Towards an Understanding of the Correlations in Jet Substructure

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

D. Adams[1], A. Arce[2], L. Asquith[3], M. Backovic[4], T. Barillari[5], P. Berta[6],
D. Bertolini[2], A. Buckley[8], J. Butterworth[9], R. C. Camacho Toro[10], J. Caudron[9],
Y.-T. Chien[11], J. Cogan[12], B. Cooper[9], D. Curtin[17], C. Debenedetti[18], J. Dolen[9],
M. Eklund[22], S. El Hedri[22], S. D. Ellis[22], T. Embry[22], D. Ferencek[23], J. Ferrando[24],
S. Fleischmann[16], M. Freytsis[25], M. Giulini[21], Z. Han[27], D. Hare[4], P. Harris[4],
A. Hinzmann[4], R. Hoing[4], A. Hornig[22], M. Jankowiak[4], K. Johns[28], G. Kasieczka[23],
T. Knight[24], G. Kasieczka[29], R. Kogler[30], W. Lampl[4], A. J. Larkoski[4],
C. Lee[31], R. Leone[31], P. Loch[31], D. Lopez Mateos[27], H. K. Lou[27], M. Low[27],
P. Maksimovic[32], I. Marchesini[32], S. Marzani[32], L. Masetti[33], R. McCarthy[32],
S. Menke[32], D. W. Miller[35], K. Mishra[36], B. Nachman[32], P. Nef[4], F. T. O'Grady[24],
A. Ovcharova[23], A. Picazio[37], C. Pollard[38], B. Potter Landua[29], C. Potter[29],
S. Rappoccio[39], J. Rojo[48], J. Rutherfoord[40], G. P. Salam[10,11], J. Schabinger[23],
A. Schwartzman[4], M. D. Schwartz[27], B. Shuve[43], P. Sinervo[44], D. Soper[45],
D. E. Sosa Corral[45], M. Spannowsky[32], E. Strauss[34], M. Swiatlowski[4], J. Thaler[34],
C. Thomas[34], E. Thompson[1], N. V. Tran[36], J. Tseng[36], E. Usai[36], L. Valery[36],
J. Veatch[23], M. Vos[23], W. Waalewijn[4], and C. Young[47]

[1] *Columbia University, Nevis Laboratory, Irvington, NY 10533, USA*
[2] *Duke University, Durham, NC 27708, USA*
[3] *Argonne National Laboratory, Lemont, IL 60439, USA*
[4] *SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[5] *Deutsches Elektronen-Synchrotron, DESY, D-15738 Zeuthen, Germany*
[6] *Cornell University, Ithaca, NY 14853, USA*
[7] *Lund University, Lund, SE 22100, Sweden*
[8] *University of Edinburgh, EH9 3JZ, UK*
[9] *University College London, WC1E 6BT, UK*
[10] *LPTHE, UPMC Univ. Paris 6 and CNRS UMR 7589, Paris, France*
[11] *CERN, CH-1211 Geneva 23, Switzerland*
[12] *CAFPE and U. of Granada, Granada, E-18071, Spain*
[13] *McGill University, Montreal, Quebec H3A 2T8, Canada*
[14] *Iowa State University, Ames, Iowa 50011, USA*
[15] *Rutgers University, Piscataway, NJ 08854, USA*
[16] *Bergische Universitaet Wuppertal, Wuppertal, D-42097, Germany*
[17] *YITP, Stony Brook University, Stony Brook, NY 11794-3840, USA*
[18] *University of Manchester, Manchester, M13 9PL, UK*
[19] *UNESP - Universidade Estadual Paulista, Sao Paulo, 01140-070, Brazil*
[20] *INFN and University of Naples, IT80216, Italy*
[21] *University of Geneva, CH-1211 Geneva 4, Switzerland*
[22] *University of Washington, Seattle, WA 98195, USA*
[23] *Instituto de Física Corpuscular, IFIC/CSIC-UVEG, E-46071 Valencia, Spain*
[24] *University of Glasgow, Glasgow, G12 8QQ, UK*
[25] *Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA*
[26] *Universidad de Buenos Aires, AR-1428, Argentina*
[27] *Harvard University, Cambridge, MA 02138, USA*
[28] *Weizmann Institute, 76100 Rehovot, Israel*
[29] *Universitaet Hamburg, DE-22761, Germany*
[30] *Universitaet Heidelberg, DE-69117, Germany*
[31] *University of Arizona, Tucson, AZ 85719, USA*
[32] *IPPP, University of Durham, Durham, DH1 3LE, UK*
[33] *Universitaet Mainz, DE 55099, Germany*
[34] *MIT, Cambridge, MA 02139, USA*
[35] *University of Chicago, IL 60637, USA*
[36] *Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*
[37] *Indiana University, Bloomington, IN 47405, USA*
[38] *University of California, Davis, CA 95616, USA*
[39] *Johns Hopkins University, Baltimore, MD 21218, USA*
[40] *INFN and University of Pisa, Pisa, IT-56127, Italy*
[41] *Texas A & M University, College Station, TX 77843, USA*
[42] *INFN and University of Calabria, Rende, IT-87036, Italy*
[43] *Brown University, Richmond, RI 02912, USA*
[44] *Yale University, New Haven, CT 06511, USA*
[45] *CEA Saclay, Gif-sur-Yvette, FR-91191, France*
[46] *University of Illinois, Chicago, IL 60607, USA*
[47] *University of California, Berkeley, CA 94720, USA*

Abstract Abstract for BOOST2013 report

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, 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 [1,2, 3].

The following report from BOOST 2013 thus views the study and implementation of jet substructure techniques as a fairly mature field. The report attempts to focus 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. The report is organized as follows: NEED TO GENERATE AN OUTLINE OF THE REPORT - ESPECIALLY AS I UNDERSTAND IT MYSELF.

2 Monte Carlo Samples

2.1 Quark/gluon and W tagging

Samples were generated at $\sqrt{s} = 8$ TeV for QCD dijets and W^+W^- pairs decaying hadronically off a (pseudo) scalar resonance. The QCD events were split into sub-samples of gg and $q\bar{q}$ events, allowing for tests of both W and quark-gluon discrimination.

Individual quark and gluon 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 produced in exclusive p_T bins of 100 GeV and generated using CTEQ6L1 PDFs. The slicing parameter was chosen to be the p_T of any final state parton or W. 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. These were then showered through PYTHIA8 (version 8.176) using the default tune 4C.

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 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 bin. The bins in leading jet p_T that are investigated in the W-tagging and q/g tagging studies are 300-450 GeV, 500-650 GeV, 1.0-1.2 TeV.

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, with matrix elements with 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 with jet radius $R = 1.2$. The matching scale is selected to be $Q_{\text{cut}} = 40, 60, 80$ GeV for the $p_{T\,\text{min}} = 600, 1000$, and 1500 GeV bins, respectively.

The analysis again relies on FASTJET 3.0.3 for jet clustering and calculation of jet substructure observables, with the same cuts applied after showering and clustering as for $\sqrt{s} = 8$ TeV data.

3 Jet Algorithms and Substructure Observables

3.1 Jet Clustering Algorithms

Jets were clustered using sequential jet clustering algorithms. Final state particles i, j are assigned a mu-

tual distance d_{ij} and a distance to the beam, d_{iB}. The particle pair with smallest d_{ij} are recombined and the algorithm repeated until the smallest distance is actually the distance to the beam, d_{iB}, in which case i is set aside and labelled as a jet. The distance metrics are defined as

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

$$d_{iB} = p_{Ti}^{2\gamma}, \tag{2}$$

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 algorithm ($\gamma = 0$), and the k_T algorithm ($\gamma = 1$), each of which has varying sensitivity to soft radiation in defining the jet. *add citations*

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

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

where $d_{\min}$ is the minimum distance for the usual jet clustering algorithm at a particular step. The parameter α is called the stiffness and is used to control how sharply peaked the probability distribution is around the "classical" value. Qjets uses statistical methods to extract more information from the jet than can be found in the usual cluster sequence. In our analyses, we use $\alpha = 0.1$.

3.2 Jet Substructure Observables

In this section, we describe the observables that we consider in this study. Originally we considered a larger set of observables but in the final analysis we reduced redundant observables in the final set for presentation purposes.

Quark/gluon discrimination: The list of observables is as follows:

– mass: this is the plain jet mass
– 1-subjettiness, τ_1^β: the N-subjettiness uses one-pass k_T axis optimization where we consider $\beta = 1, 2$
– 1-point energy correlation functions, C_1^β: the energy correlation functions consider $\beta = 0, 1, 2$
– Qjet volatility, Γ_{Qjet}: the number of trees considered is $N_{\mathrm{trees}} = 25$, the rigidity factor is $\alpha = 0.1$, the truncation factor is 0.01, and the pruning parameters are $D_{\mathrm{cut}} = 0.5$ and $z_{\mathrm{cut}} = 0.1$
– number of constituents (N_{constits})

W vs. gluon discrimination: The list of observables is as follows:

– mass: same as in the q vs. g case
– trimmed mass, m_{trimmed}: the parameter values are $f_{\mathrm{cut}} = 0.03$ and $r_{\mathrm{filt}} = 0.2$
– pruned mass, m_{pruned}: the parameter values are $D_{\mathrm{cut}} = 0.5$ and $z_{\mathrm{cut}} = 0.1$
– soft drop mass, $m_{\mathrm{softdrop}}^\beta$: z_{cut} is set always to 0.1, we consider $\beta = 0, 2$ where $\beta = 0$ is a generalization of the modified mass drop tagger
– 2-point energy correlation functions, $C_2^{\beta=1}$: we also considered $\beta = 2$ but it showed poor discrimination power
– N-subjettiness ratio, $\tau_2/\tau_1(\beta = 2)$: the N-subjettiness uses one-pass k_T axis optimization, we also considered $\beta = 2$ but it showed poor discrimination power
– Qjet volatility: same as in the q vs. g case

Top vs. QCD discrimination: We now describe the list of observables/taggers considered for top tagging. Note that for trimming, the subjet identification is optimized for identifying soft radiation, *not* for reconstructing the hard decay products of the top. Pruning does not even contain an inherent subjet identification step. For both trimming and pruning, we introduce an arbitrary method for reconstructing the subjets corresponding to the b and W decay products for a fair comparison with other top taggers, but the W reconstruction is consequently poorer than for algorithms that are optimized for W identification inside the top.

Johns Hopkins Tagger: Re-cluster the jet using the Cambridge/Aachen algorithm. The jet is iteratively declustered, and at each step the softer prong is discarded if its p_T is less than $\delta_p p_{T\,\mathrm{jet}}$. This continues until both prongs are harder than the p_T threshold, both prongs are softer than the p_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_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 θ_{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.

HEPTopTagger: Re-cluster the jet using the Cambridge/Aachen algorithm. The jet is iteratively declustered, 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 $R_{\text{filt}} = \min(0.3, \Delta R_{ij})$ (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 θ_{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.

Trimming: Re-cluster the jet using the k_{T} algorithm and radius R_{trim}. Discard all subjets with $p_{\text{T sj}}/p_{\text{T jet}} < f_{\text{cut}}$. A W candidate is reconstructed as follows: if there are two subjets, the highest-mass subjet is the W candidate; 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.

Pruning: Re-cluster the jet using the Cambridge-Aachen algorithm. At each step, discard the softer branch if $\min(p_{\text{T1}}, p_{\text{T2}})/p_{\text{T12}} < z_{\text{cut}}$ and $\Delta R_{12} > 2R_{\text{cut}} m_{\text{jet}}/p_{T,\text{jet}}$. Subjets are found by de-clustering the pruned jet by up to three splittings. A W candidate is reconstructed as follows: if there are two subjets, the highest-mass subjet is the W candidate; 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.

4 Multivariate Analysis Techniques

Put in description of BDT methods here (Nhan?)

5 Quark-Gluon Discrimination

In this section we examine the differences between quark and gluon initiated jets in terms of the substructure variables, and to what extent these variables are correlated. Along the way, we attempt to provide some theoretical understanding of these observations. The motivation for these studies comes not only from the desire to "tag" a jet as being quark or gluon initiated, but also from the point of view of understanding the quark and gluon components of the QCD background to boosted boson and boosted top tagging.

5.1 Methodology

These studies use the qq and gg samples, described previously in Section 2. Jets are reconstructed using the anti-k_{T} algorithm with radius parameters of 0.4, 0.8 and 1.2, and have various jet grooming approaches applied,

as described in Section **??**. Only leading and subleading jets in each sample are used.

Figure 1 shows a comparison of the quark and gluon samples p_T and η distributions for the sample used to study jets of $p_T = 500 - 600$ GeV. The differences in the p_T distributions can be attributed to different out-of-cone radiation patterns for quark and gluons, while the different η distributions are related to the different parton density functions involved in qq and gg production. The qualitative features of the η distributions do not change as the R parameter is changed. As the p_T increases, the η distributions peak more strongly near zero, as expected. Differences in the p_T distributions between the leading and sub-leading (and quark and gluon-induced) jets become smaller as the R parameter is increased, as expected from the physics behind these differences, outlined above.

5.2 Single Variable Discrimination

Figure 2 shows the mass of jets in the quark and gluon samples when using different groomers, and Figure 3 shows similar comparisons for different substructure variables. Jets built with the anti-k_{T} algorithm with R=0.8 and with $p_T = 500 - 650$ GeV are used. Qualitatively, the application of grooming shifts the mass distributions towards lower values 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. Few variations are observed as the radius parameter of the jet reconstruction is increased in the two highest p_T bins. 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.

Among the different substructure variables explored, n_{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 [**add citations**]. The evolution of some of these distributions with p_T and R is a bit more interesting than what was discussed for the jet mass. In particular, changing the R parameter at high p_T changes significantly the C_a^{β} for $\beta > 0$ and the n_{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 $p_T = 1 - 1.2$ 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
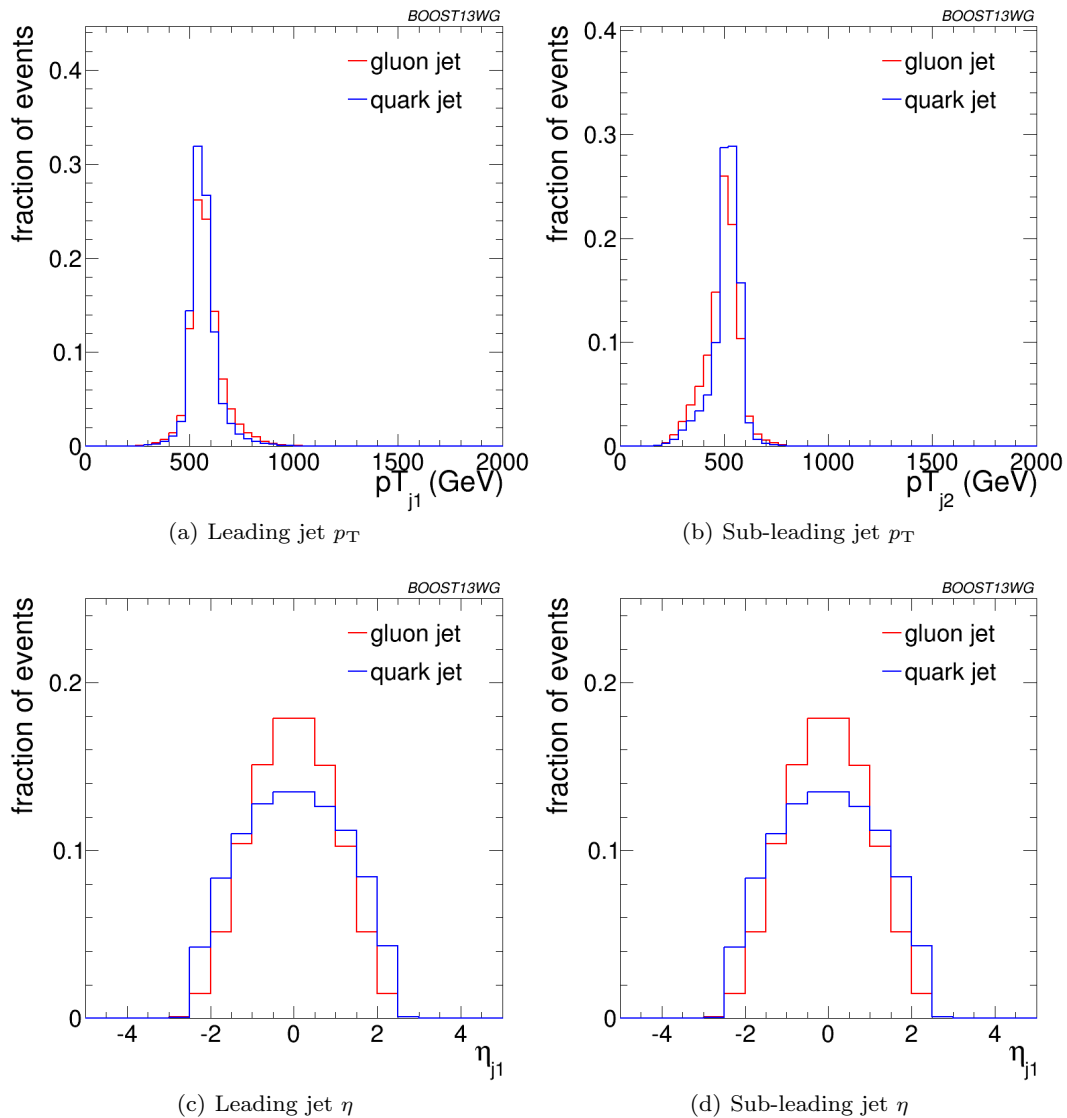
(a) Leading jet p_T

(b) Sub-leading jet p_T

(c) Leading jet η

(d) Sub-leading jet η

Fig. 1 Comparisons of quark and gluon p_T and η distributions in the sample used for the jets of $p_T = 500 - 600$ GeV bin using the anti-k_T R=0.8 algorithm.

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 Sudakov form factors are larger for gluon jets, the low-volatility peak is higher for gluon jets by about the color factor C_A/C_F.

