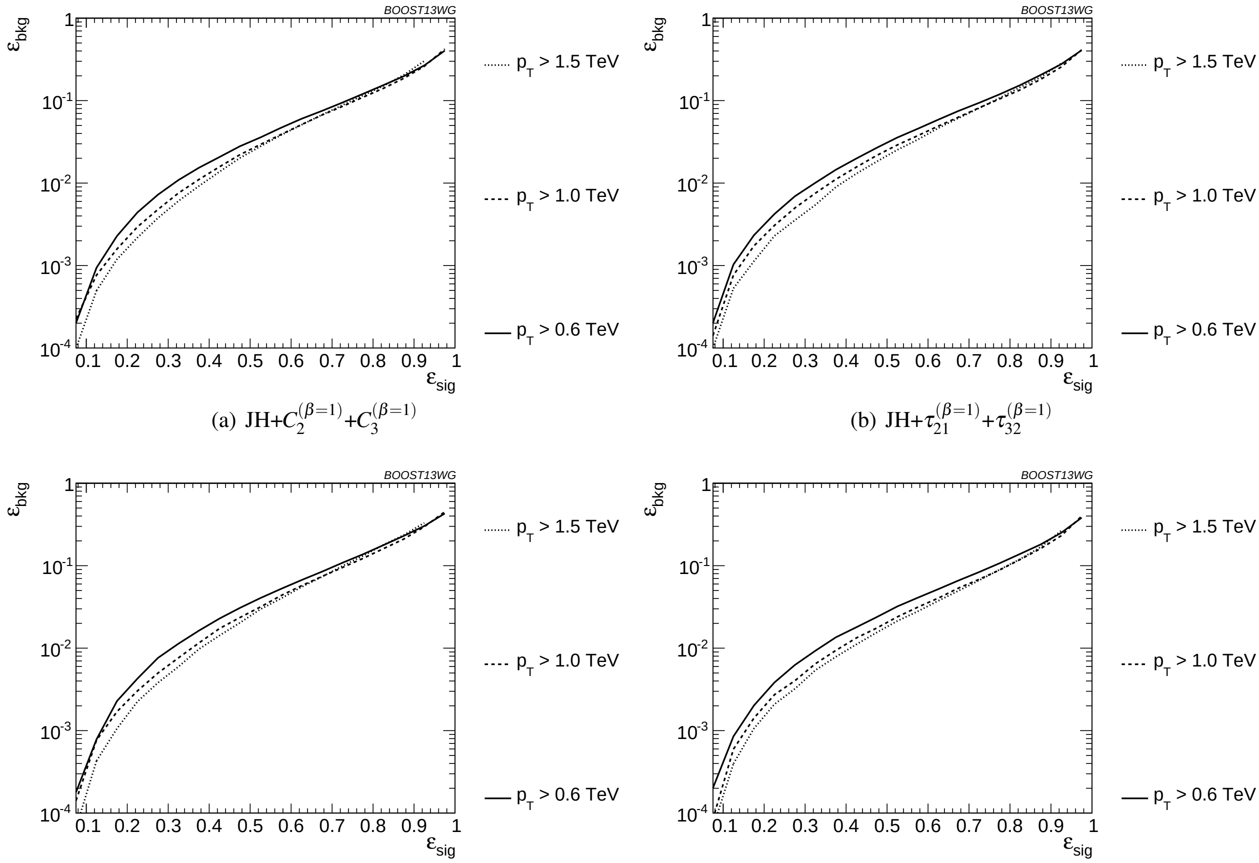
observables.

Optimizing at a single efficiency: The strongest assumption we have made so far is that the taggers can be reoptimized for each signal efficiency point. This is useful for making a direct comparison of the power of different top tagging algorithms, but is not particularly practical for the LHC analyses. We now consider the effects when the tagger inputs are optimized once, in the $\varepsilon_S = 0.3-0.35$ bin, and then used to determine the full ROC curve. We do this in the $p_T 1-1.1$ TeV bin and with $R = 0.8$.

The performance of each tagger, normalized to its performance optimized in each bin, is shown in Figure 43 for cuts on the top mass and W mass, and in Figure 44 for BDT combinations of tagger outputs and shape variables. In both plots, it is apparent that optimizing the taggers in the 0.3-0.35 efficiency bin gives comparable performance over efficiencies ranging from 0.2-0.5, although performance degrades at small and large signal efficiencies. Pruning appears to give especially robust signal-background discrimination without re-optimization, possibly due to the fact that there are no absolute distance or p_T scales that appear in the algorithm. Figures 43 and 44 suggest that, while optimization at all signal efficiencies is a useful tool for comparing different algorithms, it is not crucial to achieve good top-tagging performance in experiments.

7.5 Conclusions

We have studied the performance of various jet substructure observables, groomed masses, and top taggers to study the performance of top tagging at different p_T and jet radius parameter. At each p_T, R, and signal efficiency working point, we optimize the parameters for those observables with tuneable inputs. Overall, we have found that these techniques, individually and in combination, continue to perform well at high p_T, which is important for future LHC running. In general, the John Hopkins tagger performs best, while jet grooming algorithms under-perform relative to the best top taggers due to the lack of an optimized W-identification step. Tagger performance can be improved by a further factor of 2-4 through combination with jet substructure observables such as τ_{32}, C_3, and Qjet mass volatility; when combined with jet substructure observables, the performance of vari-



(a) JH+$C_2^{(\beta=1)}$+$C_3^{(\beta=1)}$

(b) JH+$\tau_{21}^{(\beta=1)}$+$\tau_{32}^{(\beta=1)}$

(c) JH + Qjet mass volatility

(d) JH + all

Fig. 35 Comparison of BDT combination of JH tagger + shape at different p_T using the anti-k_T R=0.8 algorithm.

