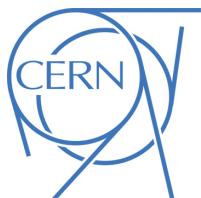




ATLAS Note

KIP-2016



Draft version 0.1

1

2 Jet Observables using Subjet-assisted Tracks

3

The ATLAS Collaboration¹, Oleg Brandt^a, Sascha Dreyer^a, Fabrizio
4 Napolitano^a

5

^a*Heidelberg University*

6

15th December 2016

7

This note presents the details of the Monte-Carlo studies on the subjet-assisted observables for groomed large-radius jet. In particular the observables for the Energy Correlation Functions and n-Subjettiness variables used by the ATLAS collaboration, C_2 , D_2 , τ_{21} and τ_{32} are discussed using subjet-assisted tracks; the mass observable constructed with this technique, m^{TAS} , is presented and discussed with a modified four-momentum prescription. In all the variables studied, large improvement have been found using this novel techniques, the first ones evaluating in terms of QCD event rejection in W/Z boson, top quark and Higgs boson tagging; the second one in terms of precision reconstruction of the large-radius jet mass.

15

© 2016 CERN for the benefit of the ATLAS Collaboration.

Reproduction of this article or parts of it is allowed as specified in the CC-BY-4.0 license.

16 **Contents**

17	1 Introduction	2
18	2 ATLAS detector	2
19	3 Monte Carlo Samples	2
20	4 Object Definition	3
21	4.1 Large-radius jet mass definitions	4
22	4.1.1 Substructure: Grooming Techniques	4
23	4.1.2 Trimming	4
24	4.1.3 Calorimeter Mass	5
25	4.1.4 Track Mass	5
26	4.1.5 Performance Figure of Merit (FoM)	7
27	4.1.6 Gaussian Fit	7
28	4.1.7 InterQuantile-Range	8
29	4.2 Energy Correlation Functions	9
30	4.3 n-Subjettiness	10
31	4.4 Receiver Operator Characteristics	11
32	5 Track-assisted subjet mass	11
33	5.1 Track-Assisted Mass (m^{TA})	12
34	5.2 Advantages and Limitation of m^{TA}	13
35	5.2.1 Performance in $W \rightarrow q'\bar{q}$ Decays	13
36	5.2.2 Performance in $t \rightarrow q'\bar{q}b$ Decays	14
37	5.2.3 Performance in $h \rightarrow b\bar{b}$ Decays	14
38	5.3 The Track-Assisted Sub-jet Mass (m^{TAS})	15
39	5.4 Observable Definition: Inputs	16
40	5.4.1 Tracks	16
41	5.4.2 Sub-jets	16
42	5.5 Observable Definition: Procedure	17
43	5.6 Performance in $W \rightarrow q'\bar{q}$ Decays	18
44	5.7 Performance in $t \rightarrow q'\bar{q}b$ Decays	18
45	5.8 Performance in $h \rightarrow b\bar{b}$ Decays	18
46	5.9 Performance in QCD Multijet Events	19
47	5.10 Performance in Massive $\tilde{W} \rightarrow q'\bar{q}$ Decays with $m_{\tilde{W}} = m_t$	21
48	5.11 Other Stability Quantifiers	21
49	5.12 Sub-jet Calibration	22
50	5.13 Preliminary Studies on Sub-jet Calibration	22
51	5.13.1 Perfect Calibration	23
52	5.13.2 Simple Sub-jet Calibration	23
53	5.14 Limitation of m^{TAS}	24
54	5.15 Alternative Observable Definitions	26
55	6 Combining the mass observables	27
56	6.1 Combination $m^{TA} - m^{calo}$	27
57	6.2 Combination $m^{TAS} - m^{calo}$	28

58	6.2.1	Procedure	28
59	6.3	Performance in $W \rightarrow q'\bar{q}$ Decays	29
60	6.4	Performance in $t \rightarrow q'\bar{q}b$ Decays	29
61	6.5	Performance in $h \rightarrow b\bar{b}$ Decays	30
62	7	Energy Correlation Functions and n-Subjettiness	30
63	7.1	Sub-jet track assistance	30
64	7.1.1	Event weighting and Mass-Cut	31
65	7.2	Track Selection	33
66	7.3	Performance with default β	34
67	7.3.1	Performance for W boson tagging	34
68	7.3.2	Un-assisted tracks and TAS at very high p_T	34
69	7.3.3	Correlation with p_T	36
70	7.3.4	Performance for Higgs boson tagging	37
71	7.3.5	Performance for Top quark tagging	39
72	7.4	Optimisation of β	39
73	7.4.1	Optimisation for W boson jets	40
74	7.4.2	Optimisation for Higgs boson jets	43
75	7.4.3	Optimisation for Top quark jets	45
76	8	Conclusions & Outlook	48
77	8.1	Jet mass observables	48
78	8.1.1	Outlook	48
79	8.2	Energy Correlation Functions and n-Subjettiness	49
80	Appendix		50
81	A	m^{TAS} distributions, boosted W/Z	59
82	B	m^{TAS} distributions, boosted tops	71
83	C	m^{TAS} distributions, boosted higgs	82
84	D	m_{TAS}^{comb} response distributions, boosted W/Z	93
85	E	m_{TAS}^{comb} response distributions, boosted tops	96
86	F	m_{TAS}^{comb} response distributions, Higgs	99
87	G	W boson Tagging	102
88	H	Higgs Boson Tagging	102
89	I	Top Quark Tagging	102
90	J	W Distributions	140
91	K	Higgs Distributions	140

94

1 Introduction

95 Jets are collimated streams of particles resulting from quarks and gluons fragmentation and hadronization.
 96 The distribution of energy inside a jet contains information about the initiating particle. When a massive
 97 particle such as a top quark, Higgs boson or W/Z bosons is produced with significant Lorentz boost and
 98 decays into quarks, the entire hadronic decay may be captured inside a single jet. The mass of such jets
 99 (jet mass) is one of the most powerful tools for distinguishing massive particle decays from the continuum
 100 multijet background; the Energy Correlation Functions and n-Subjettiness C2, D2, τ_{21} and τ_{32} provide an
 101 ad-hoc tool pupusely developed for the multijet background and constitue a fundamental part of many for
 102 boson taggers. This note documents the so-called subjet-assisted techniques with the ATLAS detector
 103 *insref*. The track-assisted subjet mass m^{TAS} definition is presented and confronted with the standard
 104 development in ATLAS, m^{comb} and m^{TA} . Energy Correlation Functions and n-Subjettiness with the
 105 modified subjet-assisted technique is presented and confronted with the standard one in ATLAS. The note
 106 ends with conclusions and future outlook in *insref*.

107

2 ATLAS detector

108 ATLAS (A Toroidal ApparatuS) is a multi-purpose particle detector with nearly 4π coverage in solid angle.
 109 A lead/liquid-argon sampling electromagnetic calorimeter is split into barrel ($|\eta| < 1.5$) and end-cap ($1.5 < |\eta| < 3.2$) sections. A steel/scintillating-tile hadronic calorimeter covers the barrel region ($|\eta| < 1.7$)
 110 and two end-cap copper/liquid-argon sections extend to higher pseudo-rapidity ($1.5 < |\eta| < 3.2$). Finally,
 111 the forward region ($3.1 < |\eta| < 4.9$) is covered by a liquid-argon calorimeter with Cu (W), absorber in the
 112 electromagnetic (hadronic) section. Inside the calorimeters there is a 2 T solenoid that surrounds an inner
 113 tracking detector which measures charged particle trajectories covering a pseudo-rapidity range $|\eta| < 2.5$
 114 with pixel and silicon micro-strip detectors (SCT) and additionally which covers the region $|\eta| < 2.0$ with
 115 a straw-tube transition radiation tracker (TRT). Outside the calorimeter there is a muon spectrometer: a
 116 system of detectors for triggering up to $|\eta| < 2.4$ and precision tracking chambers up to $|\eta| < 2.7$ inside a
 117 magnetic field supplied by three large superconducting toroid magnets.
 118

119 A breakdown of the ATLAS sub-detector performance is shown in Table ??.

120

3 Monte Carlo Samples

121 *refrseMT* The samples used are divided into two main groups: SM background and beyond SM signal.
 122 The SM background includes the QCD multijet samples, produced with a falling p_T spectrum. The
 123 beyond SM signals are $W' \rightarrow WZ \rightarrow q\bar{q}'q\bar{q}$, $Z' \rightarrow t\bar{t}$ (top quarks considered in the full hadronic channel
 124 ($t \rightarrow W(\rightarrow q\bar{q}')b$)) and RS-Graviton $\rightarrow hh \rightarrow b\bar{b}b\bar{b}$, i.e. final states have only jets in all the samples.

ATLAS	Description and performance
Magnetic field	2 T solenoid; 0.5 T toroid barrel and 1 T toroid end-cap
Tracker	Inner detector: IBL, Silicon pixel and strips, TRT $\sigma_{p_T}/p_T \simeq 5 \times 10^{-4} p_T \otimes 1\%$
EM calorimeter	EMB, EMEC and pre-sampler (Liquid Argon and lead) $\sigma_E/E \simeq 10\%/\sqrt{E} \otimes 0.7\%$
Hadronic calorimeter	Tile (Fe and scintillating tiles) and HEC (Cu and LAr) $\sigma_E/E \simeq 50\%/\sqrt{E} \otimes 3\%$
Muons	Inner detector and muon spectrometers $\sigma_{p_T}/p_T \simeq 2\% \text{ at } 50 \text{ GeV}$ $\sigma_{p_T}/p_T \simeq 10\% \text{ at } 1 \text{ TeV}$
Trigger	L1 and HLT (L2 and EF) Rates from ~ 40 MHz to ~ 75 kHz (L1) and to ~ 200 Hz (HLT)

Process	ME Generator & Fragmentation	ME PDFs	UE Tune	Resonance Masses
QCD multijet	Pythia 8	NNPDF23LO	A14	N/A
$W' \rightarrow WZ$	Pythia 8	NNPDF23LO	A14	1.5, 2.5, 3, 4, 5 TeV
$Z' \rightarrow t\bar{t}$	Pythia 8	NNPDF23LO	A14	1.5, 1.75, 2.5, 3, 4, 5 TeV
$G_{RS} \rightarrow hh(\rightarrow b\bar{b})$	Pythia 8	NNPDF23LO	A14	0.5, 1, 1.5, 2, 2.5, 3 TeV
$W' \rightarrow WW$ with $m_{\tilde{W}} = m_t$	Pythia 8	NNPDF23LO	A14	1.5, 2.5, 3, 4, 5 TeV

Table 1: Overview of the Monte Carlo Samples used. The first line shows QCD standard model process, the second, the third and the forth the beyond SM samples considered; the last line the “massive W/Z ” sample.

125 The details of the samples are given in Table 1; the masses considered span from 0.5 to 5 TeV to improve
126 and diversify the kinematic space covered.

127 A set of kinematic distributions for the W' is shown in Figure ??: on the left the p_T distribution where
128 the kinks correspond to the Jacobian peak of the mass considered and the η distribution on the right. The
129 green dots represent the distribution before the selection, which is $p_T > 250$ GeV and $|\eta| < 2.0$ and the red
130 dots after this selection. This selection typical for many searches for BSM physics. All the other samples
131 and the background can be found in the Appendix. In what follows, it will also be used the nomenclature
132 *boosted W/Z* for the W' sample, *boosted tops* for the Z' sample, *boosted Higgs* for the G_{RS} sample and
133 *massive W* for the $W' \rightarrow \tilde{W}\tilde{W}$ with $m_{\tilde{W}} = m_t$. *refrseMT*

134 4 Object Definition

135 This section gives an overview of the objects used for the subjet-assisted variables, which are the large-
 136 radius jet mass, the Energy Correlation Functions and the n-Subjettiness, as they are used within ATLAS,
 137 the next section will give the details of the modified approach of the subjet-assisted techniques.

138 4.1 Large-radius jet mass definitions

139 Large-radius jet, or arge- R jets are jets constructed with a radius parameter of the reclustering algorithm
 140 much bigger than the standard 0.4; within ATLAS the size of large- R jets is 1.0 for anti- k_t and 1.2 for C/A
 141 (the area of C/A is $\sim 20\%$ smaller than anti- k_t).

142 It is worth noting that, for a standard anti- k_t 0.4 jet the active area [**antiktalgo**] is $A_{\text{anti}-k_t} = \pi R^2 \simeq 0.5$,
 143 while it is $\simeq 3.14$ for 1.0 jet, i.e. around six times bigger.

144 Already from this “geometrical” point of view, the necessity of further techniques can be understood: the
 145 effect of soft radiation contamination from Pile-Up (PU) and Underlying Event (UE) will be in this case
 146 six times bigger and spoil the efficiency of the jet mass measurements.

147 4.1.1 Substructure: Grooming Techniques

148 This section is based on the 7 TeV article on jet Substructure [**substructure1**]. In order to use large- R
 149 jets, it is necessary to gain additional information on the interior of these objects, i.e. using techniques
 150 that exploit its substructure allowing a jet-by-jet discrimination of the energy deposit most likely coming
 151 from the hard-scattering to other soft radiation.

152 A common feature in substructure is the use of *sub-jet*, i.e. jets obtained from a parent jet (e.g. the large- R
 153 jet), using its constituent but running the jet reclustering algorithm with a smaller radius parameter; in one
 154 large- R jet, typically there are two or more sub-jets depending on the originating process and its p_T .

155 Techniques have been developed, both using sub-jets or directly constituents of a jet, which are referred
 156 to as *grooming* algorithms.

157 Grooming algorithms are designed to retain the characteristic substructure within such a large- R jet while
 158 reducing the impact of the fluctuations of the parton shower and the UE, thereby improving the mass
 159 resolution and mitigating the influence of pile-up.

160 The grooming algorithms presented here are the most important ones in ATLAS: the *Trimming*; other
 161 used as well, the *Split-Filtering* and the *Pruning* can be found in the Appendix.

162 4.1.2 Trimming

163 The trimming algorithm is the most important in ATLAS and the one mainly used in the work presented in
 164 this thesis. It takes advantage of the fact that contamination from soft radiation has a much lower p_T with
 165 respect to the hard-scattering component. Therefore uses a transverse momentum balance to distinguish
 166 among those. The algorithm works on a two-dimensional parameter space: R_{sub} and f_{cut} . The steps are
 167 as follows:

- k_t algorithm (but of course other choices are also possible) is used to create sub-jets with a smaller radius R_{sub} , aiming at separating the soft radiation from the hard one in different sub-jets. Typical choices are 0.2 and 0.3 (0.2 is used as standard);
- for each sub-jet, the ratio f_{cut} of its p_T with the parent jet p_T^{jet} is calculated: if then this ratio is below a certain value, the sub-jet is removed. Standard choice is $f_{cut} = \frac{p_T}{p_T^{jet}} = 0.05$;
- the sub-jets which survived this procedure are the only one which compose the trimmed jet.

The trimming procedure is also explained in Figure 1, an example of performance in simulation with standard parameters is shown in Appendix (Figure ??).

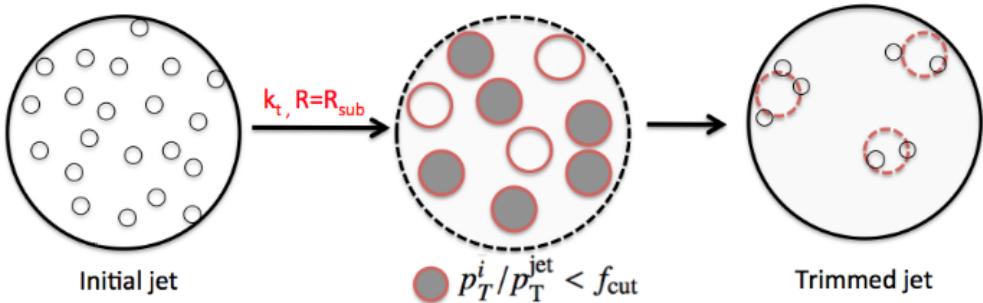


Figure 1: Schematic of the trimming algorithm.

4.1.3 Calorimeter Mass

Once the collection of constituents from the large- R jet is groomed, it is possible to use them for the measure of physical related properties such as the jet mass, since the possible sources of soft radiation from PU and UE have been reduced.

The *calorimeter mass* or m^{calo} is a widely used variable which takes as input the topo-cluster information. Given that each topo-cluster i has a 3D information on the energy deposit, E_i , the mass can be simply calculated from 4-vector properties:

$$m^{calo} = \sqrt{\left(\sum_{i \in J} E_i\right)^2 - \left(\sum_{i \in J} p_{T,i}\right)^2}$$

where J labels the Large- R jet.

4.1.4 Track Mass

This section briefly presents the tracks and their relation with the large- R jet's properties. There are significant advantages and few disadvantages of their usage for precise jet mass reconstruction, which are inherited both from the detector experimental properties and from the underlying physical processes.

185 First of all the excellent performance of track reconstruction and angular separation at low p_T is intrinsically
 186 better than the calorimeter one (see the Chapter 2. and Table ??). The second main advantage is that tracks
 187 can be associated with the primary vertex, thus simply excluding those from PU or other beam-induced
 188 soft radiation background (this is not the case for the UE).

189 The requirement made on tracks to achieve optimal performance are grouped into two categories, the
 190 quality of the track, i.e. if it was fully reconstructed from the detector and separated from others with no
 191 ambiguities, and the association conditions with the primary vertex:

- 192 • $p_T^{track} > 400$ MeV;
- 193 • $|\eta| < 2.5$;
- 194 • Maximum 7 hits in the Pixel and STC sub-detectors;
- 195 • Maximum 1 Pixel hole;
- 196 • Maximum 2 silicon holes;
- 197 • Less than 3 shared modules;
- 198 • Maximum 2 mm of displacement along beam axis (z_0) from the primary vertex;
- 199 • Maximum 2.5 mm of distance in x-y plane from the primary vertex and point of closest approach
 200 (d_0).

201 Given the set of tracks which pass this selection, the mass m^{track} is calculated summing up the 4-momenta
 202 of those tracks which are ghost associated to the groomed jet.

203 Apart from this benefits which derive from the tracker system, there is also an important disadvantage
 204 which comes from the underlying physics: it is completely blind to the electrically neutral component
 205 (mostly π^0) of the jet. As seen in Figure 2, the track mass (red distribution) is not only shifted towards
 206 lower values than the calorimeter mass (green distribution), but its width also degrades.

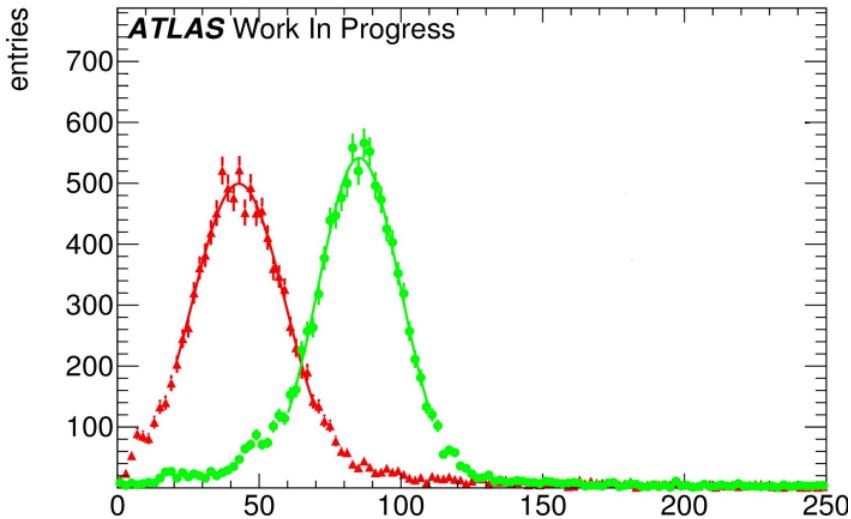


Figure 2: Mass distribution boosted W/Z : in green the m^{calo} and in red the m^{track} .

207 Tracks could be used either for independent mass reconstruction (and in this section is shown how this is
 208 not the case), or, most importantly, as an ulterior information to the calorimeter measurement.

209 **4.1.5 Performance Figure of Merit (FoM)**

210 Since we already introduced the calorimeter and track mass, a concrete, quantitative feature has to be
 211 defined in order to understand which observable is “better”, in the sense that we would prefer one or the
 212 other according to this criterion. This is often referred to as *Figure of Merit* or simply FoM.

213 There are few ways to look at the FoM: one can e.g. naively think about the mean of the mass distribution,
 214 since closer values of the mean to the e.g. W or Z mass (if we are speaking about W/Z decays), indicate a
 215 more correct mass reconstruction. However, this does not take into account the width of this distribution,
 216 as a large width spoils the reconstruction in terms of percentage of jets misreconstructed. Moreover, the
 217 mean is not as important since it can be rescaled to the desired value in a calibration procedure.

218 **4.1.6 Gaussian Fit**

219 The important feature to keep in mind, in fact, is the underlying physics which brings us to calculate the
 220 mass of a jet. In figure 3 this is made clear: if the width of the invariant mass distribution of the jet is
 221 smaller (highlighted), it allows a bigger background rejection, here shown as the QCD dijet, and a higher
 222 signal efficiency, by means of a simple mass requirement.

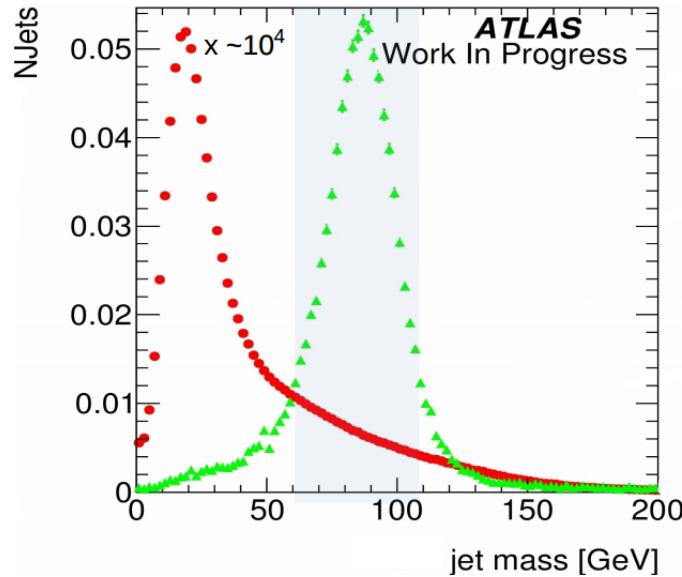


Figure 3: Mass distributions: in red the QCD dijet background rescaled, in green the W/Z from the W' sample.
 Highlighted the width of the W/Z distribution.

223 The width σ of the distribution, which can be obtained from a fit to the Gaussian core, is already a valid
 224 FoM, which has an underlying physical feature. Moreover, in order to be independent from the mean of
 225 the distribution, the width can be divided by the mean itself. This was in fact the FoM which was used
 226 at the beginning of the work for this thesis, since it provided a simple and fast solution. However, special

care must be used both in the procedure of fitting Gaussian cores of responses, since they are asymmetric, and to how the tails are treated.

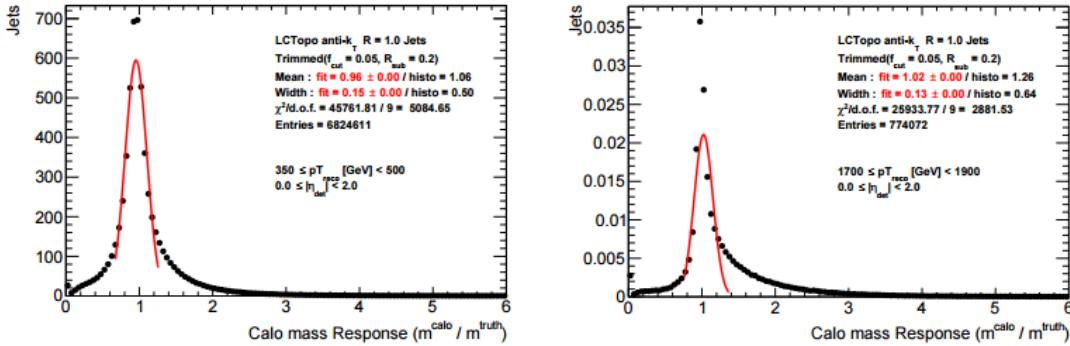


Figure 4: Mass Response distributions for the QCD multijet for various p_T ranges: on the right the failure of the Gaussian fit shows the limitation of this approach to evaluate the Figure of Merit. On the plot the fit parameters and transverse momentum ranges.

The situation is depicted e.g. in Figure 4, where a mass response is shown for calorimeter mass for QCD multijet: here the presence of a right-handed tail which enhances going from low to high transverse momenta makes the Gaussian fit clearly not the tool which provides the stability needed. The ideal tool should take care of managing the presence of at least tails outside the Gaussian core and should converge to the intuition of the standard deviation for a perfect Gaussian distribution. The closest tool to this idea was found to be the *InterQuantile Range*, which was therefore preferred and presented in the next section.

4.1.7 InterQuantile-Range

Another way to look at the mass FoM is half of the 68% of the InterQuantile range (IQnR) (here defined such as it corresponds to a sigma of a “perfect” Gaussian distribution: $q84\% - q16\%$ where $q84\%$ is the 84th percentile and $q16\%$ is the 16th, not to be confused with the InterQuartile Range (IQR) which is the $q75\% - q25\%$ and does not correspond to the sigma) divided by the Median ($\frac{1}{2} \times 68\% \text{ IQnR}/\text{median}$). It provides stability and high sensitivity to left-hand-side and right-hand-side tails.

Another important FoM, used for the work in this thesis, is the response distribution: given the reconstructed mass (calorimeter, track or whichever method) one can compare it to its *truth* mass (m^{truth}), computed from the particle at MC level before the interaction with the detector:

$$R_m = \frac{m^{reco}}{m^{truth}}$$

Standard descriptor of the FoM e.g. in [art35] and here is the IQnR of the R_m .

In Figure 5 a mass response for a single range of transverse momentum is shown, for the calorimeter mass. On the plot the contours of a standard deviation and of $q16\%$ and $q84\%$ are drawn with dashed and solid lines, respectively, showing the difference induced by the tail. This sort of plot is the key when looking quantitatively to the observable performance and can be found in the Appendix for each of the process studied in every p_T range considered. In this chapter will be shown, however, the quantity which describes

250 this FOM, the IQnR, as a function of p_T , in order to get an understanding of the behavior in the entire
 251 spectrum and assure the exclusion of local sub-optimalities.

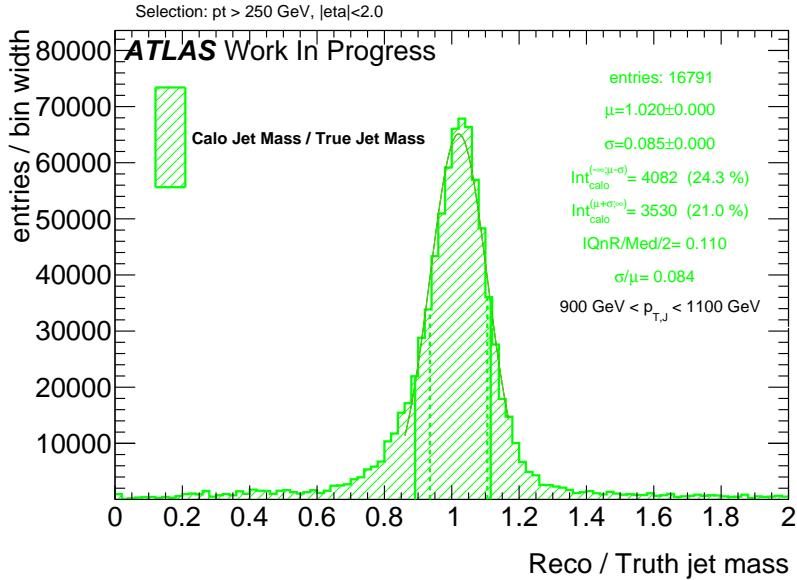


Figure 5: Calorimeter mass response plot for boosted W/Z . On the plot, right, are shown: the number of entries, the mean and the width of the fit to the Gaussian core, the integral from 0 to $\mu - \sigma$ and the one from $\mu + \sigma$ to $+\infty$, the values $\frac{1}{2} \times 68\%$ IQnR/median and σ/μ . On the distribution the dashed vertical lines represent the points $\mu - \sigma$ and $\mu + \sigma$ and the solid lines represent the $q16\%$ and $q84\%$.

252 4.2 Energy Correlation Functions

253 Information about the substructure of large-R jets can be used to discriminate between different event
 254 topologies. These are one, two and respectively three hard substructures (or prongs) inside the large-R
 255 jet. QCD jets are characterized by one hard substructure, jets originated by W or Z bosons feature two
 256 and Top quark jets feature three substructures (hadronic decay channels).

257 The ENERGY CORRELATION FUNCTIONS ECF(N, β) or N -point correlators, described in Reference [bib:ECF],
 258 explore the substructure of a jet using a sum over the constituents. The correlation between pairs and
 259 triples of constituents is considered by the product of their p_T , multiplied by the angular weighting, which
 260 is defined by the product of the pairwise angular distances of the considered constituents. This angular part
 261 can be scaled against the momentum part via an exponent β . The default value for β is 1, corresponding
 262 to angular and momentum parts being weighted equally.

$$\begin{aligned}
 \text{ECF1} &= \sum_{\text{constituents}} p_T \\
 \text{ECF}(2, \beta) &= \sum_{i=1}^n \sum_{j=i+1}^n p_{T,i} p_{T,j} \Delta R_{ij}^\beta \\
 (\text{ECF}(3, \beta)) &= \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n p_{T,i} p_{T,j} p_{T,k} (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^\beta
 \end{aligned} \tag{1}$$

263 The ECF(N) variables can be expanded straightforwardly to larger values of N by considering this
 264 definition. With this, ECF(2) uses pairwise correlation and is sensitive to two-prong structures, whereas
 265 ECF3 relies on triple-wise correlations to identify three-prong structures. ECF(1) corresponds to the p_T
 266 of the whole jet by a summation over the constituents p_T , thereby serving as normalization to minimize
 267 the energy scale dependence.

268 The ECF(N) variable tends to very small values for collinear or soft configurations of N constituents and
 269 is defined to be zero for jets with less than N constituents. For ECF(2), only pairs of constituents that
 270 are angular separated but not soft result in sum terms that are non-negligible, which directly leads to the
 271 picture of two hard substructures inside the jet. A similar conclusion can be made for ECF(3) and three
 272 hard substructures. Resulting from this, a jet with N or more hard substructures features a high ECFN
 273 value while a jet with fewer than N substructures has a lower ECF(N) value. Consequently, one can define
 274 ratios of Energy Correlation Functions. Two of them, called C2 and D2 are found to be very powerful to
 275 distinguish between one- and two-prong like jets, see e.g. Reference [bib:power_counting].

$$\begin{aligned} C2 &= \frac{\text{ECF}(3) \cdot \text{ECF}(1)}{\text{ECF}(2)^2} \\ D2 &= \frac{\text{ECF}(3) \cdot \text{ECF}(1)^3}{\text{ECF}(2)^3} \end{aligned} \quad (2)$$

276 E.g. a jet originated from a W boson features a small ECF(3) but a high ECF(2) value resulting in small
 277 C2/D2, corresponding to a high agreement with the two-prong hypothesis. QCD jets feature a very small
 278 ECF(3) and a small ECF(2) value. This results, considering the power of ECF(2) in the definitions, in a
 279 higher C2/D2 value as for a W boson jet. These variables are IRC-safe for $\beta > 0$ and theoretically very
 280 well understood, see Reference [bib:analytic_ECF]. D2 was found to perform slightly better for tagging
 281 W boson jets as C2 in Reference [bib:w_tagging], most notably due to a more p_T robust cut value and a
 282 somewhat higher background rejection.

283 4.3 n-Subjettiness

284 The n-Subjettiness variable τ_N , introduced in Reference [bib:nsub], quantifies the level of agreement
 285 between a given large-R jet and a certain number N of sub-jet axes. Several possibilities to define the
 286 sub-jet axes exist. Two often used definitions are k_T -axes and the k_T -WTA (Winner Takes All) definition.
 287 In both cases, the jet is reclustered with an exclusive k_T -algorithm, that is running the recombination just
 288 until N sub-jets are clustered. The k_T -axes are defined by the four-momenta of the k_T -sub-jets, WTA
 289 correspond to the four-momentum of the hardest constituent in each k_T -sub-jet. Used in this study is the
 290 k_T -WTA axis definition.

291 As C2 and D2, N-Subjettiness is a measure for the whole jet, calculated via a sum over the jets constituents
 292 (calorimeter clusters as default).

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})^\beta \quad (3)$$

293 For each term, the constituents p_T is multiplied by the distance to the nearest sub-jet axes. The overall
 294 value is normalized with a sum over the constituents p_T times the characteristic radius parameter R of the
 295 large jet.

$$d_0 = \sum_k p_{T,k} R_0 \quad (4)$$

296 Similar to ECF(N, β), the angular measure ΔR_{ij} can be scaled relative to the p_T factor via the exponent β .
 297 N-Subjettiness is an IRC-safe variable for values of $\beta \geq 0$.

298 Small values of τ_N correspond to a jet with all constituents more or less aligned or near to the given
 299 N sub-jet axes, hence the jet is compatible with the assumption to be composed of N or fewer sub-jets.
 300 A higher value in contrast indicates a consistency with more than N sub-jets as a non negligible part is
 301 located apart of the N sub-jet axes. Consequently, W/Z or Higgs boson jets are likely to feature a small
 302 τ_2 and a high τ_1 value. QCD jets with their one-prong structure result in a high τ_2 and a small τ_1 value.
 303 While τ_1 and τ_2 alone provide only slightly separation, the ratio

$$\tau_{21} = \frac{\tau_2}{\tau_1} \quad (5)$$

304 is an effective discrimination variable.

305 The extension to three-prong like jet identification and discrimination from one and two-prong structures
 306 follows quite naturally by taking the ratio of τ_3 and τ_2 .

$$\tau_{32} = \frac{\tau_3}{\tau_2} \quad (6)$$

307
 308 Consequently, the hadronic decay of top quarks via $t \rightarrow Wb$ and the W decaying into two quarks can be
 309 tagged using the τ_{32} variable.

310 4.4 Receiver Operator Characteristics

311 The separation power of discrimination variables can be studied quite intuitively by comparing the signal
 312 and background distributions of a certain variable. Another used figure of merit for the performance,
 313 especially for comparisons of different variables, is to use RECEIVER OPERATOR CHARACTERISTICS (ROCs)
 314 which show the achieved background rejection for different values of signal efficiency (signal fraction left
 315 after performing a cut). Each point is calculated from the underlying signal and background distributions
 316 by integrating the background distribution from zero ¹ to the point where the desired signal fraction is
 317 achieved. The fraction of background events contained in this region are kept when cutting at this signal
 318 efficiency, hence the inverse of this fraction, $\frac{1}{\epsilon_{background}}$ is an estimate for the background rejection. The
 319 lower the fraction of background events in the region, the better is the achieved exclusion. Accordingly, a
 320 good discrimination variable is represented by a ROC with preferably high values of background rejection
 321 up to high signal efficiencies.

322 5 Track-assisted subjet mass

323 The track-assisted subjet mass takes inspiration from the simpler development which is already imple-
 324 mented within ATLAS, the track-assisted mass which is described briefly below for completeness.

¹ If the signal distribution lies at lower values as the background.

325 **5.1 Track-Assisted Mass (m^{TA})**

326 The main limitation of the calorimeter mass comes from the angular resolution of the topo-clusters, which,
 327 for extreme kinematic regimes, start approaching each other at the point that they hit the granularity of
 328 the detector. The main advantage is that on the contrary the relative energy resolution increases at higher
 329 energies.

330 The tracks instead have a very good angular resolution, but p_T relative resolution degrades linearly with
 331 the transverse momentum.

332 One could then think about creating a variable which exploits the advantages of both and minimizes the
 333 disadvantages. As seen, the track mass is missing the neutral component, i.e. each measurement is missing
 334 the fraction $\frac{\text{neutral}+\text{charged}}{\text{charged}}$, but it could be corrected on a jet-by-jet basis: this leads to the definition of
 335 the *track-assisted mass* (m^{TA}):

$$m^{TA} = \frac{p_T^{\text{calo}}}{p_T^{\text{track}}} \times m^{\text{track}} \quad (7)$$

336 It can be intuitively understood as follows: the term m^{track} has the superior angular resolution, but misses
 337 the neutral component; the ratio $p_T^{\text{calo}}/p_T^{\text{track}}$, representing exactly the $(\text{neutral} + \text{charged})/\text{charged}$
 ratio, “restores” the correct value of the mass back to *charged + neutral*.

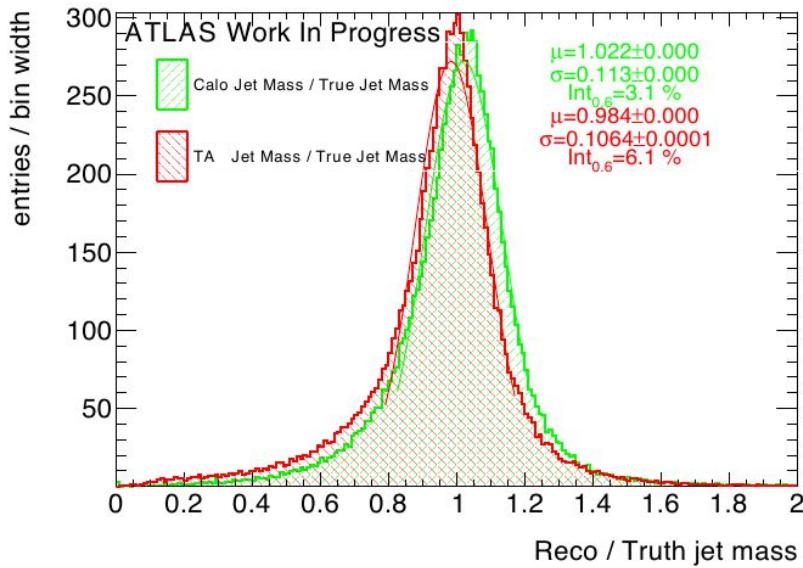


Figure 6: Track-assisted mass response plot for boosted W/Z : in green the calorimeter mass, in red the track-assisted mass. On the right are shown properties of the fit to the Gaussian core; it can be seen that the width of the m^{TA} distribution is smaller, and the mean is slightly below the calorimeter mass.

338

339 From Figure 6 the comparison of the track-assisted mass and the calorimeter mass; the width of the
 340 distribution is smaller, making this observable a good candidate for usage.

341 5.2 Advantages and Limitation of m^{TA}

342 The m^{TA} has a good handle on boosted W/Z , looking at all the transverse momentum spectrum for these
 343 results.

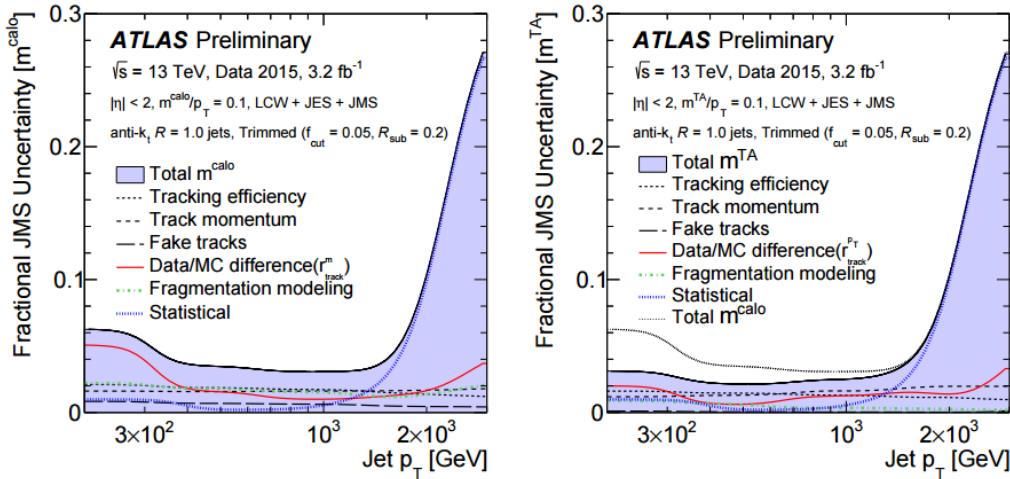


Figure 7: Comparison of the uncertainties for m^{calo} , on the left, and m^{TA} , on the right the rise on the high jet p_T is due to statistics. From the [art35].

344 Another big advantage which supports the use of the track-assisted mass is the relatively small uncertainties:
 345 in Figure 7 the comparison of m^{calo} (left) and m^{TA} (right) fractional uncertainties on the JMS, shows how
 346 the tracking uncertainties are much smaller because of the ratio m^{track}/p_T^{track} . On the right plot the black
 347 line indicates the JMS fractional uncertainty for the m^{calo} , and is always above the m^{TA} . Of course this
 348 introduces another argument in the development of new techniques, which is to look for a good balance
 349 between performance and small uncertainties: a perfect observable in terms of behavior which has very
 350 big uncertainties is not really useful.

351 When looking in the extreme kinematic regime, at very high p_T , as in the top plot in Figure 8, the
 352 m^{TA} shows its real strength, achieving much smaller value of the IQnR. However, there are some severe
 353 limitations which are worth noting, especially looking at the performance in different regions of transverse
 354 momentum: this is shown in the bottom plot of Figure 8, where at a low p_T it exhibits a much worse
 355 behavior.

356 5.2.1 Performance in $W \rightarrow q' \bar{q}$ Decays

357 The performance in all the bins of p_T can be studied looking at Figure 9; these plots have as horizontal
 358 axis the transverse momentum and as vertical one the value of the $\frac{1}{2} \times 68\%$ IQnR/median calculated
 359 from the correspondingly response. For W/Z jets, there is a crossing point around $p_T \sim 1$ TeV, which can
 360 be understood as the point in which the two sub-jet present start merging (sub-jet multiplicity shown in
 361 Figure 47 in Appendix).

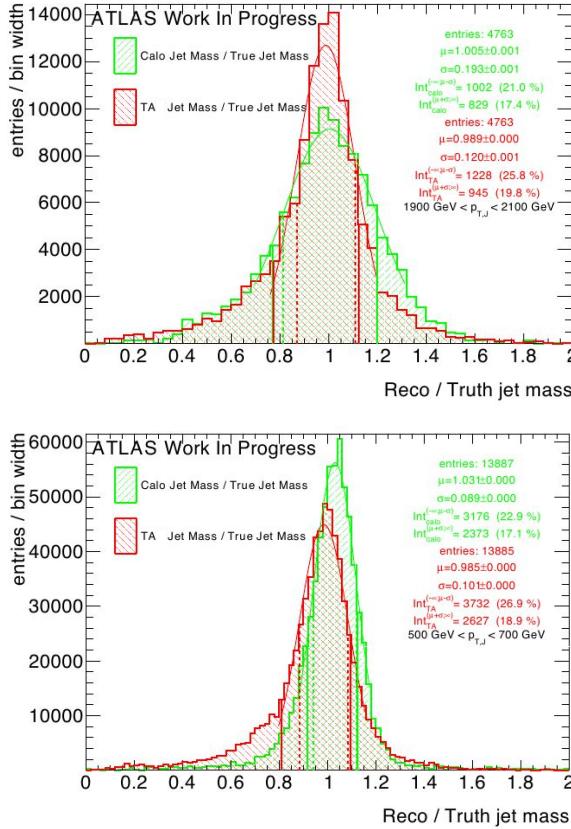


Figure 8: Mass response plots for selected ranges of p_T : on the bottom, a “low” range, $500 \text{ GeV} < p_T < 700 \text{ GeV}$, on the top an high p_T , $1900 \text{ GeV} < p_T < 2100 \text{ GeV}$. A difference in performance can be clearly seen.

362 5.2.2 Performance in $t \rightarrow q' \bar{q} b$ Decays

363 For top quarks the situation is much different: with respect to W/Z jets, in fact, there are two main
 364 disparities: on one side, the mass of the top quark is much higher than the one of the electroweak bosons,
 365 hence making the separation $\Delta R = \frac{2m}{p_T}$ bigger; on the other side, the decay is not anymore two-prong
 366 (two-sub-jet-like) but rather a three-prong (three-sub-jet-like) decay, one from the b -jet and the other two
 367 from the W decay. m^{TA} is here never performing better than m^{calo} , as can be seen e.g. in Figure 9,
 368 right.

369 5.2.3 Performance in $h \rightarrow b\bar{b}$ Decays

370 For boosted Higgs the m^{calo} outperforms the m^{TA} in the spectrum of transverse momentum. Although
 371 the decay is two-pronged, the mass of the Higgs is higher than the electroweak bosons, moreover another
 372 difference lays in light quarks initiated jets and heavy quarks initiated ones, like the b -quarks from Higgs
 373 decay.

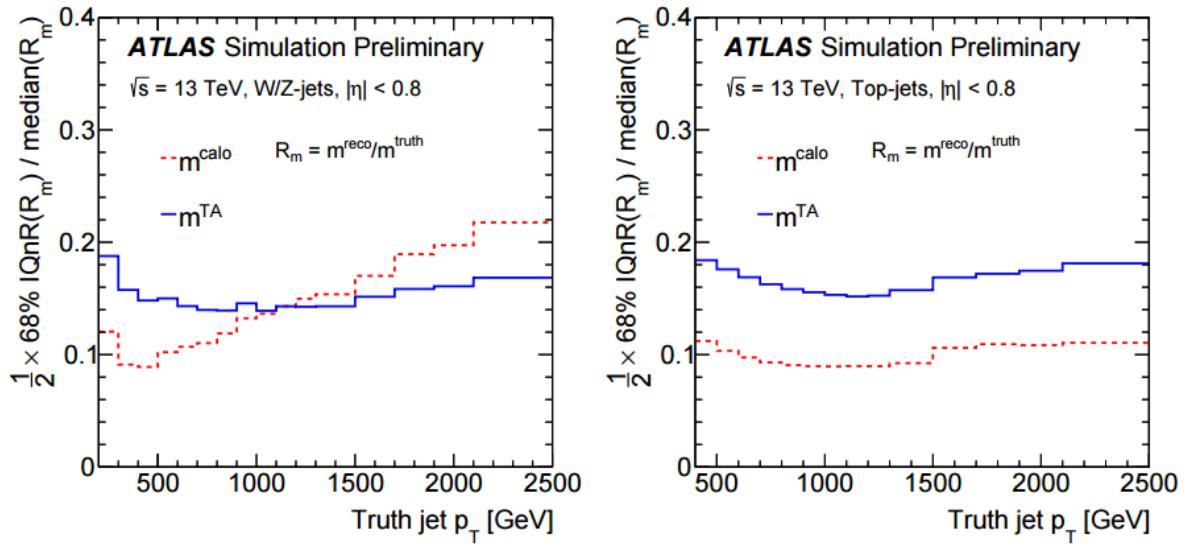


Figure 9: The comparison between the performance of m^{calo} and m^{TA} for W/Z jets (left) and top jets (right); on the x-axis the transverse momentum and on the y-axes the $\frac{1}{2} \times 68\%$ IQnR/median of the mass distribution, from [art35]. A better observable has lower values on the y-axis.

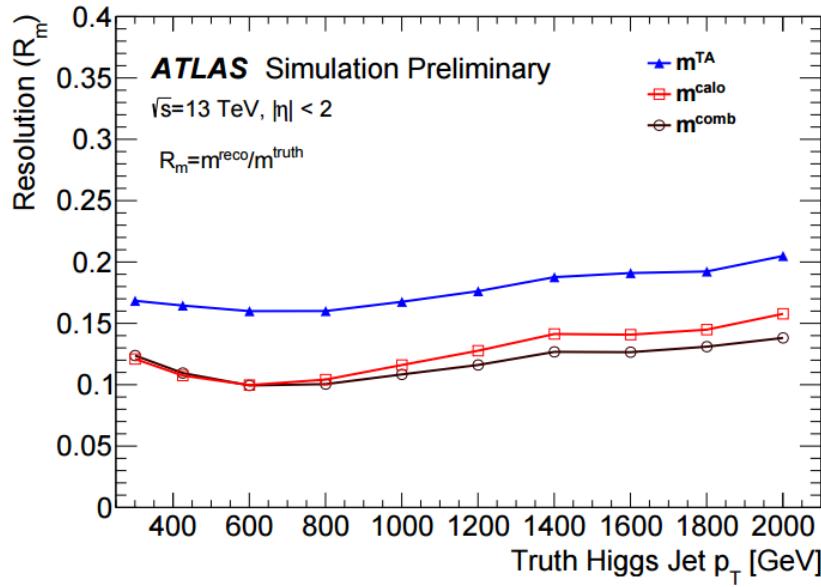


Figure 10: Performance of the m^{TA} with the boosted Higgs sample; the m^{TA} is the blue line, the m^{comb} will be described later in this chapter. From [art39]. The FoM here is the resolution of the Response.

³⁷⁴ 5.3 The Track-Assisted Sub-jet Mass (m^{TAS})

³⁷⁵ In this section the main outcome of the work of this thesis is presented: the *track-assisted sub-jet mass*
³⁷⁶ (m^{TAS}). The main idea takes inspiration from the track-assisted mass: if one can use the tracks to exploit
³⁷⁷ the better angular resolution and correct the missing neutral component jet-by-jet, there is an additional
³⁷⁸ information that can be used. The neutral fraction, in fact, varies stochastically not only per-jet basis, but

379 even per-sub-jet basis, since each sub-jet is originated from a different quark. Correcting the missed neutral
 380 component per-sub-jet, it should perform better already at an intuitive level, as it accesses information
 381 from the jet substructure. There are few question in the definition of this mass observable, whose answers
 382 are in the next section:

- 383 • Regarding the inputs:
 - 384 – How to select the set of tracks to be used?
 - 385 – Which kind of sub-jet should be used?
- 386 • Regarding the procedure
 - 387 – How to associate the tracks to a sub-jet?
 - 388 – How to correct for the missed neutrals on a sub-jet basis?
 - 389 – How to add everything back together?

390 Those details are given in the next subsection.

391 **5.4 Observable Definition: Inputs**

392 There are two inputs to the m^{TAS} : the tracks and the sub-jets. The definition of the standard inputs are
 393 give here; alternative approaches are given in subsection [5.15](#).

394 **5.4.1 Tracks**

395 Only the tracks that satisfy the quality criteria and primary vertex association, described in the previous
 396 section [4.1.4](#), are used. The tracks taken additionally are required to be ghost associated to the sub-jets
 397 of the groomed jet; namely only the sub-jets which survived the trimming procedure and are described in
 398 the next subsection. Ghost association provides a one-to-one correspondence to the sub-jets set, and was
 399 therefore chosen and preferred to other kind of assignments.

400 **5.4.2 Sub-jets**

401 The choice of sub-jets must follow a simple requirement: of course we want to take those which most likely
 402 come from the hard-scattering. This means that the choice of taking them after grooming is forced.

403 As grooming technique used, the trimming was preferred as being the standard in ATLAS and the most
 404 flexible one for optimization studies.

405 The standard version of the trimming uses the k_t reclustering algorithm with radius of 0.2, with the
 406 transverse momentum ratio f_{cut} at 5%.

407 As shown later, this is also the optimal configuration for sub-jets.

⁴⁰⁸ **5.5 Observable Definition: Procedure**

⁴⁰⁹ Having tracks and sub-jets now well defined, we can describe the recipe to produce the m^{TAS} . For brevity
⁴¹⁰ we will call the sub-jets SJ in the formulae below.

⁴¹¹ As said, the tracks are the one ghost-associated to the sub-jets; however, tracks which fall inside the area
⁴¹² of the large- R jet, but not inside the sub-jets area, are still much probably coming from the hard-scattering.
⁴¹³ They are then associated again to the closest sub-jets via ΔR association.

⁴¹⁴ Each sub-jet will have at this point some tracks associated via ghost-association and some other via ΔR
⁴¹⁵ (which are maximally 5%). We call this set of tracks, a “custom” Track-Jet or TJ.

⁴¹⁶ At this point, the one-to-one correspondence is still preserved (for each SJ there is one and only one TJ),
⁴¹⁷ and we can move on correcting the neutral fraction.

⁴¹⁸ Getting inspired from the formula $m^{TA} = p_T^{calo}/p_T^{track} \times m^{track}$, we would like to replicate this at sub-jet
⁴¹⁹ level, i.e.

$$m^{TAS} = \sum_{SJ} \frac{p_T^{SJ}}{p_T^{TJ}} \times m^{TJ}$$

⁴²⁰ Since now we are working inside the sub-jets we need to change the sub-jet’s 4-vector itself and not only
⁴²¹ the mass: if we call p_μ^{TJ} the Lorentz vector of the track-jet,

$$p_\mu^{TJ} = \begin{pmatrix} m^{TJ} \\ p_T^{TJ} \\ \eta^{TJ} \\ \phi^{TJ} \end{pmatrix} \rightarrow p_\mu^{TA} = \begin{pmatrix} m^{TJ} \times \frac{p_T^{SJ}}{p_T^{TJ}} \\ p_T^{SJ} \\ \eta^{TJ} \\ \phi^{TJ} \end{pmatrix}$$

⁴²² where p_μ^{TA} is the track-assisted sub-jet’s 4-vector. If we label i the i -th track-jet of the N ones present in
⁴²³ the large- R jet,

$$m^{TAS} = \sqrt{\left(\sum_i^N p_\mu^{TA} \right)_\mu \left(\sum_i^N p_\mu^{TA} \right)^\mu}$$

⁴²⁴ An important remark is that, in the case of a large- R jet with only one sub-jet, the m^{TAS} has exactly
⁴²⁵ the same definition of the m^{TA} . This implies, since the angular separation of the decay product scales
⁴²⁶ inversely with p_T , that the performance should approach the one of the m^{TA} in the extreme kinematic
⁴²⁷ regime. However, the space for improvement is precisely in the low-middle p_T regime, as seen in the m^{TA}
⁴²⁸ section.

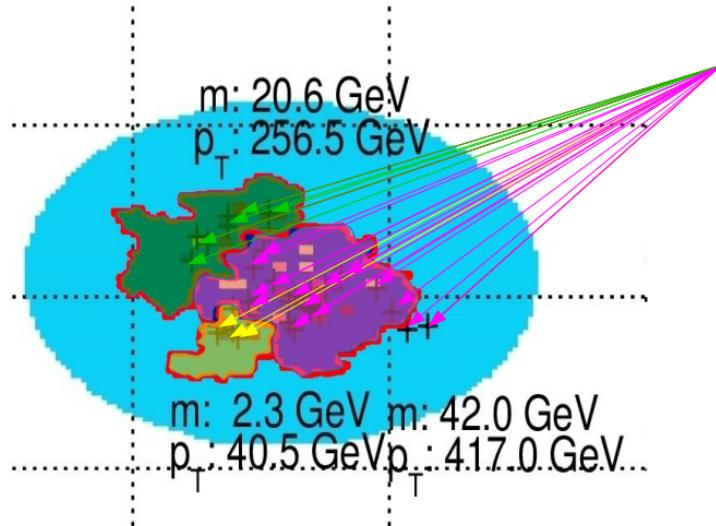


Figure 11: Pictorial event display showing the $\eta \phi$ region of a large- R jet, (in blue the catchment area of the anti- k_t) showing the different k_t sub-jets: they are highlighted in green, fuchsia and yellow. The associated track-jets (here as arrows pointing the calorimeter area) are colored with the same color of the correspondent sub-jet. Some tracks associated with ΔR procedure can be seen in the fuchsia sub-jet. The transverse momenta and mass values are also shown for the sub-jets.

429 5.6 Performance in $W \rightarrow q' \bar{q}$ Decays

430 The boosted W/Z was the first one looked at, and with which the m^{TAS} was designed. The m^{calo} shows a
 431 fast deterioration of the performance at high p_T , and, as shown in the previous section, the m^{TA} prevents
 432 this deterioration but suffers at low transverse momenta ($p_T < 1 \text{ TeV}$). The m^{TAS} has the same behavior in
 433 the extreme transverse momentum regime as the m^{TA} , since the sub-jet multiplicity peaks at one, where
 434 there are no differences between the two observables. In the low- p_T regime, on the contrary, it exploits
 435 the different charged to neutral fluctuation, achieving a better performance. This is shown in Figure 12 as
 436 a function of p_T : below $\sim 1 \text{ TeV}$ it achieves lower values of the IQnR converging from below to the m^{TA}
 437 as the number of sub-jets decreases to one.

438 5.7 Performance in $t \rightarrow q' \bar{q} b$ Decays

439 The boosted tops are shown on Figure 13; the m^{TAS} is comparable yet slightly worse than the m^{calo} in
 440 the low-middle p_T regime, while degrades at higher p_T approaching the m^{TA} , which is far beyond the
 441 track-assisted sub-jet mass in performance. As already noted, the worse performance can be ascribed both
 442 to the higher top-quark mass, and to its different and more complex decay topology.

443 5.8 Performance in $h \rightarrow b\bar{b}$ Decays

444 In the Randall-Sundrum graviton to di-Higgs to four b-quark, the performance is again problematic for the
 445 m^{TA} with respect to m^{calo} , which is far beyond the latter, while the performance of the m^{TAS} is partially

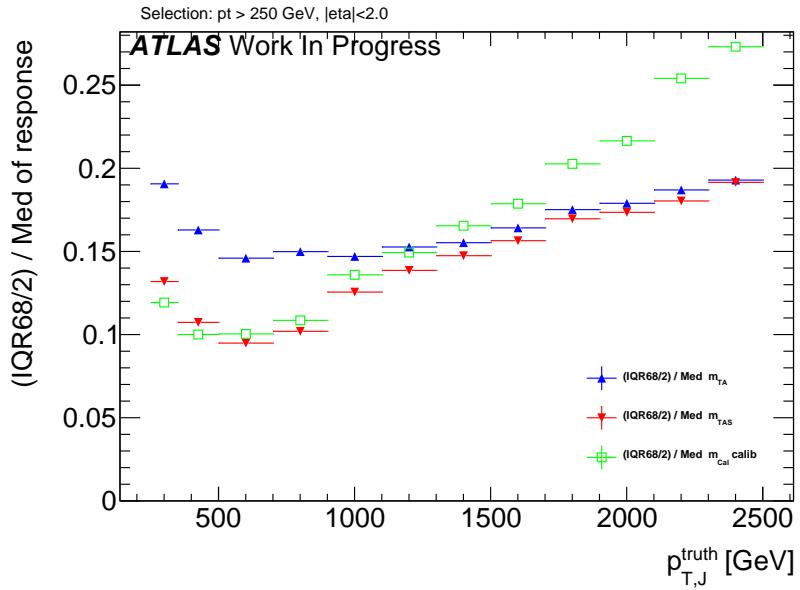


Figure 12: Performance of the m^{TAS} versus the m^{calo} and m^{TA} for the boosted W/Z sample.

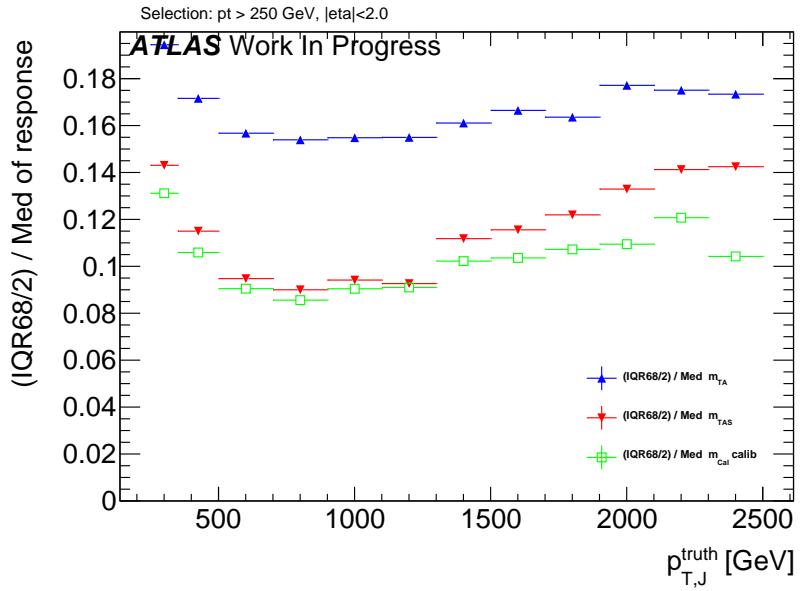


Figure 13: Performance of the m^{TAS} versus the m^{calo} and m^{TA} for the boosted top sample.

similar to the boosted top-quark sample, but degrades much more in the extreme p_T regime, following the m^{TA} . Shown in Figure 14.