In order to be more quantitative about these results and the power of each variable as a discriminator for quark/gluon tagging, ROC curves are built by scanning each distribution and plotting the background efficiency (efficiency to select gluon jets) vs the efficiency for quark selection. Figure 5 shows these ROC curves for all of the variables shown in Figure 3 and the ungroomed mass, representing the best performing mass variable, for jets of $p_T = 300 - 400$ GeV. In addition, the ROC curve for the tagger built from a BDT com-
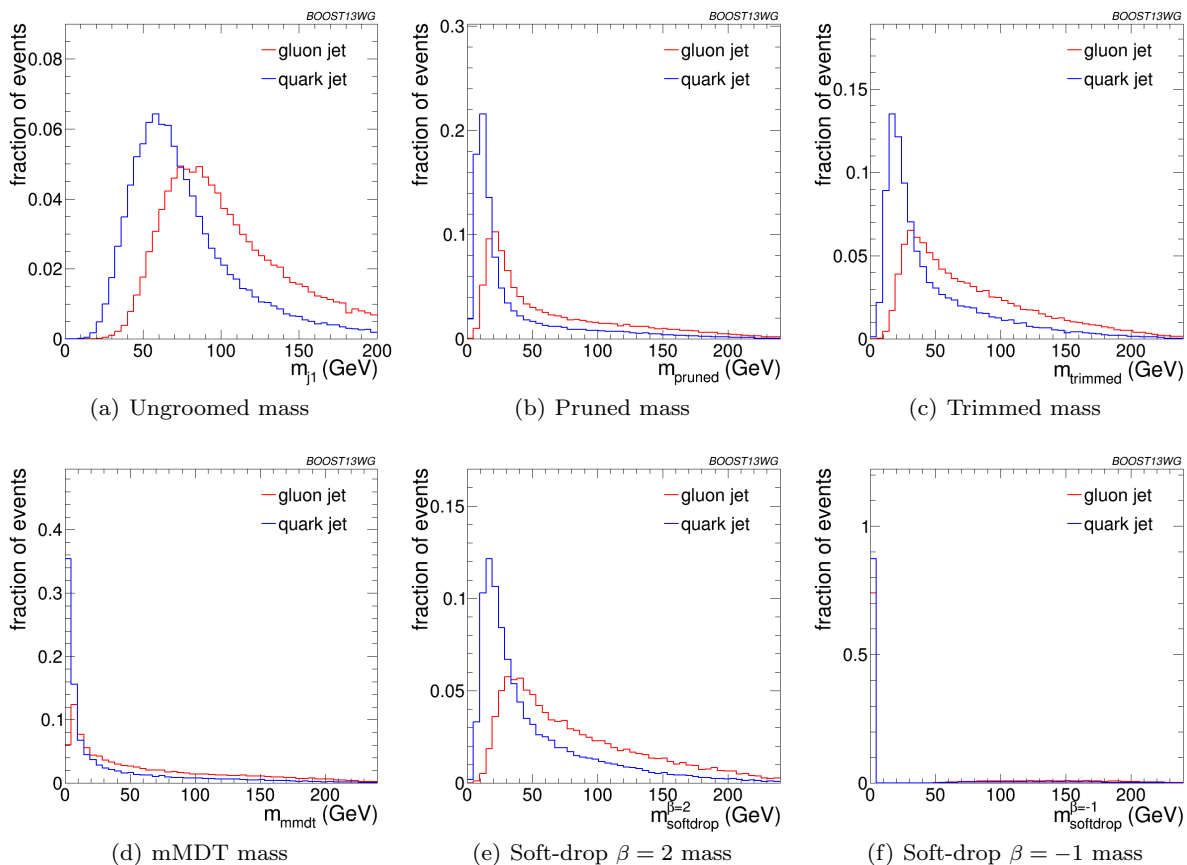
Fig. 2 Comparisons of ungroomed and groomed quark and gluon mass distributions for leading jets in the $p_T = 500 - 650$ GeV bin using the anti-k_{T} R=0.8 algorithm.

bining all the variables. The details of how the BDT is constructed are explained in Section 5.3.

Clearly, n_{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 the Q-jet volatility, which shows significantly worse discrimination. The combination of all variables shows somewhat better discrimination. The overall discriminating power decreases with increasing R, and the features discussed for this p_T bin also apply to the higher p_T bins. This statement is quantified in the next section.

5.3 Correlations and Combined Performance

The combined performance displayed in Fig. 5 is not much better than that of single variables. However, that improvement in performance can be critical for certain analyses using a quark/gluon tagger, and potentially larger in data than in Monte Carlo simulation. Furthermore, insight can be gained into the features allowing for quark/gluon discrimination if how that improvement arises is understood. For that reason,

it is worth investigating quantitatively the improvements in performance. To that purpose, quark/gluon taggers are built from every pair-wise combination of variables studied in the previous section, as well as the full set of variables using a boosted decision tree.

[Paragraph describing details of BDT]

In order to quantitatively study the value of each variable for quark/gluon tagging, the gluon rejection, defined as $1/\epsilon_{\mathrm{gluon}}$, is studied at a fixed quark selection efficiency of 50%. Figure 6 shows that rejection for each individual variable (along the diagonal of the plots) and for each pair-wise combination. The rejection for the BDT combining all variables is also shown on the bottom right of each plot. Results are shown for jets with $p_T = 1 - 1.2$ TeV and for different R parameters. As already observed in the previous section, n_{constits} is the most powerful single variable and $C_1^{(\beta=0)}$ follows closely. The combination of the two variables is also one of the most powerful combinations for the two large-R collections. However, those collections are clearly outperformed by the small-R collection, and in that case other pair-wise combinations are more pow-
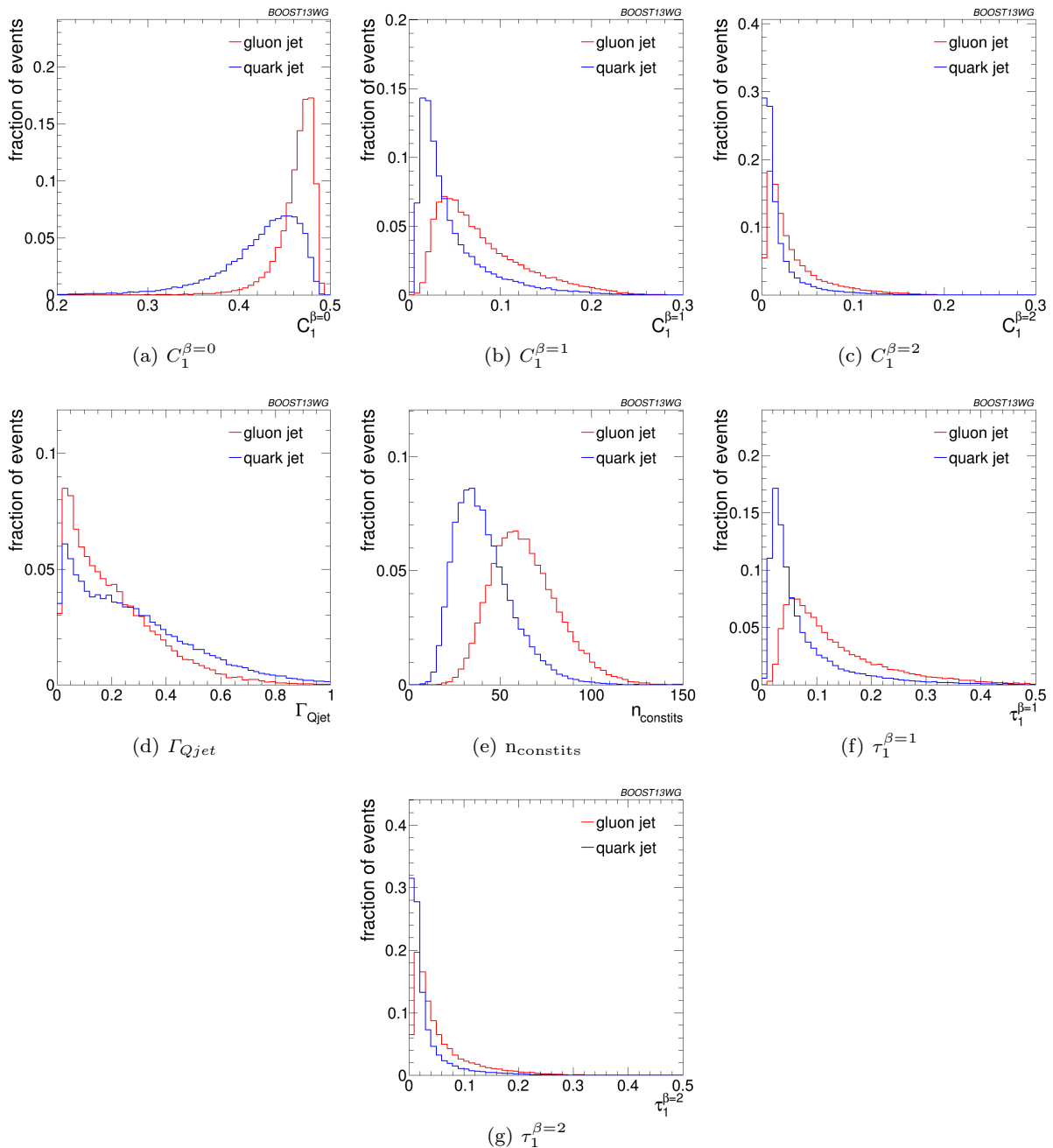
(a) $C_1^{\beta=0}$

(b) $C_1^{\beta=1}$

(c) $C_1^{\beta=2}$

(d) Γ_{Qjet}

(e) $n_{constits}$

(f) $\tau_1^{\beta=1}$

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

Fig. 3 Comparisons of quark and gluon distributions of different substructure variables for leading jets in the $p_T = 500 - 650$ GeV bin using the anti-k_T R=0.8 algorithm.

erful. 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.

The overall loss in performance with increasing R can be observed in all single variables studied, except for $C_1^{(\beta=0)}$ and the Q-jet volatility, which are quite resilient to increasing R. This is expected, since their distributions were observed to be also quite insensitive to R in the previous section. Their combination, however,

does lose performance significantly as R is increased. [**do we understand this?**] Of all the variables studied, $\beta = 2$ subjetiness and energy correlation variables are particularly sensitive to increasing R. This is understandable, because for $\beta = 2$ a larger weight is put in large-angle emissions. However, from other variables, it is understood that most of the discrimination power comes from analyzing a small-R jet, or the center of the large-R jet.

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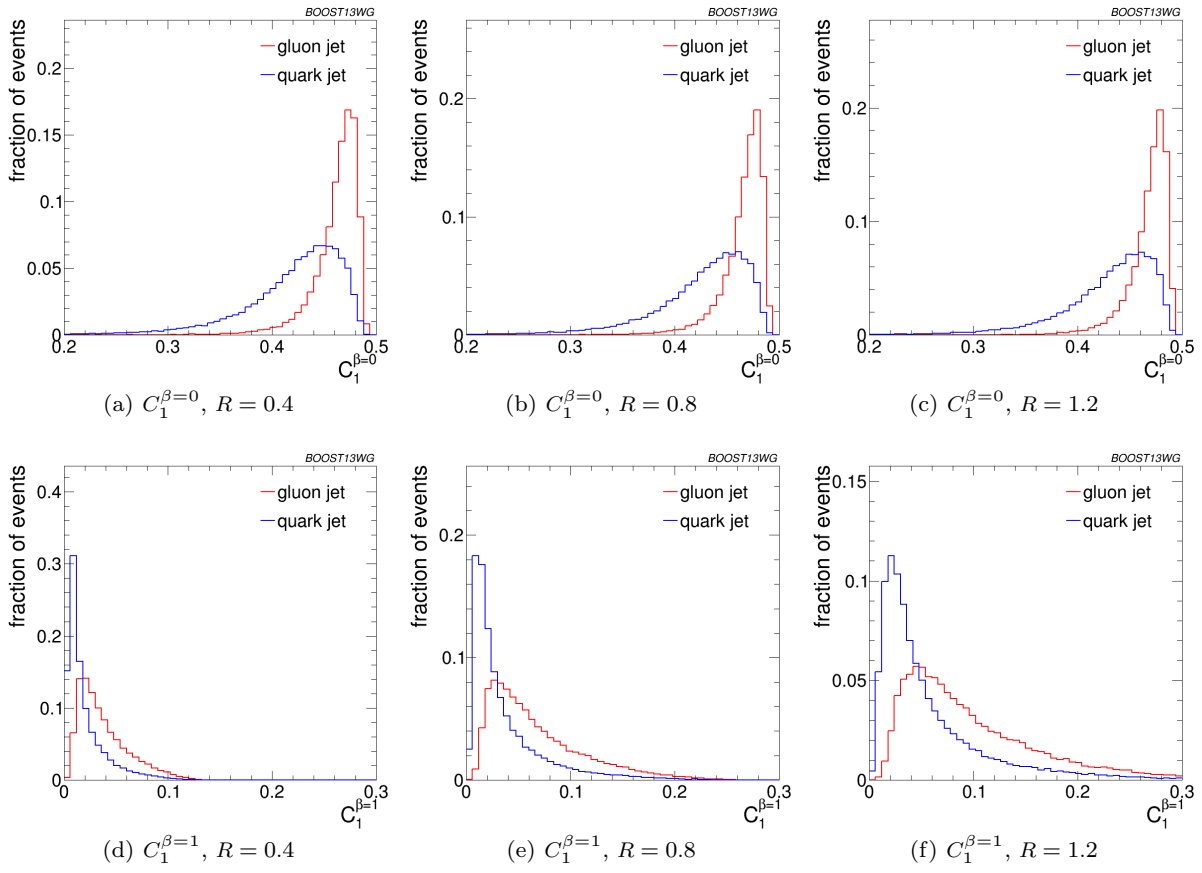
Fig. 4 Comparisons of quark and gluon distributions of $C_1^{\beta=0}$ (top) and $C_1^{\beta=1}$ (bottom) for leading jets in the $p_T = 1-1.2$ TeV bin using the anti-k_T algorithm with R=0.4,0.8 and 1.2.
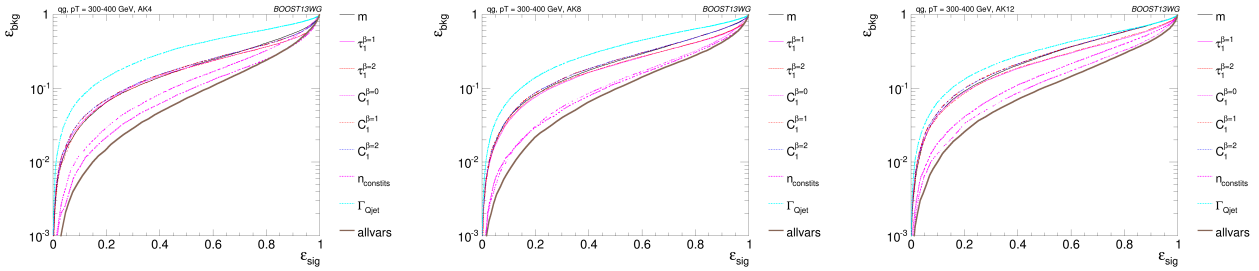

Fig. 5 The ROC curve for all single variables considered for quark-gluon discrimination in the p_T 500 GeV bin using the anti-k_T R=0.8 algorithm.

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. 7 for anti-k_T R=0.4 jets of different p_Ts. Clearly, most single variables retain their gluon rejection potential at lower p_Ts. 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?**]
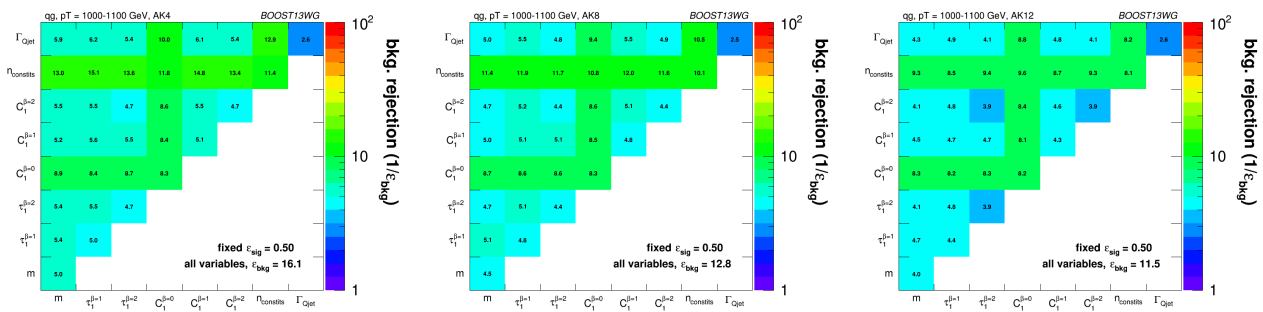
Fig. 6 Gluon rejection defined as $1/\epsilon_{\text{gluon}}$ when using each 2-variable combination as a tagger with 50% acceptance for quark jets. Results are shown for jets with $p_T = 1 - 1.2$ TeV and for different R parameters. The rejection obtained with a tagger that uses all variables is also shown in the plots.
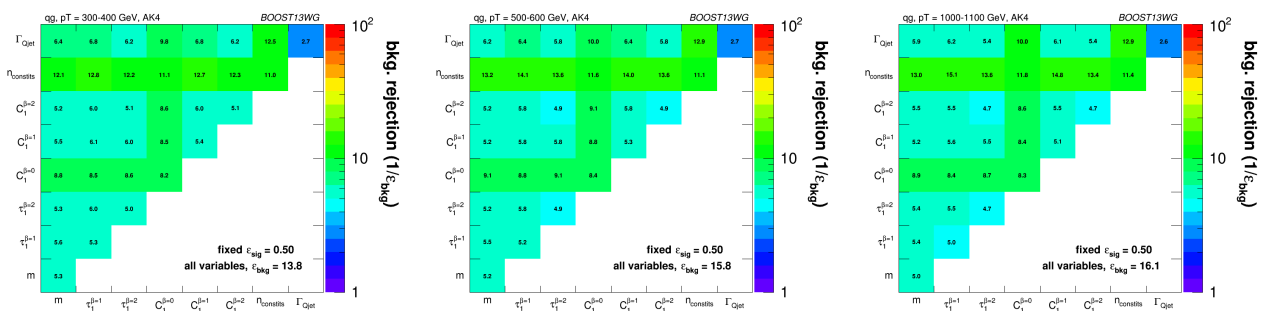

Fig. 7 Gluon rejection defined as $1/\epsilon_{\text{gluon}}$ when using each 2-variable combination as a tagger with 50% acceptance for quark jets. Results are shown for R=0.4 jets with $p_T = 300 - 400$ GeV, $p_T = 500 - 600$ GeV and $p_T = 1 - 1.2$ TeV. The rejection obtained with a tagger that uses all variables is also shown in the plots.

6 Boosted W-Tagging

In this section we study the performance of various groomed jet masses, substructure variables, and BDT combinations of groomed mass and substructure, in terms of the identification of a boosted hadronically decaying W signal aginst a gluon-gluon background. We produce Receiver Operating Characteristic (ROC) curves that elucidate the performance of the various groomed mass and substructure variables that are capable of providing discrimination between signal and background. A range of different distance parameter settings for the anti-k_T jet algorithm are explored, in a variety of kinematic regimes (lead jet p_T 300-450 GeV, 500-650 GeV, 1.0-1.2 TeV), to explore the performance as a function of jet radius and jet boost, and to see where substructure approaches may break down. The groomed mass and substructure variables are then combined in a Boosted Decision Tree (BDT), and the performance of the resulting BDT discriminant explored through ROC curves to understand the degree to which variables are correlated and exploiting the same information, and how this changes with jet boost and jet radius.