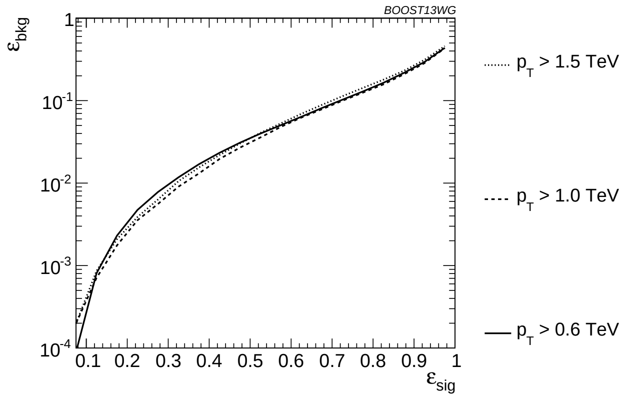
ous groomers and taggers becomes very comparable, suggesting that, taken together, the observables studied are sensitive to nearly all of the physical differences between top and QCD jets. A small improvement is also found by combining the Johns Hopkins and HEPTopTaggers, indicating that different taggers are not fully correlated.

Comparing results at different p_T and R, top tagging performance is generally better at smaller R due to less contamination from uncorrelated radiation. Similarly, most observables perform better at larger p_T due to the higher degree of collimation of radiation. Some observables fare worse at higher p_T, such as the N-subjettiness ratio τ_{32} and the Qjet mass volatility Γ, as higher-p_T QCD jets have more, harder emissions that fake the top jet substructure. The HEPTopTagger is also worse at large p_T due to the tendency of the tagger to shape backgrounds around the top mass. The p_T- and R-dependence of the multivariable combinations is dominated by the p_T- and R-dependence of the top mass reconstruction component of the tagger/groomer.

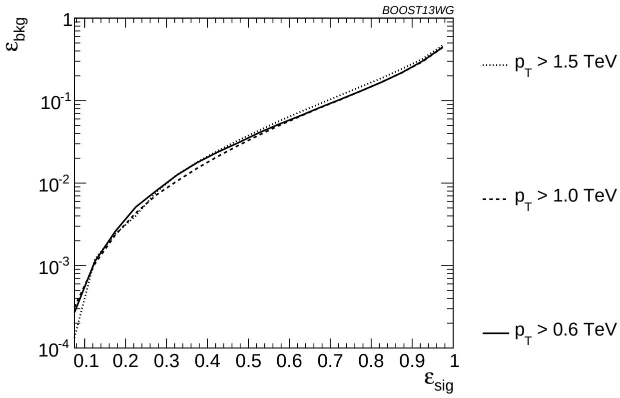
Finally, we consider the performance of various observable combinations under the more realistic assumption that the input parameters are only optimized at a single p_T, R, or signal efficiency, and then the same inputs are used at other working points. Remarkably, the performance of all observables is typically within a factor of 2 of the fully optimized inputs, suggesting that while optimization can lead to substantial gains in performance, the general behaviour found in the fully optimized analyses extends to more general applications of each variable. In particular, the performance of pruning typically varies the least when comparing suboptimal working points to the fully optimized tagger due to the scale-invariant nature of the pruning algorithm.

8 Summary & Conclusions

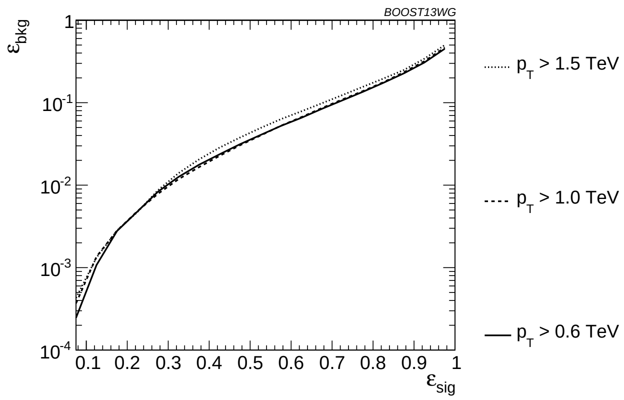
In this report we have attempted to understand the degree to which the discriminatory information in various jet substructure observables/taggers overlaps, and how this varies as a function of the parameters of the jets, such as their p_T and radius. This has been done by combining the variables into BDT discriminants, and comparing the background rejection power of this discriminant to the rejection power achieved by the individual variables. The performance of "all variables" BDT discriminants has also been investigated,



