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, 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[7], A. Buckley[8], J. Butterworth[9], R. C. Camacho Toro[10], J. Caudron[11], Y.-T. Chien[12], J. Cogan[13], B. Cooper[9], D. Curtin[14], C. Debenedetti[15], J. Dolen[16], M. Eklund[17], S. El Hedri[11], S. D. Ellis[18], T. Embry[17], D. Ferencek[19], J. Ferrando[8], S. Fleischmann[20], M. Freytsis[21], M. Giulini[22], Z. Han[23], D. Hare[24], P. Harris[25], A. Hinzmann[26], R. Hoing[27], A. Hornig[12], M. Jankowiak[28], K. Johns[17], G. Kasieczka[29], R. Kogler[27], W. Lampl[17], A. J. Larkoski[30], C. Lee[12], R. Leone[17], P. Loch[17], D. Lopez Mateos[21], H. K. Lou[31], M. Low[32], P. Maksimovic[33], I. Marchesini[27], S. Marzani[30], L. Masetti[11], R. McCarthy[34], S. Menke[5], D. W. Miller[32], K. Mishra[24], B. Nachman[13], P. Nef[13], F. T. O'Grady[17], A. Ovcharova[35], A. Picazio[10], C. Pollard[8], B. Potter-Landua[25], C. Potter[25], S. Rappoccio[16], J. Rojo[36], J. Rutherfoord[17], G. P. Salam[25,37], J. Schabinger[38], A. Schwartzman[13], M. D. Schwartz[21], B. Shuve[39], P. Sinervo[40], D. Soper[23], D. E. Sosa Corral[22], M. Spannowsky[41], E. Strauss[13], M. Swiatlowski[13], J. Thaler[30], C. Thomas[25], E. Thompson[42], N. V. Tran[24], J. Tseng[36], E. Usai[27], L. Valery[43], J. Veatch[17], M. Vos[44], W. Waalewijn[45], J. Wacker[13], and C. Young[25]

[1]*Brookhaven National Laboratory, Upton, NY 11973, USA*
[2]*Duke University, Durham, NC 27708, USA*
[3]*Argonne National Laboratory, Lemont, IL 60439, USA*
[4]*CP3, Universite catholique du Louvain, B-1348 Louvain-la-Neuve, Belgium*
[5]*Max-Planck-Institute fuer Physik, 80805 Muenchen, Germany*
[6]*Charles University in Prague, FMP, V Holesovickach 2, Prague, Czech Republic*
[7]*University of California, Berkeley, CA 94720, USA*
[8]*University of Glasgow, Glasgow, G12 8QQ, UK*
[9]*University College London, WC1E 6BT, UK*
[10]*University of Geneva, CH-1211 Geneva 4, Switzerland*
[11]*Universitaet Mainz, DE 55099, Germany*
[12]*Los Alamos National Laboratory, Los Alamos, NM 87545, USA*
[13]*SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA*
[14]*University of Maryland, College Park, MD 20742, USA*
[15]*University of California, Santa Cruz, CA 95064, USA*
[16]*University at Buffalo, Buffalo, NY 14260, USA*
[17]*University of Arizona, Tucson, AZ 85719, USA*
[18]*University of Washington, Seattle, WA 98195, USA*
[19]*Rutgers University, Piscataway, NJ 08854, USA*
[20]*Bergische Universitaet Wuppertal, Wuppertal, D-42097, Germany*
[21]*Harvard University, Cambridge, MA 02138, USA*
[22]*Universitaet Heidelberg, DE-69117, Germany*
[23]*University of Oregon, Eugene, OR 97403, USA*
[24]*Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*
[25]*CERN, CH-1211 Geneva 23, Switzerland*
[26]*Universitaet Zuerich, 8006 Zuerich, Switzerland*
[27]*Universitaet Hamburg, DE-22761, Germany*
[28]*New York University, New York, NY 10003, USA*
[29]*ETH Zuerich, 8092 Zuerich, Switzerland*
[30]*Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[31]*Princeton University, Princeton, NJ 08544, USA*
[32]*University of Chicago, IL 60637, USA*
[33]*Johns Hopkins University, Baltimore, MD 21218, USA*
[34]*YITP, Stony Brook University, Stony Brook, NY 11794-3840, USA*
[35]*Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA*
[36]*University of Oxford, Oxford, OX1 3NP, UK*
[37]*LPTHE, UPMC Univ. Paris 6 and CNRS UMR 7589, Paris, France*
[38]*Universidad Autonoma de Madrid, 28049 Madrid, Spain*
[39]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada*
[40]*University of Toronto, Toronto, Ontario M5S 1A7, Canada*
[41]*IPPP, University of Durham, Durham, DH1 3LE, UK*
[42]*Columbia University, New York, NY 10027, USA*
[43]*LPC Clermont-Ferrand, 63177 Aubiere Cedex, France*
[44]*Instituto de Física Corpuscular, IFIC/CSIC-UVEG, E-46071 Valencia, Spain*
[45]*University of Amsterdam, 1012 WX Amsterdam, Netherlands*

[1]Address(es) of author(s) should be given

Received: date / Accepted: date

Abstract Over the past decade, a large number of jet substructure observables have been proposed in the literature, and explored at the LHC experiments. Such observables attempt to utilise the internal structure of jets in order to distinguish those initiated by quarks, gluons, or by boosted heavy objects, such as Top quarks and W bosons. This report, originating from and motivated by the BOOST2013 workshop, presents original particle-level studies that aim to improve our understanding of the relationships between jet substructure observables, their complementarity, and their dependence on the underlying jet properties, particularly the jet radius R and jet p_T. This is explored in the context of quark/gluon discrimination, boosted W-boson tagging and boosted Top quark tagging.

Keywords boosted objects · jet substructure · beyond-the-Standard-Model physics searches · Large Hadron Collider

1 Introduction

The center-of-mass energies at the Large Hadron Collider are large compared to the heaviest of known particles, even after account for parton density functions. With the start of the second phase of operation in 2015, the center-of-mass energy will further increase from 7 TeV in 2010-2011 and 8 TeV in 2012 to 13 TeV. Thus, even the heaviest states in the Standard Model (and potentially previously unknown particles) will often be produced at the LHC with substantial boosts, leading to a collimation of the decay products. For fully hadronic decays, these heavy particles will not be reconstructed as several 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 Quantum Chromodynamics (QCD) jets, as well as jets containing the decay of a heavy particle, with an array of new jet observables and detection techniques introduced and studied to distinguish the two types of jets. To allow the efficient propagation of results from these studies of jet substructure, a series of BOOST Workshops have been held on an annual basis: SLAC (2009, [1]), Oxford University (2010, [2]), Princeton University (2011, [3]), IFIC Valencia (2012 [4]), University of Arizona (2013 [5]), and, most recently, University College London (2014 [6]). Following each of these meetings, Working Groups have generated reports highlighting the most interesting new results, including studies of increasingly fine details. Previous BOOST reports can be found at [7–9].

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. In new analyses developed for the report, 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 [11] and the Johns Hopkins Tagger [12]. We also study the degree to which the discriminatory information provided by the observables and taggers overlaps by examining the extent to which the signal-background separation performance increases when two or more variables/taggers are combined in a multivariate analysis. Where possible, we provide a discussion of the physics behind the structure of the correlations and the p_T and R scaling that we observe.

We present the performance of observables in idealized simulations without pile-up and detector resolution effects, with the primary goal of studying the correlations between observables and the dependence on jet radius and p_T. The relationship between substructure observables, their correlations, and how these depend on the jet radius R and jet p_T should not be too sensitive to pile-up and resolution effects; conducting studies using idealized simulations allows us to more clearly elucidate the underlying physics behind the observed performance, and also provides benchmarks for the development of techniques to mitigate pile-up and detector effects. A full study of the performance of pile-up and detector mitigation strategies is beyond the scope of the current report, and will be the focus of upcoming studies.

The report is organized as follows: in Sections 2-4, we describe the methods used in carrying out our analysis, with a description of the Monte Carlo event sample generation in Section 2, the jet algorithms, observables and taggers investigated in our report in Section 3, and an overview of the multivariate techniques used to combine multiple observables into single discriminants in Section 4. Our results follow in Sections 5-7, with q/g-tagging studies in Section 5, W-tagging studies in Section 6, and Top-tagging studies in Section 7. Finally we offer some summary of the studies and general conclusions in Section 8.

This report presents original analyses and discussions pertaining to the performance of and correlations between various jet substructure techniques applied to quark/gluon discrimination, W-boson tagging, and Top tagging. The principal organizers of and contributors to the analyses presented in the report are: B. Cooper, S. D. Ellis, M. Freytsis, A. Hornig, A. Larkoski, D. Lopez Mateos, B. Shuve, and N. V. Tran.

2 Monte Carlo Samples

Below, we describe the Monte Carlo samples used in the q/g tagging, W tagging and Top tagging sections of this report. Note that no pile-up (additional proton-proton interactions beyond the hard scatter) are included in any samples, 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; such effects are deferred to future study.

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. The W bosons are decayed 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 [13], while W^+W^- samples were generated using the JHU GENERATOR [14–16] to allow for separation of longitudinal and transverse polarizations. Both were generated using CTEQ6L1 PDFs [17]. 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 in this report were 300-400 GeV, 500-600 GeV and 1.0-1.1 TeV. The samples were then showered through PYTHIA8 (version 8.176) [18] using the default tune 4C [19]. For each of the various samples (W,q,g) and p_T bins, 500k events were simulated.

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 [20–25], 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 top quark p_T. The QCD samples were generated with a lower 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 $Q_{\text{cut}} = 40, 60, 80$ GeV for the $p_{T\,\text{min}} = 600, 1000,$ 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 Sections 3.1, 3.2, 3.3 and 3.4, we describe the various jet algorithms, groomers, taggers and other substructure variables used in these studies. Over the course of our study, we considered a larger set of observables, but for presentation purposes we included only a subset in the final analysis, eliminating redundant observables.

3.1 Jet Clustering Algorithms

Jet clustering: Jets were clustered using sequential jet clustering algorithms [26] implemented in FASTJET 3.0.3. Final state particles i, j are assigned a mutual distance d_{ij} and a distance to the beam, $d_{i\text{B}}$. The particle pair with smallest d_{ij} are recombined and the algorithm repeated until the smallest distance is from a particle i to the beam, $d_{i\text{B}}$, in which case i is set aside and labelled as a jet. The distance metrics are defined as

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

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

where $\Delta R_{ij}^2 = (\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2$. In this analysis, we use the anti-k_t algorithm ($\gamma = -1$) [27], the Cambridge/Aachen (C/A) algorithm ($\gamma = 0$) [28, 29], and the k_t algorithm ($\gamma = 1$) [30, 31], each of which has varying sensitivity to soft radiation in the definition of the jet.

Qjets: We also perform non-deterministic jet clustering [32, 33]. 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.

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

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

in which case the merger is vetoed and the softer branch discarded. The default parameters used for pruning [34] in most studies in this report are $z_{\mathrm{cut}} = 0.1$ and $R_{\mathrm{cut}} = 0.5$. 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_{trim} with the k_t algorithm. Discard all subjets i with

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

The default parameters used for trimming [35] in most studies in this report are $R_{\mathrm{trim}} = 0.2$ and $f_{\mathrm{cut}} = 0.03$.

Filtering: Given a jet, re-cluster the constituents into subjets of radius R_{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) [36]. While we do not independently use filtering, it is an important step of the HEPTopTagger to be defined later.

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

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

discard the softer subjet and repeat. Otherwise, take j to be the final soft-drop jet [37]. Soft drop has two input parameters, the angular exponent β and the soft-drop scale z_{cut}, with default value $z_{\mathrm{cut}} = 0.1$.

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

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

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 de-clustering continues until only one branch remains, the jet is considered to have failed the tagging criteria [38]. In this study we use by default $\mu = 1.0$ (i.e. implement no mass drop criteria) and $y_{\mathrm{cut}} = 0.1$.

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_{T} is less than $\delta_p\, p_{\mathrm{T\,jet}}$. 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 [12]. 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. The two free input parameters of the John Hopkins tagger in this study are δ_p and δ_R, defined above, and their values are optimized for different jet kinematics and parameters in Section 7.

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 $R_{\mathrm{filt}} = \min(0.3, \Delta R_{ij})$, 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 [11]. The output of the tagger is m_t, m_W, and θ_{h} (as defined in the Johns Hopkins Tagger). The two free input parameters of the HEPTopTagger in this study are m and μ, defined above, and their values are optimized for different jet kinematics and parameters in Section 7.

Top Tagging with Pruning or Trimming: For comparison with the other top taggers, we add a W reconstruction step to the pruning and trimming algorithms 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

The jet substructure observables defined in this section are calculated using jet constituents prior to any grooming.

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, $\mathit{\Gamma}_{\rm Qjet}$, is defined as [32]

$$\mathit{\Gamma}_{\rm Qjet} = \frac{\sqrt{\langle m_J^2\rangle - \langle m_J\rangle^2}}{\langle m_J\rangle}, \tag{8}$$

where averages are computed over the Qjet interpretations. We use a rigidity parameter of $\alpha = 0.1$ (although other studies suggest a smaller value of α may be optimal [32, 33]), and 25 trees per event for all of the studies presented here.

N**-subjettiness:** N-subjettiness [39] 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 [40]. We use both the WTA and one-pass k_t optimization axes in our analyses.

Often, a powerful discriminant is 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 [41] 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 [42]:

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

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. Note that $C_2^{(0)}$ is identical to the variable PTD introduced by CMS in [43].

4 Multivariate Analysis Techniques

Multivariate techniques are used to combine multiple variables into a single discriminant in an optimal manner. The extent to which the discrimination power increases in a multivariable combination indicates to what extent the discriminatory information in the variables overlaps. There exist alternative strategies for studying correlations in discrimination power, such as "truth matching" [44], but these are not explored here.

In all cases, the multivariate technique used to combine variables is a Boosted Decision Tree (BDT) as implemented in the TMVA package [45]. 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

These parameter values are chosen to reduce the effect of overtraining. Additionally, the simulated data were split into training and testing samples and comparisons of the BDT output were compared to ensure that the BDT performance was not affected by overtraining.

5 Quark-Gluon Discrimination

In this section, we examine the differences between quark- and gluon-initiated jets in terms of substructure variables. At a fundamental level, the primary difference between quark- and gluon-initiated jets is the color charge of the initiating parton, typically expressed in terms of the ratio of the corresponding Casimir factors $C_F/C_A = 4/9$. Since the quark has the smaller color charge, it radiates less than a corresponding gluon and the resulting jet will contain fewer constituents. We determine the extent to which the substructure observables capturing this difference are correlated, providing some theoretical understanding of these variables and their performance. The motivation for these studies arises 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 [48, 49], 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.

Other examples of recent analytic studies of the correlations between jet observables relevant to quark jet versus gluon jet discrimination can be found in [41, 44, 46, 47].

5.1 Methodology and Observable Classes

These studies use the qq and gg MC samples described in Section 2. 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 (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:

– Number of constituents (N_constits) in the jet.
– Pruned Qjet mass volatility, Γ_Qjet.
– 1-point energy correlation functions, C_1^β with $\beta = 0, 1, 2$.
– 1-subjettiness, τ_1^β with $\beta = 1, 2$. The N-subjettiness axes are computed using one-pass k_t axis optimization.
– Ungroomed jet mass, m.

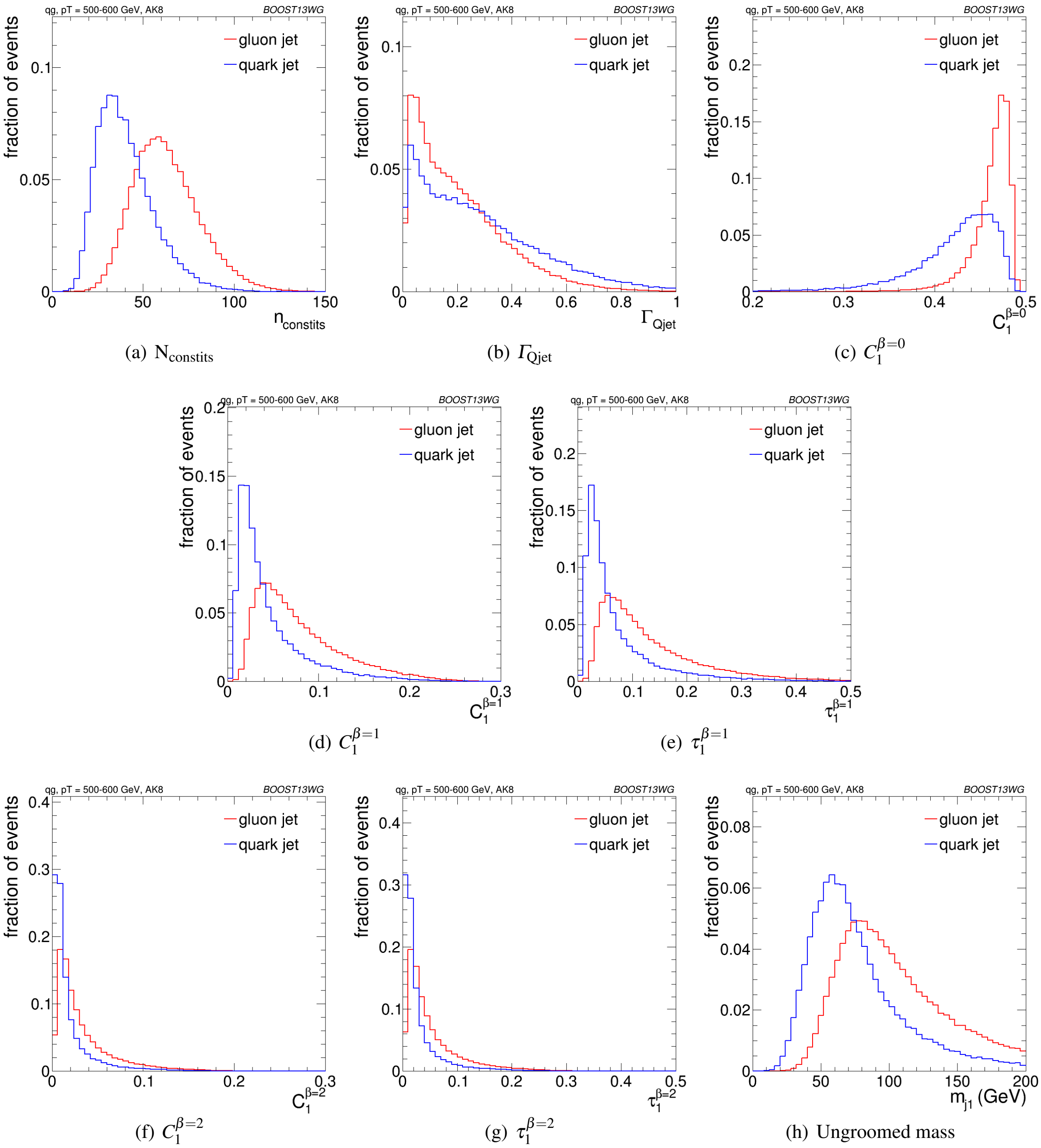
For simplicity, we hereafter refer to quark-initiated jets (gluon-initiated jets) as quark jets (gluon jets).

We will demonstrate that, in terms of their jet-by-jet correlations and their ability to separate quark jets from gluon jets, the above observables fall into five Classes. The first three observables, N_constits, Γ_Qjet and $C_1^{\beta=0}$, each constitutes a Class of its own (Classes I to III) in the sense that they each carry some independent information about a jet and, when combined, provide substantially better quark jet and gluon jet separation than any one observable alone. Of the remaining observables, $C_1^{\beta=1}$ and $\tau_1^{\beta=1}$ comprise a single class (Class IV) because their distributions are similar for a sample of jets, their jet-by-jet values are highly correlated, and they exhibit very similar power to separate quark jets and gluon jets (with very similar dependence on the jet parameters R and p_T); this separation power is not improved when they are combined. The fifth class (Class V) is composed of $C_1^{\beta=2}$, $\tau_1^{\beta=2}$ and the (ungroomed) jet mass. Again the jet-by-jet correlations are strong (even though the individual observable distributions are somewhat different), the quark versus gluon separation power is very similar (including the R and p_T dependence), and little is achieved by combining more than one of the Class V observables. This class structure is not surprising given that the observables within a class exhibit very similar dependence on the kinematics of the underlying jet constituents, and we provide more details below. For example, the members of Class V are constructed from of a sum over pairs of constituents using products of the energy of each member of the pair times the angular separation squared for the pair (this is apparent for the ungroomed mass when viewed in terms of a mass-squared with small angular separations). By the same argument, the Class IV and Class V observables will be seen to be more similar than any other pair of classes, differing only in the power (β) of the dependence on the angular separations, which produces small but detectable differences. We will return to a more complete discussion of jet masses in Section 5.4.

5.2 Single Variable Discrimination

The quark and gluon distributions of different substructure observables are shown in Figure 1 (in the $p_T = 500-600$ GeV bin and $R = 0.8$), and these illustrate some of the distinctions between the Classes made above. The fundamental difference between quarks and gluons, namely their color charge and consequent amount of radiation in the jet, is clearly indicated in Figure 1(a), suggesting that simply counting constituents provides good separation between quark and gluon jets. In fact, among the observables considered, one can see by eye that N_constits should provide the highest separation power, *i.e.*, the quark and gluon distributions are most distinct, as was originally noted in [49, 50]. Figure 1 further suggests that $C_1^{\beta=0}$ should provide the next best separation followed by $C_1^{\beta=1}$, as was also found by the CMS and ATLAS Collaborations [48, 51].

To more quantitatively study the power of each observable as a discriminator for quark/gluon tagging, Receiver Operating Characteristic (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 2 shows these ROC curves for all of the substructure variables shown in Figure 1 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. As suggested earlier,



(a) $\mathrm{N}_{\mathrm{constits}}$ (b) Γ_{Qjet} (c) $C_1^{\beta=0}$

(d) $C_1^{\beta=1}$ (e) $\tau_1^{\beta=1}$

(f) $C_1^{\beta=2}$ (g) $\tau_1^{\beta=2}$ (h) Ungroomed mass

Fig. 1 Comparisons of quark and gluon distributions of different substructure variables, organized by Class, for leading jets in the $p_T = 500-600$ GeV bin using the anti-k_{T} $R = 0.8$ algorithm. The first three plots are Classes I-III, with Class IV in the second row, and Class V in the third row.

n_{constits} is the best performing variable for all R values, although $C_1^{\beta=0}$ is not far behind, particularly for R=0.8. Most other variables have similar performance, with the main exception of Γ_{Qjet}, which shows significantly worse discrimination (which 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 [32, 33]). The combination of all variables shows somewhat better discrimination than any individual observable, and we give a more detailed discussion in Section 5.3 of the correlations between the observables and their impact on the combined discrimination power.

We now examine how the performance of the substructure observables varies with p_T and R. To present the results in a "digestible" fashion we focus on the gluon jet "rejection" factor, $1/\varepsilon_{\mathrm{bkg}}$, for a quark signal efficiency, $\varepsilon_{\mathrm{sig}}$, of

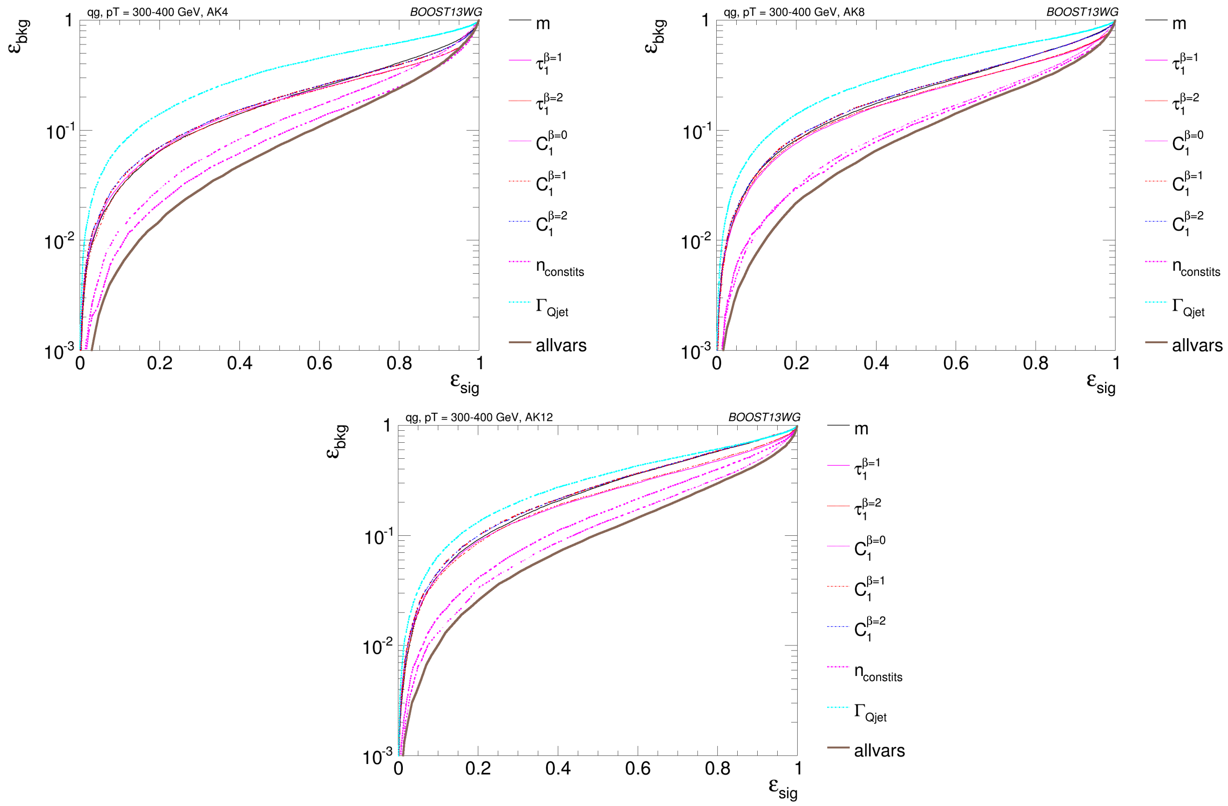


Fig. 2 The ROC curve for all single variables considered for quark-gluon discrimination in the p_T 300-400 GeV bin using the anti-k_{T} R=0.4 (top-left), 0.8 (top-right) and 1.2 (bottom) algorithm.