6.1 Methodology

These studies use the $X \rightarrow WW$ samples as signal and the gg samples to model the QCD 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. *To be checked!*

Jets are reconstructed using the anti-k_T algorithm, and have various jet grooming approaches applied, as described in Section **??**. The following event selection is then applied to these samples....(presumably this will vary depending on which kinematic bin is used, as will the actual samples used - maybe summarize in a table).

Figure 8 shows a comparison of the leading jet p_T for the signal and background in the p_T 300-450 GeV bin, for the two different anti-k_T jet algorithm distance parameters explored in this bin (R=0.8 and R=1.2). Figures 9 and 10 show the same for the p_T 500-650 GeV bin and p_T 1.0-1.2 TeV bin respectively, where for the p_T 1.0-1.2 TeV bin the distance parameter R=0.4 is also explored.
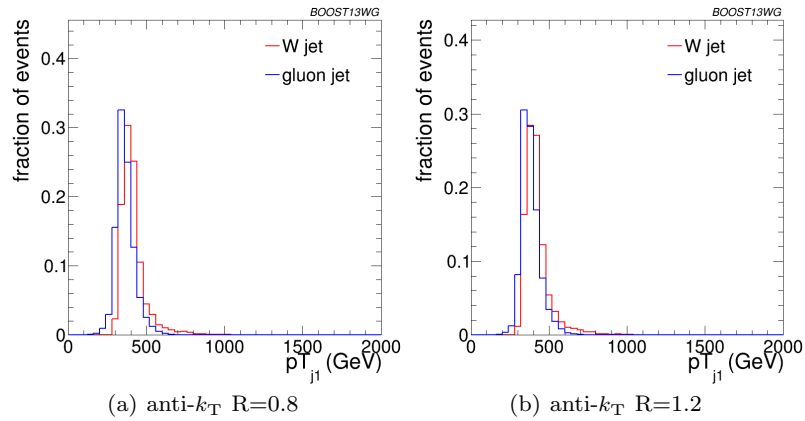
(a) anti-k_T R=0.8

(b) anti-k_T R=1.2

Fig. 8 Comparisons of the leading jet p_T spectrum of the gg background to the WW signal in the p_T 300-450 GeV bin using the different anti-k_T jet distance parameters explored.
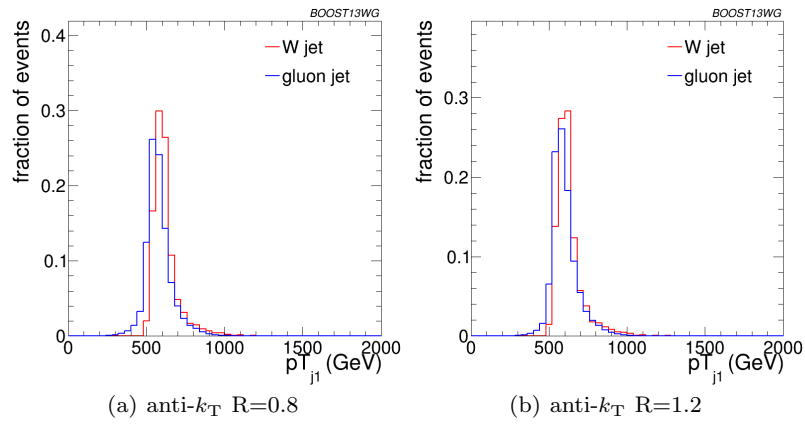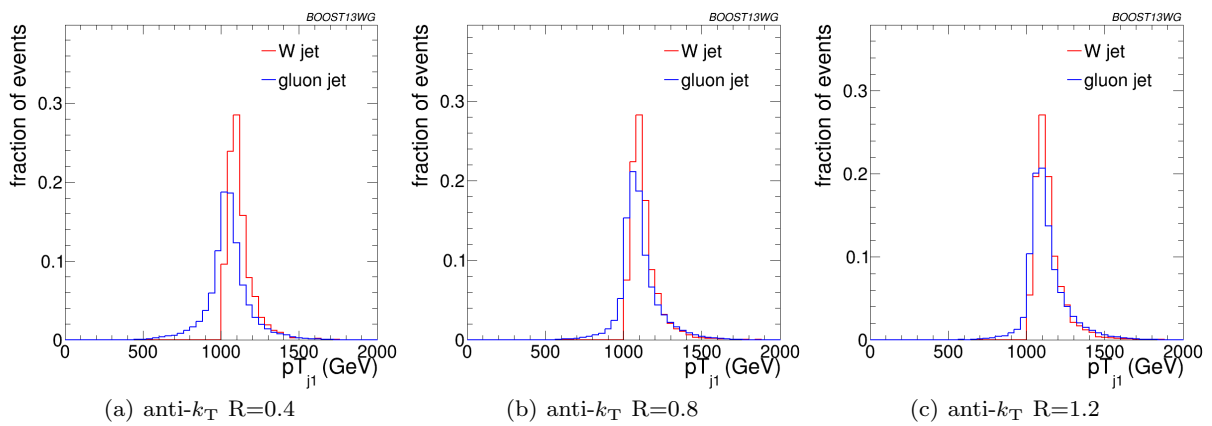

(a) anti-k_T R=0.8

(b) anti-k_T R=1.2

Fig. 9 Comparisons of the leading jet p_T spectrum of the gg background to the WW signal in the p_T 500-650 GeV bin using the different anti-k_T jet distance parameters explored.

(a) anti-k_T R=0.4

(b) anti-k_T R=0.8

(c) anti-k_T R=1.2

Fig. 10 Comparisons of the leading jet p_T spectrum of the gg background to the WW signal in the p_T 1.0-1.2 TeV bin using the different anti-k_T jet distance parameters explored.

Go on to explain how we produce the ROC curves, how the BDT training is done etc.

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 11 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 500-650 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 12 compares signal and background in the different substructure variables explored for the same jet radius and kinematic bin.

Figures 13,14 and 15 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 more performant than any of the individual single variables considered, indicating that there is considerable complementarity between the variables, that 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 16, 17 and 18 are shown matrices which give the background rejection for a signal efficiency of 50% 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. The diagonal of these plots correspond to the background rejections 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 16(a), 17(a) and 18(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 16(b), 17(b) and 18(c) we can see how performance evolves with p_T for R=1.2. The background rejection power of the groomed masses increases slowly with increasing p_T, with at most a factor two increase in rejection in going from the 300-450 GeV to 1.0-1.2 TeV bins. However, for a jet radius of R=0.8, the rejection power of $C_2^{\beta=1}$ increases dramatically with p_T, by a factor of 7 in going from the 300-450 GeV to 1.0-1.2 TeV bins. *Can we explain this?* Conversely, the background rejection of the other substructure variables (Γ_{Qjet} and $\tau_{21}^{\beta=1}$) slowly reduces with increasing p_T, at most decreasing by a factor of two.

By comparing the individual sub-figures of Figures 16, 17 and 18 we can see how the background rejection performance depends on jet radius within the same p_T bin. To within 40%, 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 of R=1.2 (more than an order of magnitude worse background rejection in the 1.0-1.2 TeV bin), and substantially worse for R=0.4. For the other jet substructure variables considered, their background rejection power also reduces for larger jet radius. *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.*

6.3 Combined Performance

The off-diagonal entries in Figures 16, 17 and 18 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. The background rejection of the most powerful mass + substructure combination comes very close to that achieved in the "all variables" case, indicating that there is little to be gained by making a BDT that is more complex, and that there is little more complementary information available, at least in terms of that which is offered by the variables considered here.
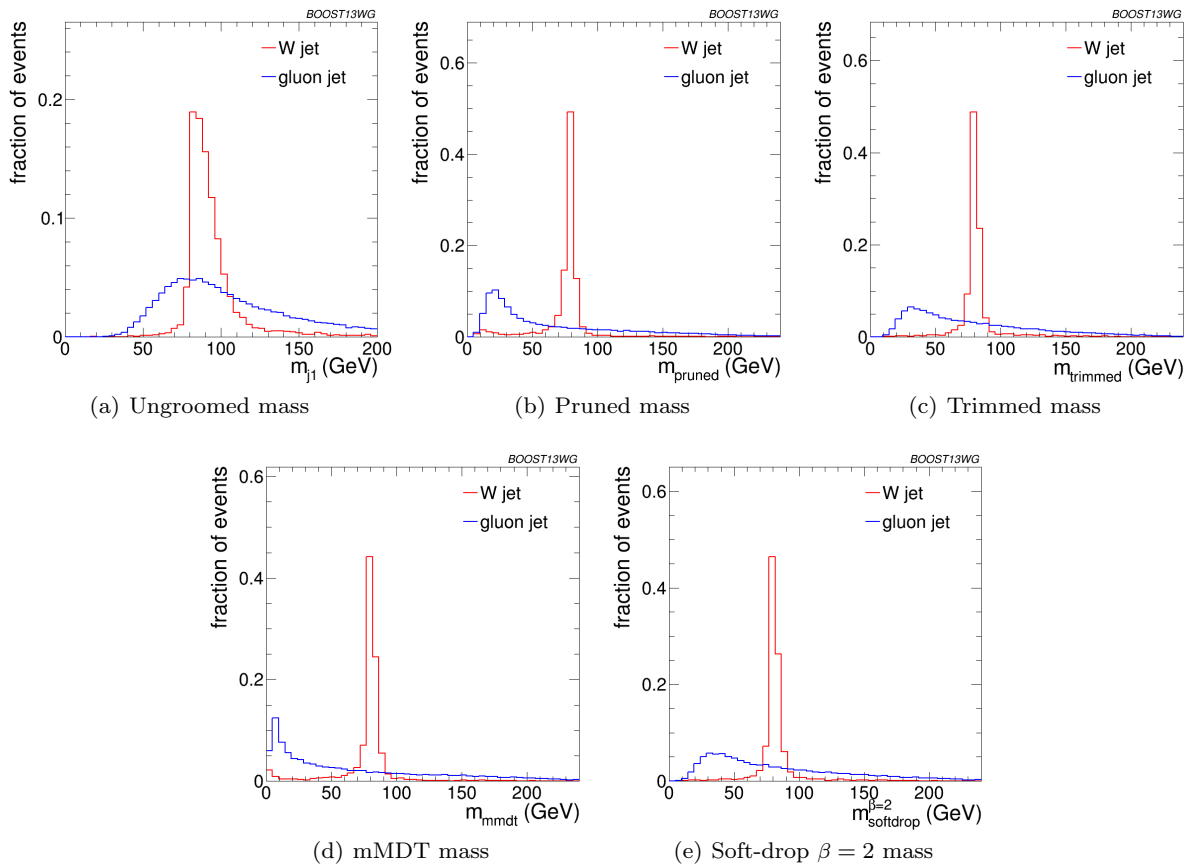
(a) Ungroomed mass

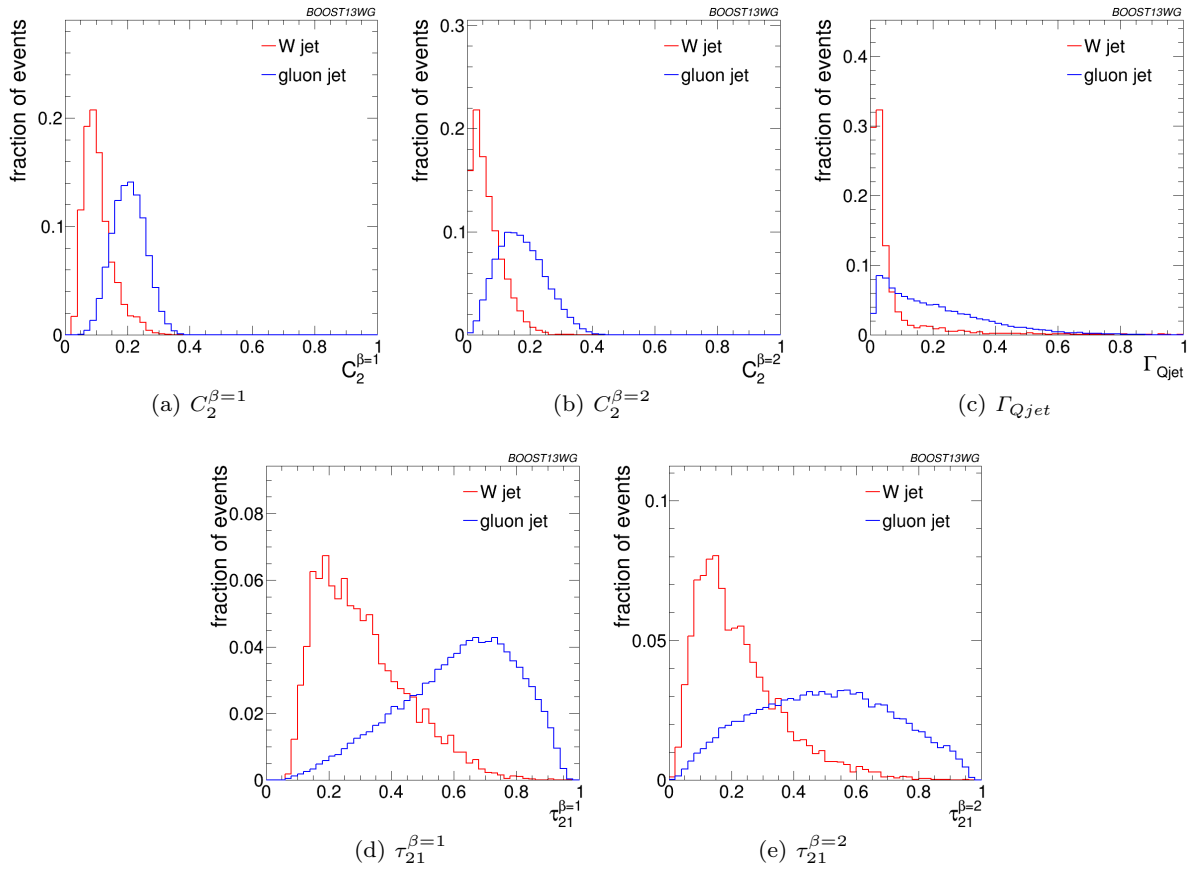(b) Pruned mass

(c) Trimmed mass

(d) mMDT mass

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

**Fig. 11** Comparisons of the QCD background to the WW signal in the p_T 500-650 GeV bin using the anti-k_T R=0.8 algorithm: leading jet mass distributions.

One can also see that there is a 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. There is also 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.

6.3.1 Dependence on p_T

6.3.2 Dependence on R

Fig. 12 Comparisons of the QCD background to the WW signal in the p_T 500-650 GeV bin using the anti-k_T R=0.8 algorithm: substructure variables.
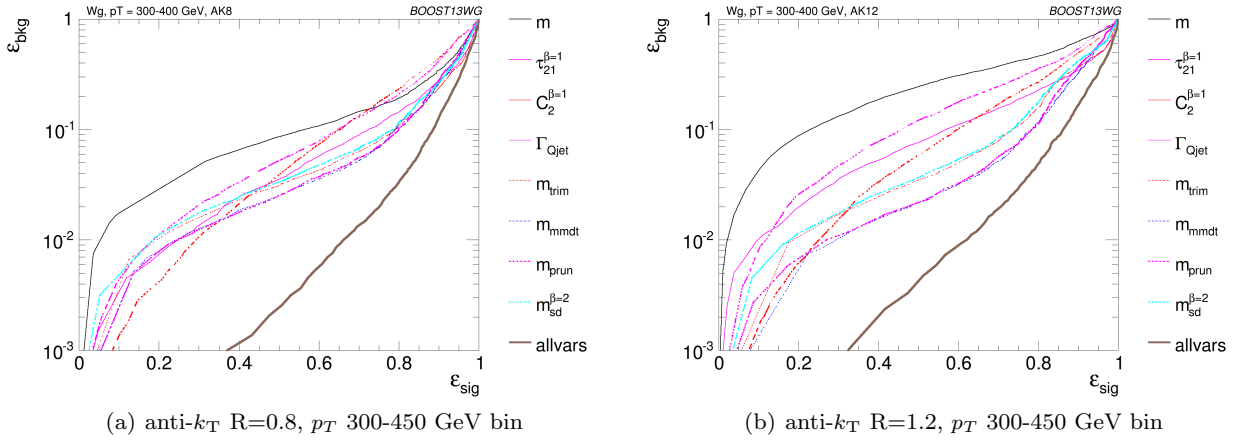
Fig. 13 The ROC curve for all single variables considered for W tagging in the p_T 300-450 GeV bin using the anti-k_T R=0.8 algorithm and R=1.2 algorithm.
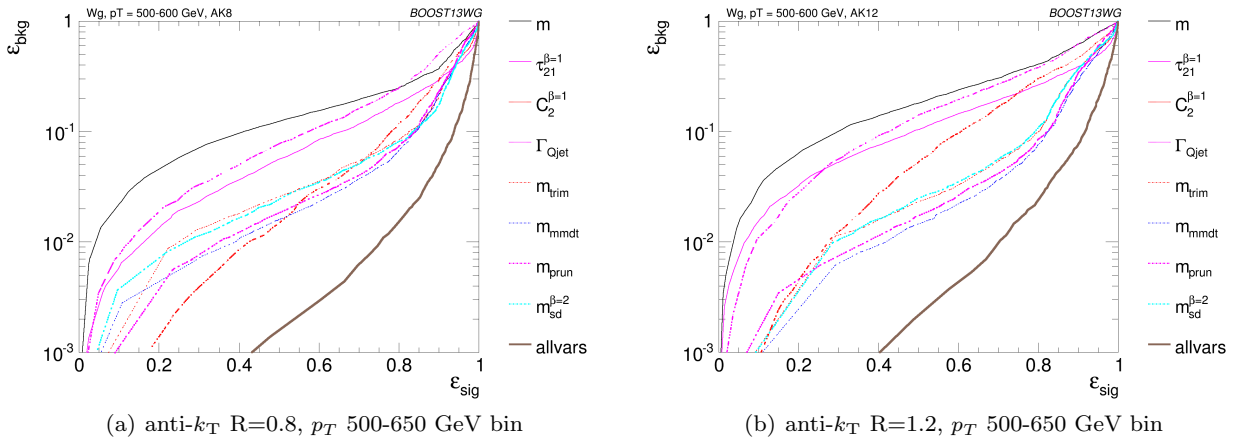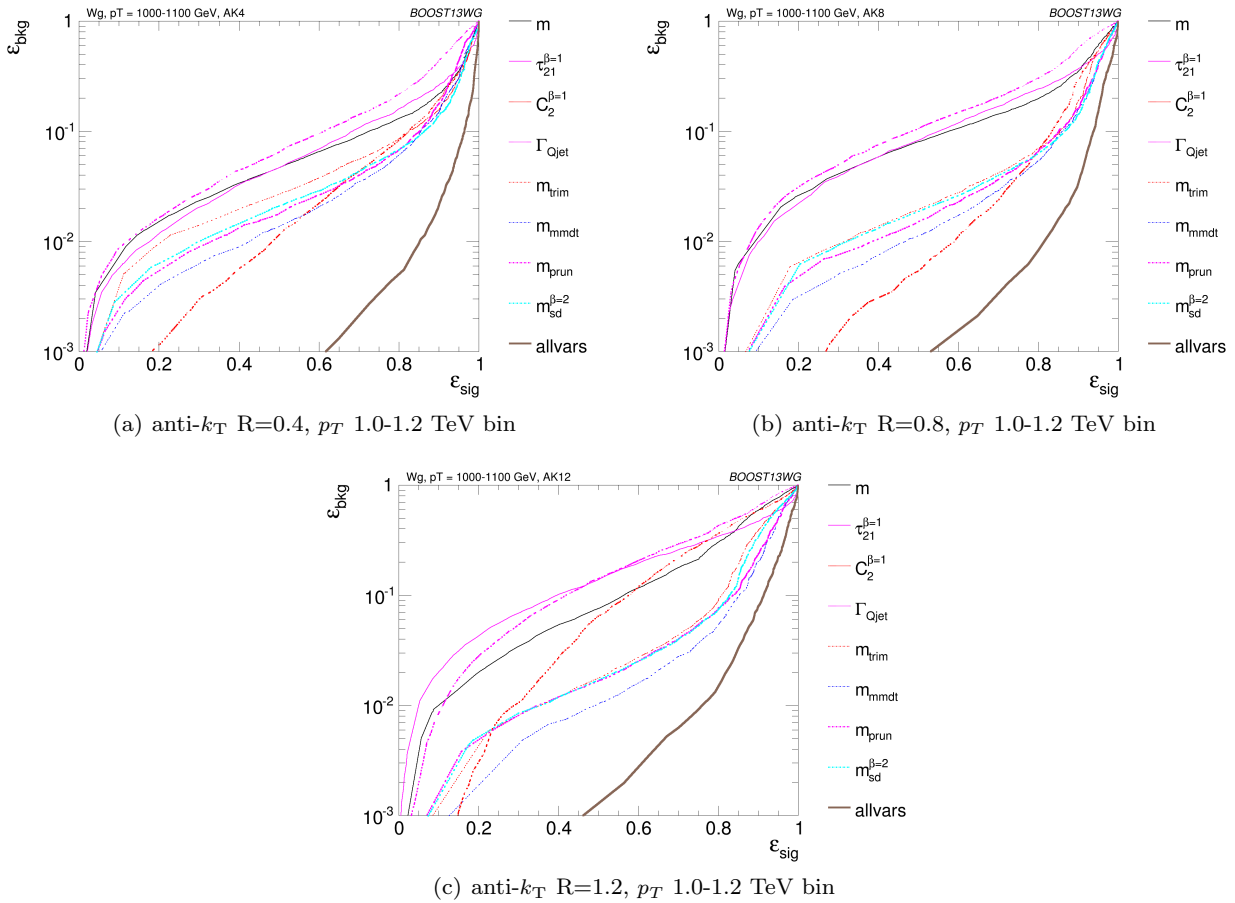
(a) anti-k_T R=0.8, p_T 500-650 GeV bin