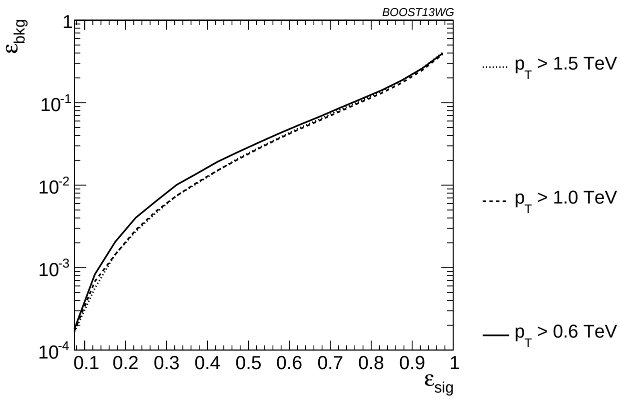
(a) HEP+$C_2^{(\beta=1)}$+$C_3^{(\beta=1)}$



(b) HEP+$\tau_{21}^{(\beta=1)}$+$\tau_{32}^{(\beta=1)}$

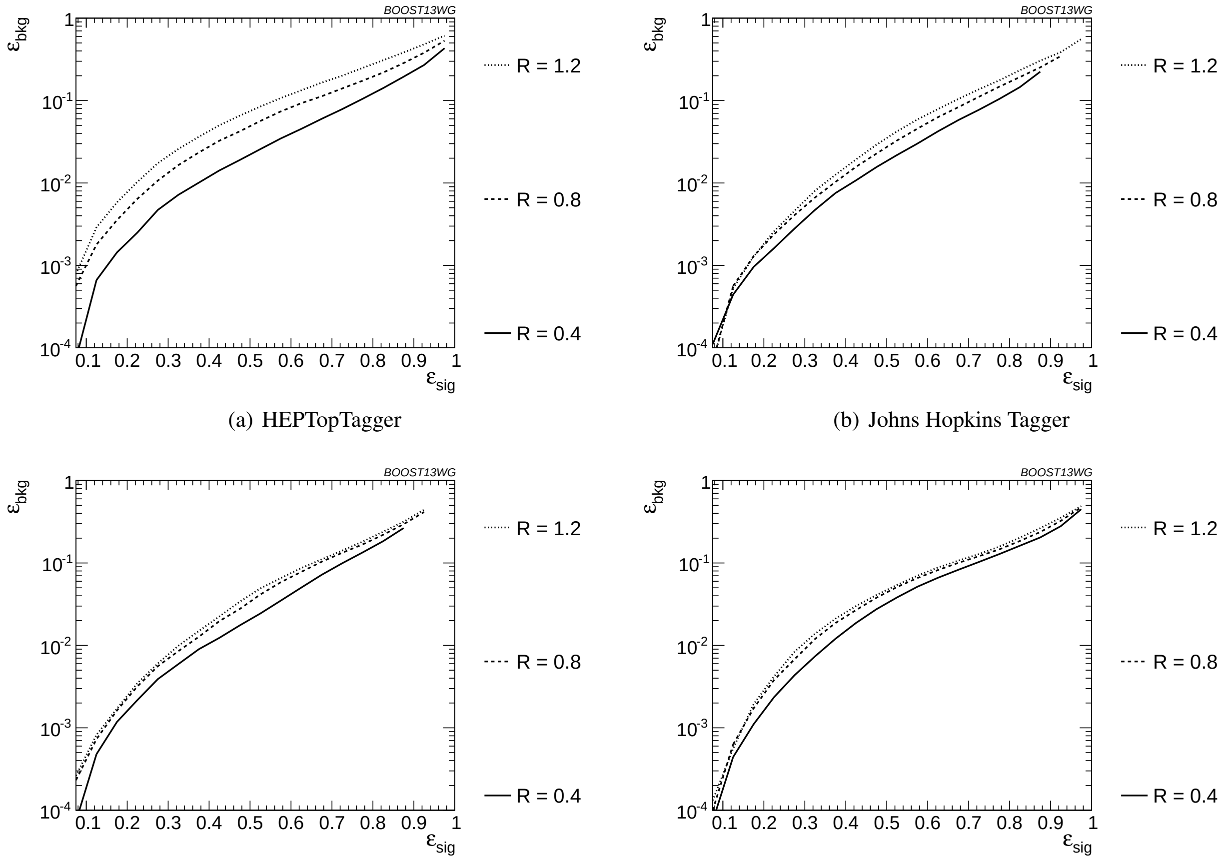


(c) HEP + Qjet mass volatility



(d) HEP + all

Fig. 36 Comparison of BDT combination of HEP tagger + shape at different p_T using the anti-k_T R=0.8 algorithm.