5.9 Performance in QCD Multijet Events

The behavior of the QCD multijet sample is similar to the boosted W/Z sample, where the m^{TA} exhibits a crossing point in the middle-low regime $p_T \simeq 900 \text{ GeV}$ and proceeds with a better performance at high

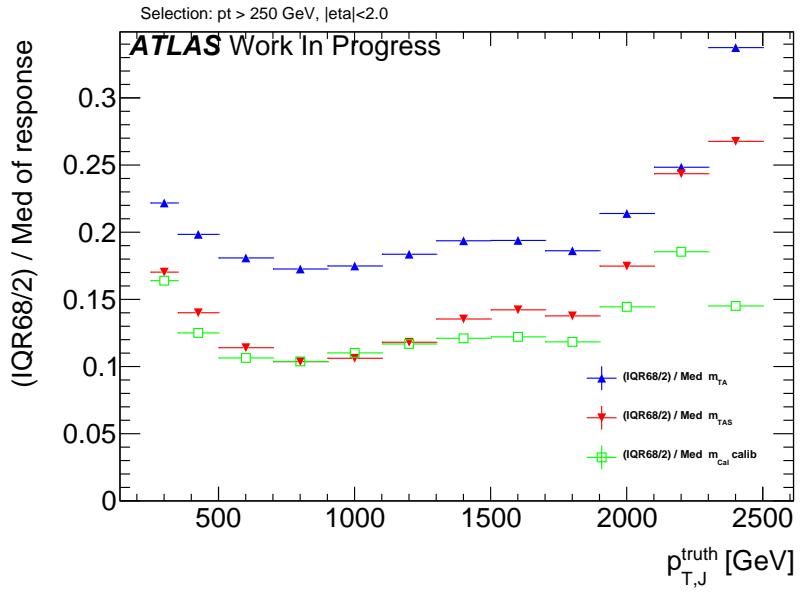


Figure 14: Performance of the m^{TAS} versus the m^{calo} and m^{TA} for the boosted Higgs sample.

transverse momenta. Again the m^{TAS} follows this similarity showing no crossing point and an optimal overall behavior, both with respect to calorimeter- and track-assisted-based mass definition. On Figure 15.

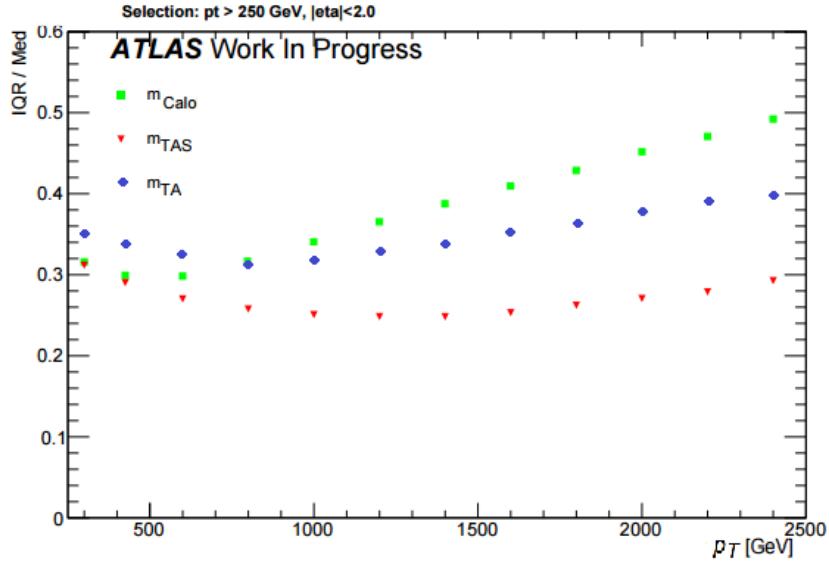


Figure 15: Performance of the m^{TAS} versus the m^{calo} and m^{TA} for the QCD multijet. Here shown IQR/Med not $\frac{1}{2} \times 68\%$ IQR/median.

454 5.10 Performance in Massive $\tilde{W} \rightarrow q' \bar{q}$ Decays with $m_{\tilde{W}} = m_t$

455 The massive W sample is a special sample which was used to understand the behavior of the boosted
 456 tops, whether its worse resolution was coming from the higher mass of the top quark or from the more
 457 complex decay topology (three-pronged instead of two-pronged decay and b-quark presence). The sample
 458 is almost identical to the boosted W/Z one ($W' \rightarrow WZ$) but in this case the SM electroweak boson are
 459 set to have the mass of the top quark $m_{\tilde{W}} = m_t$. In fact, from the rule $\Delta R = 2m/p_T$, a bigger separation is
 460 expected between the quark from the hadronic decay. The comparison with m^{calo} is shown in Figure 16,
 461 together with the boosted top-quark for comparison. As seen here, the performance of the latter is clearly
 462 worse than the former, the trend is yet very similar. This difference is interpreted in terms of different
 463 and more complex topology and hence higher sub-jet multiplicity: in the three sub-jet structure, resolving
 464 accurately the components is more challenging.

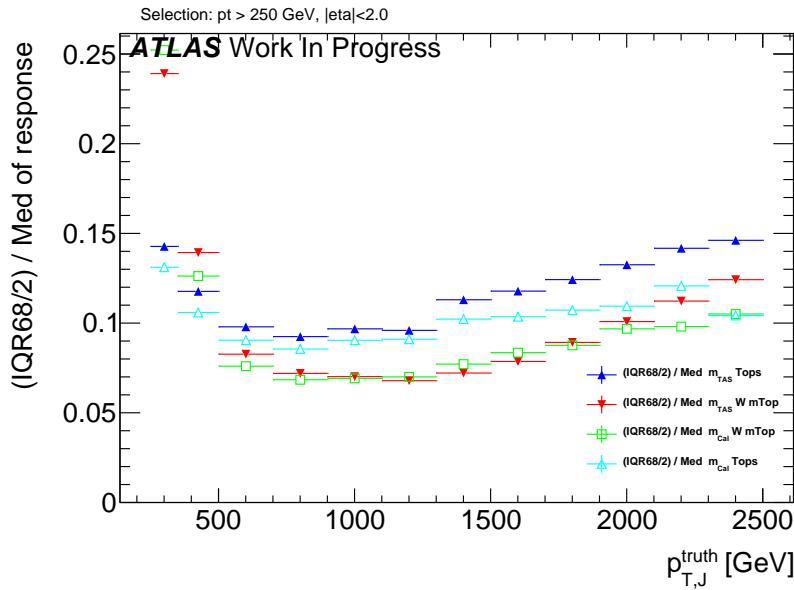


Figure 16: Performance of the m^{TAS} versus the m^{calo} for the massive W/Z (in red and green); shown on the same plot also the boosted top sample (in blue and light blue).

465 5.11 Other Stability Quantifiers

466 The stability of the m^{TAS} was checked, although the IQnR is already a good quantifier of stability, explicitly
 467 for the mean of the mass response distribution and for the left-hand-side tail, as a function of the transverse
 468 momentum. This was an important check to assure the overall gaussianity of the final distribution in the
 469 whole spectrum of p_T , and suitability in regards of the calibration step, which is not discussed in this
 470 thesis.

471 The mean of the response distribution is shown for boosted W/Z decays in Figure 17, left; as seen here,
 472 despite being the mean constantly below the unity, its behavior is much more flat and independent of
 473 p_T , especially in the low-middle regime. This is surprising since the m^{calo} is already shown after the
 474 calibration step, which is not taken instead for the m^{TAS} . Conversely the left-hand-side tail of the mass
 475 response which is shown in the same figure, right, shows a more enhanced behavior than the m^{calo} , but

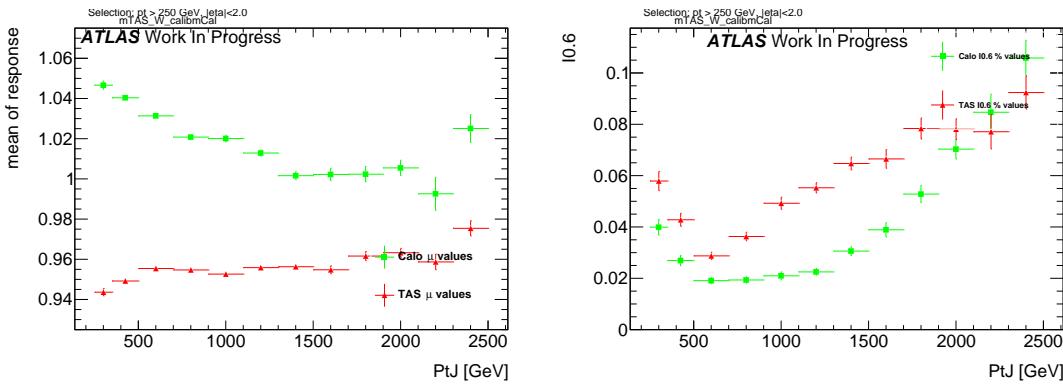


Figure 17: Stability quantifiers which were checked for the m^{TAS} : mean, on the left, and normalized left-hand side integral, on the right, of the mass response distribution. The mean is calculated from a Gaussian fit and the integral goes from 0 to 0.6.

476 still never reaches the 10%. Of course an enhancement of the tail causes a loss of gaussianity and a
 477 number of jets which are reconstructed with a lower mass than they should, but it is still comparable with
 478 the calorimeter mass.

479 Those quantifiers show analogous behavior for the other samples considered and those figures can be
 480 found in the Appendix.

481 5.12 Sub-jet Calibration

482 An additional attempt of calibrating the sub-jet was also tried and, although the results were not sub-
 483 stantially improved, it is presented in this section. This study was performed using only boosted W/Z
 484 samples.

485 5.13 Preliminary Studies on Sub-jet Calibration

486 The first attempt in calibrating the sub-jets had as start a “perfect calibration”, which means using the
 487 truth-level information from the MC sample *before* the interaction with the calorimeter. Truth-level
 488 tracks are the particles in the jet which have an electric charge and are stable, truth-level sub-jets are
 489 all the particles, charged and not, which are ghost associated to the calorimeter sub-jets. There are few
 490 possibilities in doing so, here some nomenclature for this study will be introduced:

- 491 • m^{TAS} using truth-level sub-jets and tracks; normal tracks (with all detector effects) are used to assist
 492 the truth-level sub-jets;
- 493 • m^{TAS} using truth-level tracks and truth-level sub-jets; the truth-level tracks are used to assist the
 494 truth-level sub-jets;
- 495 • m^{calo} truth, calculated using only the truth sub-jets.

496 **5.13.1 Perfect Calibration**

497 The *perfect calibration* refers to the procedure of using m^{TAS} with truth-level sub-jets and track, i.e.
498 looking at the best possible scenario with an ideal detector. The performance is of course expected to be
499 optimal, because of the use of the truth-level. This step was necessary as feasibility study, to understand
500 whether ulterior efforts in this direction were meaningful. The perfect calibration is shown in Figure
501 18; since the performance exhibits room for big improvement below ~ 1 TeV and moderate to small
502 improvement above this value, the second step of a simple calibration was tried.

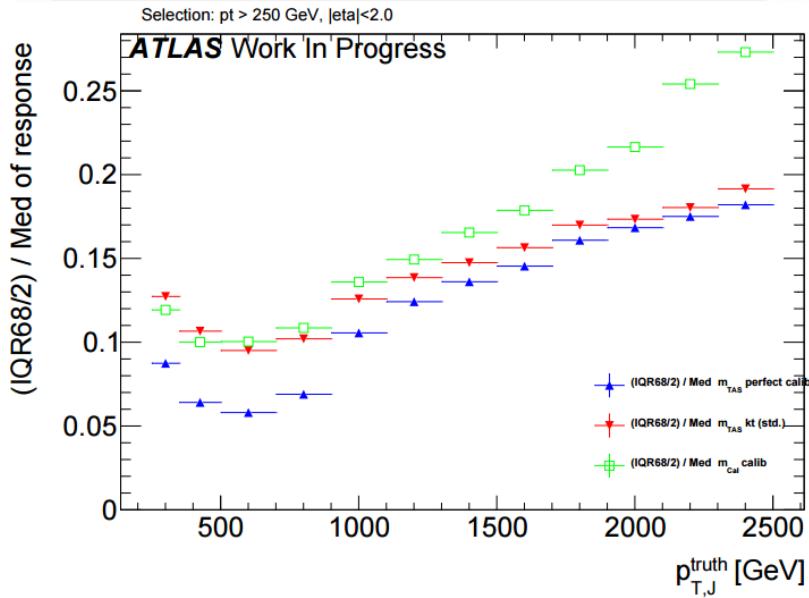


Figure 18: Performance of the perfect calibration. It shows room for improvement especially at low-middle p_T .

503 **5.13.2 Simple Sub-jet Calibration**

504 Following the example of calibration of jets in general, a simple approach to emulate this procedure was
505 tried, constructing in various bins of transverse momenta the responses of the sub-jet's energy to derive
506 the weights factors to be applied. The detailed procedure is as follows:

- 507 1. Responses in energy $R_E = E^{\text{reco}}/E^{\text{truth}}$ were built in several bins of p_T , spanning to the whole
508 transverse momentum range;
- 509 2. The mean μ_R of this response was calculated via a fit to the Gaussian core;
- 510 3. Those values (*scale factors*) were stored and applied again to the sub-jets before the computation of
511 the m^{TAS} via 4-momentum correction $E' = E/\mu_R$; the p_T (the value which only enters the m^{TAS}
512 variable) was changed then correspondingly to keep the sub-jet's mass constant.

513 This procedure was called *poor man's calibration* or PM calibration or *simple calibration*. A check on
514 the p_T response before and after calibration together with the mean of the entire Large- R jet response is
515 shown in Figure 53 and 54 in Appendix.

516 The results are on Figure 19; there are only marginal improvements in few ranges of low transverse
 517 momentum where the scale factors are further away from unity, and the overall observable is not performing
 518 better than the standard m^{TAS} . This is interpreted both in terms of a missing calibration as a function of
 519 the η variables (having hence a befit from the crack region) and because the correction done on average
 520 does not provide the sufficient handle in a jet-by-jet basis, especially when all the sub-jets are rescaled by
 521 similar factors (which translates into a similarity of p_T s of the sub-jets, often the case for e.g. boosted
 522 W/Z , less for boosted tops entirely contained in the large- R jet).

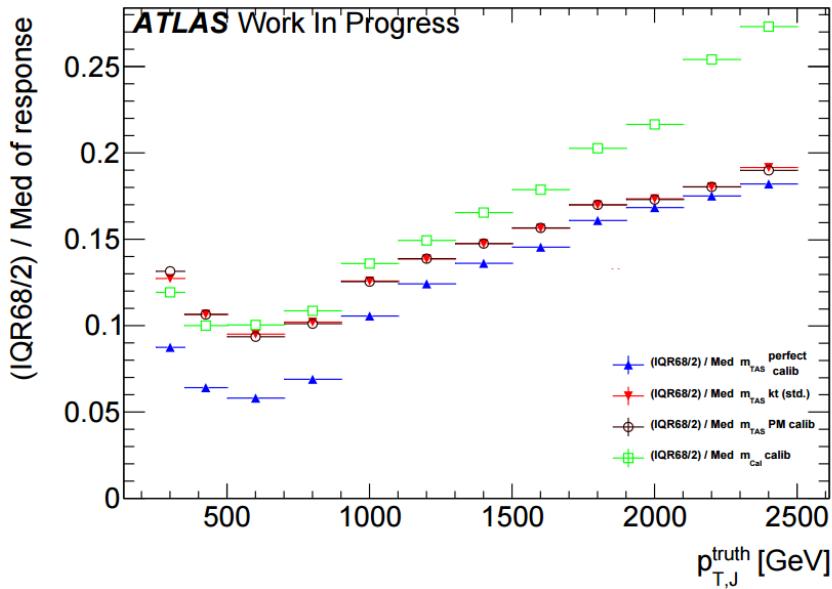


Figure 19: Performance of the poor man’s calibration. The improvement is marginal throughout the entire transverse momentum space.

523 5.14 Limitation of m^{TAS}

524 The final effort to understand the various and competing effects, which take place in the m^{TAS} and which
 525 was inspired by the perfect calibration procedure, brought to a final study on the variable to understand
 526 the reason for the worsening of the resolution at high transverse momenta, using again the truth MC
 527 information.

528 The preliminary investigation in this direction was then the study on the track-resolution: since the track
 529 relative resolution of the transverse momentum is expected to worsen linearly with this variable, a response
 530 of the mass of the tracks was constructed, using the truth-level tracks.

531 The result is shown on Figure 20: for the samples considered, it shows a linear degradation of the mass of
 532 the tracks, both for massive and SM W/Z .

533 The hypothesis of the degradation of the m^{TAS} driven by the tracks is also supported by the Figure 55
 534 in Appendix, where the truth-level tracks are used instead of real tracks to compute the variable; it can
 535 be seen the flat behavior at high p_T , hence ascribing the worsening of the resolution to tracks at higher
 536 transverse momenta.

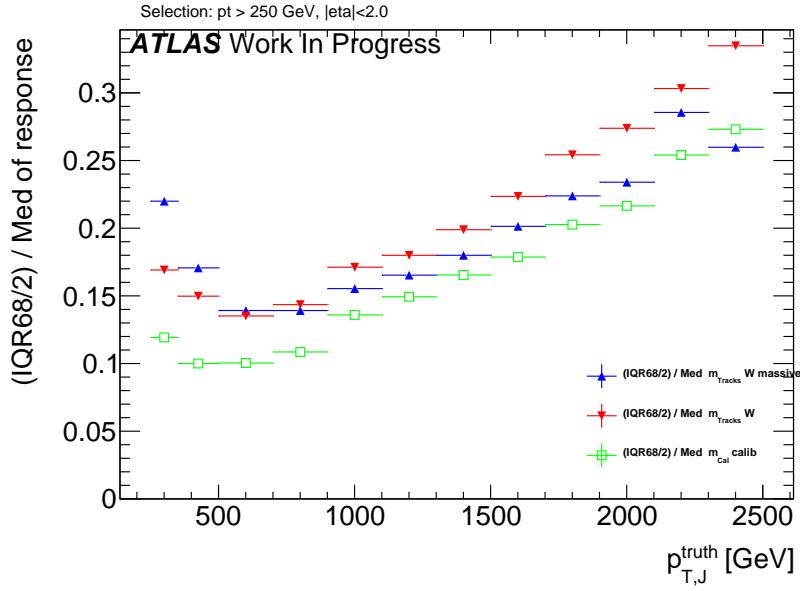


Figure 20: The performance of the track mass in blue and red for massive W sample and boosted W/Z respectively; for reference in green the calorimeter mass of the large- R jet.

537 A complete breakdown of the variable in terms of truth-level particles is given in Figure 21, where all the
 538 different components are separated. In particular the black dots show the m^{TAS} using truth-level sub-jets
 539 but real tracks for the track assistance procedure. Even combining this truth-level information, in fact, it
 540 shows a large worsening of the performance (truth-level sub-jets only are shown as blue dots).

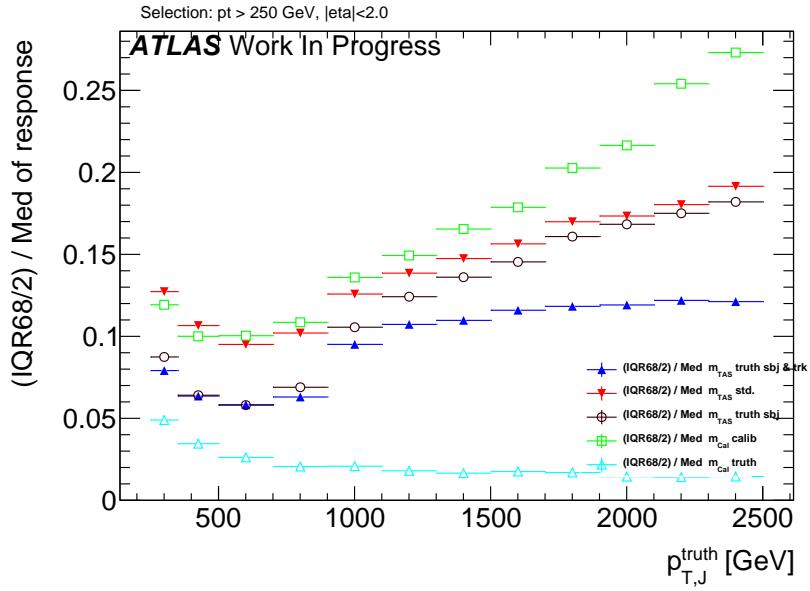


Figure 21: Breakdown of the m^{TAS} in its component using truth-level information for boosted W/Z decays.

541 Other results using truth-level information on boosted tops are shown and described in the Appendix.

542 5.15 Alternative Observable Definitions

543 There are quite a few ways to modify the track-assisted sub-jet mass; however, all the alternative approaches
 544 showed worse performance, and they are mentioned here for completeness only.

545 Alternatives considered were:

- 546 • for the tracks:
 - 547 – use of tracks not as input directly, but only taking those belonging to anti- k_t reclustered
 track-jet with radius of 0.3 or 0.2;
 - 549 – tighter or looser quality conditions were explored;
 - 550 – tighter or looser primary vertex association requirement were explored.
- 551 • for the sub-jets:
 - 552 – the trimming procedure was modified: various radii R_{sub} of the sub-jets were tested;
 - 553 – the sub-jets were reclustered using not only the standard k_t , but also anti- k_t and C/A.
- 554 • for the procedure: different 4-momentum correction scheme was also explored.

555 The different reclustering algorithm choice has a deep impact and was studied in details, since it changes
 556 the topo-cluster added to the sub-jets and the tracks associated to them. The situation is depicted in the
 557 event-display in Figure 22; the display on the left shows the standard choice of k_t , the one on the right
 558 shows the modified approach anti- k_t .

559 In the Appendix, figure 48 49 50 the performance for boosted W/Z , tops and Higgs are shown, respectively.
 560 It can be seen that the k_t algorithm provides the best observable definition, in all the samples considered.
 561 However, the anti- k_t algorithm provides similar performances; this was an important check as the jet
 562 calibration procedure currently going on in ATLAS, the *R-Scan* procedures includes the anti- k_t algorithm
 563 with radius of $R=0.2$ and aims at providing the calibration and uncertainties that could be used directly in
 564 the computation of the m^{TAS} .

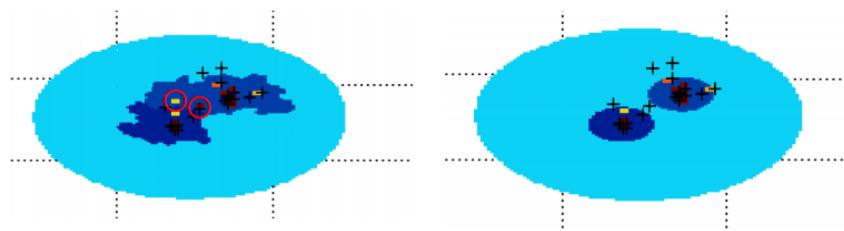


Figure 22: An example of event-display shows the differences in the reclustering algorithm used for the sub-jets: on the right k_t and on the left anti- k_t . Highlighted some constituents trimmed away with the second choice.

565 6 Combining the mass observables

566 Since the calorimeter large- R jet mass is not explicitly used in the track-assisted (sub-jet) mass, it may be
 567 possible to improve the performance creating a new observable which combines both mass definitions.

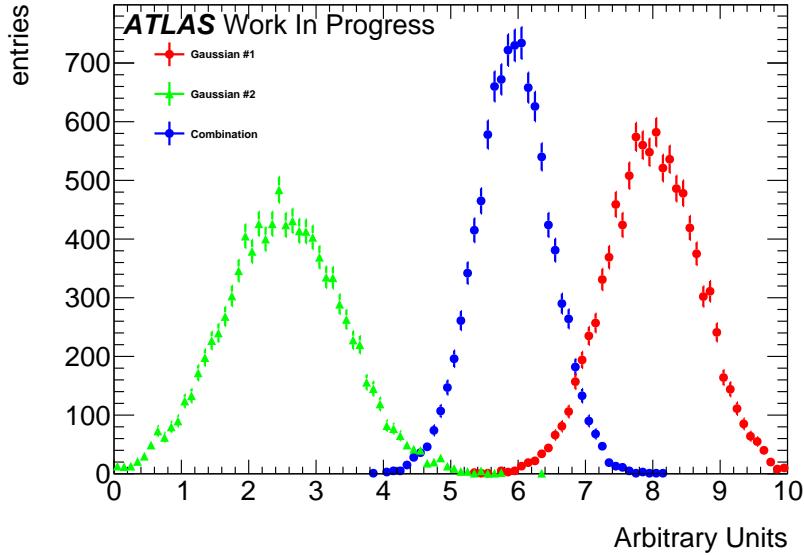


Figure 23: A toy example of the combination of two independent Gaussian observables, in red and green, and their combination, in blue. It can be seen that the combination has a smaller width.

568 This is true for both the m^{TA} and the m^{TAS} ; they are introduced in the next subsections. Provided that the
 569 two observables are nearly independent (correlation coefficient are $\sim 10\%$, see Figure 51 in the Appendix),
 570 due to the Gaussian nature of the p_T and mass response, the optimal combination of the two is linear¹. An
 571 example is provided in Figure 23.

572 6.1 Combination $m^{TA} - m^{calo}$

573 For the $m^{TA} - m^{calo}$ combination the observables are considered nearly independent, then

$$m^{comb} = a \times m^{calo} + b \times m^{TA},$$

$$a = \frac{\sigma_{calo}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} \quad b = \frac{\sigma_{TA}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} \quad (8)$$

574 where σ_{calo} and σ_{TA} are the m^{calo} 's and m^{TA} 's resolution functions. The m^{comb} then is the $m^{TA} - m^{calo}$
 575 combination.

¹ If the joint distribution of the responses is Gaussian, then one can write their probability distribution function as $f(x, y) = h(x, y) \times \exp[A(\mu) + T(x, y)\mu]$, where x is the calorimeter-based jet mass response, y is the track-assisted jet mass response, μ is the common average response, and h, A, T are real-valued functions. This form shows that the distribution is from the exponential family and therefore T is a sufficient statistic. Since the natural parameter space is one-dimensional, T is also complete. Therefore, the unique minimal variance unbiased estimator of μ is the unique unbiased function of $T(x, y) = x/\sigma_x^2 + y/\sigma_y^2$. See e.g. Ref. [statistic] and [art35] for details.

576 **6.2 Combination $m^{TAS} - m^{calo}$**

577 There is a main difference between the m^{TAS} and m^{TA} when it comes to combination: since the m^{TAS} is
 578 using sub-jet level information but m^{TA} not, the correlation with the m^{calo} is expected to be higher. This
 579 can be seen e.g. in the plots in Figure 24 (additional plots shown in Figure 52 in Appendix), where the
 580 correlation is not only higher for the simple W/Z and Higgs jets, but above 50% for tops. The assumption
 581 of independent variables here falls, forcing a more complete approach. The Ansatz is to take into account
 582 the correlation via the formula:

$$m_{TAS}^{comb} = w \times m^{calo} + (1 - w) \times m^{TAS},$$

$$w = \frac{\sigma_{TAS}^2 - \rho \sigma_{calo} \sigma_{TAS}}{\sigma_{calo}^2 + \sigma_{TAS}^2 - 2\rho \sigma_{calo} \sigma_{TAS}} \quad (9)$$

583 where now m_{TAS}^{comb} is the new $m^{TAS} - m^{TA}$ combination. This expression reduces then to the form:

$$m_{TAS}^{comb} = a \times m^{calo} + b \times m^{TAS},$$

$$a = \frac{\sigma_{TAS}^2 - \rho \sigma_{calo} \sigma_{TAS}}{\sigma_{calo}^2 + \sigma_{TAS}^2 - 2\rho \sigma_{TAS} \sigma_{calo}} \quad b = \frac{\sigma_{calo}^2 - \rho \sigma_{calo} \sigma_{TAS}}{\sigma_{calo}^2 + \sigma_{TAS}^2 - 2\rho \sigma_{TAS} \sigma_{calo}} \quad (10)$$

584 which reduces to equation (8) after simple algebra for the case when $\rho = 0$. Of course, this value can be
 585 set to the value of the specific sample considered, or to an average of 0.3 if one wants to give a definition
 586 generally valid for all the cases considered; in this case, the performance would be slightly sub-optimal.

587 **6.2.1 Procedure**

588 The procedure of producing the m_{TAS}^{comb} is defined as follows:

- 589 1. For the given sample, the m^{TAS} and m^{calo} are produced;
- 590 2. The mass responses are also produced for the given ranges of p_T ;
- 591 3. For each of these responses, the value of the IQnR as defined previously is calculated and stored;
- 592 4. The average correlation factor of 0.3 is assumed;
- 593 5. With the formula 9, m_{TAS}^{comb} is calculated using the m^{TAS} , m^{calo} and the values stored from before.

594 A remark on the procedure: the step 3. uses values of the IQnR because this was showed to be a more
 595 robust way to look at the response and fit-independent. For step 4. the correlation factor was decided to
 596 be an average of the samples considered.

597 Additionally, the IQnR weights are produced for each sample specifically. In order to give a sample-
 598 independent definition of the m_{TAS}^{comb} , following also the procedure adopted for the m^{comb} , these weights
 599 could be taken from a QCD multijet sample and applied indiscriminately to the particular case. Here of
 600 course the performance would be again sub-optimal, since the variable was not developed in an ad-hoc
 601 way.

602 Throughout the results presented in the following sections, both observables were calculated with ad-hoc
 603 weights. Quantitative statements between them would still hold in the case of QCD weights. However,
 604 when confronting e.g. m^{TAS} with them it has to be kept in mind that in this case their performance is
 605 overestimated, since this choice, although being more general, would perform slightly worse.

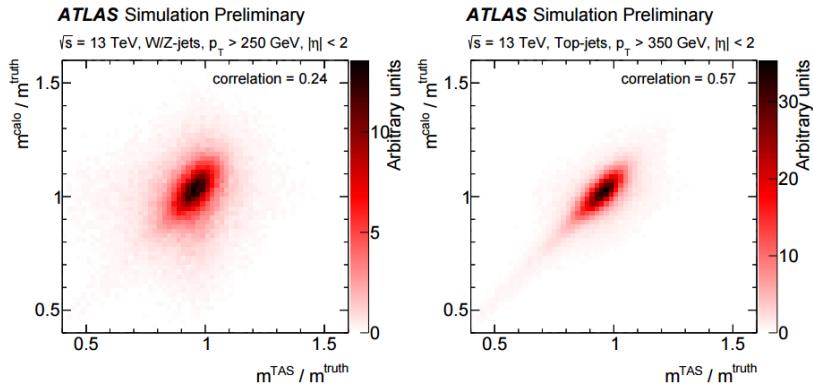


Figure 24: The calorimeter based jet mass mass response versus the track-assisted sub-jet mass response, on the left for boosted W/Z on the right for boosted tops.

606 6.3 Performance in $W \rightarrow q' \bar{q}$ Decays

607 On the boosted W/Z s sample, the performance of the m_{TAS}^{comb} outperforms all the other definitions
 608 throughout all the transverse momentum space; on Figure 25 they are shown for reference together with
 609 the m_{TAS} . It can be noted here that the track-assisted sub-jet mass, although being sub-optimal, has
 610 comparable performance, yet presenting fewer complications due to the combination procedure.

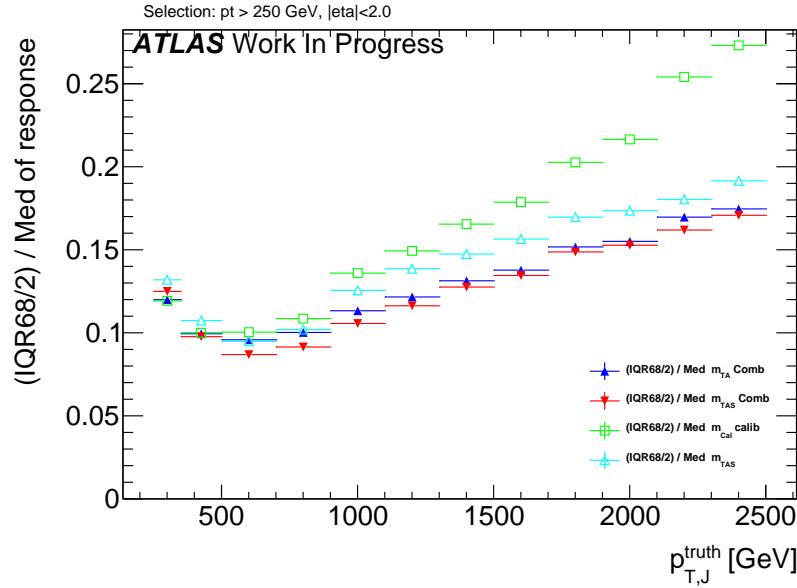


Figure 25: Performance of the combined mass on W/Z samples; here shown the two definitions of the combined mass, m^{comb} and m_{TAS}^{comb} , together with the calorimeter mass and the track-assisted sub-jet mass.

611 6.4 Performance in $t \rightarrow q' \bar{q} b$ Decays

612 The boosted top sample remains the most challenging one also with the combined mass; as seen on Figure
 613 26, the m^{comb} performs quite similarly to the calorimeter based mass definition, yet behaving considerably

614 better than the m^{TAS} especially at high transverse momentum. The m_{TAS}^{comb} , however, outperforms all the
 615 other definitions, and shows its optimal observable strength at middle p_T i.e. in the range $1 < p_T < 1.6$
 616 TeV.

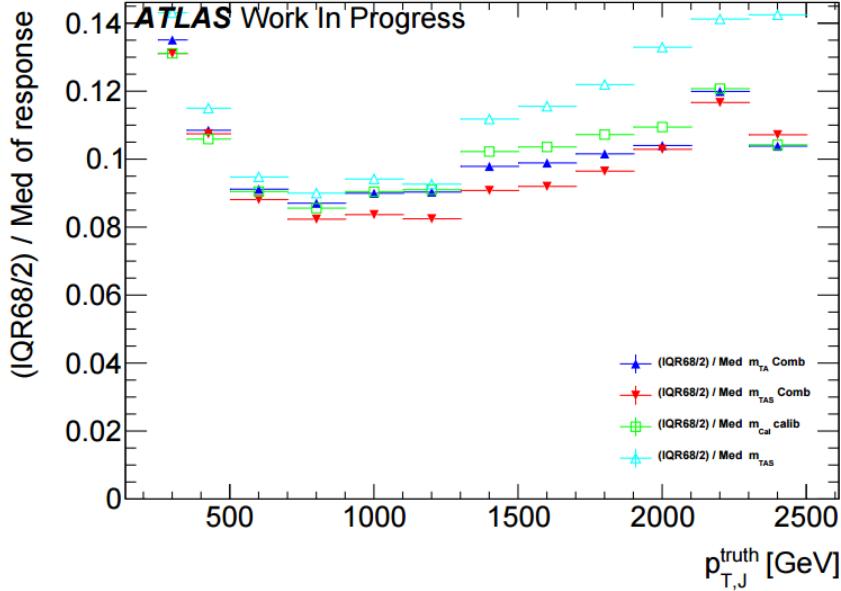


Figure 26: Performance of the combined mass on the top sample; here shown the two definitions of the combined mass, m^{comb} and m_{TAS}^{comb} , together with the calorimeter mass and the track-assisted sub-jet mass.

617 6.5 Performance in $h \rightarrow b\bar{b}$ Decays

618 Again, for the Higgs decay there are similarities as for the top sample; on Figure 27 the two definitions of
 619 the combined mass, together with the simpler m^{TAS} . Although this variable is lightly sub-optimal yet still
 620 comparable in the low to intermediate range in transverse momenta, where the tracks are driving a decrease
 621 in performance for the high to very-high p_T . The m_{TAS}^{comb} uses this advantage to achieve optimal behavior
 622 in the entire transverse momentum spectrum, outperforming both m^{calo} and m^{comb} almost everywhere.

623 7 Energy Correlation Functions and n-Subjettiness

624 7.1 Sub-jet track assistance

625 Tracks and their angular resolution could not only improve the jet mass definition but also the performance
 626 of tagging variables such as the Energy Correlation Functions or n-Subjettiness. These variables are
 627 usually calculated with calorimeter clusters as input, studied here are tracks and assisted tracks as input in
 628 comparison with the default method using clusters. In contrast to the m^{TA} variable introduced in Section
 629 5.1, not the mass but the p_T of each track is scaled, since C_2 , D_2 , τ_{21} and τ_{32} are calculated with the
 630 constituents p_T .

631 The concept of track assisting with the p_T ratio of the whole jet is without effect for the studied substructure
 632 variables. This can be understood from the definitions of the weighted p_T sums. If corrected with only

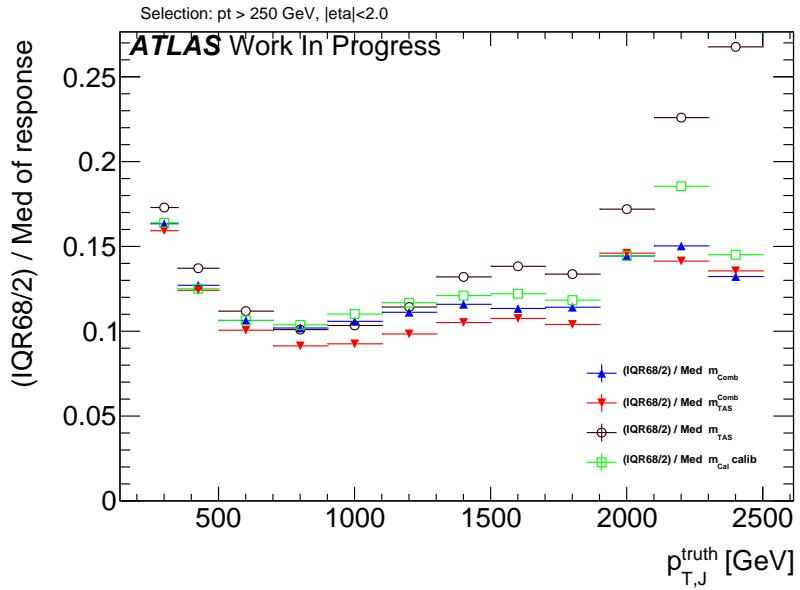


Figure 27: Performance of the combined mass on the Higgs decay; here shown the two definitions of the combined mass, m^{comb} and m_{TAS}^{comb} , together with the calorimeter mass and the track-assisted sub-jet mass.

one ratio, all tracks are scaled by the same factor c , which then can be put in front of the sum and cancels as soon as the ratios τ_{21} and τ_{32} , respectively C2 and D2 are formed.

$$\begin{aligned}\tau_N &= \frac{1}{d_0} \sum_k p_{T,k} c \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})^\beta \\ &= \frac{c}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})^\beta\end{aligned}\quad (11)$$

Track assisting with ghost association to subjets (TAS), see Section 5.3 for m^{TAS} works with different scaling factors depending on the corresponding sub-jet c_k , which also affect ratios:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} c_k \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})^\beta \quad (12)$$

This leads to the following adaption of the TAS procedure:

$$\begin{pmatrix} m_{track} \\ p_{T,track} \\ \eta_{track} \\ \phi_{track} \end{pmatrix} \rightarrow \begin{pmatrix} m_{track} \\ p_{T,track} \frac{m_{track}}{\sum_{ga\,tracks} p_{T,track}} \\ \eta_{track} \\ \phi_{track} \end{pmatrix} \quad (13)$$

Where the sum combines the p_T of all tracks that are associated to a given sub-jet.

7.1.1 Event weighting and Mass-Cut

The substructure variables are compared via their QCD (multi-jet) rejection performance. While the p_T distribution of the multi-jet sample falls exponentially, the p_T of the signal samples features characteristic

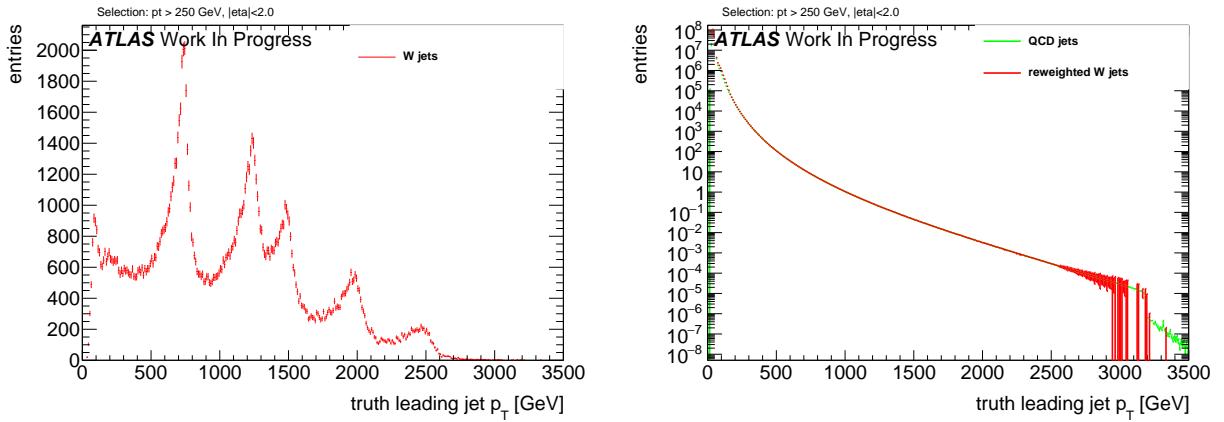


Figure 28: Exemplary p_T distributions of (left) W boson jets and (right) QCD jets from multi-jet events with reweighted W boson events

p_T [GeV]	W boson		Higgs boson		Top quark	
	Mass [GeV]	$\frac{1}{\epsilon_{bgr}}$	Mass [GeV]	$\frac{1}{\epsilon_{bgr}}$	Mass [GeV]	$\frac{1}{\epsilon_{bgr}}$
250 - 500	63 - 85	10.8	56 - 167	3.8	77 - 191	6.3
500 - 800	72 - 92	13.6	92 - 150	7.3	117 - 205	6.9
800 - 1200	76 - 104	9.6	98 - 143	9.5	122 - 218	6.5
1200 - 1600	77 - 107	7.3	103 - 149	9.0	122 - 227	6.3
1600 - 2000	79 - 115	5.6	91 - 170	4.4	121 - 235	5.6
> 2000	80 - 126	4.2	/	/	123 - 251	4.8

Table 2: Studied p_T regions and corresponding calculated 68% mass intervals along with the background rejections from the mass cut for W boson, Higgs boson and Top quark jets.

peaks related to the different resonance masses, see Figure 28. To avoid bias in the comparison, the signal sample is given weights such that the truth p_T distribution of the leading jet matches the one of the background sample. Furthermore, the spectrum is split into six different p_T regions to study the behavior with rising energy.

Tagging variables such as C2, D2, τ_{21} and τ_{32} are usually used after applying a mass cut around the interval that contains 68% of the signal events. Therefore, a cut is applied on the calibrated mass of the large-R calorimeter jet which is calculated to cover the smallest interval around the peak mass that contains 68% of the signal events. The comparison is performed in six different p_T regions to study the behavior connected with rising energy of the decaying particle. These regions are presented in the left part of Table 2. In case of the Higgs boson study, there is not enough statistics to derive a conclusive result for $p_T > 2000$ GeV, since the highest resonance mass of the $G^* \rightarrow HH$ samples is 3000 GeV in contrast to 5000 GeV for the $Z' \rightarrow tt$ and $W' \rightarrow WZ$ samples. Hence this study is restricted to the five lower p_T bins. Prior to tagging with the n-Subjettiness or C2/D2 variables, a cut on the calibrated calorimeter jet mass is applied, given that the mass is the main discriminant in QCD jet rejection. This cut is defined to choose the smallest interval around the peak mass containing 68% of the signal. However, the reconstructed mass depends on the p_T region, therefore a different cut was calculated for every region to meet the requirements.

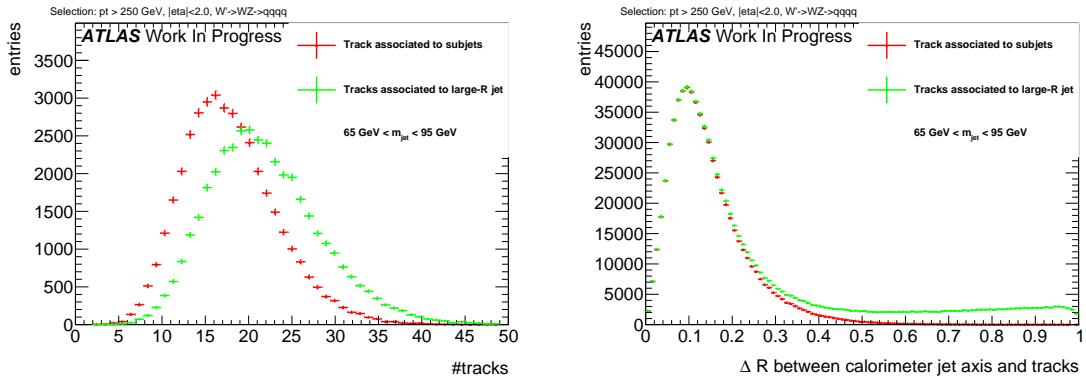


Figure 29: The number of tracks ghost associated to the large-R jet and to the sub-jets (left) and angular distance of associated tracks to the large-R calorimeter jet axis (right). Signal events were not reweighted at this step.

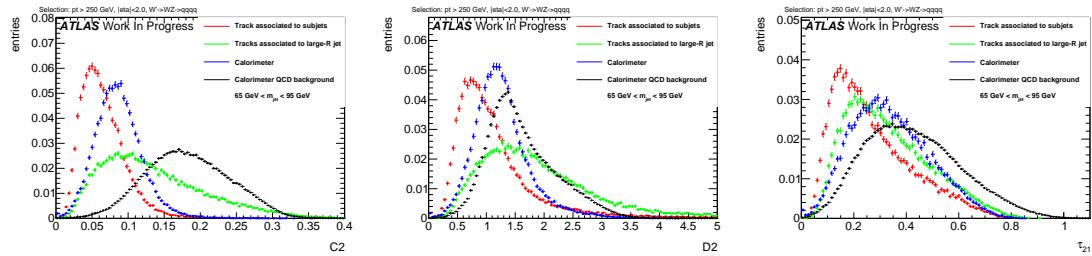


Figure 30: Substructure variables (left) C2, (right) D2 and (below) τ_{21} calculated with calorimeter clusters as well as tracks associated to sub-jets and to the large-R jet. Signal events were not reweighted at this step.

658 7.2 Track Selection

659 There are different collections of tracks that could be used to calculate substructure variables. Compared
660 here are tracks that are ghost associated to the ungroomed large-R jet with the collection which is also
661 used for the m^{TAS} , see Section 5.5, which is ghost association to k_T -subjets and ΔR matching of tracks
662 close to sub-jets.

663 The distributions showing the number of tracks associated to a calorimeter jet, see the left side of Figure
664 29, indicate, that on average around four tracks less are associated to the sub-jets compared to the
665 ungroomed jet. The right side of Figure 29 shows the angular distance ΔR between the single tracks and
666 the axis of the large-R calorimeter jet. Both distributions are aligned in the lower ΔR region while the
667 histogram representing the tracks associated to the ungroomed jet shows an enhancement towards larger
668 ΔR . Accordingly, these additional tracks feature an angular separation from the jet axis of more than
669 0.3, and are in consequence distributed primarily around the outer regions of the large-R jet. Given the
670 required primary vertex association, it is unlikely that these tracks originate from pile-up. Instead, the
671 origin might be found in final- or initial state radiation.

672 Figure 30 shows the signal distributions of the C2/D2, and τ_{21} , calculated with both selections of tracks
673 for W boson jets. The large ΔR to the jet axis of the differing tracks push the substructure variables to
674 higher, more background like values. The broader distributions are a result of the variating nature of
675 these tracks. C2 and D2 are more sensitive to tracks with a large ΔR to the jet axis, because the angular
676 distance between all pairs and triples of tracks is considered, among tracks on possibly opposite ends of
677 the large-R jet, whereas τ_{21} uses distances to k_T -WTA axes. For comparison, the signal and background

678 distributions for the variables calculated with calorimeter clusters are shown as well. It is possible to
 679 anticipate that the performance of variables calculated with tracks and assisted tracks is not worse than
 680 cluster base variables. In contrast to the previously studied jet mass variable, ratios of ECF(N) and τ_N
 681 are rather energy scale independent and are found to not be as sensitive to the missing neutral fraction
 682 with un-assisted tracks. Starting from this observations, the performance of substructure techniques is
 683 compared with the following objects as input:

- 684 • Calorimeter clusters, labeled 'calo'.
 685 • Tracks selected as described in Section 5.5, labeled 'tracks'.
 686 • The same collection of tracks, assisted as defined in Section 7.1, labeled 'TAS'.

687 7.3 Performance with default β

688 The performance of track and TAS based tagging variables with the default angular weighting of $\beta =$
 689 1 is compared to the corresponding calorimeter variables for W boson, Higgs boson and Top quark
 690 identification. The stated signal efficiencies are calculated after the mass cut plus tagging with n-
 691 Subjettiness or C2/D2. Therefore, the endpoint of the ROCs is at 68% signal efficiency, the fraction kept
 692 after the mass cut. Consequently, it is required to achieve a tagging only signal efficiency of $\frac{0.5}{0.68} \sim 0.74$
 693 for a signal efficiency of 0.5 after mass cut and tagging. Similarly, the stated and compared background
 694 rejections result from the multiplication of both, thus representing the QCD rejection of the combined
 695 mass- and tagging variable-cut. The complete set of signal and background distributions for different
 696 inputs can be found in the Appendix ??.

697 7.3.1 Performance for W boson tagging

698 Shown in Figure 31 are exemplary signal and background distributions in intermediate p_T bins compared
 699 for different inputs. This shows throughout narrower signal distributions rising slightly sharper for tracks
 700 and assisted tracks compared to calorimeter clusters which can be explained by the high angular resolution.
 701 The right handed tails of the signal distributions are similar to the calorimeter variables. Similarly, the
 702 background distributions shift as well, but not as distinct as seen for the signal.

703 The ROCs in Figure 32, 33 and 34 show the actual achieved background rejection at different p_T values.
 704 For lower p_T values, TAS perform comparably to calorimeter clusters. Tracks without assisting achieve
 705 a considerably lower background rejection with D2 and τ_{21} for lower energies. Tracks and TAS perform
 706 equally well at high energies for D2 and τ_{21} and for C2 over the whole studied range. At higher boosts,
 707 the angular resolution of the tracks becomes more and more relevant as the separation between jet
 708 constituents shrinks. Consequently, tracks and TAS start to outperform calorimeter based variables and
 709 become increasingly effective with rising energy.

710 7.3.2 Un-assisted tracks and TAS at very high p_T

711 The C2 variable was found to perform equally well with tracks and TAS as input. This variable seems to
 712 be relative insensitive to the track assisting and tracks alone already perform well. D2 and τ_{21} in contrast,

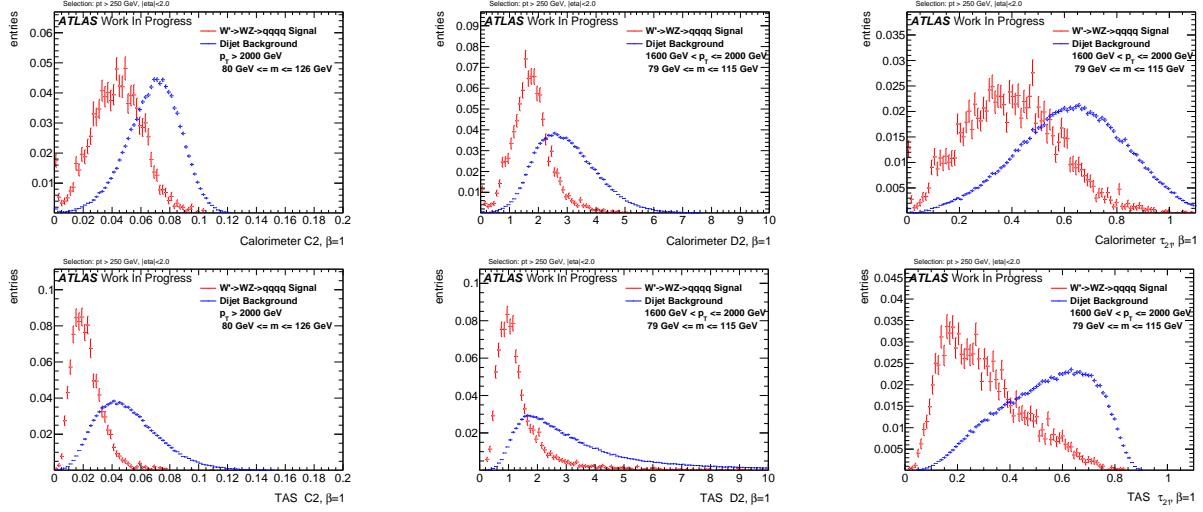


Figure 31: W boson signal and QCD background distributions for calorimeter (left) and TAS (right) at $\beta = 1$ with C2 (top) for more than 2000 GeV and D2 (middle) and τ_{21} (bottom) for 1200-1600 GeV

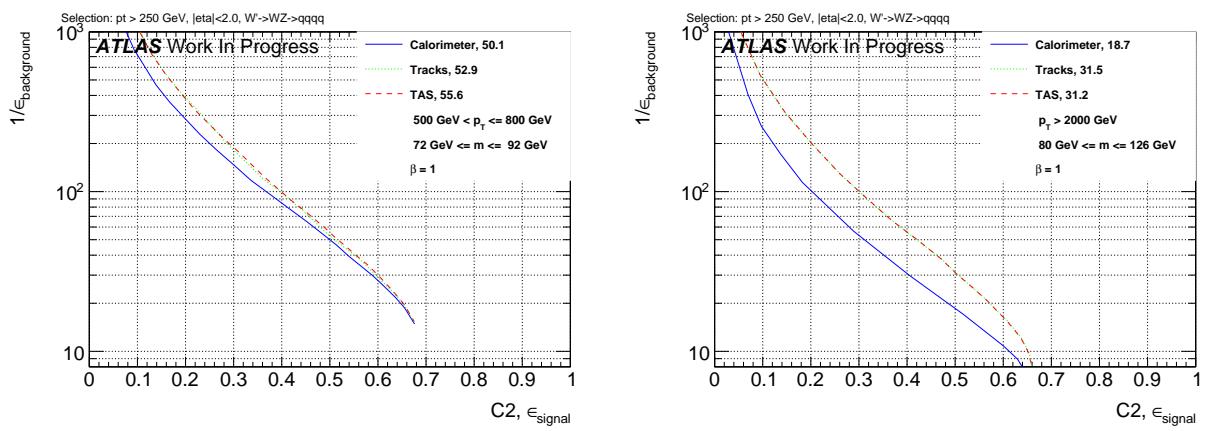


Figure 32: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and colorimeter C2 at $\beta = 1$ for 500-800 GeV (left) and >2000 GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

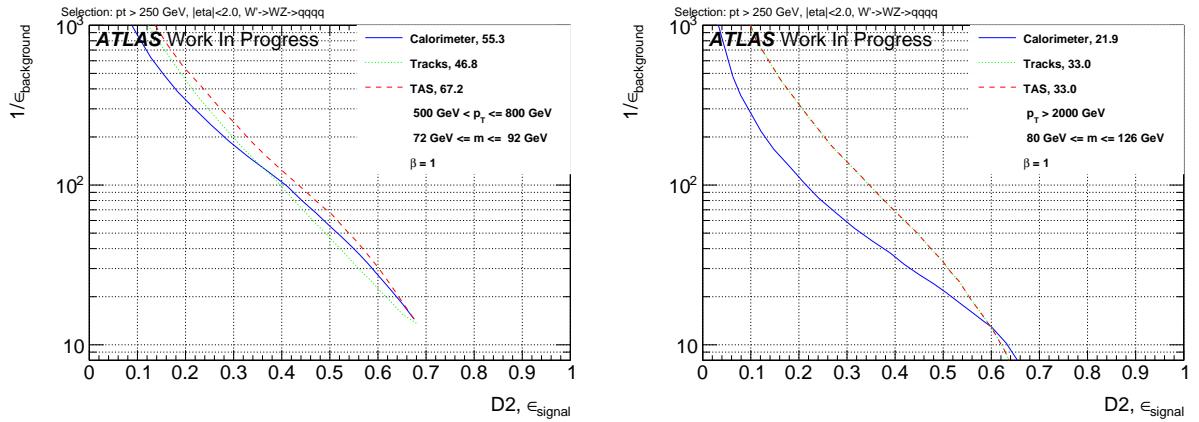


Figure 33: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and calorimeter D2 at $\beta = 1$ for 500-800 GeV (left) and >2000 GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

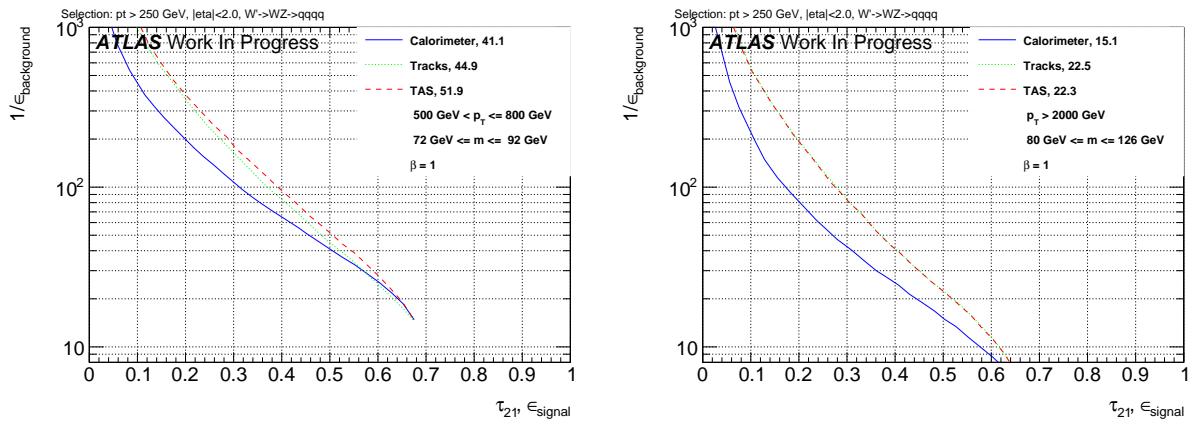


Figure 34: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and colorimeter τ_{21} at $\beta = 1$ for 500-800 GeV (left) and >2000 GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

713 feature a visibly worse separation with tracks than with assisted tracks. In these cases, the scale difference
714 due to the missing neutral fraction seems to have a greater influence.

715 For very high p_T values however, it is often the case that the large-R calorimeter jet features only one
716 $R = 0.2$ sub-jet after trimming due to the now small separation of constituents. A single sub-jet results in
717 the TAS procedure to fall back to TA. As stated in Section 7.1, TA has no impact on the ratios. Therefore,
718 $C2/D2$ and τ_{21} perform equally well when calculated with tracks or TAS for events with only one sub-jet
719 and thereby the difference between both decreases for very high energies.

720 7.3.3 Correlation with p_T

721 Due to the rapidly falling p_T spectrum and hence low weights for high p_T are the correlation plots divided
722 into the six different p_T regions. For $C2$, see Figure 35, one can observe a strong trend to lower values
723 for signal and background with calorimeter clusters as well as TAS. Furthermore, it is possible to observe
724 that the TAS distributions concentrate at lower values compared to calorimeter counterparts.

725 In the cases of D2, Figure 36, and τ_{21} , Figure 37, there is a small upward trend of the calorimeter variables
 726 visible in the lower p_T regions which, with rising boost, slows down for D2 and τ_{21} and ends for τ_{21} in
 727 a broader distribution. This verifies the higher p_T dependence of the C2 variable in comparison to D2
 728 and τ_{21} . The TAS counterparts feature an even more robust signal with the background moving to higher
 729 values, hence improving separation. The p_T dependence of variables calculated with tracks is very similar
 730 to the ones with TAS, therefore they are omitted.