(b) anti-k_T R=1.2, p_T 500-650 GeV bin

Fig. 14 The ROC curve for all single variables considered for W tagging in the p_T 500-650 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 1.0-1.2 TeV bin

(b) anti-k_T R=0.8, p_T 1.0-1.2 TeV bin

(c) anti-k_T R=1.2, p_T 1.0-1.2 TeV bin

Fig. 15 The ROC curve for all single variables considered for W tagging in the p_T 1.0-1.2 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 300-450 GeV bin

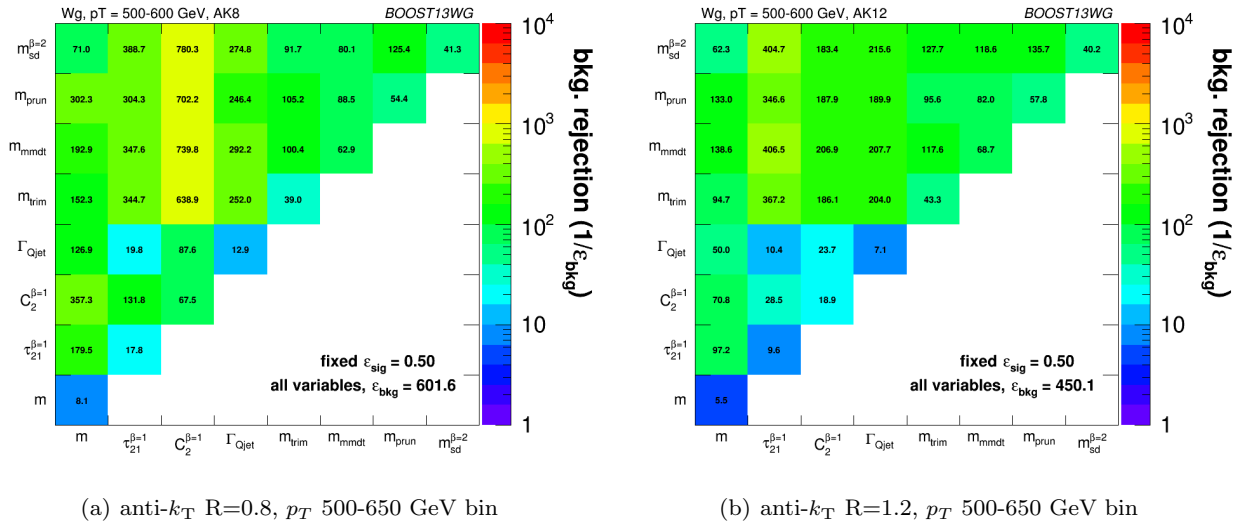(b) anti-k_{T} R=1.2, p_T 300-450 GeV bin
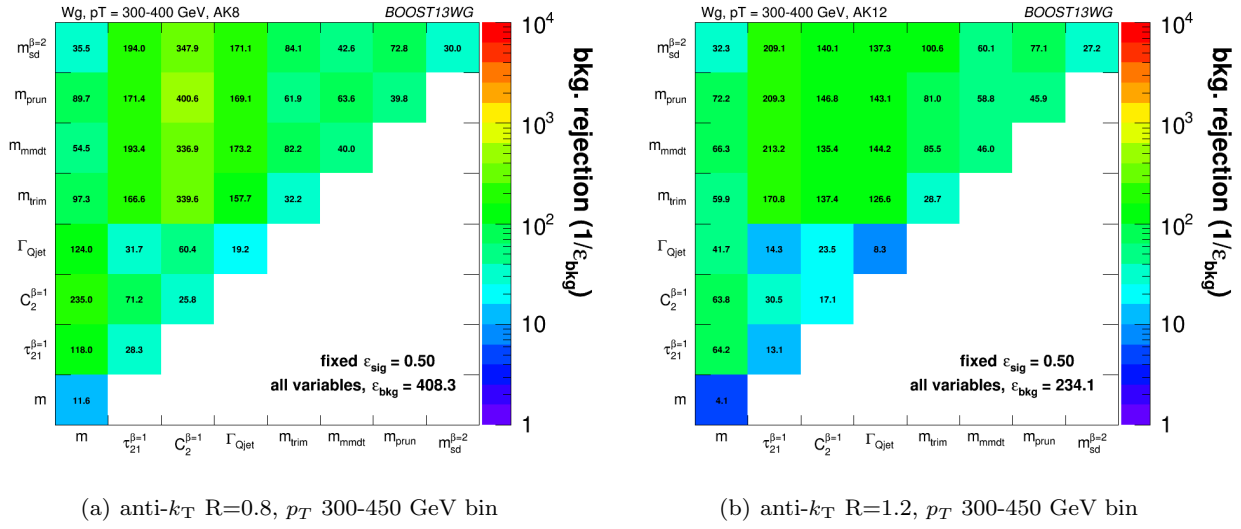
Fig. 16 The background rejection for a fixed signal efficiency (50%) of each BDT combination of each pair of variables considered, in the p_T 300-450 GeV bin using the anti-k_{T} R=0.8 algorithm and R=1.2 algorithm. Also shown is the background rejection for a BDT combination of all of the variables considered.
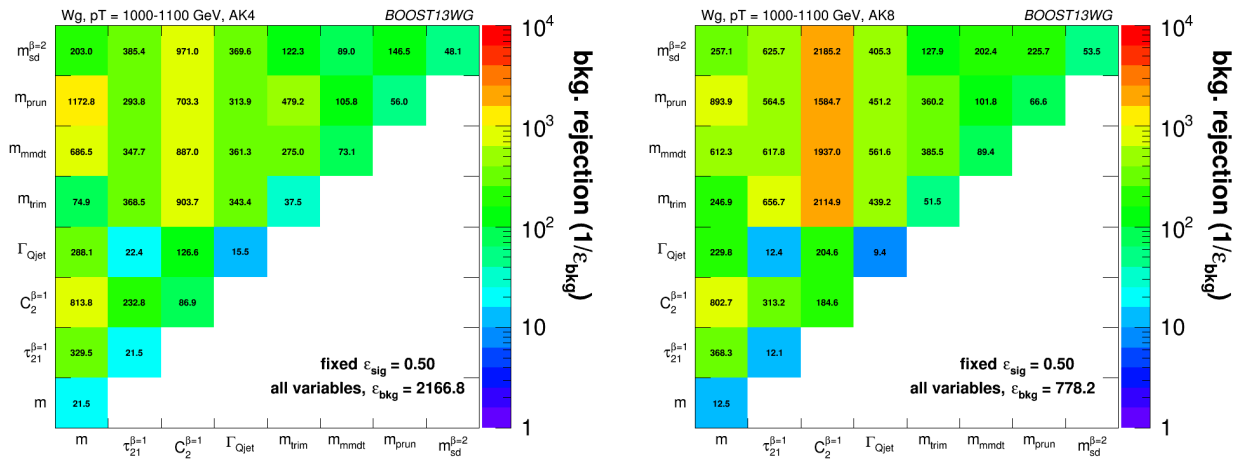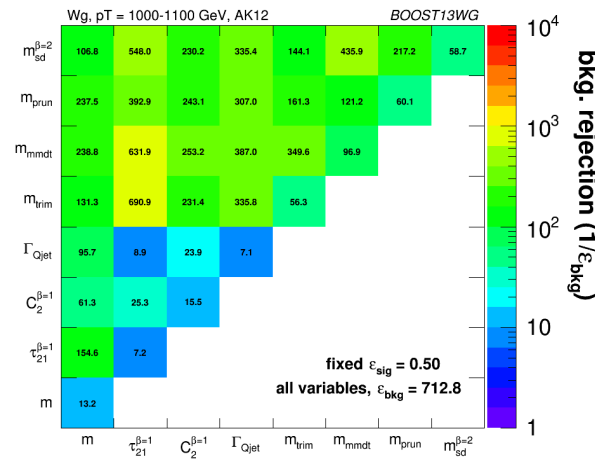
(a) anti-k_{T} R=0.8, p_T 500-650 GeV bin

(b) anti-k_{T} R=1.2, p_T 500-650 GeV bin

Fig. 17 The background rejection for a fixed signal efficiency (50%) of each BDT combination of each pair of variables considered, in the p_T 500-650 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=0.4, p_T 1.0-1.2 TeV bin

(b) anti-k_T R=0.8, p_T 1.0-1.2 TeV bin

(c) anti-k_T R=1.2, p_T 1.0-1.2 TeV bin

Fig. 18 The background rejection for a fixed signal efficiency (50%) of each BDT combination of each pair of variables considered, in the p_T 1.0-1.2 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.

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 allows 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.

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, while the grooming algorithms (trimming and pruning) do not incorporate a W-identification step. For a level playing field, we construct a W candidate from the three leading subjets by taking 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. All of the above taggers and groomers incorporate a step to remove pile-up and other soft radiation.

We also consider the performance of jet shape observables. In particular, we consider the N-subjettiness ratios $\tau_{32}^{\beta=1}$ and $\tau_{21}^{\beta=1}$, energy correlation function ratios $C_3^{\beta=1}$ and $C_2^{\beta=1}$, and the Qjet mass volatility Γ. In addition to the jet shape performance, we combine the jet shapes with the mass-reconstruction methods listed 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 inputs (list, or maybe refer back to Section 3), we scan over reasonable values of the inputs to determine the optimal value 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:** $f_{\text{cut}} \in [0.02, 0.14]$, $R_{\text{trim}} \in [0.1, 0.5]$
- **Pruning:** $z_{\text{cut}} \in [0.02, 0.14]$, $R_{\text{cut}} \in [0.1, 0.6]$

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. 19 shows the ROC curves for each of the top-tagging observables, with the bare jet mass also plotted for comparison. Unlike W tagging, the jet shape observables perform more poorly than jet mass. *(Check reasoning: this argument due to Andrew Larkoski).* As an example illustrating 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.19(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, the Johns Hopkins (JH) tagger out-performs the HEPTopTagger in its signal-to-background separation of both the top and W candidate masses, with larger discrepancy at higher p_T and larger jet radius. In Fig. 20, we show the histograms for the top mass output from the JH and HEPTop-Tagger for different R, optimized at a signal efficiency of 30%. The likely reason for this behavior 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 requirement 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.* [4] 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. Note that both the JH tagger and

the HEPTopTagger are superior at using the W candidate inside of the top for signal discrimination; this is because the the pruning and trimming algorithms do not have inherent W-identification steps and are not optimized for this purpose.

We also directly compare the performance of top mass and jet shape observables for different jet p_T and radius. The input parameters of the taggers, groomers, and shape variables are separately optimized for each p_T and radius:

p_T **comparison:** We compare various top tagging observables for jets in different p_T bins and $R = 0.8$ in Figs. 21 and 24. The tagging performance of jet shapes do not change substantially with p_T. $\tau_{32}^{(\beta=1)}$ and the Qjet volatility Γ have the most variation and tend to degrade with higher p_T (see Fig. 22-23). 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, 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 degrades at higher p_T, likely in part due to the background-shaping effects discussed earlier.

R **comparison:** We compare various top tagging observables for jets of different R and $p_T = 1.5 - 1.6$ TeV in Figs. 25-29. Most of the top-tagging parameters perform best for smaller radius; this is 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. As we show in Figs. 26-28, the distributions for both signal 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. The discriminating power generally gets worse with increasing R, except for $C_3^{(\beta=1)}$, which performs optimally at $R = 0.8$; in this case, the signal and background happen to have the same distribution around $R = 0.4$, and so $R = 0.8$ gives superior performance. *Is this really due to lack of 3-pronged structure in jet, or is it just luck?*

7.3 Performance of multivariable combinations

We now consider various combinations of the observables from Section 7.2. In particular, we consider the performance of individual taggers such as the JH tagger and HEPTopTagger, which output information about the t and W candidate masses and the helicity angle; groomers, such as trimming and pruning, which remove
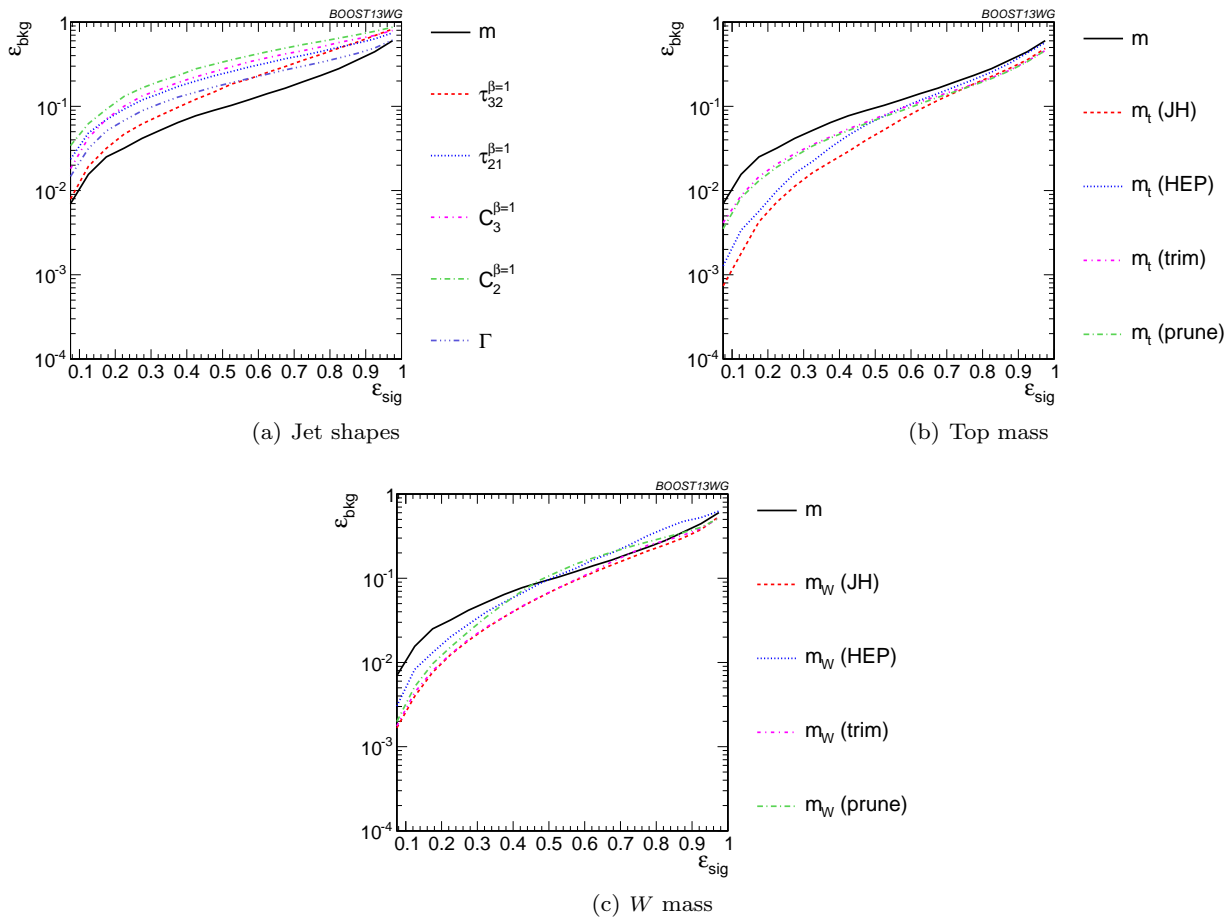
(a) Jet shapes

(b) Top mass

(c) W mass

Fig. 19 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.