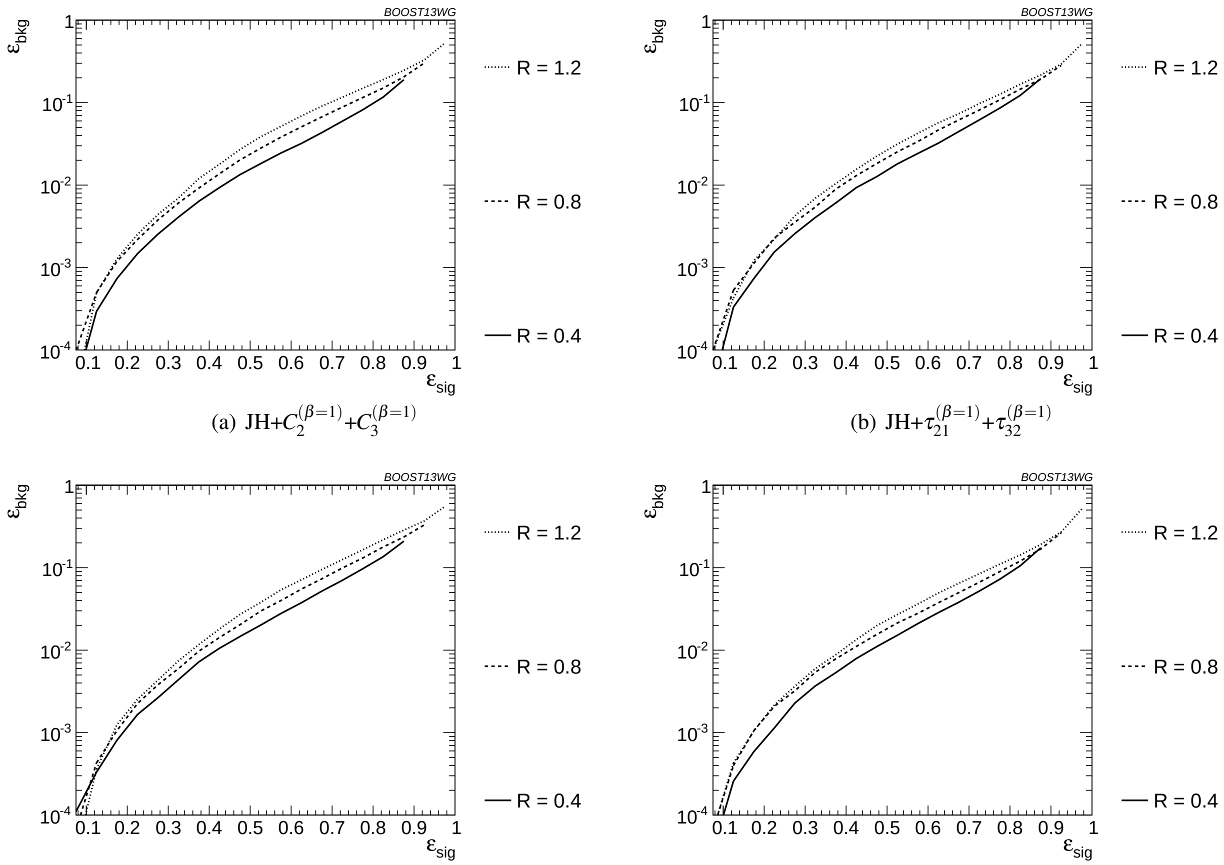
(a) HEPTopTagger

(b) Johns Hopkins Tagger

(c) Trimming

(d) Pruning

Fig. 37 Comparison of tagger and jet shape performance at different radius at p_T = 1.5-1.6 TeV.



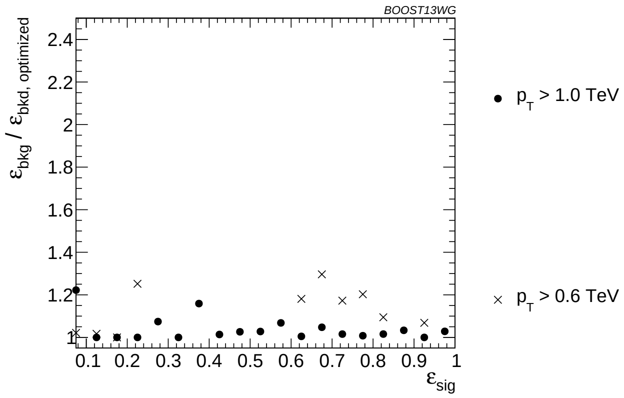
(a) JH+$C_2^{(\beta=1)}$+$C_3^{(\beta=1)}$

(b) JH+$\tau_{21}^{(\beta=1)}$+$\tau_{32}^{(\beta=1)}$

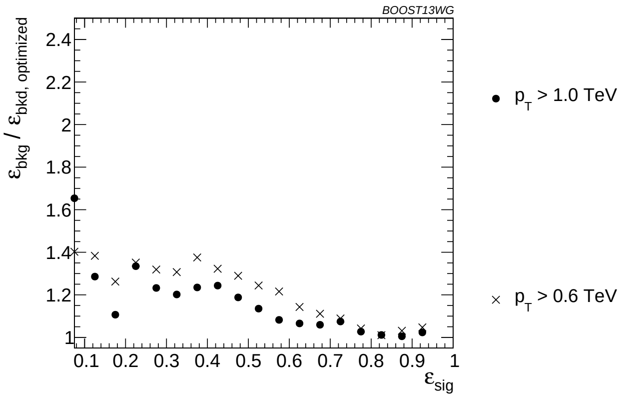
(c) JH + Qjet mass volatility

(d) JH + all

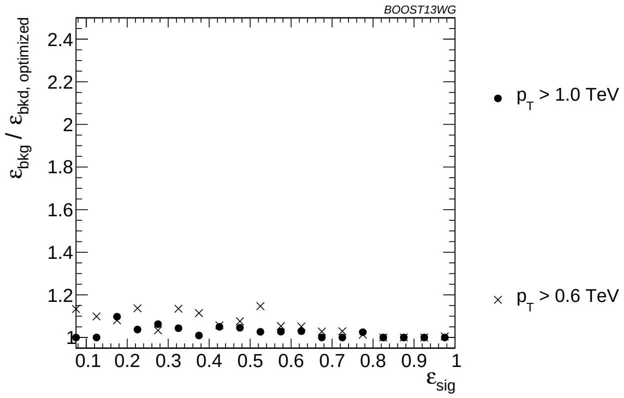
Fig. 38 Comparison of BDT combination of JH tagger + shape at different radius at p_T = 1.5-1.6 TeV.