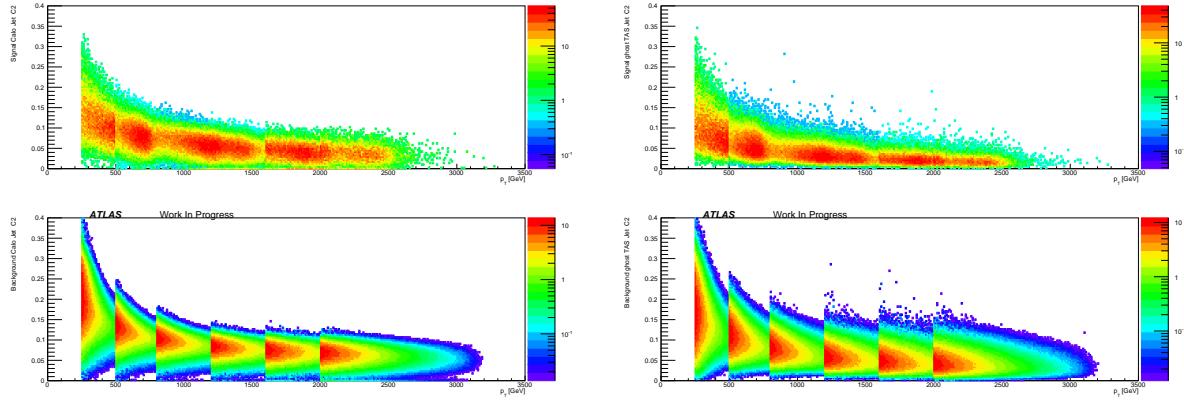


Figure 35: Correlation between C2 at $\beta = 1$ and p_T applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

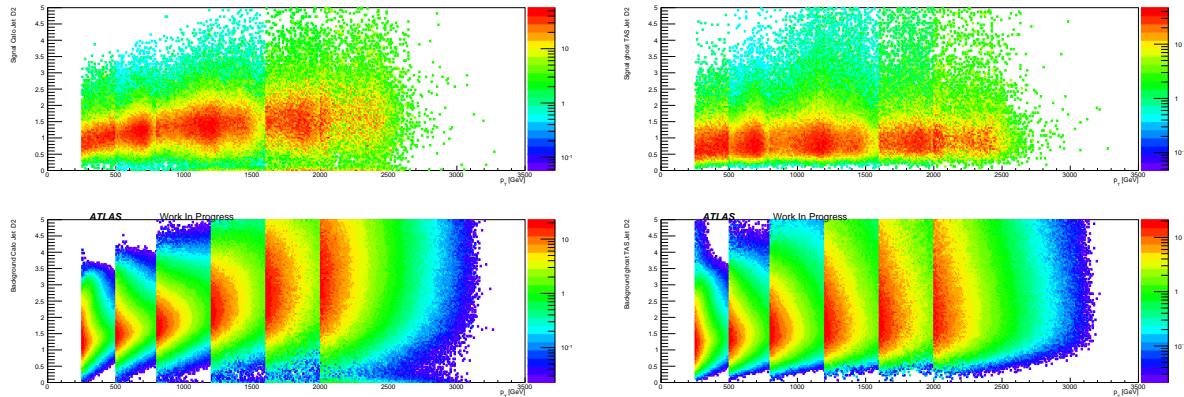


Figure 36: Correlation between D2 at $\beta = 1$ and p_T applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

731 7.3.4 Performance for Higgs boson tagging

732 The Higgs boson is heavier than the W or Z boson, resulting in a higher angular separation of the jet
 733 constituents considering the rule of thumb $\delta R \sim \frac{2m}{p_T}$ for decay products. As a result, angular resolution
 734 effects won't have the same impact as for the W boson. This can be verified by the performance of
 735 track-based variables in the ROCs found in Figure 38.

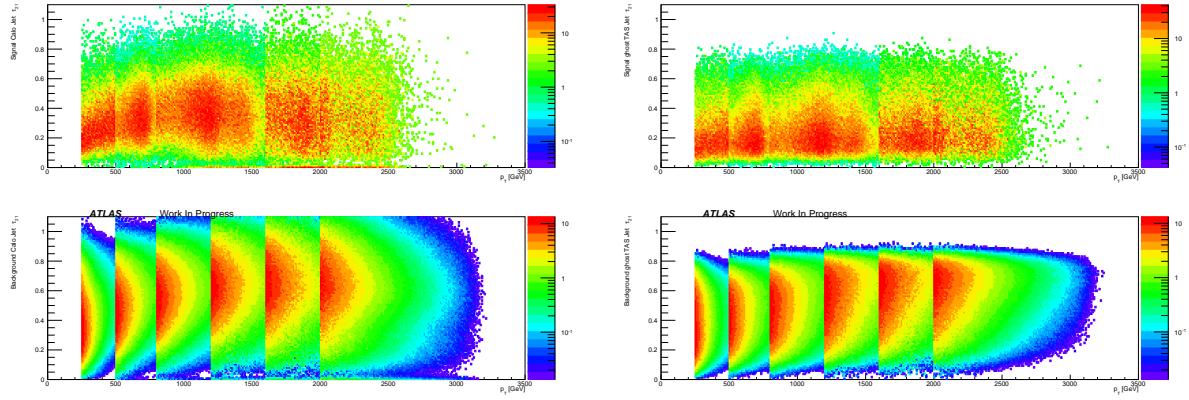


Figure 37: Correlation between τ_{21} at $\beta = 1$ and p_T applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

736 For Higgs boson tagging and an angular weight of $\beta = 1$, found were no distinct improvements with TAS
 737 or tracks compared to calorimeter clusters. The C2 variable performs better with calorimeter clusters,
 738 D2 yields an equal QCD discrimination with TAS and calorimeter clusters. The n-Subjettiness ratio τ_{21}
 739 benefits from TAS in some p_T regions, while the calorimeter pendant performs better in the other regions.
 Furthermore, tracks and TAS perform comparable over the whole studied p_T range.

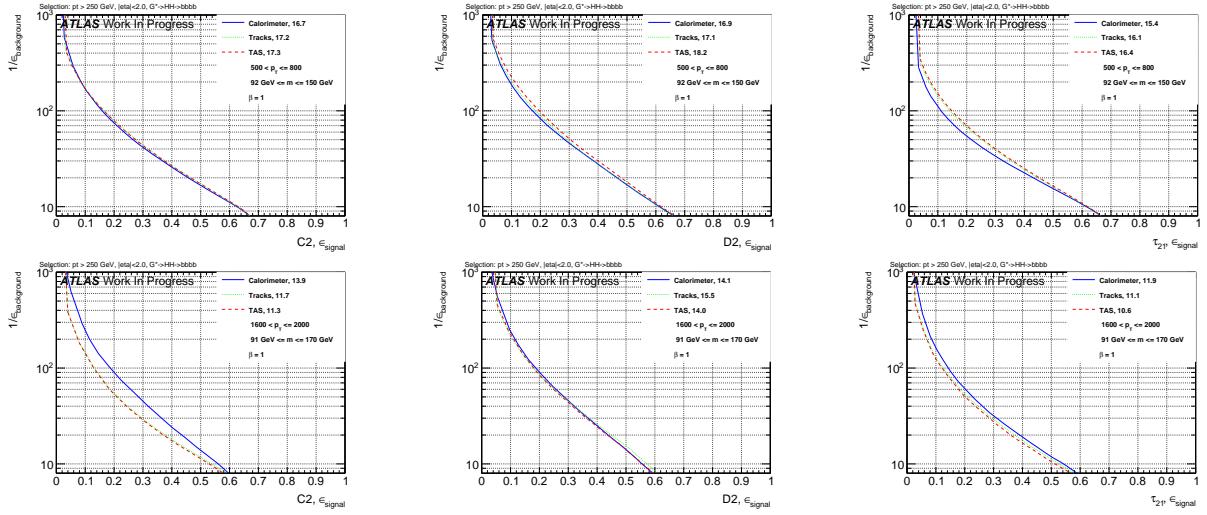
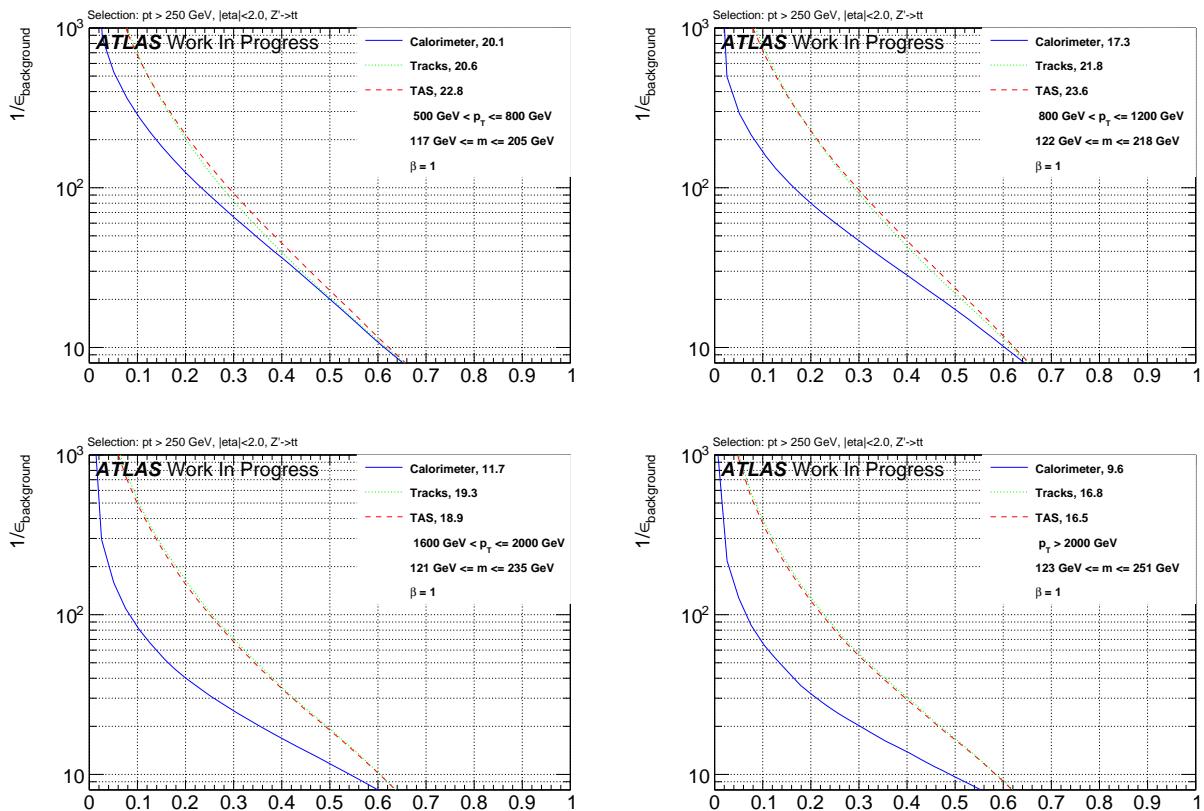


Figure 38: ROCs showing QCD rejection against Higgs boson efficiency for tracks, TAS and colorimeter. C2 (left), D2 (middle) and τ_{21} (right) at $\beta = 1$. Shown is the energy range between 500-800 GeV (top) and 1600-2000 GeV (bottom). The numbers in the legend second p_T bin (left) and highest bin (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

741 7.3.5 Performance for Top quark tagging

742 The top quark features a characteristic three body decay and a very high mass around 173 GeV. Studied
 743 here is the n-Subjettiness ratio τ_{32} to distinguish the three prong like top quark jets and QCD background
 744 jets.

745 The ROCs in Figure 39 show the accompanying improvements in the separation power of τ_{32} possible
 746 with TAS. Tagging tops quark events with τ_{32} is found to greatly benefit from the excellent angular
 747 resolution of tracks. This is especially the case for high p_T where the limitation of the calorimeter cell
 748 size clearly diminishes the possible identification of three distinct substructures inside a large-R jet. The
 749 enhancements are not as articulated for the low p_T regions, nevertheless TAS τ_{32} performs here at least
 750 equally well as calorimeter τ_{32} . Furthermore, tracks are observed to perform slightly worse in comparison
 with TAS for the lower p_T regions, but match the TAS performance for very large boosts as expected.



751 Figure 39: ROCs showing QCD rejection against top quark efficiency for tracks, TAS and calorimeter τ_{32} at $\beta = 1$, p_T ordering from upper left to lower right.

752

7.4 Optimisation of β

753 The observed background rejection of variables calculated with TAS and tracks is at least as high as of
 754 calorimeter cluster based variables due to the high angular resolution of tracks. Therefore, studied are the

755 effects of a higher weighting of the angular part of the substructure variables. For completeness, considered
 756 as well is a lower weighting. Previous studies of default calorimeter variables for W boson tagging, see e.g.
 757 Reference [bib:w_tagging], found $\beta = 1$ to maximize the separation power of calorimeter variables.

758 A scan over the values $\beta = 0.5, 1, 1.7, 2, 3$ is performed in order to identify the best variables for the specific
 759 scenarios of tagging W boson, Higgs boson or top quark jets. The background rejections, achieved at the
 760 50 % working point after mass cut and tagging are summarized in tables. The corresponding ROCs can
 761 be found in Appendix F. Pseudo-experiments were used to propagate the uncertainties on the signal and
 762 background distributions due to the finite size of the MC samples to the background rejections.

763 7.4.1 Optimisation for W boson jets

764 The results of the optimisation for W boson jets are shown in Table 3. As expected, tracks and TAS
 765 perform visibly worse with a low angular weighting. For higher values of β , tracks and TAS gain in
 766 separation power, verifying the significance of the angular part for track based variables. Nevertheless, the
 767 separation is observed to degrade for angular weightings too high compared to the p_T part, here $\beta = 3$.

768 A β of around 2 maximizes the separation power of tracks and TAS. The advantages of $\beta = 2$ compared to
 769 $\beta = 1$ are found at higher p_T values, minor losses are visible in the lowest energy regions. A slightly lower
 770 value of $\beta = 1.7$ was able to retain the great background rejection of a large angular weighting at high p_T
 771 while still performing well at lower energies. Variables calculated with clusters are not as sensitive to a
 772 variation of the angular weighting.

773 C2 is found to be relatively insensitive to the track assisting, whereas separation with τ_{21} and D2 (for
 774 lower energies) is visibly poorer with tracks compared to TAS. Starting around 1200 GeV, tracks perform
 775 comparably and for even higher energies equally well as TAS due to the rising amount of large-R jets with
 776 only one sub-jet. Comparing variables independently for the different inputs, τ_{21} performs worse than C2
 777 and D2. However, e.g. τ_{21} , TAS can outperform $C2_{calo}$ and $D2^{calo}$.

778 The variables achieving the highest background rejections for W boson tagging are $D2_{TAS}^{(\beta=1.7)}$ and $C2_{TAS}^{(\beta=1.7)}$,
 779 depending on the energy. D2 cut values were shown to be more p_T robust, C2 could also be used with
 780 tracks instead of TAS, which saves the assistance. For TAS variables, uncertainties on the sub-jets,
 781 whose derivation is being worked at, need to be propagated while tracks feature already well-known
 782 uncertainties.

783 Shown in Figure 40 are the cut values for 50 % and 25 % signal efficiency for $D2_{TAS}^{(\beta=1.7)}$ and $C2_{TAS}^{(\beta=1.7)}$.
 784 $D2_{TAS}^{(\beta=1.7)}$. As for the default cluster variables, the $D2_{TAS}^{(\beta=1.7)}$ cut is more p_T robust than the cut on $C2_{TAS}^{(\beta=1.7)}$.

785
 786 Table 4 lists the background rejections for $D2_{TAS}^{(\beta=1.7)}$, $C2_{TAS}^{(\beta=1.7)}$ and the currently used $D2_{calo}^{\beta=1}$ along with
 787 the corresponding improvements. For lower energies, $D2_{TAS}^{(\beta=1.7)}$ is the best choice. For very high boosts of
 788 the W boson, $C2_{TAS}^{(\beta=1.7)}$ performs superior, especially for 25 % ϵ_{signal} , where the background rejection with
 789 $C2_{TAS}^{(\beta=1.7)}$ is around 3.5 times as large as the QCD rejection with $D2_{calo}^{(\beta=1)}$. These enormous improvements
 790 at lower ϵ_{signal} are due to the signal distributions for TAS and tracks rising much steeper than for clusters.
 791 The tail to higher, background like values in contrast, is more comparable, leading to an alignment of the
 792 background rejection for very large ϵ_{signal} . The improvements due to TAS lie around 50 % for D2 and
 793 up to a 100 % for C2 in the upper p_T regions and 50 % W boson efficiency. For the lower working point,
 794 they can reach 200 % for D2 and around 250 % for C2, again for very large boosts of the W boson.

p_T [GeV]	Calorimeter						τ_{21}								
	C2			D2			C2			D2					
	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	29.7(1.5)	31.7(1.9)	31.4(1.6)	30.7(1.9)	28.5(1.4)	27.2(2.0)	35.0(2.0)	33.0(1.8)	25.7(1.2)	33.1(1.8)	27.6(1.3)	26.2(1.4)	25.1(1.2)	22.4(0.8)	
500 - 800	44.2(1.8)	50.1(2.0)	49.6(1.9)	48.6(1.8)	42.6(1.9)	40.3(2.2)	55.3(2.6)	56.3(2.4)	52.5(2.1)	39.3(1.3)	49.4(2.0)	41.1(1.4)	43.3(1.7)	41.3(1.6)	
800 - 1200	32.0(1.5)	37.5(1.7)	35.4(1.5)	33.4(1.5)	26.8(0.9)	34.0(2.1)	41.1(2.0)	38.5(1.6)	34.9(1.3)	25.4(0.7)	30.5(1.2)	30.9(1.2)	32.5(1.3)	28.1(0.9)	
1200 - 1600	30.1(1.3)	34.4(1.8)	29.4(1.3)	26.8(1.0)	20.7(0.8)	34.1(1.8)	38.1(1.9)	31.4(1.4)	27.6(1.2)	19.3(0.5)	23.1(0.9)	27.3(1.1)	31.1(1.2)	29.9(1.3)	
1600 - 2000	20.9(1.3)	22.4(1.5)	18.2(1.2)	16.5(0.9)	12.9(0.6)	26.4(1.7)	25.4(1.3)	19.3(1.1)	16.9(0.9)	11.9(0.5)	16.4(1.0)	19.1(1.1)	21.1(1.1)	19.9(1.0)	
> 2000	16.9(1.4)	18.7(1.4)	14.1(0.9)	12.6(0.8)	9.9(0.7)	23.3(1.9)	21.9(1.7)	15.7(1.1)	13.5(0.9)	9.2(0.4)	12.3(1.1)	15.5(1.1)	17.2(1.2)	15.7(1.1)	
TAS	C2						D2						τ_{21}		
p_T [GeV]	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	29.4(1.9)	30.1(1.9)	28.9(1.5)	28.5(1.3)	27.7(1.3)	28.6(2.0)	37.7(2.1)	35.4(2.3)	29.4(1.2)	36.2(2.2)	31.5(1.6)	26.8(1.3)	25.4(1.4)	24.0(1.0)	
500 - 800	48.2(2.0)	55.5(2.7)	58.6(2.6)	59.1(2.7)	56.8(2.0)	42.8(2.3)	67.2(3.1)	67.6(3.2)	63.7(3.0)	52.6(2.3)	55.7(2.6)	51.9(2.1)	45.5(2.0)	44.0(1.9)	
800 - 1200	31.0(1.2)	44.6(1.9)	54.6(2.8)	55.2(2.8)	53.0(3.2)	26.1(1.3)	47.6(2.3)	54.9(2.4)	52.6(2.8)	43.1(1.5)	36.4(1.8)	37.3(1.7)	36.2(1.8)	41.3(1.5)	
1200 - 1600	20.9(0.7)	39.1(1.9)	53.8(2.6)	55.1(3.0)	50.1(1.6)	22.7(1.4)	42.1(2.4)	50.8(1.8)	49.6(2.3)	41.1(1.2)	27.9(1.3)	31.4(1.5)	33.4(1.6)	35.5(1.6)	
1600 - 2000	16.7(0.7)	36.9(2.9)	50.9(4.3)	50.3(4.4)	42.2(2.4)	18.7(1.7)	32.7(3.3)	37.8(2.0)	36.1(2.4)	28.7(1.2)	20.5(1.2)	24.8(1.6)	26.1(2.0)	33.0(1.8)	
> 2000	11.6(0.6)	31.2(3.2)	46.1(4.7)	45.5(5.2)	35.5(3.8)	17.8(2.0)	33.0(4.0)	36.3(2.0)	34.0(2.5)	27.4(1.3)	16.4(1.3)	22.3(2.0)	24.2(2.2)	24.4(2.5)	
Tracks	C2						D2						τ_{21}		
p_T [GeV]	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	27.1(1.2)	28.1(1.5)	28.7(1.9)	28.2(1.7)	21.6(1.2)	28.9(2.0)	29.5(1.8)	29.1(1.6)	28.1(1.3)	28.7(1.8)	28.0(1.7)	25.6(1.3)	25.1(1.3)	24.2(0.9)	
500 - 800	46.5(1.9)	52.9(2.4)	57.7(2.6)	58.1(2.7)	55.8(2.5)	30.1(1.8)	46.8(2.4)	53.4(2.2)	52.1(2.3)	46.6(1.7)	46.1(2.3)	44.9(1.8)	41.7(2.1)	40.6(1.8)	
800 - 1200	30.3(1.1)	44.5(2.2)	54.8(2.8)	56.4(3.0)	53.7(3.6)	24.5(1.5)	42.3(2.3)	48.6(2.5)	47.5(1.2)	42.4(1.2)	34.5(1.6)	36.2(1.8)	36.0(1.8)	35.7(1.5)	
1200 - 1600	20.7(0.6)	39.0(1.9)	54.2(2.7)	55.5(3.3)	50.9(1.7)	22.7(1.3)	41.0(2.2)	50.0(1.6)	47.6(2.2)	41.4(1.2)	27.7(1.2)	31.3(1.4)	33.3(1.6)	33.2(1.8)	
1600 - 2000	16.6(0.7)	36.7(2.3)	51.7(5.2)	51.6(4.0)	43.1(2.3)	18.5(1.7)	32.1(3.0)	37.0(1.9)	35.9(2.3)	29.3(1.2)	20.5(1.3)	24.6(1.7)	26.2(1.8)	25.9(2.2)	
> 2000	11.6(0.5)	31.5(3.0)	46.8(5.7)	46.0(4.2)	36.1(4.3)	17.8(2.2)	33.0(3.3)	35.9(2.1)	34.2(2.6)	28.1(1.0)	16.4(1.4)	22.5(1.8)	24.5(2.4)	24.7(2.6)	

Table 3: Listing of the QCD background rejection for W boson signal achieved with C2, D2 and τ_{21} together with different angular weightings β and for calorimeter cluster, tracks and TAS.

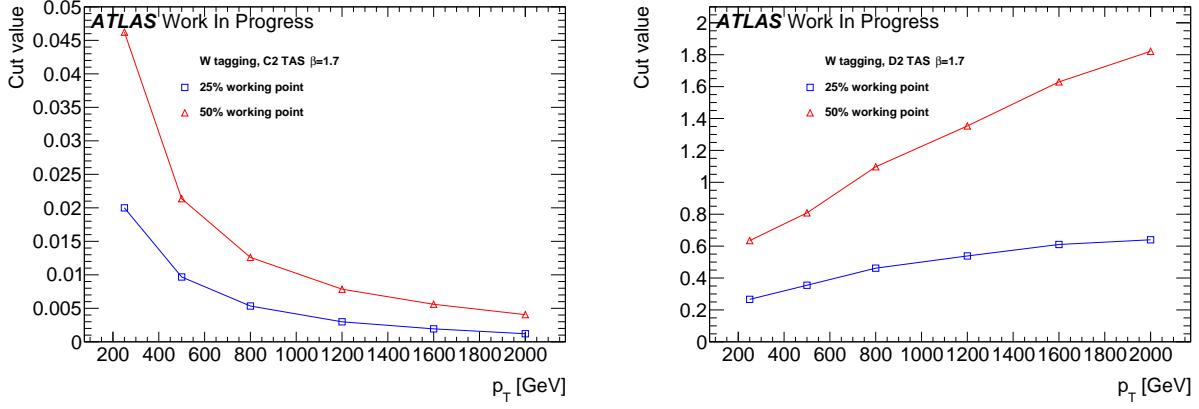


Figure 40: Cut values for $C2_{\text{TAS}}^{(\beta=1.7)}$ (left) and $D2_{\text{TAS}}^{(\beta=1.7)}$ (right) to achieve 50 % and 25 % W boson efficiency.

50% ϵ_{signal}		W tagging		
p_T [GeV]		$D2_{\text{calo}}^{(\beta=1)}$	$D2_{\text{TAS}}^{(\beta=1.7)}$	$C2_{\text{TAS}}^{(\beta=1.7)}$
250 - 500		35.0 ± 2.0	$35.4 \pm 2.3 (+1 \pm 9\%)$	$28.9 \pm 1.5 (-17 \pm 6\%)$
500 - 800		55.3 ± 2.6	$67.6 \pm 3.2 (+22 \pm 8\%)$	$58.6 \pm 2.6 (+6 \pm 7\%)$
800 - 1200		41.1 ± 2.0	$54.9 \pm 2.4 (+34 \pm 9\%)$	$54.6 \pm 2.8 (+33 \pm 9\%)$
1200 - 1600		38.1 ± 1.9	$50.8 \pm 1.8 (+33 \pm 8\%)$	$53.8 \pm 2.7 (+41 \pm 10\%)$
1600 - 2000		25.4 ± 1.3	$37.8 \pm 2.0 (+49 \pm 11\%)$	$50.9 \pm 4.3 (+100 \pm 20\%)$
> 2000		21.9 ± 1.7	$36.3 \pm 2.0 (+66 \pm 16\%)$	$46.1 \pm 4.7 (+111 \pm 27\%)$

25% ϵ_{signal}		W tagging		
p_T [GeV]		$D2_{\text{calo}}^{(\beta=1)}$	$D2_{\text{TAS}}^{(\beta=1.7)}$	$C2_{\text{TAS}}^{(\beta=1.7)}$
250 - 500		139.6 ± 9.8	$146.0 \pm 12.4 (+5 \pm 12\%)$	$108.2 \pm 7.5 (-22 \pm 8\%)$
500 - 800		243.7 ± 13.2	$360.1 \pm 21.1 (+48 \pm 12\%)$	$298.4 \pm 15.9 (+22 \pm 9\%)$
800 - 1200		181.0 ± 8.8	$308.5 \pm 19.3 (+70 \pm 14\%)$	$313.2 \pm 24.4 (+78 \pm 16\%)$
1200 - 1600		156.9 ± 8.3	$295.4 \pm 17.8 (+88 \pm 15\%)$	$354.6 \pm 25.6 (+126 \pm 20\%)$
1600 - 2000		84.6 ± 5.7	$219.6 \pm 10.9 (+160 \pm 22\%)$	$320.5 \pm 31.4 (+279 \pm 45\%)$
> 2000		78.9 ± 7.6	$233.5 \pm 14.7 (+196 \pm 34\%)$	$288.4 \pm 33.3 (+266 \pm 55\%)$

Table 4: Listing of the background rejections after the jet mass cut and tagging at 50% and 25% W boson efficiency for the identified best variables $D2_{\text{TAS}}^{(\beta=1.7)}$ & $C2_{\text{TAS}}^{(\beta=1.7)}$ together with the improvements over the standard choice $D2_{\text{calo}}^{(\beta=1)}$

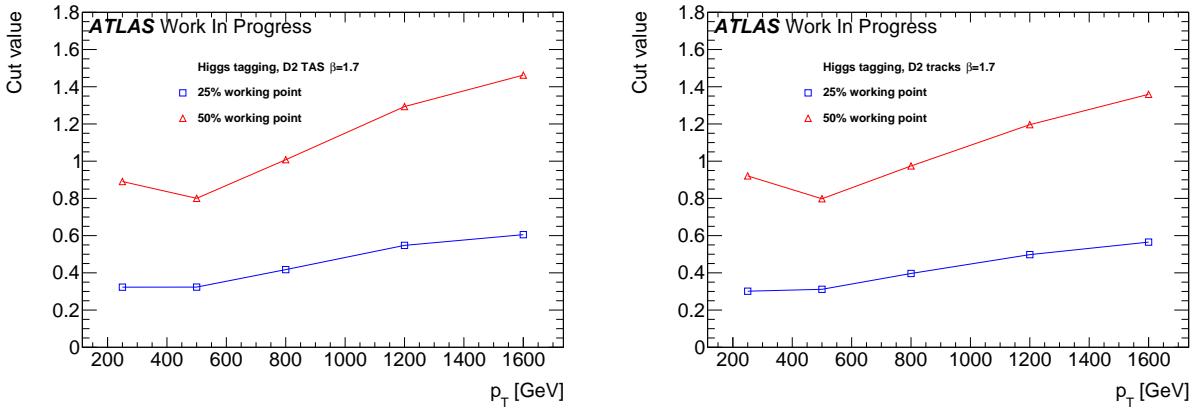


Figure 41: Cut values for $D2_{TAS}^{(\beta=1.7)}$ (left) and $D2_{track}^{(\beta=1.7)}$ (right) to achieve 50% and 25% Higgs boson efficiency.

7.4.2 Optimisation for Higgs boson jets

The results of the optimisation for Higgs boson jets are shown in Table 5. The study of $\beta = 1$ in the Higgs boson case, see section 7.3.4, showed no improvements in the rejection of QCD events due to tracks and TAS as input. As for the W boson, the performance of tracks and TAS diminishes considerably with an angular weighting of $\beta = 0.5$.

No improvement of τ_{21} is observed with tracks or TAS, clusters perform equally well for lower p_T and slightly better at high energies. Again, the QCD rejection achieved with τ_{21} is exceeded by C2 and D2. The discrimination with clusters profits from a slightly higher angular weighting, although the gain is not as significant as for tracks and TAS. This consistently shows the lower sensitivity to a variation of the angular weight. The small gain is connected to the higher separation of the Higgs decay products compared to the W boson case.

For boosted Higgs tagging, D2 outperforms C2 over the whole studied energy range. Values of $\beta = 1.7 \& 2$ yield the highest background rejection for track and TAS based D2. $D2_{TAS}^{(\beta=1.7,2)}$ and $D2_{track}^{(\beta=1.7,2)}$ perform superior to $D2_{calo}$ at high boosts, due to the low angular separation of constituents, and equally well at lower energies.

The differences between $\beta = 1.7$ and $\beta = 2$ are inconclusive with minor advantages at high and slight inferiorities at low p_T for $\beta = 2$. Tracks perform slightly worse than TAS for lower energies but similarly better in the two highest studied p_T regions. Chosen for further examination are $D2_{TAS}^{(\beta=1.7)}$ and $D2_{track}^{(\beta=1.7)}$.

Shown in Figure 41 are the cut values for 50 % and 25 % signal efficiency for $D2_{TAS}^{(\beta=1.7)}$ and $D2_{track}^{(\beta=1.7)}$. The cut value shows a slight upward trend for rising p_T . Moreover, cut values for the first bin are higher as for the second, in contrast to the overall upward trend of D2. This is the result of the low boost in the lowest p_T region resulting in a left shoulder of the mass distributions representing large-R jets containing only part of the Higgs boson decay. These jets feature one-prong structure and result in background-like D2 values. The TAS D2 cut is marginally higher than the corresponding track D2 cut since the assisted tracks have a higher p_T and the D2 cut features a rising tendency with p_T .

Listed in Table 6 are the background rejections for $D2_{TAS}^{(\beta=1.7)}$, $D2_{track}^{(\beta=1.7)}$, and for the best calorimeter variable, which is $D2_{calo}^{(\beta=1)}$, with the corresponding improvements due to the use of TAS respectively

Calorimeter	$p_T [GeV]$	C2						D2						τ_{21}		
		$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	4.6(0.1)	5.0(0.1)	5.2(0.1)	5.3(0.1)	5.5(0.1)	5.7(0.1)	7.3(0.2)	8.4(0.2)	8.4(0.2)	8.4(0.2)	7.6(0.2)	8.0(0.2)	7.9(0.2)	7.8(0.2)	7.5(0.2)	
500 - 800	15.7(0.3)	16.7(0.4)	17.0(0.4)	16.9(0.4)	16.2(0.4)	13.6(0.3)	16.9(0.4)	17.7(0.4)	17.2(0.4)	15.2(0.3)	16.7(0.4)	15.4(0.3)	15.2(0.3)	14.8(0.3)	14.0(0.3)	
800 - 1200	22.1(0.5)	23.8(0.5)	25.0(0.6)	25.0(0.6)	23.4(0.5)	18.4(0.4)	23.7(0.6)	26.3(0.6)	25.6(0.6)	22.3(0.5)	22.8(0.5)	21.9(0.5)	22.6(0.5)	22.1(0.5)	20.9(0.5)	
1200 - 1600	24.0(0.6)	26.0(0.8)	26.4(0.8)	25.9(0.7)	23.0(0.6)	19.3(0.6)	24.9(0.7)	27.0(0.8)	26.1(0.7)	21.9(0.5)	21.3(0.5)	22.6(0.6)	24.0(0.6)	23.7(0.6)	22.2(0.5)	
1600 - 2000	12.1(0.7)	13.9(0.8)	14.3(0.7)	14.0(0.7)	12.3(0.6)	11.1(0.7)	14.1(0.9)	14.9(0.8)	14.2(0.6)	11.8(0.5)	10.3(0.5)	11.9(0.5)	13.1(0.6)	13.1(0.7)	12.3(0.7)	

TAS	$p_T [GeV]$	C2						D2						τ_{21}		
		$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	4.8(0.1)	5.2(0.1)	5.5(0.1)	5.6(0.1)	5.8(0.1)	5.9(0.1)	7.6(0.2)	8.5(0.2)	8.6(0.2)	8.5(0.2)	7.6(0.2)	8.0(0.2)	7.7(0.2)	7.6(0.2)	7.4(0.2)	
500 - 800	16.1(0.4)	17.3(0.4)	17.7(0.4)	17.6(0.4)	17.7(0.4)	14.0(0.3)	18.2(0.4)	18.7(0.4)	18.3(0.4)	16.9(0.4)	16.2(0.4)	16.4(0.4)	15.4(0.4)	15.1(0.3)	14.6(0.3)	
800 - 1200	20.6(0.5)	23.5(0.5)	26.2(0.6)	26.9(0.7)	27.7(0.6)	18.8(0.4)	25.6(0.6)	28.5(0.7)	28.4(0.7)	26.8(0.6)	21.7(0.5)	22.4(0.5)	22.1(0.5)	22.0(0.5)	21.8(0.5)	
1200 - 1600	18.6(0.4)	22.6(0.6)	27.4(0.7)	28.7(0.8)	30.0(0.7)	17.9(0.4)	24.3(0.7)	28.9(0.7)	29.3(0.6)	28.1(0.7)	19.3(0.5)	20.0(0.5)	20.7(0.5)	21.0(0.6)	21.9(0.5)	
1600 - 2000	8.0(0.3)	11.3(0.5)	15.4(0.9)	16.5(1.0)	17.8(0.7)	10.0(0.5)	14.0(0.8)	17.7(0.8)	18.1(0.9)	17.9(0.6)	9.8(0.4)	10.6(0.5)	11.4(0.6)	11.8(0.6)	12.6(0.6)	

Tracks	$p_T [GeV]$	C2						D2						τ_{21}		
		$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	4.9(0.1)	5.2(0.1)	5.5(0.1)	5.6(0.1)	5.9(0.1)	5.8(0.1)	7.4(0.2)	8.3(0.2)	8.3(0.2)	8.5(0.2)	7.4(0.2)	7.9(0.2)	7.8(0.2)	7.7(0.2)	7.6(0.2)	
500 - 800	15.6(0.3)	17.2(0.4)	17.8(0.4)	17.9(0.4)	17.7(0.4)	13.5(0.3)	17.1(0.4)	17.9(0.4)	17.7(0.4)	16.8(0.4)	15.7(0.3)	16.1(0.4)	15.5(0.3)	15.3(0.3)	14.8(0.1)	
800 - 1200	20.1(0.5)	24.0(0.5)	26.9(0.6)	27.7(0.7)	28.4(0.6)	18.8(0.4)	25.3(0.6)	28.0(0.7)	28.0(0.7)	26.9(0.6)	22.0(0.5)	22.7(0.5)	22.5(0.5)	22.4(0.5)	22.4(0.3)	
1200 - 1600	18.5(0.5)	23.8(0.6)	28.8(0.8)	30.0(0.8)	31.1(0.7)	19.4(0.5)	26.3(0.7)	30.0(0.8)	30.3(0.8)	29.2(0.7)	20.8(0.5)	21.4(0.5)	21.9(0.6)	22.3(0.6)	23.0(0.5)	
1600 - 2000	8.0(0.3)	11.7(0.5)	16.1(0.9)	17.1(0.9)	18.3(0.9)	11.0(0.7)	15.5(0.7)	18.5(0.8)	18.7(0.8)	18.4(0.6)	10.4(0.5)	11.1(0.5)	12.0(0.6)	12.4(0.7)	13.2(0.6)	

Table 5: Listing of the QCD background rejection for Higgs signal achieved with C2, D2 and τ_{21} together with different angular weightings β and for calorimeter cluster, tracks and TAS.

50% ϵ_{signal}		Higgs tagging		
p_T [GeV]	D2 _{calo} ^($\beta=1$)	D2 _{TAS} ^($\beta=1.7$)	D2 _{track} ^($\beta=1.7$)	
250 - 500	8.4 ± 0.2	$8.5 \pm 0.2 (+1 \pm 4\%)$	$8.3 \pm 0.2 (-1 \pm 3\%)$	
500 - 800	17.7 ± 0.4	$18.7 \pm 0.4 (+6 \pm 3\%)$	$17.9 \pm 0.4 (+1 \pm 3\%)$	
800 - 1200	26.3 ± 0.6	$28.5 \pm 0.7 (+8 \pm 4\%)$	$28.0 \pm 0.7 (+6 \pm 4\%)$	
1200 - 1600	27.0 ± 0.8	$28.9 \pm 0.7 (+7 \pm 4\%)$	$30.0 \pm 0.8 (+11 \pm 4\%)$	
1600 - 2000	14.9 ± 0.8	$17.7 \pm 0.8 (+19 \pm 8\%)$	$18.5 \pm 0.8 (+24 \pm 9\%)$	

25% ϵ_{signal}		Higgs tagging		
p_T [GeV]	D2 _{calo} ^($\beta=1$)	D2 _{TAS} ^($\beta=1.7$)	D2 _{track} ^($\beta=1.7$)	
250 - 500	25.1 ± 0.6	$28.9 \pm 0.7 (+15 \pm 4\%)$	$30.5 \pm 0.8 (+22 \pm 4\%)$	
500 - 800	54.1 ± 1.4	$69.6 \pm 1.9 (+29 \pm 5\%)$	$64.9 \pm 1.8 (+20 \pm 5\%)$	
800 - 1200	90.8 ± 2.5	$121.3 \pm 3.4 (+34 \pm 5\%)$	$117.9 \pm 3.2 (+30 \pm 5\%)$	
1200 - 1600	97.6 ± 3.1	$117.7 \pm 3.8 (+21 \pm 5\%)$	$122.4 \pm 4.2 (+25 \pm 6\%)$	
1600 - 2000	54.6 ± 3.5	$74.0 \pm 5.7 (+36 \pm 14\%)$	$75.0 \pm 5.1 (+37 \pm 13\%)$	

Table 6: Listing of the background rejections after the jet mass cut and tagging at 50% and 25% Higgs signal efficiency for the identified best variables D2_{TAS}^($\beta=1.7$) & D2_{track}^($\beta=1.7$) together with the improvements over the best variable with clusters which is D2_{calo}^($\beta=1$).

823 tracks instead of clusters. At very high energies, the angle between the $b\bar{b}$ pair is small despite the high
 824 Higgs boson mass and the effect of the calorimeter cell size becomes significant. The improvements for
 825 D2 calculated with TAS instead of clusters are single-digit percentages for low p_T and up to 20 % for the
 826 highest studied p_T bin at 50 % Higgs boson efficiency. For the lower working point, they reach around
 827 30 % of the QCD rejection achieved with cluster based D2.

828 7.4.3 Optimisation for Top quark jets

829 The results of the optimisation for Top quark jets are shown in Table 7. Studied was τ_{32} with values of
 830 $\beta \geq 1$, since the W boson and Higgs boson parts affirmed the expected lower performance of track and
 831 TAS based variables with an angular weighting of $\beta \leq 1$. The calorimeter τ_{32} variable profits from a
 832 higher angular weighting up to around $\beta = 2$, but degrades in performance for $\beta = 3$. Since the involved
 833 three prong structure of the top quark decay requires a good angular separation of the jet constituents to
 834 be resolved, tracks and TAS perform superior to clusters. A higher angular weighting does not improve
 835 the separation power of track and TAS variables, $\beta = 2$ already diminishes the performance. The best
 836 discrimination is achieved with TAS and $\beta = 1, 1.7$. The marginal differences between both values of
 837 β depend on the considered p_T region. Track $\tau_{32, track}$ achieves lower separation as $\tau_{32, TAS}$, except for
 838 regions with very high boosts, but as well outperforms the cluster variable.

839 Shown in Figure 42 are the cut values for 50 % and 25 % signal efficiency for $\tau_{32, TAS}^{(\beta=1.7)}$ and $\tau_{32, track}^{(\beta=1.7)}$. The
 840 crack between the first and second p_T bin is more evident since the top quark with its much higher mass is
 841 here very unlikely to be reconstructed into a single large-R jet, resulting in background like signal events.
 842 Furthermore, τ_{32} ($\beta = 1.7$) needs to be cut at lower values as τ_{32} ($\beta = 1$) to achieve a certain signal
 843 efficiency. This is the result of the higher angular weighting that shifts the overall distributions to lower
 844 values, because the angular distance between two constituents inside a (highly) boosted large-R jet is in

Calorimeter		τ_{32}			
$p_T [GeV]$		$\beta = 1$	1.7	2	3
250 - 500		9.7 ± 0.2	9.5 ± 0.2	9.5 ± 0.4	9.4 ± 0.2
500 - 800		20.1 ± 0.5	22.2 ± 0.6	22.4 ± 0.6	22.0 ± 0.6
800 - 1200		17.3 ± 0.4	20.3 ± 0.5	20.6 ± 0.5	20.3 ± 0.5
1200 - 1600		14.3 ± 0.3	16.4 ± 0.4	16.6 ± 0.5	16.1 ± 0.5
1600 - 2000		11.7 ± 0.3	13.3 ± 0.4	13.3 ± 0.4	12.6 ± 0.3
> 2000		9.6 ± 0.3	11.0 ± 0.4	10.9 ± 0.4	10.1 ± 0.3

TAS		τ_{32}			
$p_T [GeV]$		$\beta = 1$	1.7	2	3
250 - 500		10.7 ± 0.2	10.1 ± 0.2	9.9 ± 0.2	9.6 ± 0.2
500 - 800		22.8 ± 0.6	22.8 ± 0.6	22.5 ± 0.6	21.6 ± 0.6
800 - 1200		23.6 ± 0.6	24.1 ± 0.6	23.6 ± 0.6	22.2 ± 0.5
1200 - 1600		22.0 ± 0.6	22.3 ± 0.6	21.7 ± 0.6	19.8 ± 0.6
1600 - 2000		18.9 ± 0.6	18.8 ± 0.6	17.9 ± 0.5	16.0 ± 0.5
> 2000		16.5 ± 0.7	15.7 ± 0.7	15.2 ± 0.7	13.1 ± 0.6

Tracks		τ_{32}			
$p_T [GeV]$		$\beta = 1$	1.7	2	3
250 - 500		10.5 ± 0.2	9.8 ± 0.2	9.6 ± 0.2	9.4 ± 0.2
500 - 800		20.6 ± 0.5	21.3 ± 0.6	21.1 ± 0.5	20.3 ± 0.5
800 - 1200		21.8 ± 0.6	22.9 ± 0.6	22.6 ± 0.6	21.4 ± 0.6
1200 - 1600		21.7 ± 0.6	22.1 ± 0.6	21.6 ± 0.6	19.5 ± 0.6
1600 - 2000		19.3 ± 0.6	19.0 ± 0.6	18.2 ± 0.6	16.0 ± 0.5
> 2000		16.8 ± 0.7	15.8 ± 0.7	15.1 ± 0.7	13.0 ± 0.5

Table 7: Listing of the QCD background rejection for top signal achieved with τ_{32} together with different angular weightings β and for calorimeter cluster, tracks and TAS.

the majority of cases lower than one. Thus, the angular part of τ_{32} decreases with $\beta > 1$. The TAS τ_{32} cut value is observed to be robust against variations of p_T , in accordance to the results of the p_T correlation plots, see 37.

Listed in Table 8 are the background rejections for $\tau_{32, \text{TAS}}^{(\beta=1)}$, $\tau_{32, \text{TAS}}^{(\beta=1.7)}$ and the best cluster based variable, $\tau_{32, \text{calo}}^{(\beta=2)}$. The differences between both values of β with TAS are marginal, as well for lower signal efficiencies. Improvements due to the use of TAS instead of clusters are possible for Top quark tagging over the whole studied p_T range. These enhancements are, as expected, rising with the boost of the Top quark and can reach around 50 % for the 50 % working point and even 100 % for 25 % Top efficiency.

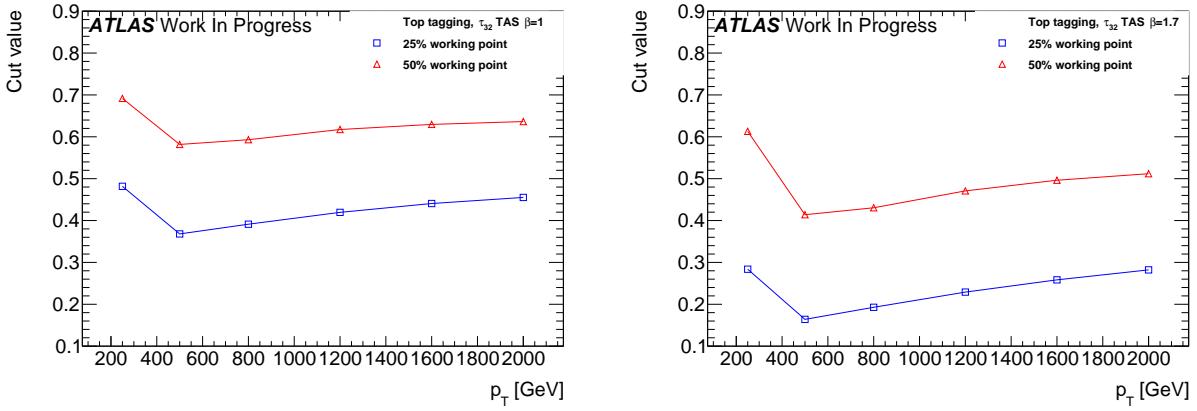


Figure 42: Cut values for $\tau_{32, \text{TAS}}^{(\beta=1)}$ (left) and $\tau_{32, \text{TAS}}^{(\beta=1.7)}$ (right) to achieve 50% and 25% Top quark efficiency

50% ϵ_{signal}			
Top Tagging			
p_T [GeV]	$D2_{\text{calo}}^{(\beta=2)}$	$\tau_{32, \text{TAS}}^{(\beta=1)}$	$\tau_{32, \text{TAS}}^{(\beta=1.7)}$
250 - 500	9.5 ± 0.2	$10.7 \pm 0.2 (+13 \pm 3 \%)$	$10.1 \pm 0.2 (+6 \pm 3 \%)$
500 - 800	22.4 ± 0.6	$22.8 \pm 0.6 (+2 \pm 4 \%)$	$22.8 \pm 0.6 (+2 \pm 4 \%)$
800 - 1200	20.6 ± 0.5	$23.6 \pm 0.6 (+15 \pm 4 \%)$	$24.1 \pm 0.6 (+17 \pm 4 \%)$
1200 - 1600	16.6 ± 0.4	$22.0 \pm 0.6 (+33 \pm 5 \%)$	$22.3 \pm 0.6 (+34 \pm 5 \%)$
1600 - 2000	13.3 ± 0.4	$18.9 \pm 0.6 (+42 \pm 6 \%)$	$18.8 \pm 0.6 (+41 \pm 6 \%)$
> 2000	10.9 ± 0.4	$16.5 \pm 0.7 (+51 \pm 8 \%)$	$15.7 \pm 0.7 (+44 \pm 8 \%)$

25% ϵ_{signal}			
Top Tagging			
p_T [GeV]	$D2_{\text{calo}}^{(\beta=2)}$	$\tau_{32, \text{TAS}}^{(\beta=1)}$	$\tau_{32, \text{TAS}}^{(\beta=1.7)}$
250 - 500	33.7 ± 1.0	$37.6 \pm 1.4 (+12 \pm 5 \%)$	$36.7 \pm 1.2 (+9 \pm 5 \%)$
500 - 800	114.7 ± 3.3	$138.0 \pm 4.3 (+20 \pm 5 \%)$	$139.1 \pm 4.2 (+21 \pm 5 \%)$
800 - 1200	97.0 ± 2.7	$144.6 \pm 4.9 (+49 \pm 7 \%)$	$149.6 \pm 5.2 (+54 \pm 7 \%)$
1200 - 1600	68.6 ± 2.1	$133.2 \pm 4.6 (+94 \pm 9 \%)$	$134.7 \pm 5.1 (+96 \pm 10 \%)$
1600 - 2000	47.5 ± 1.6	$100.3 \pm 4.2 (+111 \pm 11 \%)$	$99.9 \pm 4.4 (+110 \pm 12 \%)$
> 2000	36.3 ± 1.6	$80.2 \pm 5.0 (+121 \pm 17 \%)$	$75.5 \pm 4.9 (+108 \pm 16 \%)$

Table 8: Listing of the background rejections after the jet mass cut and tagging at 50% and 25% top signal efficiency for the identified best variables $\tau_{32, \text{TAS}}^{(\beta=1,1.7)}$ together with the improvements over the best variable with clusters which is $D2_{\text{calo}}^{(\beta=2)}$.

8 Conclusions & Outlook

8.1 Jet mass observables

The m^{TAS} variable was developed for the large- R jet mass; it combines the information of the tracker- and calorimeter-system to achieve an higher precision in the jet mass reconstruction, correcting the missed neutral fraction which is absent in the tracker but not in the calorimeter. With respect to the m^{TA} , it applies this correction at sub-jet by sub-jet level and not at jet by jet level, therefore providing a more accurate reconstruction. It was shown in Monte Carlo simulation to be a very good observable confronting quantitatively with the other definitions which are either standard or in preparation, m^{calo} , m^{TA} and m^{comb} . In fact, it behaves better in terms of $\frac{1}{2} \times 68\%$ IQnR/median and all the other ways to look at the figure of merit, the mass response, for the boosted W/Z and QCD sample; is always better than the m^{TA} and similar to the m^{calo} for the boosted tops and Higgs. Moreover, it is a slightly worse observable than the m^{comb} , yet being comparable, and avoiding the development of ad-hoc weights. The optimal configuration of m^{TAS} is shown and confronted with different approaches, in particular in terms of different trimming procedure of the large- R jet to be used as an input. All the components of the observable have been studied with the use of truth Monte Carlo information without detector effect, in order to evaluate quantitatively its limits and strengths; the track p_T measure degradation was found to be the cause of the variable decreasing performance at higher transverse momenta.

The m^{comb} is the logical extension of the m^{TAS} , which improves by construction the results beyond the m^{calo} and the m^{TAS} , combining these two variables on the same way of the m^{comb} , but taking into account the higher correlation factor which is inherited from the sub-jet usage. Weights for its construction can be in both cases either derived specifically for the sample considered, or constructed on average with the QCD sample, in this case getting a sub-optimal performance. In all the cases studied, it has a better behavior than the m^{comb} , m^{calo} and m^{TA} .

For the very conclusion, both the variables constructed in the work of this thesis, m^{TAS} and m^{comb} , exhibit a better performance of their counterparts, m^{TA} and m^{comb} , which are now ready to be use or in preparation within the ATLAS collaboration, and share the same advantages -and disadvantages. Further steps are necessary to get this observables to usage: calibration and uncertainties.

8.1.1 Outlook

The outlook of the m^{TAS} and m^{comb} variables follows two main scenarios, concerning the calibration and uncertainties determination which are necessary to get this observables ready to be used. The procedure involved are already fully understood, since the the same was applied or is being applied for the m^{TA} and m^{comb} .

For the simple scenario here the procedure that would take place is the direct Monte Carlo calibration of the m^{TAS} , aiming at correcting the reconstructed jet mass to the particle-level jet mass by applying the calibration factors derived from QCD multijet events, an analogous procedure to the one described in Section ?? for the jet energy scale.

The more complex scenario considers an additional calibration to the sub-jets with $R=0.2$, which is already at an advanced stage within ATLAS for anti- k_t reclustering algorithm (it has a slightly worse performance than k_t , as presented previously).

892 The uncertainties are expected to be similar to the one which were derived for the m^{TA} and which are
893 compared to the m^{calo} on Figure 7; the tracking uncertainties are smaller for the track-assisted mass
894 because of the ratio m^{track}/p_T^{track} and will be smaller as well for the track-assisted sub-jet mass since it
895 uses the same ratio.

896 In-situ uncertainties were derived for the m^{TA} with a sample of enriched top-quark; the same technology
897 used here can be applied to the m^{TAS} .

898 In the more complex scenario, the uncertainties could be derived for the sub-jets R=0.2 reclustered with
899 anti- k_t .

900 **8.2 Energy Correlation Functions and n-Subjettiness**

901 ***here sascha conclusions***

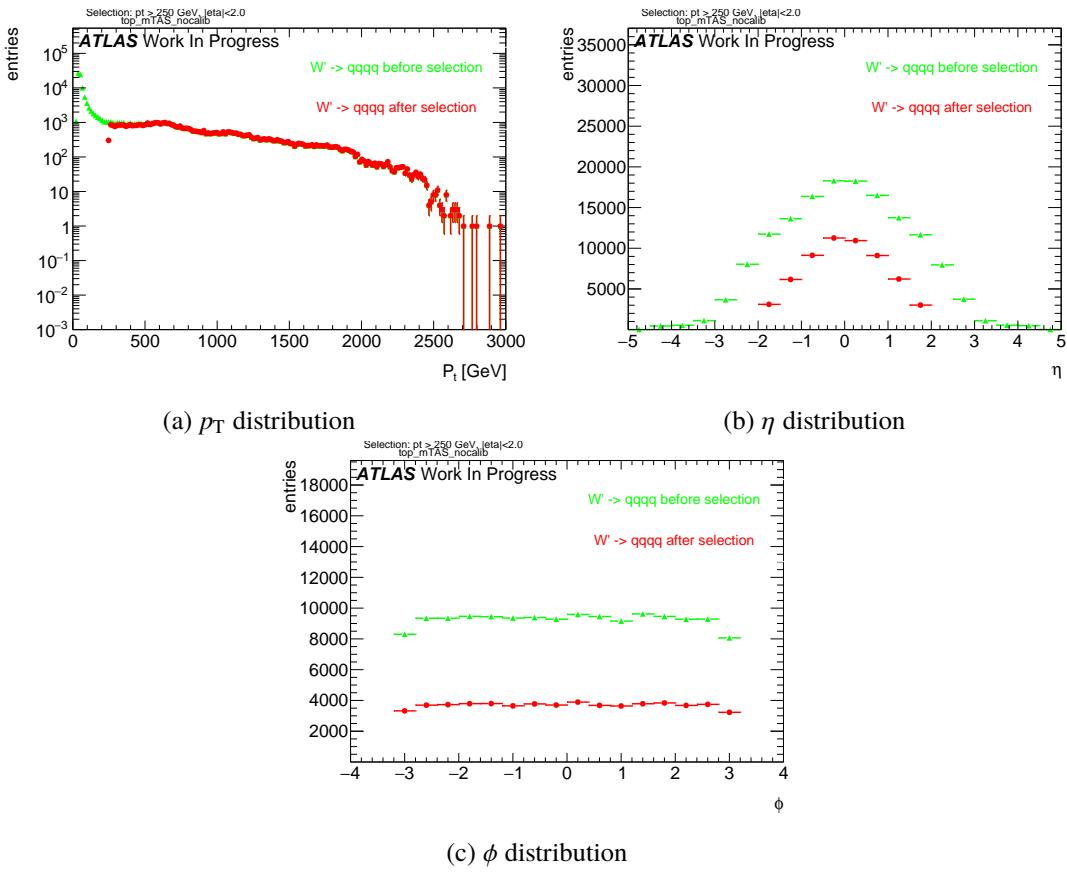


Figure 43: Boosted tops kinematic distribution.

902 Appendix

903 inputappendixA1.tex

904 Jet Mass Observable Distribution Kinematic distribution for all the samples, p_T η and ϕ is shown.

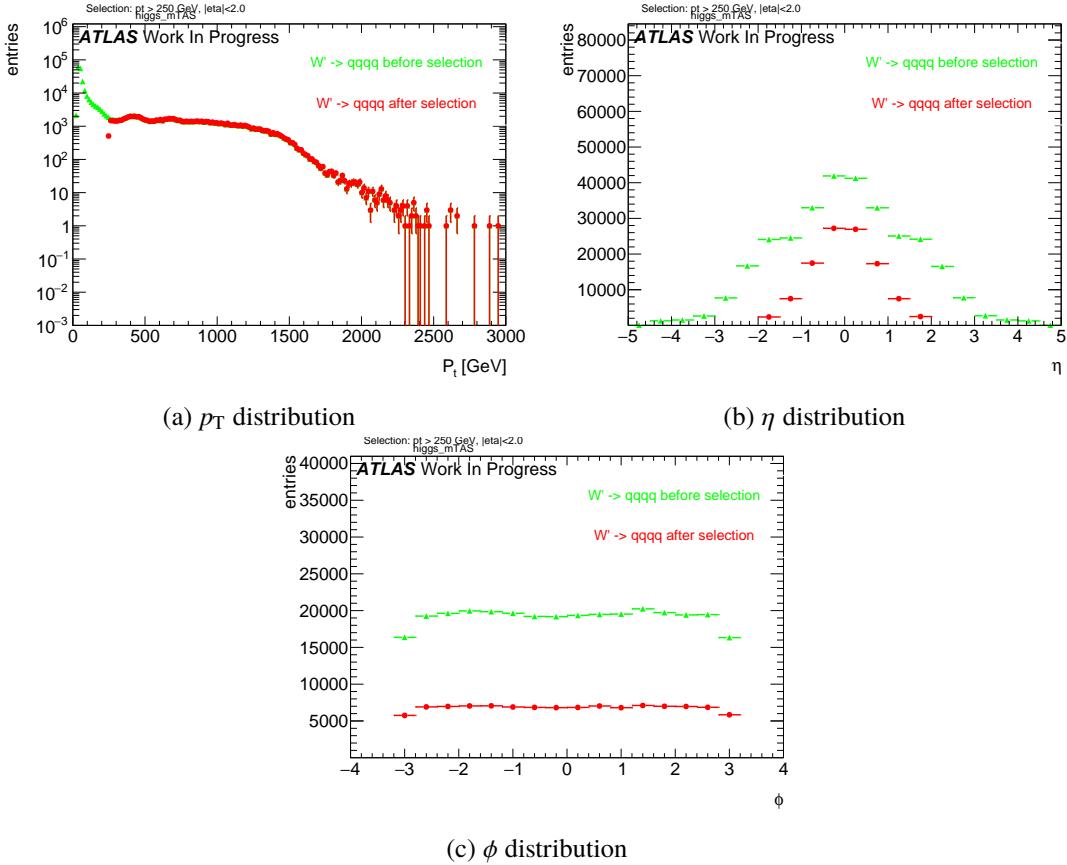


Figure 44: RS-Graviton kinematic distribution.

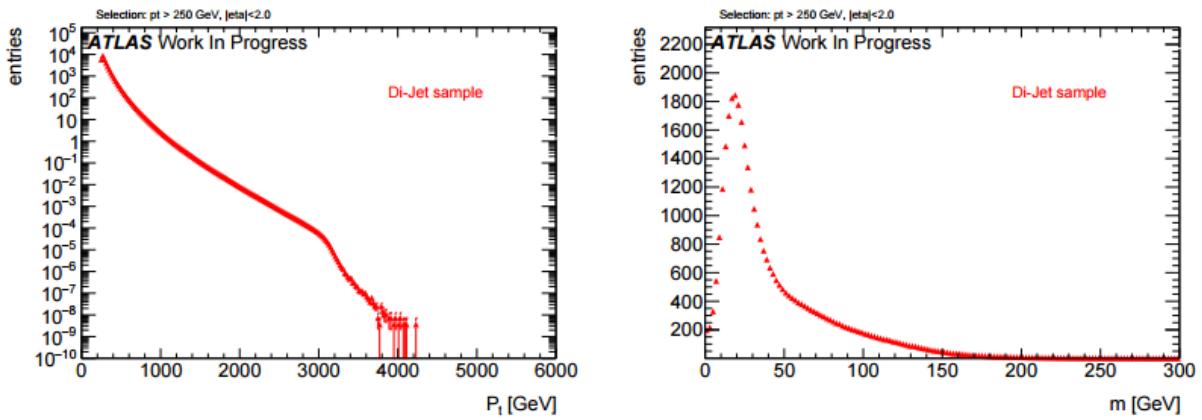


Figure 45: QCD dijet transverse momentum and mass distributions.