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 above taggers/groomers 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.

Link to discussion of BDT methods

Fig. 30 shows our main results for the multivariable combinations; in all cases, we also show the ungroomed jet mass as a baseline comparison. In Fig. 30(a), we directly compare the performance of the HEPTopTagger, the JH tagger, trimming, and grooming. 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 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 t and W for different taggers contains complementary information.

In Fig. 30(b)-(d), we present the results for multivariable combinations of 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 2-3 improvement in signal discrimination. We see that, when combined with the tagger outputs, both the energy cor-
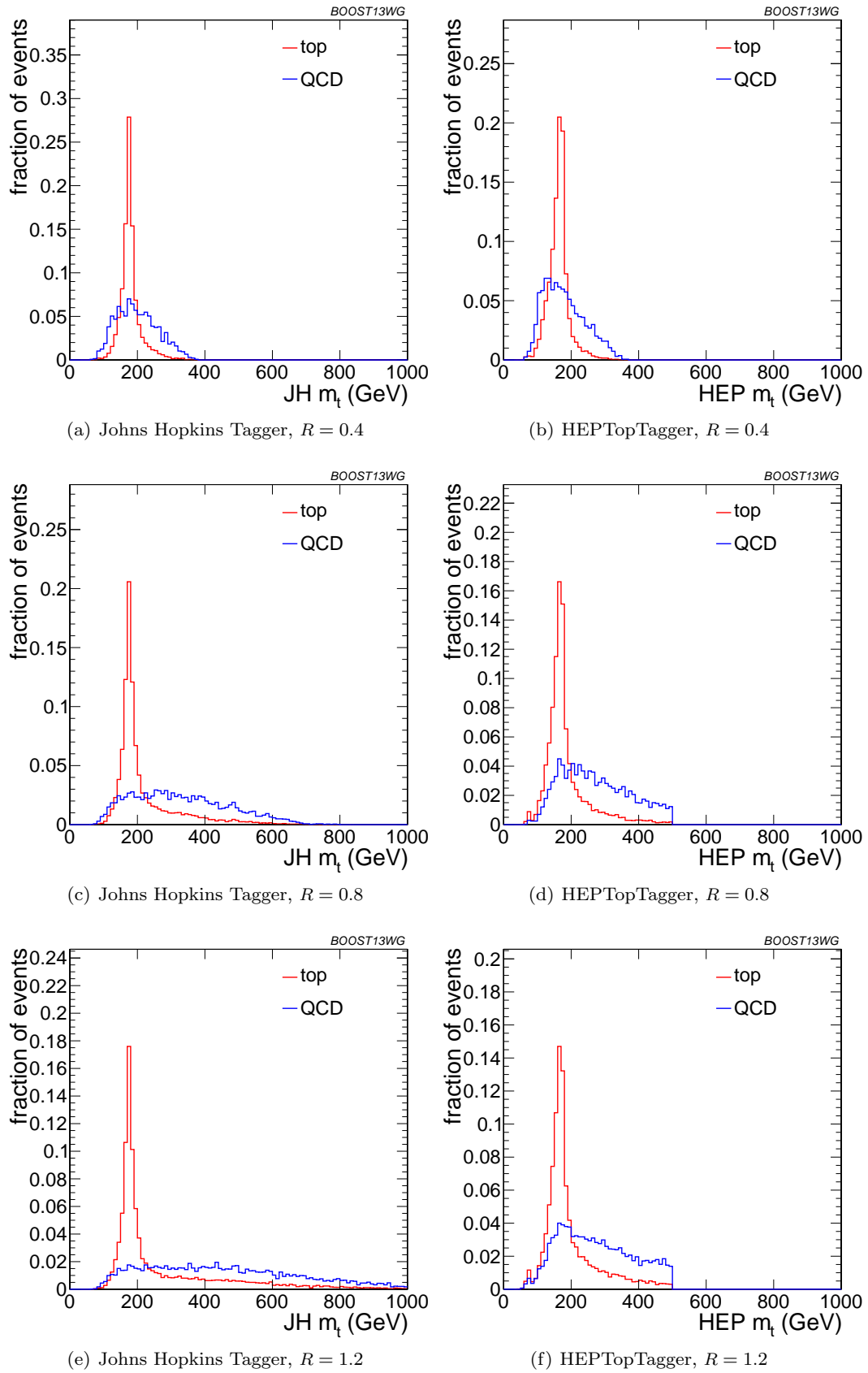
(a) Johns Hopkins Tagger, $R = 0.4$

(b) HEPTopTagger, $R = 0.4$

(c) Johns Hopkins Tagger, $R = 0.8$

(d) HEPTopTagger, $R = 0.8$

(e) Johns Hopkins Tagger, $R = 1.2$

(f) HEPTopTagger, $R = 1.2$

Fig. 20 Comparison of top mass reconstruction with the JH and HEPTopTaggers 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 at signal efficiency 0.3 and is normalized to the fraction of events passing the tagger.

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

(b) $C_3^{(\beta=1)}$

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

(d) $\tau_{32}^{(\beta=1)}$

(e) Qjet mass volatility
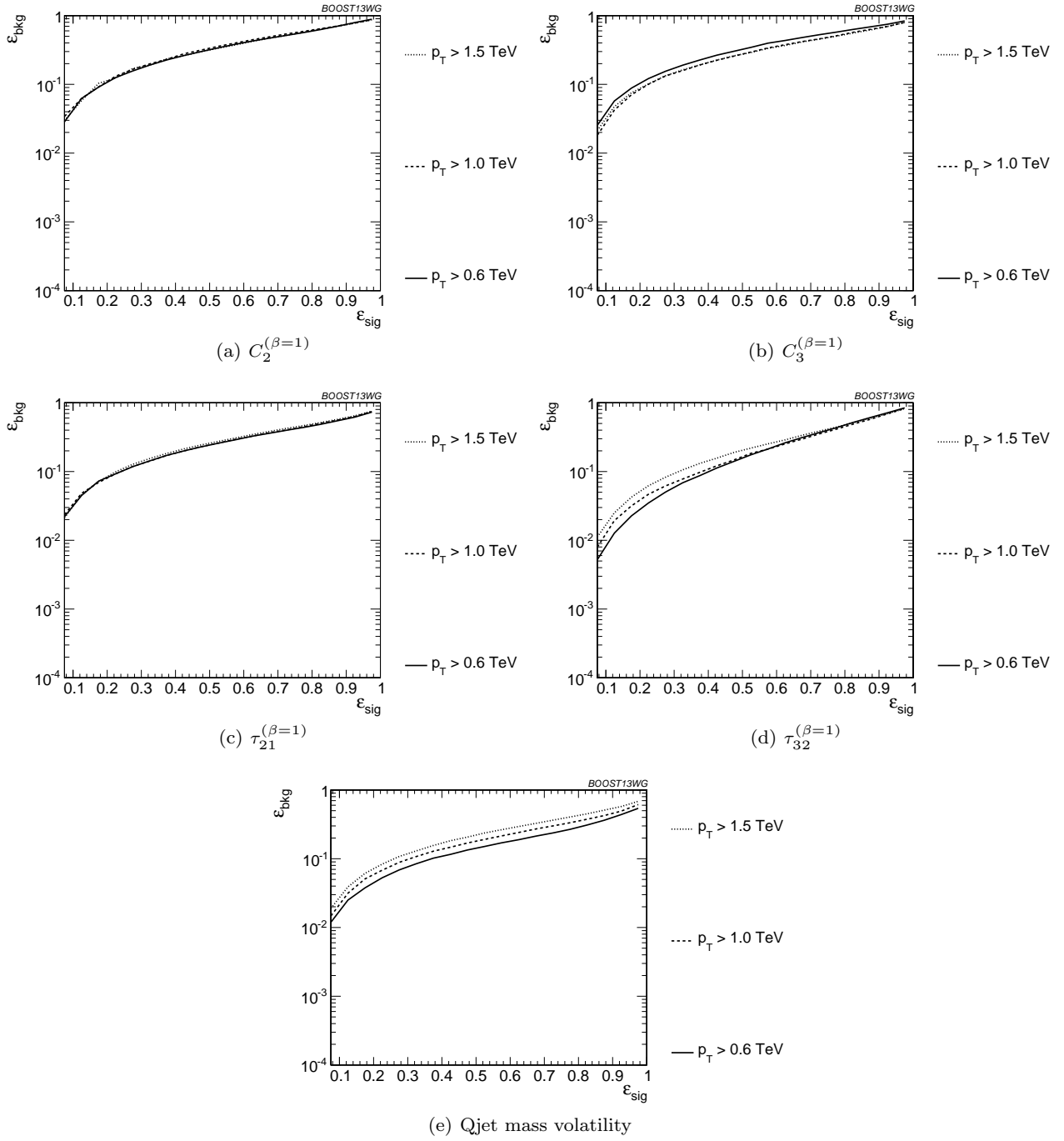
Fig. 21 Comparison of individual jet shape performance at different p_T using the anti-k_T R=0.8 algorithm.

relation 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. *OK?* Combining all shape observables with a single top tagger provides even more enhancement in discrimination power.

We directly compare the performance of the JH and HEPTopTaggers in Fig. 30(d). Combining the taggers with shape information nearly erases the difference between the tagging methods observed in Fig. 30(a); 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 discriminat-
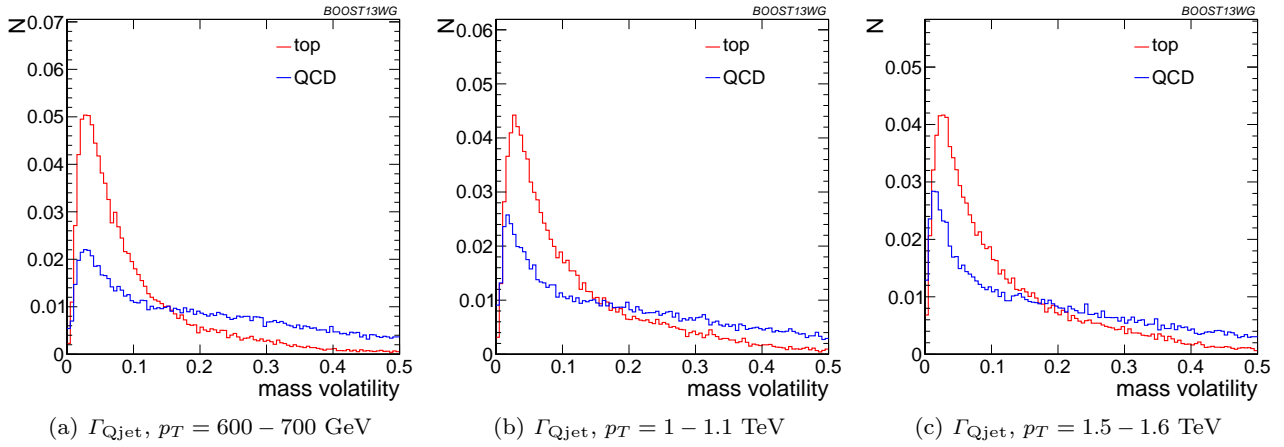
Fig. 22 Comparison of Γ_{Qjet} at $R = 0.8$ and different values of the p_T.

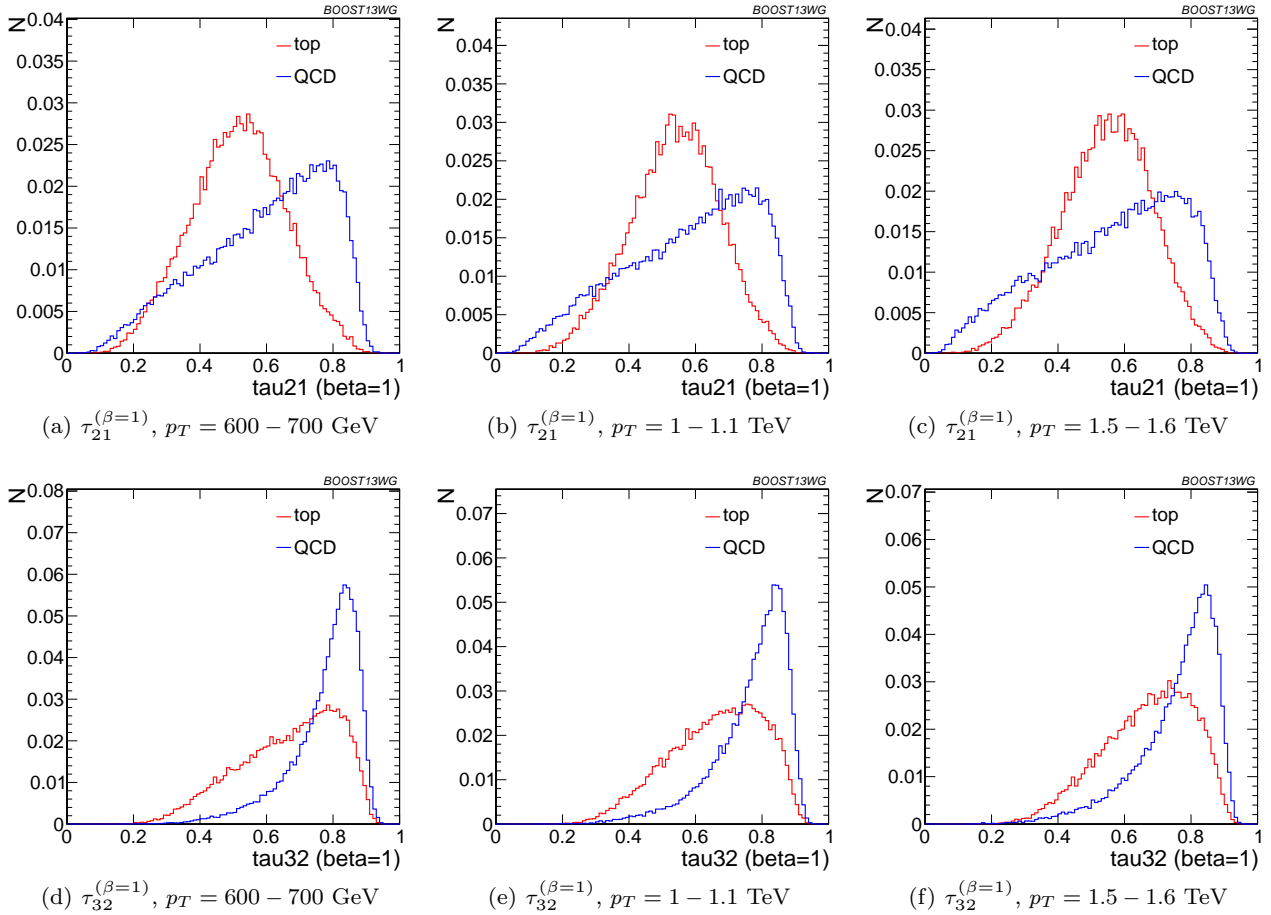**Fig. 23** Comparison of $\tau_{21}^{\beta=1}$ and $\tau_{32}^{\beta=1}$ with $R = 0.8$ and different values of the p_T.
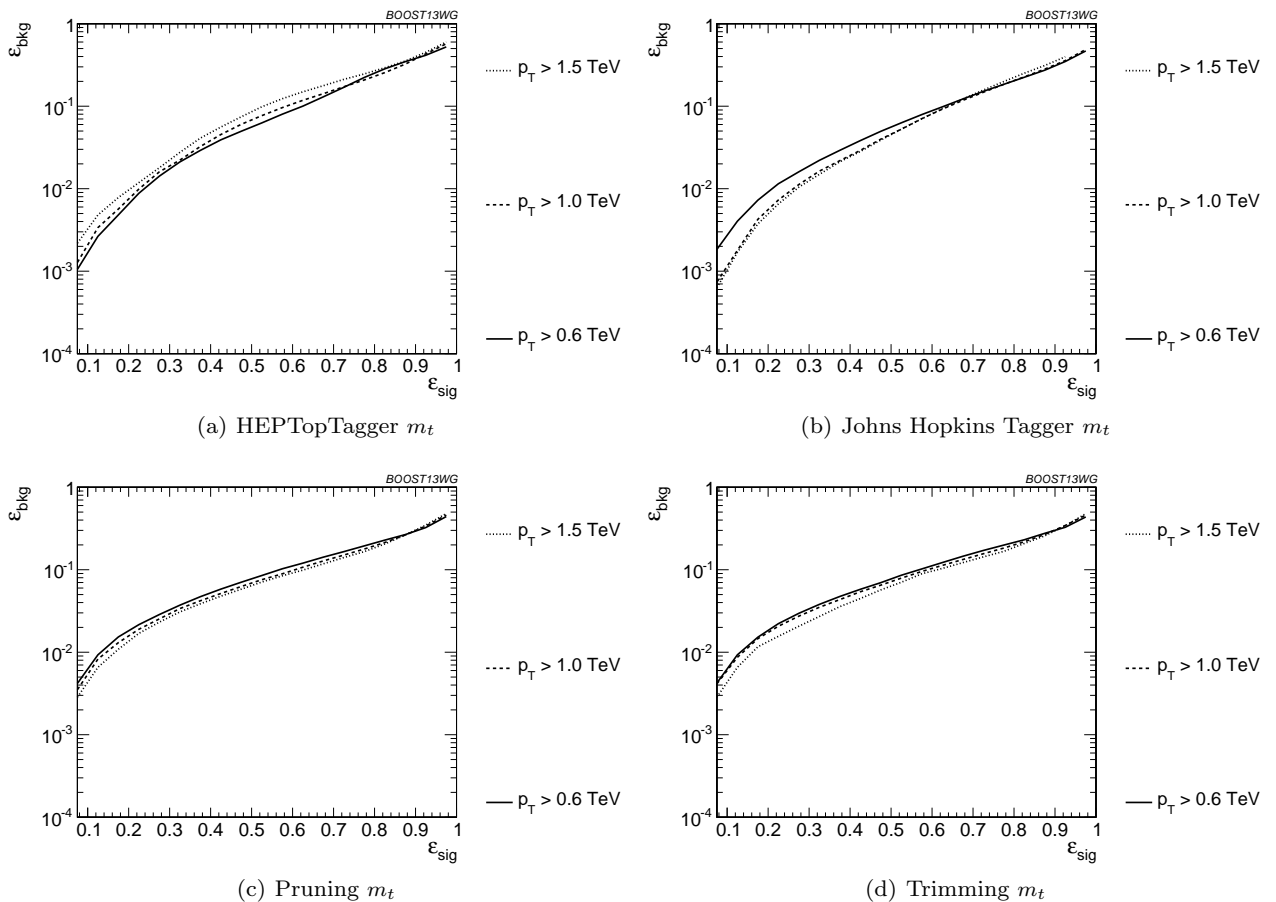
(a) HEPTopTagger m_t

(b) Johns Hopkins Tagger m_t

(c) Pruning m_t

(d) Trimming m_t

Fig. 24 Comparison of top mass performance of different taggers at different p_T using the anti-k_T R=0.8 algorithm.

ing power may be minimal, as various multivariable combinations are converging to within a factor of 20% or so.