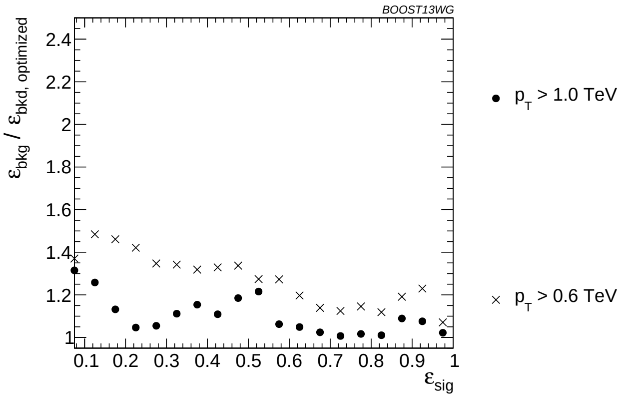
(a) HEPTopTagger m_t



(b) Johns Hopkins Tagger m_t

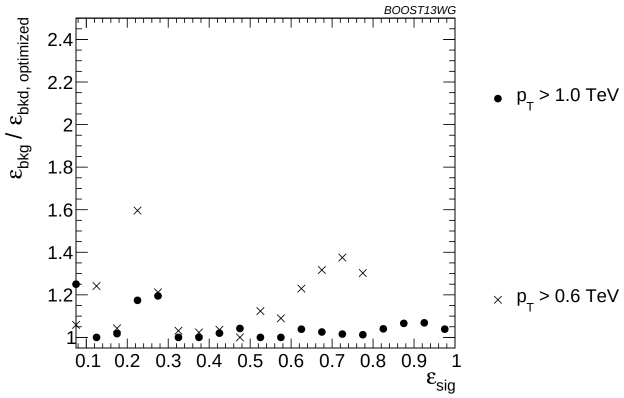


(c) Pruning m_t

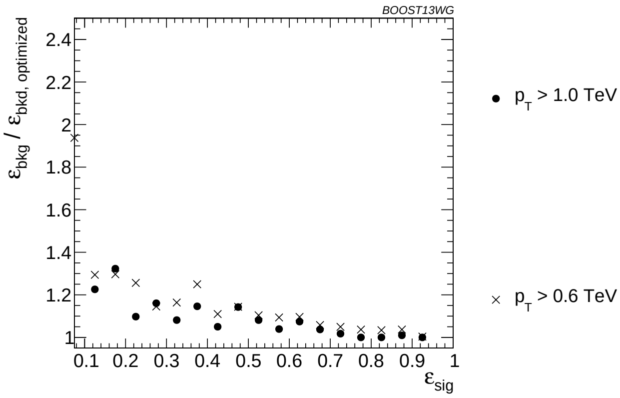


(d) Trimming m_t

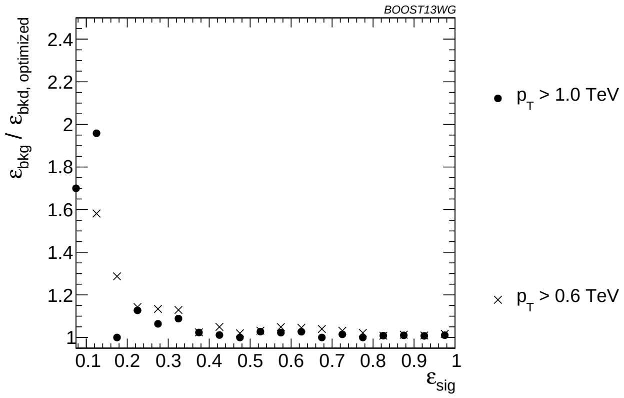
Fig. 39 Comparison of top mass performance of different taggers at different p_T using the anti-k_{T} R=0.8 algorithm; the tagger inputs are set to the optimum value for $p_T = 1.5 - 1.6$ TeV.