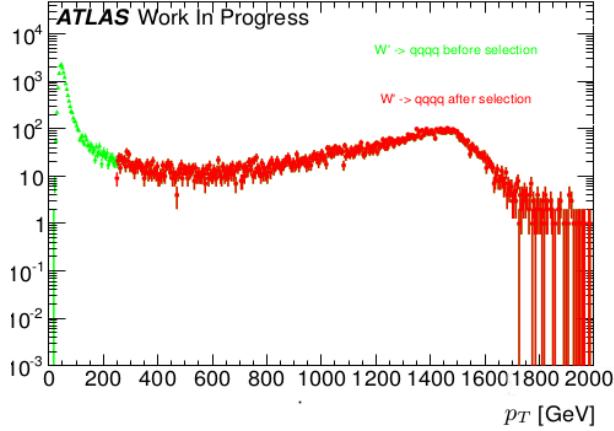


Figure 46: The p_T distribution of a 3 TeV resonance from the hadronically decaying W or Z , in logarithmic plot. As can be seen, the jacobian peak is around $p_T \simeq m_{W'}/2 \simeq 1.5$ TeV.

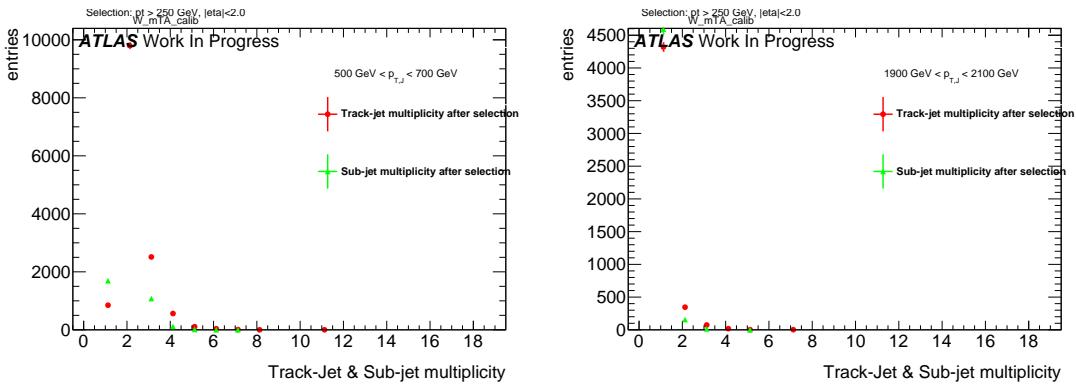


Figure 47: Sub-jet and Track-jet (jets created having tracks as input) multiplicity, for selected bins of transverse momentum.

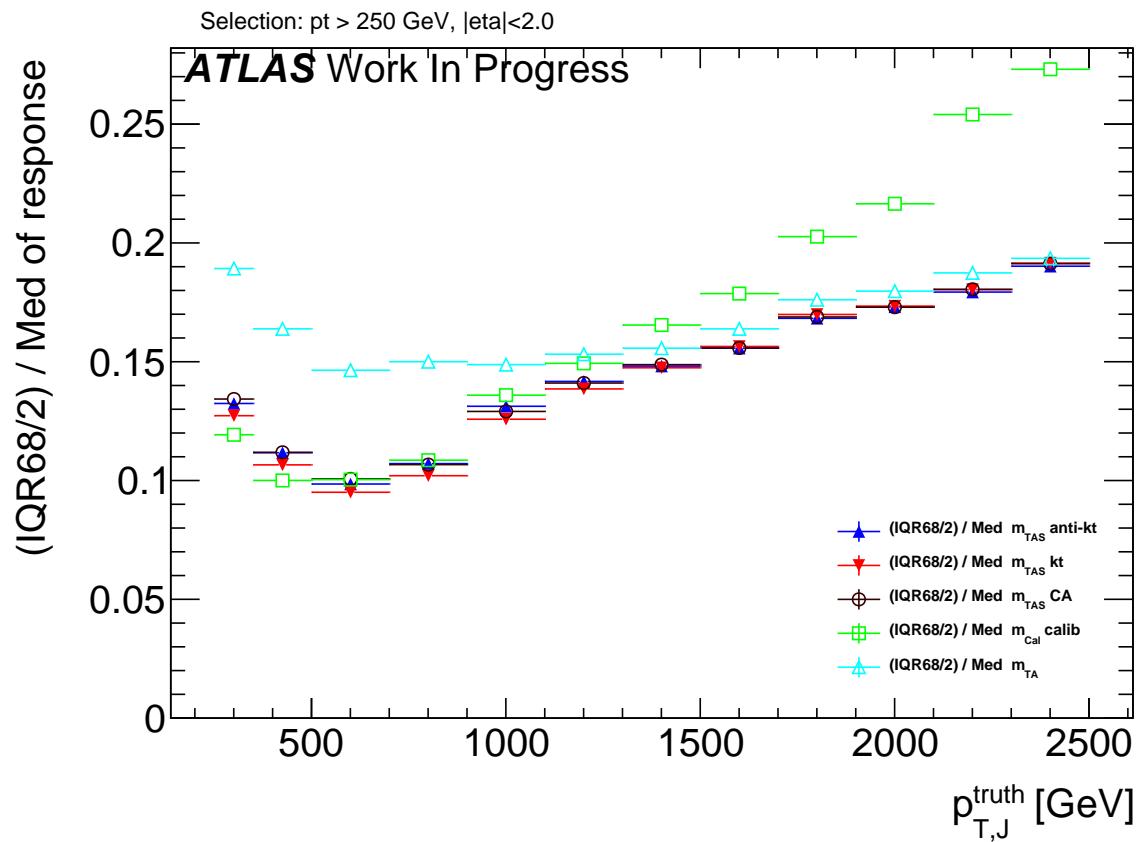


Figure 48: Performance of m^{TAS} with different reclustering algorithm for the sub-jets: anti- k_t , k_t and C/A. Boosted W/Z sample.

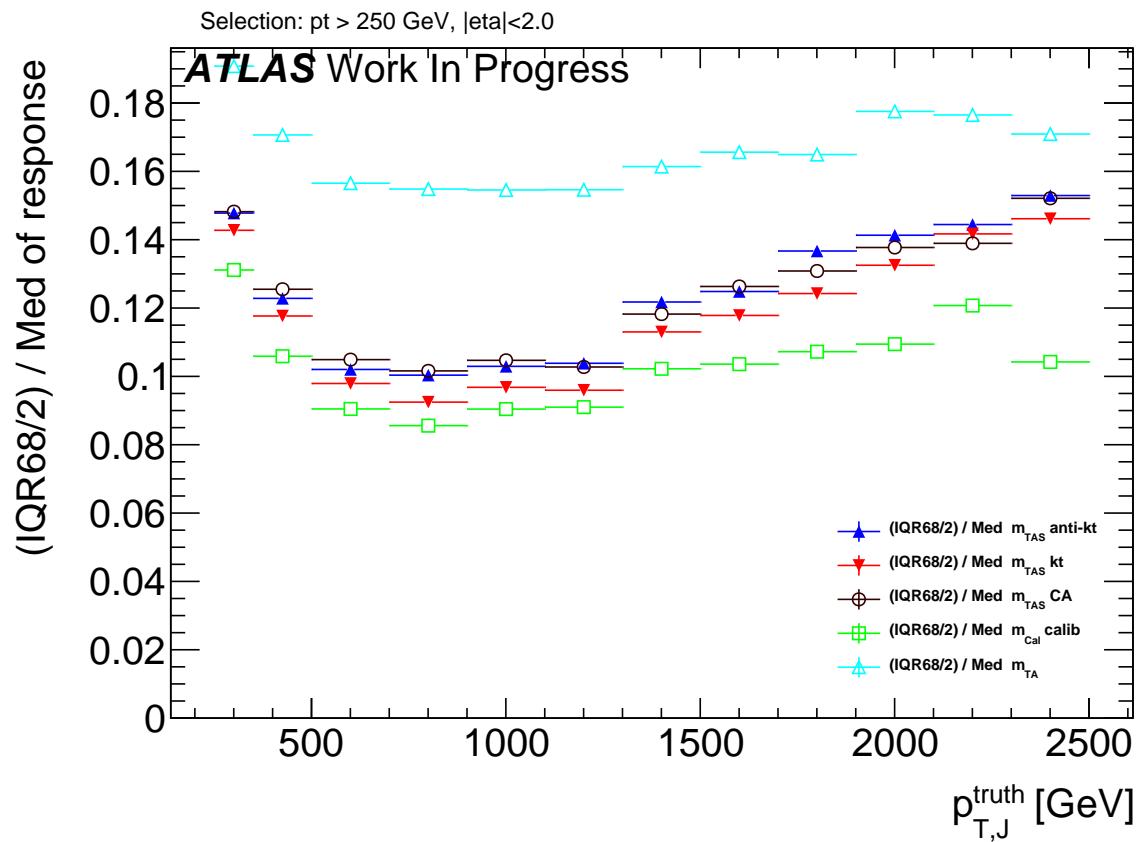


Figure 49: Performance of m^{TAS} with different reclustering algorithm for the sub-jets: anti- k_t , k_t and C/A. Boosted top sample.

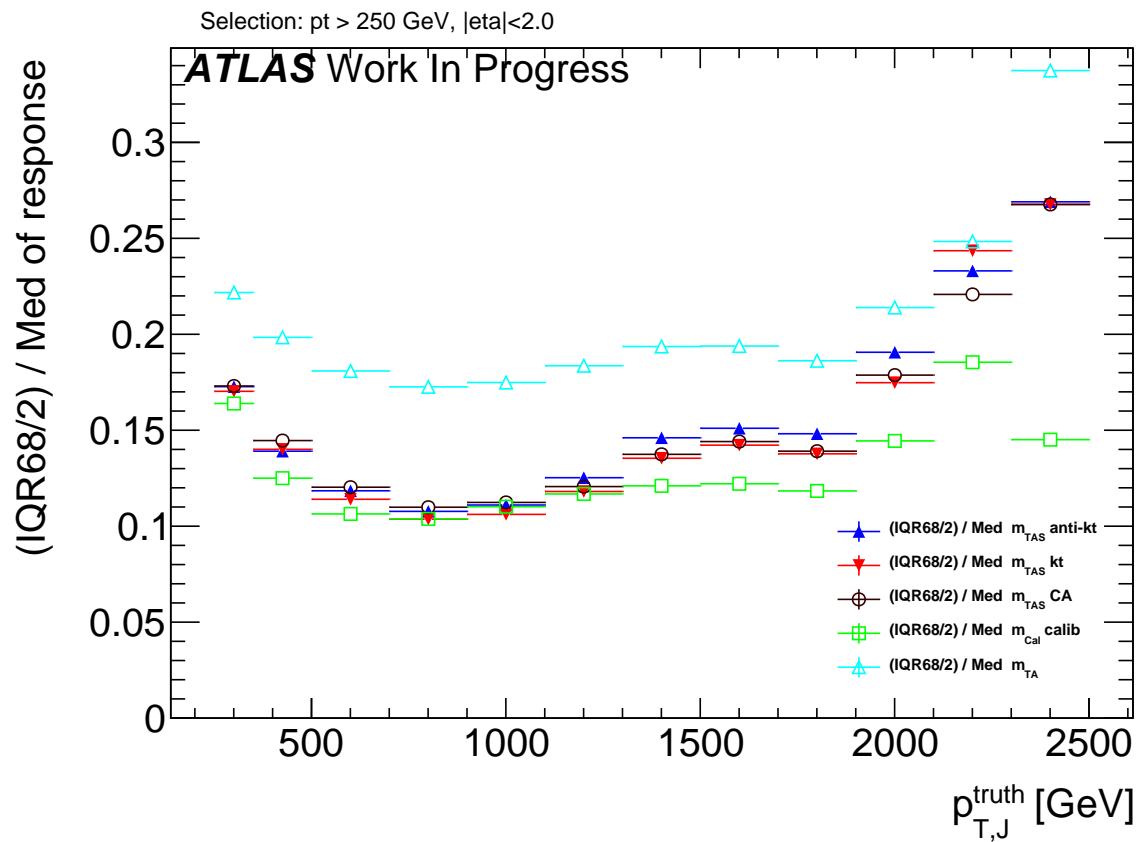


Figure 50: Performance of m^{TAS} with different reclustering algorithm for the sub-jets: anti- k_t , k_t and C/A. Boosted higgs sample.

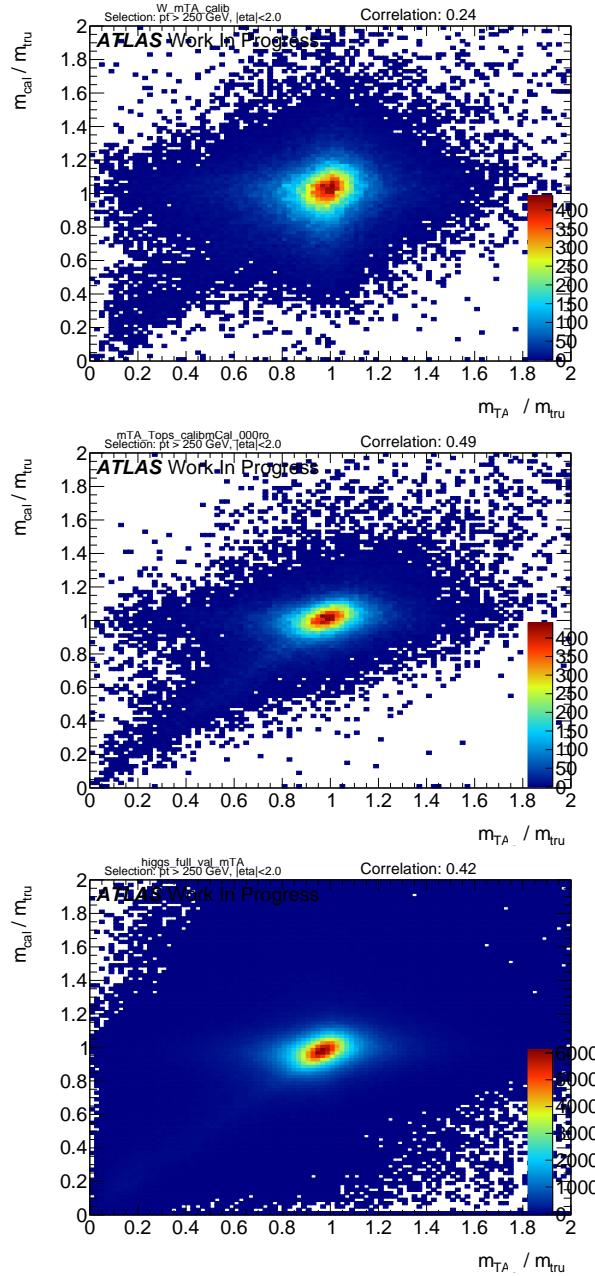


Figure 51: Calorimeter based jet mass response vs the track-assisted mass response for the three signal samples. Correlation coefficient is indicated on the top right.

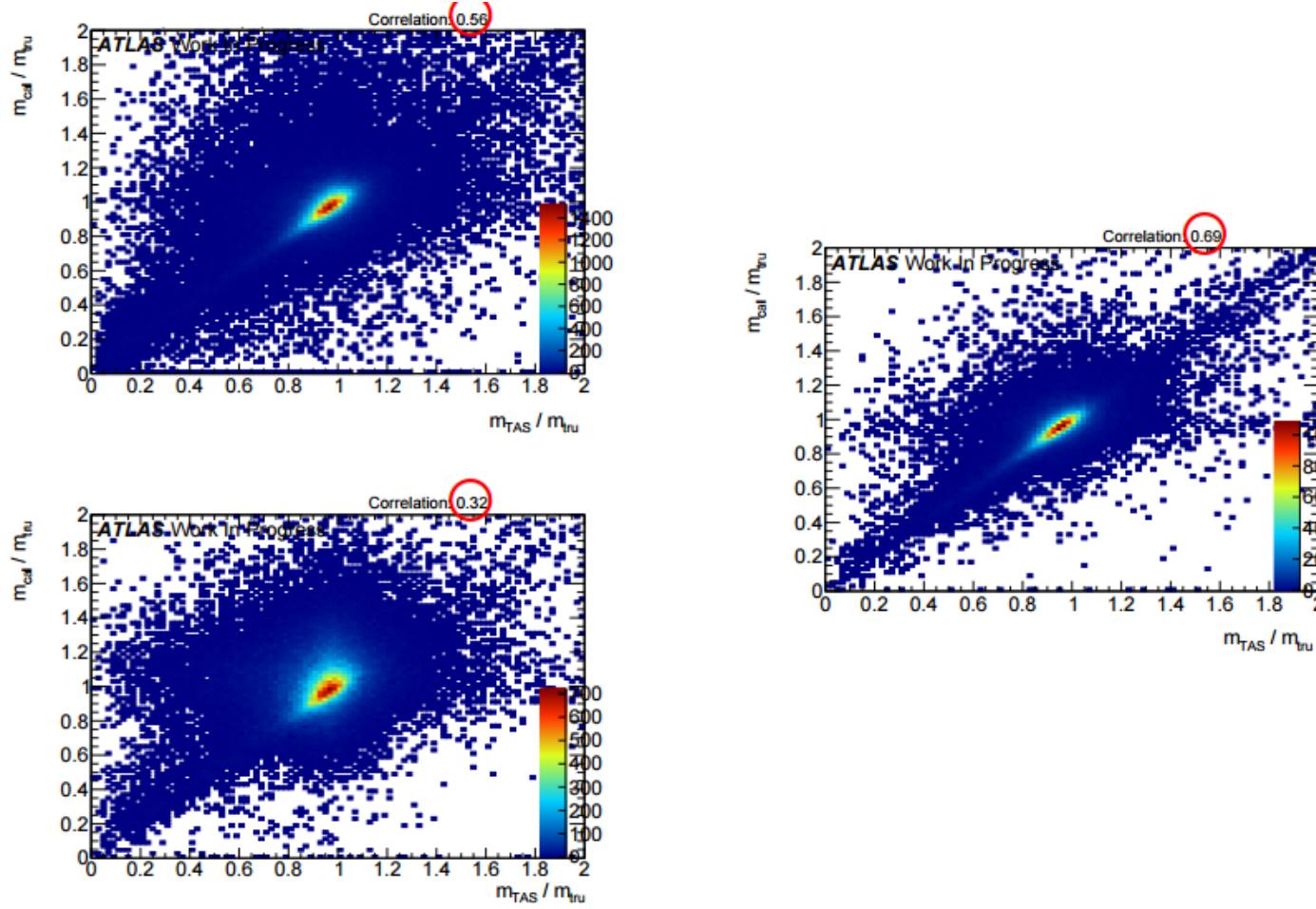


Figure 52: Calorimeter based jet mass response vs the track-assisted sub-jet mass response for the three signal samples. Correlation coefficient is indicated on the top right and highlighted. On the left, top, the higgs sample, bottom, the W/Z ; on the right the top-quark sample.

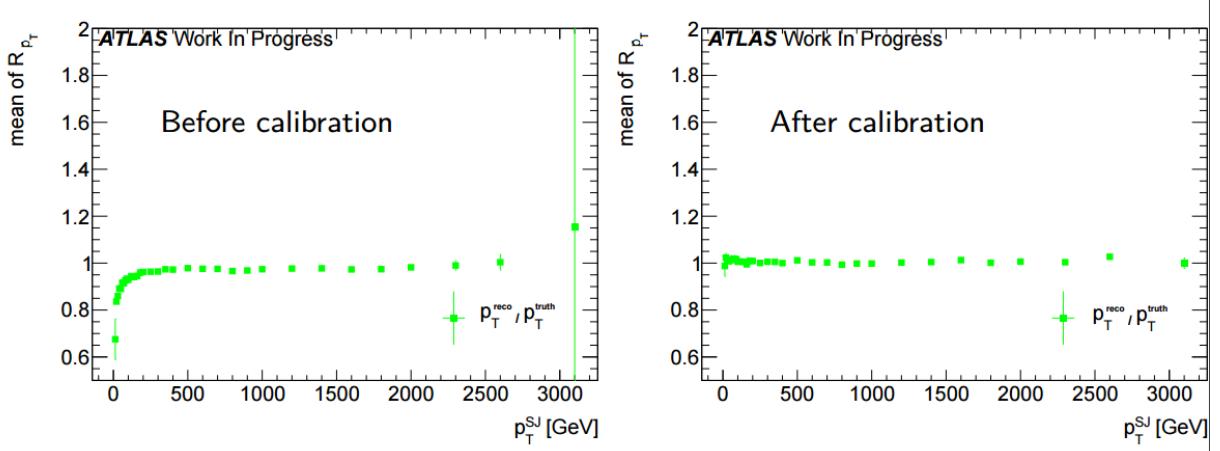


Figure 53: Poor's man calibration effect on mean of transverse momentum's response of the sub-jet, before, left, and after, right, the procedure.

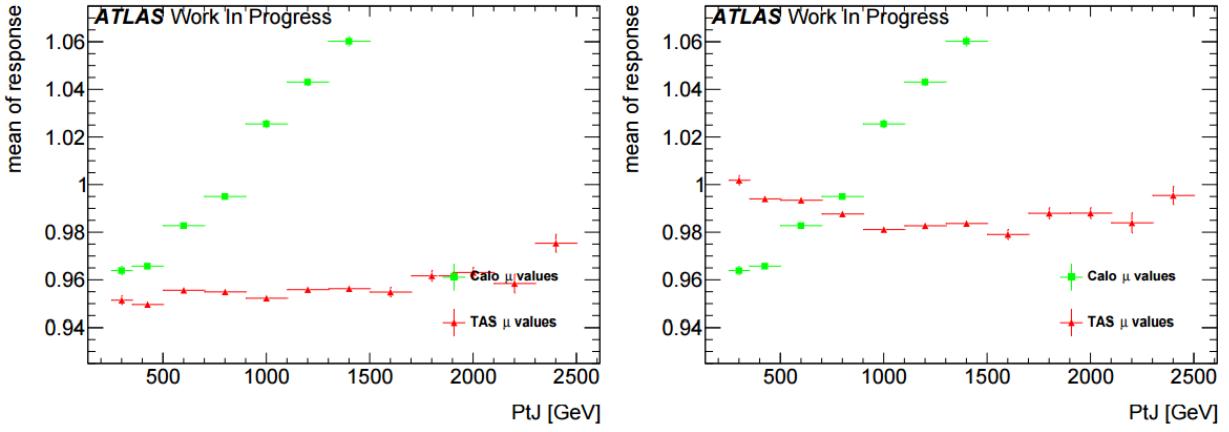


Figure 54: Poor's man calibration effect on the mean of the mass response of the large-R jet, before, left, and after, right, the procedure.

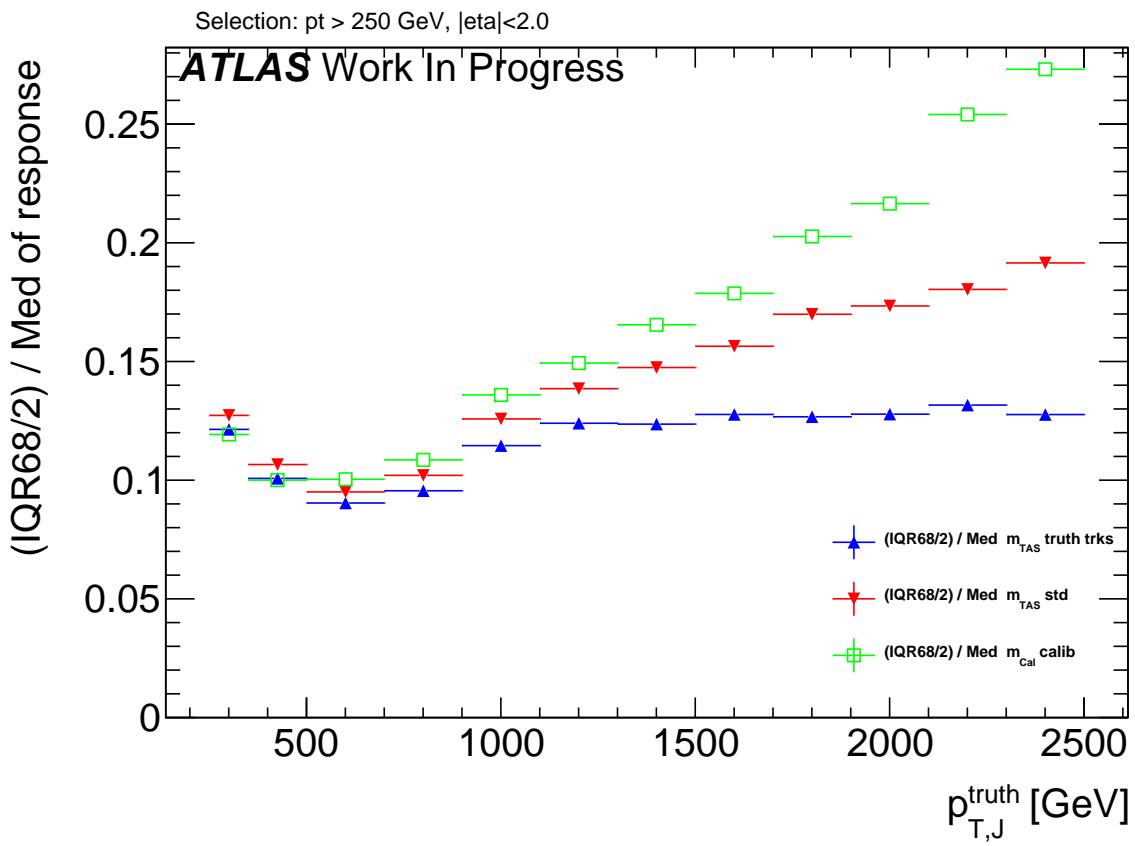
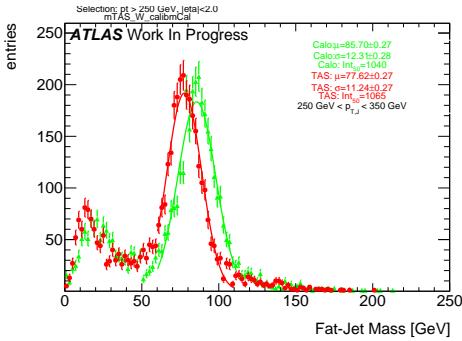
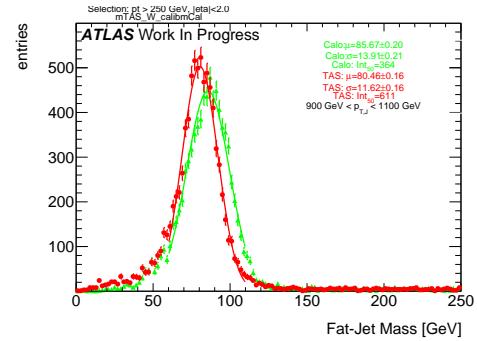
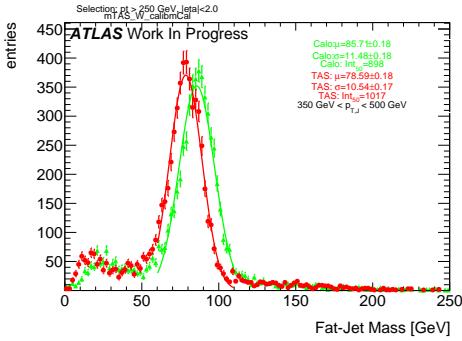
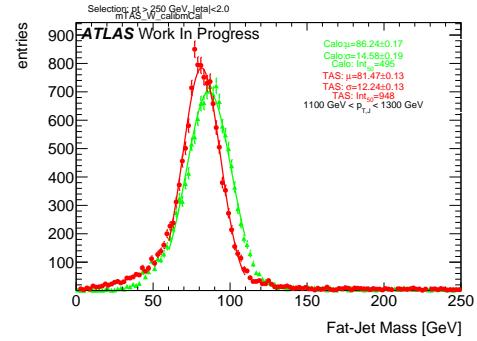
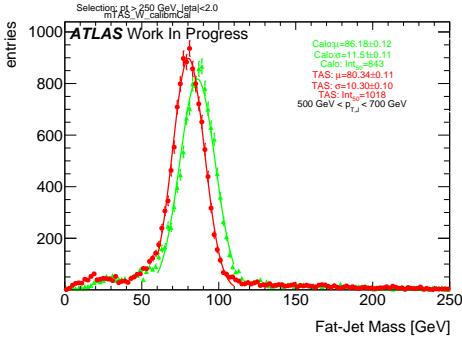
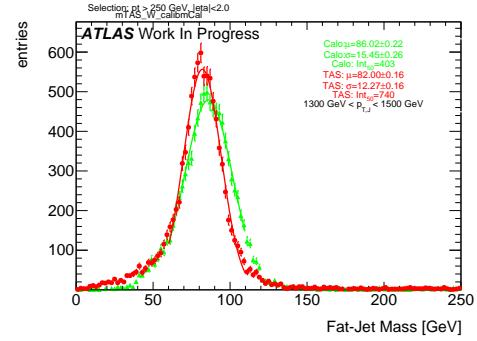
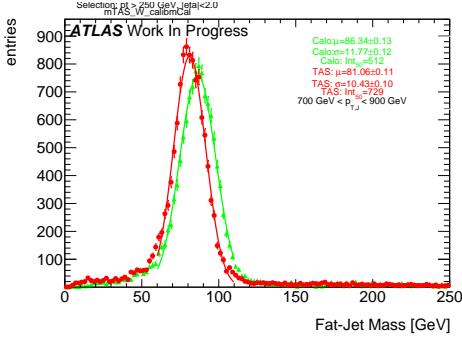
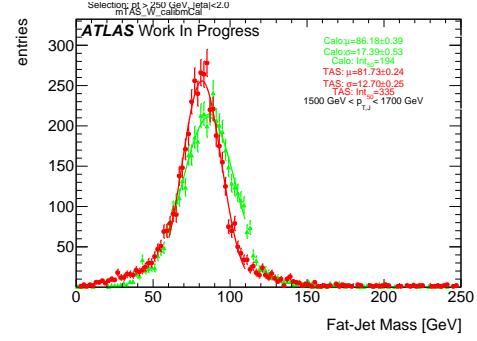


Figure 55: Comparison of the m^{TAS} and the same variable using truth-level information for the tracks.

₉₀₅ A m^{TAS} distributions, boosted W/Z

Figure 56: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 60: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 57: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 61: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 58: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 62: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 59: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 63: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

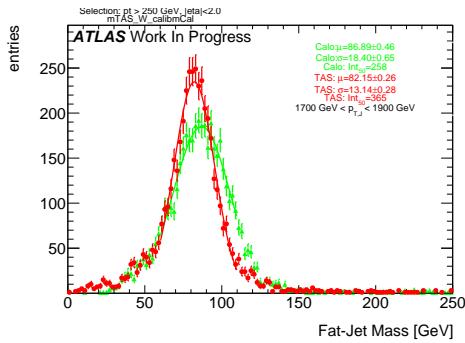


Figure 64: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

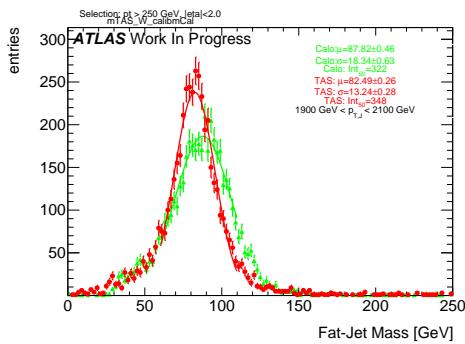


Figure 65: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

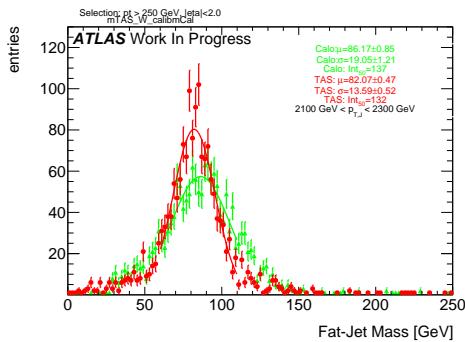


Figure 66: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

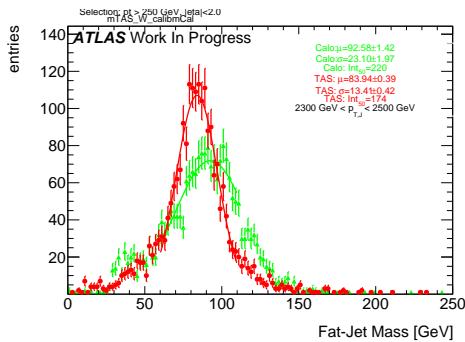


Figure 67: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

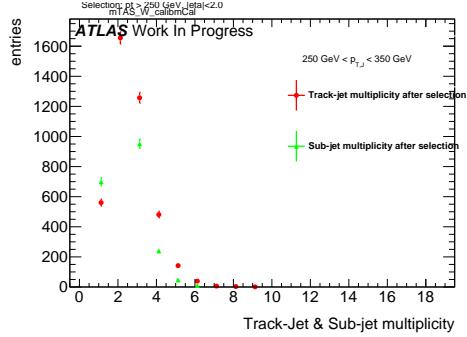


Figure 68: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

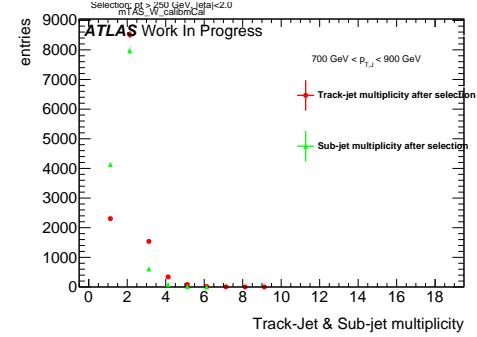


Figure 71: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

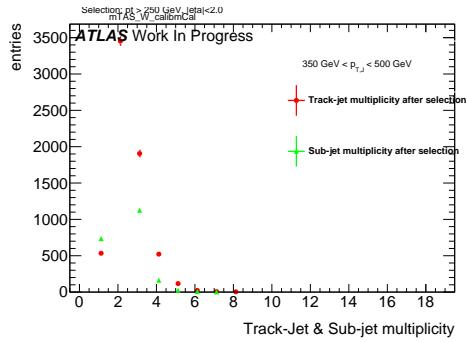


Figure 69: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

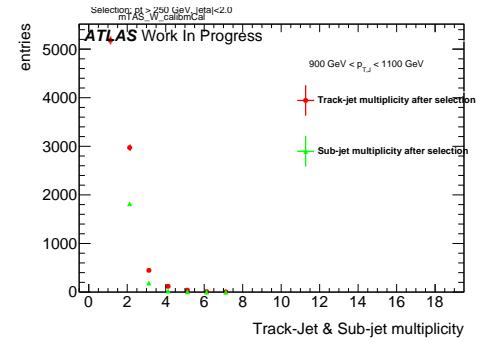


Figure 72: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

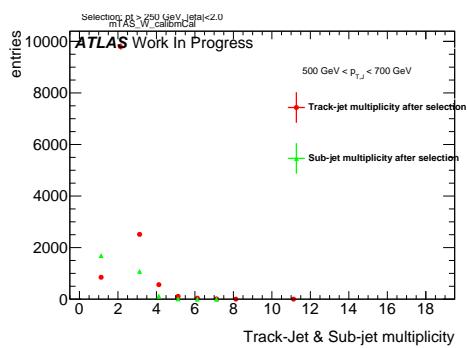


Figure 70: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

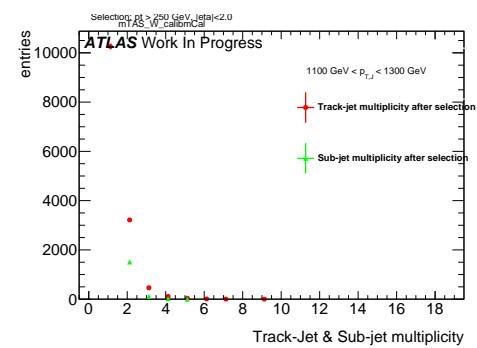


Figure 73: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

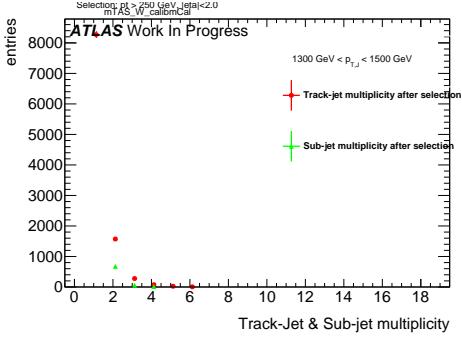


Figure 74: Track-jet $R=0.2$ and sub-jet multiplicity for p_T^J bin (indicated on plot)

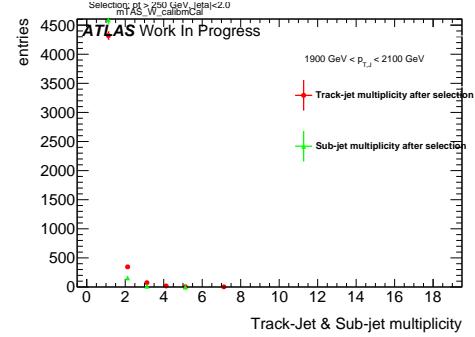


Figure 77: Track-jet $R=0.2$ and sub-jet multiplicity for p_T^J bin (indicated on plot)

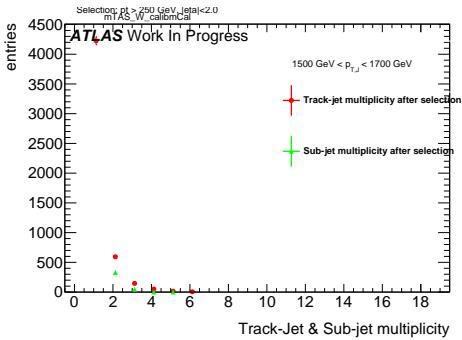


Figure 75: Track-jet $R=0.2$ and sub-jet multiplicity for p_T^J bin (indicated on plot)

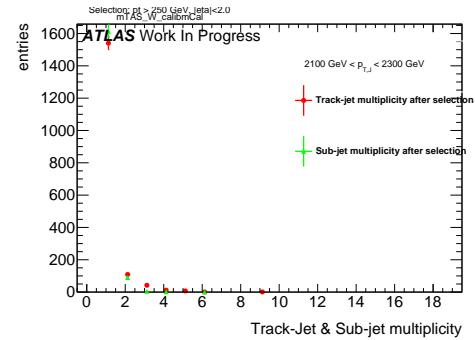


Figure 78: Track-jet $R=0.2$ and sub-jet multiplicity for p_T^J bin (indicated on plot)

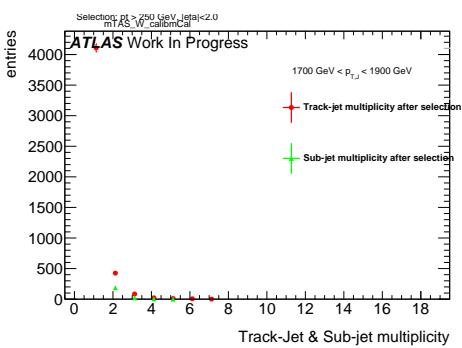


Figure 76: Track-jet $R=0.2$ and sub-jet multiplicity for p_T^J bin (indicated on plot)

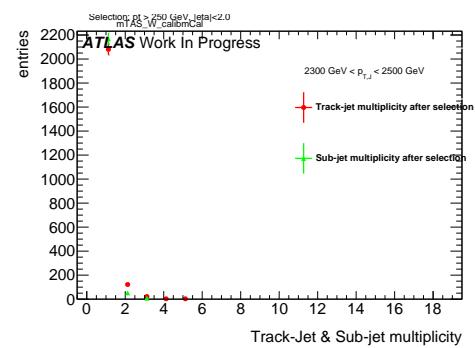


Figure 79: Track-jet $R=0.2$ and sub-jet multiplicity for p_T^J bin (indicated on plot)

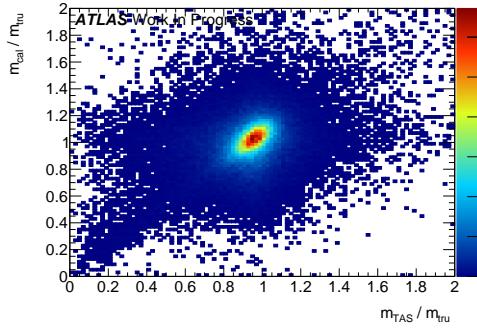
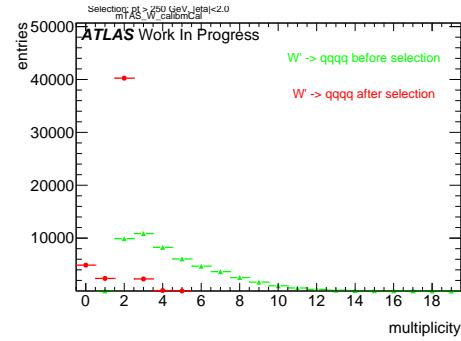
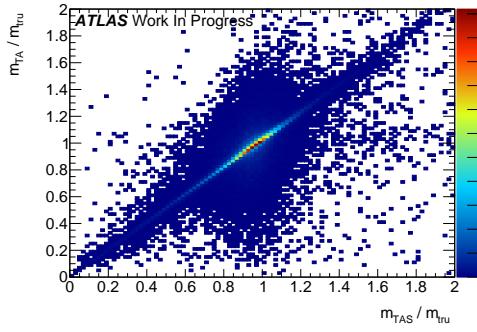
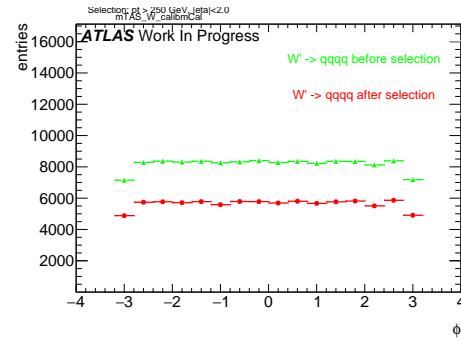
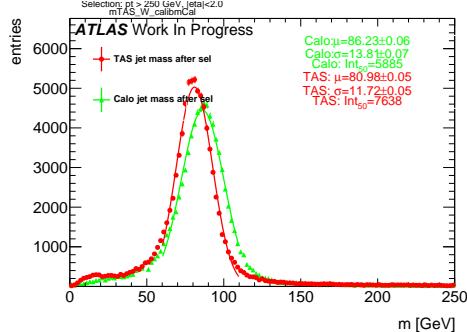
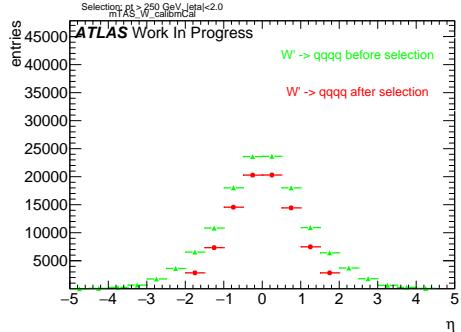
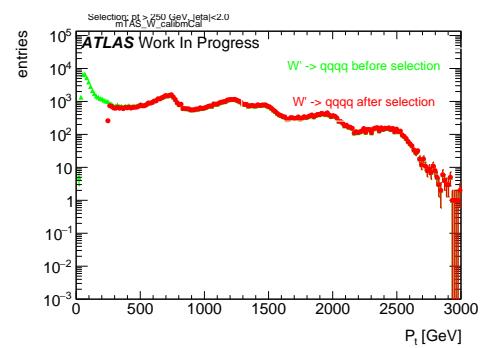
Figure 80: Scatter plot m^{TAS} versus m^{calo} responses

Figure 84: large-R jet Multiplicity, before and after selection

Figure 81: Scatter plot m^{TAS} versus m^TA responsesFigure 85: ϕ distribution of the large-R jet, before and after selectionFigure 82: m^{TAS} distribution in all the p_T binsFigure 83: η distribution of the large-R jet, before and after selectionFigure 86: p_T distribution of the large-R jet, before and after selection

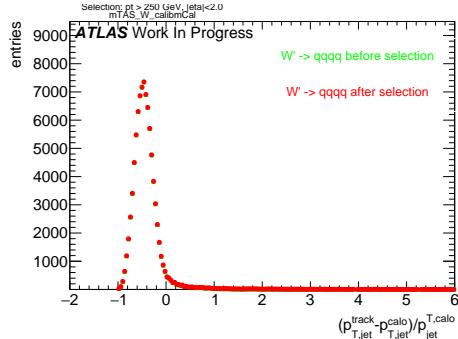


Figure 87: p_T resolution: $\frac{p_{T,jet}^{track} - p_{T,jet}^{calo}}{p_{T,jet}^{calo}}$, before and after selection

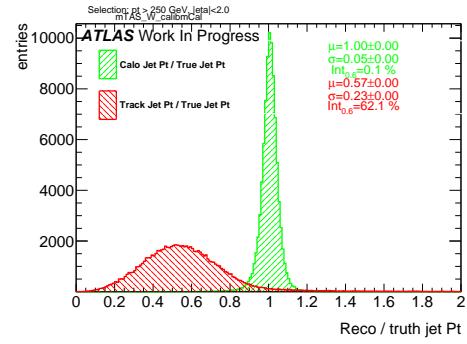


Figure 90: Transverse momentum response p_T^{Reco} / p_T^{Truth} for calorimeter and tracks

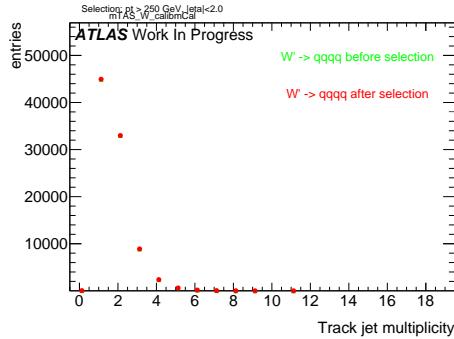


Figure 88: Multiplicity of track-jets R=0.2 per large-R jet

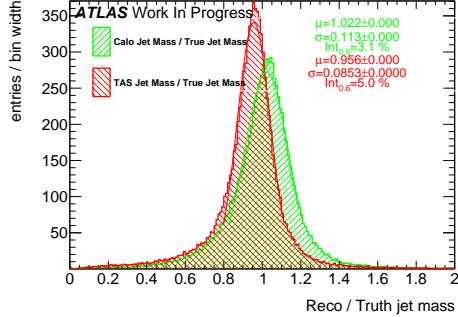


Figure 89: Response m^{Reco} / m^{Truth} for all the p_T bins

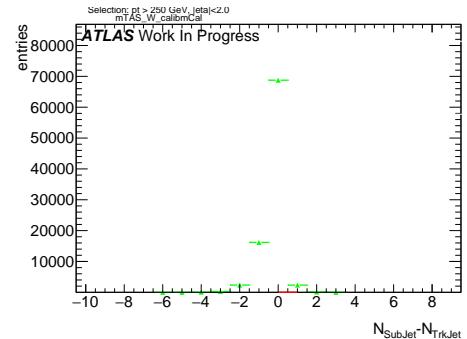


Figure 91: sub-jet - track-jet Multiplicity

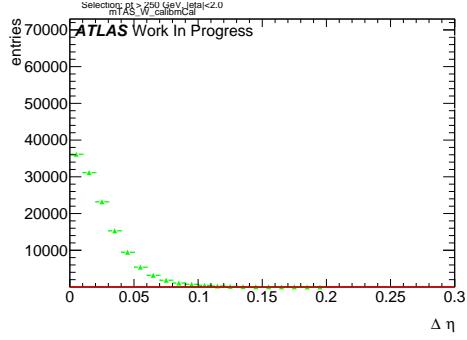


Figure 92: $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

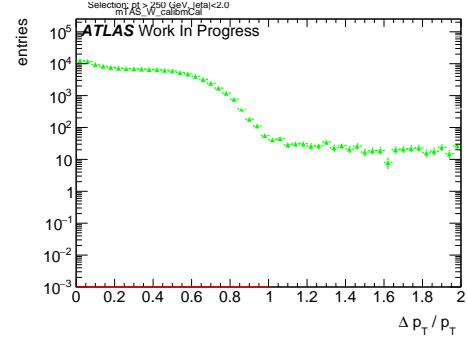


Figure 95: $|p_{T,\text{sub-jet}} - p_{T,\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

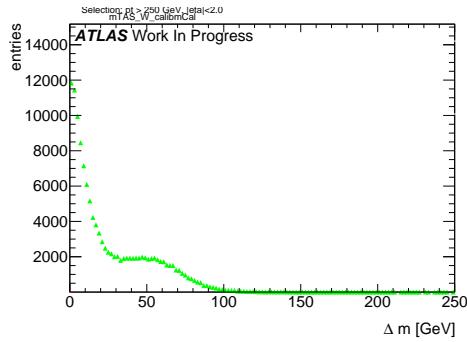


Figure 93: $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

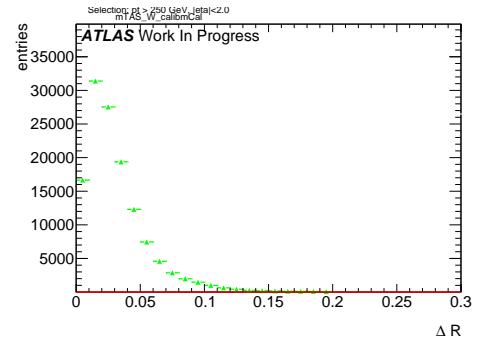


Figure 96: $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

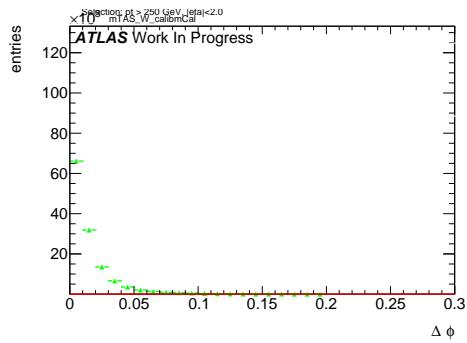


Figure 94: $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

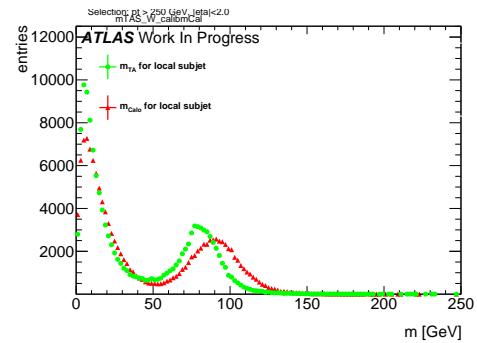


Figure 97: Mass distribution of the sub-jet, calorimeter and track-assisted

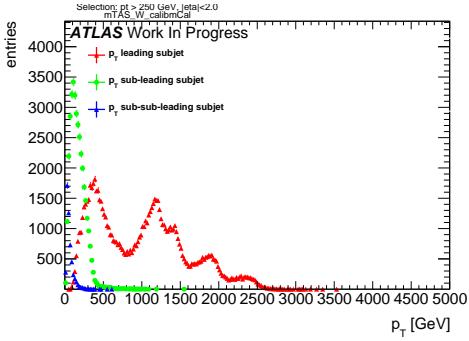


Figure 98: p_T distribution for leading, sub-leading and sub-sub-leading sub-jets

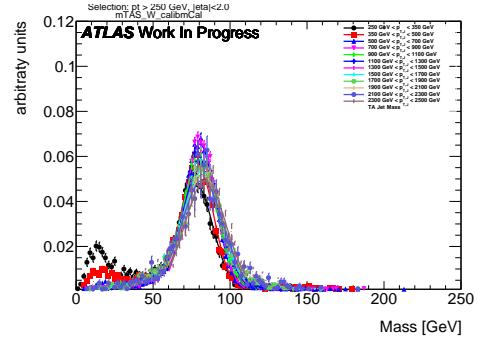


Figure 101: m^{TAS} for p_T^J bin, superimposed

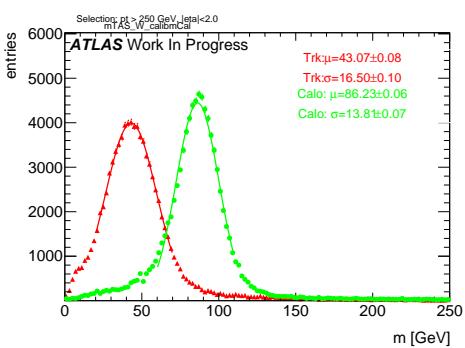


Figure 99: Mass distribution for calorimeter and tracks associated to the large- R jet

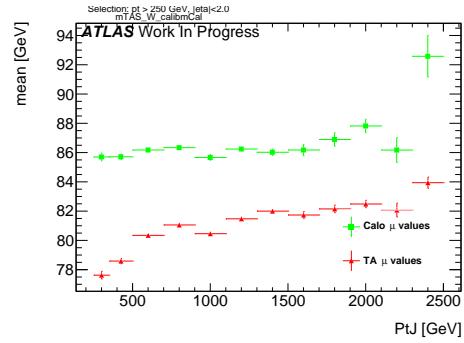


Figure 102: μ from fit of the mass distribution vs bin of p_T^J

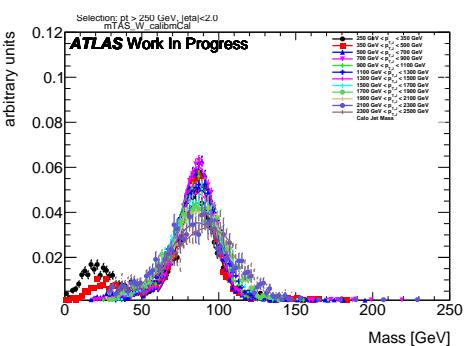


Figure 100: m^{calo} for p_T^J bin, superimposed

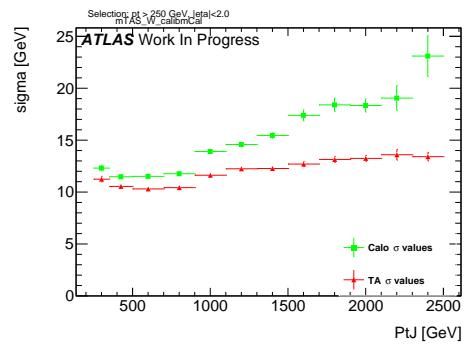
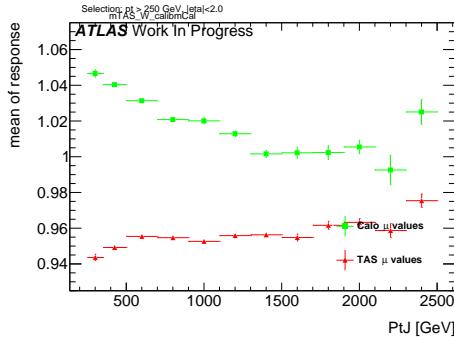
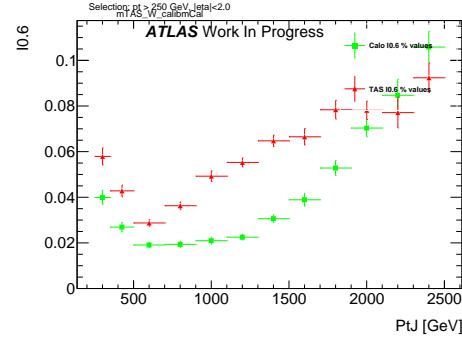
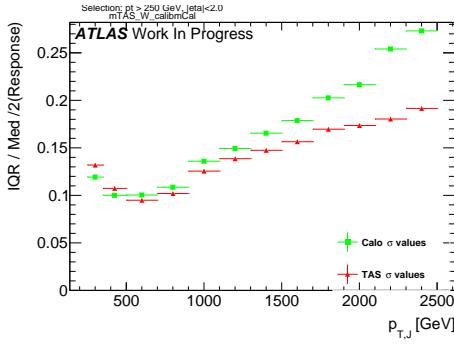
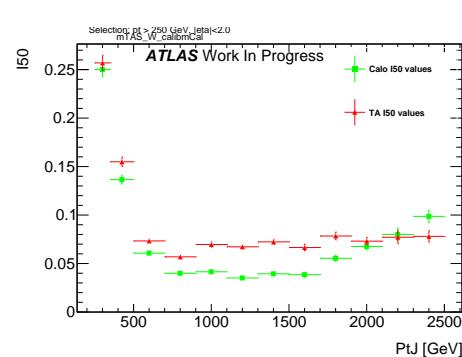
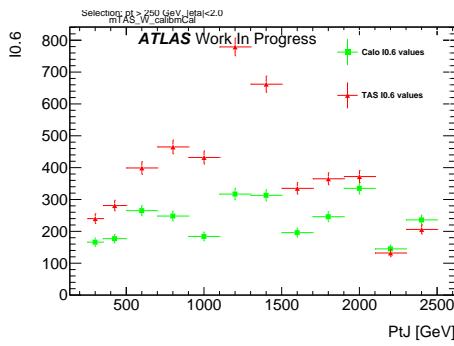
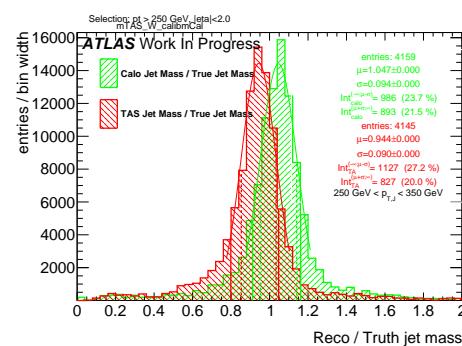
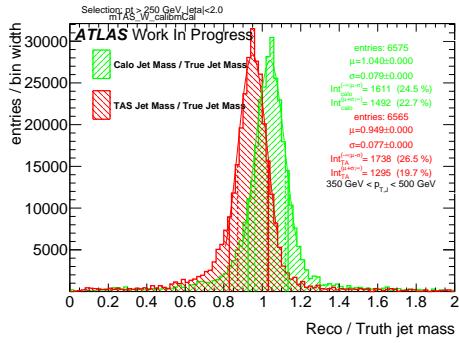
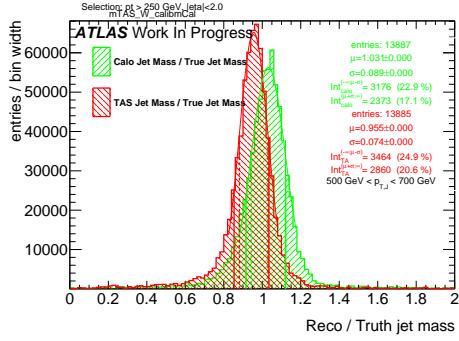
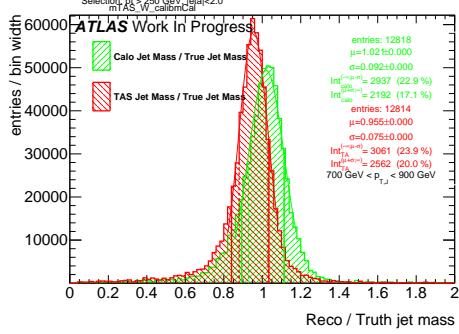
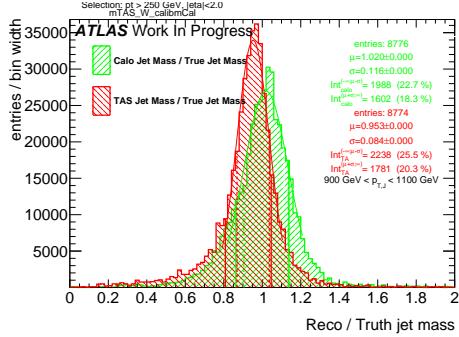
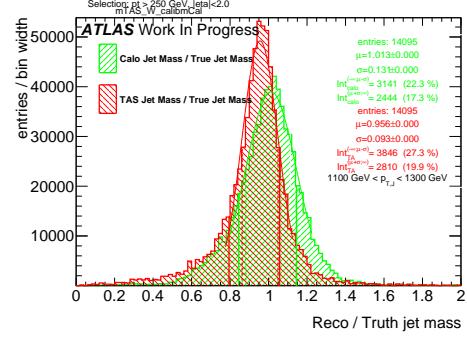
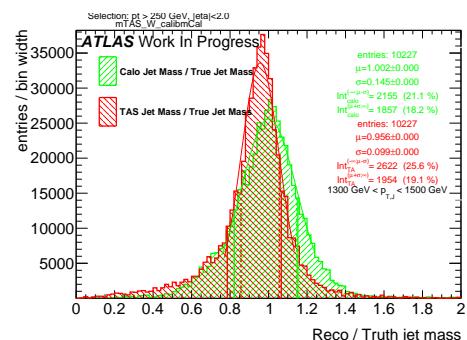
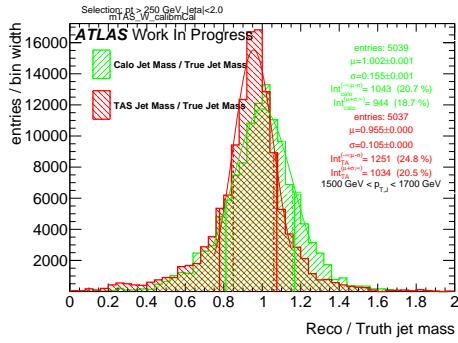
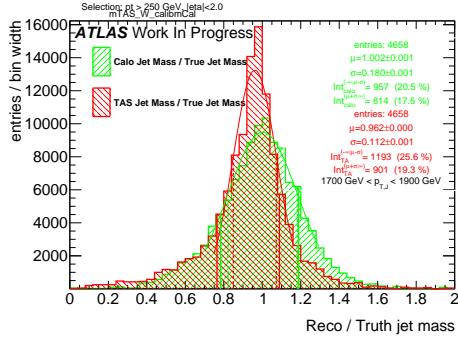
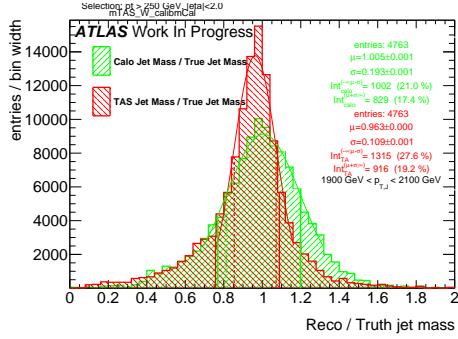
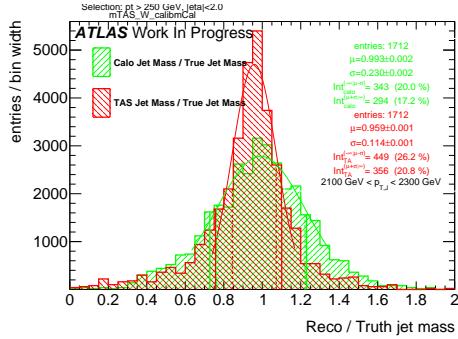
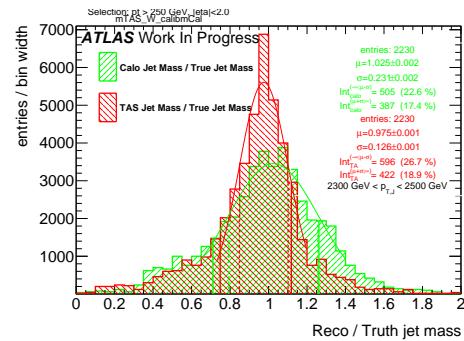