In Fig. 30(e)-(g), 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 30(g), we find that the differences between pruning and trimming are erased when combined with shape information.

p_T **comparison**: 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 Fig. 31. The behaviour with p_T is qualitatively similar to the behaviour of the m_t observable for each tagger/groomer shown in Fig. 24; 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 Fig. 32, 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
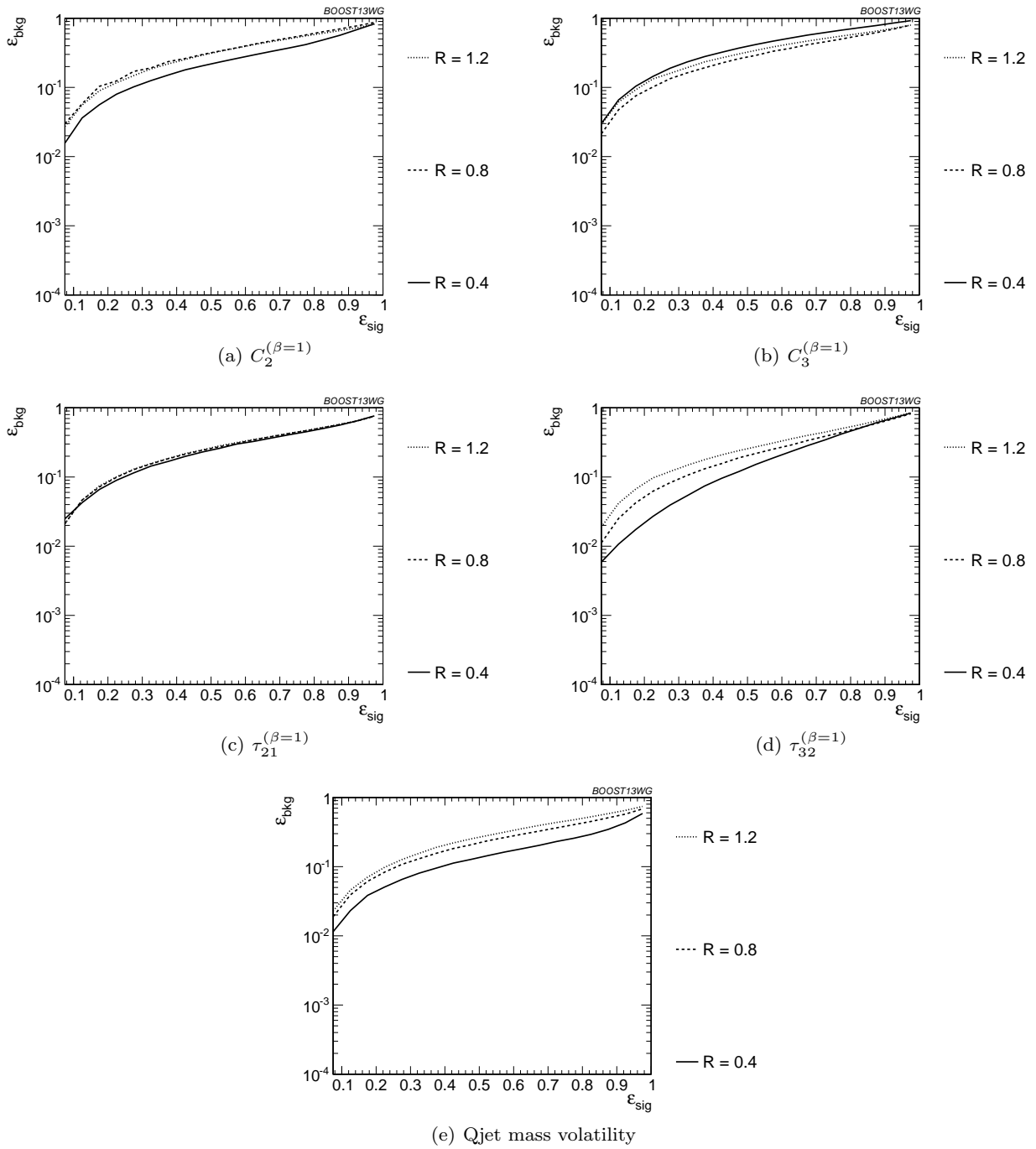
(a) $C_2^{(\beta=1)}$

(b) $C_3^{(\beta=1)}$

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

(d) $\tau_{32}^{(\beta=1)}$

(e) Qjet mass volatility

Fig. 25 Comparison of individual jet shape performance at different R in the $p_T = 1.5 - 1.6$ TeV bin.

curves, shown in Fig. 33, 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 $C_3^{(\beta=1)}$, which in Fig. 21(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.

R **comparison:** We now compare the BDT combinations of tagger outputs, with and without shape variables, at different R and $p_T = 1.5 - 1.6$ TeV. The taggers are optimized over all input parameters for each choice of R and signal efficiency, with the results shown in Fig. 34. We find that, for all taggers and groomers,
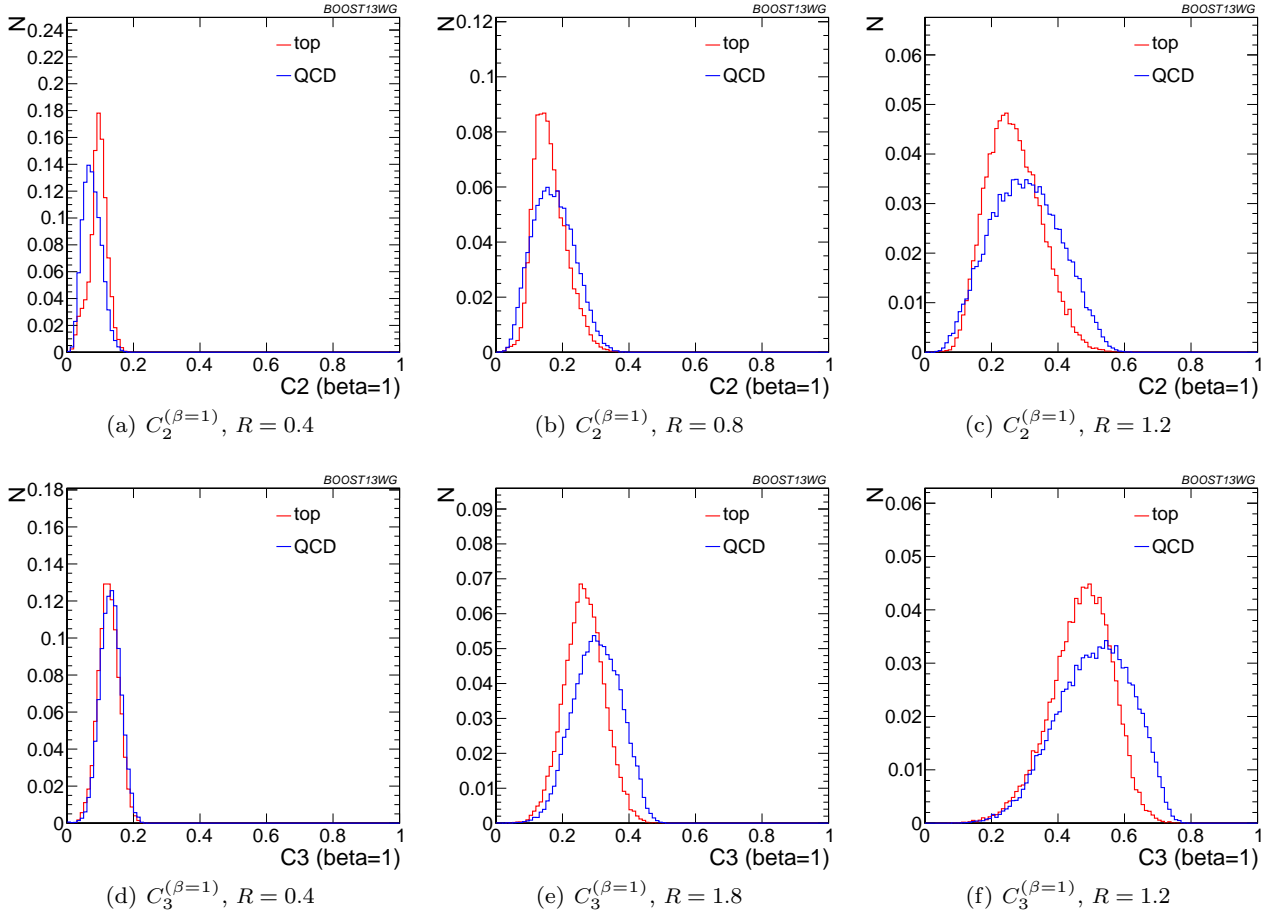
(a) $C_2^{(\beta=1)}$, $R = 0.4$ (b) $C_2^{(\beta=1)}$, $R = 0.8$ (c) $C_2^{(\beta=1)}$, $R = 1.2$

(d) $C_3^{(\beta=1)}$, $R = 0.4$ (e) $C_3^{(\beta=1)}$, $R = 1.8$ (f) $C_3^{(\beta=1)}$, $R = 1.2$

Fig. 26 Comparison of $C_2^{\beta=1}$ and $C_3^{\beta=1}$ in the $p_T = 1.5 - 1.6$ TeV bin and different values of the anti-k_{T} radius R.
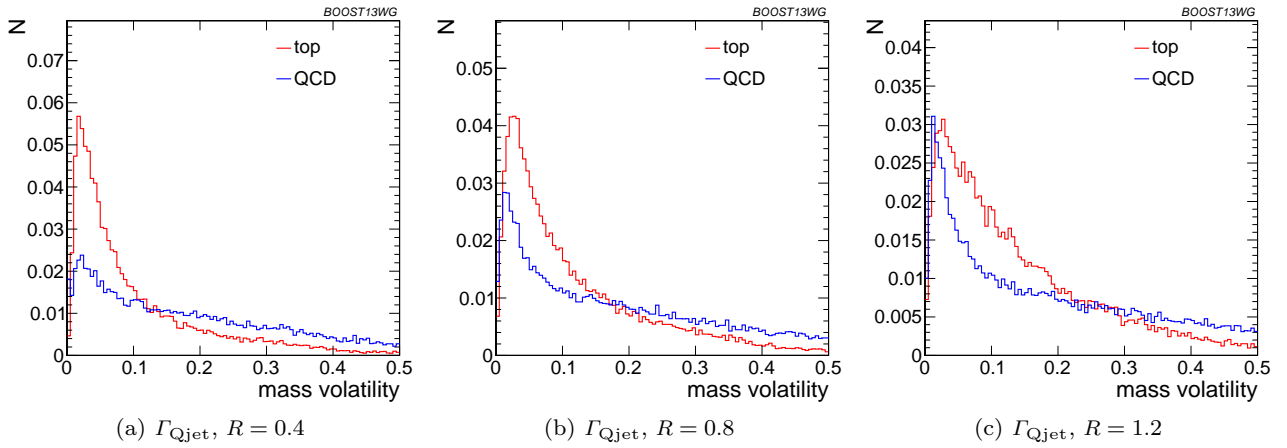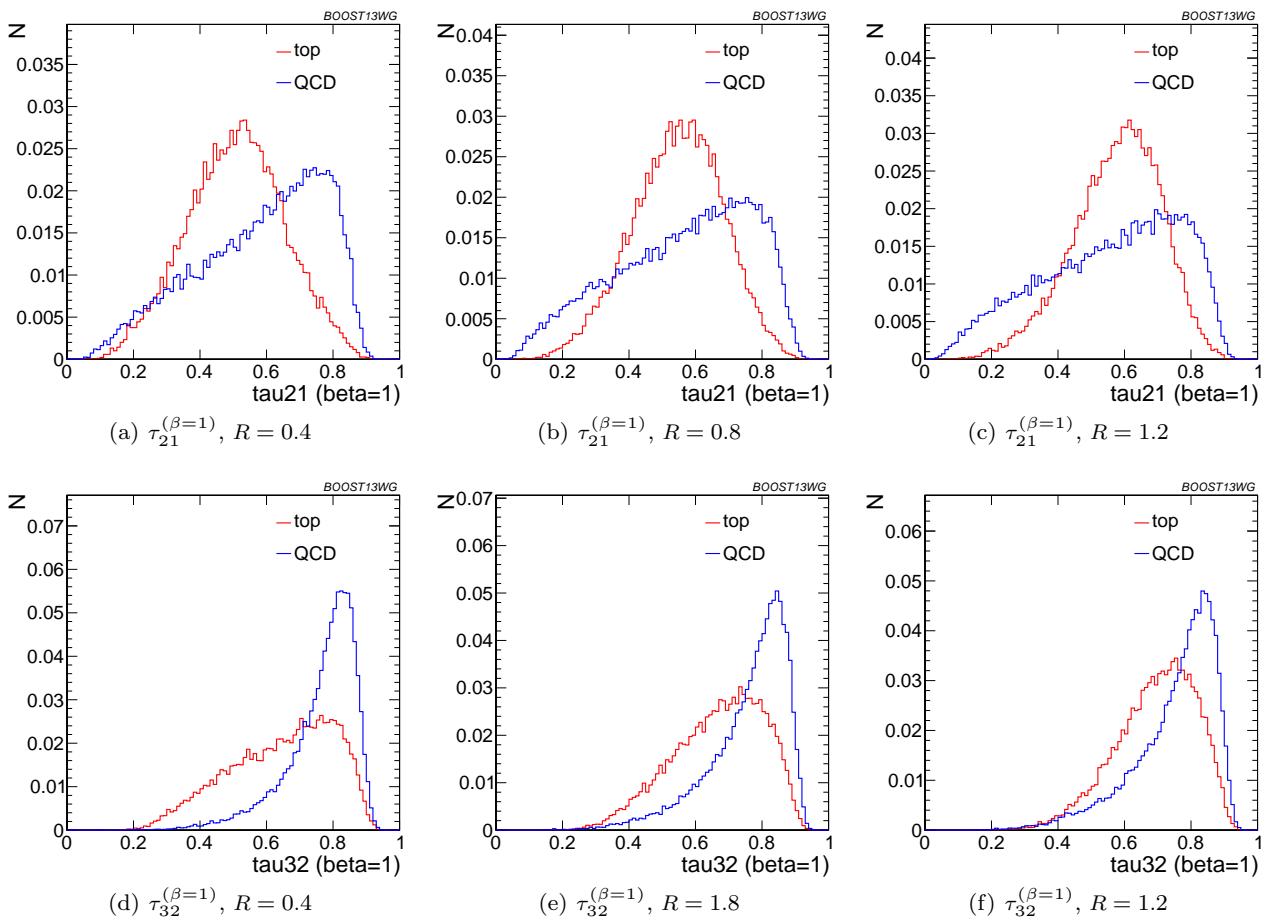

(a) Γ_{Qjet}, $R = 0.4$ (b) Γ_{Qjet}, $R = 0.8$ (c) Γ_{Qjet}, $R = 1.2$

Fig. 27 Comparison of Γ_{Qjet} in the $p_T = 1.5 - 1.6$ TeV bin and different values of the anti-k_{T} radius R.

Fig. 28 Comparison of $\tau_{21}^{\beta=1}$ and $\tau_{32}^{\beta=1}$ in the $p_T = 1.5 - 1.6$ TeV bin and different values of the anti-k_T radius R.

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 observables; for example, in the case of the JH tagger (Fig. 35), the R-dependence is identical for all combinations. The same holds true for the HEPTopTagger, trimming, and pruning.

(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.

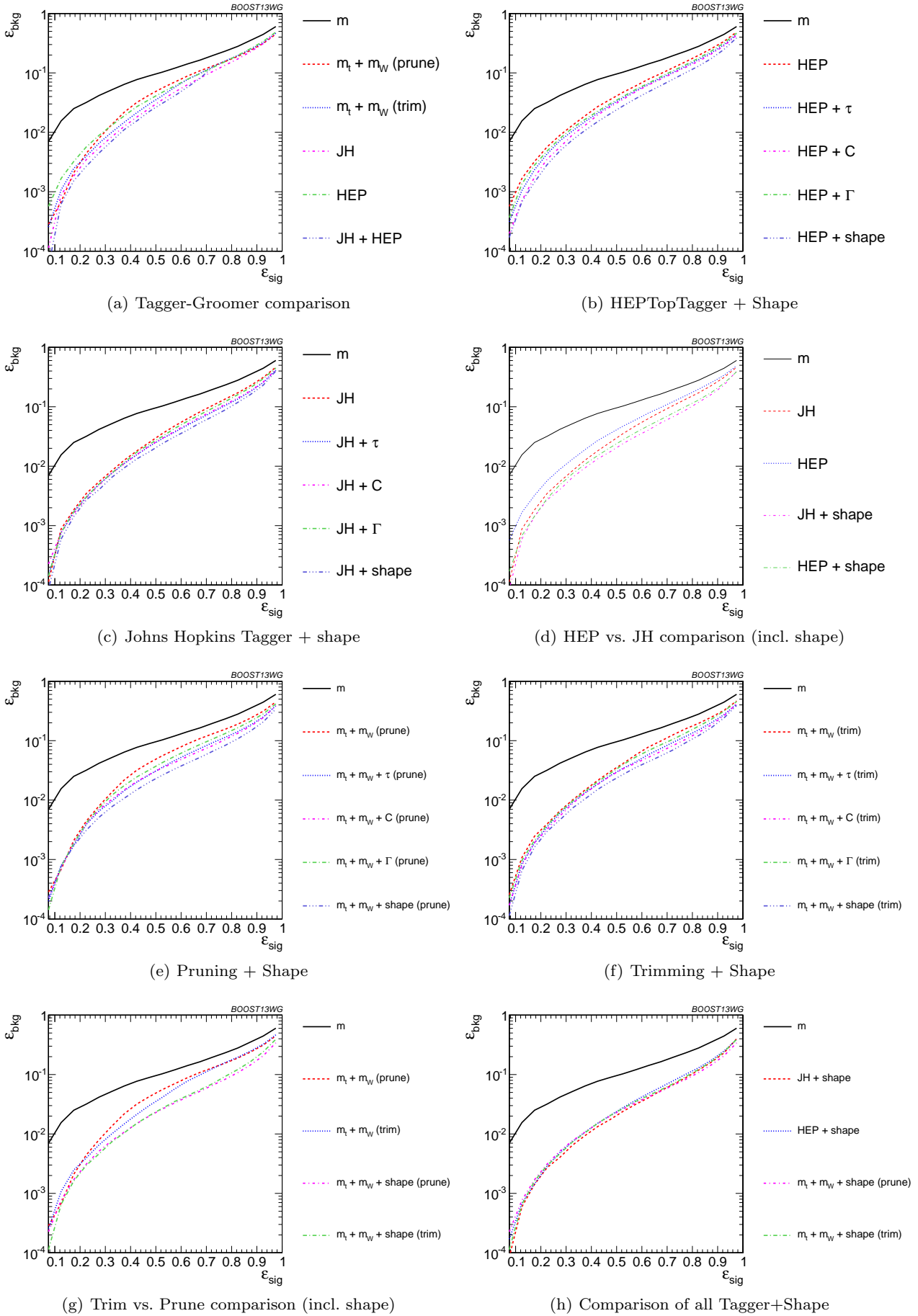(a) Tagger-Groomer comparison

(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. 30 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").