50 %. We can use the values of $1/\varepsilon_{\mathrm{bkg}}$ generated for the 9 kinematic points introduced above ($R = 0.4, 0.8, 1.2$ and the 100 GeV p_T bins with lower limits $p_T = 300\,\mathrm{GeV}$, $500\,\mathrm{GeV}$, $1000\,\mathrm{GeV}$) to generate surface plots. The surface plots in Figure 3 indicate both the level of gluon rejection and the variation with p_T and R for each of the studied single observable. We organize our results by class:

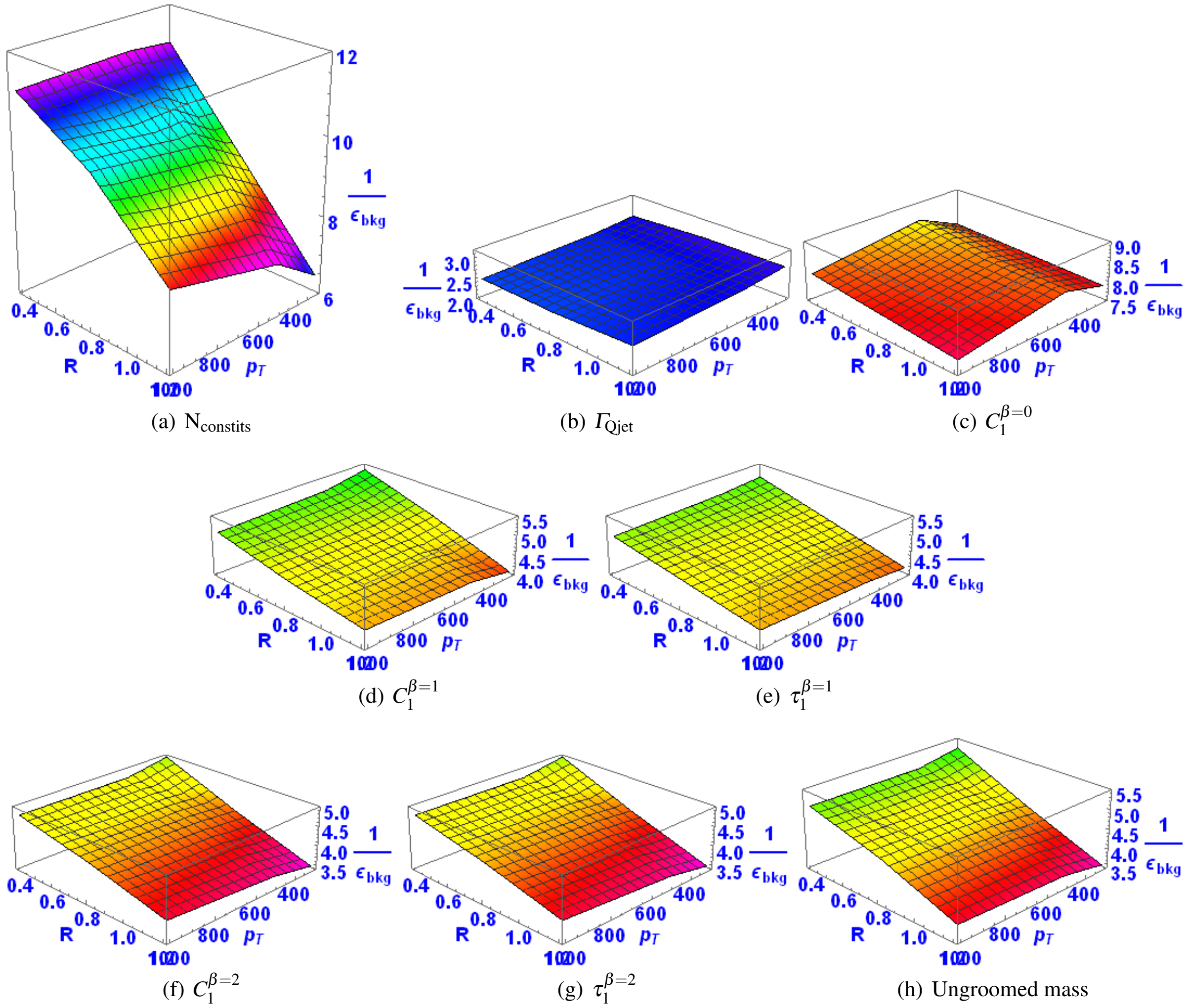
Class I: The sole constituent of this class is N_{constits}. We see in Figure 3(a) that, as expected, the numerically largest rejection rates occur for this observable, with the rejection factor ranging from 6 to 11 and varying rather dramatically with R. As R increases the jet collects more constituents from the underlying event, which are the same for quark and gluon jets, and the separation power decreases. At large R, there is some improvement with increasing p_T due to the enhanced QCD radiation, which is different for quarks vs. gluons.

Class II: The variable Γ_{Qjet} constitutes this class. Figure 3(b) confirms the limited efficacy of this single observable (at least for our parameter choices) with a rejection rate only in the range 2.5 to 2.8. On the other hand, this observable probes a very different property of jet substructure, *i.e.*, the sensitivity to detailed changes in the grooming procedure and this difference is suggested by the distinct R and p_T dependence illustrated in Figure 3(b). The rejection rate increases with increasing R and decreasing p_T, since the distinction between quark and gluon jets for this observable arises from the relative importance of the one "hard" gluon emission configuration. The role of this contribution is enhanced for both decreasing p_T and increasing R.

Class III: The only member of this class is $C_1^{\beta=0}$. Figure 3(c) indicates that this observable can itself provide a rejection rate in the range 7.8 to 8.6 (intermediate between the two previous observables), and again with distinct R and p_T dependence. In this case the rejection rate decreases slowly with increasing R, which follows from the fact that $\beta = 0$ implies no weighting of ΔR in the definition of $C_1^{\beta=0}$, greatly reducing the angular dependence. The rejection rate peaks at intermediate p_T values, an effect visually enhanced by the limited number of p_T values included.

Class IV: Figures 3(d) and (e) confirm the very similar properties of the observables $C_1^{\beta=1}$ and $\tau_1^{\beta=1}$ (as already suggested in Figures 1(d) and (e)). They have essentially identical rejection rates (4.1 to 5.4) and identical R and p_T dependence (a slow decrease with increasing R and an even slower increase with increasing p_T).

Class V: The observables $C_1^{\beta=2}$, $\tau_1^{\beta=2}$, and m have similar rejection rates in the range 3.5 to 5.3, as well as very similar R and p_T dependence (a slow decrease with increasing R and an even slower increase with increasing p_T).



(a) $\mathrm{N_{constits}}$ (b) Γ_{Qjet} (c) $C_1^{\beta=0}$

(d) $C_1^{\beta=1}$ (e) $\tau_1^{\beta=1}$

(f) $C_1^{\beta=2}$ (g) $\tau_1^{\beta=2}$ (h) Ungroomed mass

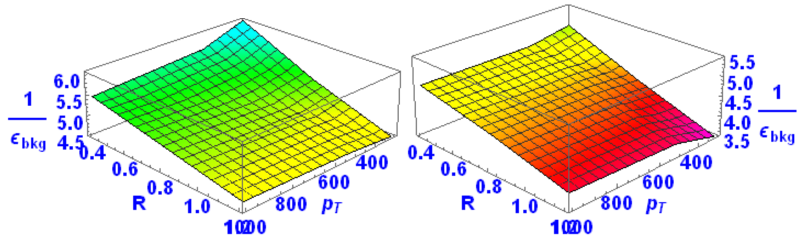
Fig. 3 Surface plots of $1/\varepsilon_{\text{bkg}}$ for all single variables considered for quark-gluon discrimination as functions of R and p_T. The first three plots are Classes I-III, with Class IV in the second row, and Class V in the third row.

Arguably, drawing a distinction between the Class IV and Class V observables is a fine point, but the color shading does suggest some distinction from the slightly smaller rejection rate in Class V. Again the strong similarities between the plots within the second and third rows in Figure 3 speaks to the common properties of the observables within the two classes.

In summary, the overall discriminating power between quark and gluon jets tends to decrease with increasing R, except for the Γ_{Qjet} observable, presumably in large part due to the contamination from the underlying event. Since the construction of the Γ_{Qjet} observable explicitly involves pruning away the soft, large angle constituents, it is not surprising that it exhibits different R dependence. In general the discriminating power increases slowly and monotonically with p_T (except for the Γ_{Qjet} and $C_1^{\beta=0}$ observables). This is presumably due to the overall increase in radiation from high p_T objects, which accentuates the differences in the quark and gluon color charges and providing some increase in discrimination. In the following section, we study the effect of combining multiple observables.

5.3 Combined Performance and Correlations

Combining multiple observables in a BDT can give further improvement over cuts on a single variable. Since the improvement from combining correlated observables is expected to be inferior to that from combining uncorrelated observables, studying the performance of multivariable combinations gives insight into the correlations between substructure variables and the physical features allowing for quark/gluon discrimination. Based on our discussion of the correlated properties of observables within a single class, we expect little improvement in the rejection rate when combining observables from the same class, and substantial improvement when combining observables from different classes. Our clas-



(a) $C_1^{\beta=1}+\tau_1^{\beta=1}$ (b) $C_1^{\beta=2}+\tau_1^{\beta=2}$

Fig. 4 Surface plots of $1/\varepsilon_{\rm bkg}$ for the indicated pairs of variables from (a) Class IV and (b) Class V considered for quark-gluon discrimination as functions of R and p_T.

sification of observables for quark/gluon tagging therefore motivates the study of particular combinations of variables for use in experimental analyses.

To quantitatively study the improvement obtained from multivariate analyses, we build quark/gluon taggers from every pair-wise combination of variables studied in the previous section; we also compare the pair-wise performance with the all-variables combination. To illustrate the results achieved in this way, we use the same 2D surface plots as in Figure 3. Figure 4 shows pair-wise plots for variables in (a) Class IV and (b) Class V, respectively. Comparing to the corresponding plots in Figure 3, we see that combining $C_1^{\beta=1}+\tau_1^{\beta=1}$ provides a small ($\sim 10\%$) improvement in the rejection rate with essentially no change in the R and p_T dependence, while combining $C_1^{\beta=2}+\tau_1^{\beta=2}$ yields a rejection rate that is essentially identical to the single observable rejection rate for all R and p_T values (with a similar conclusion if one of these observables is replaced with the ungroomed jet mass m). This confirms the expectation that the observables within a single class effectively probe the *same* jet properties.

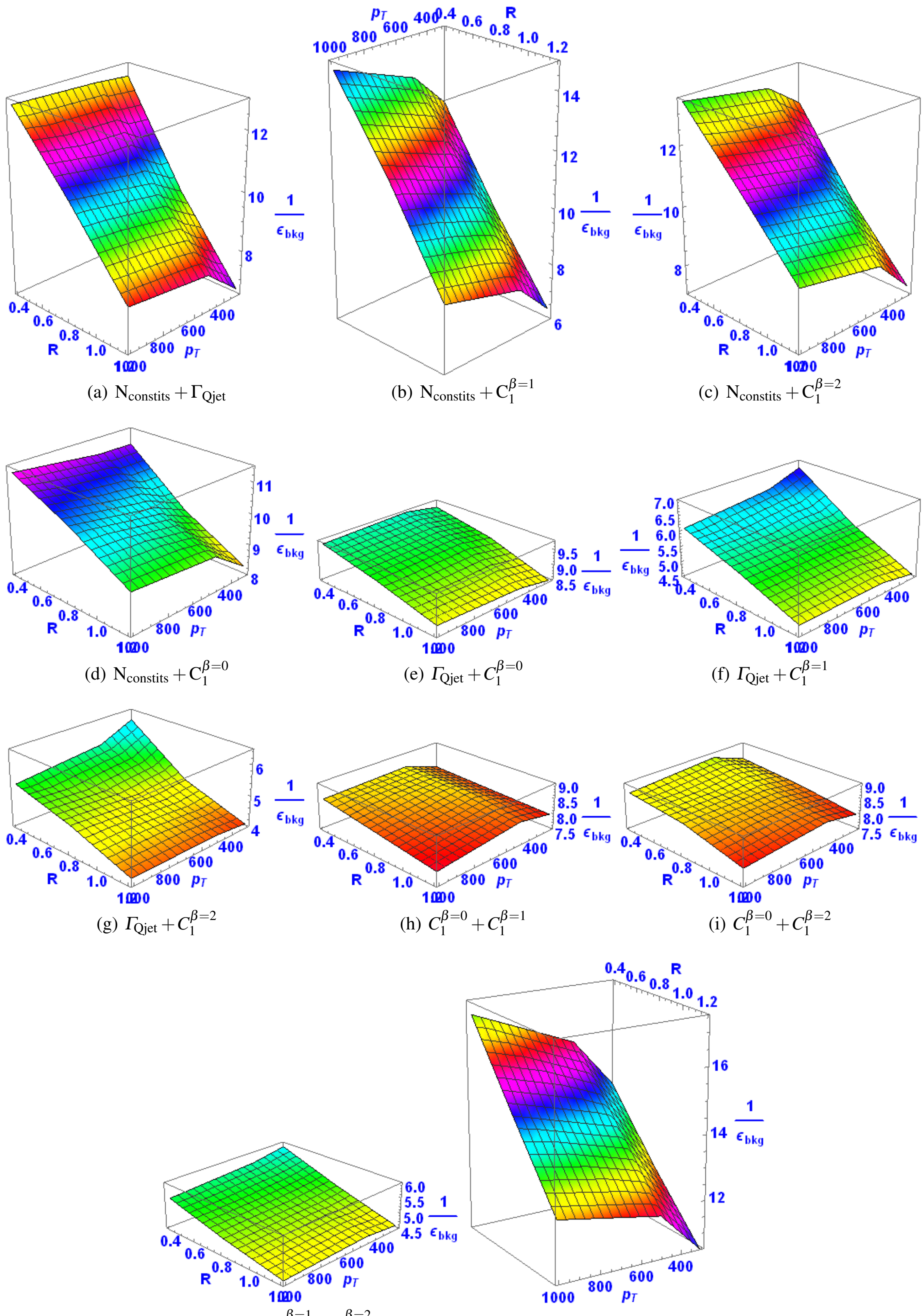
Next, we consider cross-class pairs of observables in Figure 5, where for each class we only use a single observable for illustrative purposes. Since $\mathrm{N_{constits}}$ is the best performing single variable, the largest rejection rates are obtained from combining another observable with $\mathrm{N_{constits}}$ (Figures 5(a) to (d)). In general, the rejection rates are larger for the pair-wise case than for the single variable case. In particular, the pair $\mathrm{N_{constits}}+\mathrm{C}_1^{\beta=1}$ yields rejection rates in the range 6.4 to 14.7 with the largest values at small R and large p_T. The other pairings with $\mathrm{N_{constits}}$ (except with $\tau_1^{\beta=1}$) yield smaller rejection rates and smaller dynamic range. The pair $\mathrm{N_{constits}}+\mathrm{C}_1^{\beta=0}$ (Figure 5(d)) exhibits the smallest range of rates (8.3 to 11.3), suggesting that the differences between these two observables serve to substantially reduce the R and p_T dependence for the pair, but this also reduces the possible optimization. The other pairs shown exhibit similar behavior.

The R and p_T dependence of the pair-wise combinations is generally similar to the single observable with the most dependence on R and p_T. The smallest R and p_T variation always occurs when pairing with $C_1^{\beta=0}$. Changing any of the observables in these pairs with a different observable in the same class (*e.g.*, $C_1^{\beta=2}$ for $\tau_1^{\beta=2}$) produces very similar results. Figure 5(k) shows the result of a BDT analysis including all of the current observables with rejection rates in the range 10.5 to 17.1. This is a somewhat narrower range than in Figure 5(b) but with larger maximum values.

Some features are more easily seen with an alternative presentation of the data: we fix R and p_T and simultaneously show the single- and pair-wise observables performance in a single matrix, and these matrices are shown in Figures 6 and 7. The numbers in each cell are the same rejection rate for gluons used earlier, $1/\varepsilon_{\rm bkg}$, with $\varepsilon_{\rm sig}=50\%$ (quarks). Figure 6 shows the results for $p_T = 1-1.1$ TeV and $R = 0.4, 0.8, 1.2$, while Figure 7 is for $R = 0.4$ and the 3 p_T bins. The single observable rejection rates appear on the diagonal, and the pairwise results are off the diagonal. The correlations indicated by the shading should be largely understood as indicating the organization of the observables into the now-familiar classes. The all-observable (BDT) result appears as the number at the lower right in each plot.

5.4 QCD Jet Masses

To close the discussion of q/g-tagging, we provide some insight into the behavior of the masses of QCD jets initiated by both kinds of patrons, with and without grooming. Recall that, in practice, an identified jet is simply a list of constituents, *i.e.*, final state particles. To the extent that the masses of these individual constituents can be neglected (due to the constituents being relativistic), each constituent has a "well- defined" 4-momentum from its energy and direction. It follows that the 4-momentum of the jet is simply the sum of the 4-momenta of the constituents and its square is the jet mass squared. Simply on dimensional grounds, we know that jet mass must have an overall linear scaling with p_T, with the remaining p_T dependence arising predominantly from the running of the coupling, $\alpha_s(p_T)$. The R dependence is also crudely linear as the jet mass scales ap-



(a) $\mathrm{N_{constits}} + \Gamma_{\mathrm{Qjet}}$ (b) $\mathrm{N_{constits}} + \mathrm{C}_1^{\beta=1}$ (c) $\mathrm{N_{constits}} + \mathrm{C}_1^{\beta=2}$

(d) $\mathrm{N_{constits}} + \mathrm{C}_1^{\beta=0}$ (e) $\Gamma_{\mathrm{Qjet}} + C_1^{\beta=0}$ (f) $\Gamma_{\mathrm{Qjet}} + C_1^{\beta=1}$

(g) $\Gamma_{\mathrm{Qjet}} + C_1^{\beta=2}$ (h) $C_1^{\beta=0} + C_1^{\beta=1}$ (i) $C_1^{\beta=0} + C_1^{\beta=2}$

(j) $C_1^{\beta=1} + C_1^{\beta=2}$ (k) All

Fig. 5 Surface plots of $1/\varepsilon_{\mathrm{bkg}}$ for the indicated pairs of variables from different classes considered for quark-gluon discrimination as functions of R and p_T.

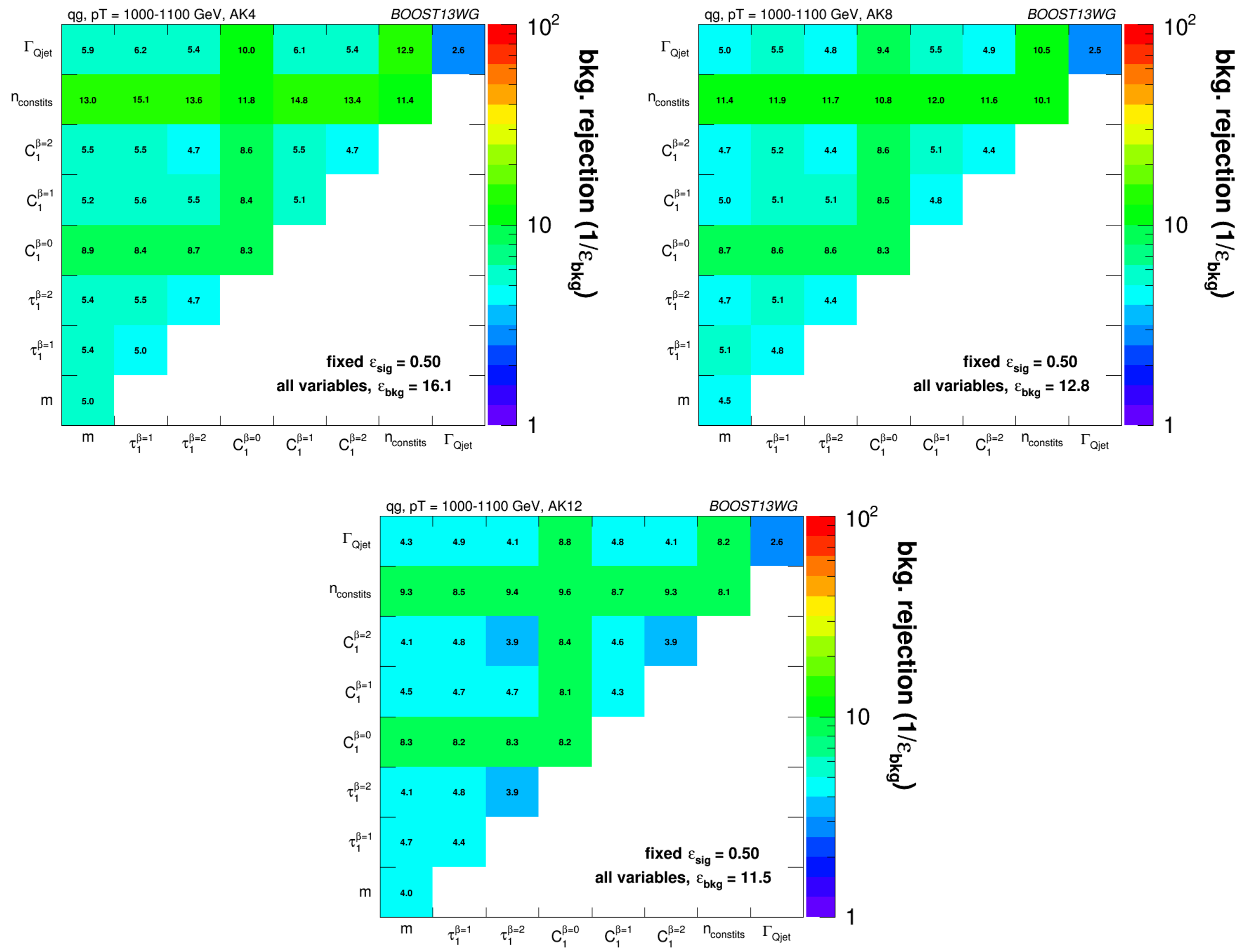


Fig. 6 Gluon rejection defined as $1/\varepsilon_{\text{gluon}}$ when using each 2-variable combination as a tagger with 50% acceptance for quark jets. Results are shown for jets with $p_T = 1-1.1$ TeV and for (top left) $R = 0.4$; (top right) $R = 0.8$; (bottom) $R = 1.2$. The rejection obtained with a tagger that uses all variables is also shown in the plots.

proximately with the largest angular opening between any 2 constituents, which is set by R.

To demonstrate this universal behavior for jet mass, we first note that if we consider the mass distributions for many kinematic points (various values of R and p_T), we observe considerable variation in behaviour. This variation, however, can largely be removed by plotting versus the scaled variable $m/p_T/R$. The mass distributions for quark and gluon jets versus $m/p_T/R$ for all of our kinematic points are shown in Figure 8, where we use a logarithmic scale on the y-axis to clearly exhibit the behavior of these distributions over a large dynamic range. We observe that the distributions for the different kinematic points do approximately scale as expected, *i.e.*, the simple arguments above capture most of the variation with R and p_T. We will consider shortly an explanation of the residual non-scaling. A more rigorous quantitative understanding of jet mass distributions requires all-orders calculations in QCD, which have been performed for ungroomed jet mass spectra at high logarithmic accuracy, both in the context of direct QCD resummation [52, 53] and Soft Collinear Effective Theory [54, 55].

Several features of Figure 8 can be easily understood. The distributions all cut off rapidly for $m/p_T/R > 0.5$, which is understood as the precise limit (maximum mass) for a jet composed of just 2 constituents. As expected from the soft and collinear singularities in QCD, the mass distribution peaks at small mass values. The actual peak is "pushed" away from the origin by the so-called Sudakov form factor. Summing the corresponding logarithmic structure (singular in both p_T and angle) to all orders in perturbation theory yields a distribution that is highly damped as the mass vanishes. In words, there is precisely *zero* probability that a color parton emits *no* radiation (and the resulting jet has zero mass). Above the Sudakov-suppressed part of phase space, there are two structures in the distribution: the "shoulder" and the "peak". The large mass shoulder ($0.3 < m/p_T/R < 0.5$) is driven largely by the presence of a single large angle, energetic emission in the underlying QCD shower, *i.e.*,

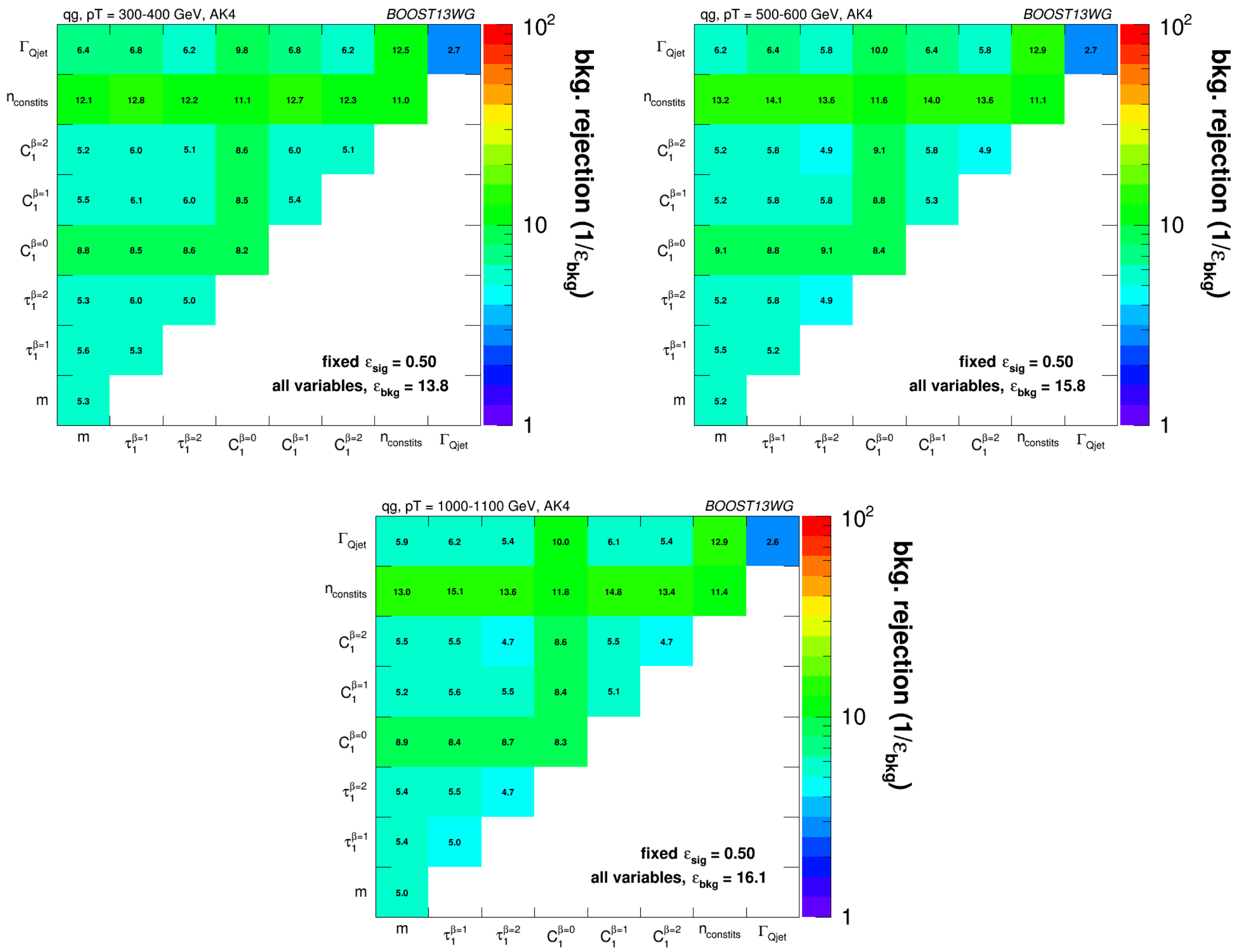
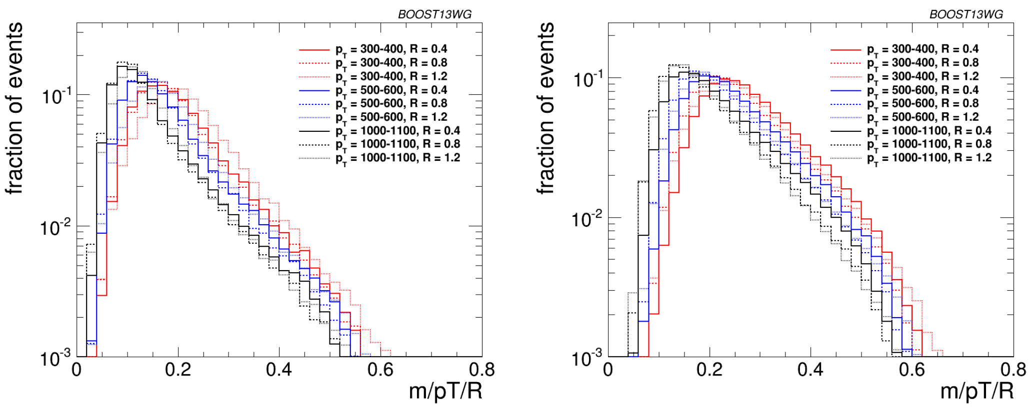


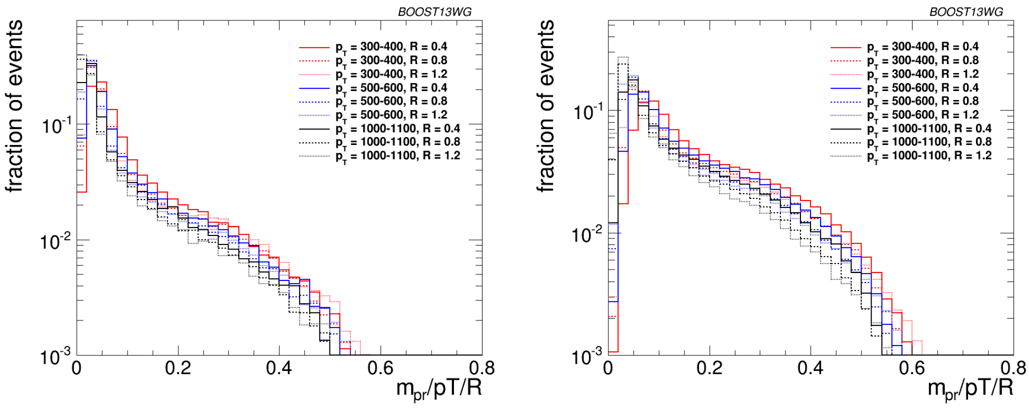
Fig. 7 Gluon rejection defined as $1/\varepsilon_{\text{gluon}}$ when using each 2-variable combination as a tagger with 50% acceptance for quark jets. Results are shown for R=0.4 jets with (top left) $p_T = 300-400$ GeV, (top right) $p_T = 500-600$ GeV and (bottom) $p_T = 1-1.1$ TeV. The rejection obtained with a tagger that uses all variables is also shown in the plots.