Figure 103: σ from fit of the mass distribution vs bin of p_T^J

Figure 104: μ from fit of the mass Response vs bin of p_T^J Figure 107: Left integral normalized, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J Figure 105: σ from fit of the mass Response vs bin of p_T^J Figure 108: $\int_0^{50 GeV}$ from fit of the mass distribution vs bin of p_T^J (normalized)Figure 106: Left integral, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J Figure 109: Response in bin of p_T^J (indicated on plot)

Figure 110: Response in bin of p_T^J (indicated on plot)Figure 111: Response in bin of p_T^J (indicated on plot)Figure 112: Response in bin of p_T^J (indicated on plot)Figure 113: Response in bin of p_T^J (indicated on plot)Figure 114: Response in bin of p_T^J (indicated on plot)Figure 115: Response in bin of p_T^J (indicated on plot)

Figure 116: Response in bin of p_T^J (indicated on plot)Figure 117: Response in bin of p_T^J (indicated on plot)Figure 118: Response in bin of p_T^J (indicated on plot)Figure 119: Response in bin of p_T^J (indicated on plot)Figure 120: Response in bin of p_T^J (indicated on plot)

906 **B m^{TAS} distributions, boosted tops**

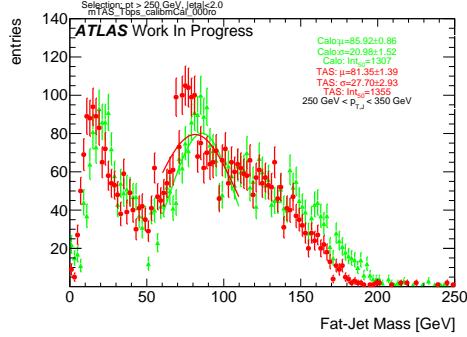


Figure 121: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

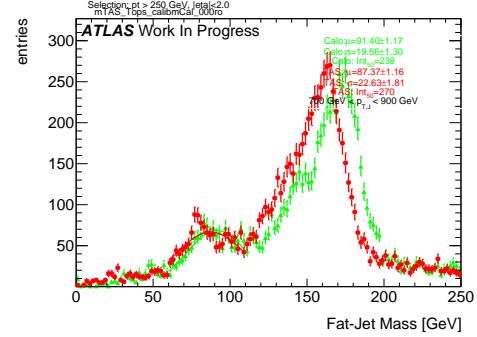


Figure 124: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

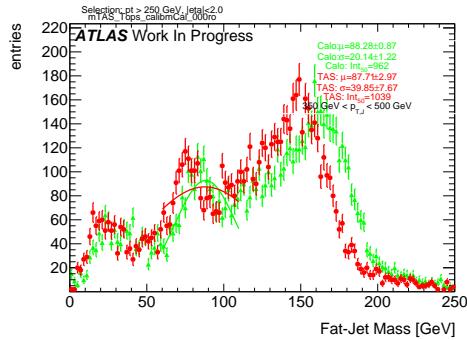


Figure 122: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

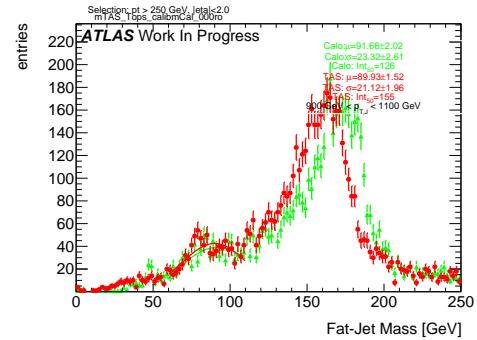


Figure 125: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

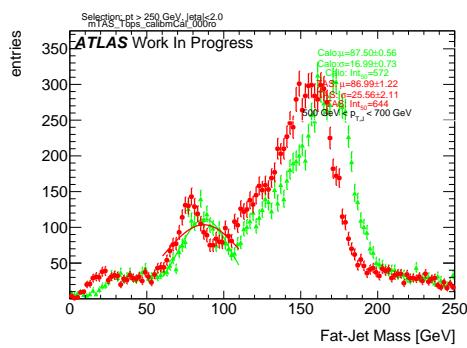


Figure 123: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

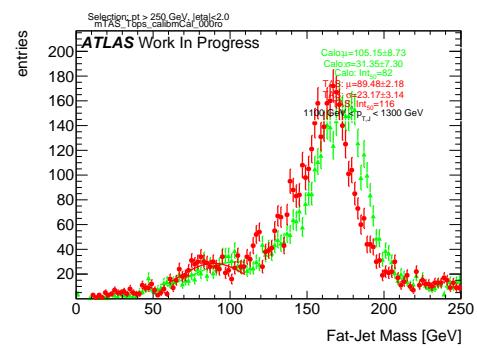


Figure 126: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

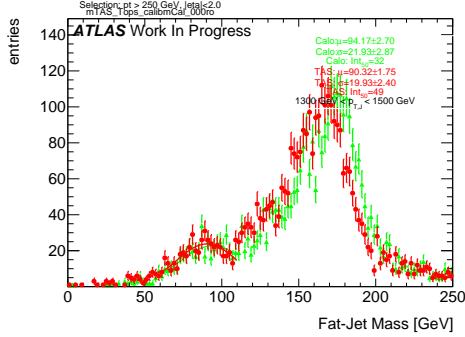


Figure 127: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

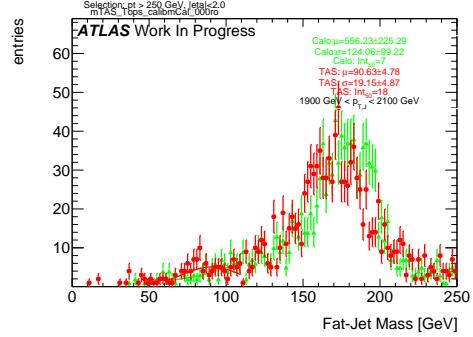


Figure 130: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

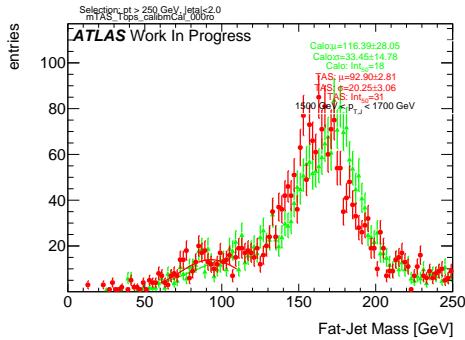


Figure 128: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

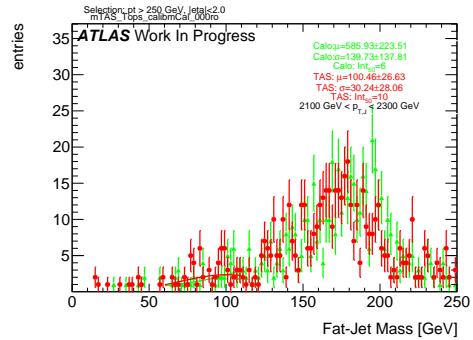


Figure 131: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

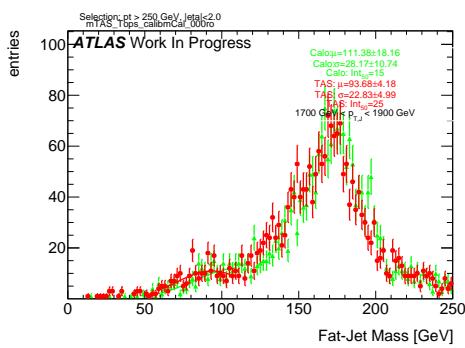


Figure 129: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

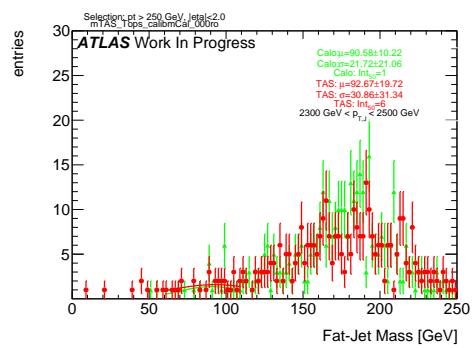


Figure 132: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

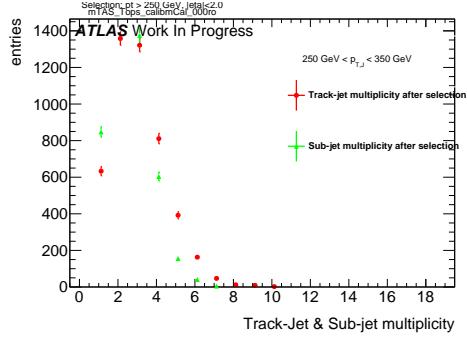


Figure 133: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

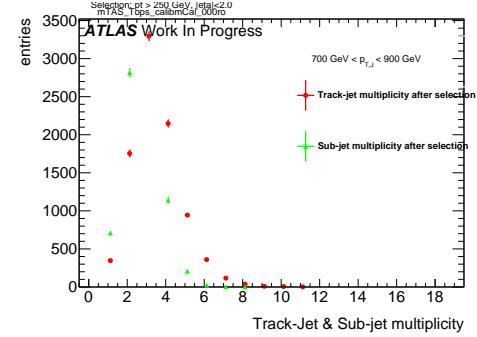


Figure 136: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

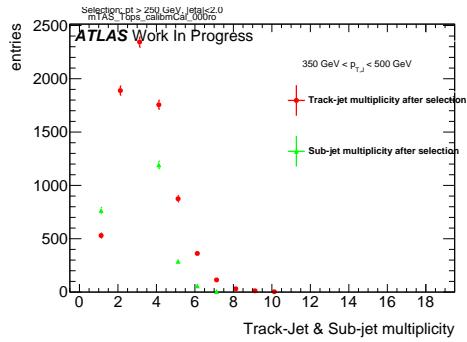


Figure 134: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

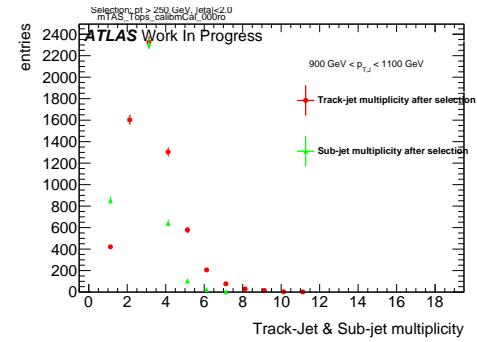


Figure 137: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

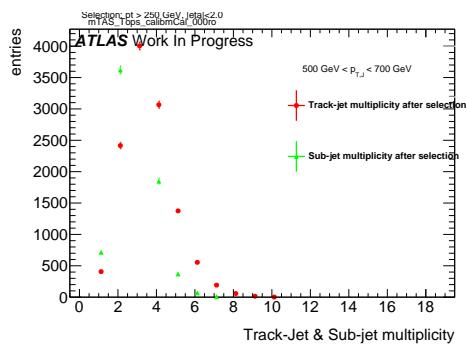


Figure 135: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

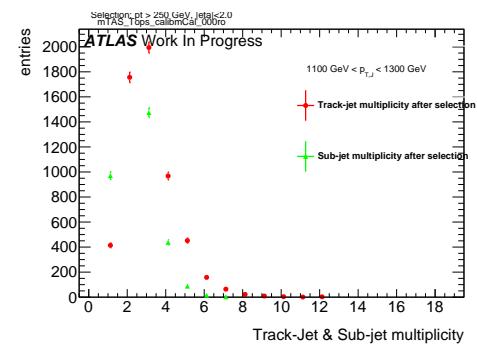


Figure 138: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

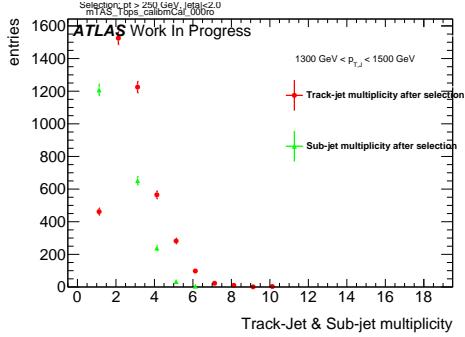


Figure 139: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

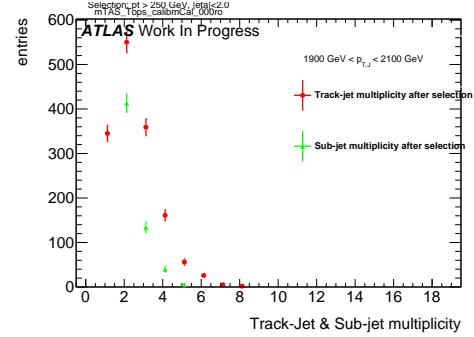


Figure 142: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

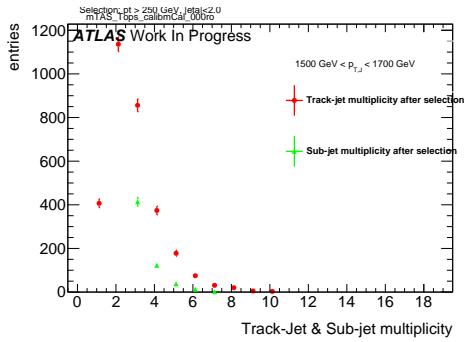


Figure 140: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

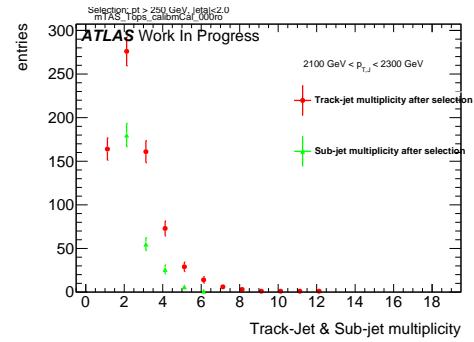


Figure 143: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

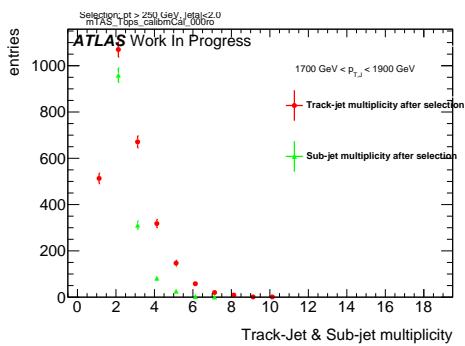


Figure 141: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

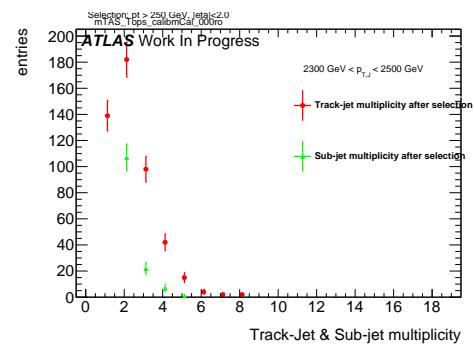


Figure 144: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

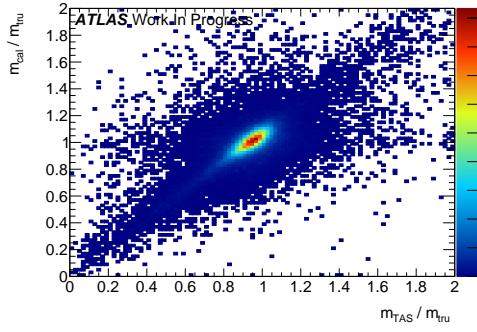
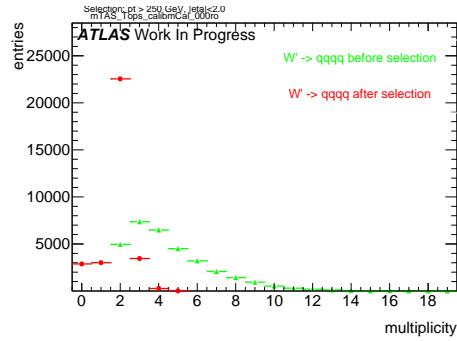
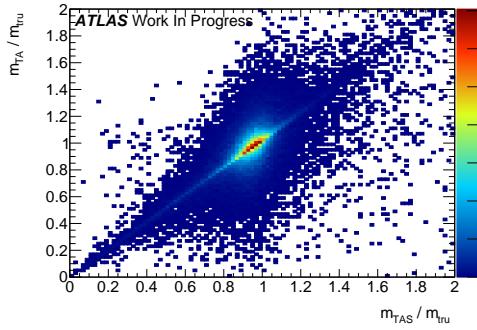
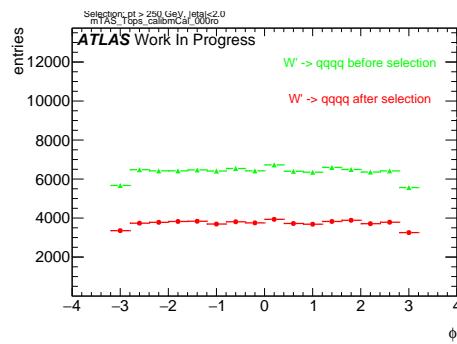
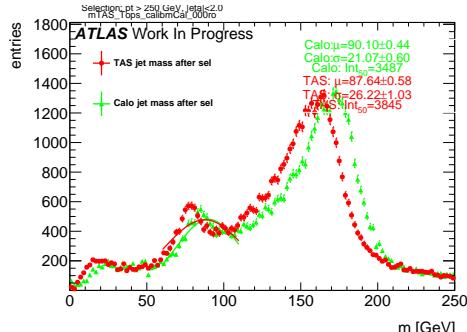
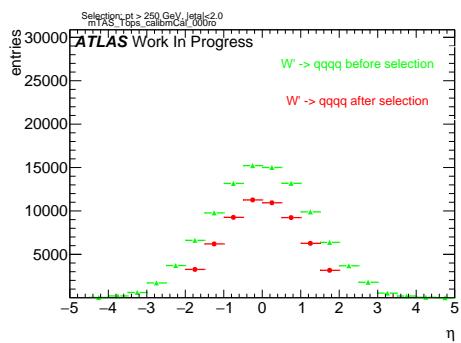
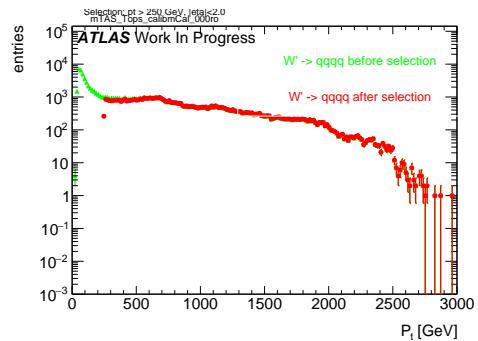
Figure 145: Scatter plot m^{TAS} versus m^{calo} responses

Figure 149: large-R jet Multiplicity, before and after selection

Figure 146: Scatter plot m^{TAS} versus m^A responsesFigure 150: ϕ distribution of the large-R jet, before and after selectionFigure 147: m^{TAS} distribution in all the p_T binsFigure 148: η distribution of the large-R jet, before and after selectionFigure 151: p_T distribution of the large-R jet, before and after selection

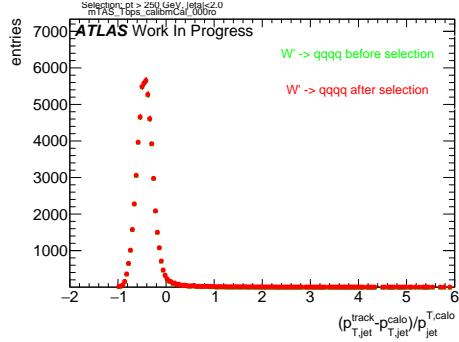


Figure 152: p_T resolution: $\frac{p_{T,jet}^{track} - p_{T,jet}^{fat}}{p_{T,jet}^{fat}}$, before and after selection

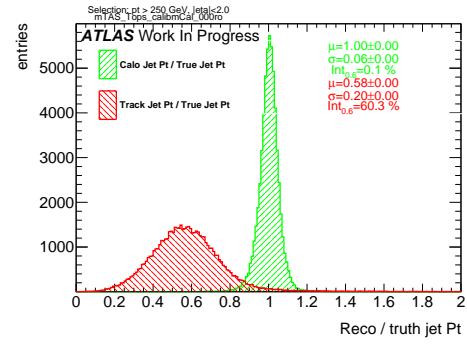


Figure 155: Transverse momentum response p_T^{Reco} / p_T^{Truth} for calorimeter and tracks

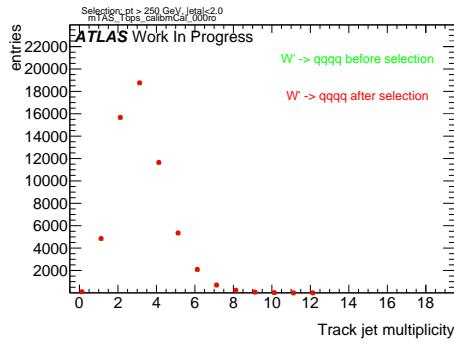


Figure 153: Multiplicity of track-jets R=0.2 per large-R jet

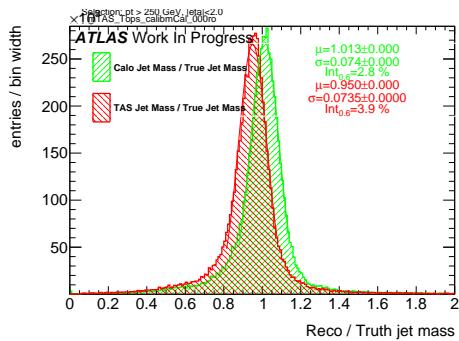


Figure 154: Response m^{Reco} / m^{Truth} for all the p_T bins

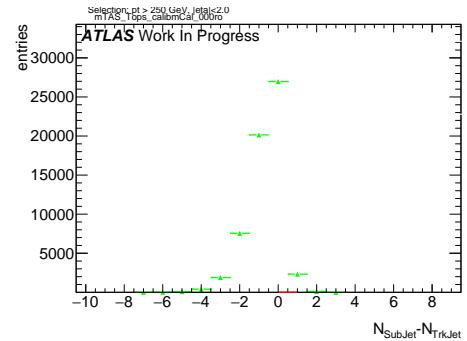


Figure 156: sub-jet - track-jet Multiplicity

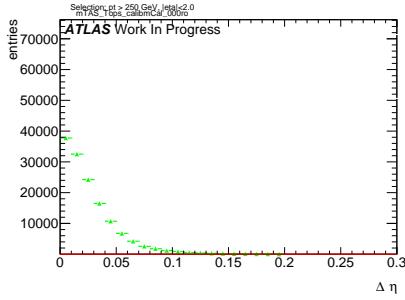


Figure 157: $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

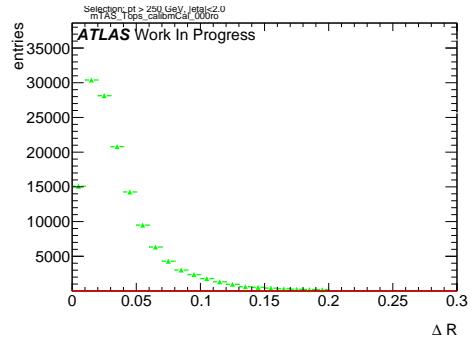


Figure 161: $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

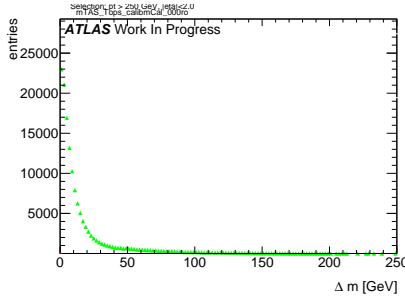


Figure 158: $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

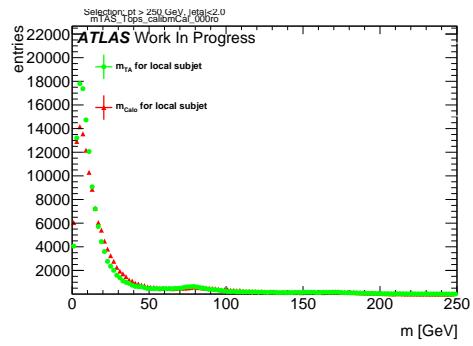


Figure 162: Mass distribution of the sub-jet, calorimeter and track-assisted

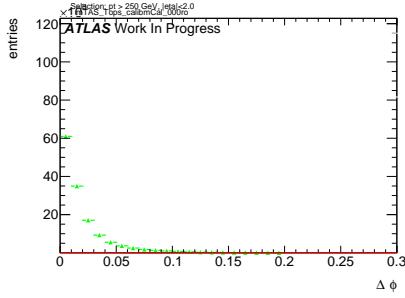


Figure 159: $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

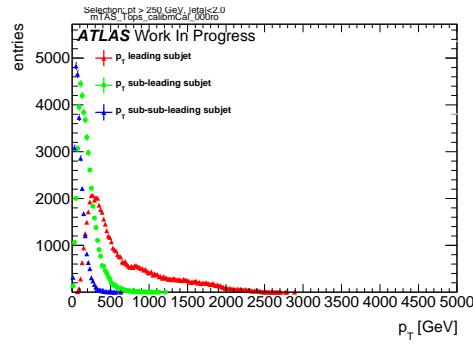


Figure 163: p_T distribution for leading, sub-leading and sub-sub-leading sub-jets

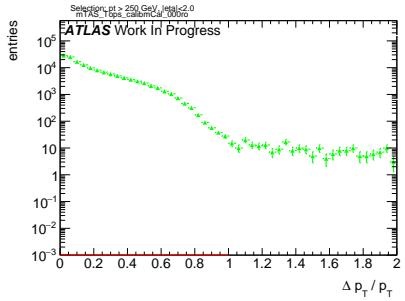


Figure 160: $|p_{T,\text{sub-jet}} - p_{T,\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

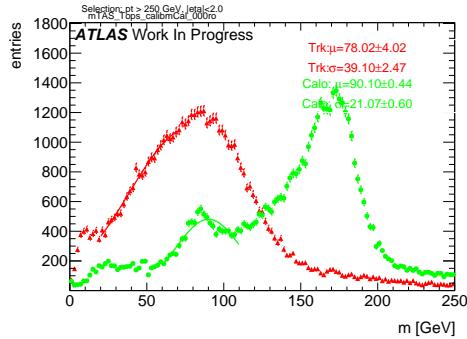


Figure 164: Mass distribution for calorimeter and tracks associated to the large-R jet

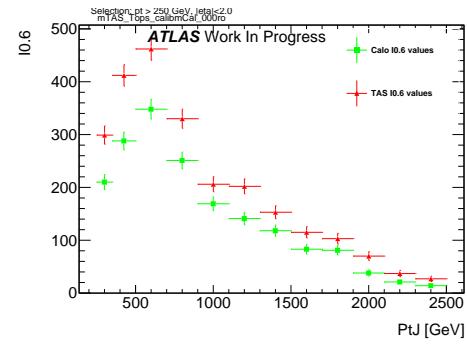


Figure 167: Left integral, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

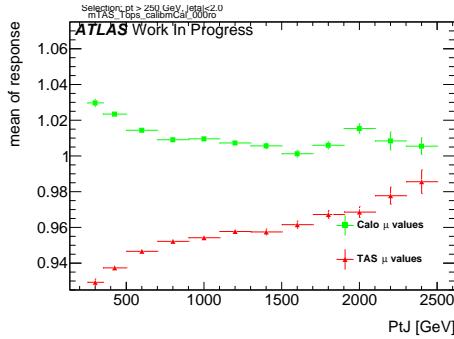


Figure 165: μ from fit of the mass Response vs bin of p_T^J

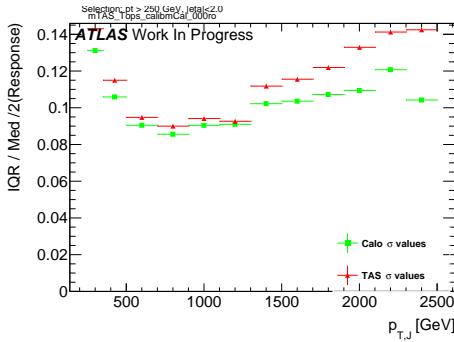


Figure 166: σ from fit of the mass Response vs bin of p_T^J

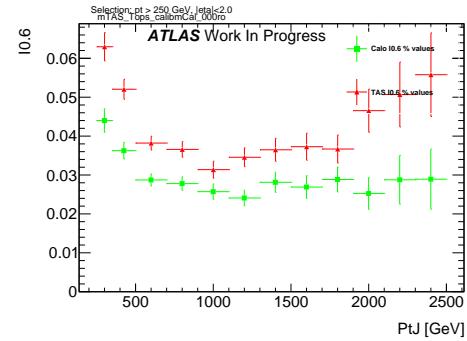
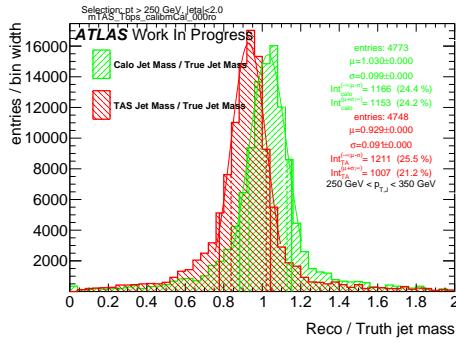
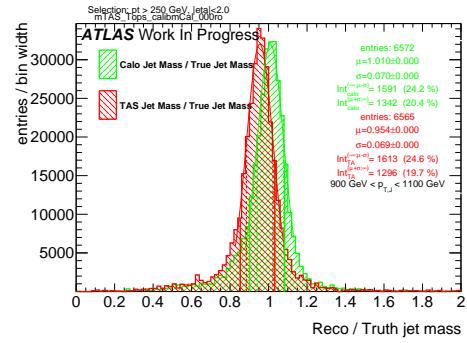
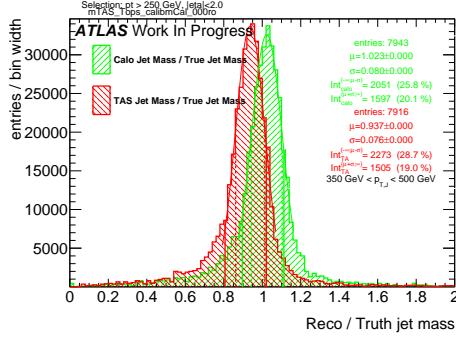
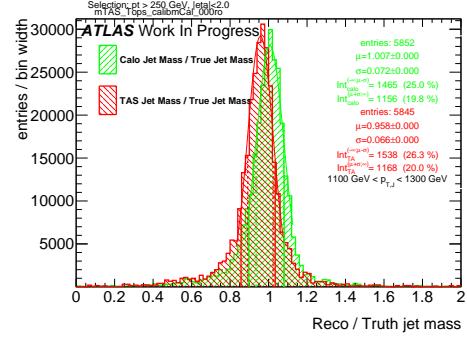
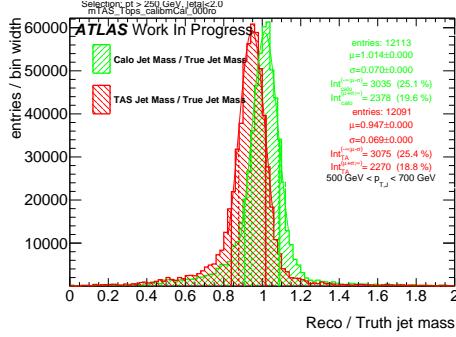
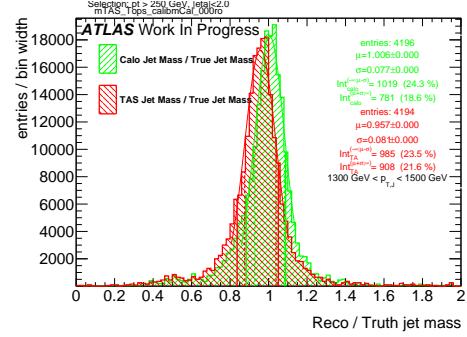
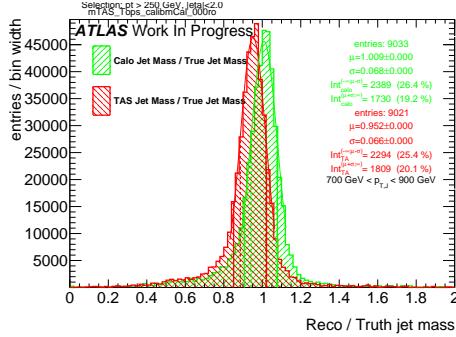
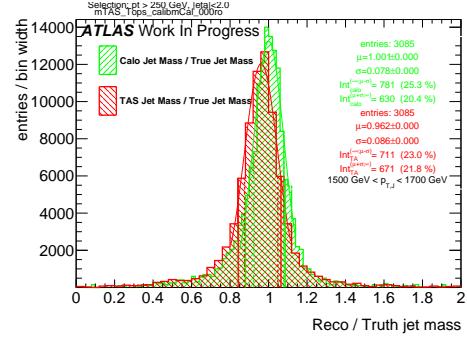
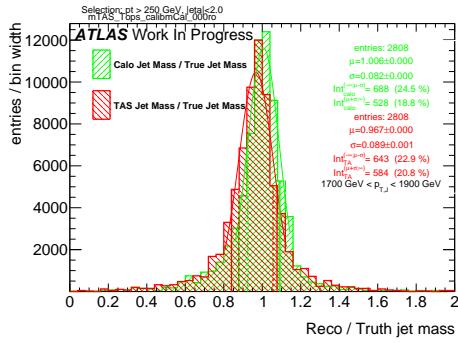
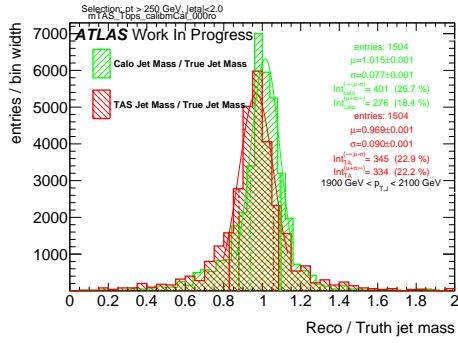
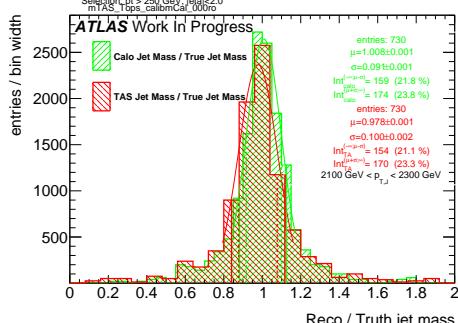
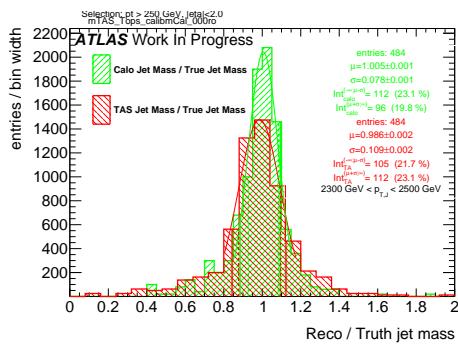


Figure 168: Left integral normalized, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

Figure 169: Response in bin of p_T^J (indicated on plot)Figure 173: Response in bin of p_T^J (indicated on plot)Figure 170: Response in bin of p_T^J (indicated on plot)Figure 174: Response in bin of p_T^J (indicated on plot)Figure 171: Response in bin of p_T^J (indicated on plot)Figure 175: Response in bin of p_T^J (indicated on plot)Figure 172: Response in bin of p_T^J (indicated on plot)Figure 176: Response in bin of p_T^J (indicated on plot)

Figure 177: Response in bin of p_T^J (indicated on plot)Figure 178: Response in bin of p_T^J (indicated on plot)Figure 179: Response in bin of p_T^J (indicated on plot)Figure 180: Response in bin of p_T^J (indicated on plot)

907 **C m^{TAS} distributions, boosted higgs**

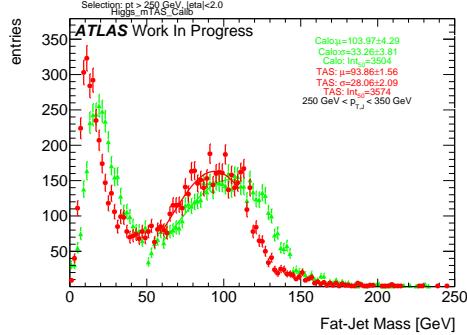


Figure 181: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

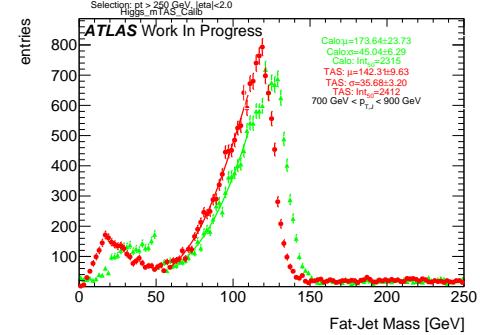


Figure 184: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

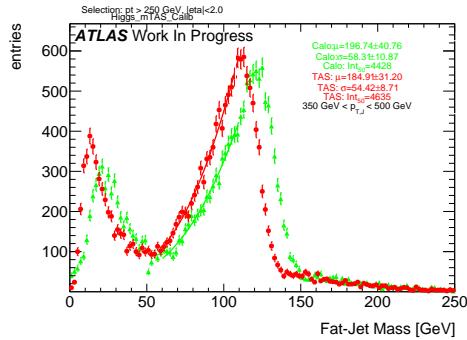


Figure 182: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

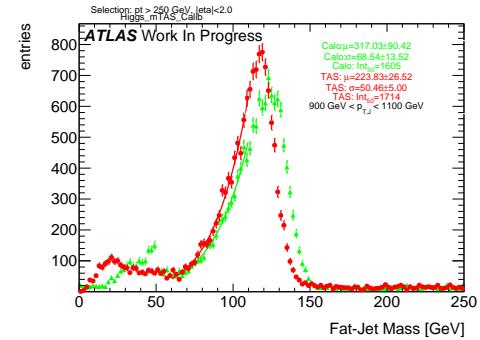


Figure 185: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

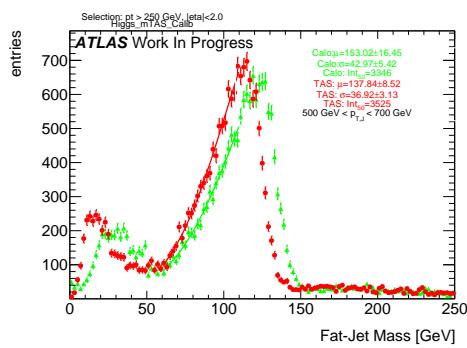


Figure 183: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

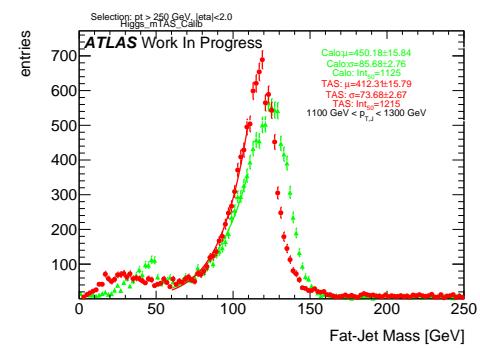


Figure 186: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

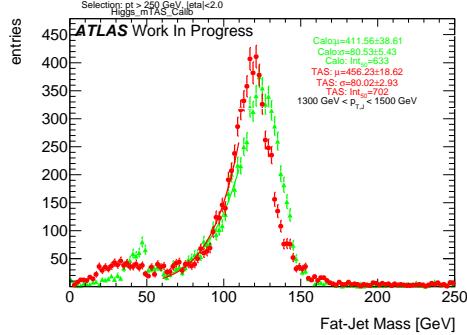


Figure 187: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

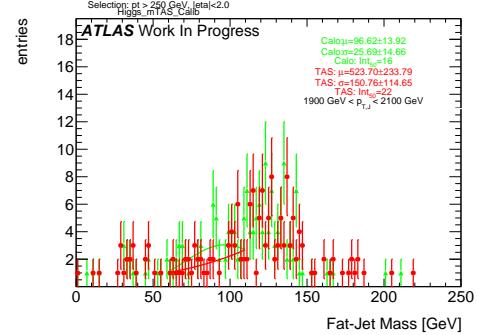


Figure 190: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

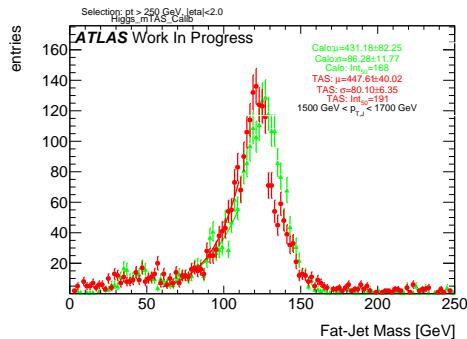


Figure 188: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

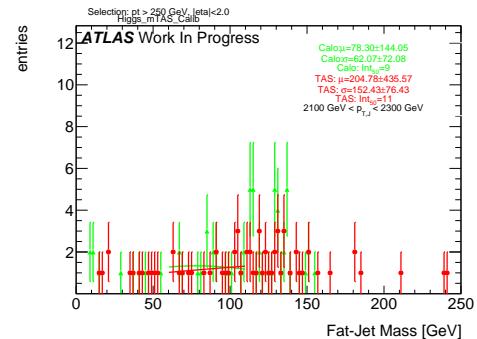


Figure 191: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

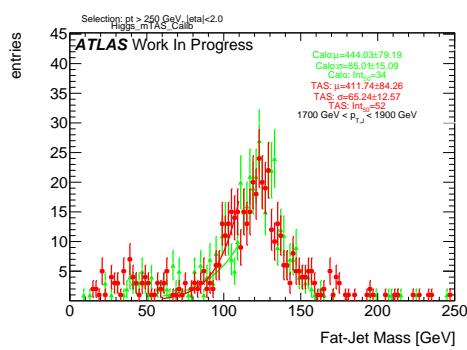


Figure 189: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

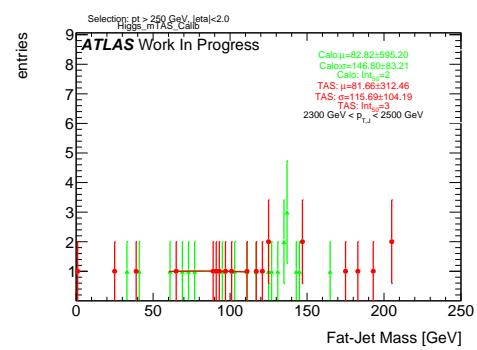


Figure 192: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

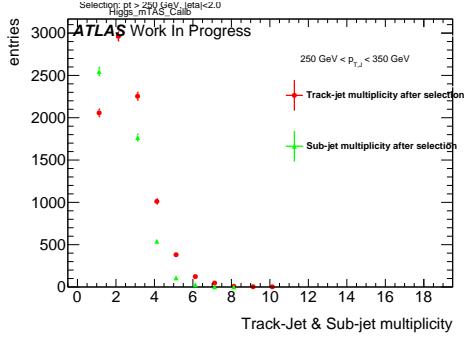


Figure 193: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

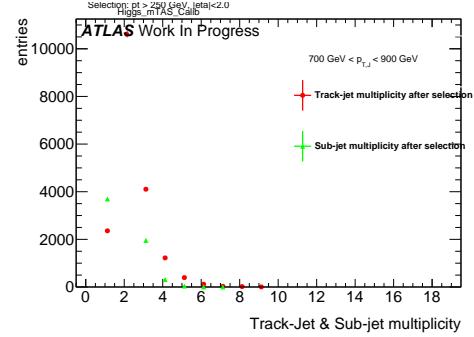


Figure 196: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

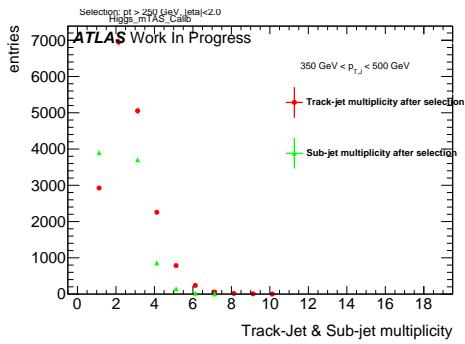


Figure 194: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

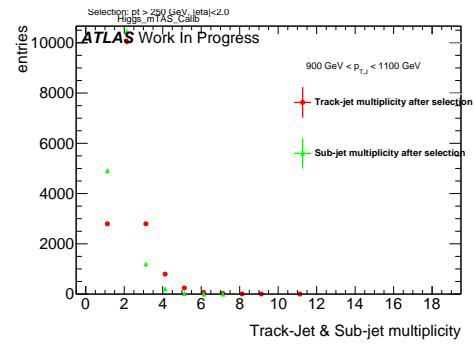


Figure 197: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

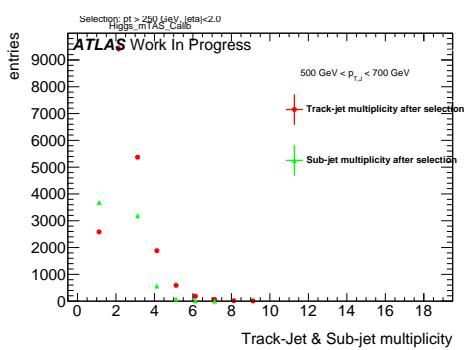


Figure 195: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

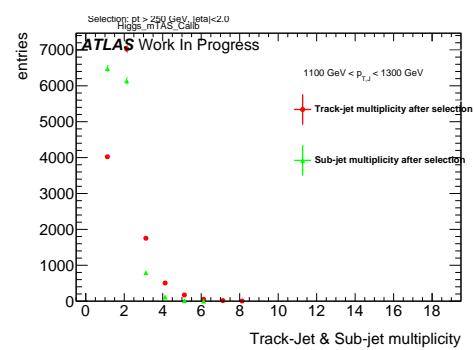


Figure 198: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

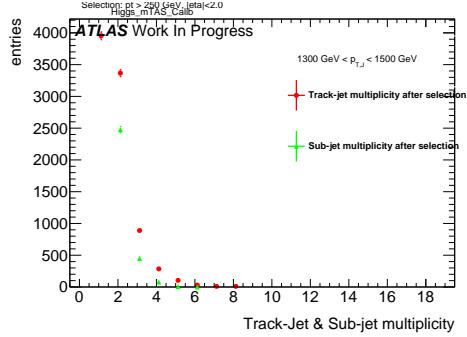


Figure 199: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

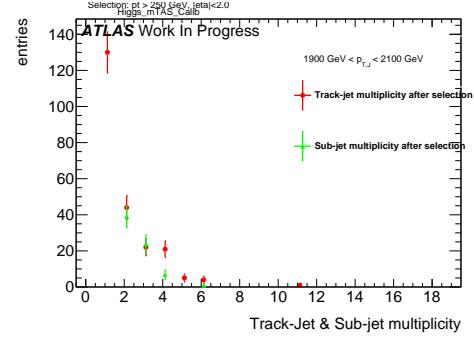


Figure 202: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

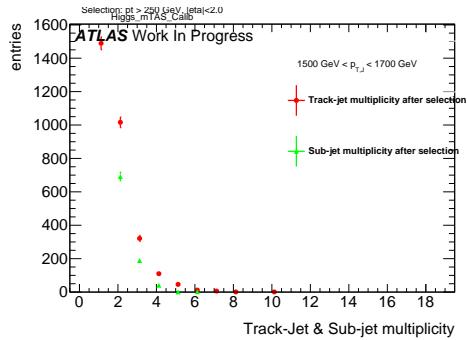


Figure 200: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

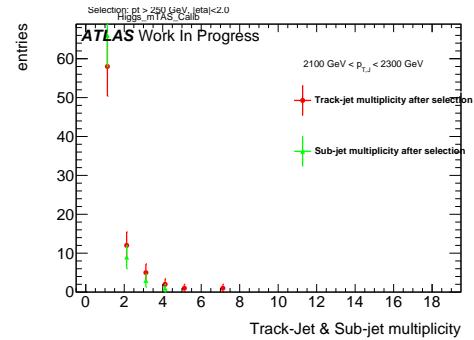


Figure 203: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

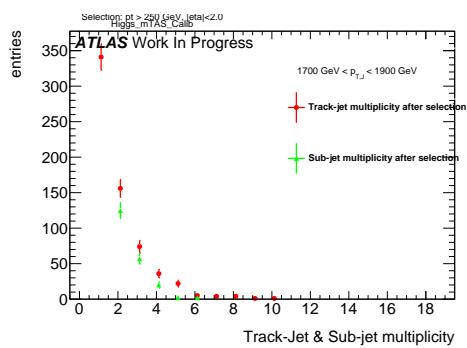


Figure 201: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

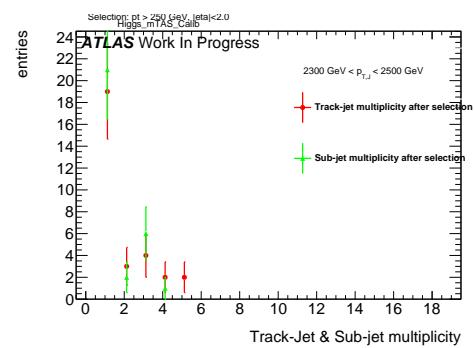


Figure 204: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

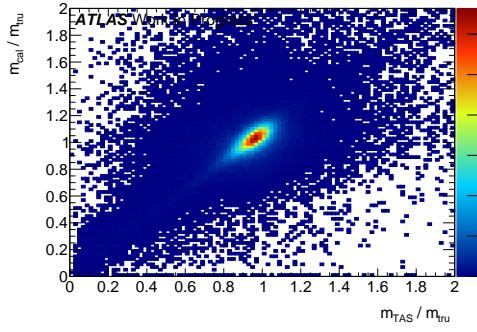
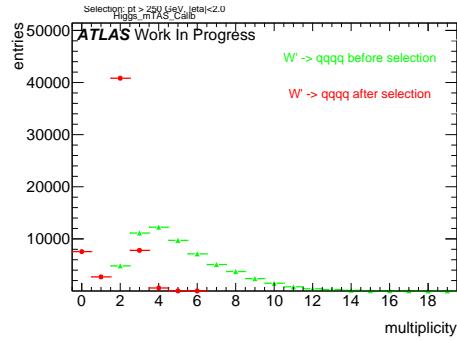
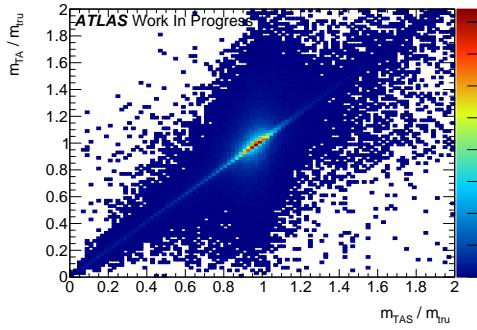
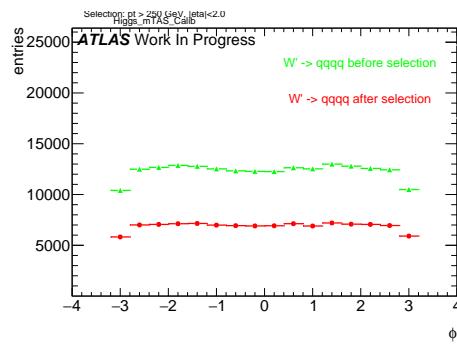
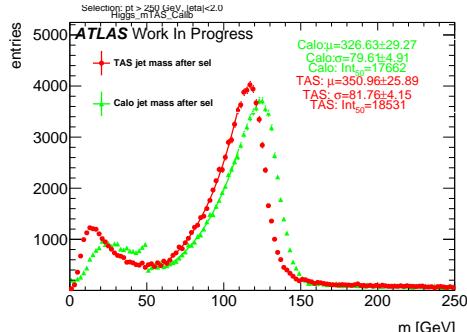
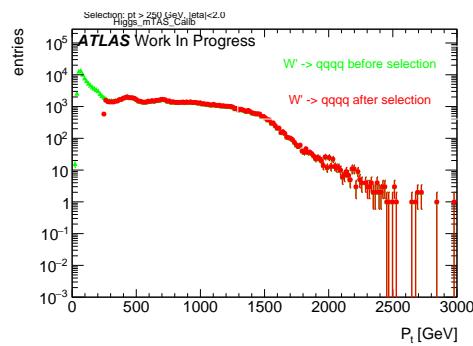
Figure 205: Scatter plot m^{TAS} versus m^{calo} responses

Figure 209: large-R jet Multiplicity, before and after selection

Figure 206: Scatter plot m^{TAS} versus m^{TA} responsesFigure 210: ϕ distribution of the large-R jet, before and after selectionFigure 207: m^{TAS} distribution in all the p_T binsFigure 211: p_T distribution of the large-R jet, before and after selectionFigure 208: η distribution of the large-R jet, before and after selection

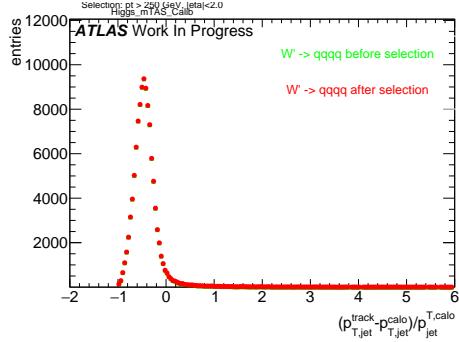


Figure 212: p_T resolution: $\frac{p_{T,jet}^{track} - p_{T,jet}^{fat}}{p_{T,jet}^{fat}}$, before and after selection

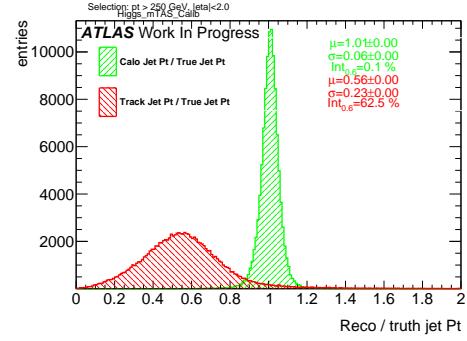


Figure 215: Transverse momentum response p_T^{Reco} / p_T^{Truth} for calorimeter and tracks

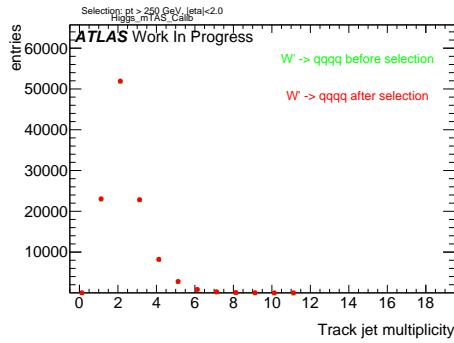


Figure 213: Multiplicity of track-jets R=0.2 per large-R jet

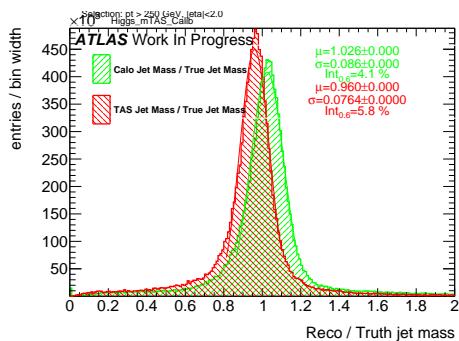


Figure 214: Response m^{Reco} / m^{Truth} for all the p_T bins

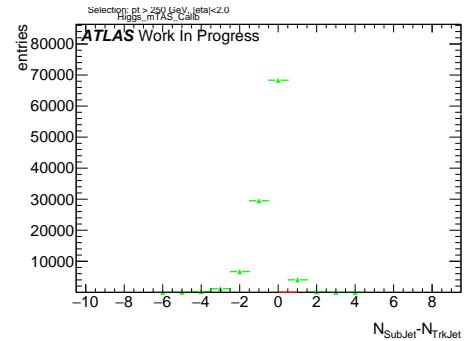


Figure 216: sub-jet - track-jet Multiplicity

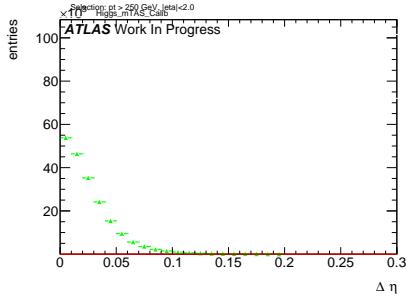


Figure 217: $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

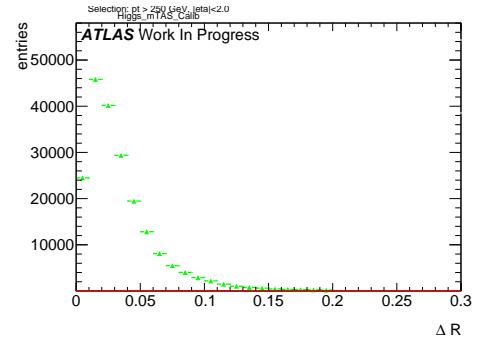


Figure 221: $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

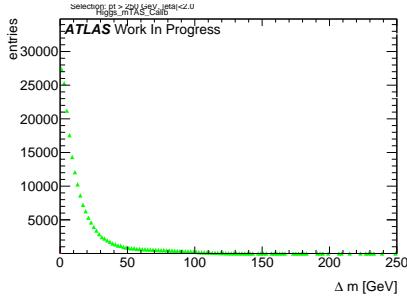


Figure 218: $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

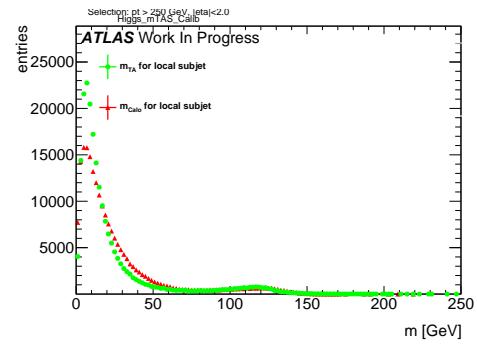


Figure 222: Mass distribution of the sub-jet, calorimeter and track-assisted

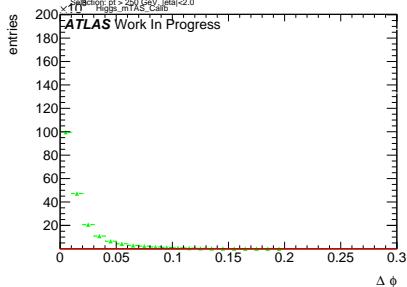


Figure 219: $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

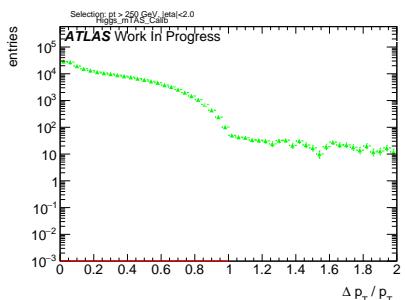


Figure 220: $|p_{\text{T,sub-jet}} - p_{\text{T,track-jet}}|$ distribution, where sub-jet and track-jet are the closest