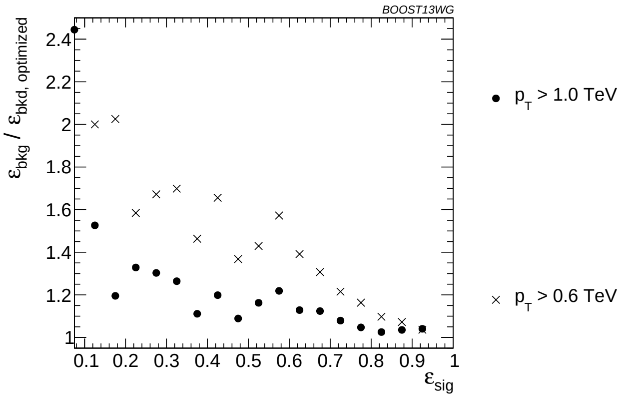
(a) HEPTopTagger



(b) Johns Hopkins Tagger

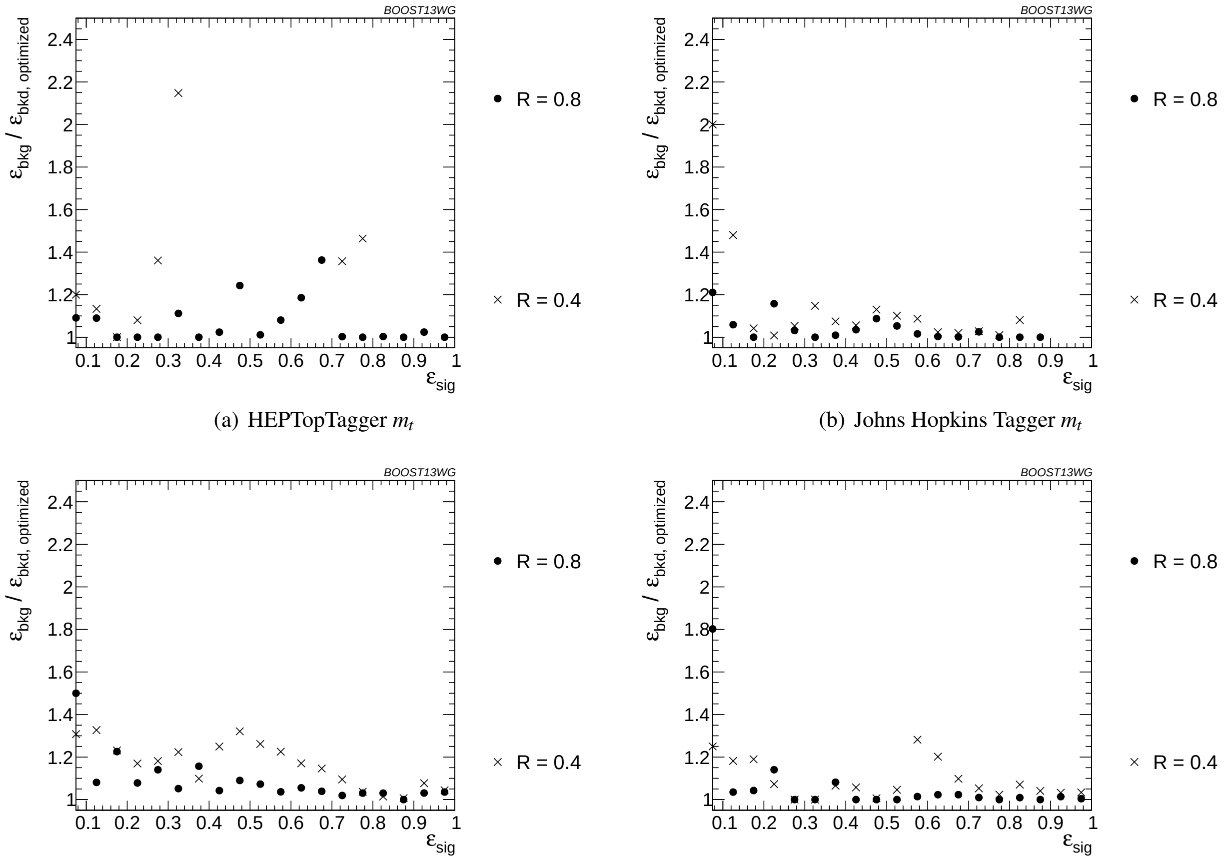


(c) Pruning



(d) Trimming

Fig. 40 Comparison of BDT combination of tagger performance at different p_T using the anti-k_T R=0.8 algorithm; the tagger inputs are set to the optimum value for $p_T = 1.5 - 1.6$ TeV.



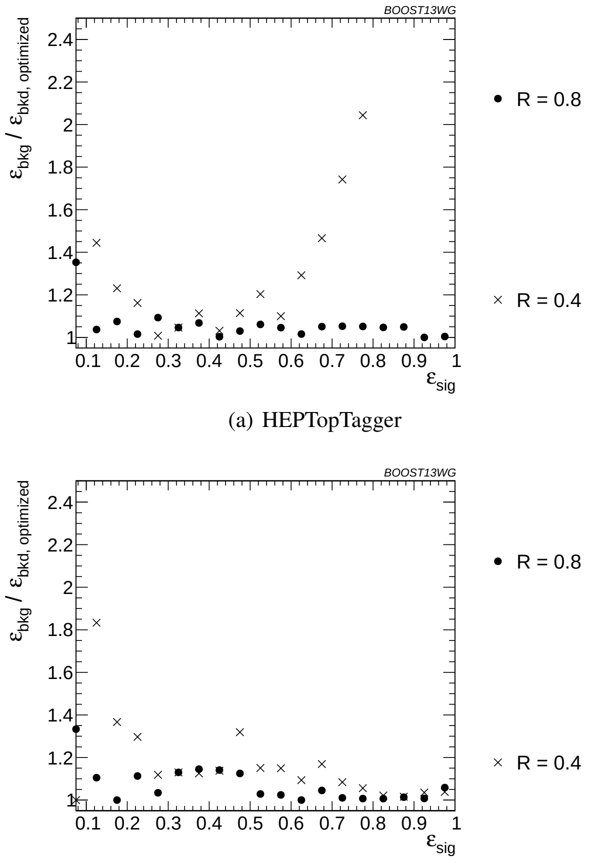
(a) HEPTopTagger m_t

(b) Johns Hopkins Tagger m_t

(c) Pruning m_t

(d) Trimming m_t

Fig. 41 Comparison of top mass performance of different taggers at different R in the $p_T = 1500 - 1600$ GeV bin; the tagger inputs are set to the optimum value for $R = 1.2$.