(a) Quark jets

(b) Gluon jets

Fig. 8 Comparisons of quark and gluon ungroomed mass distributions versus the scaled variable $m/p_T/R$.



(a) Quark jets (b) Gluon jets

Fig. 9 Comparisons of quark and gluon pruned mass distributions versus the scaled variable $m_{\mathrm{pr}}/p_T/R$.

this regime is quite well described by low-order perturbation theory[1] In contrast, we can think of the peak region as corresponding to multiple soft emissions. This simple, necessarily approximate picture provides an understanding of the bulk of the differences between the quark and gluon jet mass distributions. Since the probability of the single large angle, energetic emission is proportional to the color charge, the gluon distribution should be enhanced in this region by a factor of about $C_A/C_F = 9/4$, consistent with what is observed in Figure 8. Similarly the exponent in the Sudakov damping factor for the gluon jet mass distribution is enhanced by the same factor, leading to a peak "pushed" further from the origin. Therefore, the gluon jet mass distribution exhibits a larger average jet mass than the quark jets, with a larger relative contribution arising from the perturbative shoulder region.

Together with the fact that the number of constituents in the jet is also larger (on average) for the gluon jet simply because a gluon will radiate more than a quark, these features explain much of what we observed earlier in terms of the effectiveness of the various observable to separate quark jets from gluons jets. They also give us insight into the difference in the distributions for the observable Γ_{Qjet}. Since the shoulder is dominated by a single large angle *hard* emission, it is minimally impacted by pruning, which is designed to remove the large angle, *soft* constituents (as shown in more detail below). Thus, jets in the shoulder exhibit small volatility and they are a larger component in the gluon jet distribution. Hence gluon jets, on average, have smaller values of Γ_{Qjet} than quark jets as in Figure 1(b). Further, this feature of gluon jets is distinct from the fact that there are more constituents, explaining why Γ_{Qjet} and $\mathrm{N}_{\mathrm{constits}}$ supply largely independent information for distinguishing quark and gluon jets.

[1]The shoulder label will become more clear when examining groomed jet mass distributions.

To illustrate some of these points in more detail, Figure 9 exhibits the same jet mass distributions *after pruning* [34, 56]. Removing the large angle, soft constituents moves the peak in both of the distributions from $m/p_T/R \sim 0.1-0.2$ to the region around $m/p_T/R \sim 0.05$. This explains why pruning works to reduce the QCD background when looking for a signal in a specific jet mass bin. The shoulder feature at higher mass is much more apparent after pruning, as is the larger shoulder for the gluon jets. A quantitative (all-orders) understanding of groomed mass distributions is also possible. For instance, resummation of the pruned mass distribution was achieved in [38, 57]. Figure 9 serves to confirm the physical understanding of the relative behavior of Γ_{Qjet} for quark and gluon jets.

Our final topic in this section is the residual R and p_T dependence exhibited in Figures 8 and 9, which deviates from the linear scaling removed with the variable $m/p_T/R$. As already suggested, the residual p_T dependence can be understood as arising primarily from the slow decrease of the strong coupling $\alpha_s(p_T)$ as p_T increases. This leads to a corresponding decrease in the (largely perturbative) shoulder regime for both distributions asat higher p_T. At the same time, and for the same reason, the Sudakov damping is less strong with increasing p_T and the peak moves in towards the origin. Thus the overall impact of increasing p_T for both distributions is a (gradual) shift to smaller values of $m/p_T/R$. This is just what is observed in Figures 8 and 9, although the numerical size of the effect is reduced in the pruned case.

The R dependence is more complicated as there are effectively three different contributions to the mass distribution. The perturbative large angle, energetic single emission contribution largely scales in the variable $m/p_T/R$, which is why we see little residual R dependence in either figure at higher masses ($m/p_T/R > 0.4$). The large angle, soft emissions can contribute in two ways: by contributing to mass values that scale like R, and by increasing the number of large angle, soft emissions included in the jet as R increases (*i.e.*, as the area of the jet grows as R^2). Such contributions yield a distribution that shifts to the right with increasing R and presumably explain the behavior at small p_T in Figure 8. Since pruning largely removes this contribution, we observe no such behavior in Figure 9. The contribution of small angle, soft emissions will be at fixed m values and thus shift to the left versus the scaled variable as R increases. This presumably explains the small shifts in this direction observed in both figures.

5.5 Conclusions

In Section 5 we have seen that a variety of jet observables provide information about the jet that can be employed to effectively separate quark-initiated from gluon-initiated jets. Further, when used in combination, these observables can provide superior separation; since the improvement depends on the correlation between observables, we use the multivariable performance to separate the observables into different classes, with each class containing highly correlated observables. We saw that the best performing single observable is simply the number of constituents in the jet, N_{constits}, while the largest further improvement comes from combining with $C_1^{\beta=1}$ (or $\tau_1^{\beta=1}$), but the smallest R and p_T dependence arises from combining with $C_1^{\beta=0}$. On the other hand, some of the commonly used observables are highly correlated and do not provide extra information and enhanced tagging when used together. In addition to demonstrating these correlations, we have provided a discussion of the physics behind the structure of the correlation. Using the jet mass as an example, we have given arguments to explicitly explain the differences between jet observables initiated by each type of parton.

Finally, we remind the reader that the numerical results were derived for a particular color configuration (qq and gg events), in a particular implementation of the parton shower and hadronization. Color connections in more complex event configurations, or different Monte Carlo programs, may well exhibit somewhat different efficiencies and rejection factors. The value of our results is that they indicate a subset of variables expected to be rich in information about the partonic origin of final-state jets. These variables can be expected to act as valuable discriminants in searches for new physics, and could also be used to define model-independent final-state measurements which would nevertheless be sensitive to the short-distance physics of quark and gluon production.

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. 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 are not expected to be substantially different for quark backgrounds; we will see that the differences in the substructure properties of quark- and gluon-initiated jets, explored in the last section, are significantly smaller than the differences between W-initiated and gluon-initiated jets.

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}).
- 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. Since we have not attempted to optimise the grooming parameter settings of each grooming algorithm, we do not want to place too much emphasis here on the relative performance of the groomed masses, but instead concentrate on how their performance changes depending on the kinematic bin and jet radius considered.

Figure 10 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 11 compares signal and background in the different substructure variables explored for the same jet radius and kinematic bin.

Figures 12, 13 and 14 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.

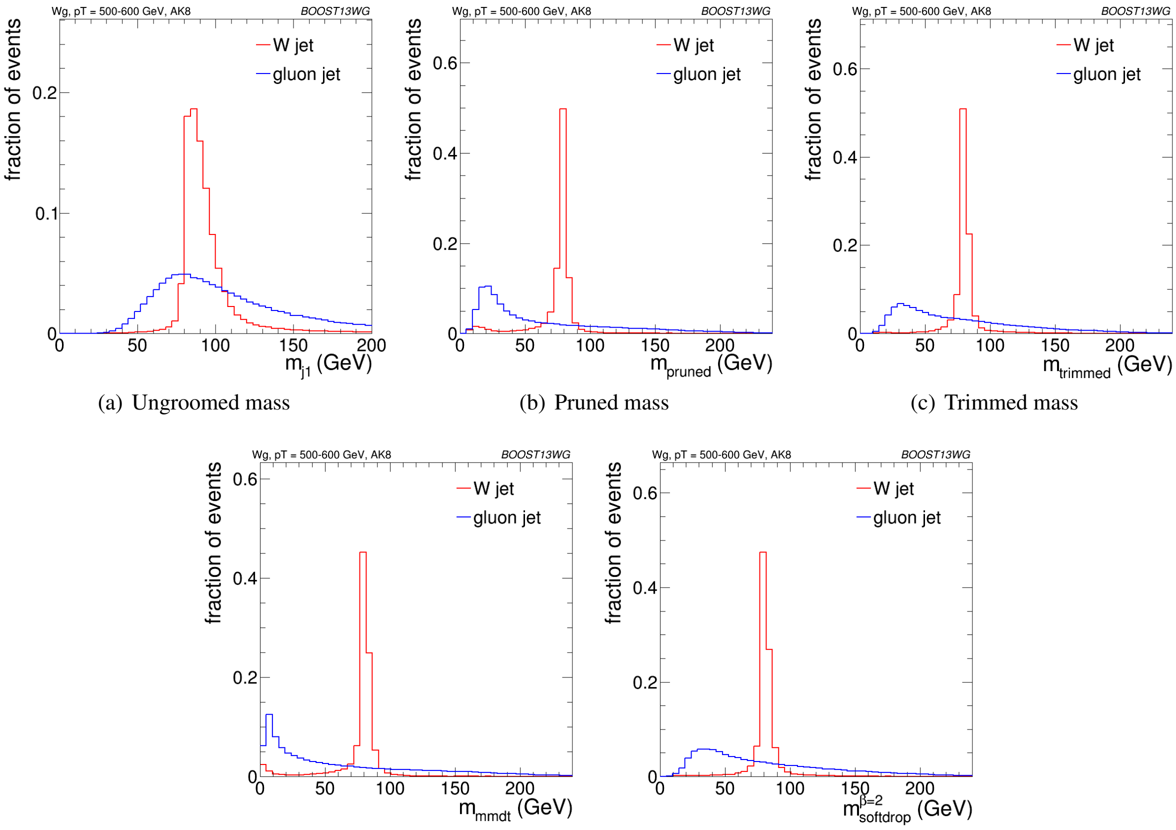
Although the ROC curves give all the relevant information, it is hard to compare performance quantitatively. In Figures 15, 16 and 17 are shown matrices which give the background rejection for a signal efficiency of 70% when two variables (that on the x-axis and that on the y-axis) are combined in a 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 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 15(a), 16(a) and 17(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 15(b), 16(b) and 17(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. In Figure 18 we show the Soft-drop $\beta=2$ groomed mass and the pruned mass for signal and background in the p_T 300-400 and p_T 1.0-1.1 TeV bins for R=1.2 jets. Two effects result in the improved performance of the groomed mass at high p_T. Firstly, as is evident from the figure, the resolution of the signal peak after grooming improves, because the groomer finds it easier to pick out the hard signal component of the jet against the softer components of the underlying event when the signal is boosted. Secondly, one can see from Figure 9 that as p_T increases the perturbative shoulder of the gluon distribution decreases in size, as discussed in Section 5.4, and thus there is a slight decrease (or at least no increase) in the level of background in the signal mass region (m/p_T/R $\sim$ 0.5).

However, one can see from the Figures 15(b), 16(b) and 17(c) that 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. In Figure 19 we show the $\tau_{21}^{\beta=1}$ and $C_2^{\beta=1}$ distributions for signal and background in the p_T 300-400 and p_T 1.0-1.1 TeV bins for R=0.8 jets. For $\tau_{21}^{\beta=1}$ one can see that in moving from the lower to the higher p_T bin, the signal peak remains fairly unchanged, whereas the background peak shifts to smaller $\tau_{21}^{\beta=1}$ values, reducing the discrimination power of the variable. This is expected, since jet substructure methods explicitly relying on identifying hard prongs would expect to work better at low p_T, where the prongs would tend to be more separated. However, $C_2^{\beta=1}$ does not rely on the explicit identification of subjets, and one can see from Figure 19 that the discrimination power visibly increases with increasing p_T. This is in line with the observation in [42] that $C_2^{\beta=1}$ performs best when m/p_T is small.

By comparing the individual sub-figures of Figures 15, 16 and 17 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. Figure 20 shows how the groomed mass changes for varying jet radius in the p_T 1.0-1.1 TeV bin. One can see that the signal mass peak remains unaffected by the increased radius, as expected, since grooming removes the soft contamination which could otherwise increase the mass of the jet as the radius increased. The gluon background in the signal mass region also remains largely unaffected, as expected from Figure 9, which shows very little dependence of the



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

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

Fig. 10 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.

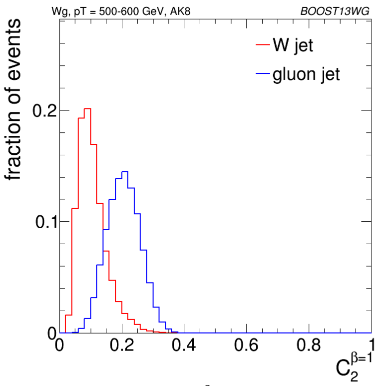
groomed gluon mass distribution on R in the signal region ($\mathrm{m}/p_T/\mathrm{R} \sim 0.5$). This is discussed further in Section 5.4.

However, we again see rather different behaviour versus R 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 (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. Figure 21 shows the $\tau_{21}^{\beta=1}$ and $C_2^{\beta=1}$ distributions for signal and background in the 1.0-1.1 TeV p_T bin for R=0.8 and R=1.2 jet radii. One can clearly see that for the larger jet radius the $C_2^{\beta=1}$ distribution of both signal and background get wider, and consequently the discrimination power decreases. For $\tau_{21}^{\beta=1}$ there is comparitively little change in the distributions with increasing jet radius. The increased sensitivity of C_2 to soft wide angle radiation in comparison to τ_{21} is a known feature of this variable [42], and a useful feature in discriminating coloured versus colour singlet jets. However, at very large jet radii (R$\sim$1.2), this feature becomes disadvantageous; the jet can pick up a significant amount of initial state or other uncorrelated radiation, and C_2 is more sensitive to this than is τ_{21}. This uncorrelated radiation has no (or very little) dependence on whether the jet is W- or gluon-initiated, and so sensitivity to this radiation means that the discrimination power will decrease.

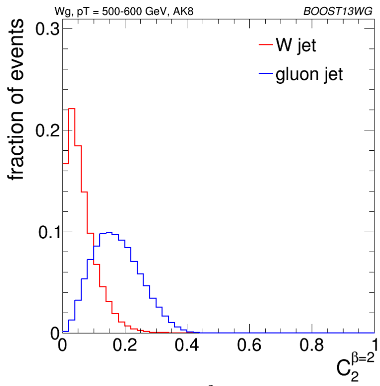
6.3 Combined Performance

The off-diagonal entries in Figures 15, 16 and 17 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.

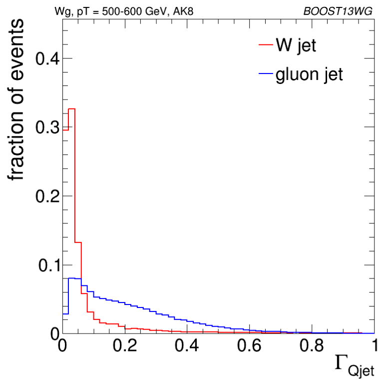
One can see that in general the most powerful two-variable combinations involve a groomed mass and a non-mass substructure variable ($C_2^{\beta=1}$, Γ_{Qjet} or $\tau_{21}^{\beta=1}$). Two-variable combinations of the substructure variables are not powerful in



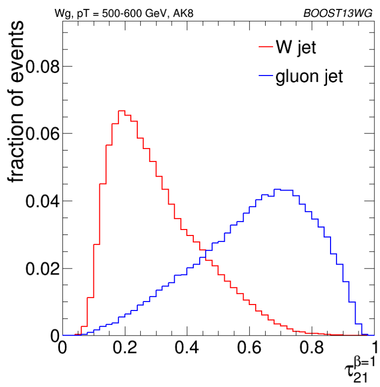
(a) $C_2^{\beta=1}$



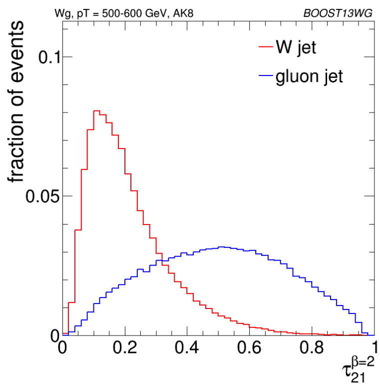
(b) $C_2^{\beta=2}$



(c) Γ_{Qjet}

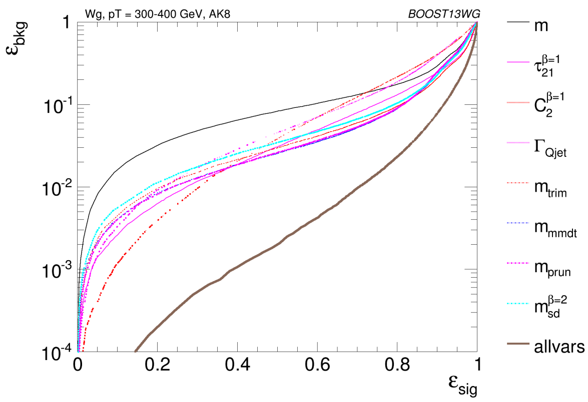


(d) $\tau_{21}^{\beta=1}$

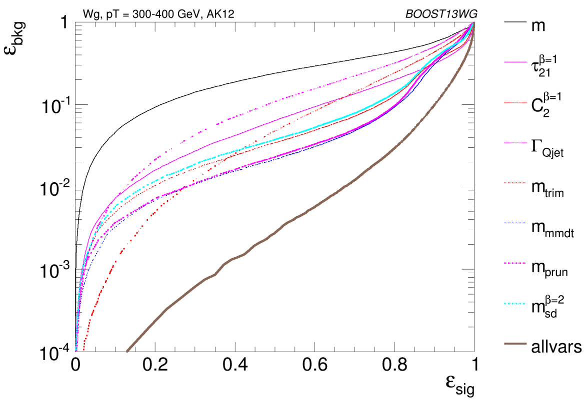


(e) $\tau_{21}^{\beta=2}$

Fig. 11 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.

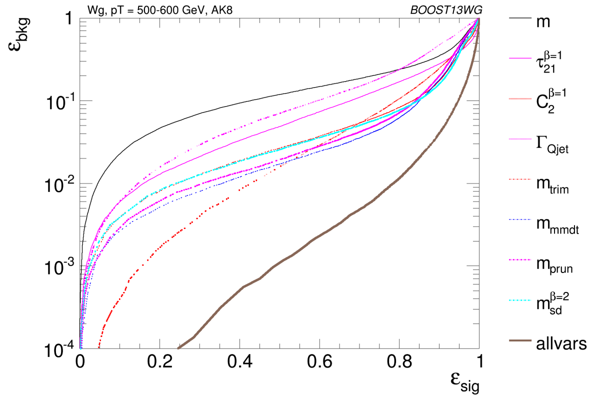


(a) anti-k_T R=0.8, p_T 300-400 GeV bin

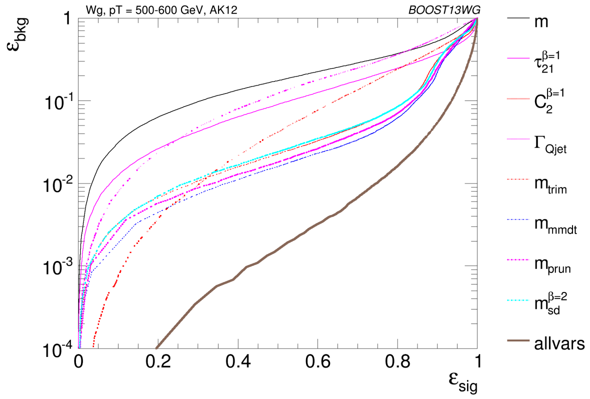


(b) anti-k_T R=1.2, p_T 300-400 GeV bin

Fig. 12 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.

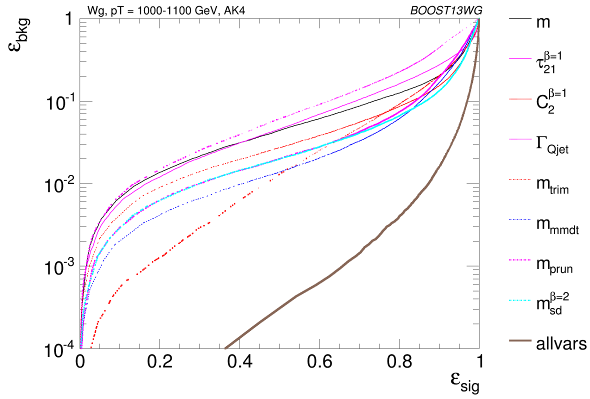


(a) anti-k_T R=0.8, p_T 500-600 GeV bin

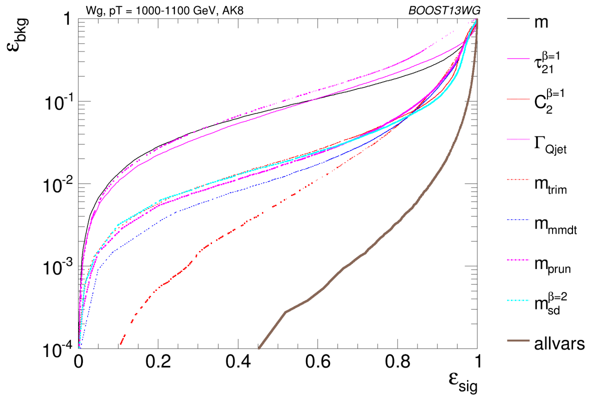


(b) anti-k_T R=1.2, p_T 500-600 GeV bin

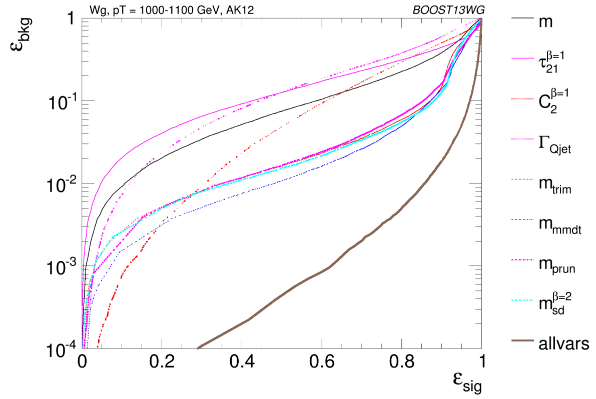
Fig. 13 The ROC curve for all single variables considered for W tagging in the p_T 500-600 GeV bin using the anti-k_T R=0.8 algorithm and R=1.2 algorithm.



(a) anti-k_T R=0.4, p_T 1.0-1.1 TeV bin

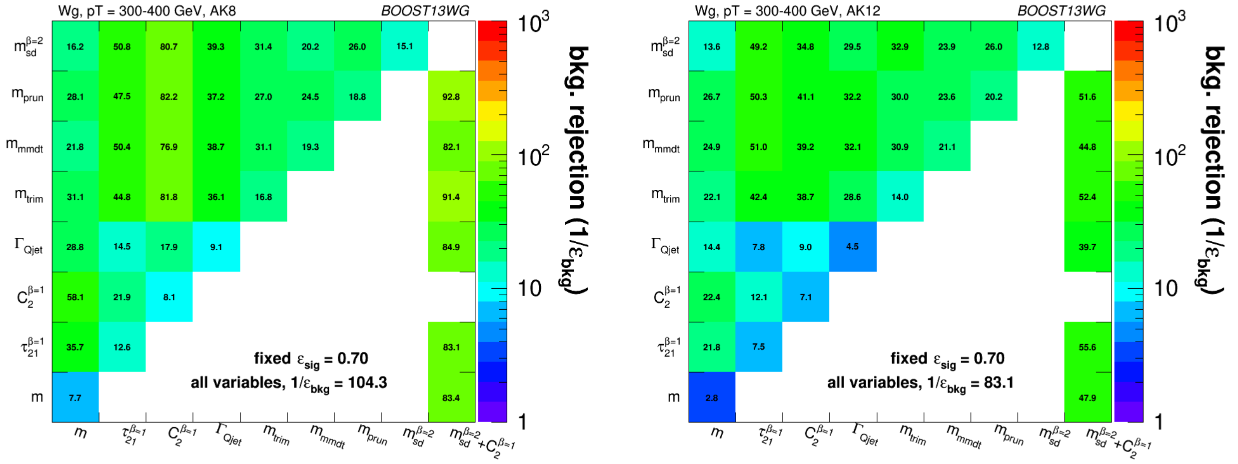


(b) anti-k_T R=0.8, p_T 1.0-1.1 TeV bin



(c) anti-k_T R=1.2, p_T 1.0-1.1 TeV bin

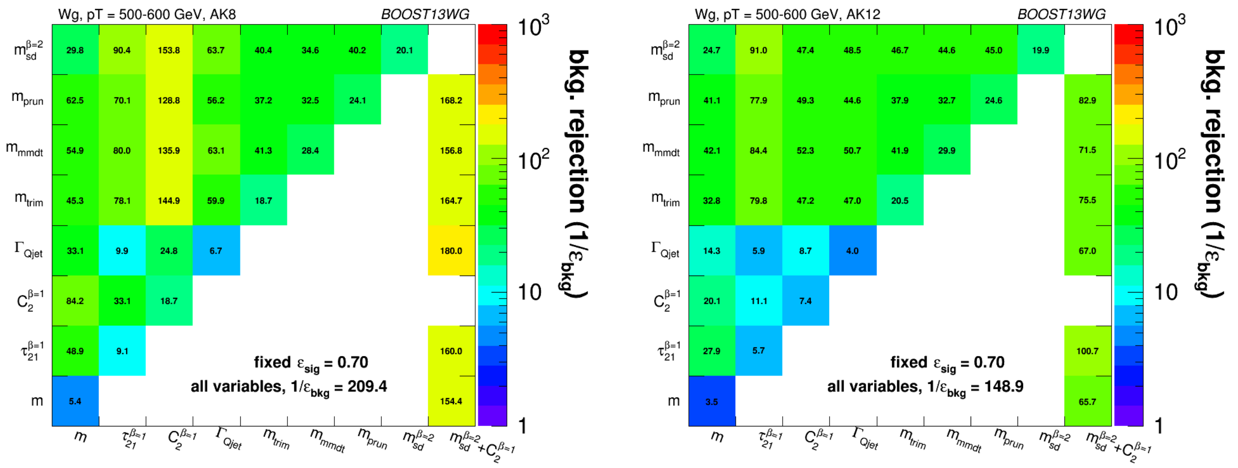
Fig. 14 The ROC curve for all single variables considered for W tagging in the p_T 1.0-1.1 TeV bin using the anti-k_T R=0.4 algorithm, anti-k_T R=0.8 algorithm and R=1.2 algorithm.