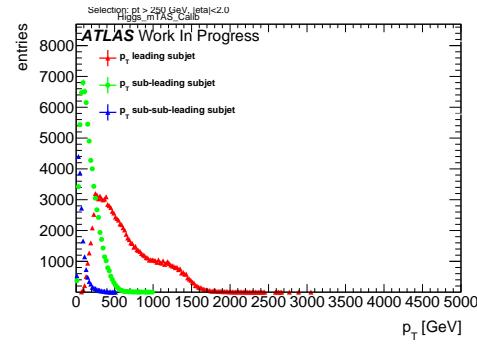


Figure 223: p_{T} distribution for leading, sub-leading and sub-sub-leading sub-jets

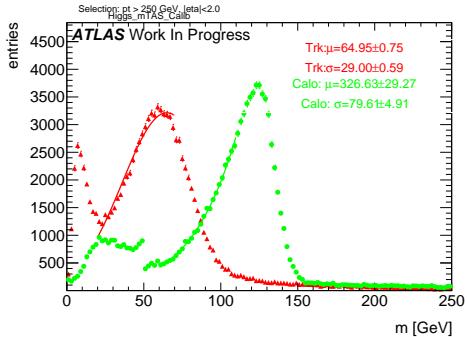


Figure 224: Mass distribution for calorimeter and tracks associated to the large-R jet

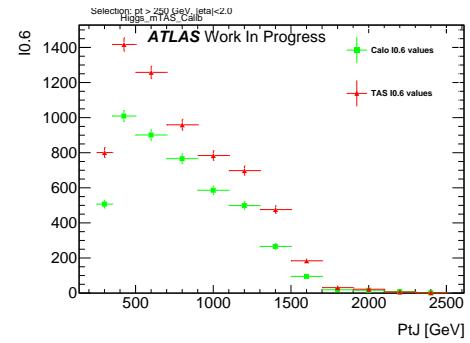


Figure 227: Left integral, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

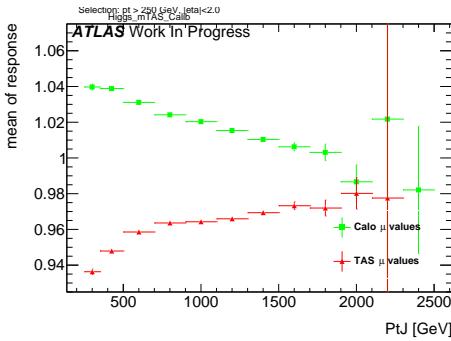


Figure 225: μ from fit of the mass Response vs bin of p_T^J

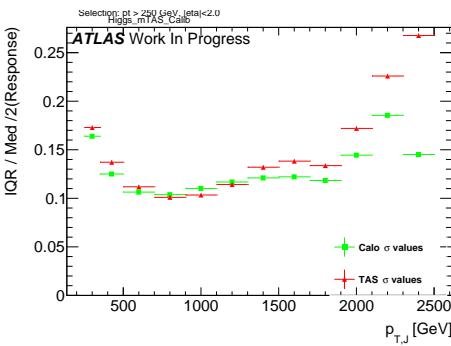


Figure 226: σ from fit of the mass Response vs bin of p_T^J

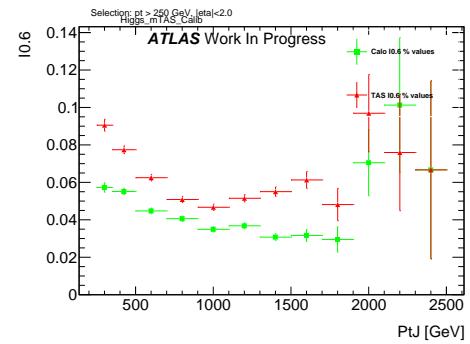
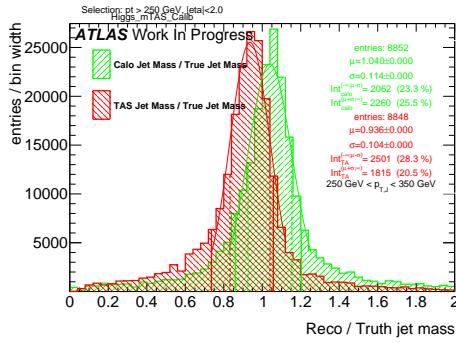
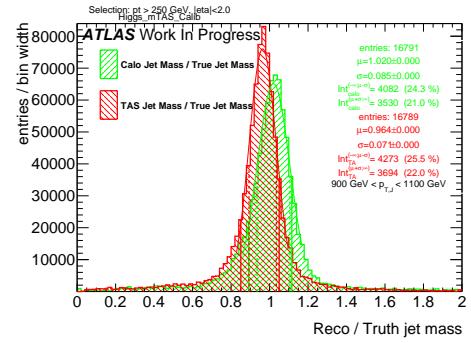
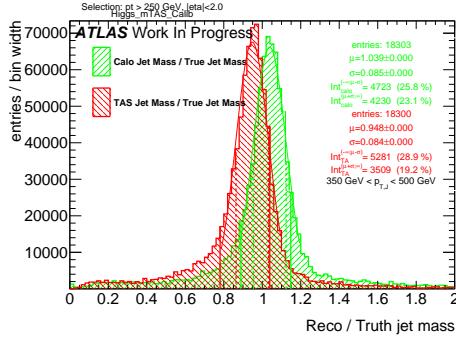
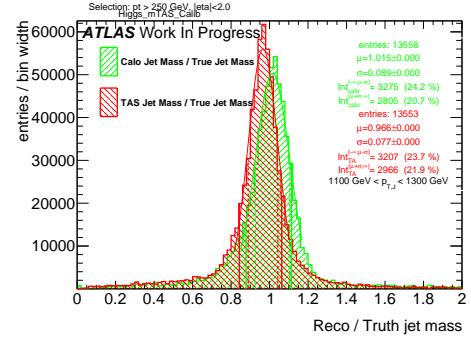
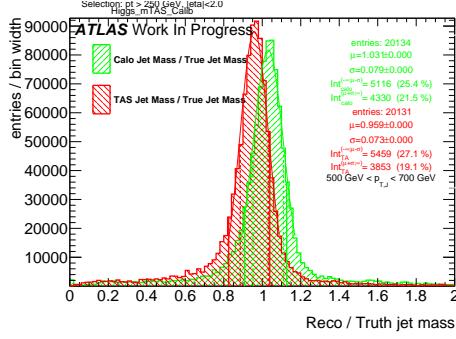
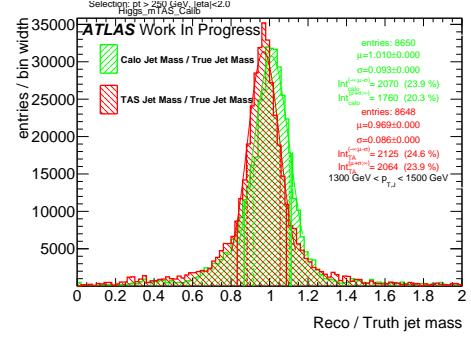
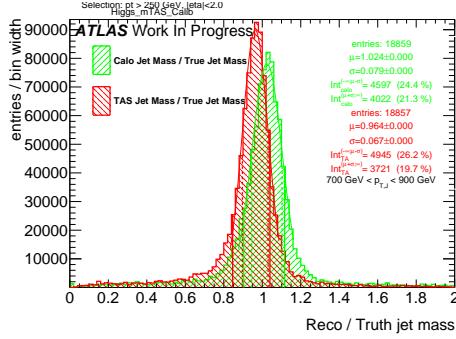
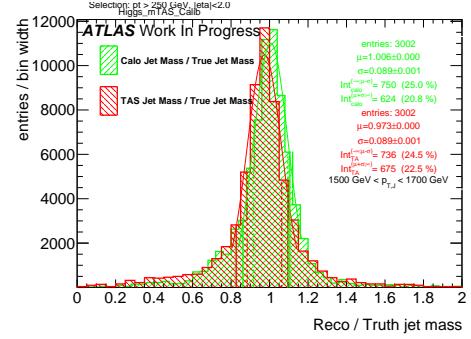


Figure 228: Left integral normalized, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

Figure 229: Response in bin of p_T^J (indicated on plot)Figure 233: Response in bin of p_T^J (indicated on plot)Figure 230: Response in bin of p_T^J (indicated on plot)Figure 234: Response in bin of p_T^J (indicated on plot)Figure 231: Response in bin of p_T^J (indicated on plot)Figure 235: Response in bin of p_T^J (indicated on plot)Figure 232: Response in bin of p_T^J (indicated on plot)Figure 236: Response in bin of p_T^J (indicated on plot)

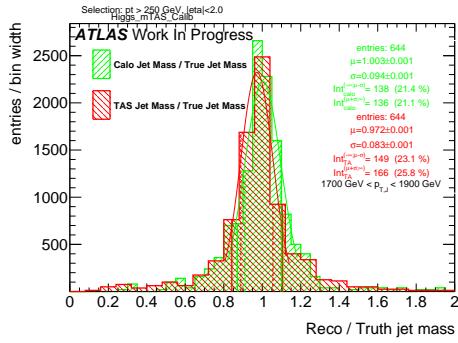


Figure 237: Response in bin of p_T^J (indicated on plot)

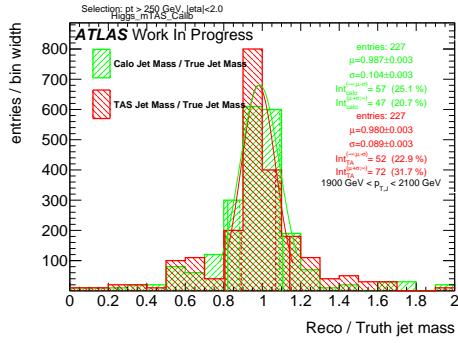


Figure 238: Response in bin of p_T^J (indicated on plot)

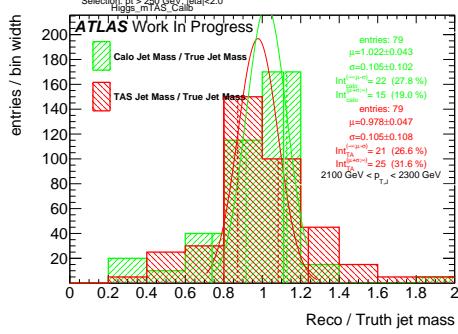


Figure 239: Response in bin of p_T^J (indicated on plot)

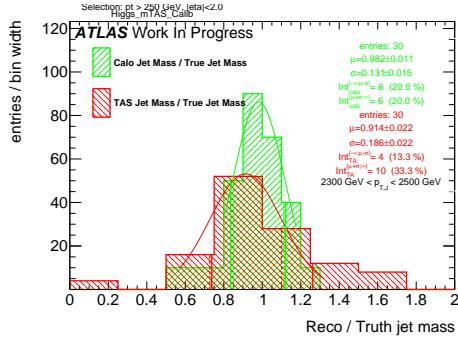
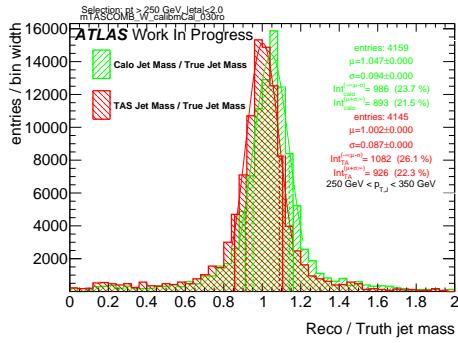
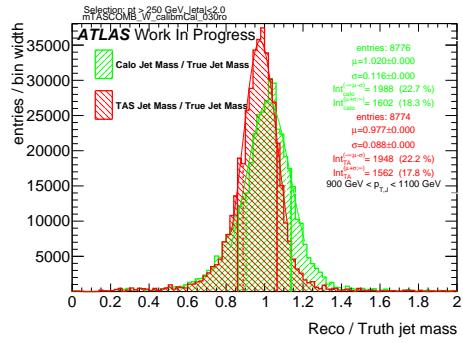
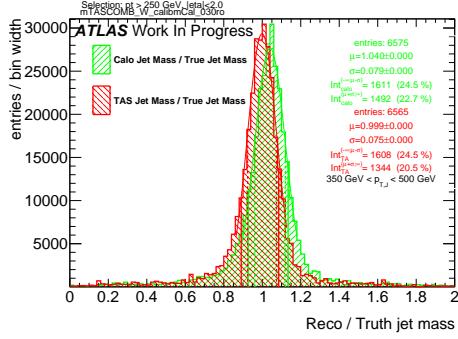
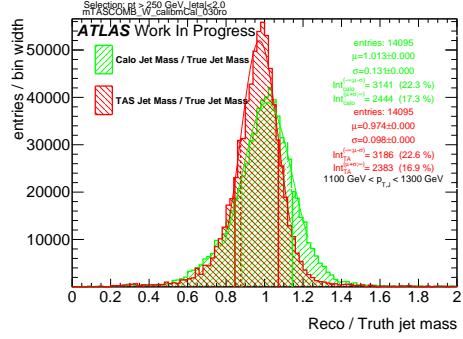
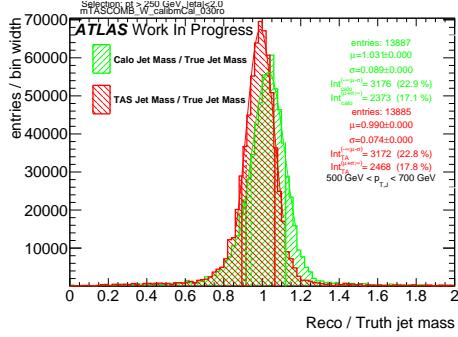
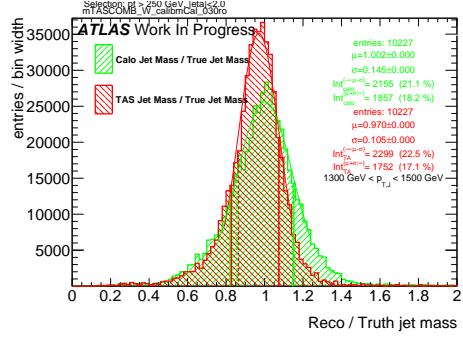
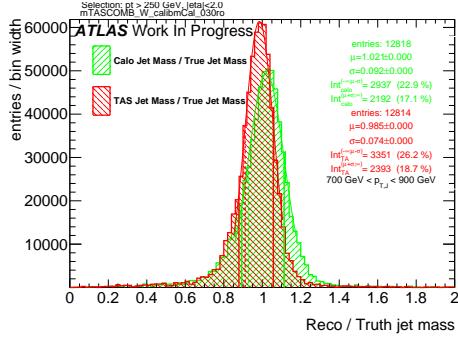
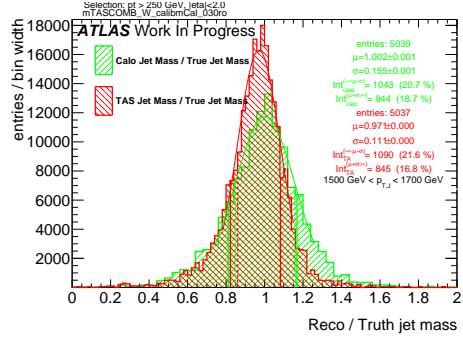
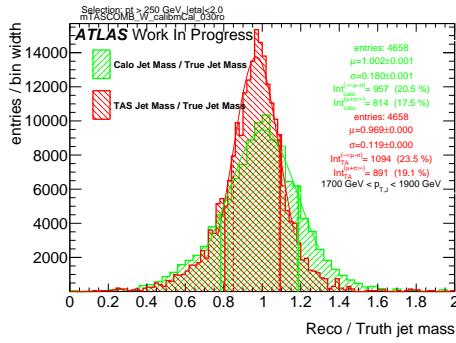
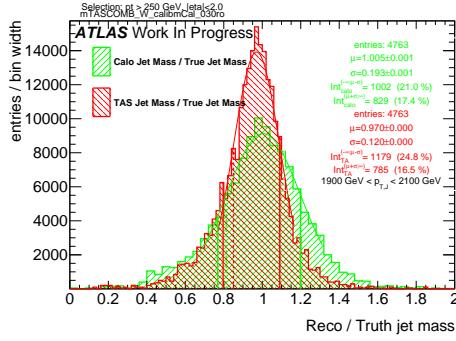
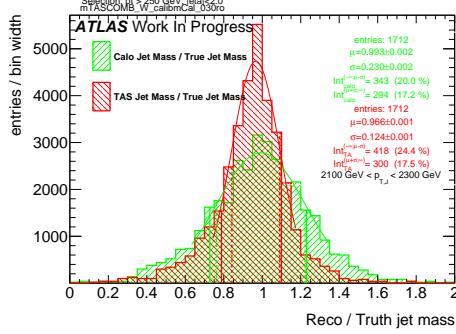
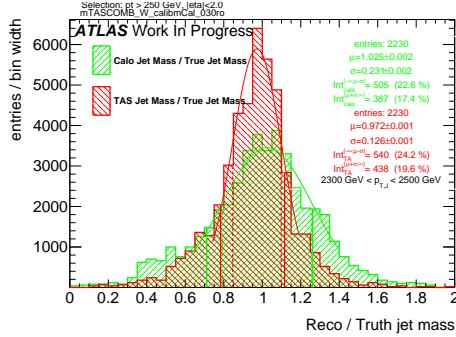


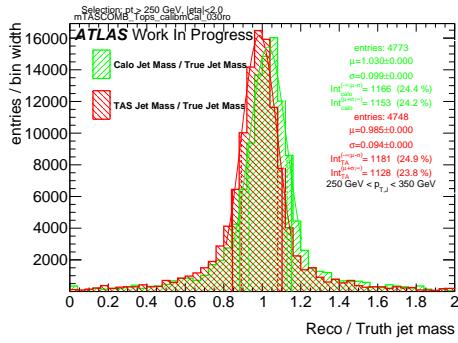
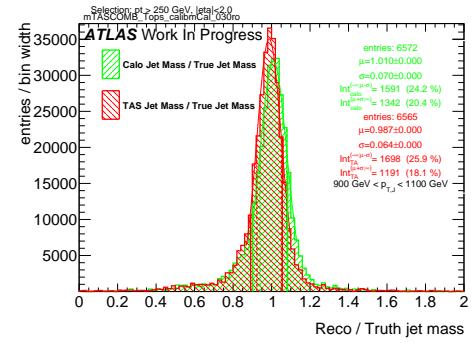
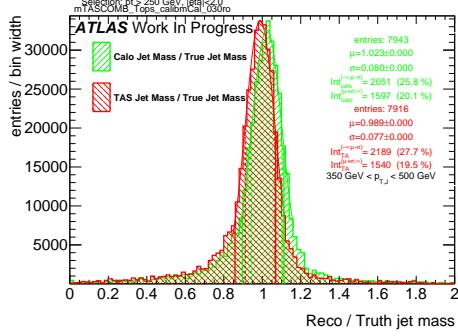
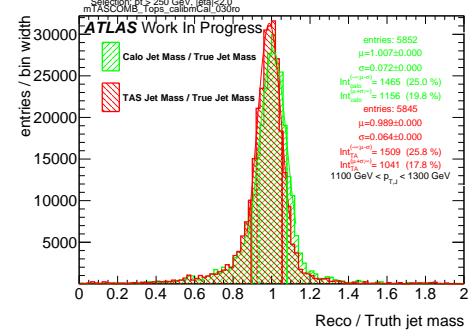
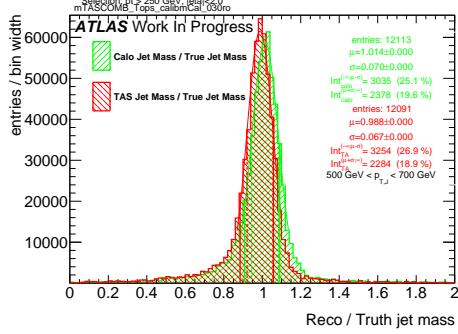
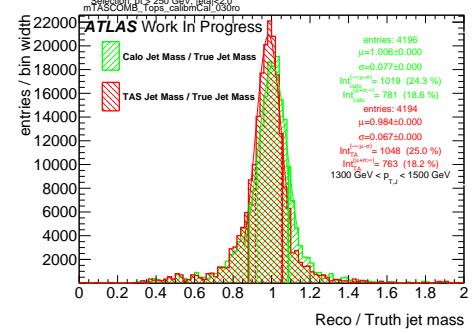
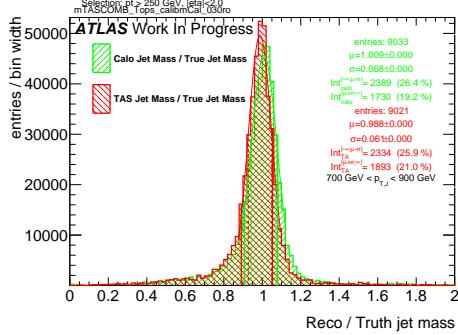
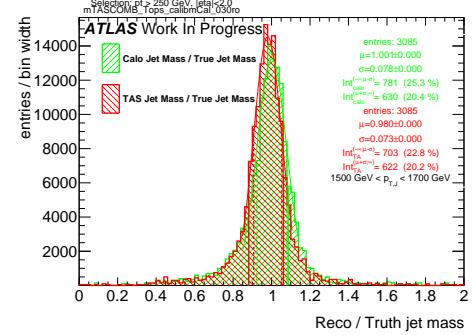
Figure 240: Response in bin of p_T^J (indicated on plot)

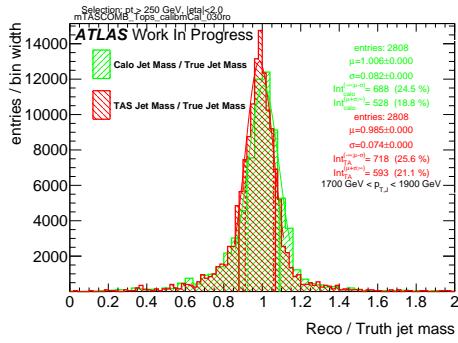
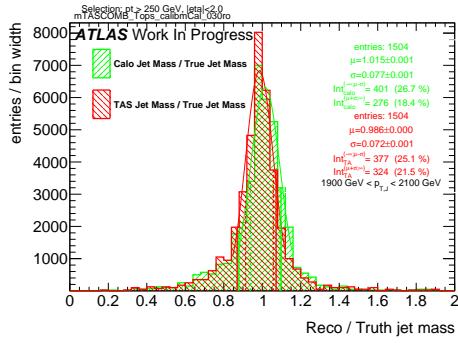
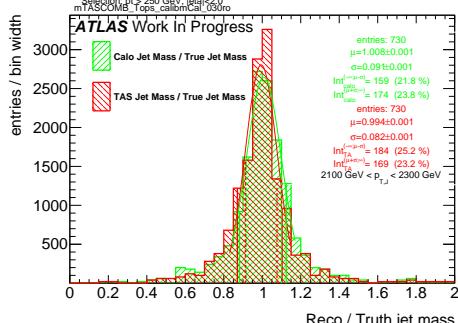
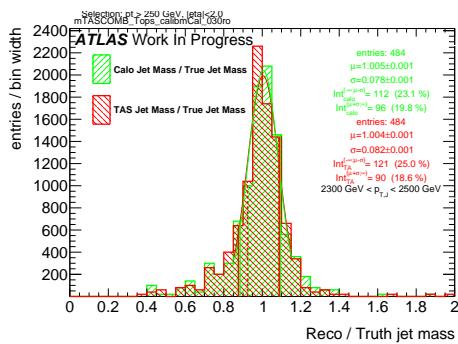
908 **D m_{TAS}^{comb} response distributions, boosted W/Z**

Figure 241: Response in bin of p_T^J (indicated on plot)Figure 245: Response in bin of p_T^J (indicated on plot)Figure 242: Response in bin of p_T^J (indicated on plot)Figure 246: Response in bin of p_T^J (indicated on plot)Figure 243: Response in bin of p_T^J (indicated on plot)Figure 247: Response in bin of p_T^J (indicated on plot)Figure 244: Response in bin of p_T^J (indicated on plot)Figure 248: Response in bin of p_T^J (indicated on plot)

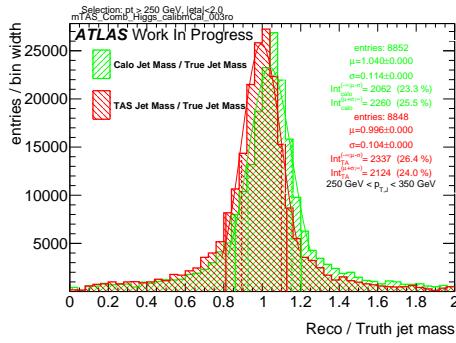
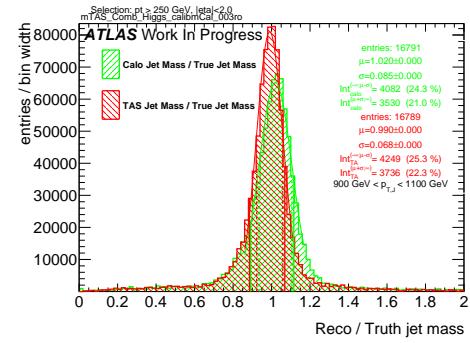
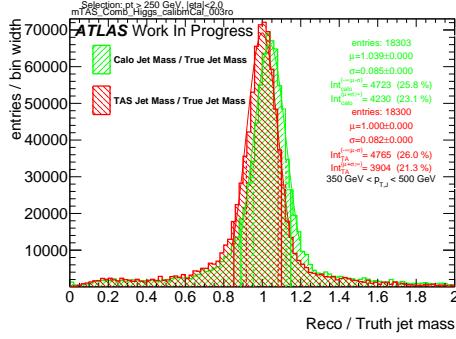
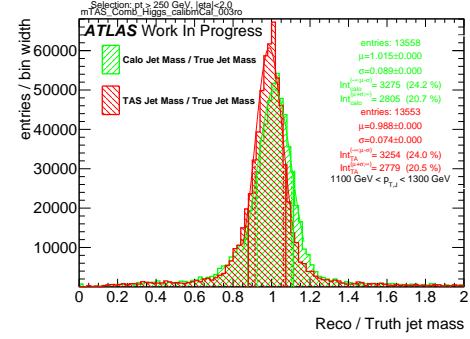
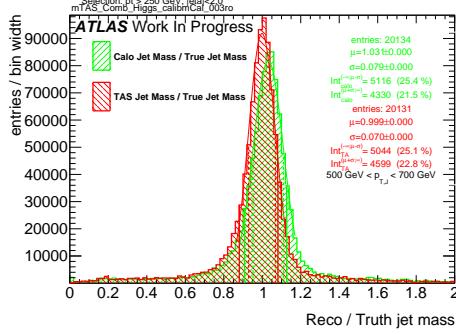
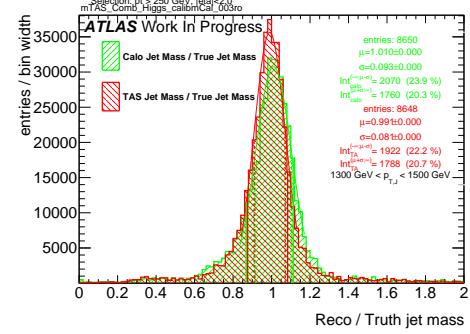
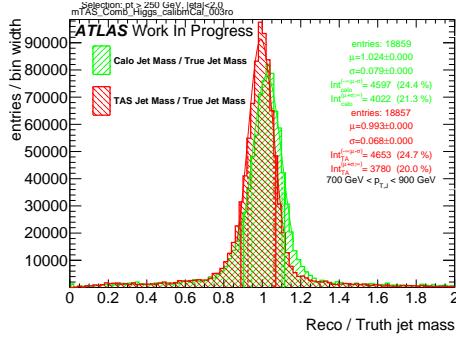
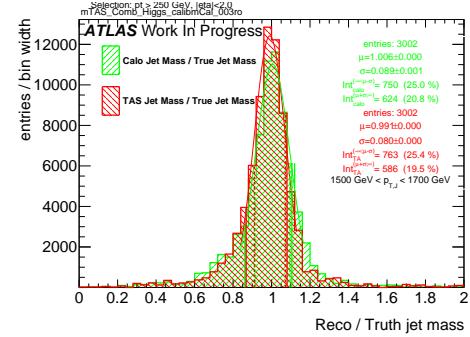
Figure 249: Response in bin of p_T^J (indicated on plot)Figure 250: Response in bin of p_T^J (indicated on plot)Figure 251: Response in bin of p_T^J (indicated on plot)Figure 252: Response in bin of p_T^J (indicated on plot)

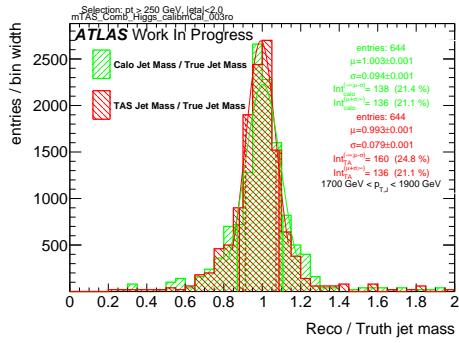
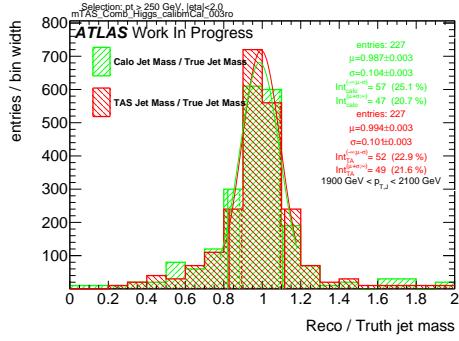
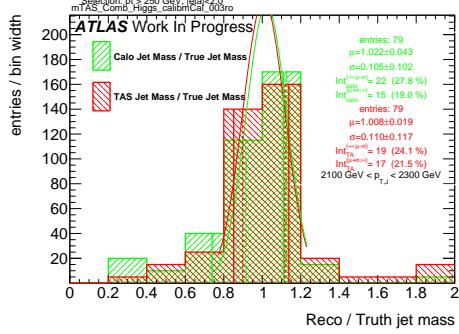
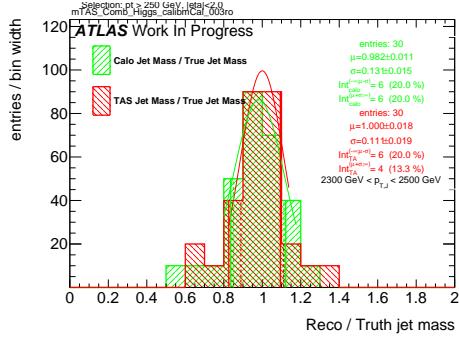
⁹⁰⁹ E m_{TAS}^{comb} response distributions, boosted tops

Figure 253: Response in bin of p_T^J (indicated on plot)Figure 257: Response in bin of p_T^J (indicated on plot)Figure 254: Response in bin of p_T^J (indicated on plot)Figure 258: Response in bin of p_T^J (indicated on plot)Figure 255: Response in bin of p_T^J (indicated on plot)Figure 259: Response in bin of p_T^J (indicated on plot)Figure 256: Response in bin of p_T^J (indicated on plot)Figure 260: Response in bin of p_T^J (indicated on plot)

Figure 261: Response in bin of p_T^J (indicated on plot)Figure 262: Response in bin of p_T^J (indicated on plot)Figure 263: Response in bin of p_T^J (indicated on plot)Figure 264: Response in bin of p_T^J (indicated on plot)

₉₁₀ F m_{TAS}^{comb} response distributions, Higgs

Figure 265: Response in bin of p_T^J (indicated on plot)Figure 269: Response in bin of p_T^J (indicated on plot)Figure 266: Response in bin of p_T^J (indicated on plot)Figure 270: Response in bin of p_T^J (indicated on plot)Figure 267: Response in bin of p_T^J (indicated on plot)Figure 271: Response in bin of p_T^J (indicated on plot)Figure 268: Response in bin of p_T^J (indicated on plot)Figure 272: Response in bin of p_T^J (indicated on plot)

Figure 273: Response in bin of p_T^J (indicated on plot)Figure 274: Response in bin of p_T^J (indicated on plot)Figure 275: Response in bin of p_T^J (indicated on plot)Figure 276: Response in bin of p_T^J (indicated on plot)

911 ROCs for the Best Variables

912 **G W boson Tagging**

913 **H Higgs Boson Tagging**

914 **I Top Quark Tagging**

915 Signal and Background Distributions

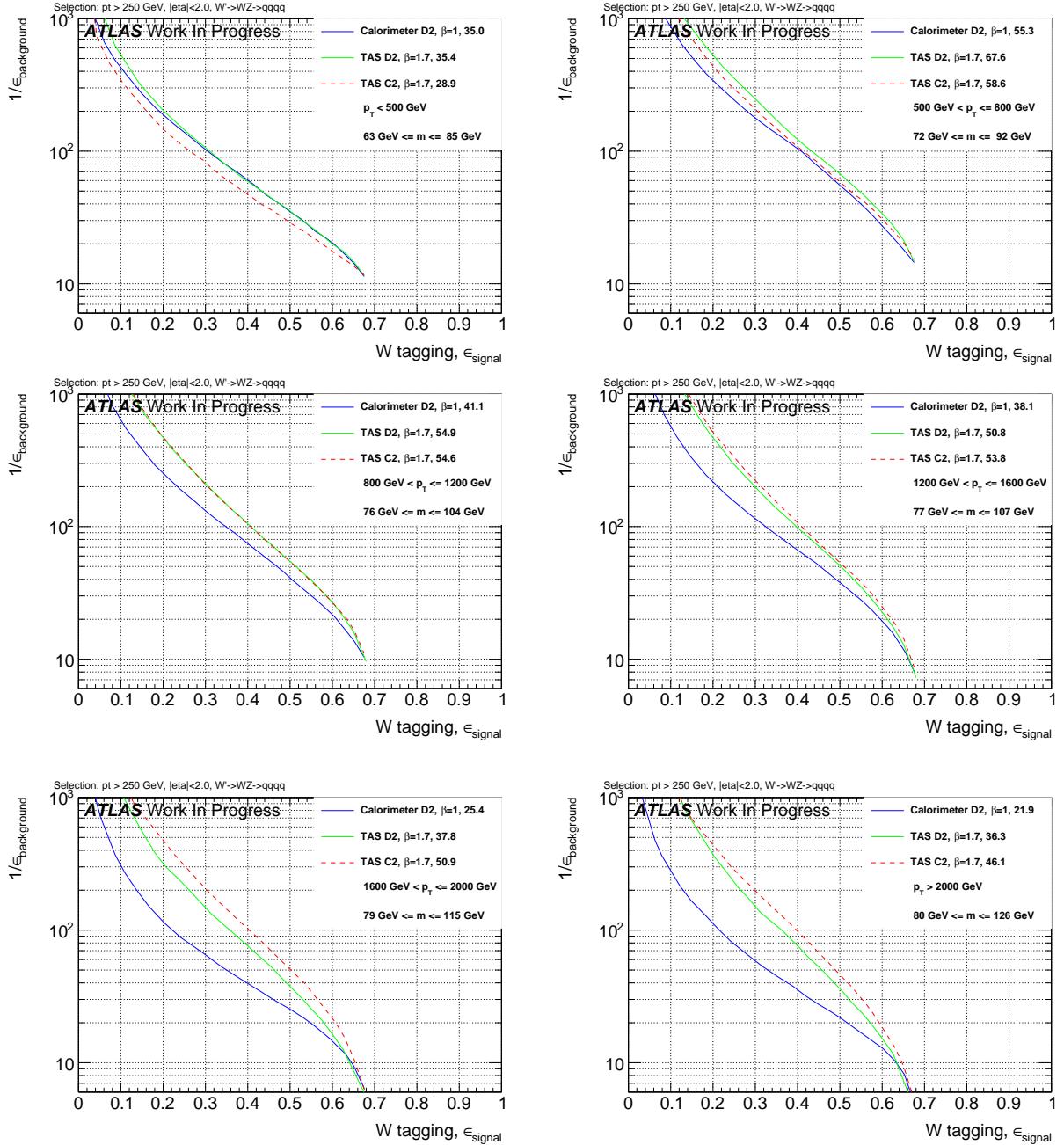


Figure 277: ROCs showing QCD rejection against W jet efficiency for $D2_{\text{TAS}}^{(\beta=1.7)}$ & $C2_{\text{TAS}}^{(\beta=1.7)}$ compared to $D2_{\text{calo}}^{(\beta=1)}$.

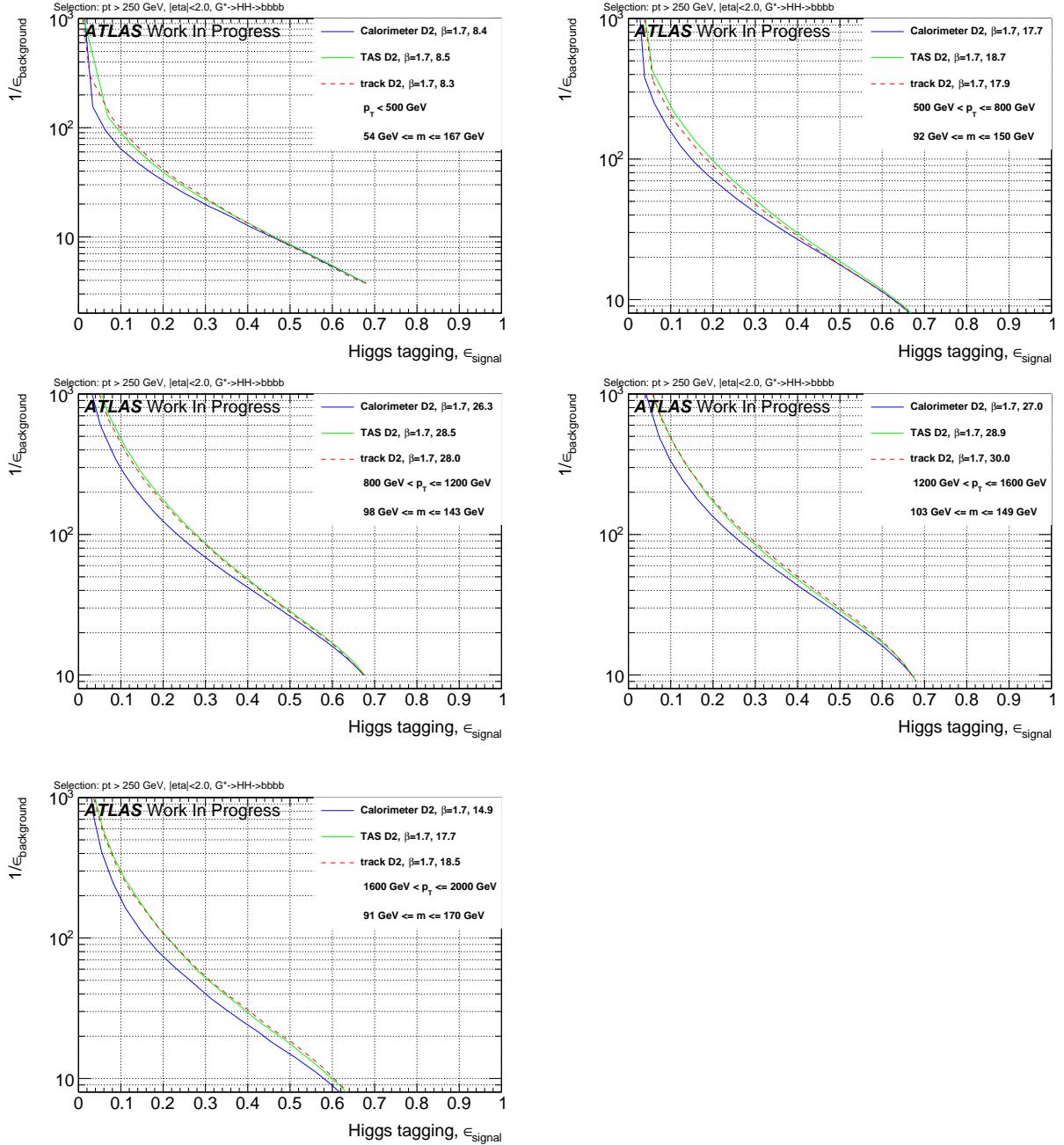


Figure 278: ROCs showing QCD rejection against Higgs jet efficiency for $D2_{\text{TAS}}^{(\beta=1.7)}$ & $D2_{\text{track}}^{(\beta=1.7)}$ compared to $D2_{\text{calo}}^{(\beta=1)}$.

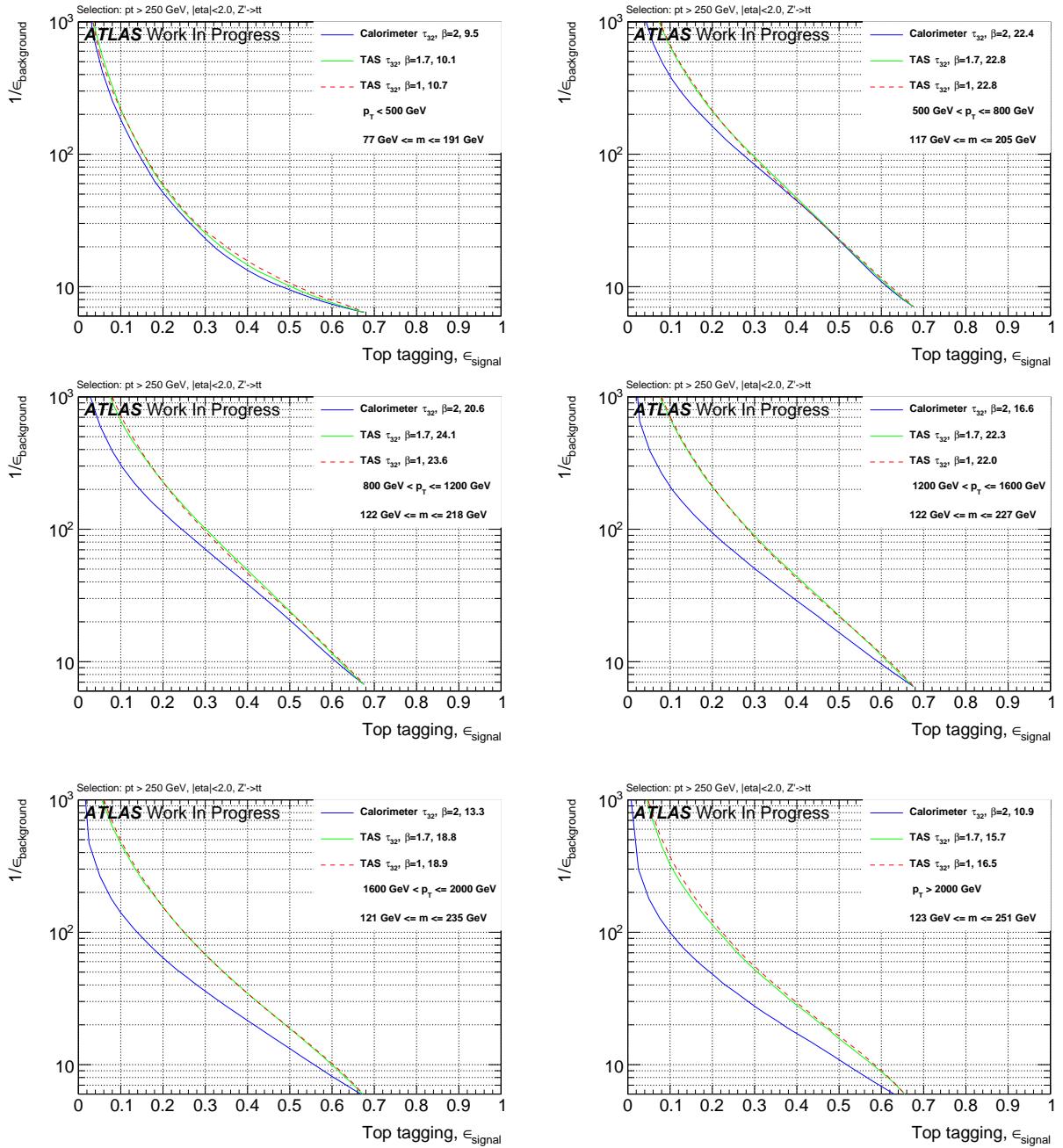
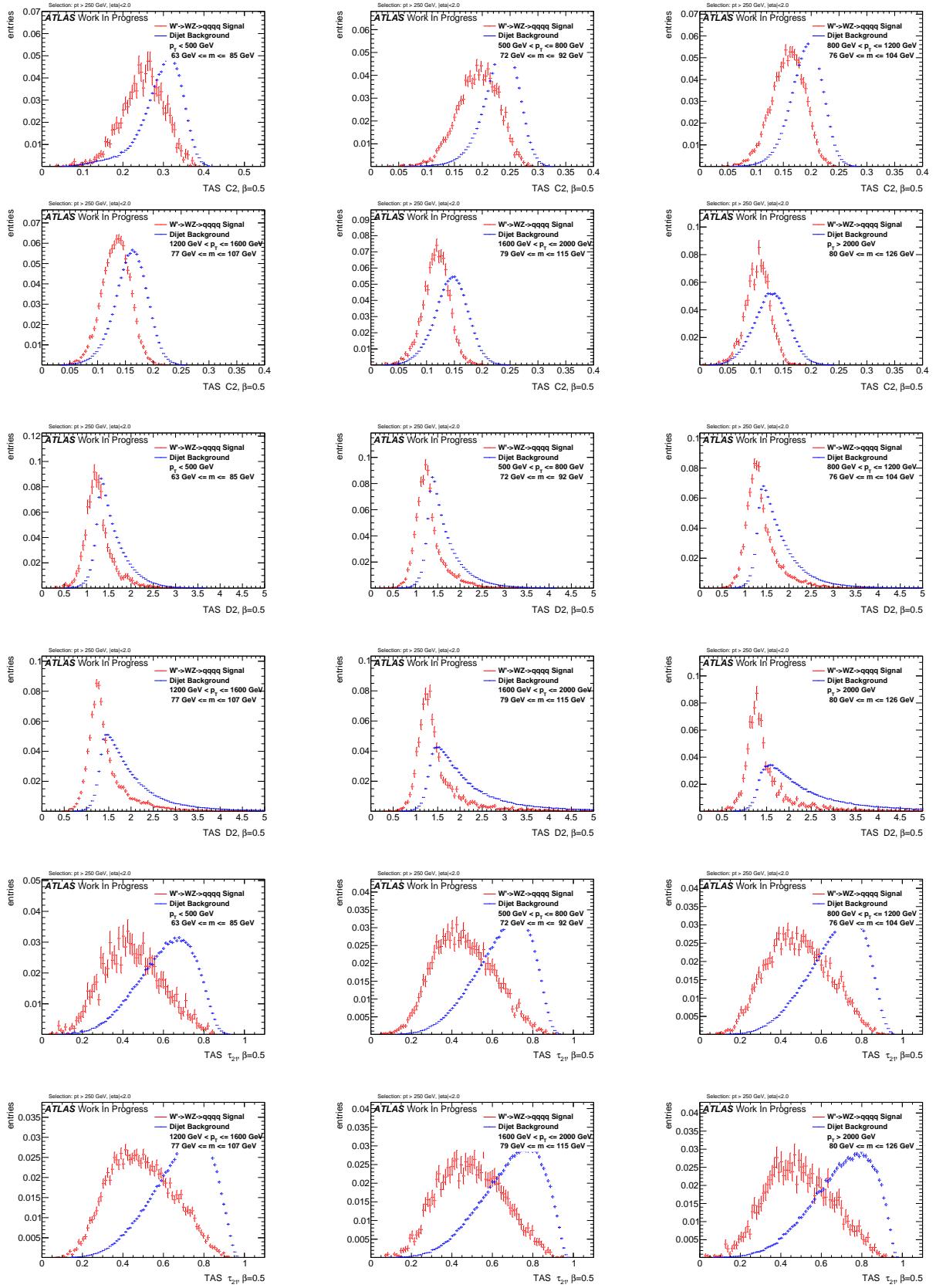
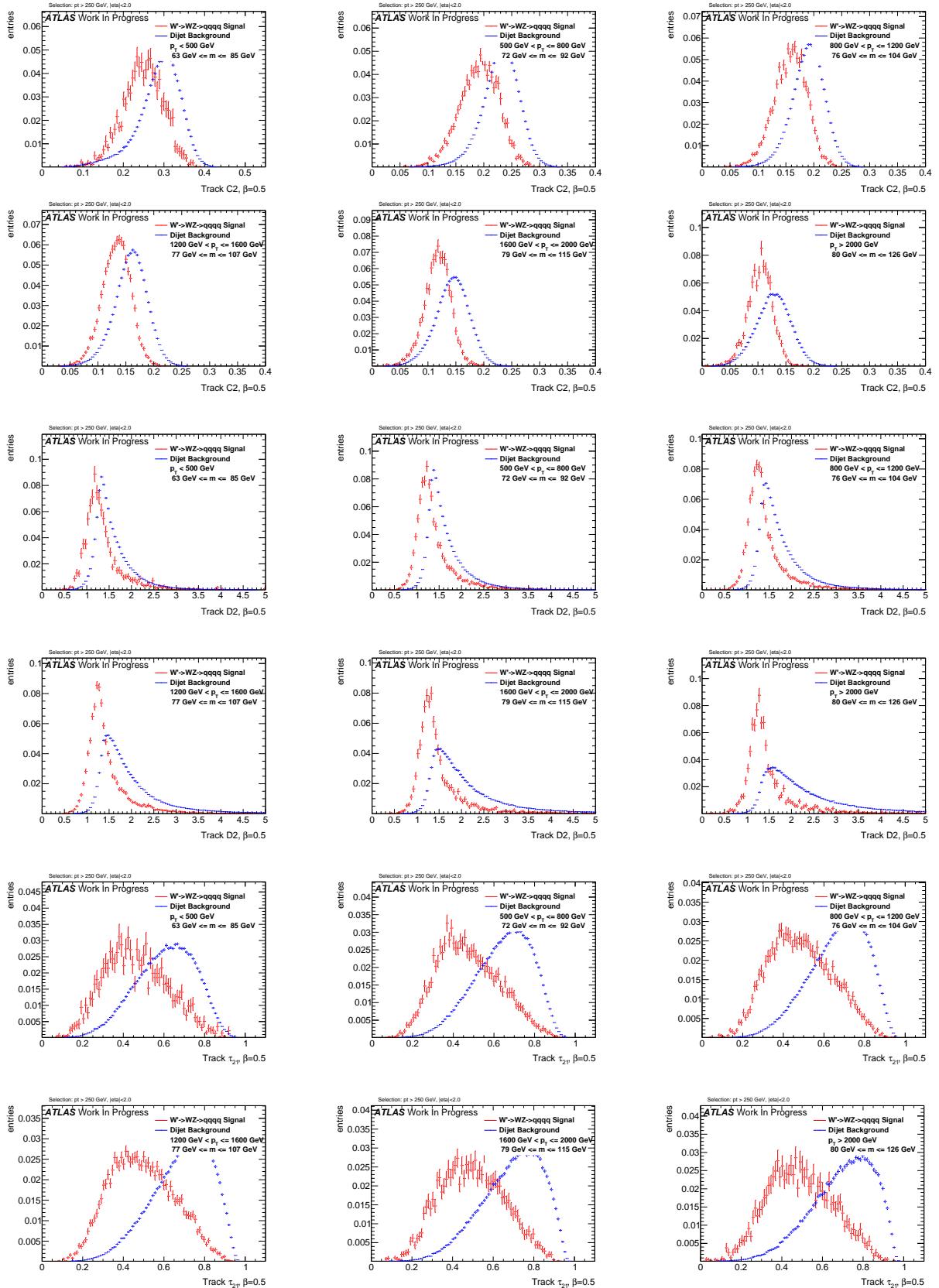
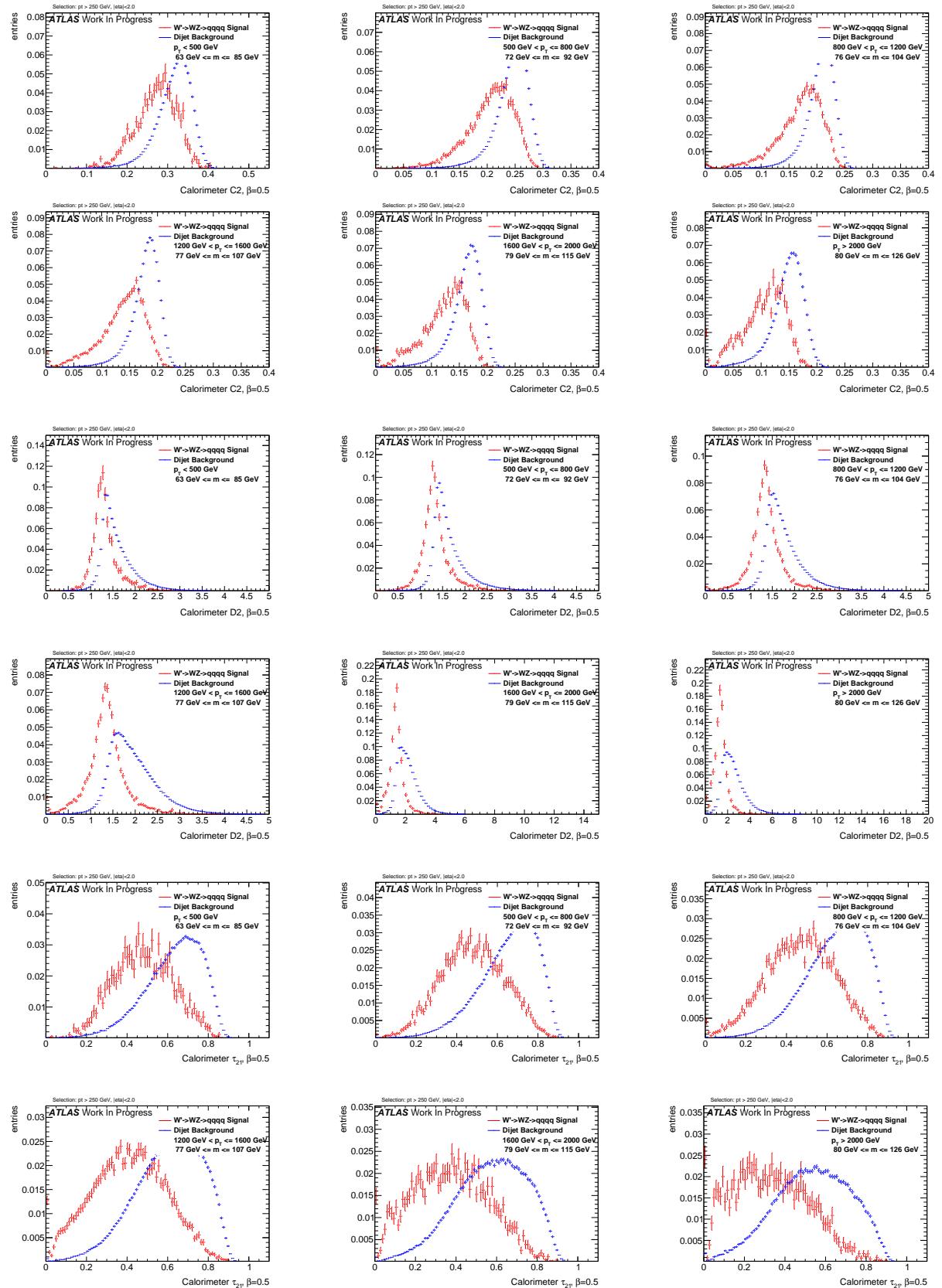
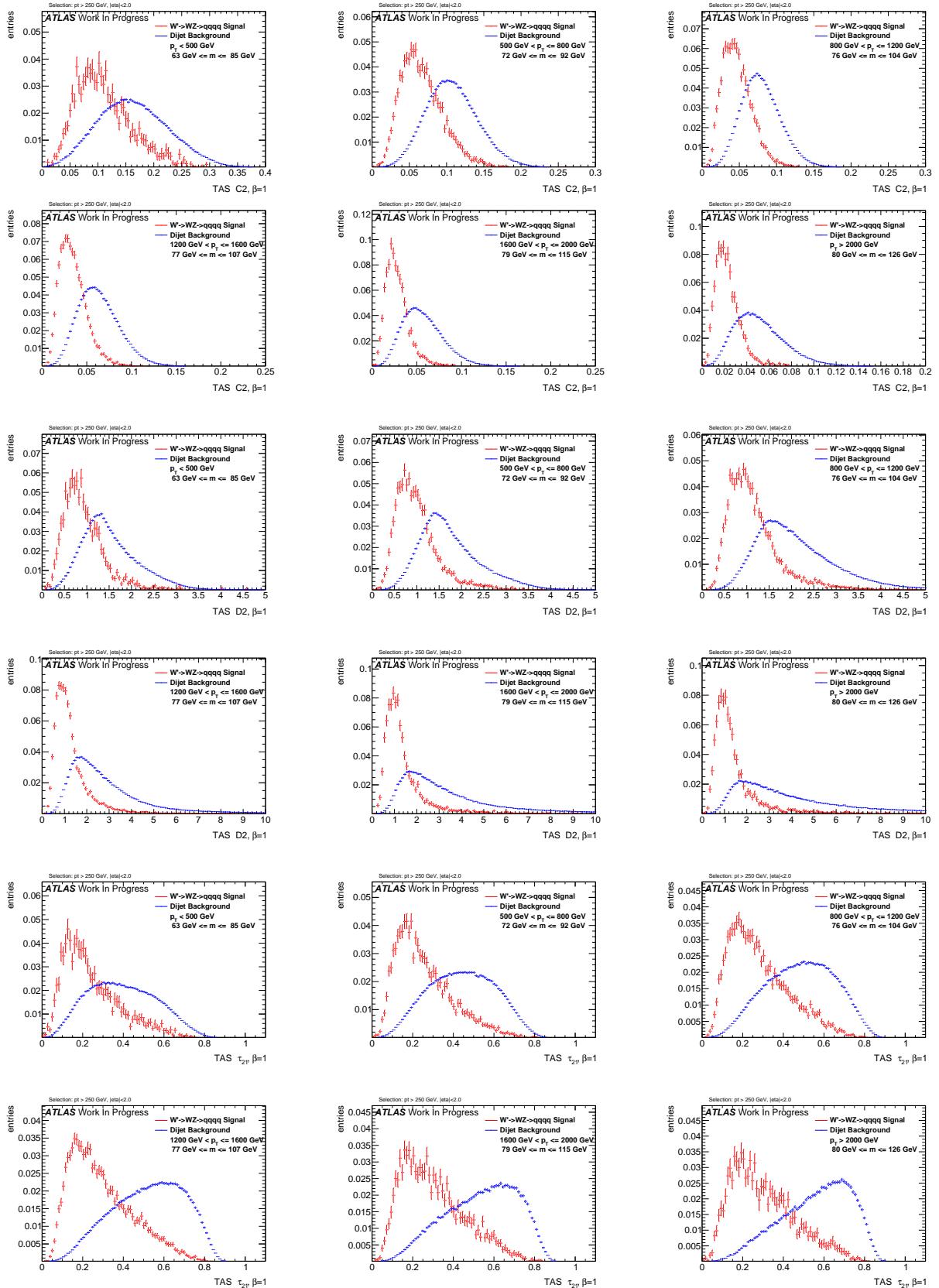


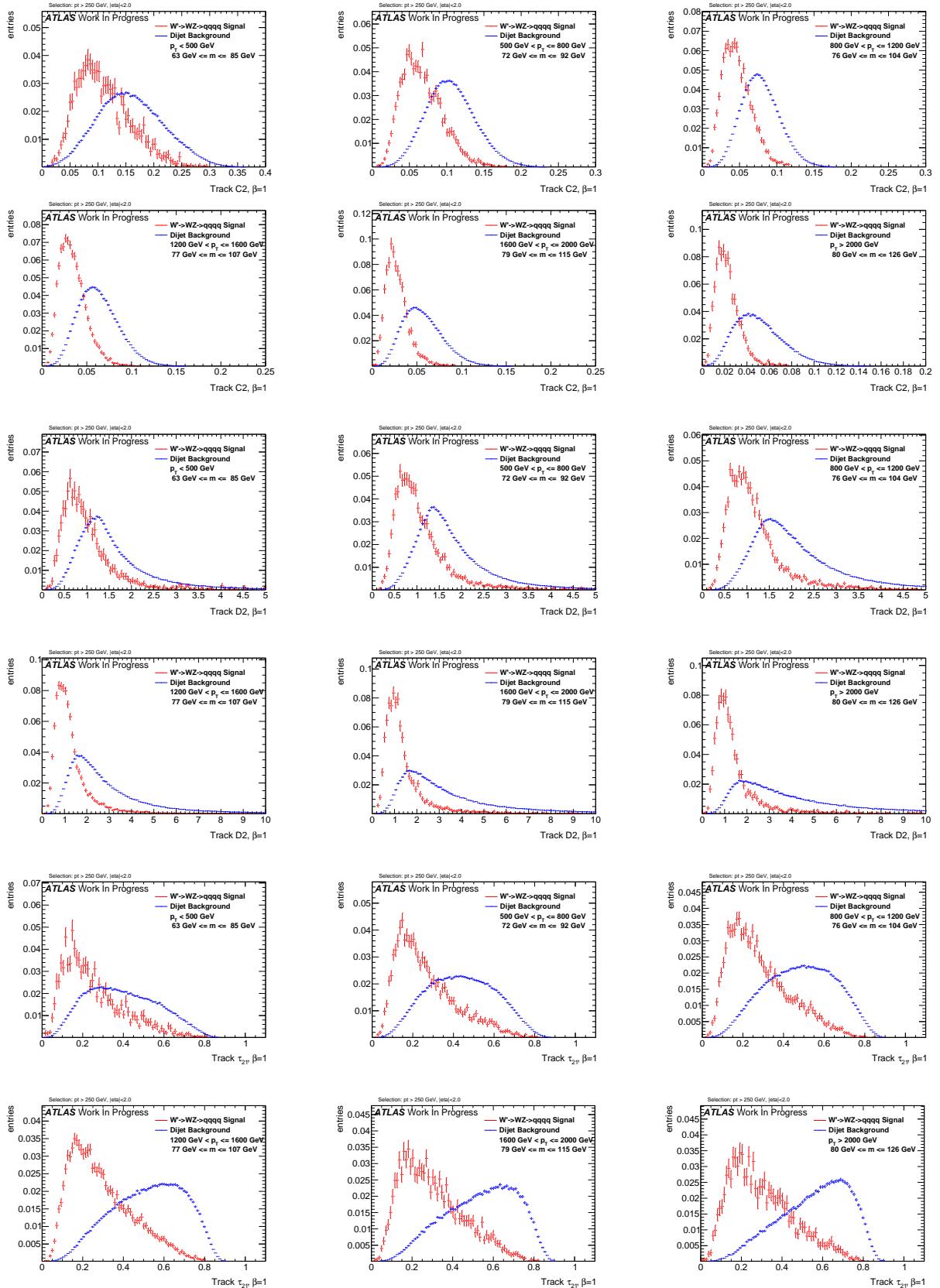
Figure 279: ROCs showing QCD rejection against Top jet efficiency for $\tau_{32, \text{TAS}}^{(\beta=1)}$ & $\tau_{32, \text{TAS}}^{(\beta=1.7)}$ compared to $\tau_{32, \text{TAS}}^{(\beta=2)}$

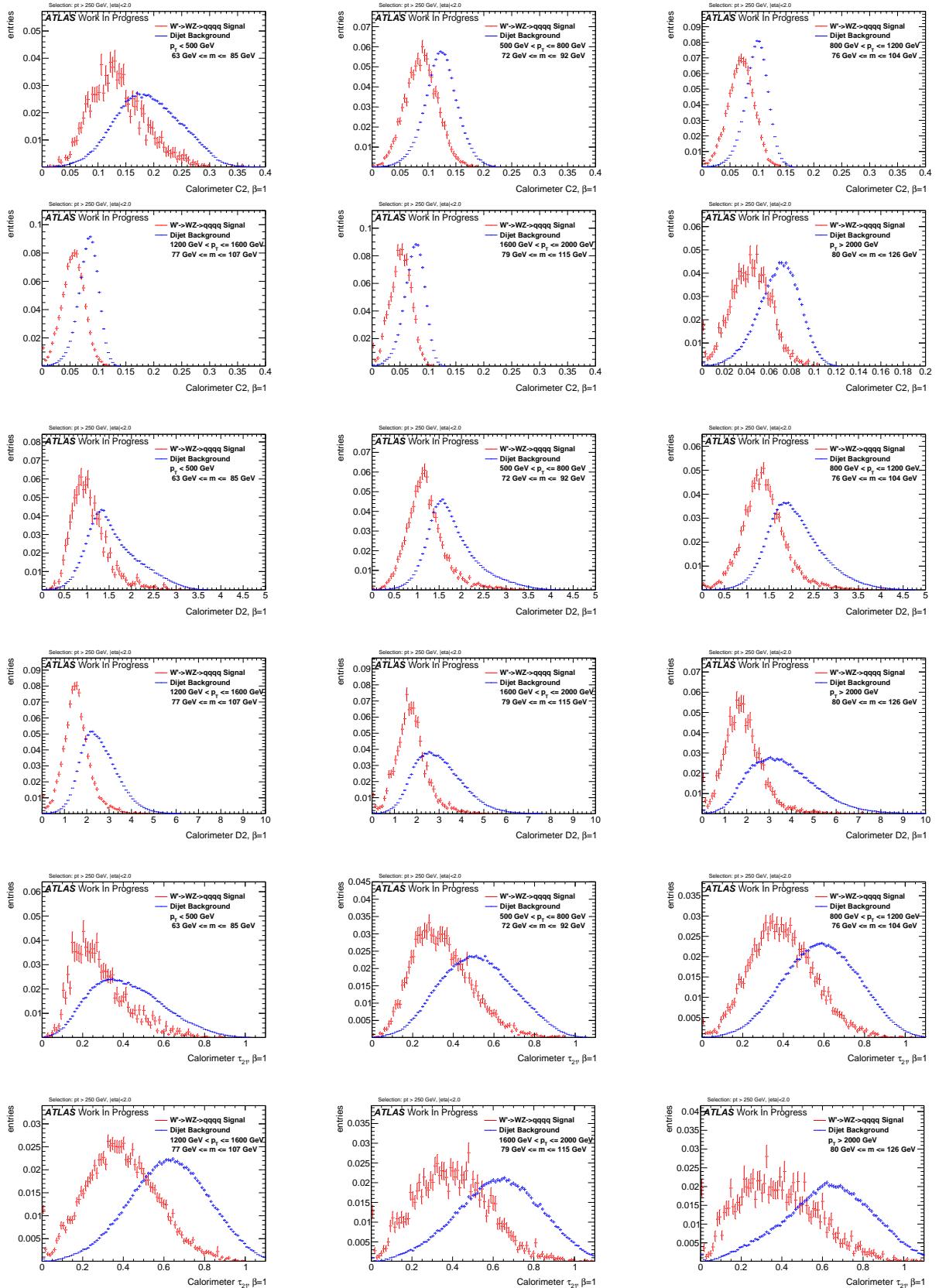
Figure 280: Distributions for W boson tagging using TAS $\beta = 0.5$. C2, D2, τ_{21} top down.