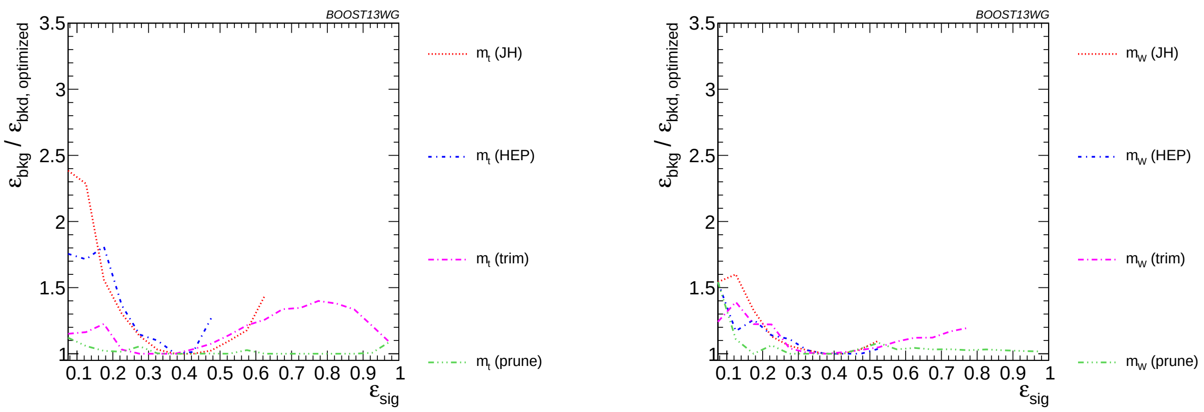
(a) HEPTopTagger

(b) Johns Hopkins Tagger

(c) Pruning

(d) Trimming

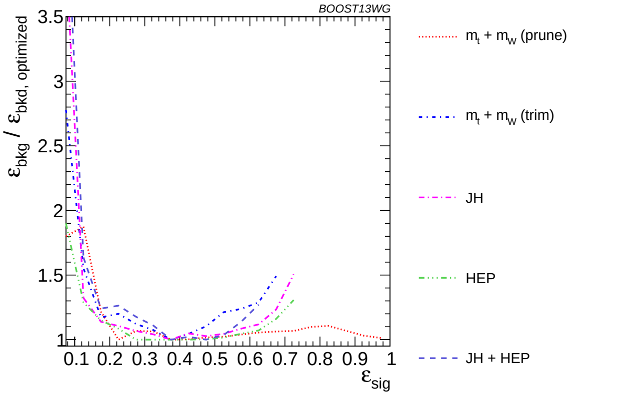
Fig. 42 Comparison of BDT combination of tagger performance at different radius at p_T = 1.5-1.6 TeV; the tagger inputs are set to the optimum value for $R = 1.2$.



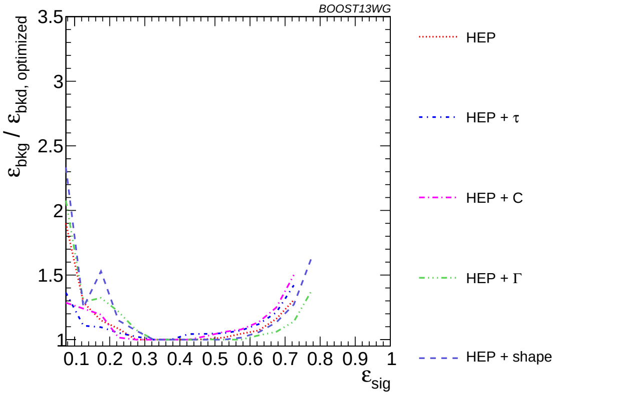
(a) Top mass

(b) W mass

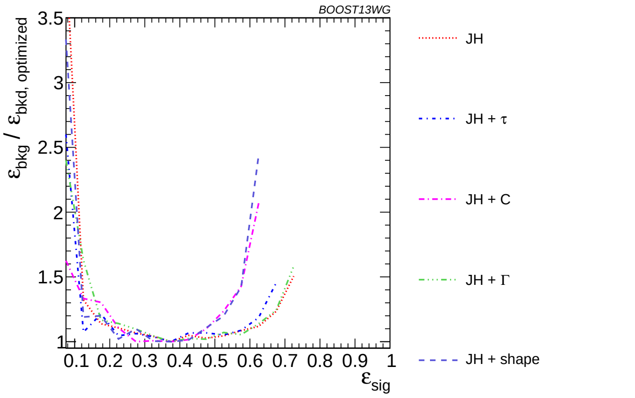
Fig. 43 Comparison of single-variable top-tagging performance in the $p_T = 1 - 1.1$ GeV bin using the anti-k_{T}, R=0.8 algorithm; the inputs for each tagger are optimized for the $\varepsilon_{\mathrm{sig}} = 0.3 - 0.35$ bin.