(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. 15 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

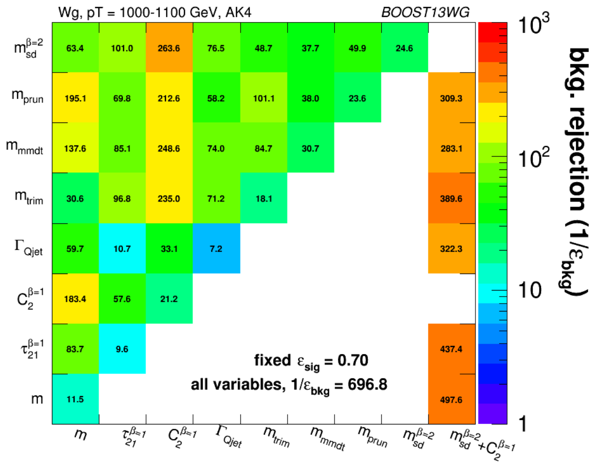
(b) anti-k_T R=1.2, p_T 500-600 GeV bin

Fig. 16 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.

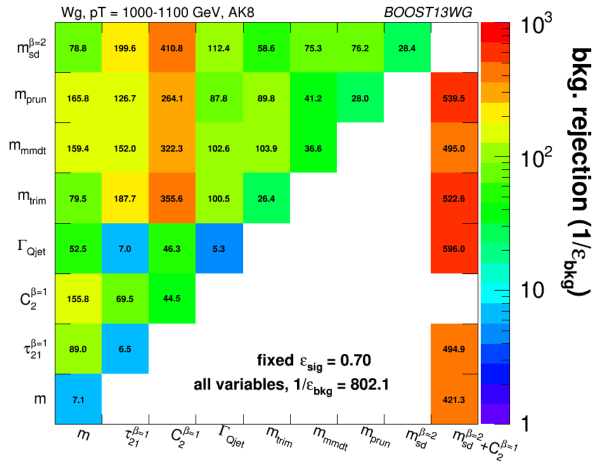
comparison. Which particular mass + substructure variable combination is the most powerful depends strongly on the p_T and R of the jet, as discussed in the sections that follow.

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.

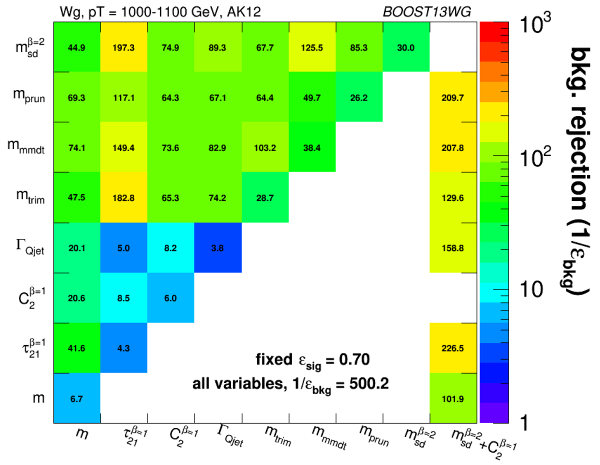
Generally one can see that the R=0.8 jets offer the best two-variable combined performance in all p_T bins explored here. This is despite the fact that in the highest 1.0-1.1 GeV p_T bin the average separation of the quarks from the W decay is much smaller than 0.8, and well within 0.4. This conclusion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



(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

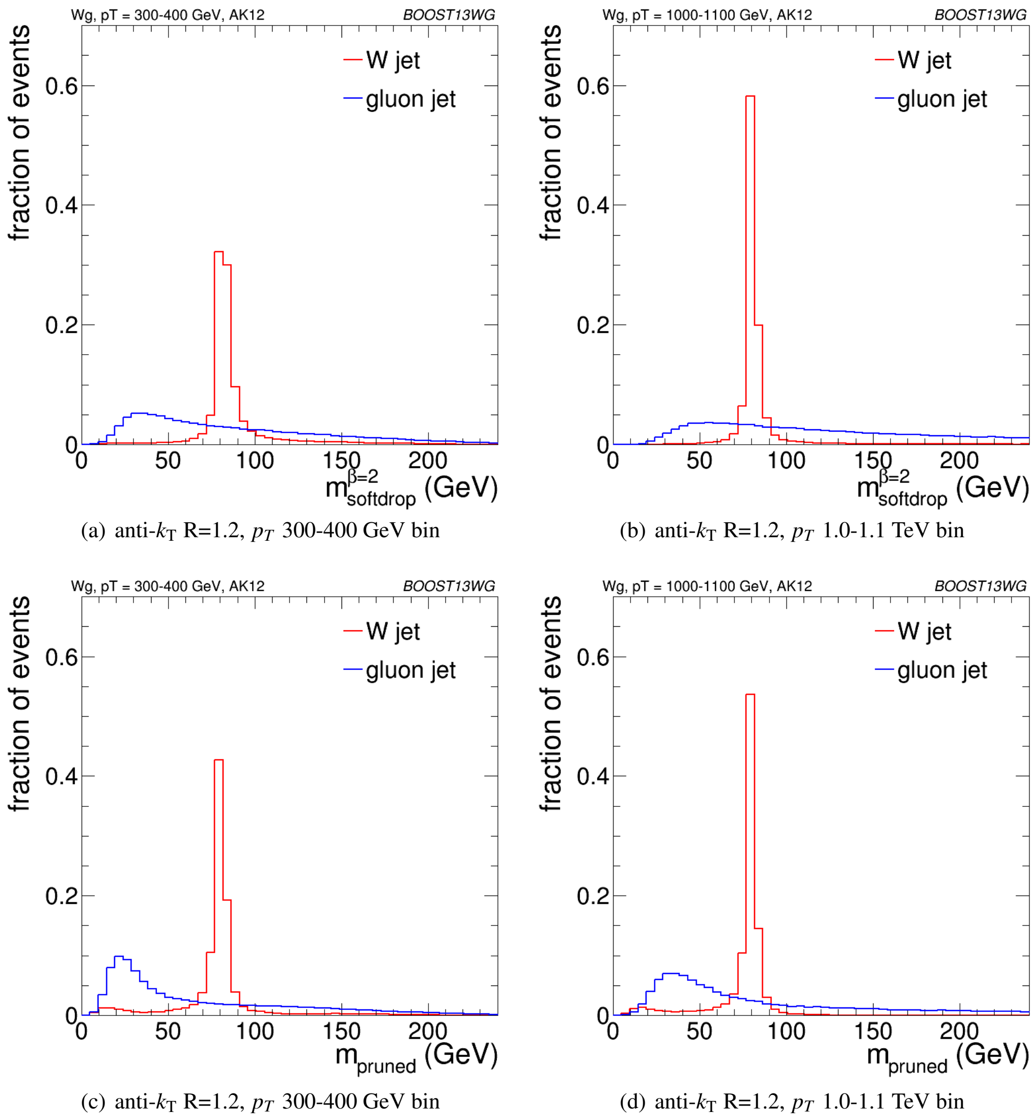
Fig. 17 The background rejection for a fixed signal efficiency (70%) of each BDT combination of each pair of variables considered, in the p_T 1.0-1.1 TeV bin using the anti-k_{T} R=0.4, R=0.8 and R=1.2 algorithm. Also shown is the background rejection for a BDT combination of all of the variables considered.

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 22 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 23 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 24 is the equivalent set of distributions for $m_{sd}^{\beta=2}$ and $\tau_{21}^{\beta=1}$. One can see from Figure 23 that, due to the sensitivity of the observable to to soft, wide-



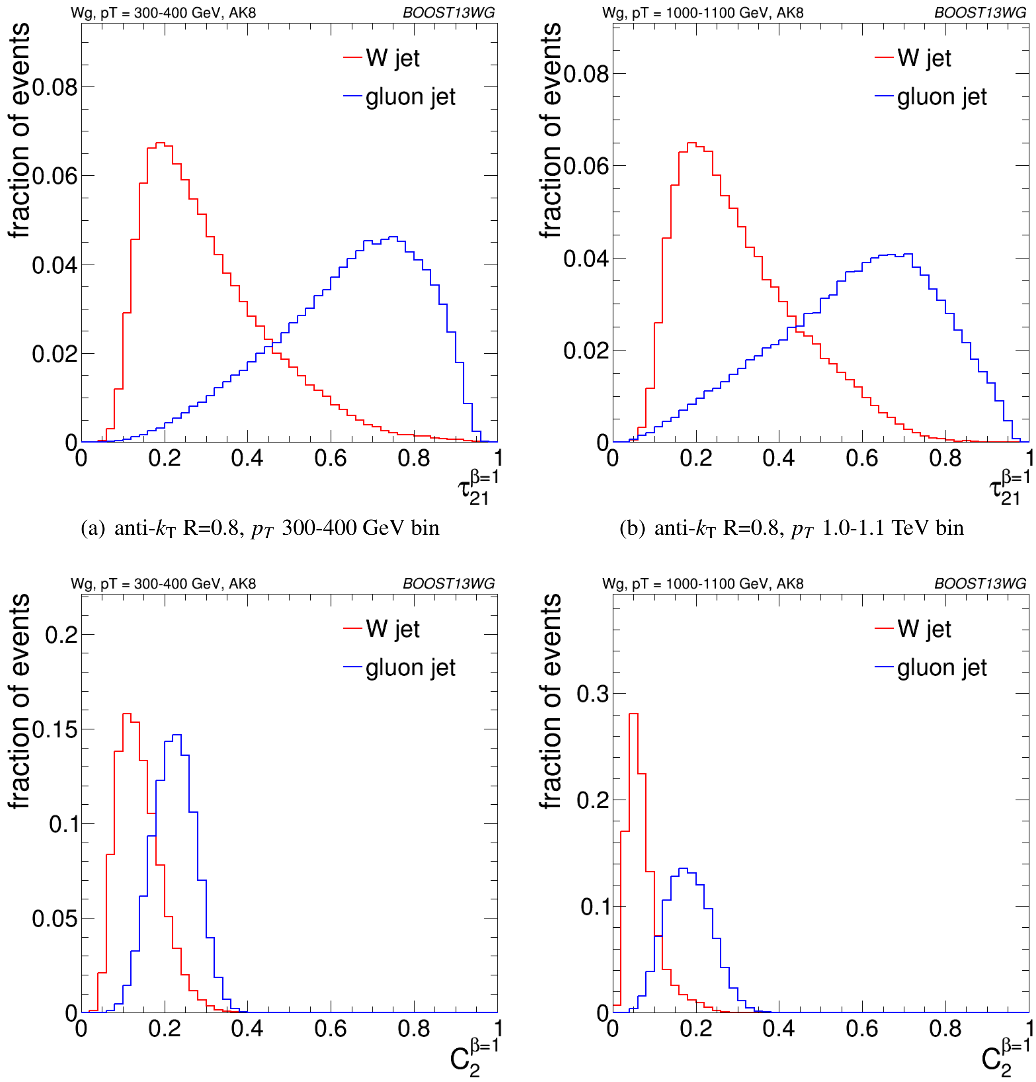
(a) anti-k_{T} R=1.2, p_T 300-400 GeV bin
(b) anti-k_{T} R=1.2, p_T 1.0-1.1 TeV bin
(c) anti-k_{T} R=1.2, p_T 300-400 GeV bin
(d) anti-k_{T} R=1.2, p_T 1.0-1.1 TeV bin

Fig. 18 The Soft-drop $\beta = 2$ and pruned groomed mass distribution for signal and background R=1.2 jets in two different p_T bins.

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 24 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 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 17 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 $\sim$two improvement for R=1.2 jets. A similar behaviour can be seen for mMDT mass. In Figures 25, 26 and 27 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 mass with the ungroomed mass, a combination that does not



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

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

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

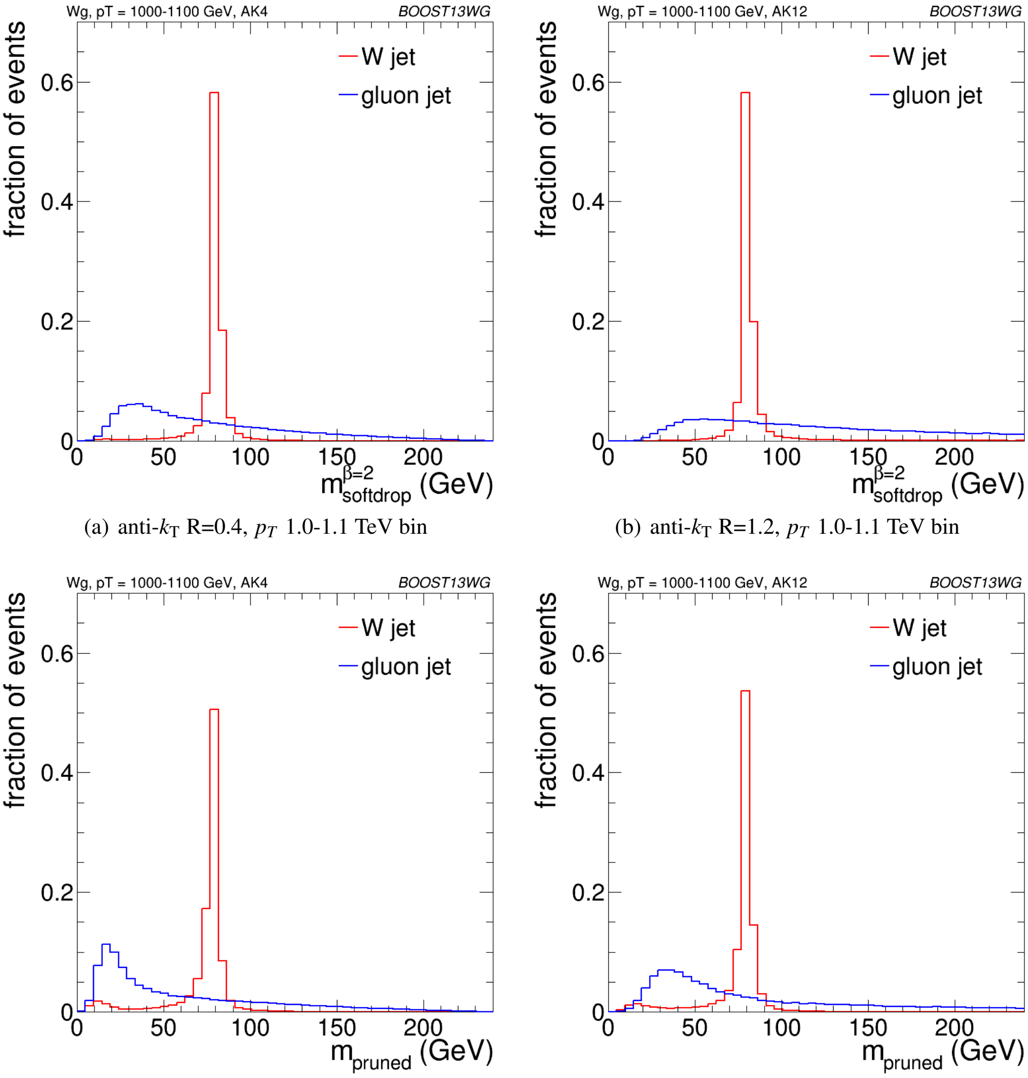
(d) anti-k_{T} R=0.8, p_T 1.0-1.1 TeV bin

Fig. 19 The $\tau_{21}^{\beta=1}$ and $C_2^{\beta=1}$ distributions for signal and background R=0.8 jets in two different p_T bins.

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 25, 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. $\sim$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 15, 16 and 17 also report the background rejection achieved by a combination of all the variables considered into a single BDT discriminant. One can see that, in all cases, the re-



(a) anti-k_T R=0.4, p_T 1.0-1.1 TeV bin

(b) anti-k_T R=1.2, p_T 1.0-1.1 TeV bin

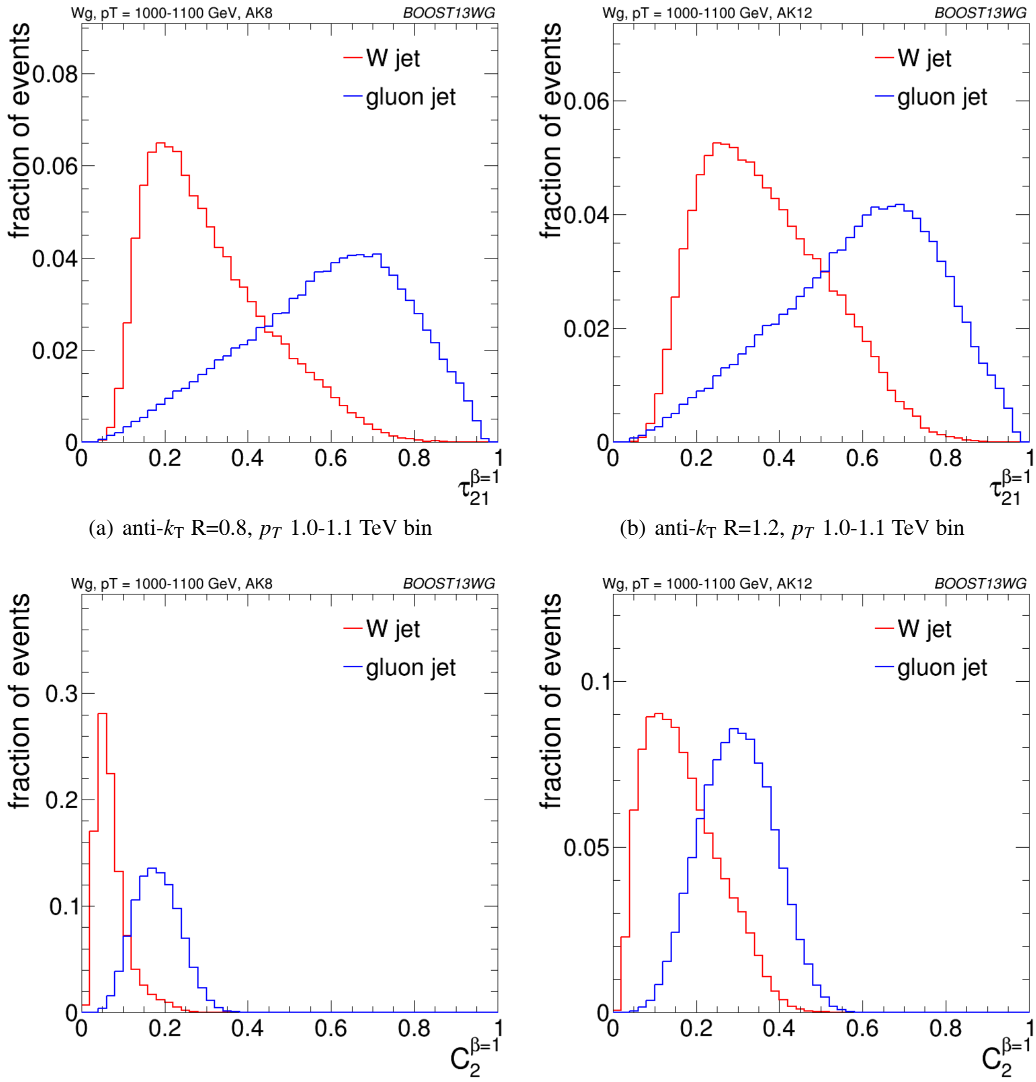
(c) anti-k_T R=0.4, p_T 1.0-1.1 TeV bin

(d) anti-k_T R=1.2, p_T 1.0-1.1 TeV bin

Fig. 20 The Soft-drop $\beta = 2$ and pruned groomed mass distribution for signal and background R=0.4 and R=1.2 jets in the 1.0-1.1 TeV p_T bin.

jection 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 15, 16 and 17 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) anti-k_{T} R=0.8, p_T 1.0-1.1 TeV bin

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

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

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

Fig. 21 The $\tau_{21}^{\beta=1}$ and $C_2^{\beta=1}$ distributions for signal and background R=0.8 and R=1.2 jets in the 1.0-1.1 TeV p_T bin.

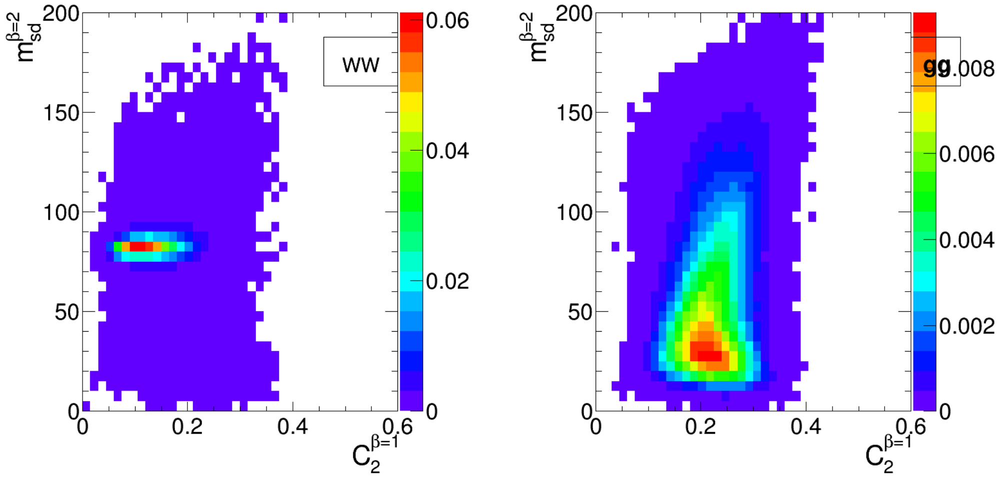
to the $m_{sd}^{\beta=2}+C_2^{\beta=1}$ combination, it is hard to see an obvious pattern; the best third variable changes depending on the p_T and R considered.

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.

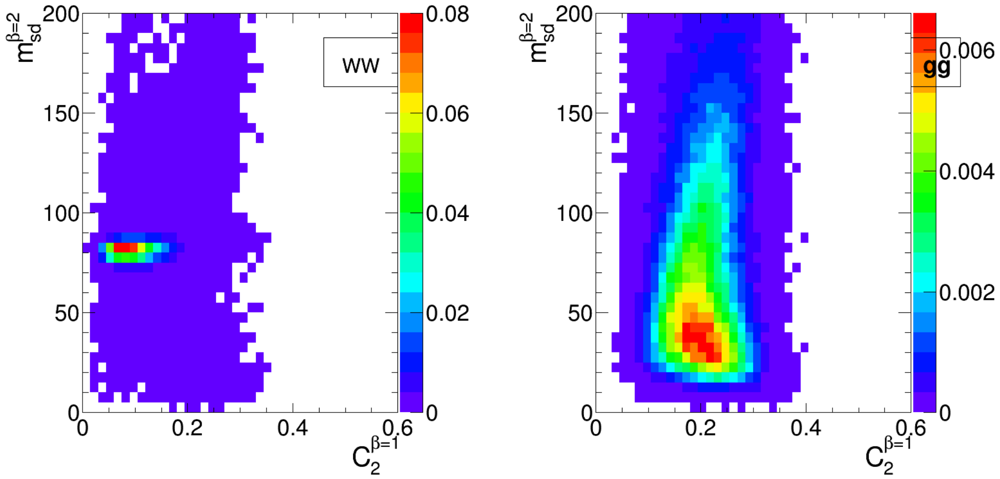
6.4 Conclusions

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.

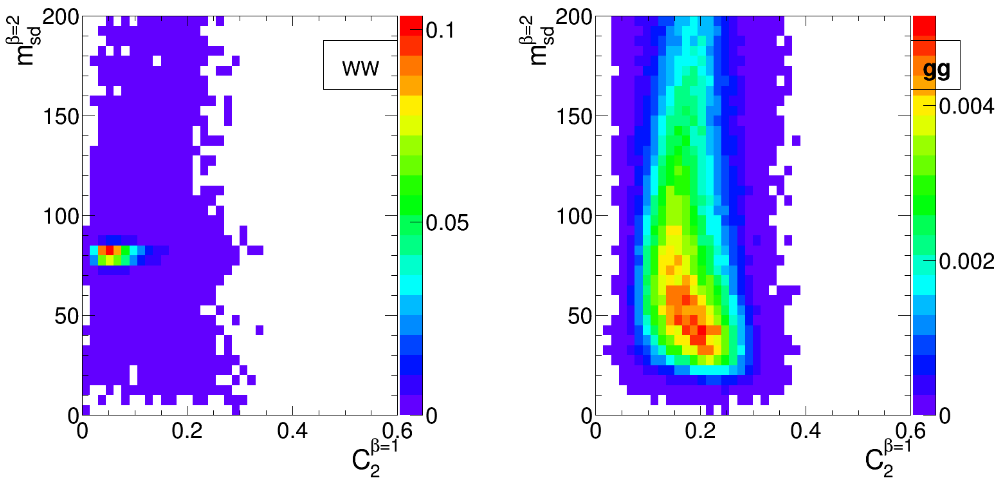
In terms of the performance of individual variables, we find that, in agreement with other studies [58], in general the groomed masses perform best, with a background rejection power that increases with increasing p_T , but which is more constant with respect to changes in R. We have explained the dependence of the groomed mass performance on p_T and R using the understanding of the QCD mass distribution gleaned in Section 5.4. 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



(a) p_T 300-400 GeV

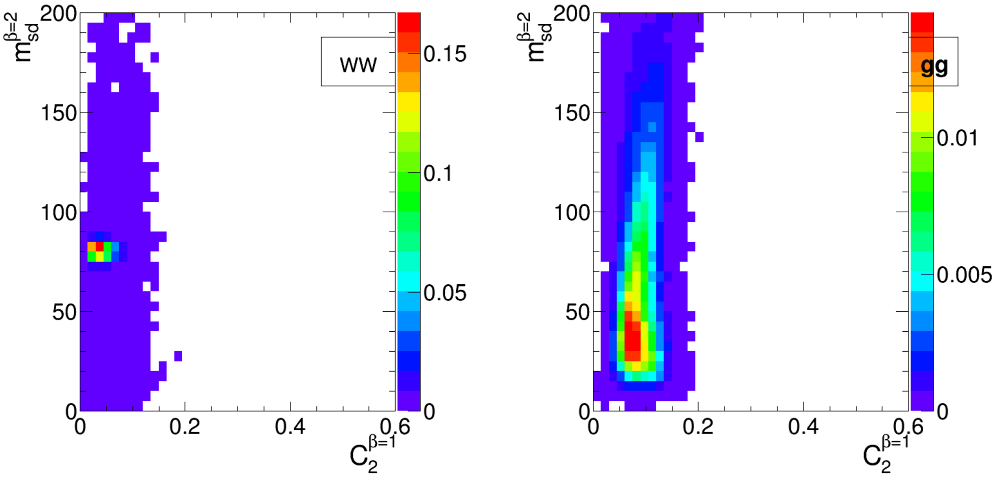


(b) p_T 500-600 GeV

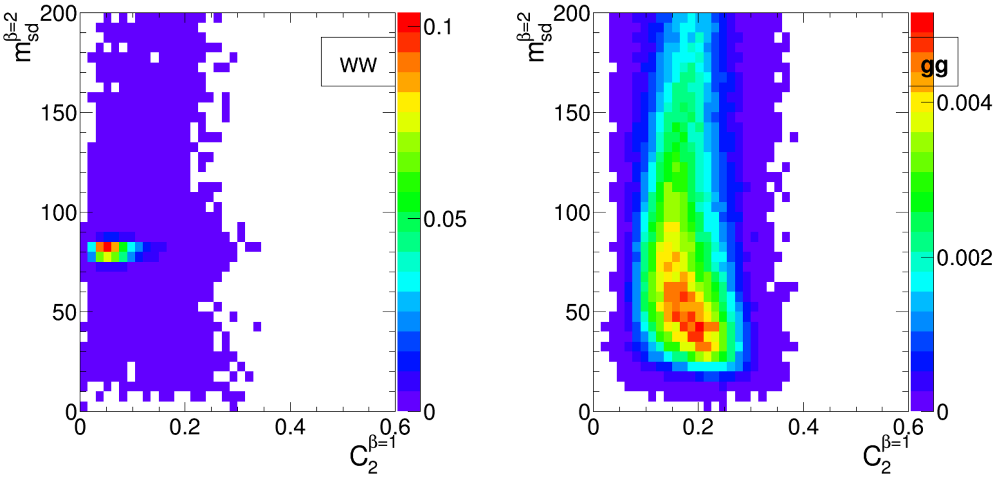


(c) p_T 1.0-1.1 TeV

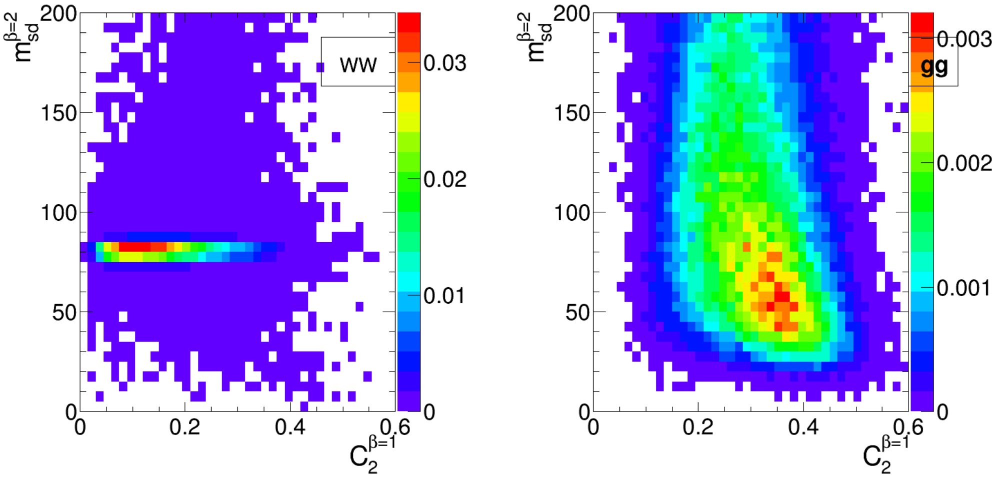
Fig. 22 2-D plots showing $m_{sd}^{\beta=2}$ versus $C_2^{\beta=1}$ for R=0.8 jets in the various p_T bins considered.