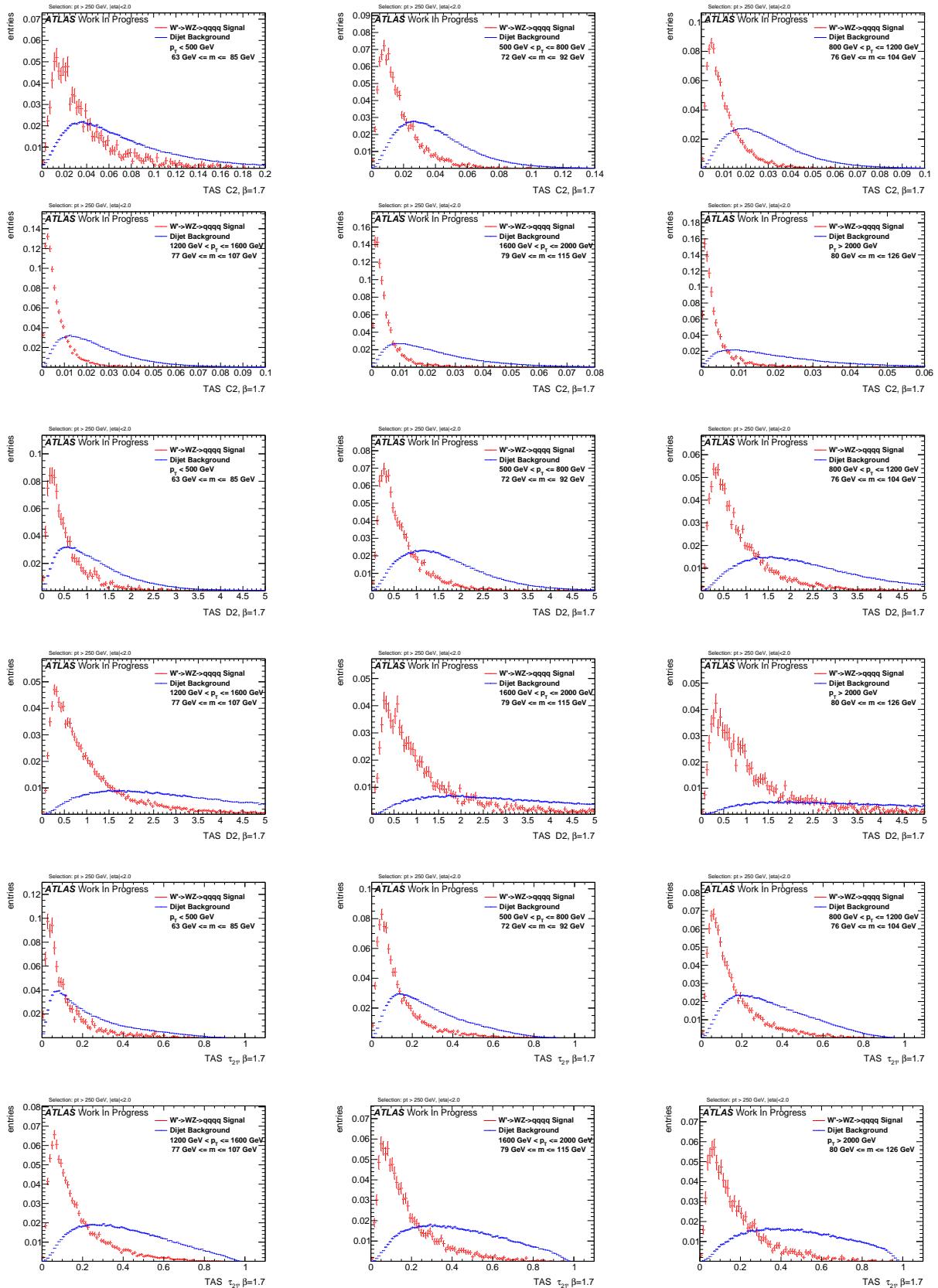
Figure 281: Distributions for W boson tagging using tracks $\beta = 0.5$. C2, D2, τ_{21} top down.

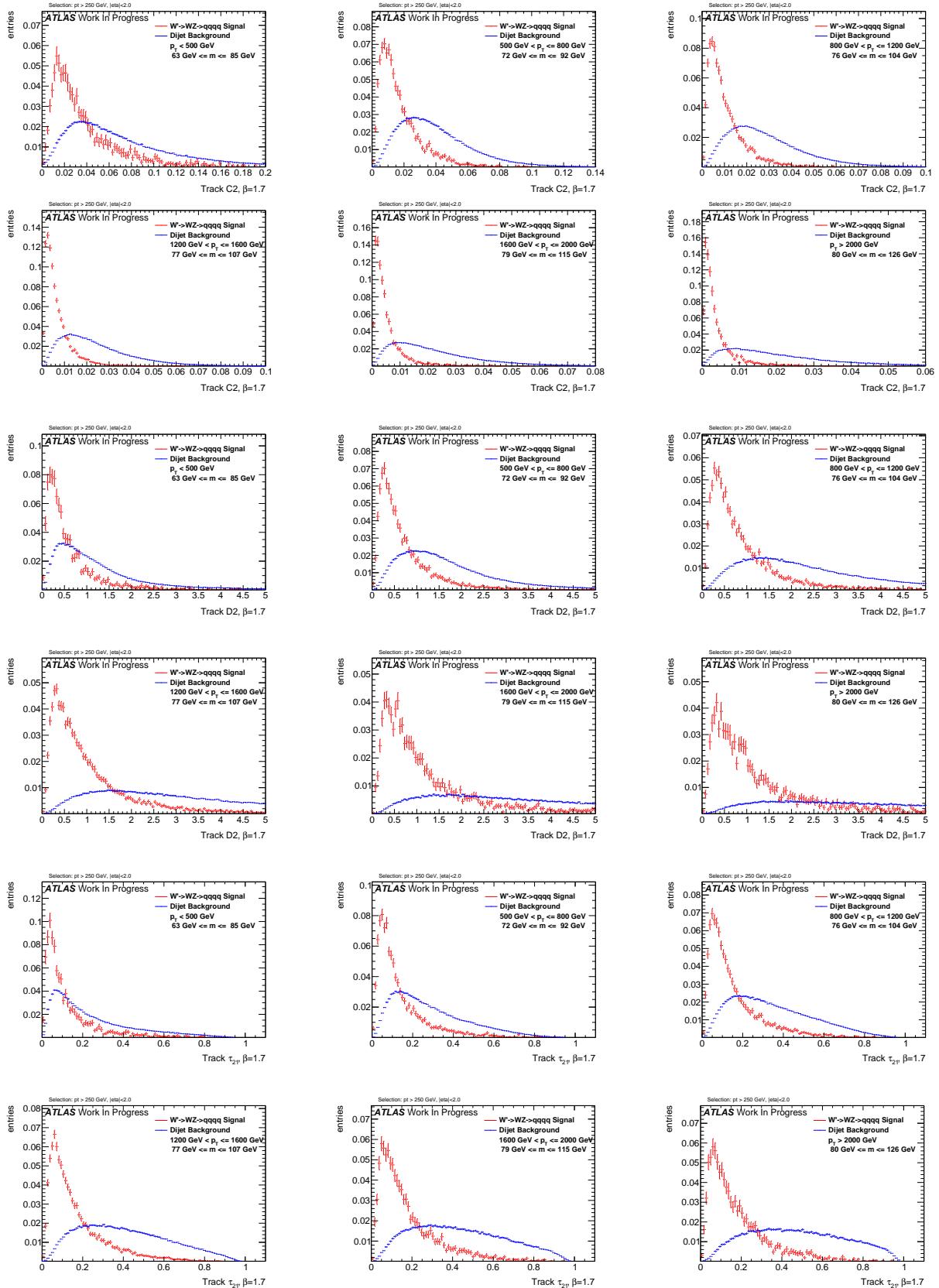
Figure 282: Distributions for W boson tagging using calorimeter clusters $\beta = 0.5$. C2, D2, τ_{21} top down.

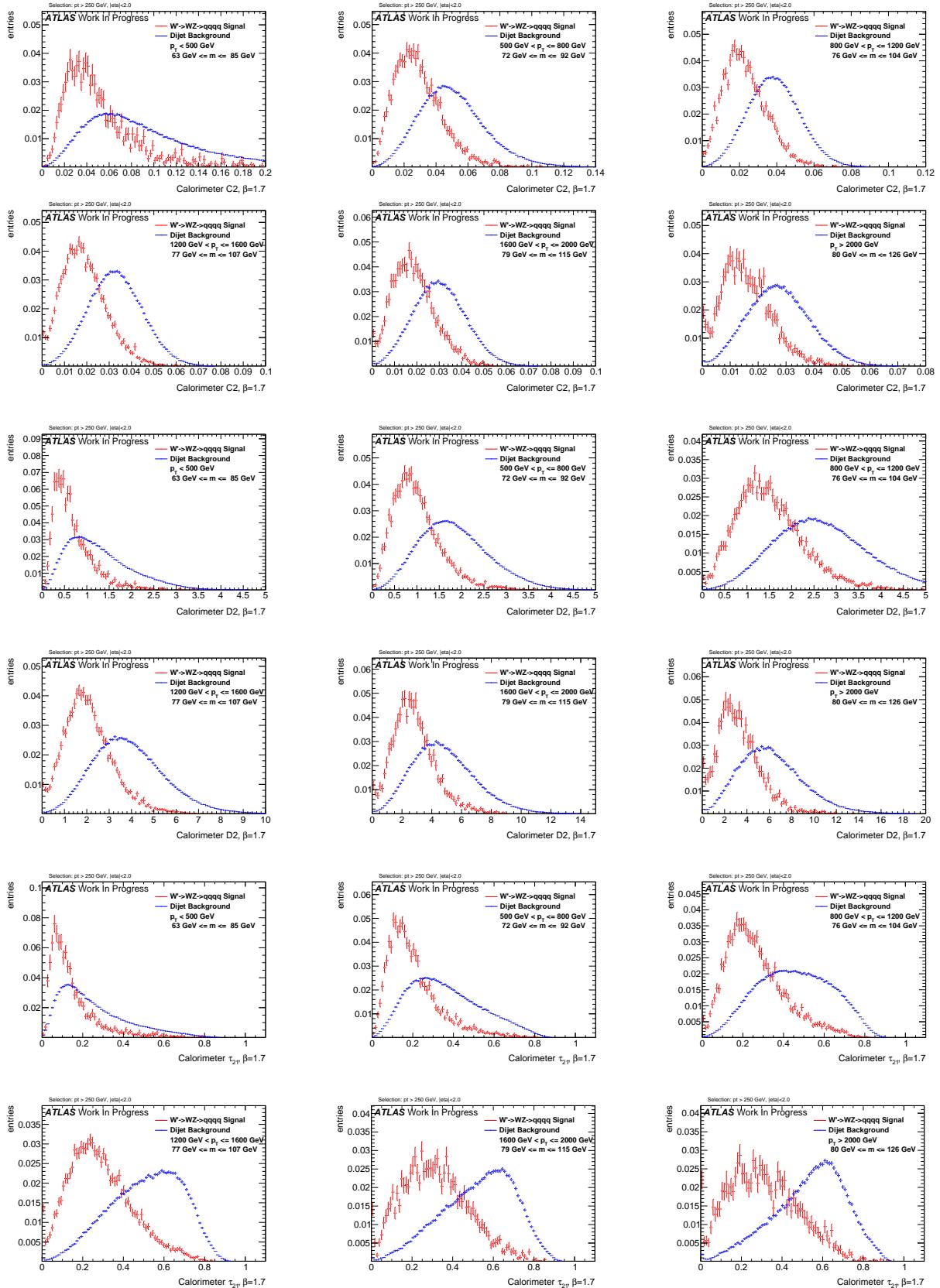
Figure 283: Distributions for W boson tagging using TAS $\beta = 1$. C2, D2, τ_{21} top down.

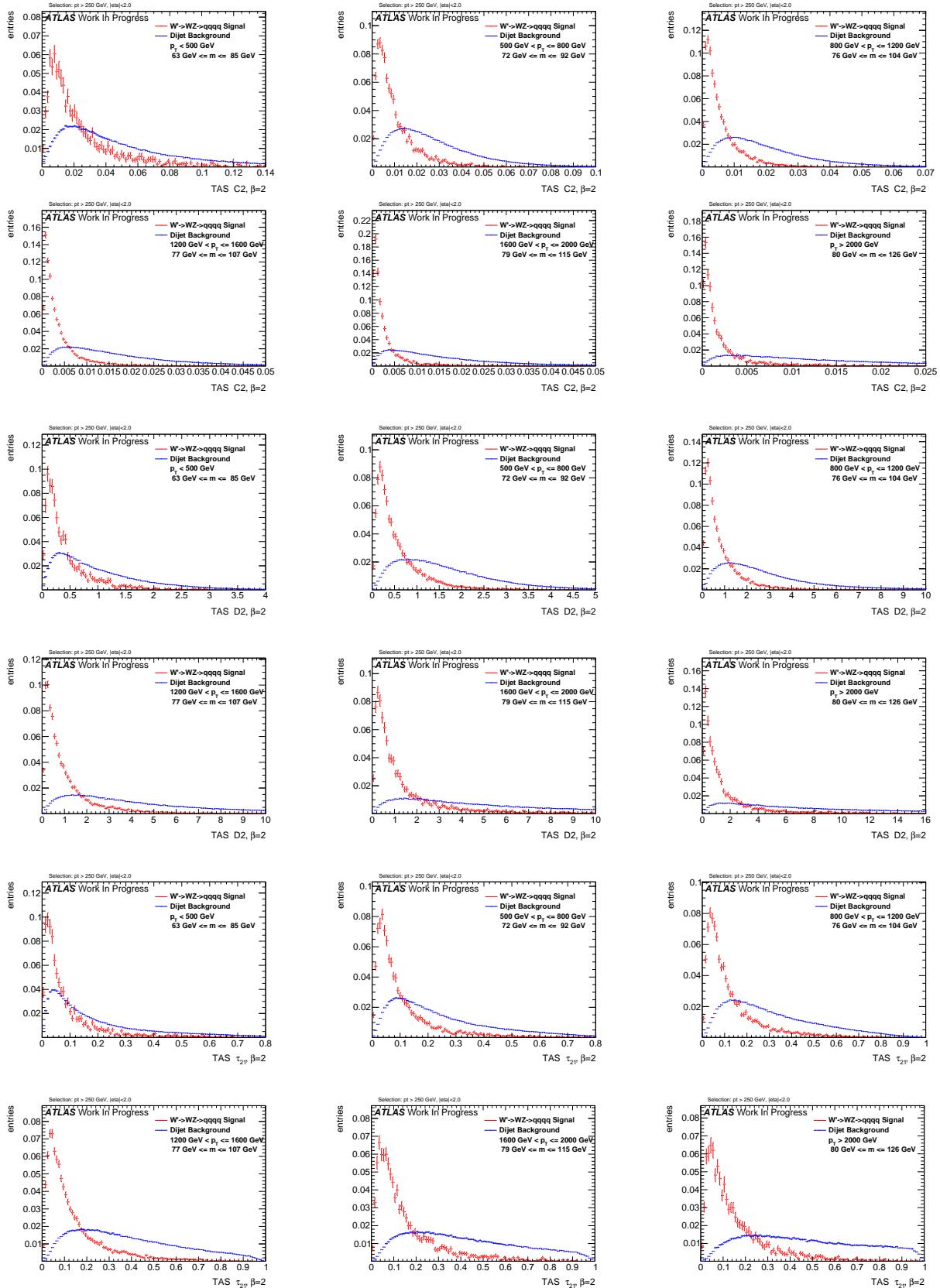
Figure 284: Distributions for W boson tagging using tracks $\beta = 1$. C2, D2, τ_{21} top down.

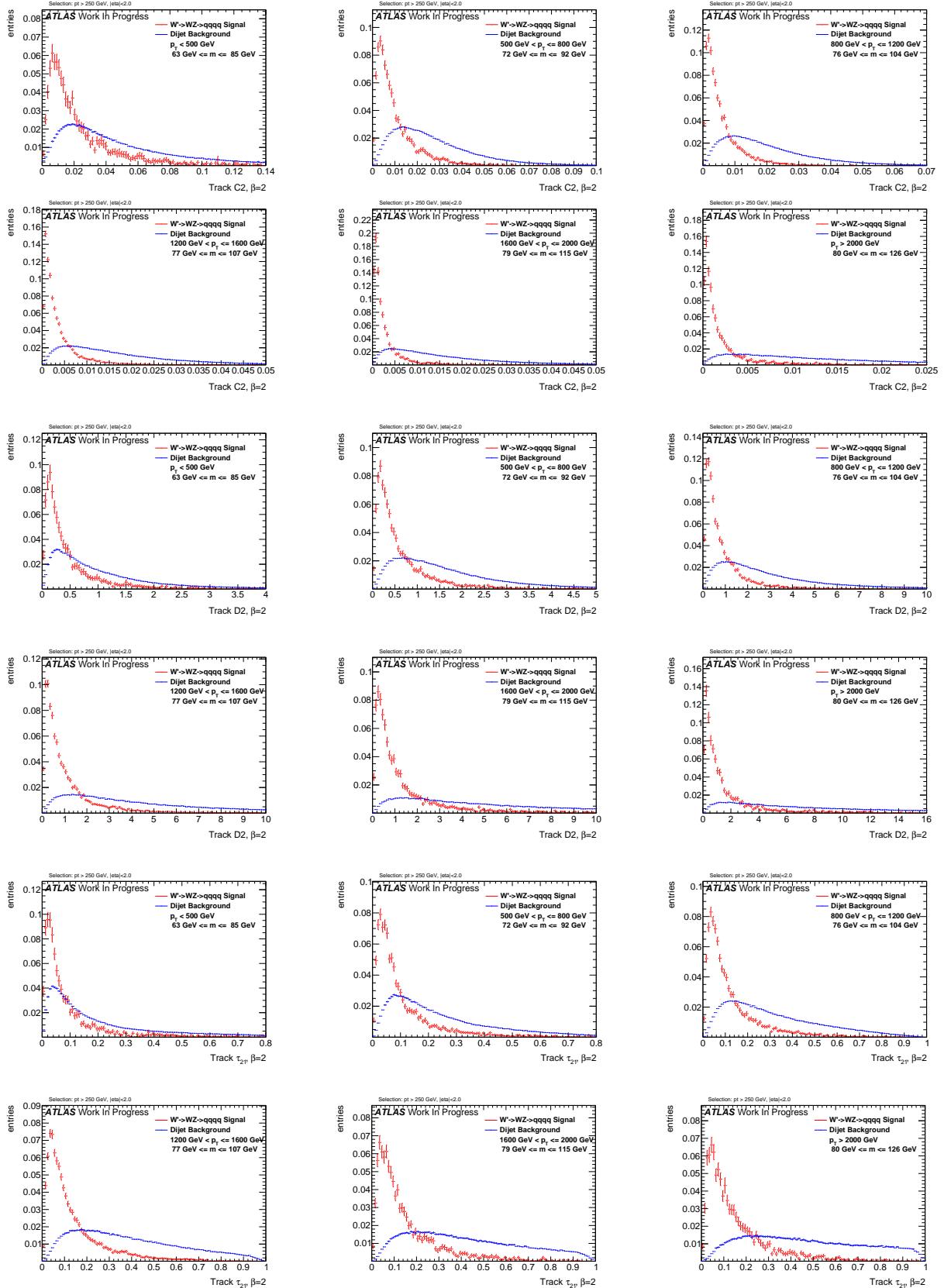
Figure 285: Distributions for W boson tagging using calorimeter clusters $\beta = 1$. C2, D2, τ_{21} top down.

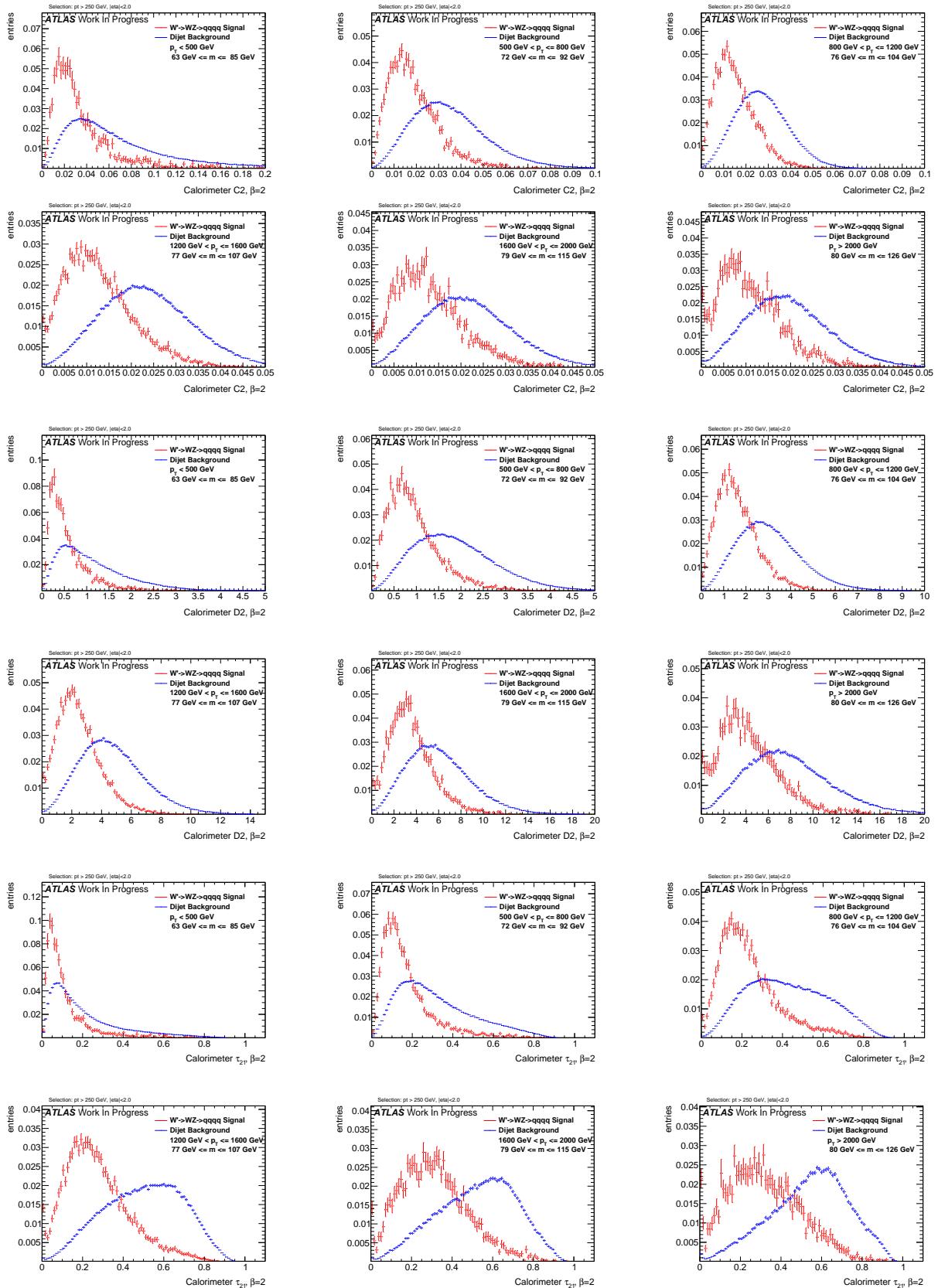
Figure 286: Distributions for W boson tagging using TAS $\beta = 1.7$. C2, D2, τ_{21} top down.

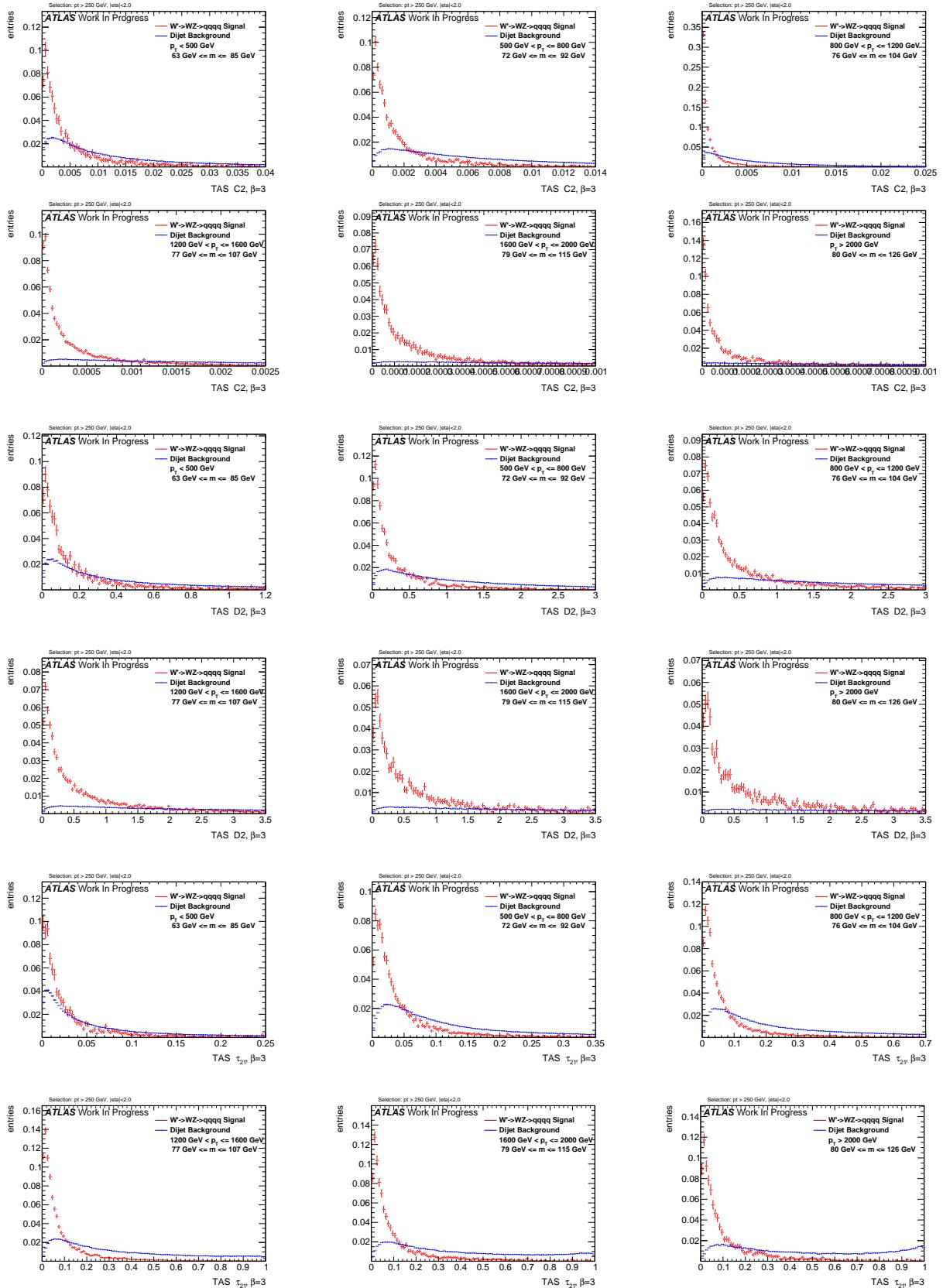
Figure 287: Distributions for W boson tagging using tracks $\beta = 1.7$. C2, D2, τ_{21} top down.

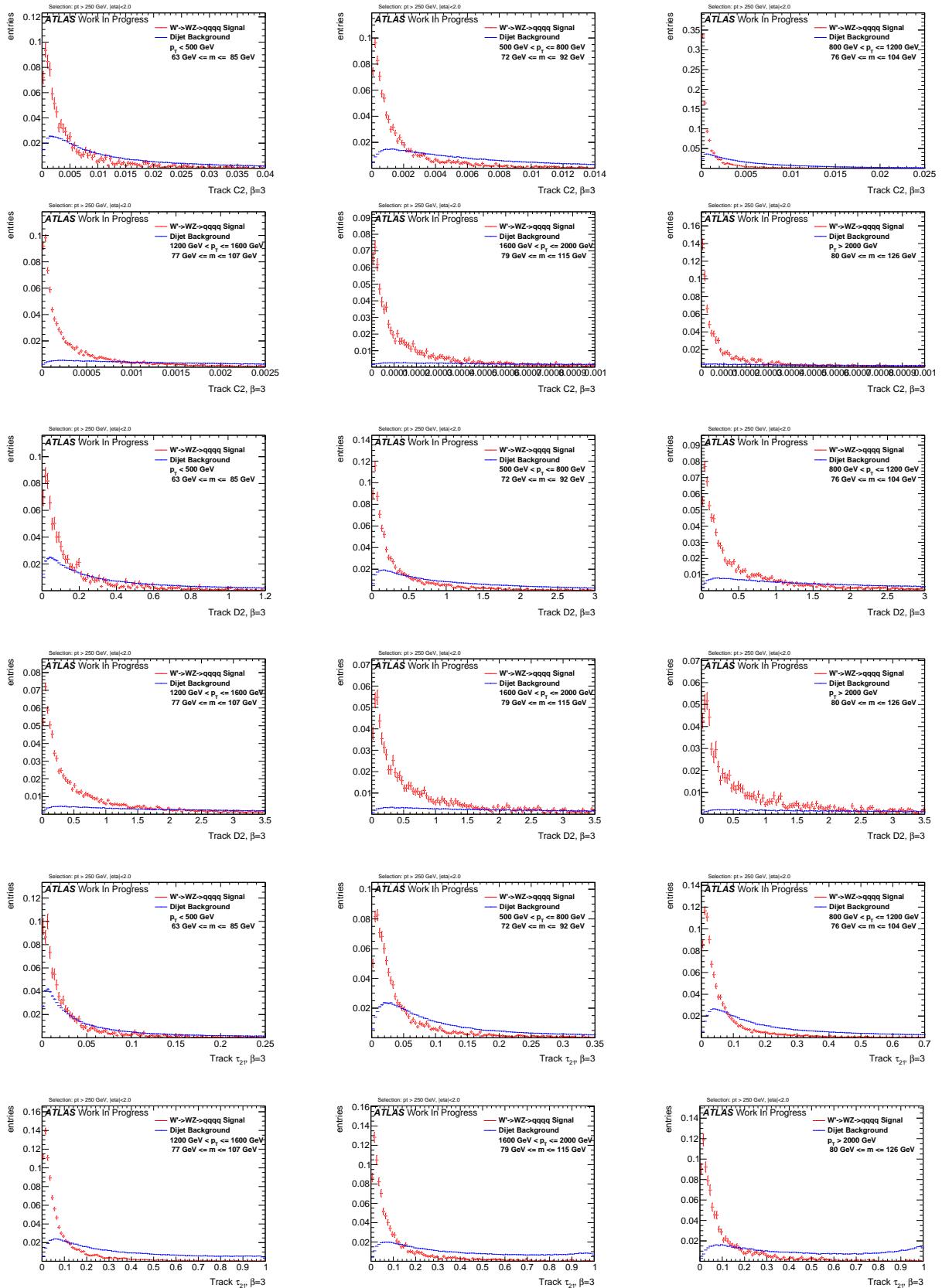
Figure 288: Distributions for W boson tagging using calorimeter clusters $\beta = 1.7$. C2, D2, τ_{21} top down.

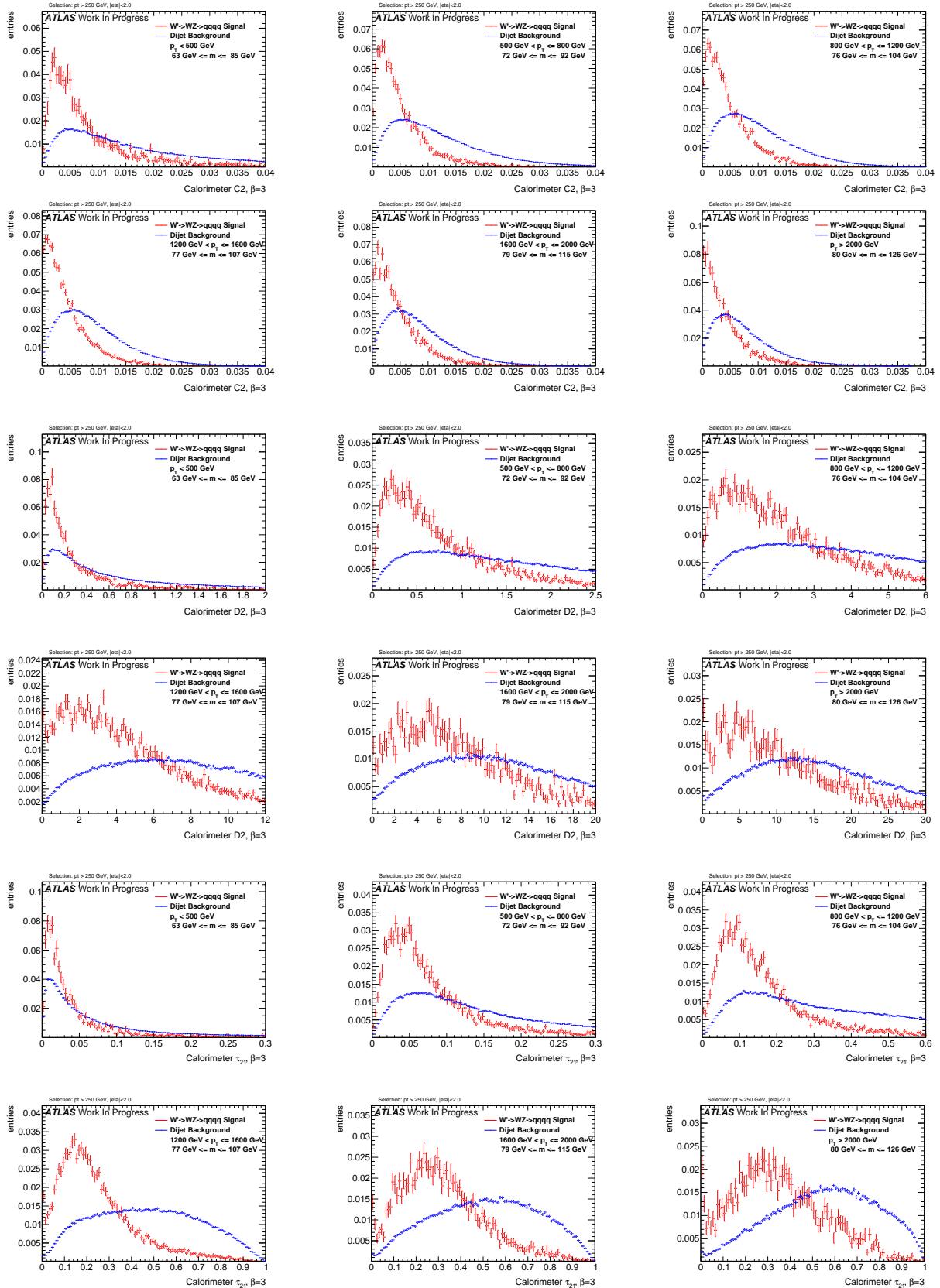
Figure 289: Distributions for W boson tagging using TAS $\beta = 2$. C2, D2, τ_{21} top down.

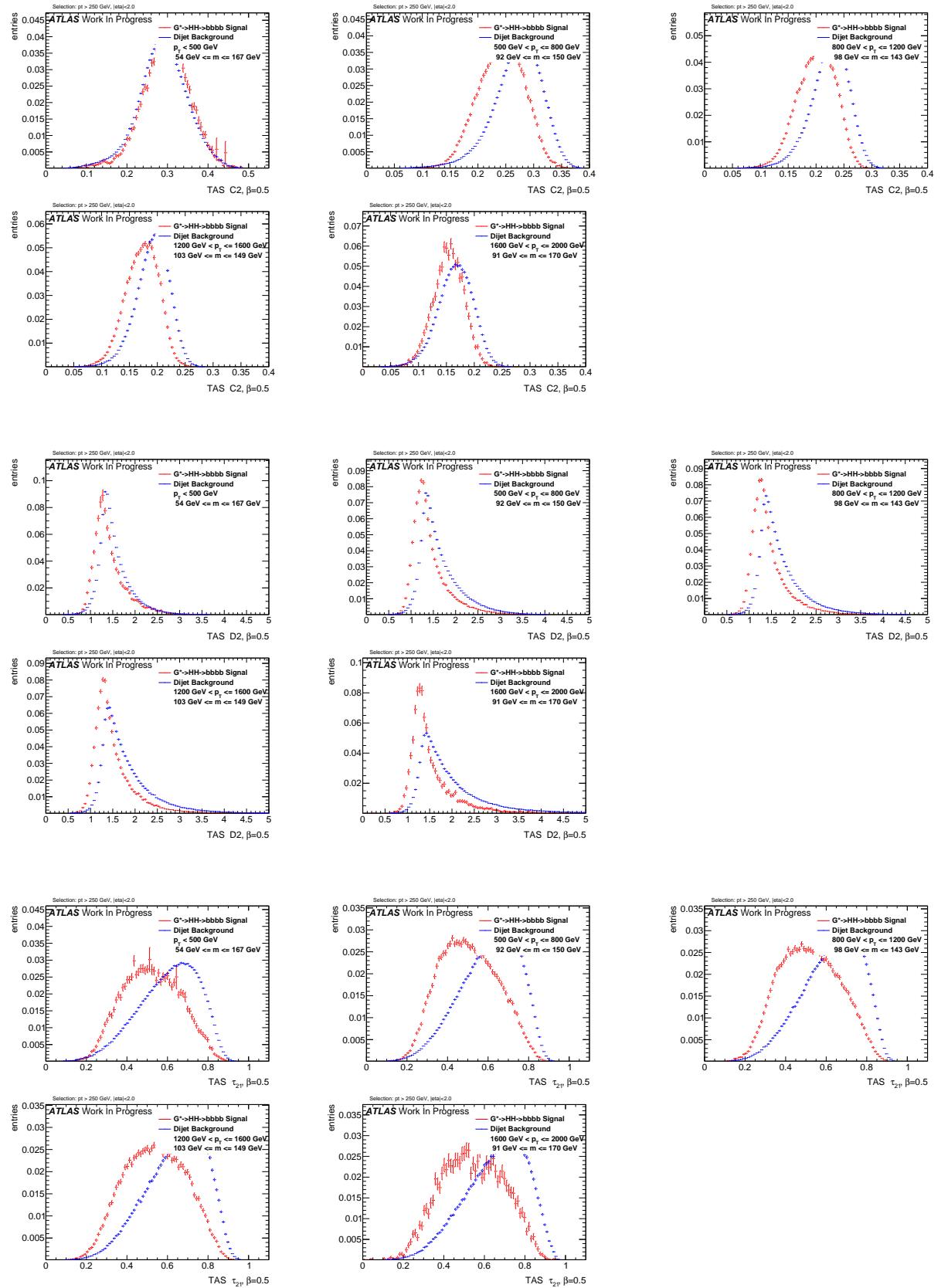
Figure 290: Distributions for W boson tagging using tracks $\beta = 2$. C2, D2, τ_{21} top down.

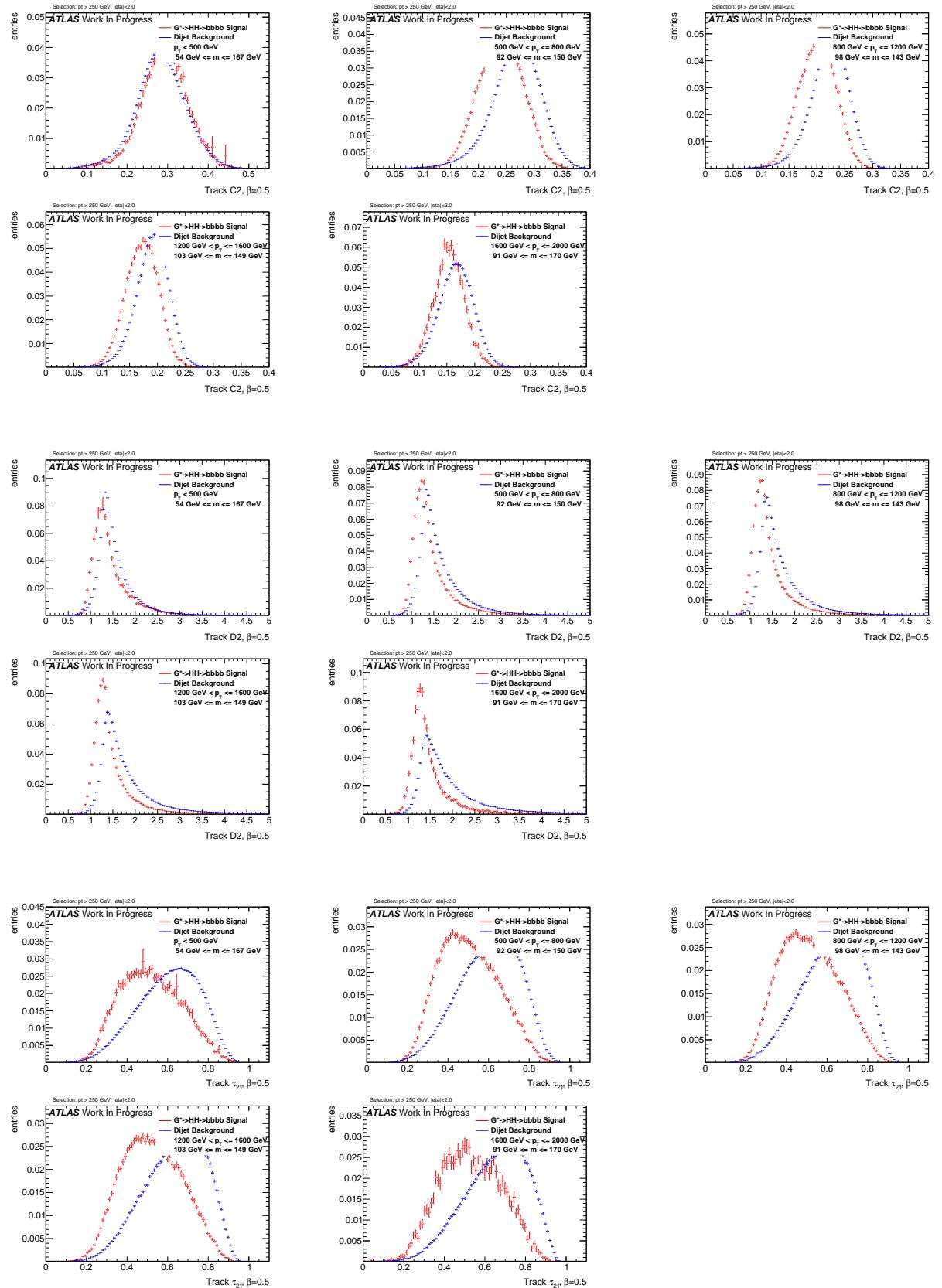
Figure 291: Distributions for W boson tagging using calorimeter clusters $\beta = 2$. C2, D2, τ_{21} top down.

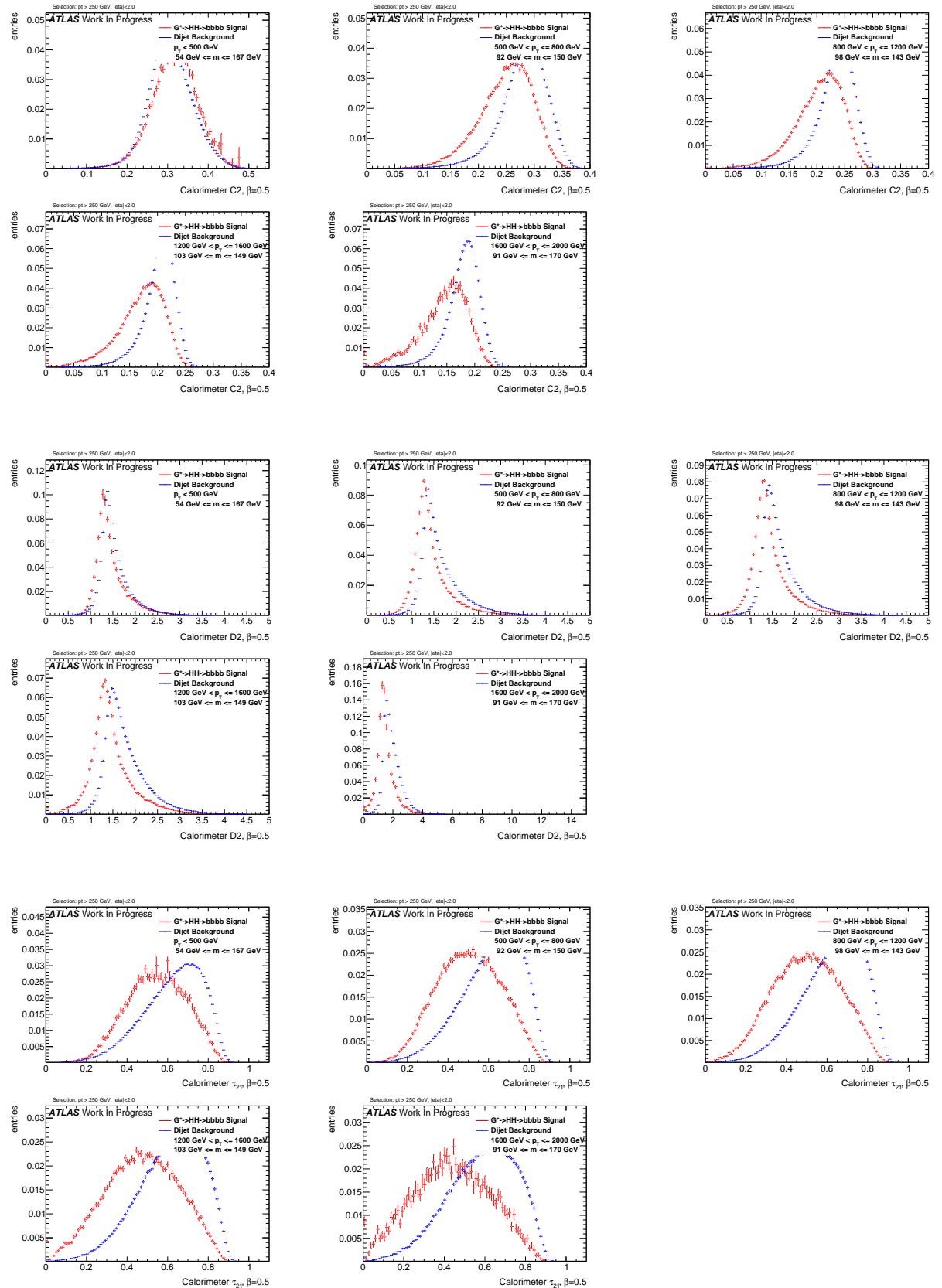
Figure 292: Distributions for W boson tagging using TAS $\beta = 3$. C2, D2, τ_{21} top down.

Figure 293: Distributions for W boson tagging using tracks $\beta = 3$. C2, D2, τ_{21} top down.

Figure 294: Distributions for W boson tagging using calorimeter clusters $\beta = 3$. C2, D2, τ_{21} top down.

Figure 295: Distributions for Higgs boson tagging using TAS $\beta = 0.5$. C2, D2, τ_{21} top down.

Figure 296: Distributions for Higgs boson tagging using tracks $\beta = 0.5$. C2, D2, τ_{21} top down.

Figure 297: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 0.5$. C2, D2, τ_{21} top down.