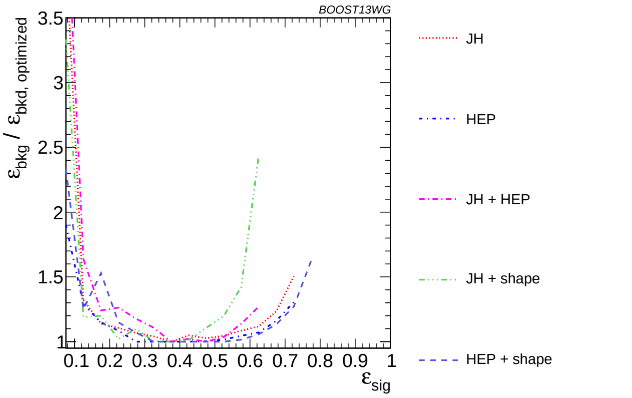
(a) Tagger-Groomer comparison



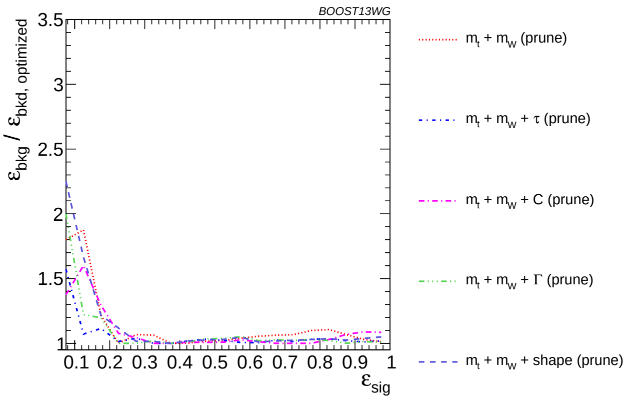
(b) HEPTopTagger + Shape



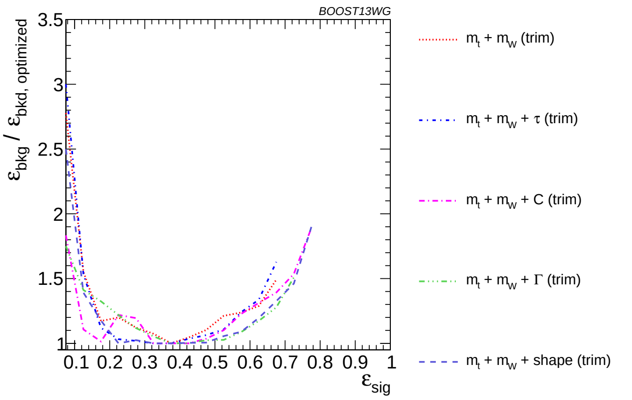
(c) Johns Hopkins Tagger + shape



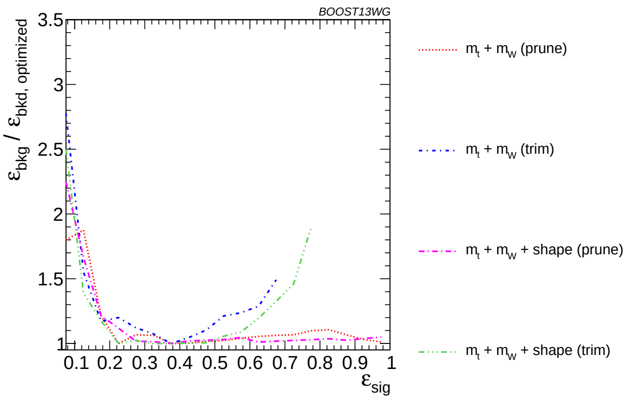
(d) HEP vs. JH comparison (incl. shape)



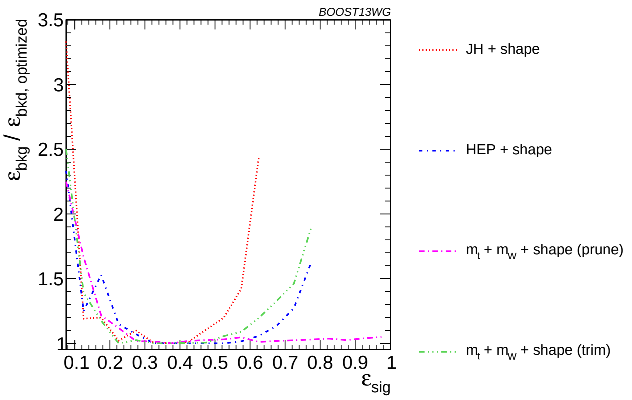
(e) Pruning + Shape



(f) Trimming + Shape



(g) Trim vs. Prune comparison (incl. shape)



(h) Comparison of all Tagger+Shape

Fig. 44 The BDT combinations in the $p_T = 1-1.1$ TeV bin using the anti-k_T R=0.8 algorithm. Taggers are combined with the following shape observables: $\tau_{21}^{(\beta=1)} + \tau_{32}^{(\beta=1)}$, $C_2^{(\beta=1)} + C_3^{(\beta=1)}$, Γ_{Qjet}, and all of the above (denoted “shape”). The inputs for each tagger are optimized for the $\varepsilon_{\mathrm{sig}} = 0.3-0.35$ bin.