(a) p_T R=0.4

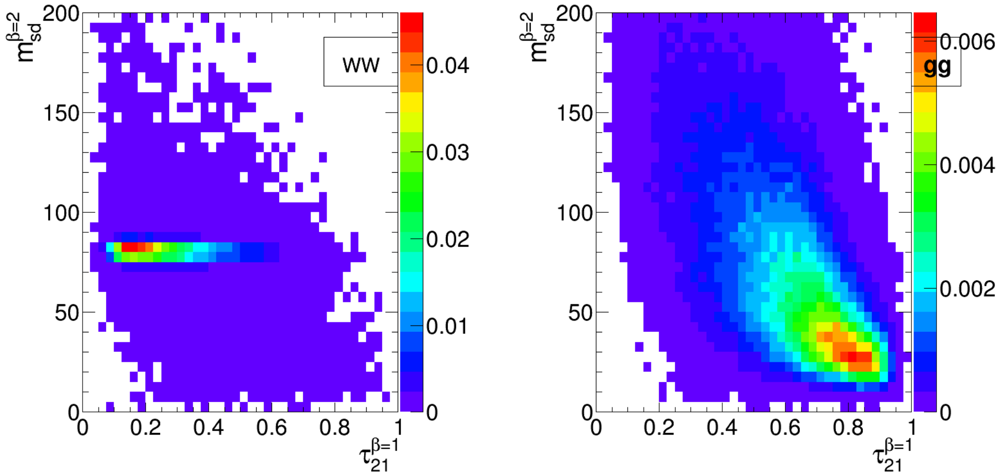


(b) p_T R=0.8

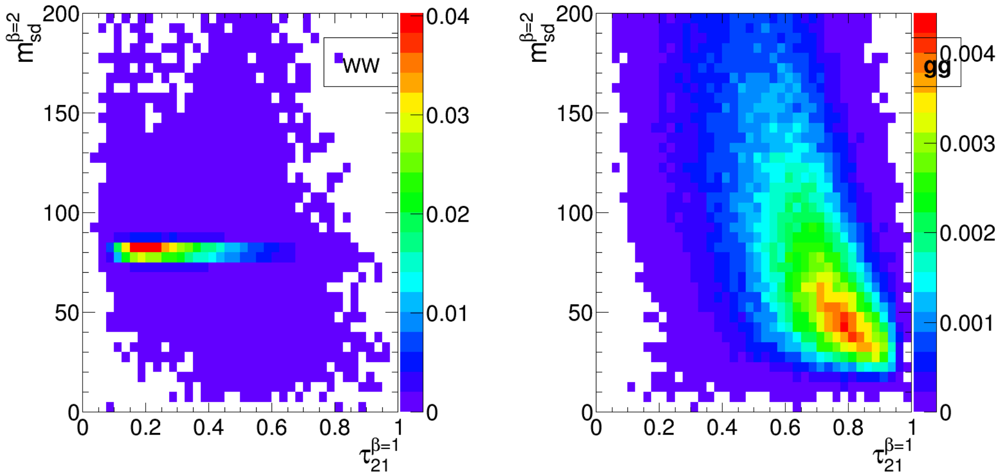


(c) p_T R=1.2

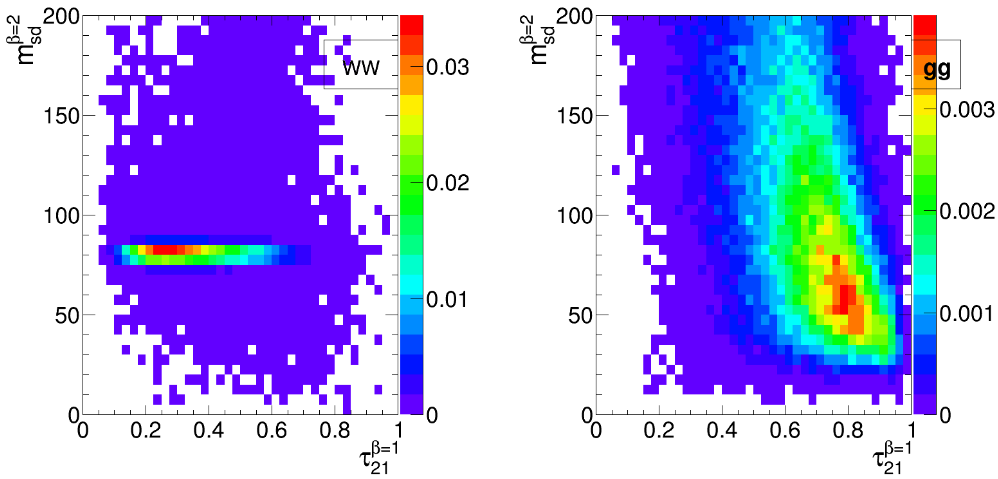
Fig. 23 2-D plots showing $m_{sd}^{\beta=2}$ versus $C_2^{\beta=1}$ for R=0.4, 0.8 and 1.2 jets in the p_T 1.0-1.1 TeV bin.



(a) p_T R=0.4

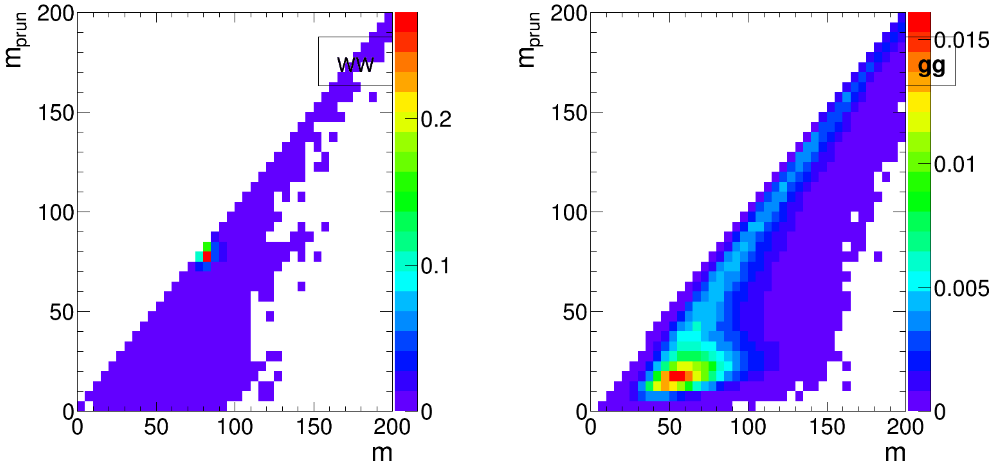


(b) p_T R=0.8

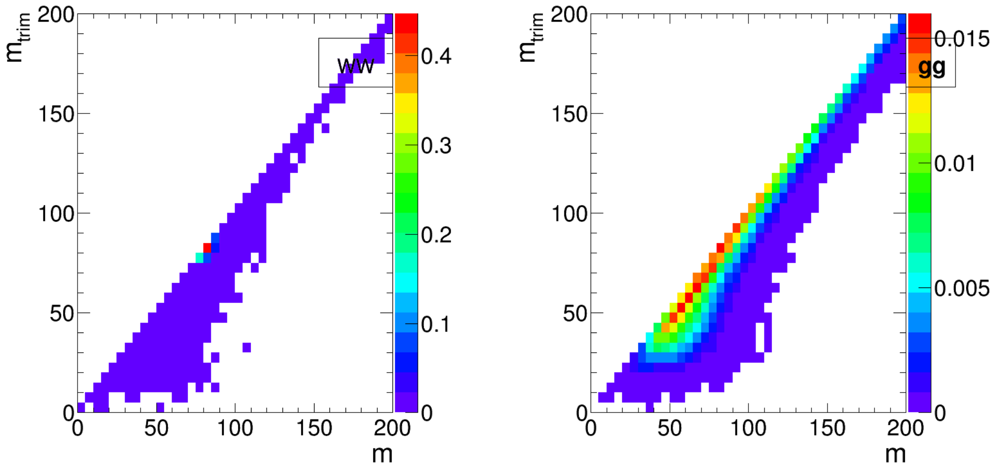


(c) p_T R=1.2

Fig. 24 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. 25 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.

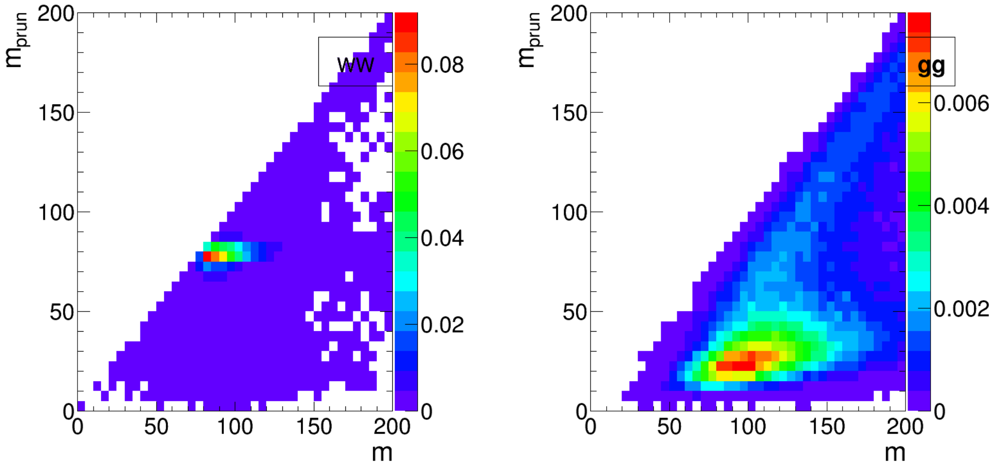
power decreasing with increasing R. This is due to the inherent sensitivity of these observables to soft, wide angle radiation.

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

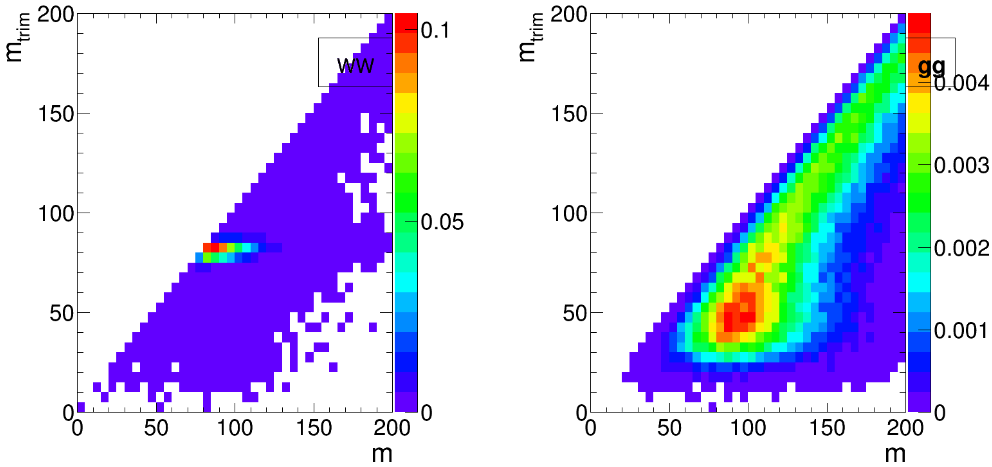
By examining the performance of a BDT combination of all the variables considered, it is clear that there are potentially substantial performance gains to be made by designing a more complex multivariate W tagger, especially at higher p_T .

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 study fully hadronic decays of the top quark.



(a) Pruned mass vs ungroomed mass



(b) Trimmed mass vs ungroomed mass

Fig. 26 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.

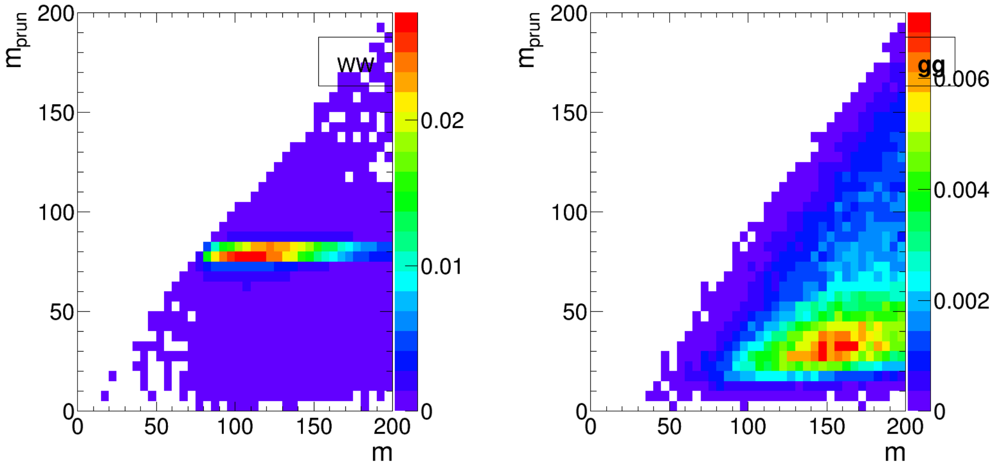
We consider top quarks with moderate boost (600-1000 GeV), and perhaps most interestingly, at high boost ($\gtrsim$ 1500 GeV). Top tagging faces several challenges in the high-p_T regime. For such high-p_T jets, the b-tagging efficiencies are no longer reliably known. Also, the top jet can also accompanied 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.

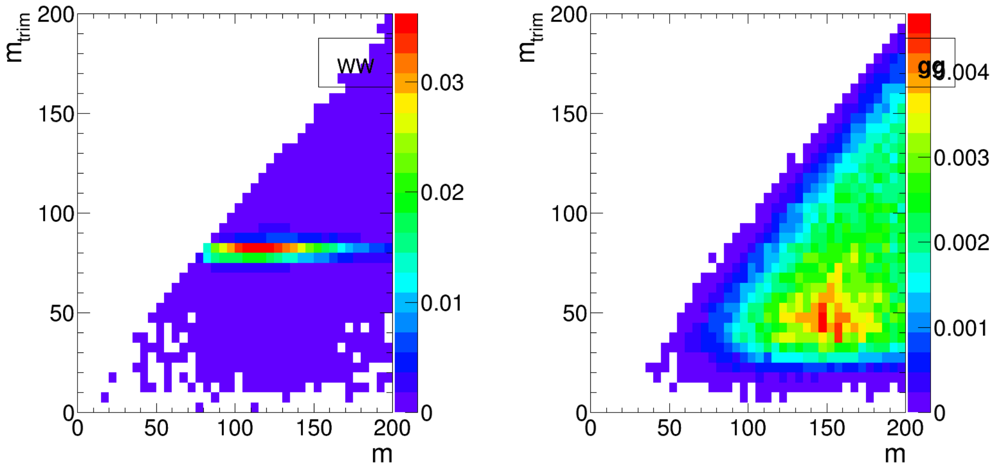
7.1 Methodology

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

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



(a) Pruned mass vs ungroomed mass



(b) Trimmed mass vs ungroomed mass

Fig. 27 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.

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, Γ_{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.

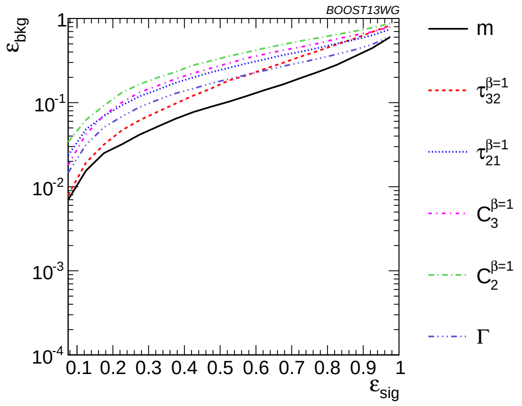
Fig. 28 shows the ROC curves for each of the top-tagging observables, with the bare (ungroomed) jet mass also plotted for comparison. The jet shape observables all perform substantially worse than jet mass, unlike W tagging for which several observables are competitive with or perform better than jet mass (see, for example, Fig. 10). 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.28(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 28 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; this is expected, as the HEPTopTagger was designed to reconstruct moderate p_T top jets in ttH events (for a proposal for a high-p_T variant of the HEPTopTagger, see [59]). In Figure 29 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 30 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 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 [60] 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. 33 and 36).

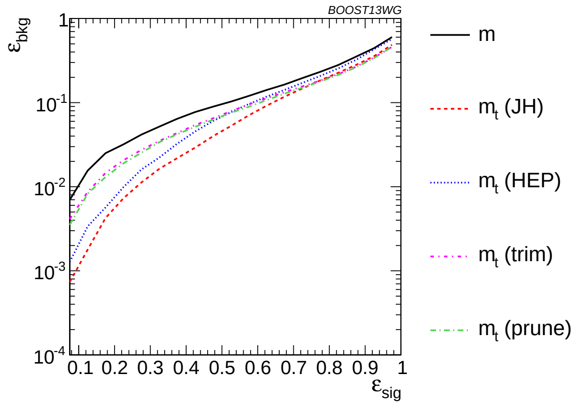
We also see in Figure 28(b) that the top mass from the JH tagger and the HEPTopTagger has superior performance relative to either of the grooming algorithms; this is because the pruning and trimming algorithms do not have inherent W-identification steps and are not optimized for this purpose. Indeed, 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 29 and 30, showing that with $\varepsilon_{\rm sig} = 0.3-0.35$, 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 31 and 33 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 31 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 32. 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 33 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 degrades at higher p_T, likely in part due to the background-shaping effects discussed earlier.

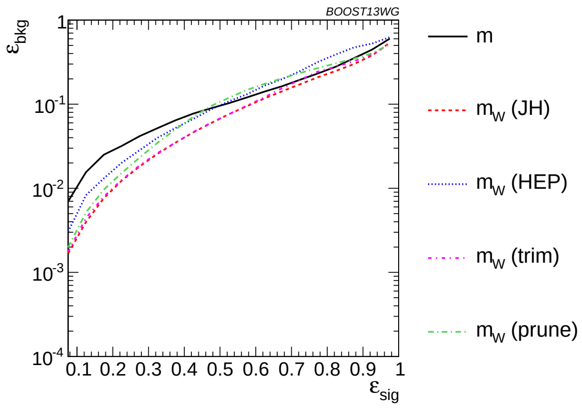
In Figures 34 and 36 we directly compare ROC curves for jet shape observable performance and top mass performance respectively for the three different jet radii considered within the p_T 1.5-1.6 TeV bin. Again, the input parameters of the taggers, groomers and shape variables are separately optimized 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



(a) Jet shapes



(b) Top mass



(c) W mass

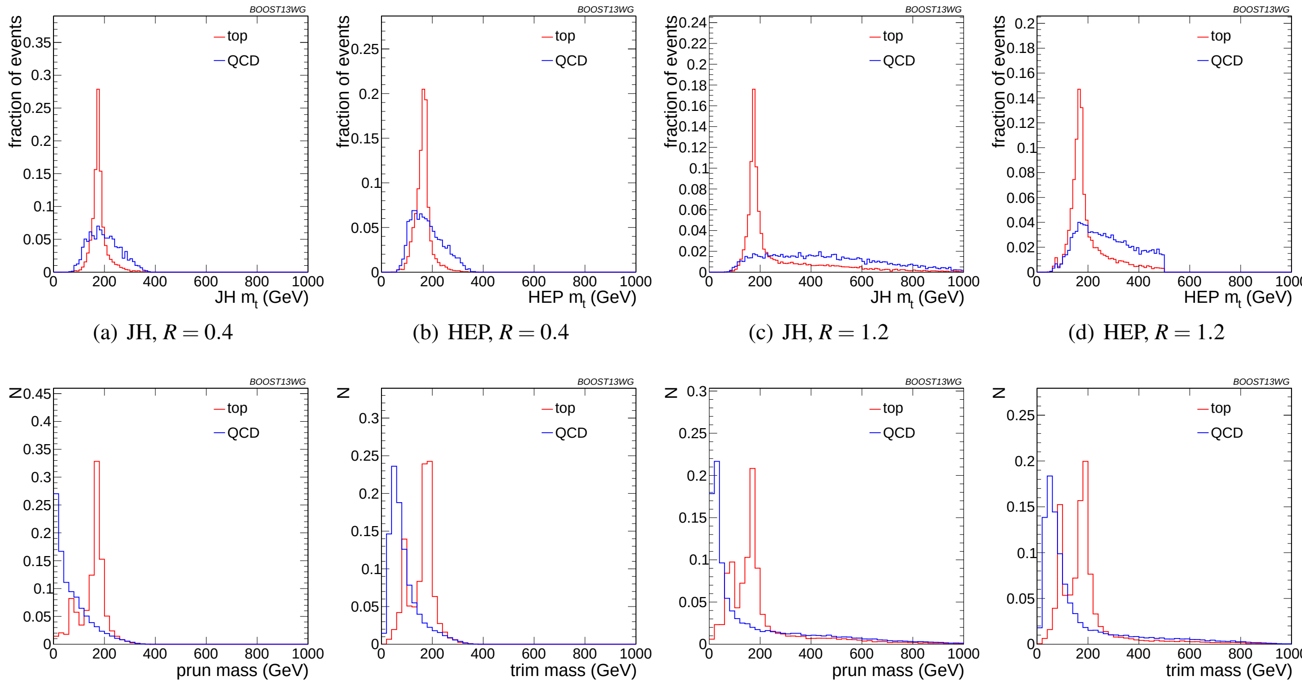
Fig. 28 Comparison of single-variable top-tagging performance in the $p_T = 1-1.1$ GeV bin using the anti-k_{T}, R=0.8 algorithm.

the 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 35, 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

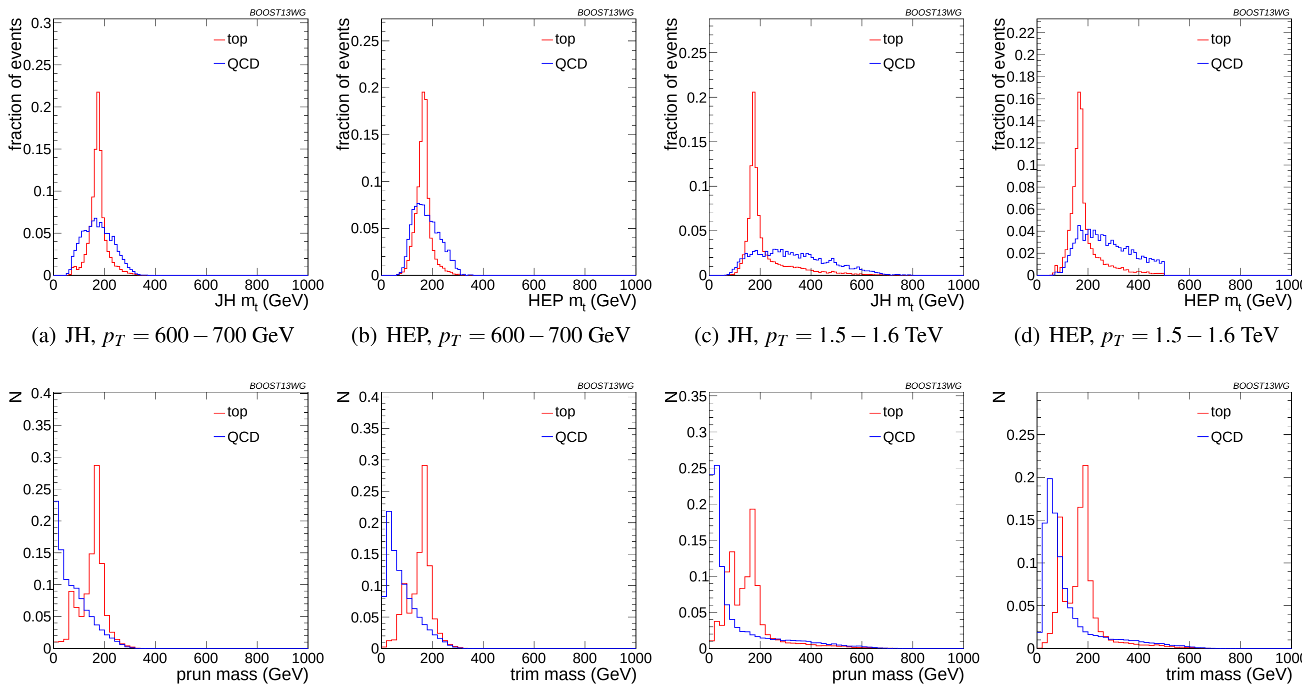
We now consider various BDT combinations of the observables from Section 7.2, using the techniques described in Section 4. In particular, we consider the performance of individual taggers such as the JH tagger and HEPTopTagger, which output information about the top and W candidate masses and the helicity 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 37, 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



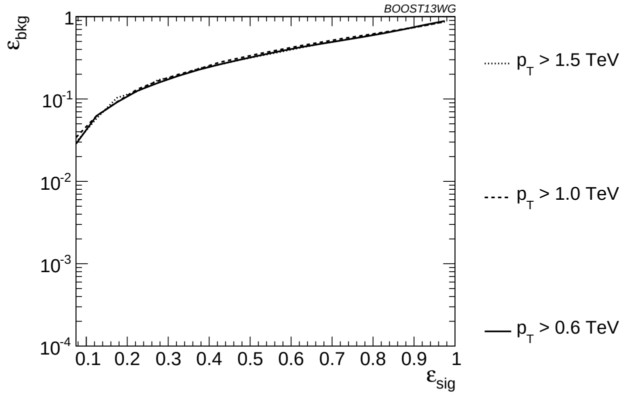
(a) JH, $R = 0.4$
(b) HEP, $R = 0.4$
(c) JH, $R = 1.2$
(d) HEP, $R = 1.2$
(e) prune, $R = 0.4$
(f) trim, $R = 0.4$
(g) prune, $R = 1.2$
(h) trim, $R = 1.2$

Fig. 29 Comparison of top mass reconstruction with the Johns Hopkins (JH), HEPTopTaggers (HEP), pruning, and trimming at different R using the anti-k_T algorithm, $p_T = 1.5 - 1.6$ 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.