(a) HEPTopTagger

(b) Johns Hopkins Tagger

(c) Trimming

(d) Pruning

Fig. 31 Comparison of BDT combination of tagger performance at different p_T using the anti-k_T R=0.8 algorithm.

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

(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
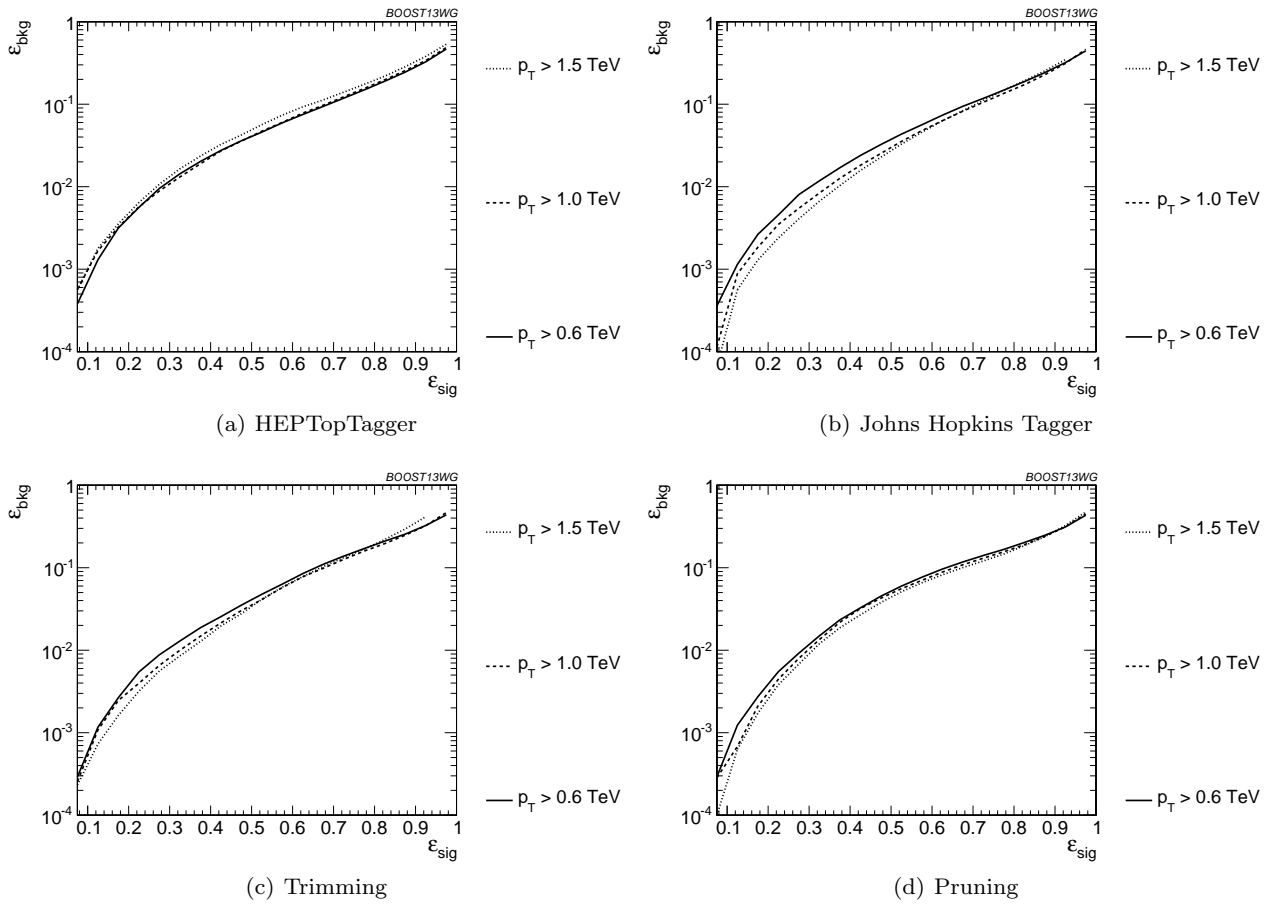
Fig. 33 Comparison of BDT combination of HEP tagger + shape at different p_T using the anti-k_T R=0.8 algorithm.
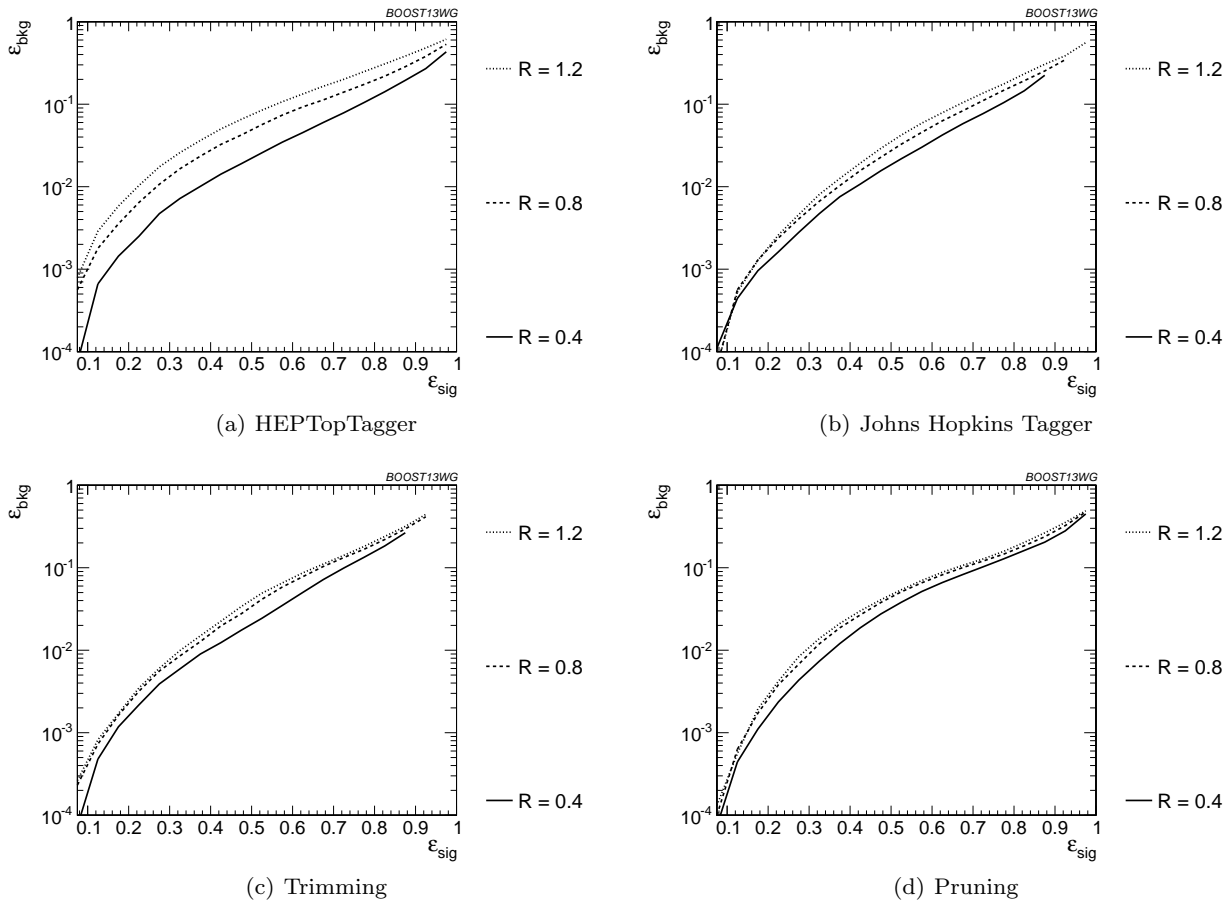
(a) HEPTopTagger

(b) Johns Hopkins Tagger

(c) Trimming

(d) Pruning

Fig. 34 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
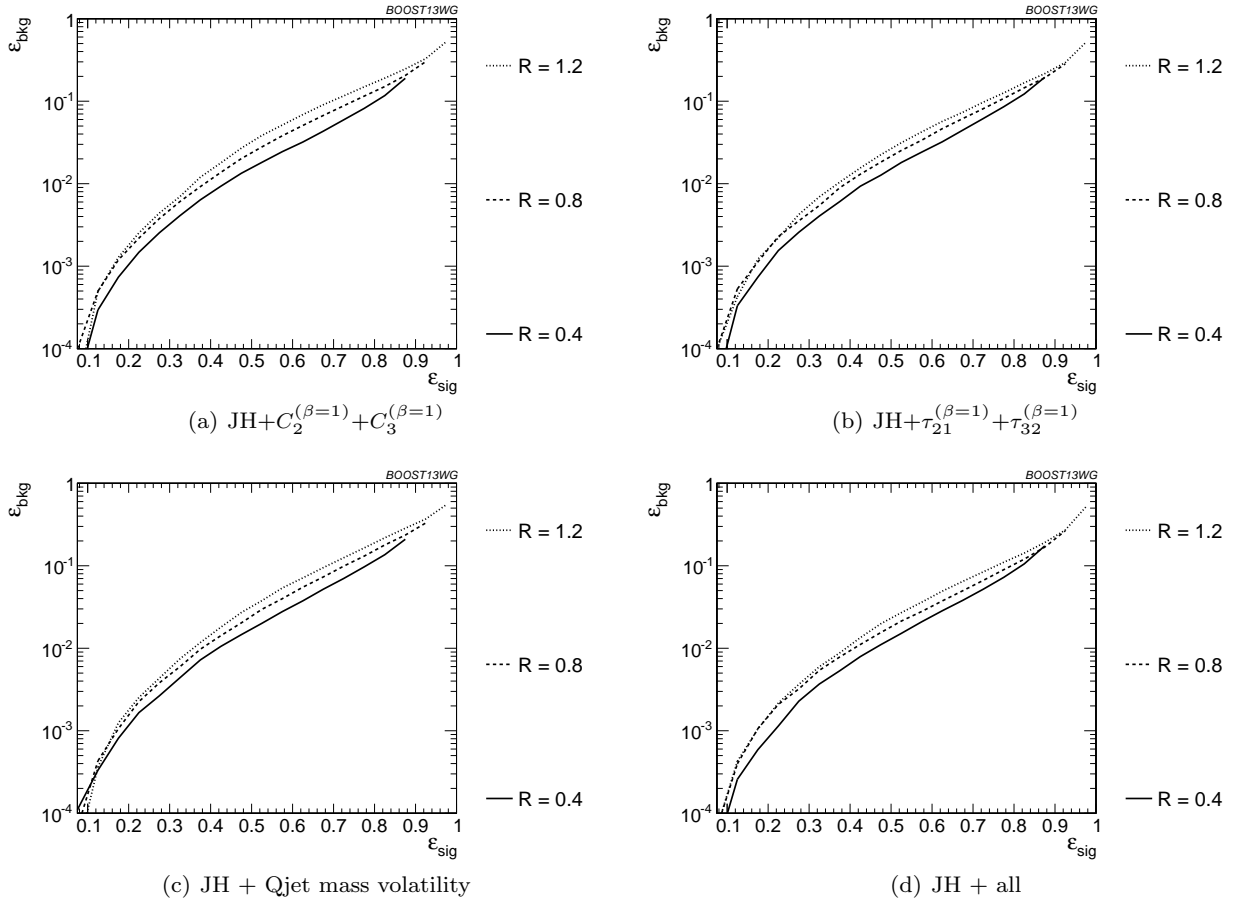
Fig. 35 Comparison of BDT combination of JH tagger + shape at different radius at $p_T = 1.5$-1.6 TeV.

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.

Optimizing at a single p_T **:** The shape observables typically do not have any input parameters to optimize. Therefore, we focus on the taggers and groomers. We show the performance of the top taggers, with all input parameters optimized to the $p_T = 1.5 - 1.6$ TeV values at each efficiency, in Fig. 36. Comparing to Fig. 29, we see that while the performance degrades slightly when

the high-p_T optimized points are used at other momenta, the ROC curves are consistent to within $O(1)$, with the performance of trimming degrading the most.

   The same holds true for the BDT combinations of the full tagger outputs (see Fig. 37). The performance for the sub-optimal taggers does not degrade substantially, with trimming seeing the largest decrease in discriminating power. However, we do observe one phenomenon: for taggers such as the HEPTopTagger and JH tagger, which 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 find enough signal candidates to reach the same efficiency at a different p_T. This explains why, in Fig. 37(1), the $p_T = 600 - 700$ GeV bin curve disappears at $\varepsilon \sim 0.75$, while the others continue up to nearly one. 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.

   Similar behaviour holds for the BDT combinations of taggers + shape observables, although we do not show the plots here because they look similar to Fig. 37.

Optimizing at a single R**:**
   We perform a similar analysis, but now optimize tagger parameters for each signal efficiency only at $R = 1.2$, and then use the same parameters for smaller R. We show the performance of the top taggers, with all input parameters optimized to the $R = 1.2$ values at each efficiency, in Fig. 38; these are to be compared with Fig. 29. For the HEPTopTagger, which is sensitive to the selected value of R, using the sub-optimal input parameters further degrades the performance at $R = 0.4$ and $R = 0.8$. 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. By contrast, the performance of the JH tagger and the grooming algorithms does not seem to suffer from using sub-optimal input parameters.

   The same holds true for the BDT combinations of the full tagger outputs (see Fig. 39). The performance for the sub-optimal taggers does not degrade substantially, and the HEPTopTagger is now more consistent with Fig. 34. The same behaviour holds for the BDT combinations of tagger outputs and shape observables.

Optimizing at a single efficiency:
   The strongest assumption so far is that the taggers are reoptimized for each signal efficiency point. This is useful for making a direct comparison of different top tagging algorithms, but is not particularly practical for
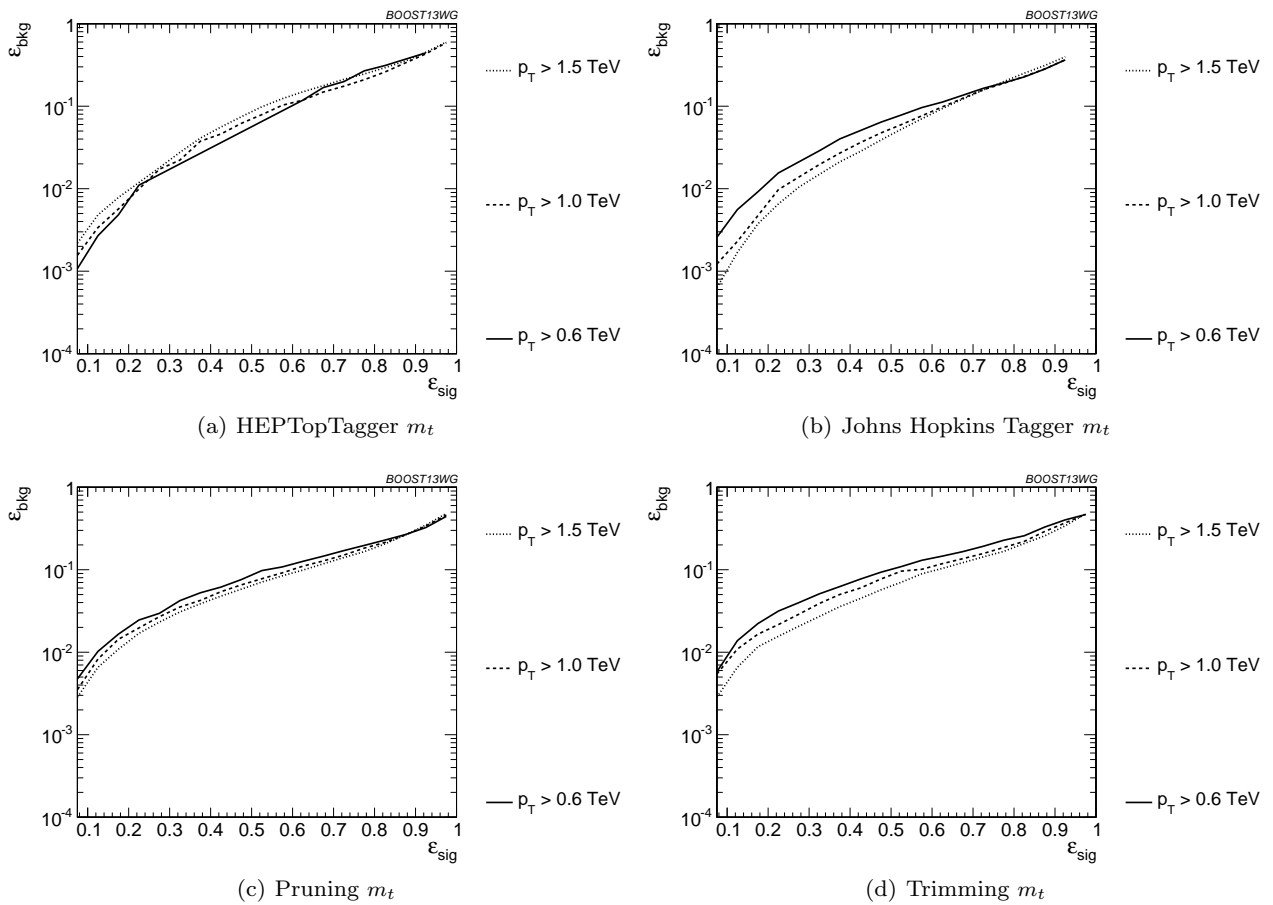
(a) HEPTopTagger m_t

(b) Johns Hopkins Tagger m_t

(c) Pruning m_t

(d) Trimming m_t

Fig. 36 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.

the experiments. We now consider the effects when the tagger inputs are optimized once, at $\varepsilon_S = 0.35$, and then used to determine the full ROC curve. We do this at $p_T = 1 - 1.1$ TeV and with $R = 0.8$.