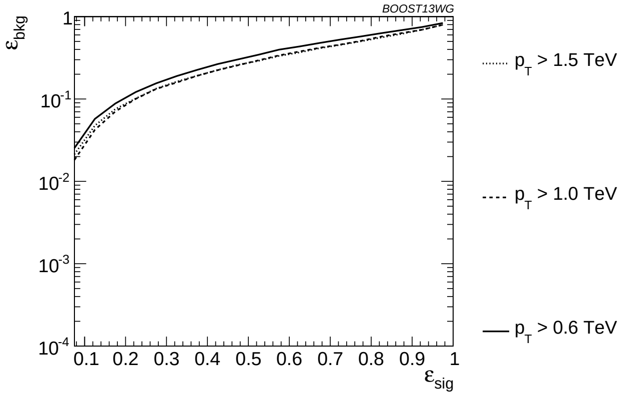


(a) JH, $p_T = 600 - 700$ GeV
(b) HEP, $p_T = 600 - 700$ GeV
(c) JH, $p_T = 1.5 - 1.6$ TeV
(d) HEP, $p_T = 1.5 - 1.6$ TeV
(e) prune, $p_T = 600 - 700$ GeV
(f) trim, $p_T = 600 - 700$ GeV
(g) prune, $p_T = 1.5 - 1.6$ TeV
(h) trim, $p_T = 1.5 - 1.6$ TeV

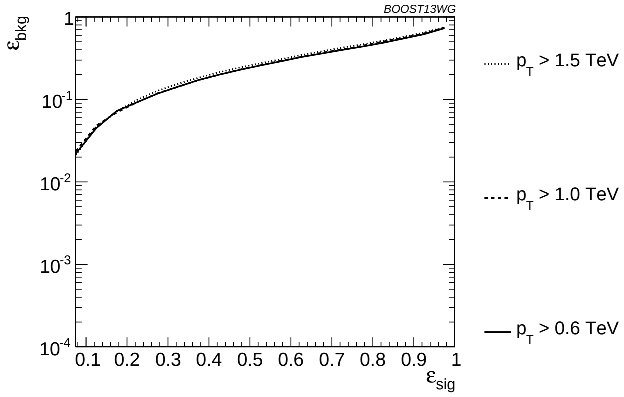
Fig. 30 Comparison of top mass reconstruction with the Johns Hopkins (JH), HEPTopTaggers (HEP), pruning, and trimming at different p_T using the anti-k_T algorithm, $R = 0.8$. Each histogram is shown for the working point optimized for best performance with m_t in the $0.3 - 0.35$ signal efficiency bin, and is normalized to the fraction of events passing the tagger.



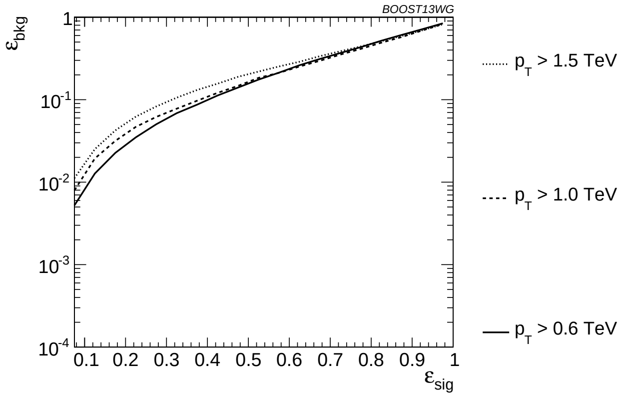
(a) $C_2^{(\beta=1)}$



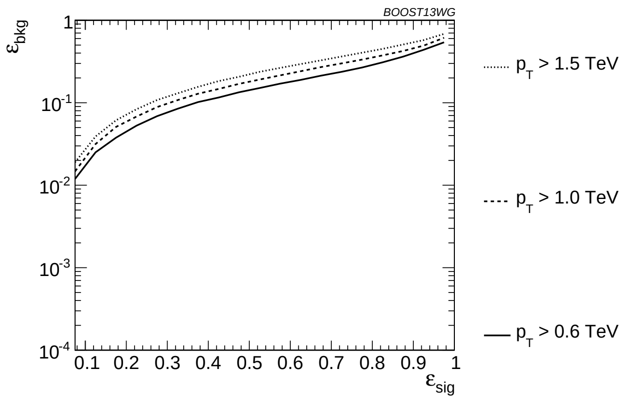
(b) $C_3^{(\beta=1)}$



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



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

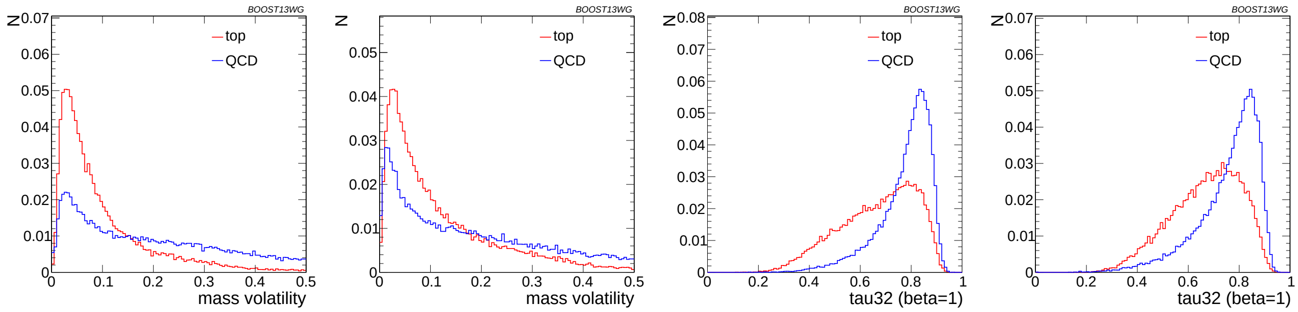


(e) Qjet mass volatility

Fig. 31 Comparison of individual jet shape performance at different p_T using the anti-k_{T} R=0.8 algorithm.

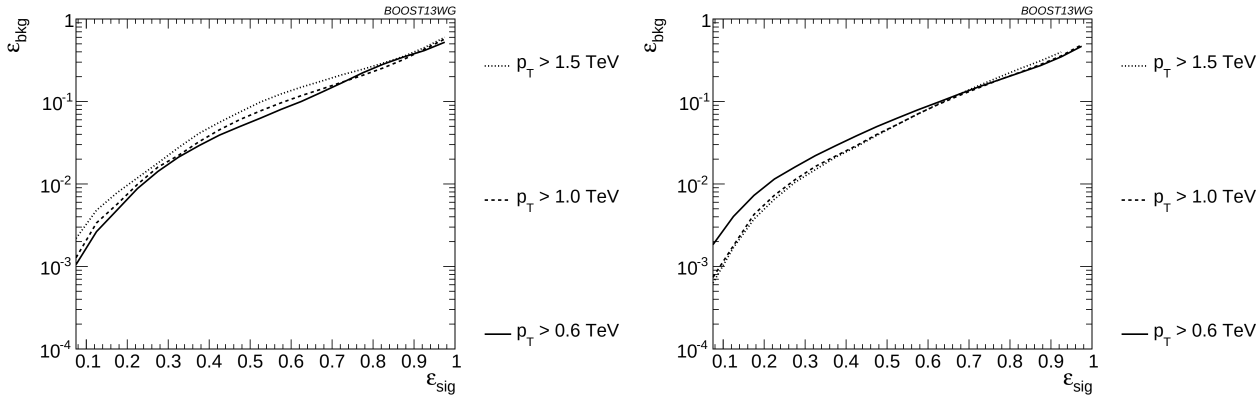
there is complementary information in the mass outputs from different top taggers, we also consider in Figure 37 a multivariable combination of all of the JH and HEPTopTagger outputs. The maximum efficiency of the combined JH and HEPTopTaggers is limited, as some fraction of signal events inevitably 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 38 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 discrimi-

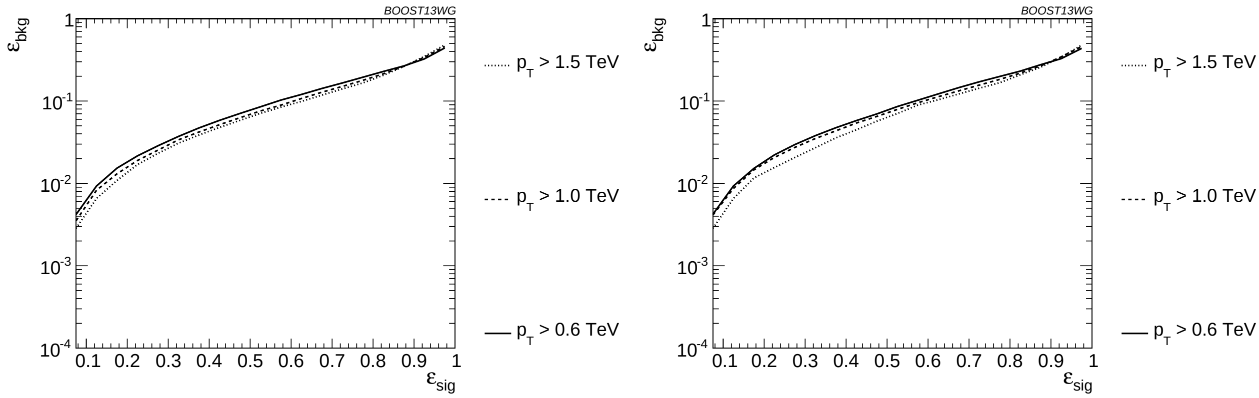


(a) $\mathit{\Gamma}_{\text{Qjet}}$, $p_T = 600-700$ GeV (b) $\mathit{\Gamma}_{\text{Qjet}}$, $p_T = 1.5-1.6$ TeV (c) $\tau_{32}^{(\beta=1)}$, $p_T = 600-700$ GeV (d) $\tau_{32}^{(\beta=1)}$, $p_T = 1.5-1.6$ TeV

Fig. 32 Comparison of $\mathit{\Gamma}_{\text{Qjet}}$ 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.



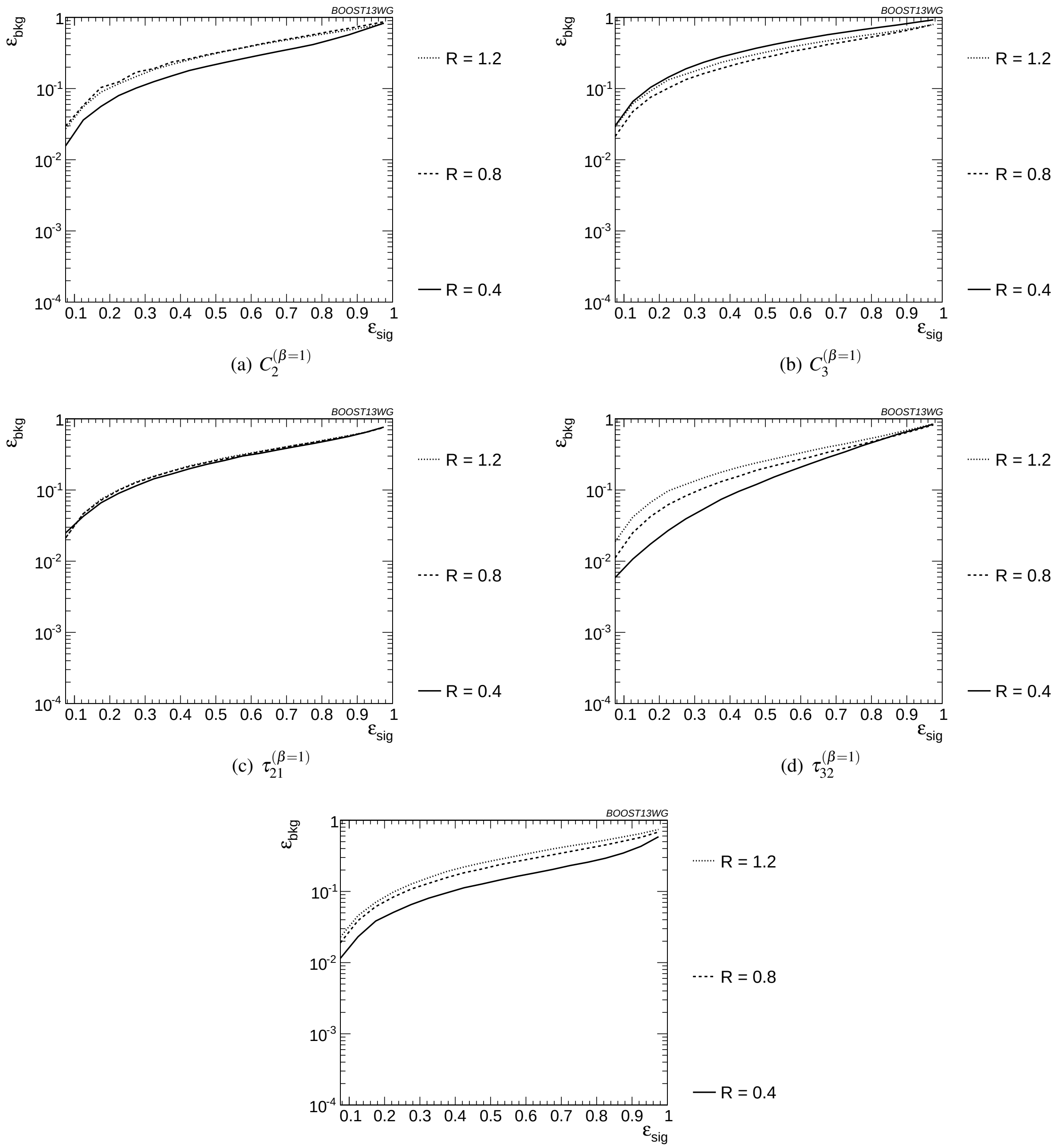
(a) HEPTopTagger m_t (b) Johns Hopkins Tagger m_t



(c) Pruning m_t (d) Trimming m_t

Fig. 33 Comparison of top mass performance of different taggers at different p_T using the anti-k_{T} R=0.8 algorithm.

nation. 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 38(c). Combining the taggers with shape information nearly erases the difference between the tagging methods observed in Figure 37; 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.

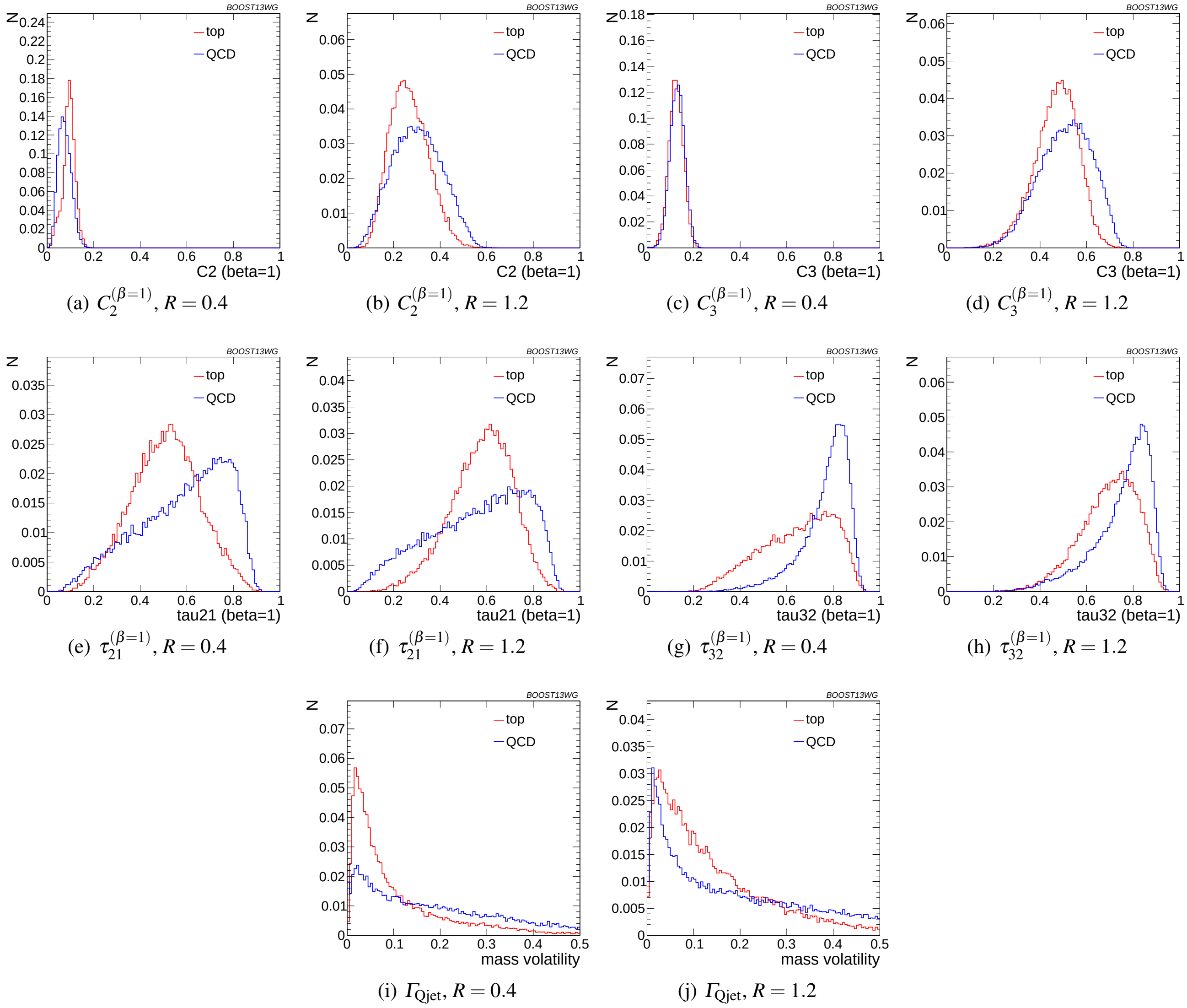


(a) $C_2^{(\beta=1)}$ (b) $C_3^{(\beta=1)}$ (c) $\tau_{21}^{(\beta=1)}$ (d) $\tau_{32}^{(\beta=1)}$ (e) Qjet mass volatility

Fig. 34 Comparison of individual jet shape performance at different R in the $p_T = 1.5-1.6$ TeV bin.

In Figure 39 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 39(c), we find that the differences between pruning and trimming are erased when combined with shape information.

Finally, in Figure 40, we compare the performance of each of the tagger/groomers when their outputs are com-



(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) $\tau_{32}^{(\beta=1)}$, $R = 1.2$

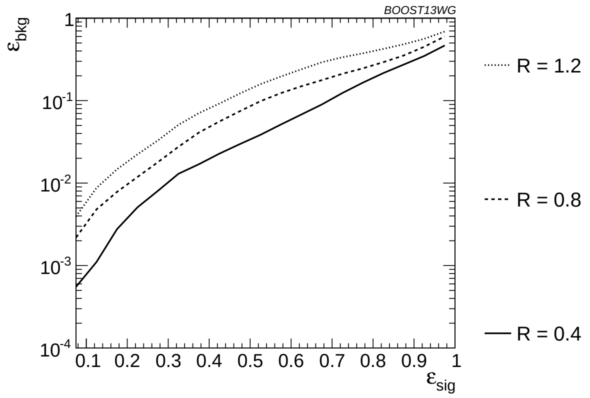
(i) $\Gamma_{\rm Qjet}$, $R = 0.4$ (j) $\Gamma_{\rm Qjet}$, $R = 1.2$

Fig. 35 Comparison of various shape observables in the $p_T = 1.5-1.6$ TeV bin and different values of the anti-$k_{\rm T}$ radius R.

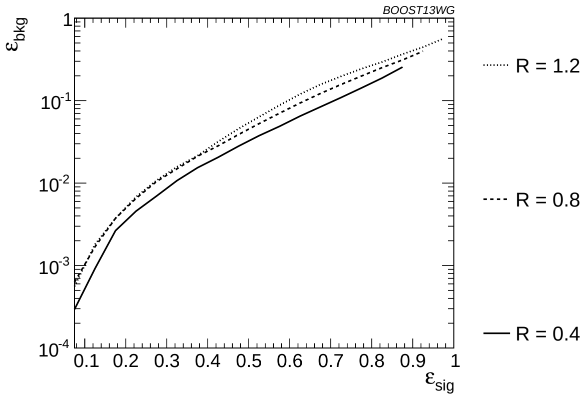
bined with all of the shape observables considered. One can see that the discrepancies between the performance of the different taggers/groomers all but vanishes, suggesting perhaps that we are here utilising all available signal-background discrmination information, and that this is the optimal top-tagging 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-$k_{\rm T}$ 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 41. The behaviour with p_T is qualitatively similar to the behaviour of the m_t observable for each tagger/groomer shown in Figure 33; 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.

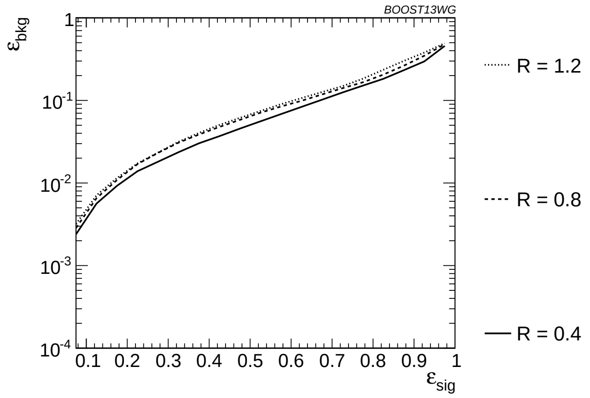
In Figure 42, we show the p_T dependence of BDT combinations of the JH tagger output combined with shape observables. We find that the curves look nearly identical: the p_T dependence is dominated by the top mass reconstruction, and combining the tagger outputs with different shape observables does not substantially change this behaviour. The same holds true for trimming and pruning. By contrast, HEPTopTagger ROC curves, shown in Figure 43, do change



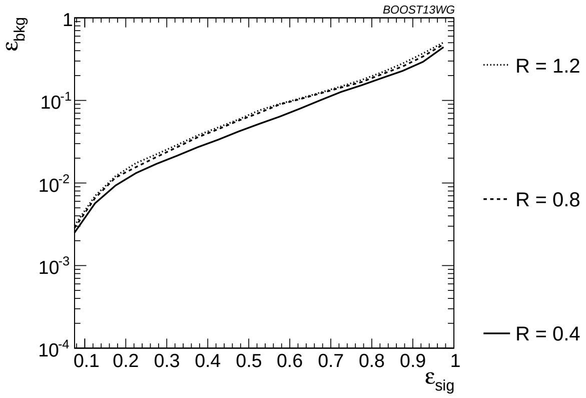
(a) HEPTopTagger m_t



(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 R in the $p_T = 1.5 - 1.6$ TeV bin.

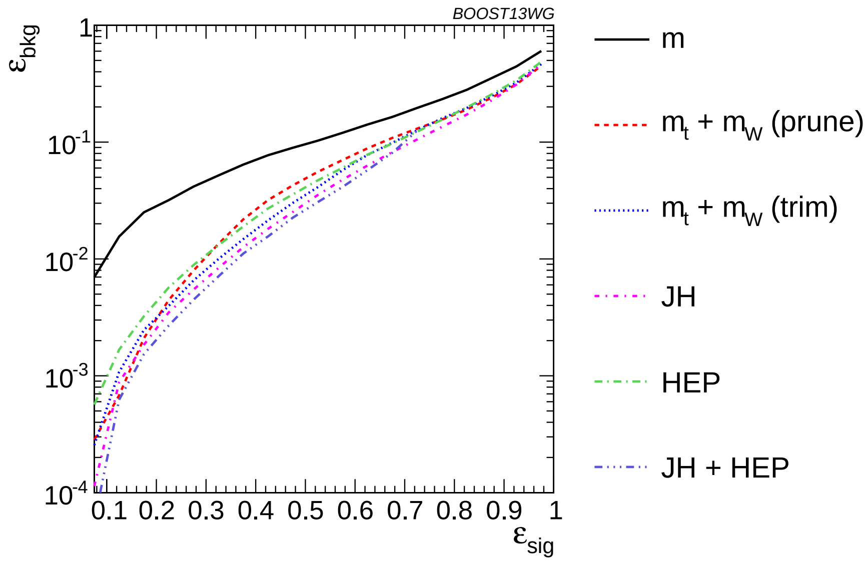
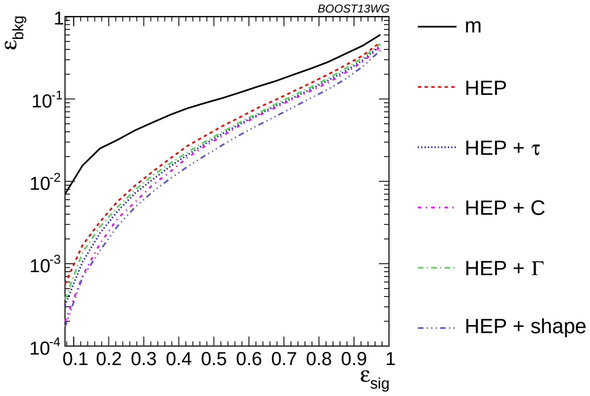
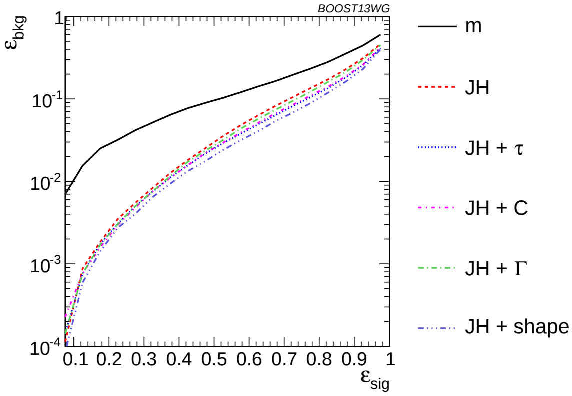


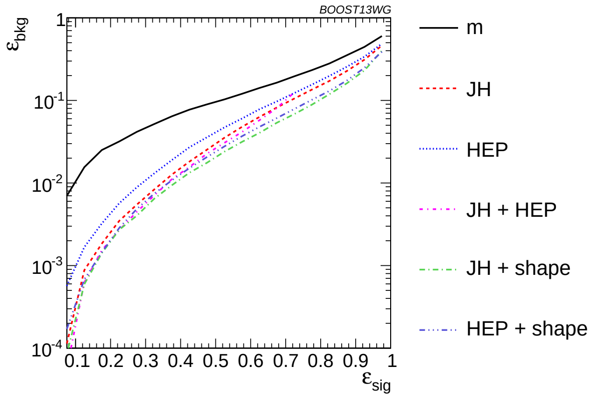
Fig. 37 The performance of the various taggers in the $p_T = 1 - 1.1$ TeV bin using the anti-k_{T} R=0.8 algorithm. For the groomers a BDT combination of the reconstructed m_t and m_W are used. Also shown is a multivariable combination of all of the JH and HEPTopTagger outputs. The ungroomed mass performance is shown for comparison.



(a) HEPTopTagger + Shape



(b) Johns Hopkins Tagger + shape



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

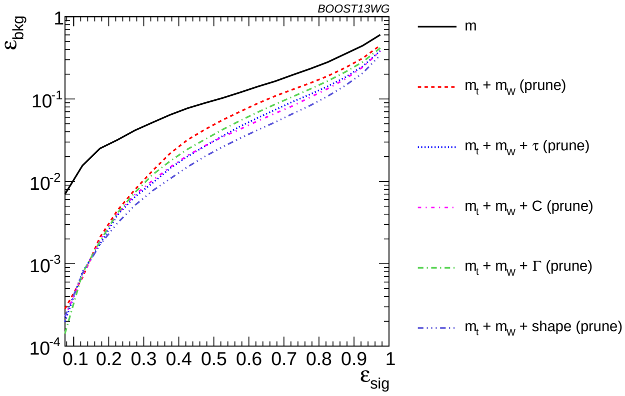
Fig. 38 The performance of BDT combinations of the JH and HepTopTagger outputs with various shape observables in the $p_T = 1-1.1$ TeV bin using the anti-k_{T} R=0.8 algorithm. Taggers are combined with the following shape observables: $\tau_{21}^{(\beta=1)} + \tau_{32}^{(\beta=1)}$, $C_2^{(\beta=1)} + C_3^{(\beta=1)}$, Γ_{Qjet}, and all of the above (denoted "shape").

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 Figure 31(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.

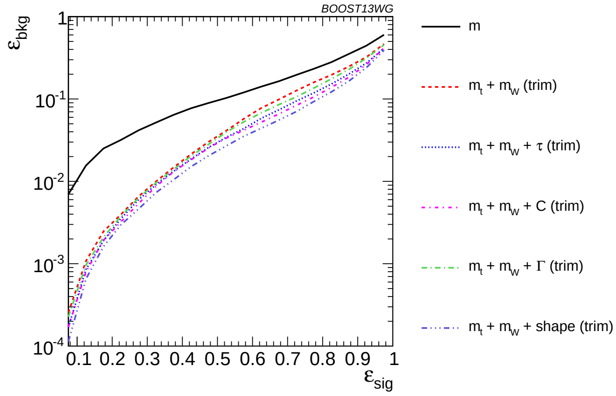
In Figure 44 we compare the BDT combinations of tagger outputs, with and without shape variables, at different jet radius R in the $p_T = 1.5-1.6$ TeV bin. The taggers are optimized over all input parameters for each choice of R and signal efficiency. We find that, for all taggers and groomers, the performance 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 45 is shown the depedence 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

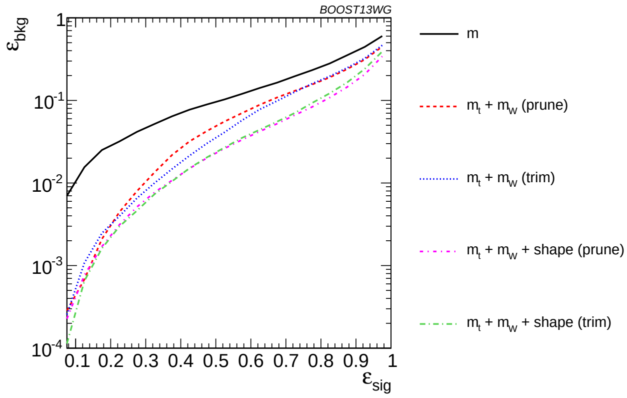
Up until now, we have re-optimized our tagger and groomer parameters for each p_T, R, and signal efficiency working point. In reality, experiments will choose a finite set of working points to use. How do our results hold up when this is taken into account? To address this concern, we replicate our analyses, 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.



(a) Pruning + Shape



(b) Trimming + Shape



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

Fig. 39 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-$k_{\rm T}$ 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").

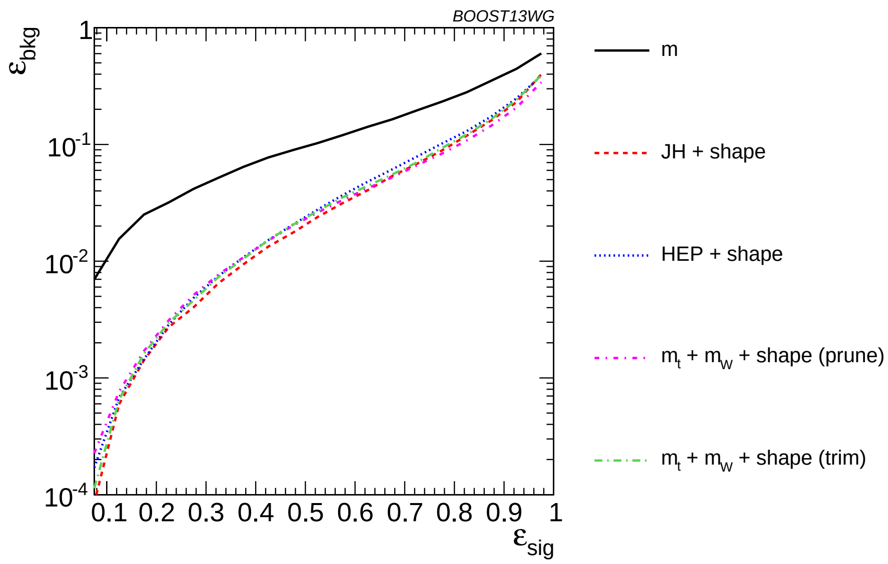
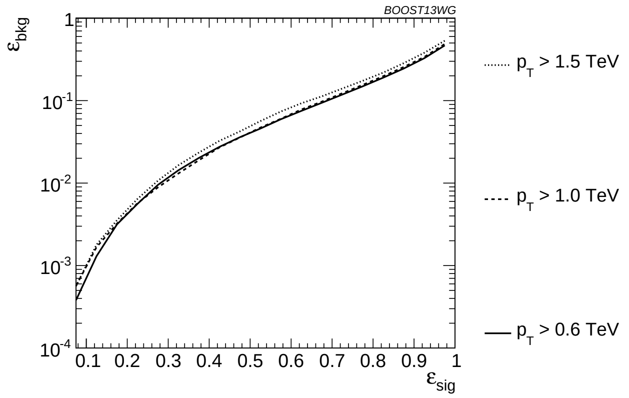
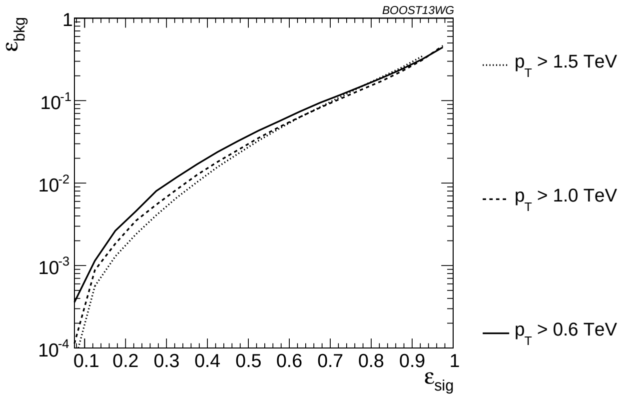


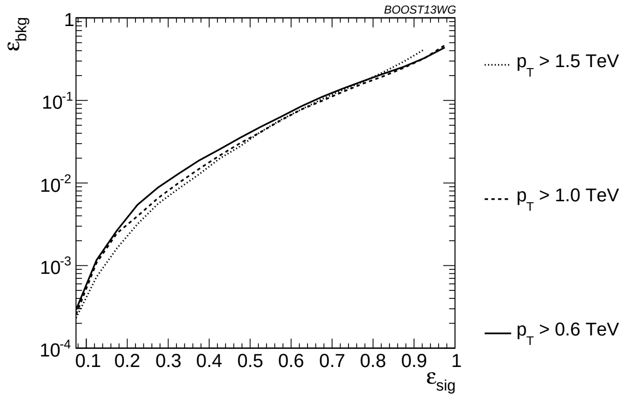
Fig. 40 Comparison of the performance of the BDT combinations of all the groomer/tagger outputs with all the available shape observables in the $p_T = 1-1.1$ TeV bin using the anti-$k_{\rm T}$ R=0.8 algorithm. Tagger/groomer outputs are combined with all of the following shape observables: $\tau_{21}^{(\beta=1)} + \tau_{32}^{(\beta=1)}$, $C_2^{(\beta=1)} + C_3^{(\beta=1)}$, $\Gamma_{\rm Qjet}$.



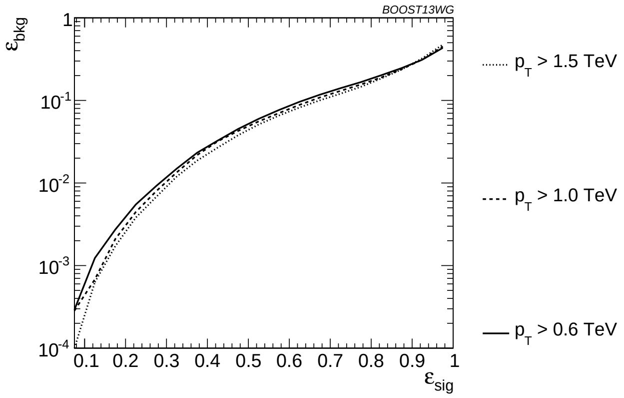
(a) HEPTopTagger



(b) Johns Hopkins Tagger



(c) Trimming



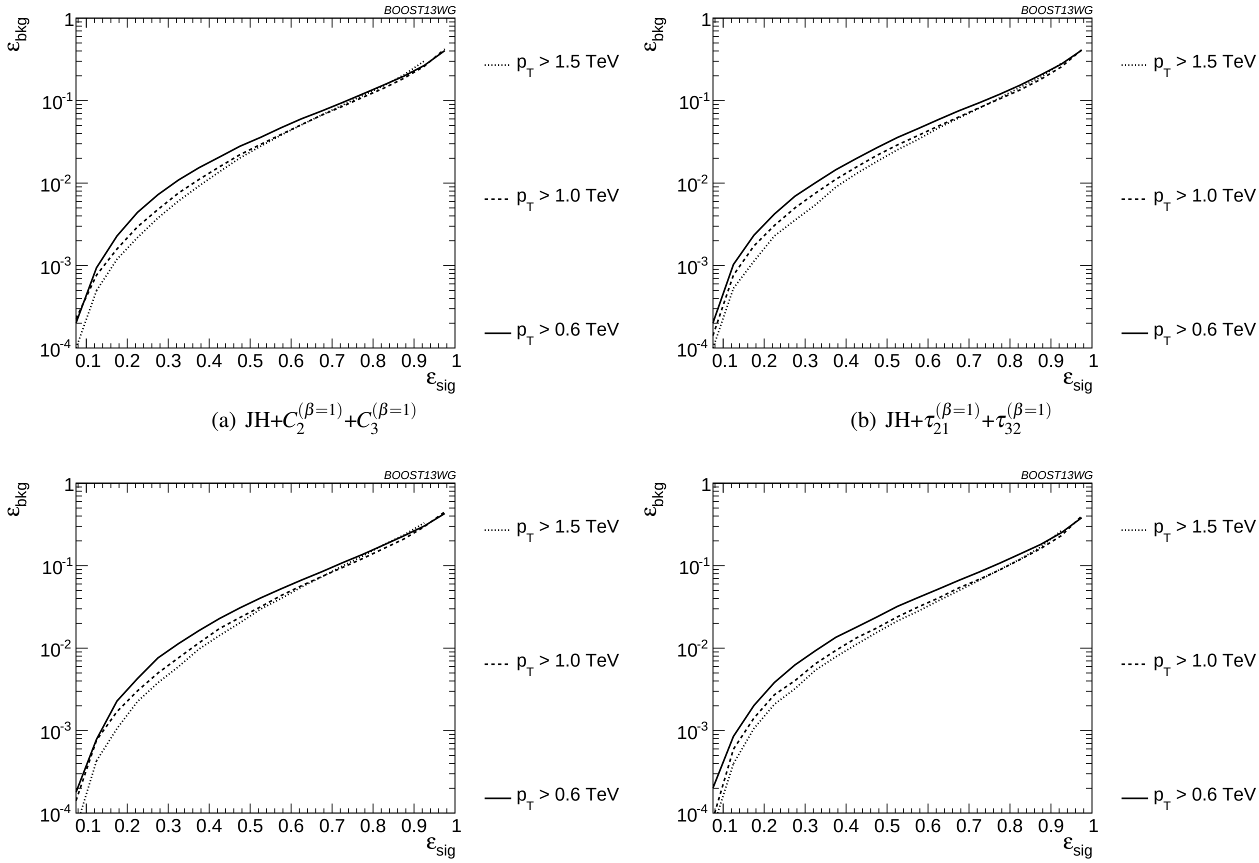
(d) Pruning

Fig. 41 Comparison of BDT combination of tagger performance at different p_T using the anti-k_T R=0.8 algorithm.

Optimizing at a single p_T : We show in Figure 46 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 associated 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 47), 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 48 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 $\varepsilon_{\rm sig}$ 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



(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. 42 Comparison of BDT combination of JH tagger + shape at different p_T using the anti-k_{T} R=0.8 algorithm.

p_T would require a more careful optimization of parameters to obtain the best performance.

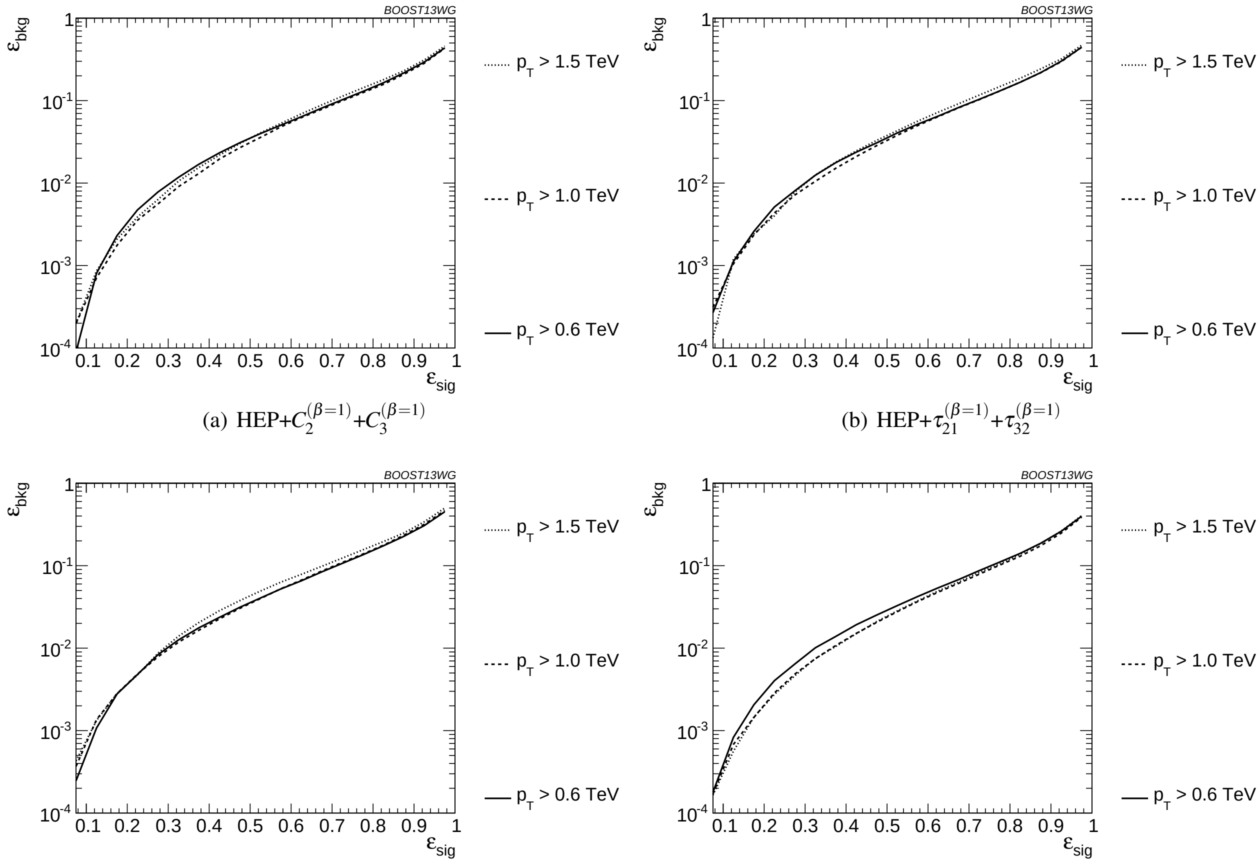
The same holds true for the BDT combinations of the full tagger outputs, shown in Figure 49). The performance for the sub-optimal taggers is still within an $O(1)$ factor of the optimized performance, and the HEPTopTagger performs better with the combination of all of its outputs relative to the performance with just m_t. The same behaviour holds for the BDT combinations of tagger outputs and shape 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 50 for cuts on the top mass and W mass, and in Figure 51 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 50 and 51 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,



(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

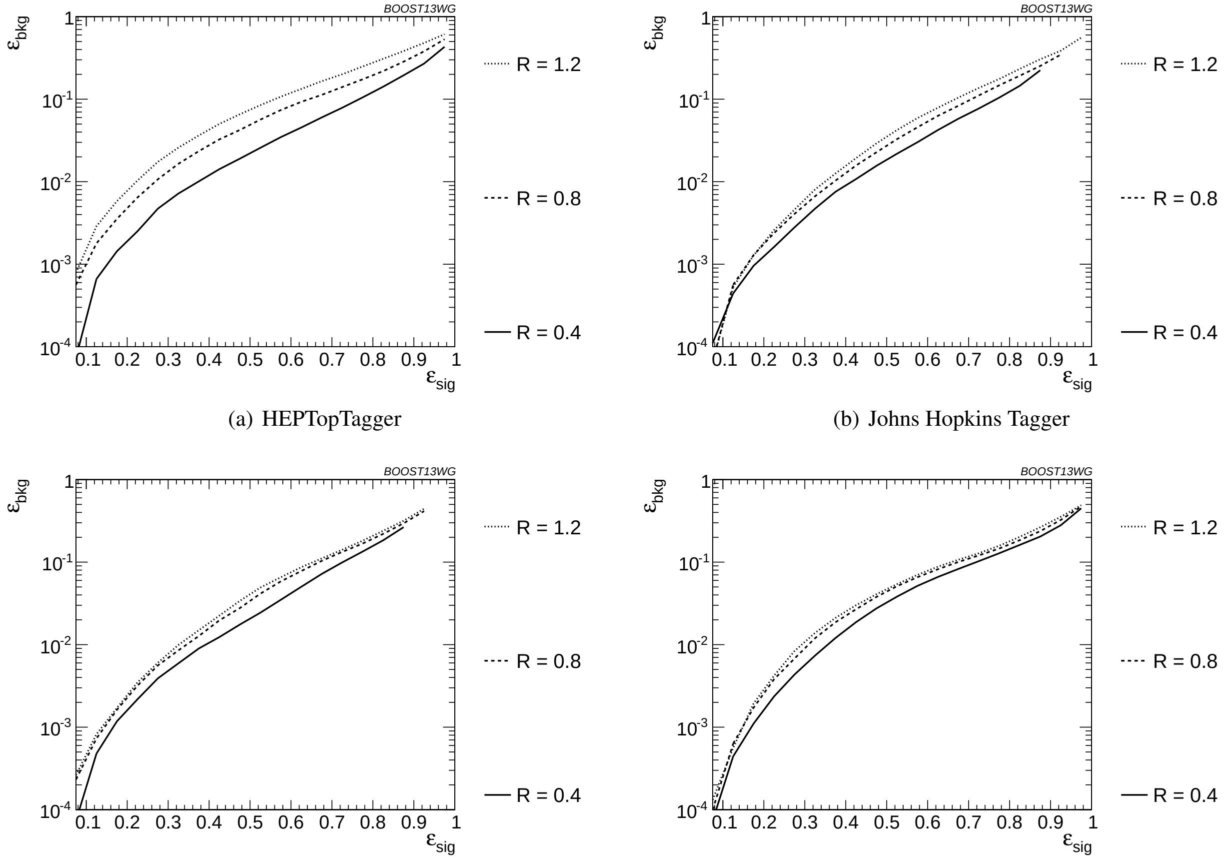
(d) HEP + all

Fig. 43 Comparison of BDT combination of HEP tagger + shape at different p_T using the anti-k_{T} R=0.8 algorithm.

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; as expected from its design, the HEPTopTagger performance degrades at high p_T . 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 various 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.



(a) HEPTopTagger

(b) Johns Hopkins Tagger

(c) Trimming

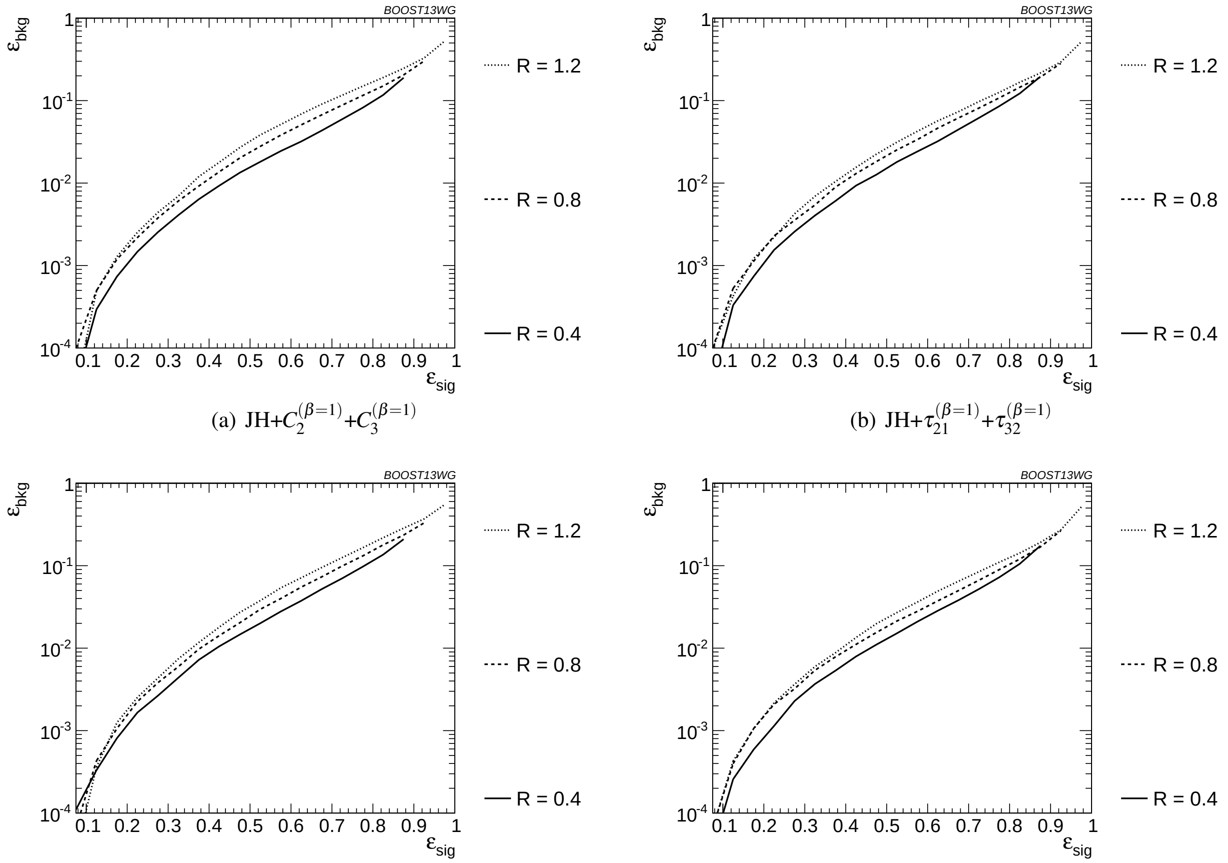
(d) Pruning

Fig. 44 Comparison of tagger and jet shape performance at different radius at p_T = 1.5-1.6 TeV.

8 Summary & Conclusions

Furthering our understanding of jet substructure is crucial to improving our understanding of QCD and enhancing the prospects for the discovery of new physical processes at Run II of the LHC. In this report we have studied the performance of jet substructure techniques over a wide range of kinematic regimes that will be encountered in Run II of the LHC. The performance of observables and their correlations have been studied 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, to understand the potential of the "ultimate" tagger where "all" available information (at least, all of that provided by the variables considered) is used.

We focused on the discrimination of quark jets from gluon jets, and the discrimination of boosted W bosons and top quarks from the QCD backgrounds. For each, we have identified the best-performing jet substructure observables, both individually and in combination with other observables. In doing so, we have also provided a physical picture of why certain sets of observables are (un)correlated. Additionally, we have investigated how the performance of jet substructure observables varies with R and p_T , identifying observables that are particularly robust against or susceptible to these changes. In the case of q/g tagging, it seems that close to the ultimate performance can be achieved by combining the most powerful discriminant, the number of constituents of a jet, with just one other variable, $C_1^{\beta=1}$ (or $\tau_1^{\beta=1}$). Many of the other variables considered are highly correlated and provide little additional discrimination. For both top and W tagging, the groomed mass is a very important discriminating variable, but one that can be substantially improved in combination with other variables. There is clearly a rich and complex relationship between the variables considered for W and top tagging, and the performance and correlations between these variables can change considerably with changing jet p_T and R. In the case of W tagging, even after combining groomed mass with two other substructure observables, we are still some way short of the ultimate tagger performance, indicating the complexity of the information available, and the complementarity between the observ-



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

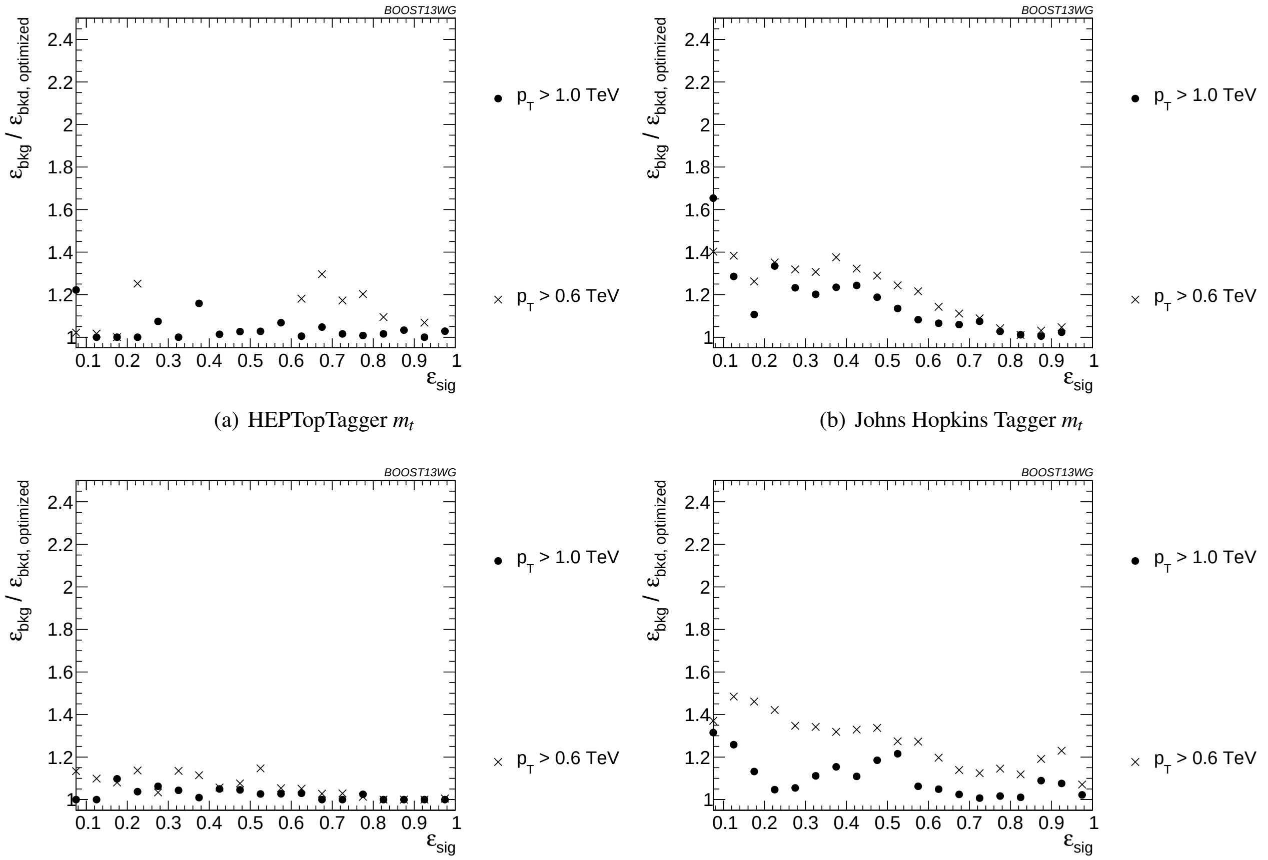
ables considered. In the case of top tagging, we have shown that the performance of both the John Hopkins and Hep Top Tagger can be improved when their outputs are combined with substructure observables such as τ_{32} and C_3, and that the performance of a discriminant built from groomed mass information plus substructure observables is very comparable to the performance of the taggers. We have optimized the top taggers for a particular value of p_T , R, and signal efficiency, and studied their performance at other working points. We have found that the performance of observables remains within a factor of two of the optimized value, suggesting that the performance of jet substructure observables is not significantly degraded when tagger parameters are only optimized for a few select benchmark points.