The performance of each tagger, optimized only once, is shown in Fig. 40 for cuts on the top mass and W mass, and in Fig. 41 for BDT combinations of tagger outputs and shape variables. In both plots, it is apparent that, except at very small and very large signal efficiency, optimizing the tagger gives comparable performance to the scenario where the tagger is re-optimized for each efficiency. 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. Figs. 40-41 suggest that, while optimization at all signal efficiencies is a useful tool for comparing different algorithms, it is not necessary to achieve good top-tagging performance in experiments.
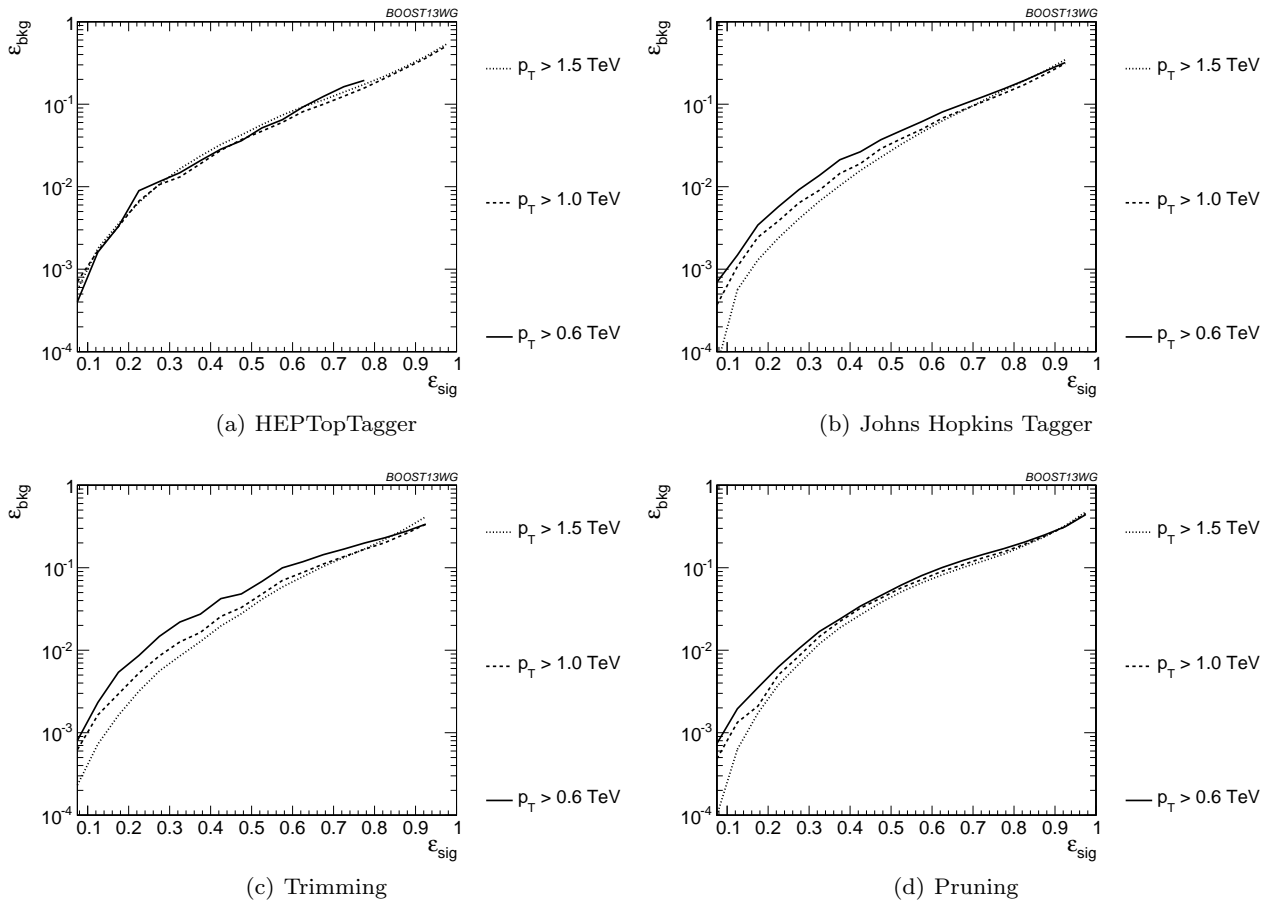
(a) HEPTopTagger

(b) Johns Hopkins Tagger

(c) Trimming

(d) Pruning

Fig. 37 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.
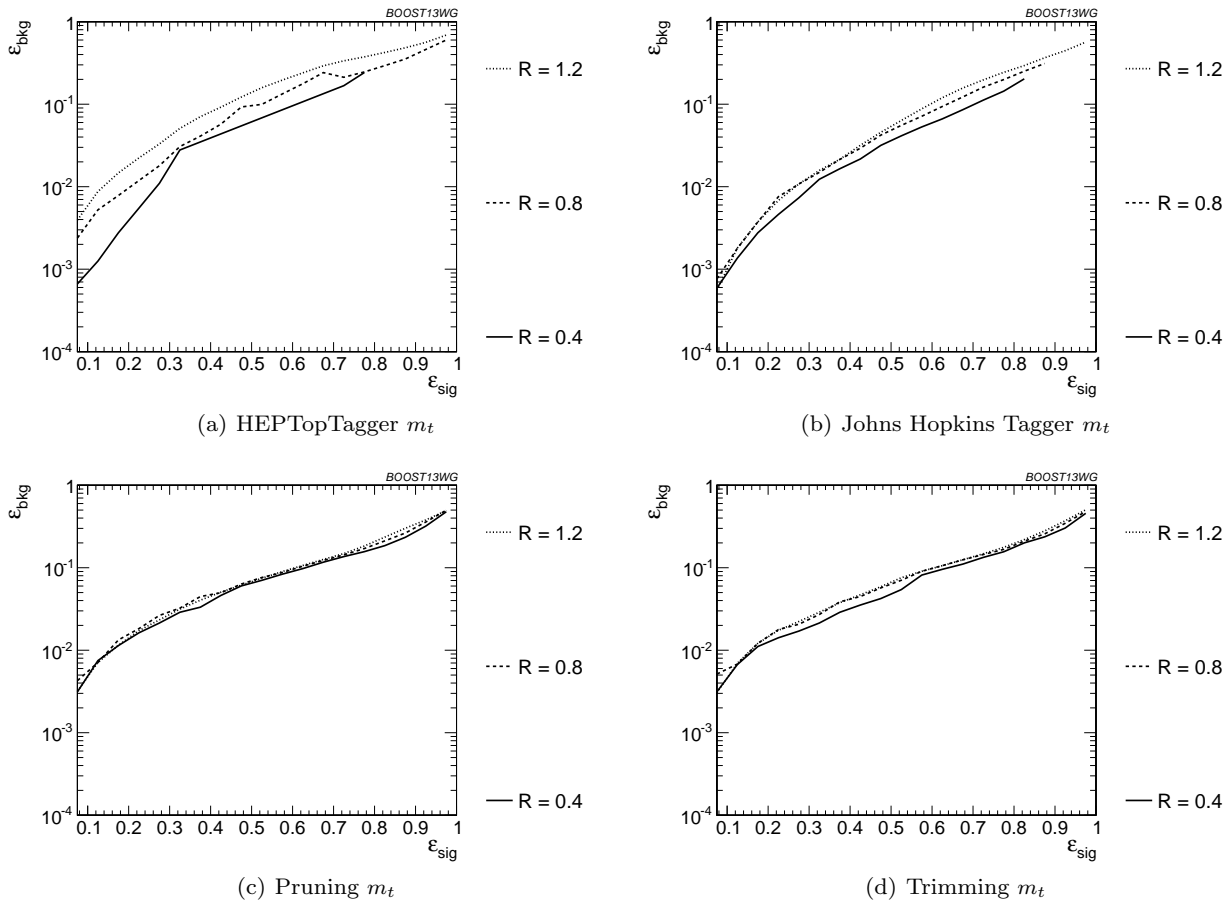
(a) HEPTopTagger m_t

(b) Johns Hopkins Tagger m_t

(c) Pruning m_t

(d) Trimming m_t

Fig. 38 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 $R = 1.2$.
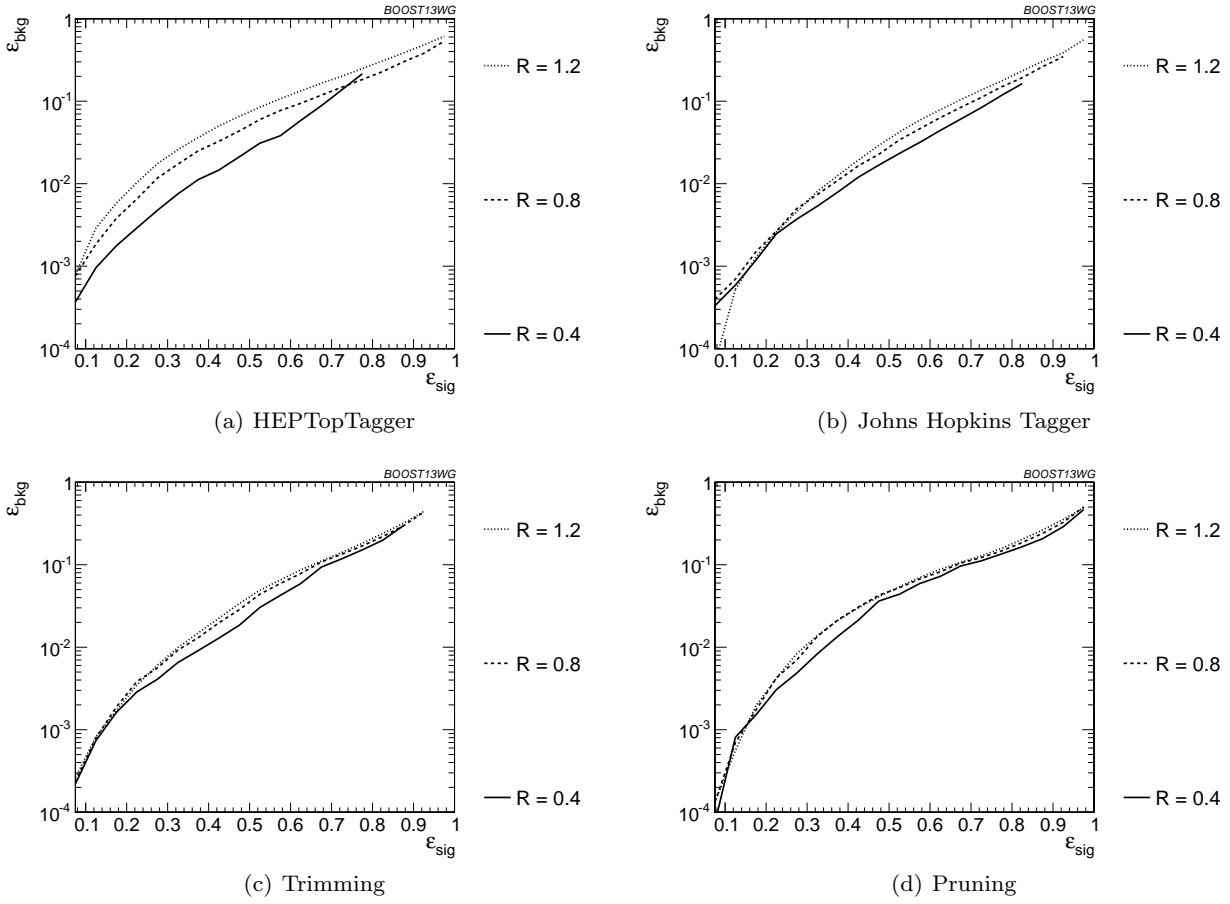
Fig. 39 Comparison of tagger and jet shape performance at different radius at $p_T = 1.5\text{-}1.6$ TeV; the tagger inputs are set to the optimum value for $R = 1.2$.
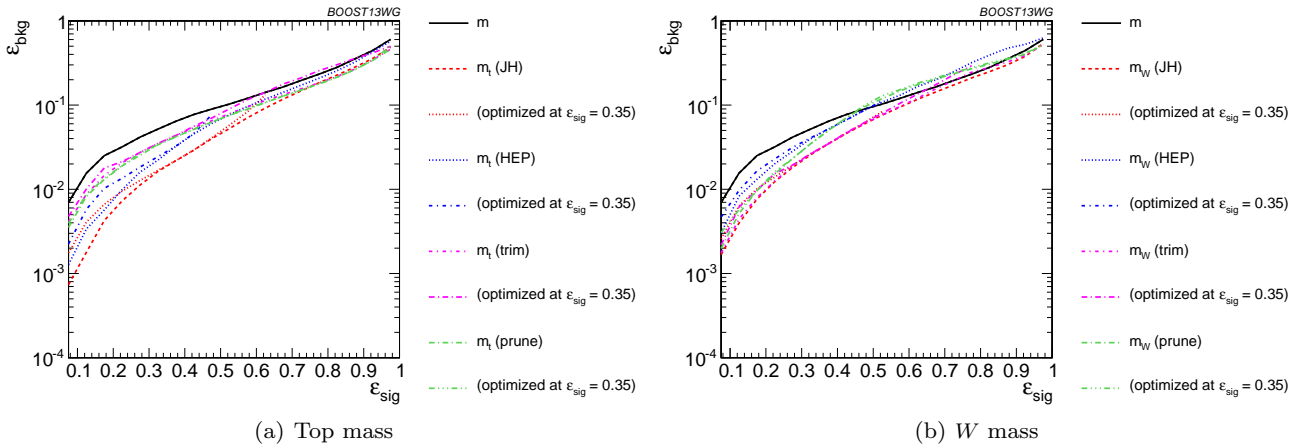

Fig. 40 Comparison of single-variable top-tagging performance in the $p_T = 1-1.1$ GeV bin using the anti-k_T, R=0.8 algorithm; the inputs for each tagger are optimized at the $\varepsilon_\mathrm{sig} = 0.35$ working point.

(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)

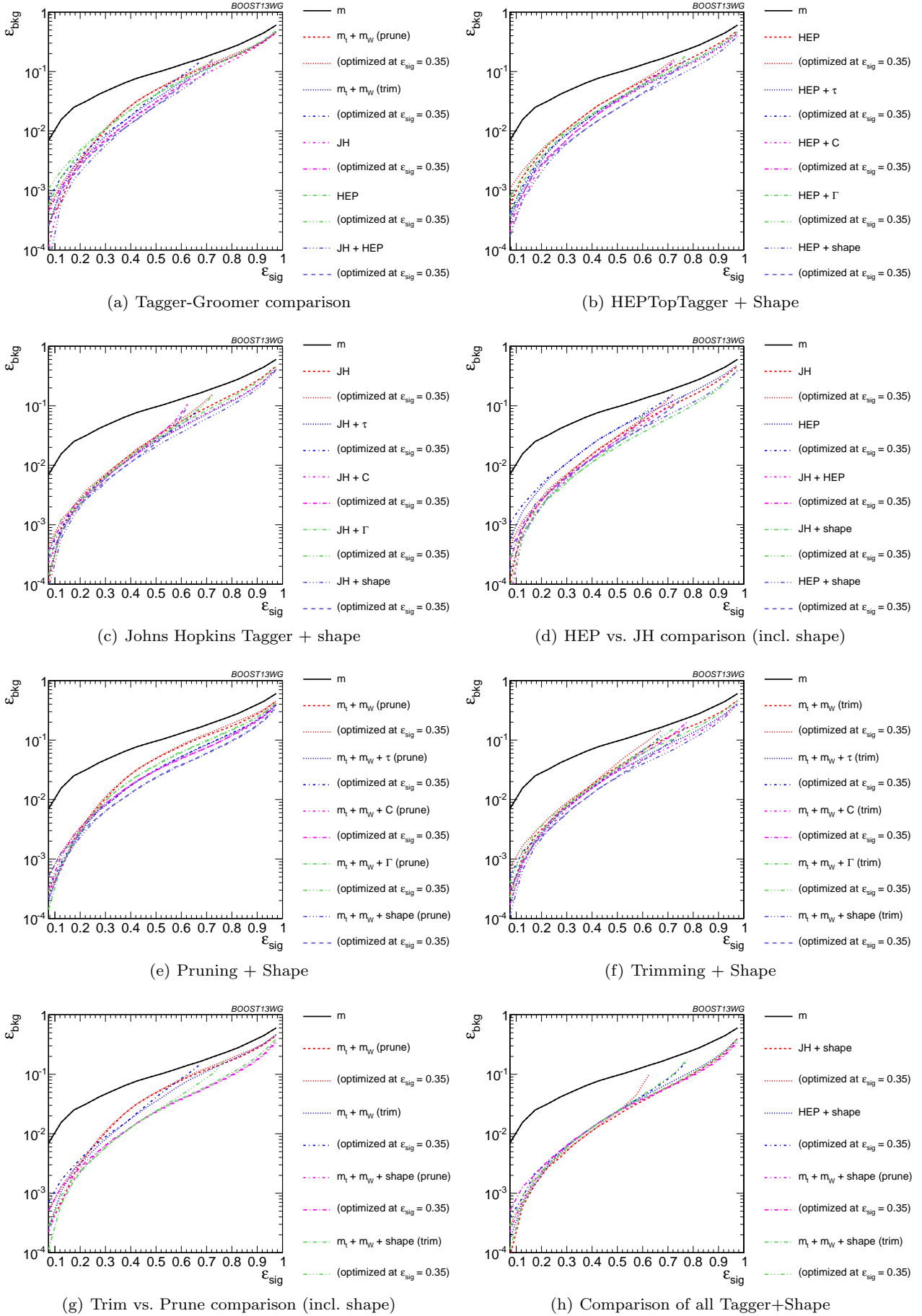(h) Comparison of all Tagger+Shape

Fig. 41 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 at the $\varepsilon_{\text{sig}} = 0.35$ working point.

8 Summary & Conclusions

This report discussed the correlations between observables and looked forward to jet substructure at Run II of the LHC at 14 TeV center-of-mass collisions eneergies.

Acknowledgements

References

1. A. Abdesselam, E. B. Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, et al., *Boosted objects: A Probe of beyond the Standard Model physics*, *Eur.Phys.J.* **C71** (2011) 1661, [`arXiv:1012.5412`].

2. A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, et al., *Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks*, *J.Phys.* **G39** (2012) 063001, [`arXiv:1201.0008`].

3. A. Altheimer, A. Arce, L. Asquith, J. Backus Mayes, E. Bergeaas Kuutmann, et al., *Boosted objects and jet substructure at the LHC*, `arXiv:1311.2708`.

4. C. Anders, C. Bernaciak, G. Kasieczka, T. Plehn, and T. Schell, *Benchmarking an Even Better HEPTopTagger*, *Phys.Rev.* **D89** (2014) 074047, [`arXiv:1312.1504`].