Our analyses were performed with ideal detector and pile-up conditions in order to most clearly elucidate the underlying physical scaling with p_T and R. At higher boosts, detector resolution effects will become more important, and with the higher pile-up expected at Run II of the LHC, pile-up mitigation will be crucial for future jet substructure studies. Future studies will be needed to determine which of the observables we have studied are most robust against pile-up and detector effects, and our analyses suggest particularly useful combinations of observables to consider in such studies.

At the new energy frontier of Run II of the LHC boosted jet substructure techniques will be more central to our searches for new physics than ever before, and by achieving a deeper understanding of the underlying structure of quark, gluon, W and Top initiated jets, and how the observables that try to elucidate this structure are related, the hope is that more sophisticated taggers can be commissioned that will extend the reach for new physics as far as possible.

References

1. *Boost2009*, SLAC National Accelerator Laboratory, 9-10 July, 2009, [http://www-conf.slac.stanford.edu/Boost2009].
2. *Boost2010*, University of Oxford, 22-25 June 2010, [http://www.physics.ox.ac.uk/boost2010].
3. *Boost2011*, Princeton University, 22-26 May 2011, [https://indico.cern.ch/event/138809/].

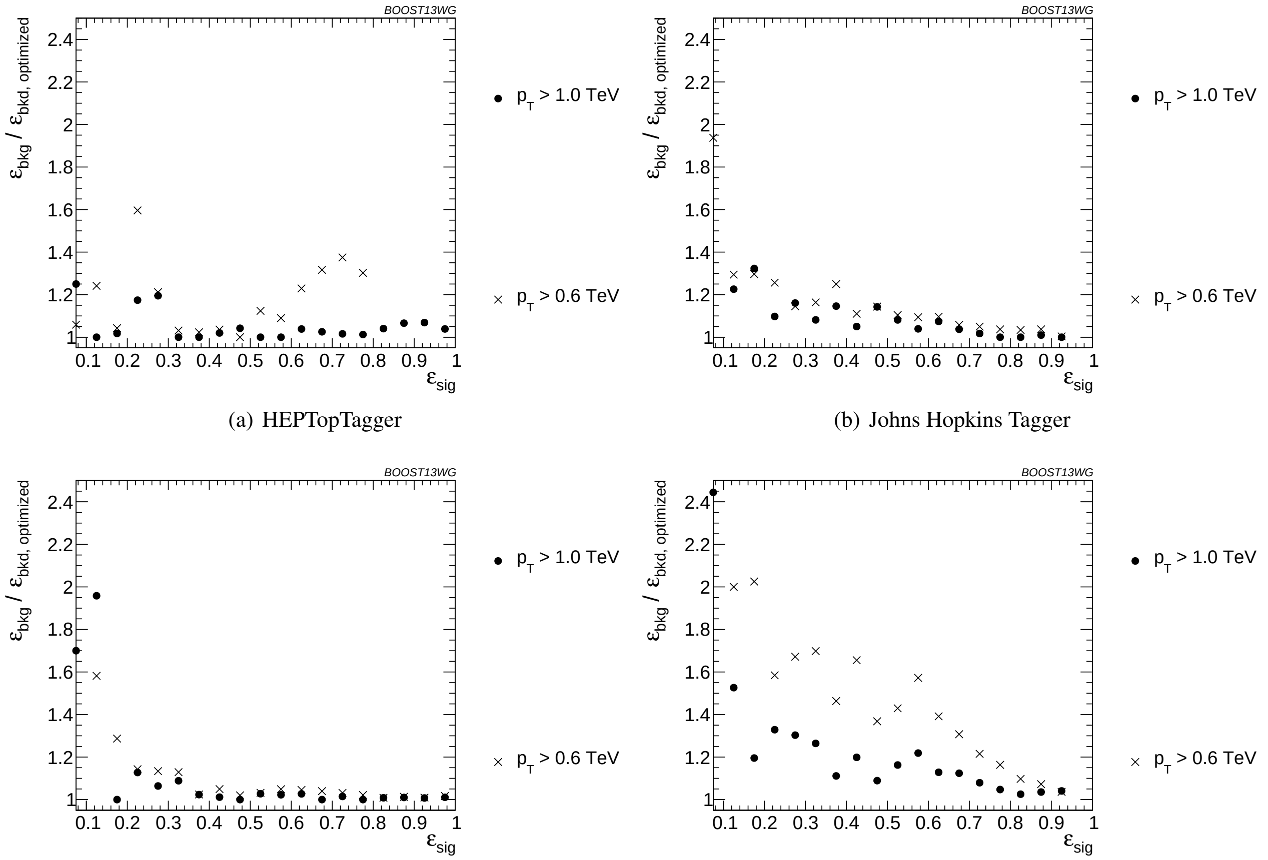


(a) HEPTopTagger m_t (b) Johns Hopkins Tagger m_t

(c) Pruning m_t (d) Trimming m_t

Fig. 46 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.

4. *Boost2012*, IFIC Valencia, 23-27 July 2012, [http://ific.uv.es/boost2012].
5. *Boost2013*, University of Arizona, 12-16 August 2013, [https://indico.cern.ch/event/215704/].
6. *Boost2014*, University College London, 18-22 August 2014, [http://http://www.hep.ucl.ac.uk/boost2014/].
7. 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].
8. 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].
9. 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.
10. M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur.Phys.J.* **C72** (2012) 1896, [arXiv:1111.6097].
11. T. Plehn, M. Spannowsky, M. Takeuchi, and D. Zerwas, *Stop Reconstruction with Tagged Tops*, *JHEP* **1010** (2010) 078, [arXiv:1006.2833].
12. D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, *Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks*, *Phys.Rev.Lett.* **101** (2008) 142001, [arXiv:0806.0848].
13. J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5 : Going Beyond*, *JHEP* **1106** (2011) 128, [arXiv:1106.0522].
14. Y. Gao, A. V. Gritsan, Z. Guo, K. Melnikov, M. Schulze, et al., *Spin determination of single-produced resonances at hadron colliders*, *Phys.Rev.* **D81** (2010) 075022, [arXiv:1001.3396].
15. S. Bolognesi, Y. Gao, A. V. Gritsan, K. Melnikov, M. Schulze, et al., *On the spin and parity of a single-produced resonance at the LHC*, *Phys.Rev.* **D86** (2012) 095031, [arXiv:1208.4018].
16. I. Anderson, S. Bolognesi, F. Caola, Y. Gao, A. V. Gritsan, et al., *Constraining anomalous HVV interactions at proton and lepton colliders*, *Phys.Rev.*



(a) HEPTopTagger

(b) Johns Hopkins Tagger

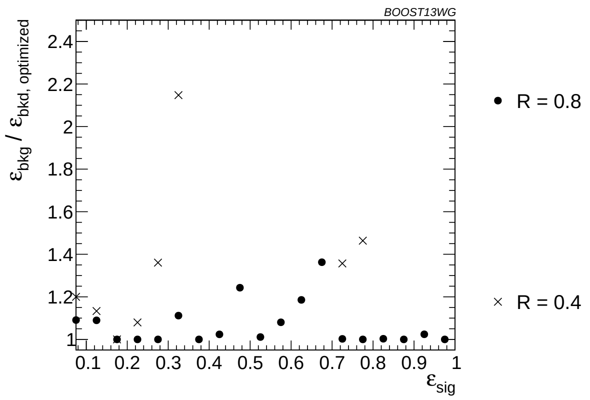
(c) Pruning

(d) Trimming

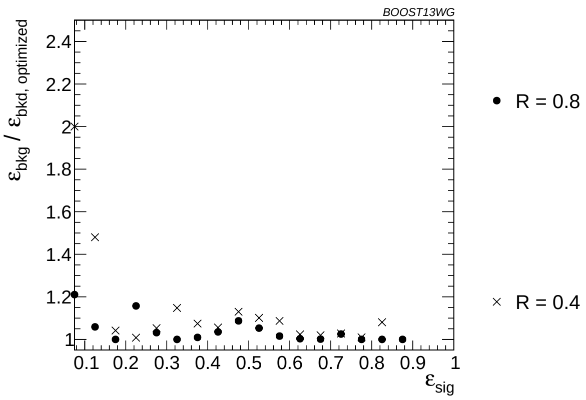
Fig. 47 Comparison of BDT combination of tagger performance at different p_T using the anti-k_T R=0.8 algorithm; the tagger inputs are set to the optimum value for $p_T = 1.5 - 1.6$ TeV.

D89 (2014) 035007, [arXiv:1309.4819].

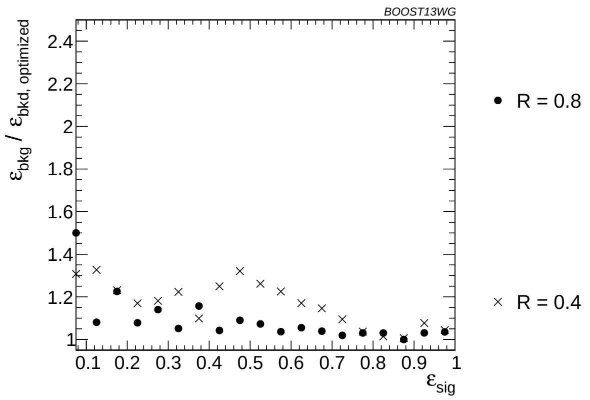
17. J. Pumplin, D. Stump, J. Huston, H. Lai, P. M. Nadolsky, et al., *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **0207** (2002) 012, [hep-ph/0201195].
18. T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput.Phys.Commun.* **178** (2008) 852–867, [arXiv:0710.3820].
19. A. Buckley, J. Butterworth, S. Gieseke, D. Grellscheid, S. Hoche, et al., *General-purpose event generators for LHC physics*, *Phys.Rept.* **504** (2011) 145–233, [arXiv:1101.2599].
20. T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, *JHEP* **0902** (2009) 007, [arXiv:0811.4622].
21. S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, *JHEP* **0803** (2008) 038, [arXiv:0709.1027].
22. F. Krauss, R. Kuhn, and G. Soff, *AMEGIC++ 1.0: A Matrix element generator in C++*, *JHEP* **0202** (2002) 044, [hep-ph/0109036].
23. T. Gleisberg and S. Hoeche, *Comix, a new matrix element generator*, *JHEP* **0812** (2008) 039, [arXiv:0808.3674].
24. S. Hoeche, F. Krauss, S. Schumann, and F. Siegert, *QCD matrix elements and truncated showers*, *JHEP* **0905** (2009) 053, [arXiv:0903.1219].
25. M. Schonherr and F. Krauss, *Soft Photon Radiation in Particle Decays in SHERPA*, *JHEP* **0812** (2008) 018, [arXiv:0810.5071].
26. **JADE Collaboration** Collaboration, S. Bethke et al., *Experimental Investigation of the Energy Dependence of the Strong Coupling Strength*, *Phys.Lett.* **B213** (1988) 235.
27. M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **0804** (2008) 063, [arXiv:0802.1189].
28. Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *Better jet clustering algorithms*, *JHEP* **9708** (1997) 001, [hep-ph/9707323].
29. M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, hep-ph/9907280.



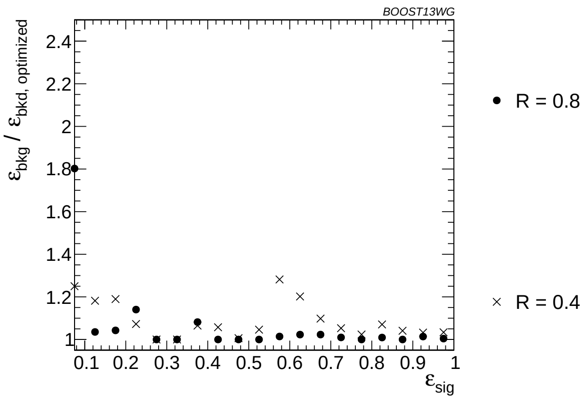
(a) HEPTopTagger m_t



(b) Johns Hopkins Tagger m_t



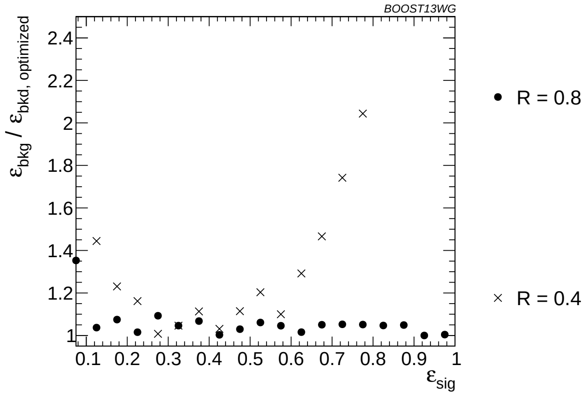
(c) Pruning m_t



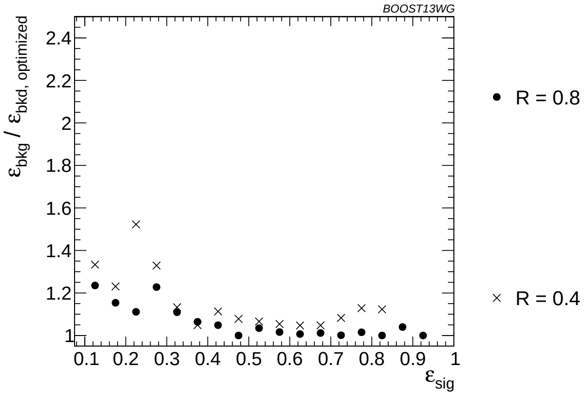
(d) Trimming m_t

Fig. 48 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 $R = 1.2$.

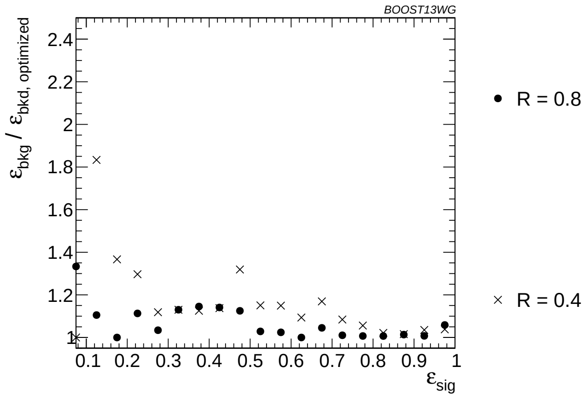
30. S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, *Longitudinally invariant K_t clustering algorithms for hadron hadron collisions*, *Nucl.Phys.* **B406** (1993) 187–224.
31. S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys.Rev.* **D48** (1993) 3160–3166, [hep-ph/9305266].
32. S. D. Ellis, A. Hornig, T. S. Roy, D. Krohn, and M. D. Schwartz, *Qjets: A Non-Deterministic Approach to Tree-Based Jet Substructure*, *Phys.Rev.Lett.* **108** (2012) 182003, [arXiv:1201.1914].
33. S. D. Ellis, A. Hornig, D. Krohn, and T. S. Roy, *On Statistical Aspects of Qjets*, *JHEP* **1501** (2015) 022, [arXiv:1409.6785].
34. S. D. Ellis, C. K. Vermilion, and J. R. Walsh, *Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches*, *Phys.Rev.* **D81** (2010) 094023, [arXiv:0912.0033].
35. D. Krohn, J. Thaler, and L.-T. Wang, *Jet Trimming*, *JHEP* **1002** (2010) 084, [arXiv:0912.1342].
36. J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Jet substructure as a new Higgs search channel at the LHC*, *Phys.Rev.Lett.* **100** (2008) 242001, [arXiv:0802.2470].
37. A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *Soft Drop*, *JHEP* **1405** (2014) 146, [arXiv:1402.2657].
38. M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, *Towards an understanding of jet substructure*, *JHEP* **1309** (2013) 029, [arXiv:1307.0007].
39. J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with N-subjettiness*, *JHEP* **1103** (2011) 015, [arXiv:1011.2268].
40. A. J. Larkoski, D. Neill, and J. Thaler, *Jet Shapes with the Broadening Axis*, *JHEP* **1404** (2014) 017, [arXiv:1401.2158].
41. A. J. Larkoski and J. Thaler, *Unsafe but Calculable: Ratios of Angularities in Perturbative QCD*, *JHEP* **1309** (2013) 137, [arXiv:1307.1699].
42. A. J. Larkoski, G. P. Salam, and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **1306** (2013) 108, [arXiv:1305.0007].
43. **CMS Collaboration** Collaboration, S. Chatrchyan et al., *Search for a Higgs boson in the decay channel H*



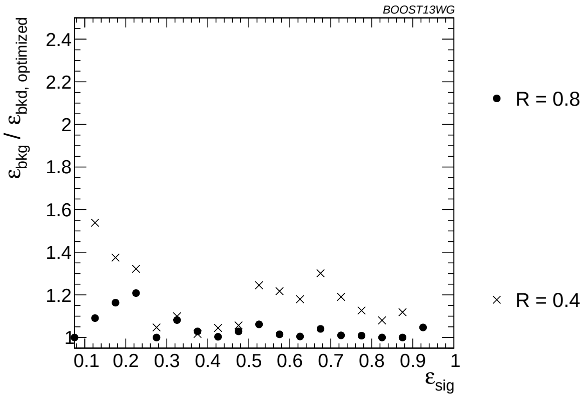
(a) HEPTopTagger



(b) Johns Hopkins Tagger

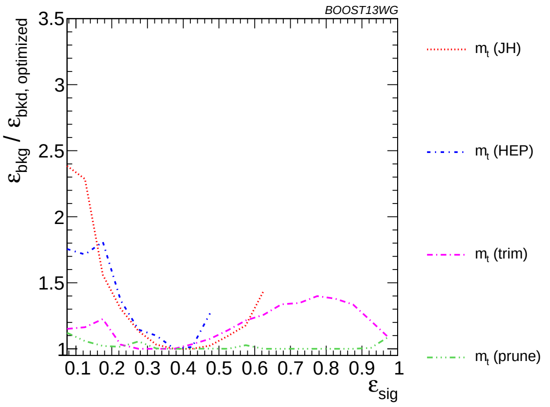


(c) Pruning

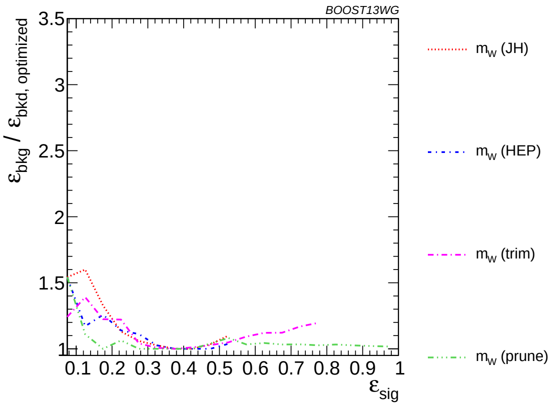


(d) Trimming

Fig. 49 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 $R = 1.2$.

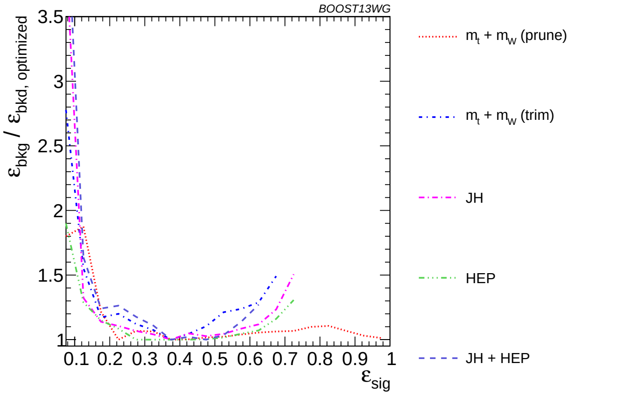


(a) Top mass

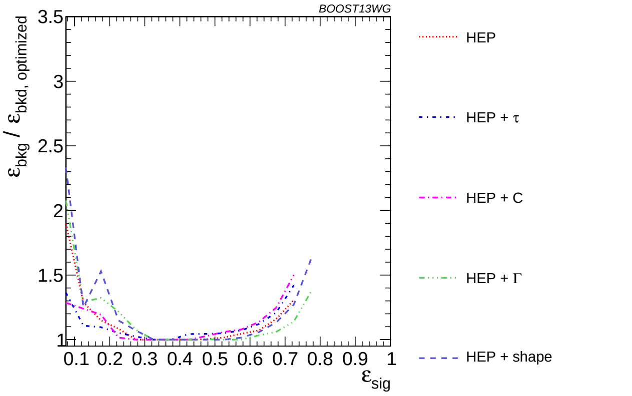


(b) W mass

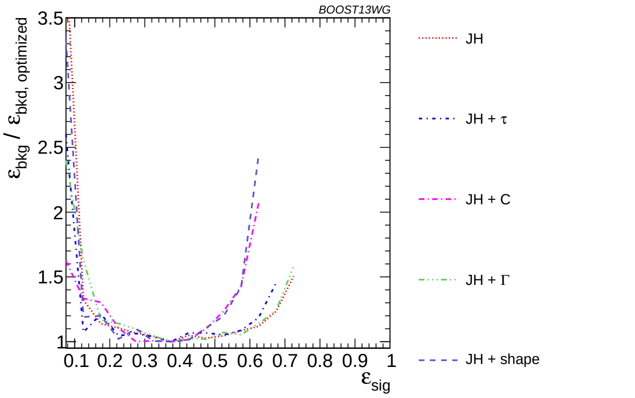
Fig. 50 Comparison of single-variable top-tagging performance in the $p_T = 1-1.1$ GeV bin using the anti-k_T, R=0.8 algorithm; the inputs for each tagger are optimized for the $\varepsilon_{\mathrm{sig}} = 0.3-0.35$ bin.



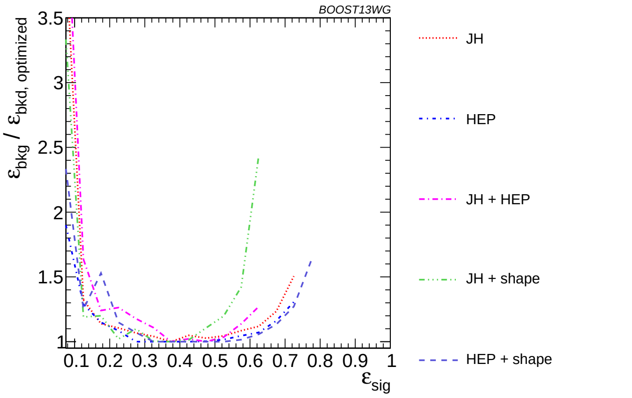
(a) Tagger-Groomer comparison



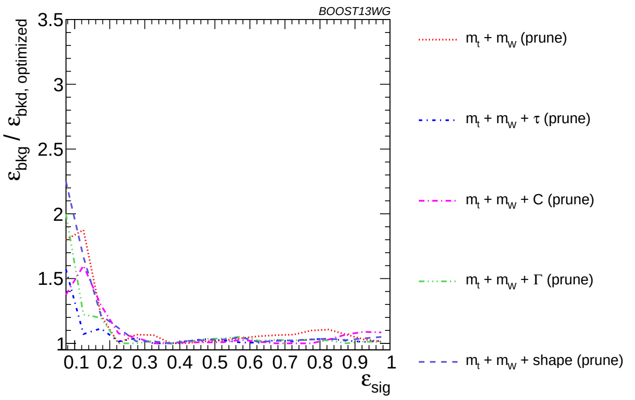
(b) HEPTopTagger + Shape



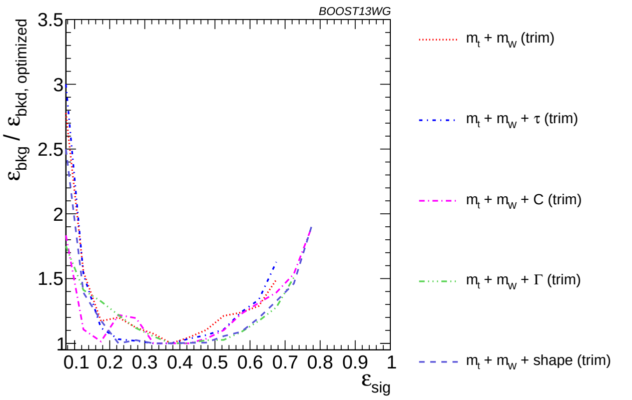
(c) Johns Hopkins Tagger + shape



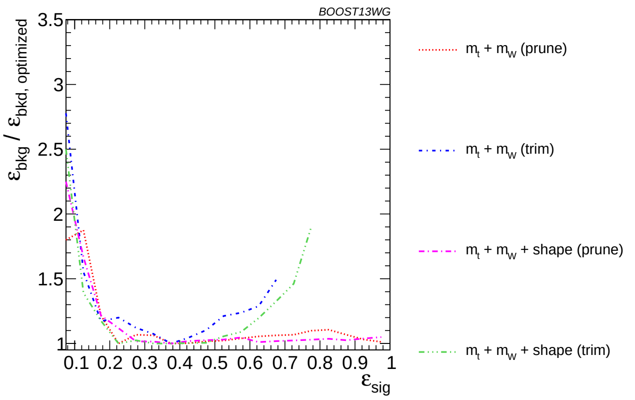
(d) HEP vs. JH comparison (incl. shape)



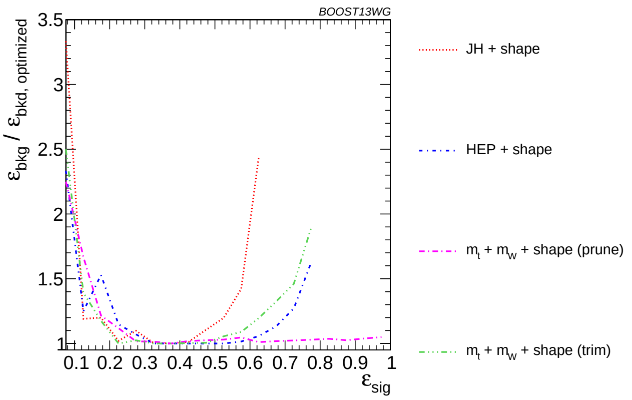
(e) Pruning + Shape



(f) Trimming + Shape



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



(h) Comparison of all Tagger+Shape

Fig. 51 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: $\tau_{21}^{(\beta=1)} + \tau_{32}^{(\beta=1)}$, $C_2^{(\beta=1)} + C_3^{(\beta=1)}$, $\Gamma_{\rm Qjet}$, and all of the above (denoted "shape"). The inputs for each tagger are optimized for the $\varepsilon_{\rm sig} = 0.3-0.35$ bin.