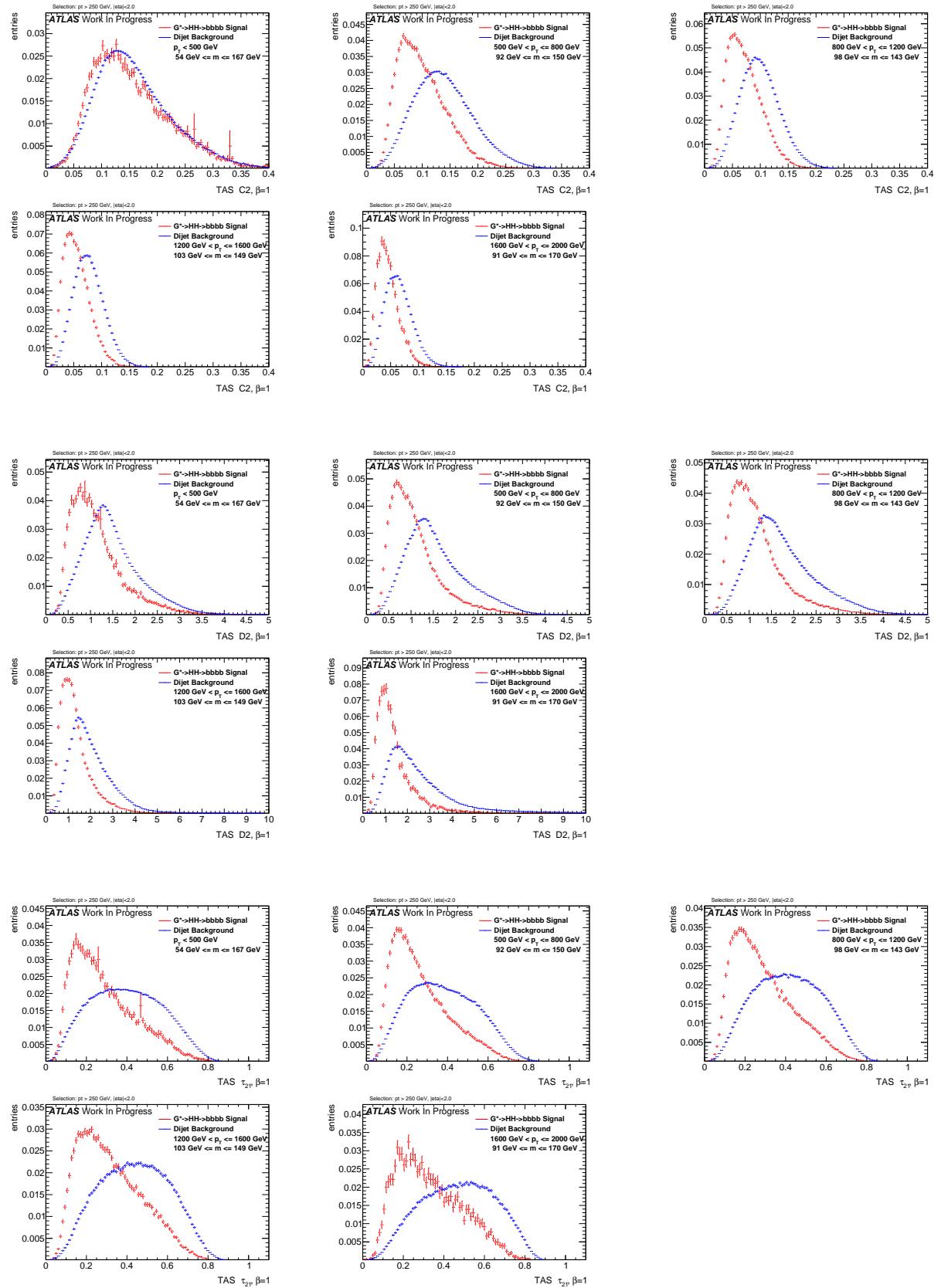
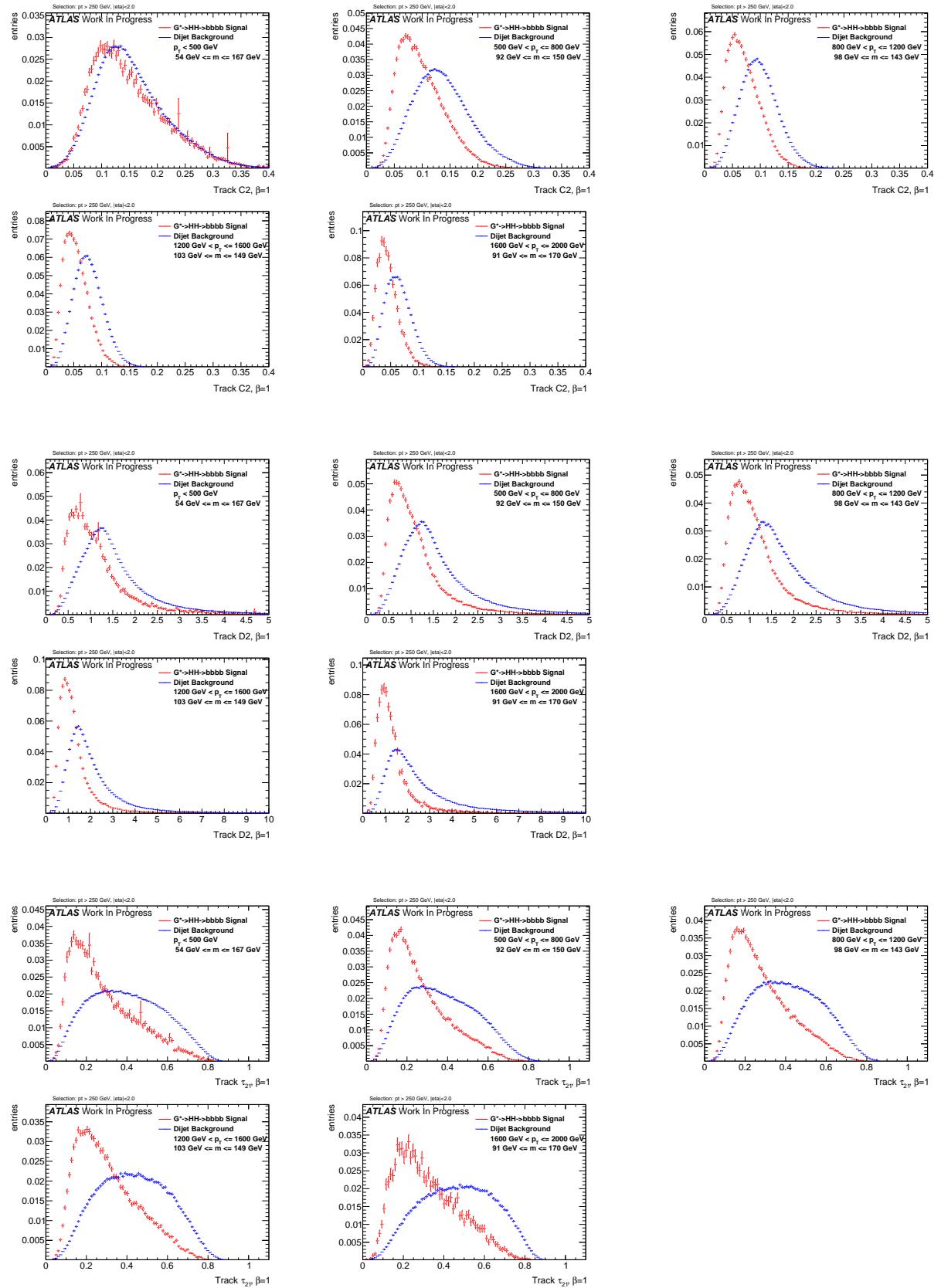
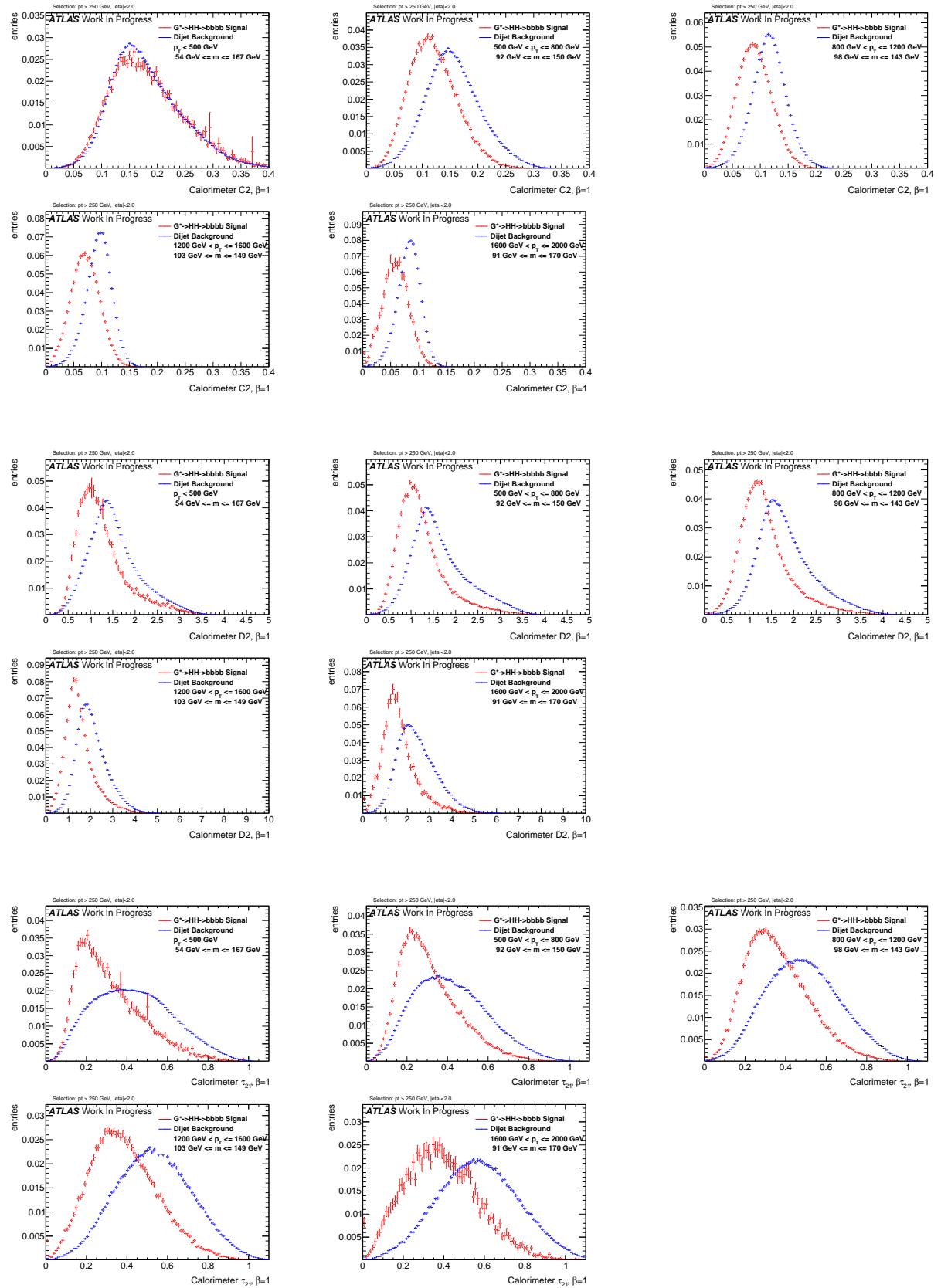
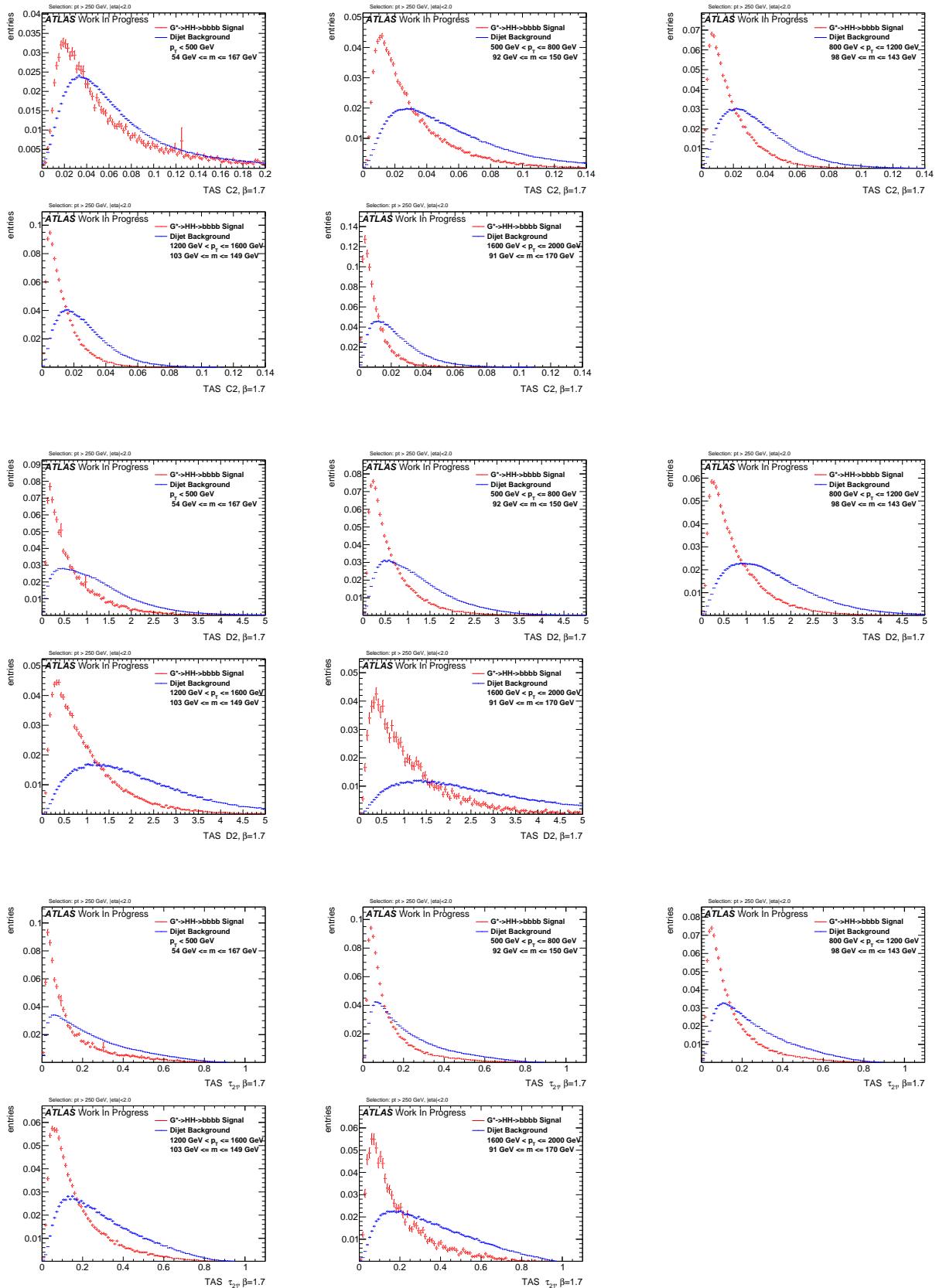
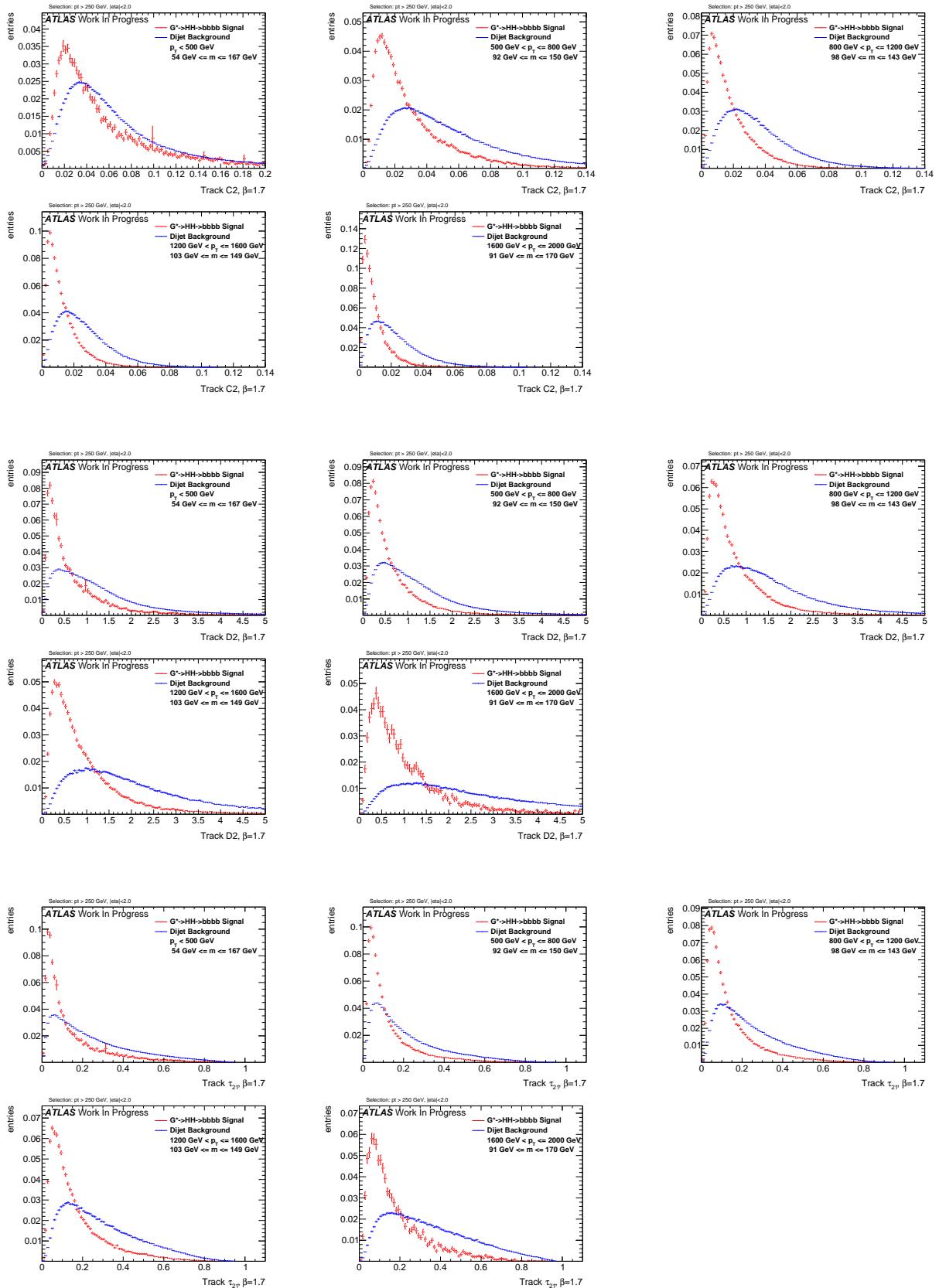


Figure 298: Distributions for Higgs boson tagging using TAS $\beta = 1$. C2, D2, τ_{21} top down.

Figure 299: Distributions for Higgs boson tagging using tracks $\beta = 1$. C2, D2, τ_{21} top down.

Figure 300: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 1$. C2, D2, τ_{21} top down.

Figure 301: Distributions for Higgs boson tagging using TAS $\beta = 1.7$. C2, D2, τ_{21} top down.

Figure 302: Distributions for Higgs boson tagging using tracks $\beta = 1.7$. C2, D2, τ_{21} top down.

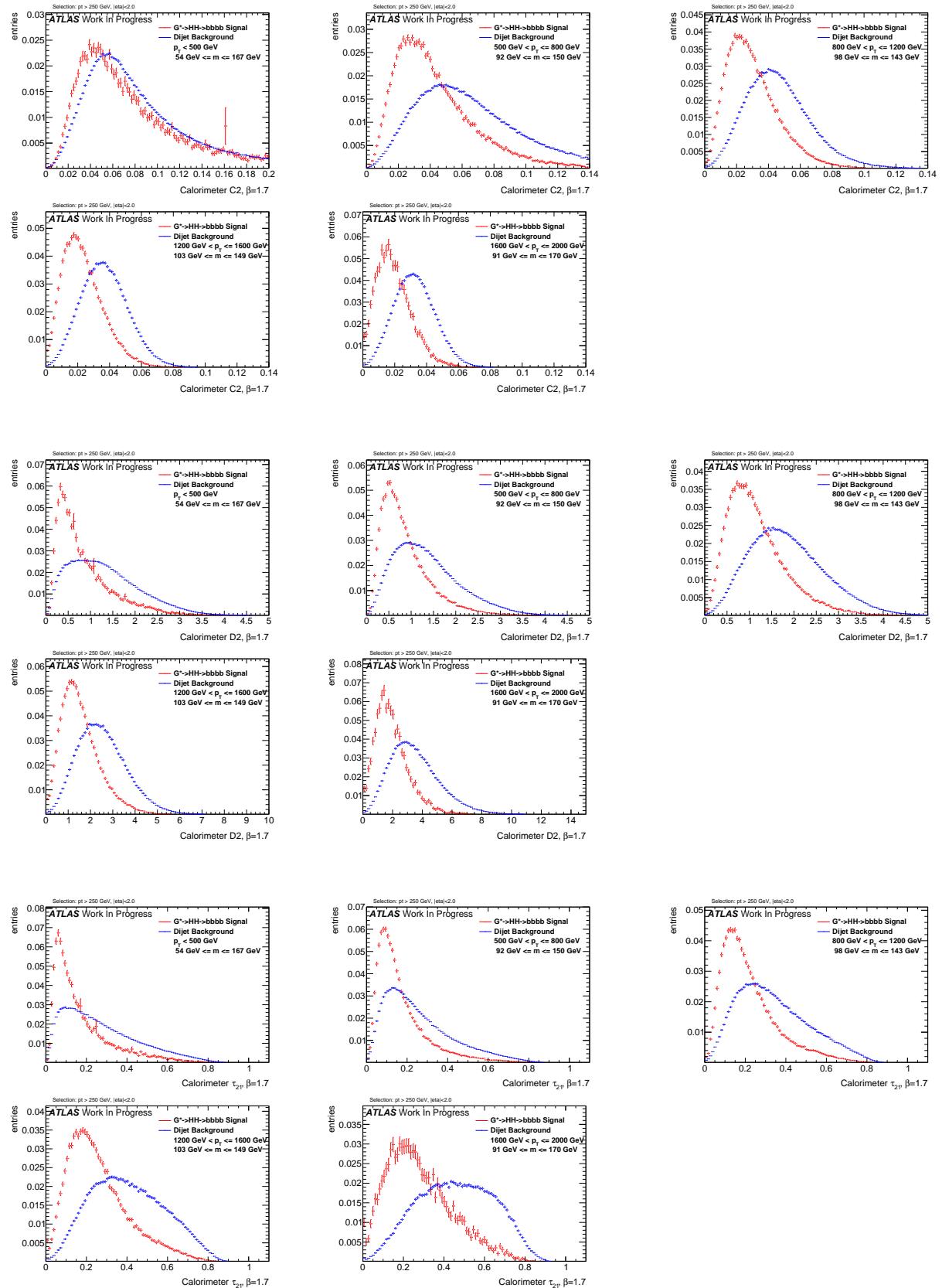
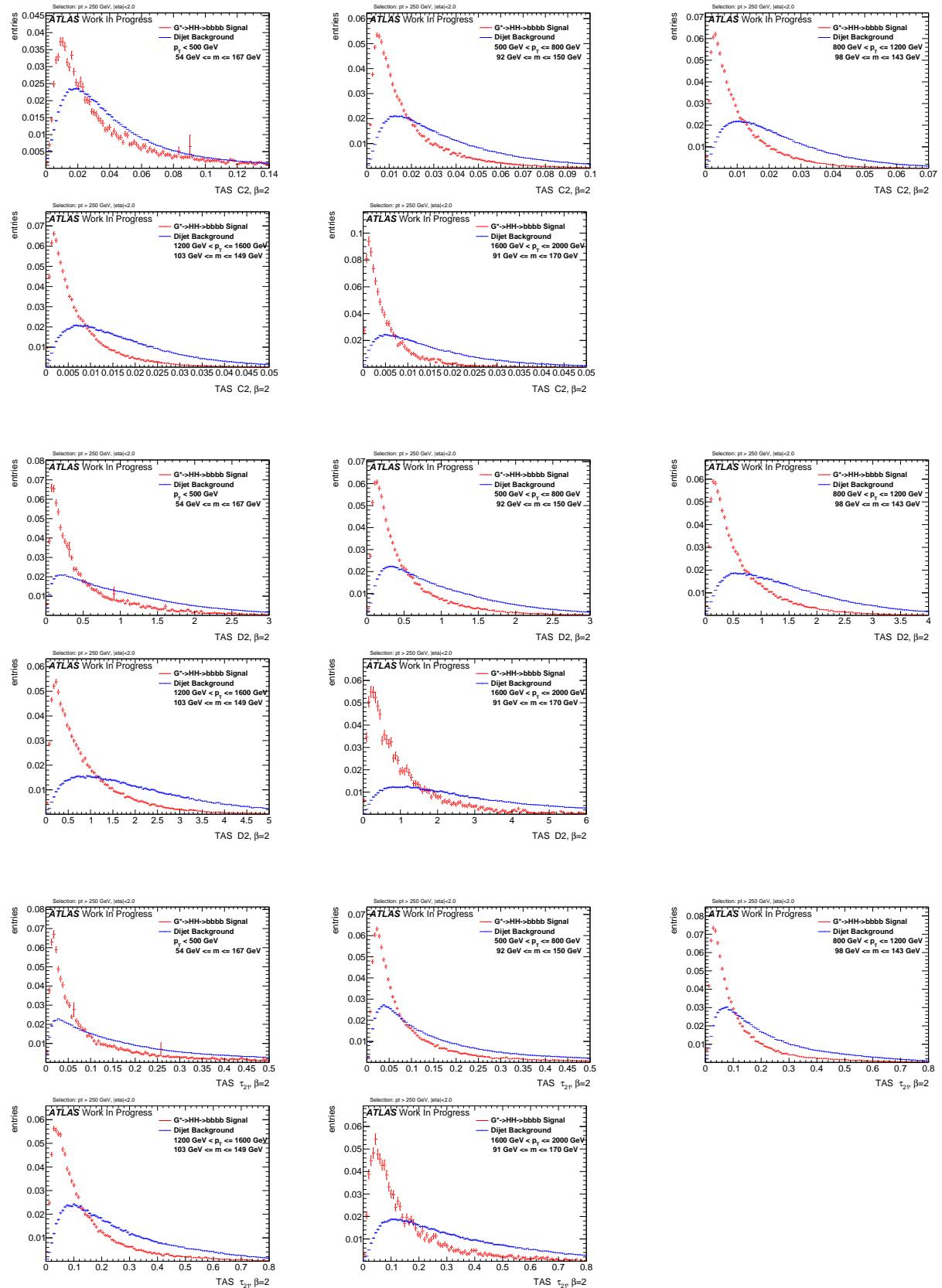
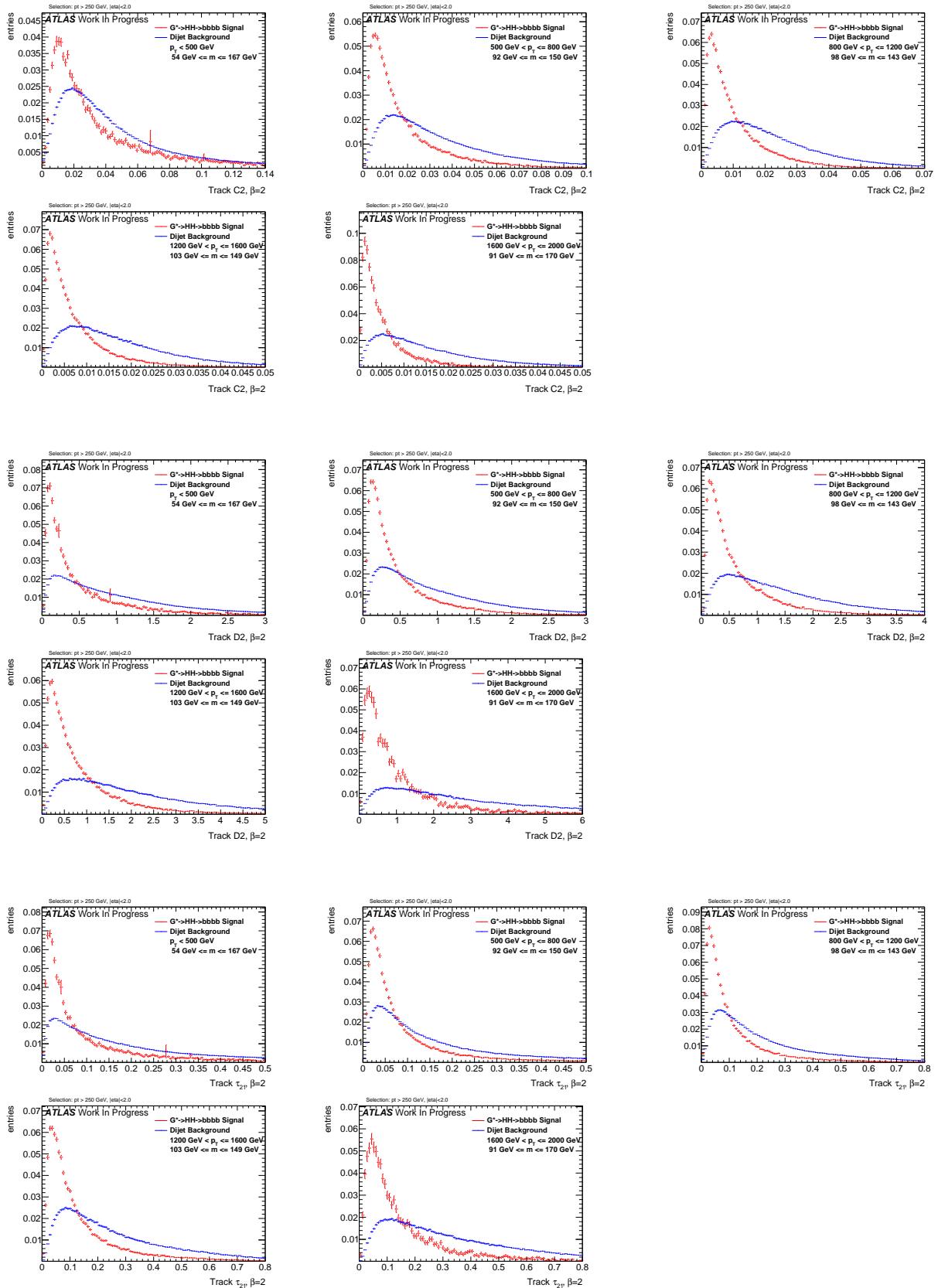
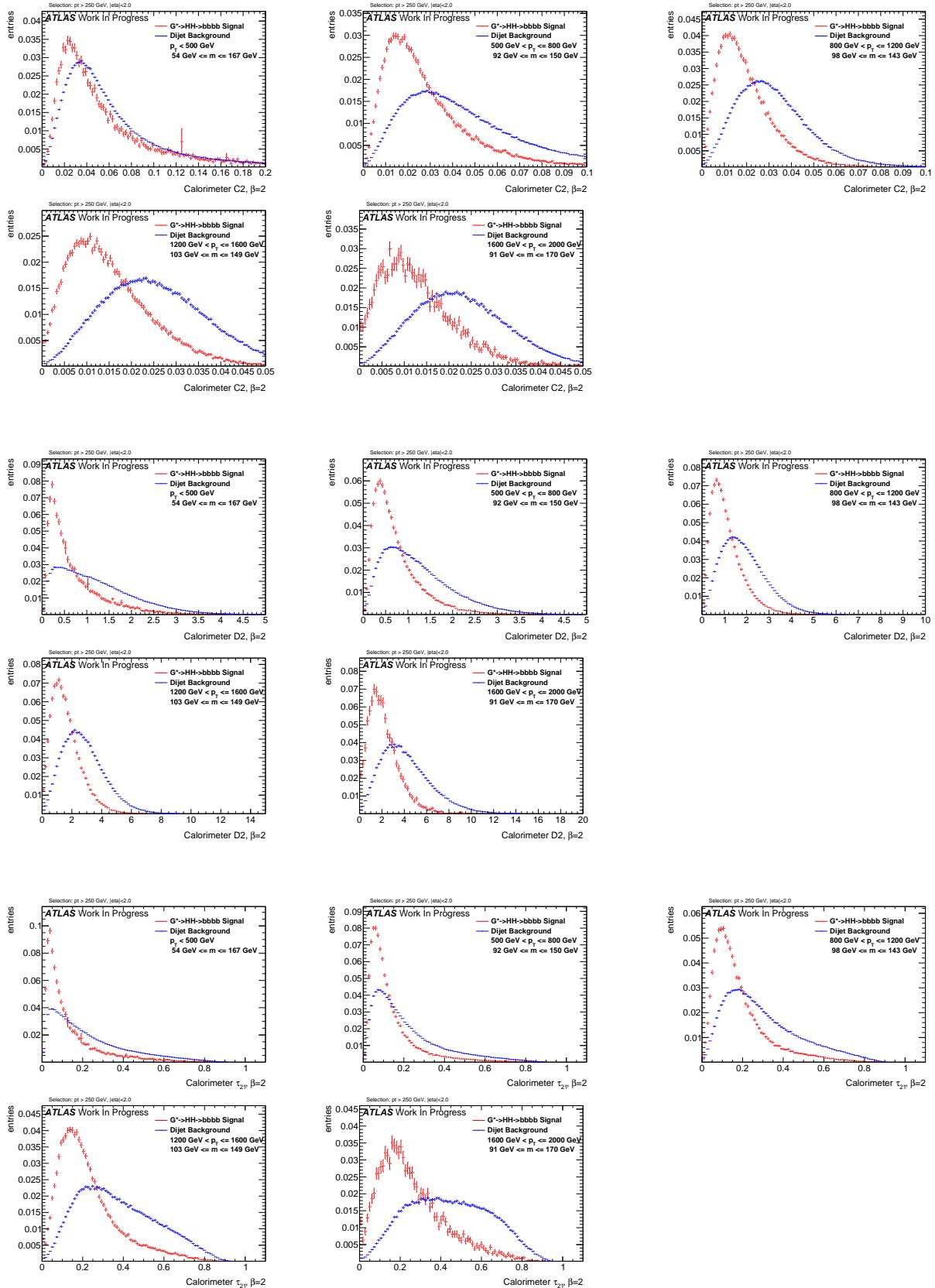
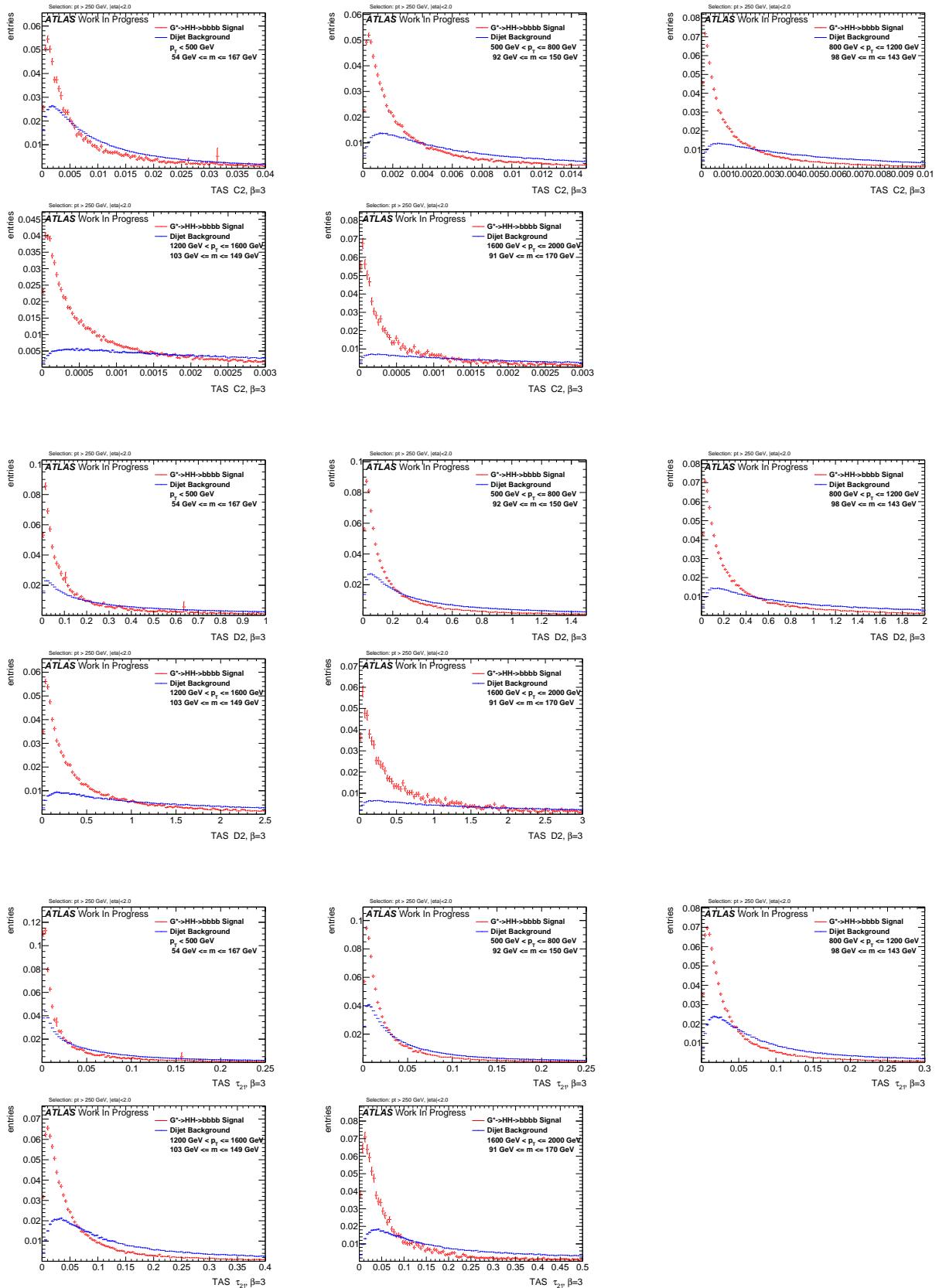


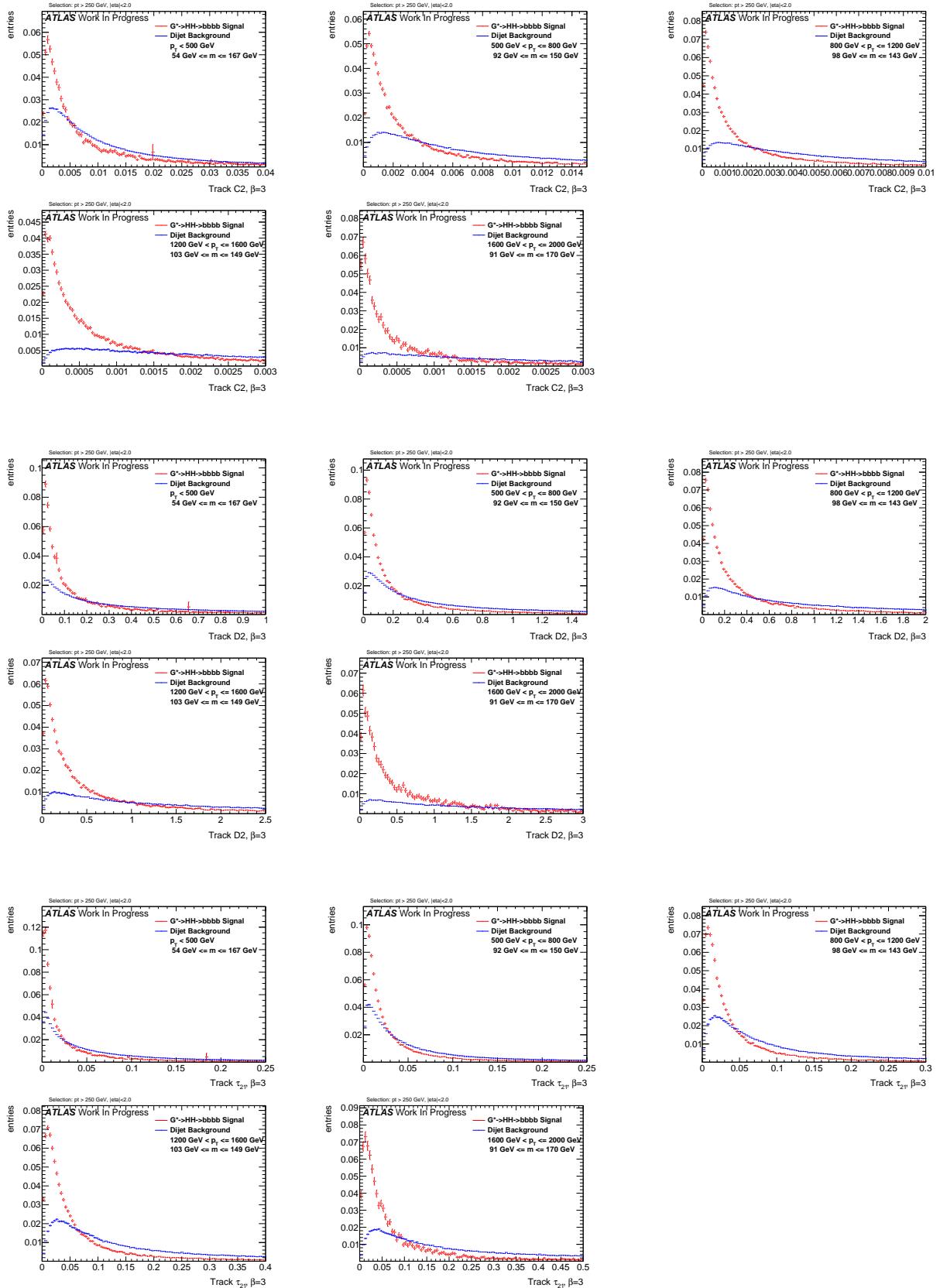
Figure 303: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 1.7$. C2, D2, τ_{21} top down.

Figure 304: Distributions for Higgs boson tagging using TAS $\beta = 2$. C2, D2, t_{21} top down.

Figure 305: Distributions for Higgs boson tagging using tracks $\beta = 2$. C2, D2, τ_{21} top down.

Figure 306: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 2$. C2, D2, τ_{21} top down.

Figure 307: Distributions for Higgs boson tagging using TAS $\beta = 3$. C2, D2, t_{21} top down.

Figure 308: Distributions for Higgs boson tagging using tracks $\beta = 3$. C2, D2, τ_{21} top down.

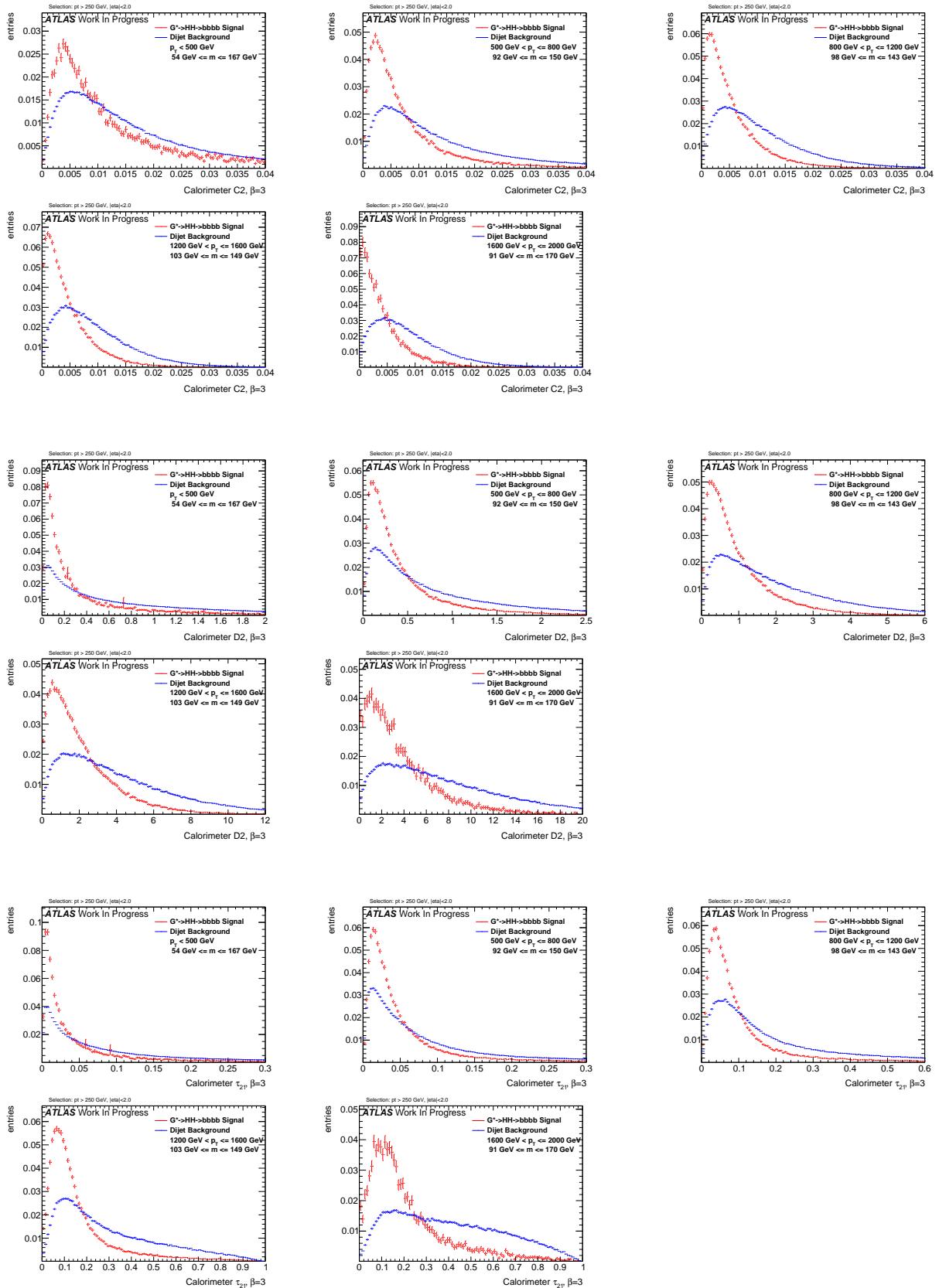


Figure 309: Distributions for Higgs boson tagging using calorimeter clusters $\beta = 3$. C2, D2, τ_{21} top down.

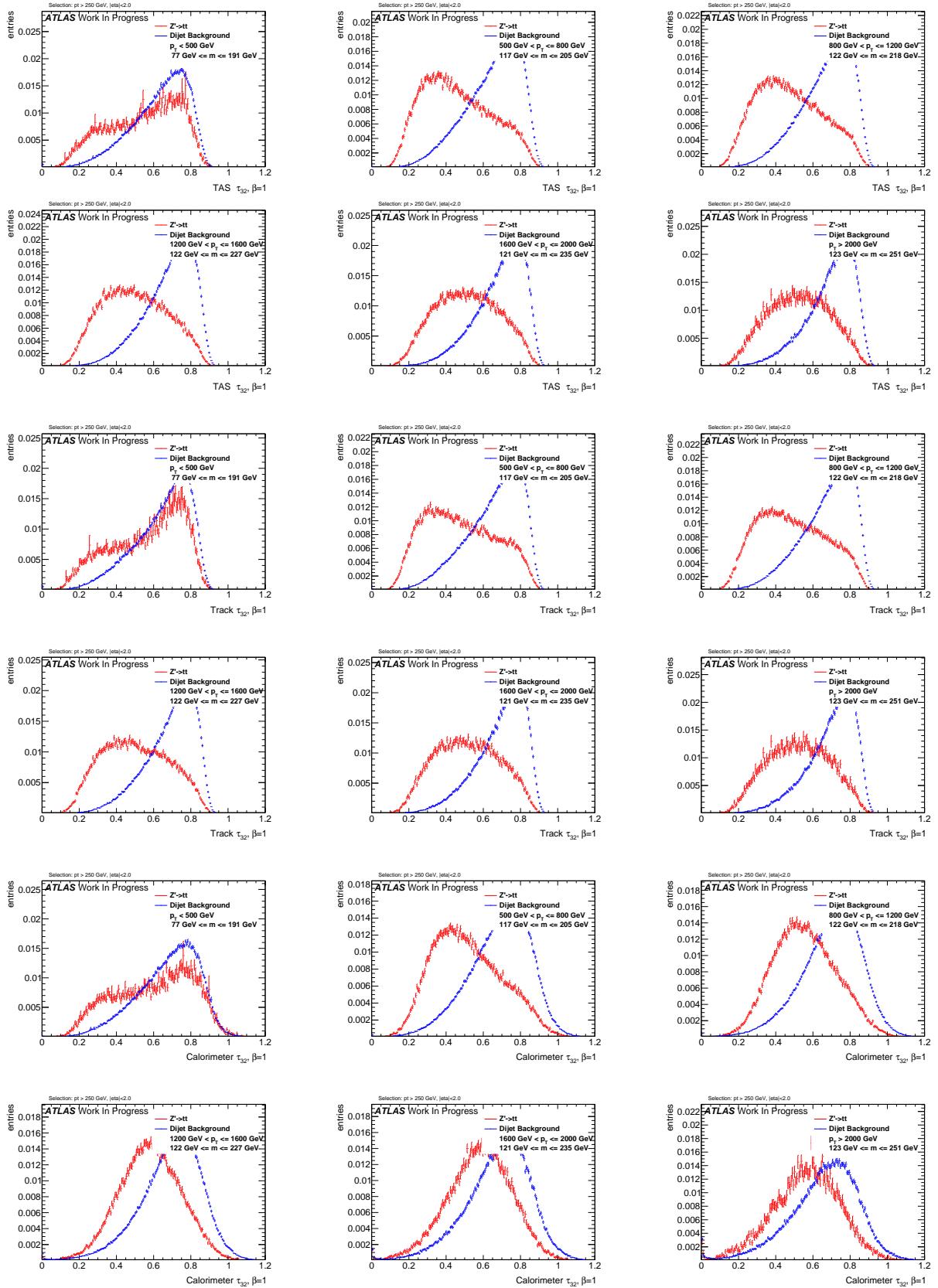
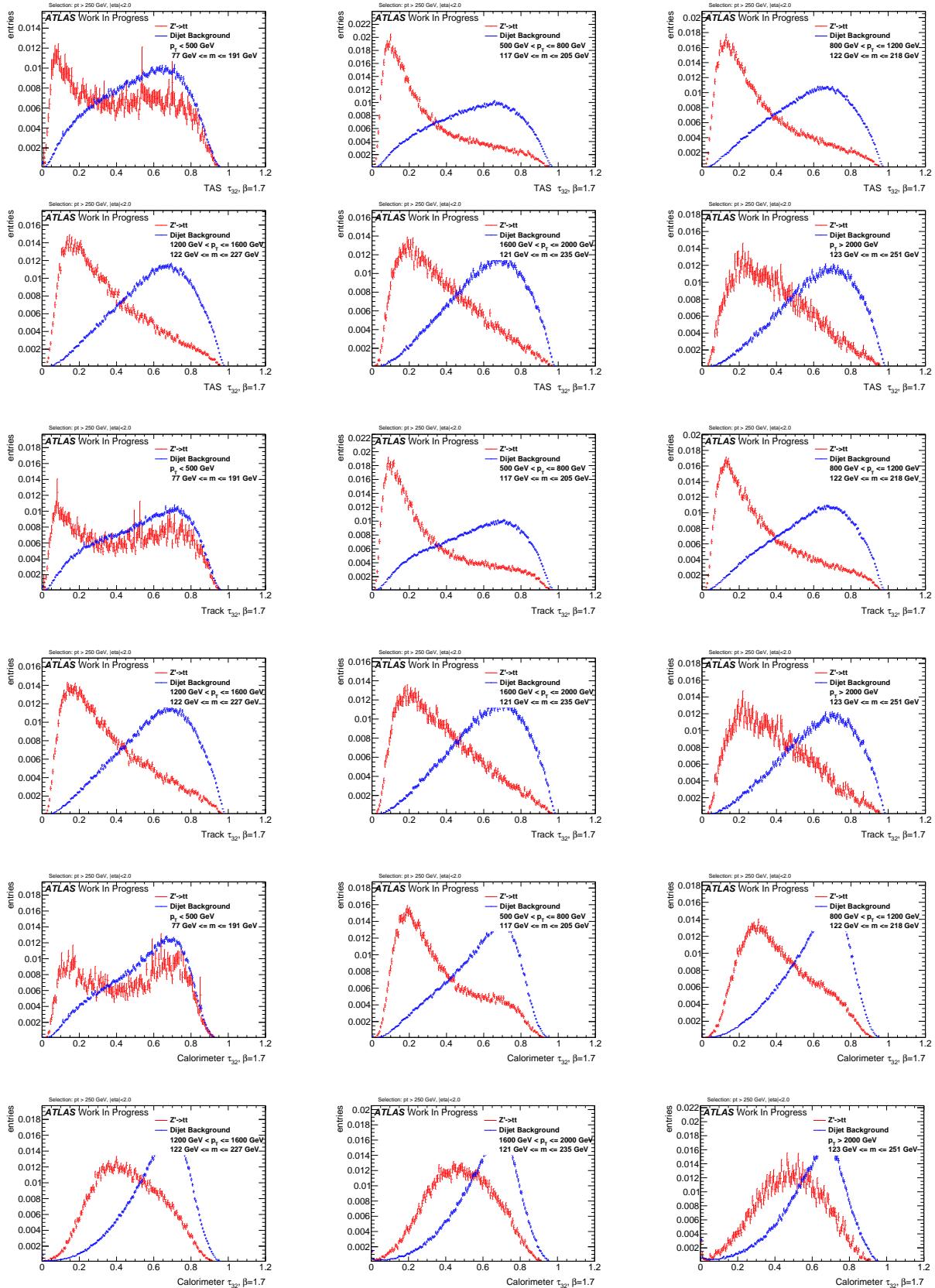
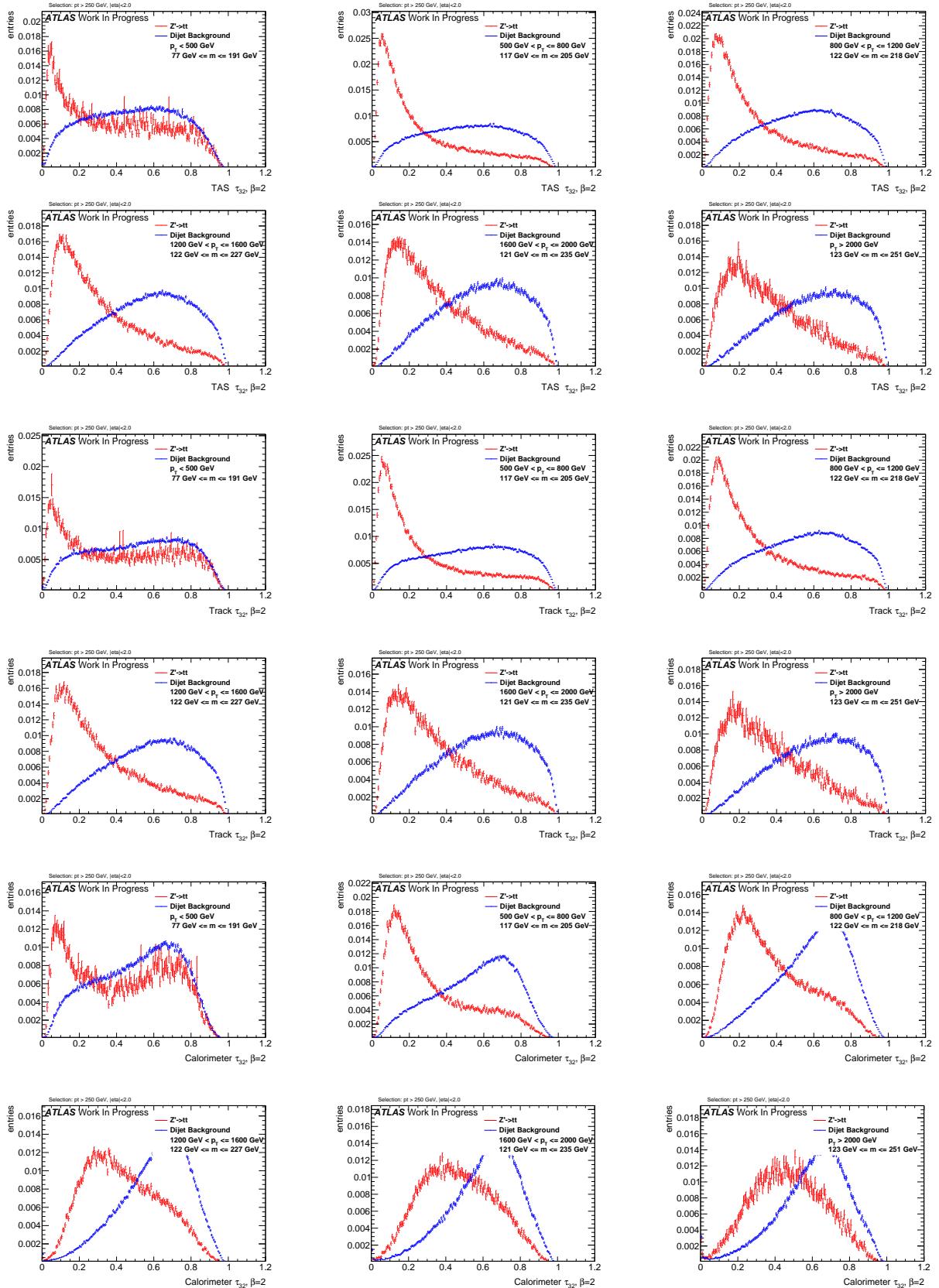
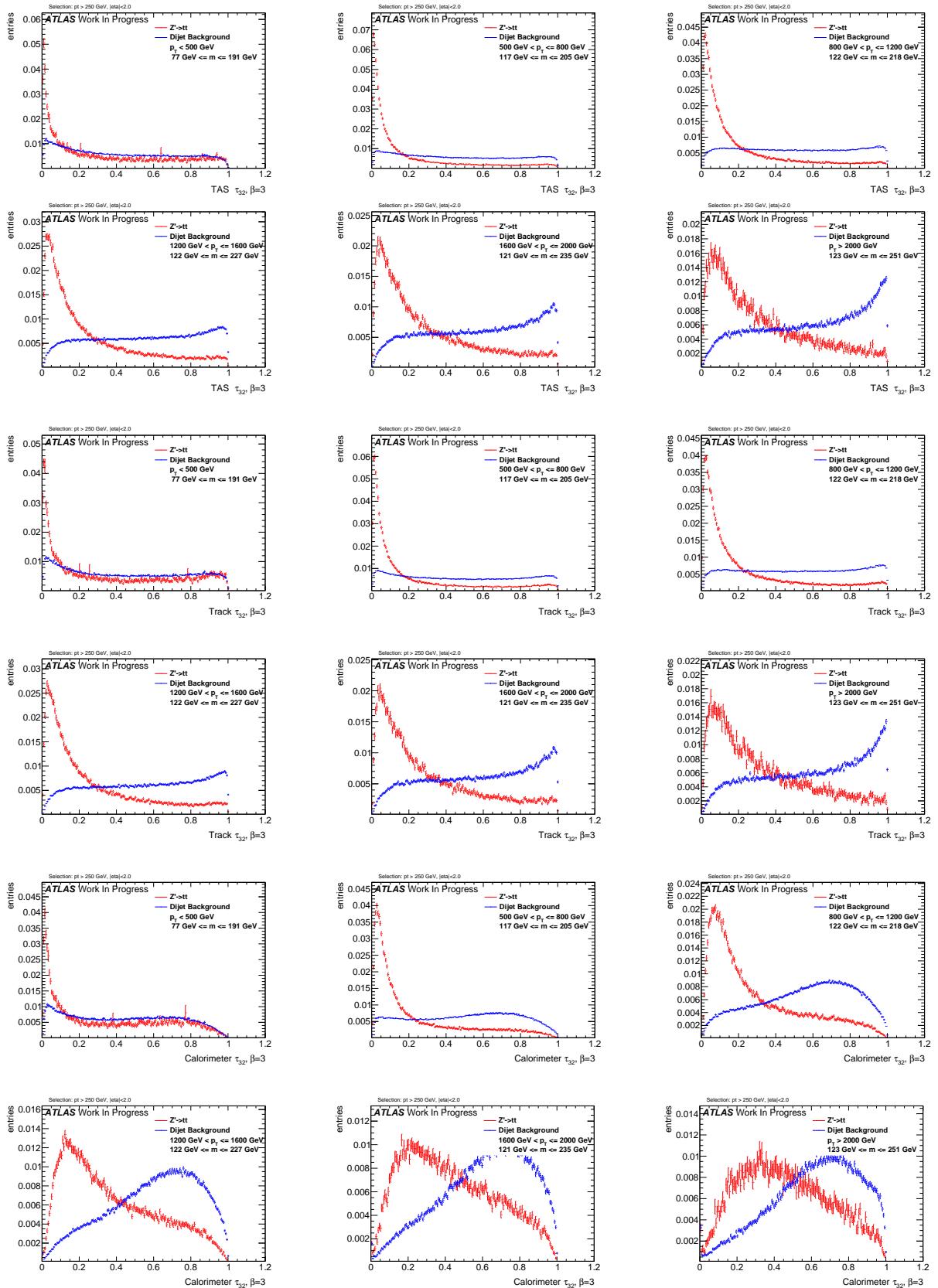


Figure 310: Distributions for Top tagging using τ_{32} ($\beta = 1$) using TAS, tracks and calorimeter clusters top down.

Figure 311: Distributions for Top tagging using τ_{32} ($\beta = 1.7$) using TAS, tracks and calorimeter clusters top down.

Figure 312: Distributions for Top tagging using τ_{32} ($\beta = 2$) using TAS, tracks and calorimeter clusters top down.

Figure 313: Distributions for Top tagging using τ_{32} ($\beta = 3$) using TAS, tracks and calorimeter clusters top down.

916 J W Distributions

917 $\beta = 0.5$

918 $\beta = 1$

919 $\beta = 1.7$

920 $\beta = 2$

921 $\beta = 3$

922 K Higgs Distributions

923 $\beta = 0.5$

924 $\beta = 1$

925 $\beta = 1.7$

926 $\beta = 2$

927 $\beta = 3$

928 L Top Distributions

929 $\beta = 1$

930 $\beta = 1.7$

931 $\beta = 2$

932 $\beta = 3$

933 References

- 934 [1] Particle Data Group Collaboration, K. Olive *et al.*, *Review of Particle Physics*, Chinese Physics C 38
935 (2014) 090001.
- 936 [2] Paul Langacker, *The Physics of Heavy Z Gauge Bosons*. In: Rev. Mod. Phys. 81 (2009), pp. 1199–1228.
937 doi: 10.1103/RevModPhys.81.1199. arXiv: 0801.1345 [hep-ph].
- 938 [3] Martin Schmaltz and Christian Spethmann, *Two Simple W' Models for the Early LHC*. In: JHEP 07
939 (2011), p. 046. doi: 10.1007/JHEP07(2011)046. arXiv: 1011. 5918 [hep-ph].
- 940 [4] Georges Aad *et al.* *Combination of searches for WW, WZ, and ZZ resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*. In: Phys. Lett. B755 (2016), pp. 285–305. doi:
941 10.1016/j.physletb.2016.02.015. arXiv: 1512.05099 [hep-ex].

- [5] The CMS Collaboration, *Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV*, CERN-PH-EP/2013-037 2014/09/03 arXiv:1405.3447v2 [hep-ex].
- [6] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016). <http://pdg.lbl.gov/2015/reviews/rpp2015-rev-wprime-searches.pdf>
- [7] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016). <http://pdg.lbl.gov/2015/reviews/rpp2015-rev-zprime-searches.pdf>
- [8] L. Randall and R. Sundrum, *A Large Mass Hierarchy from a Small Extra Dimension*, Phys. Rev. Lett. 83 (1999) 3370, arXiv: hep-ph/9905221.
- [9] H. Davoudiasl, J.L. Hewett and T.G. Rizzo, *Warped Phenomenology*, Stanford Linear Accelerator Center Stanford CA 94309, USA, arXiv:hep-ph/9909255v1 6 Sep 1999.
- [10] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1 & 2*, Tech. Rep. ATLAS TDR 14, CERN/LHCC 99-14, Geneva, 1999.
- [11] L. Evans and P. Bryant, *LHC Machine*, JINST 3 (2008) S08001.
- [12] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016). <http://pdg.lbl.gov/2010/reviews/rpp2010-rev-qcd.pdf>
- [13] The CMS Collaboration, *Measurement of the inclusive 3-jet production differential cross section in proton-proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range*, CERN-PH-EP/2013-037 2015/05/04 arXiv:hep-ph/14121633.
- [14] The ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C (2013) 73: 2518. doi:10.1140/epjc/s10052-013-2518-3.
- [15] W. Bartel *et al.* [JADE Collaboration], *Observation of Planar Three Jet Events in $e + e^-$ Annihilation and Evidence for Gluon Bremsstrahlung*, Phys. Lett. B 91, 142 (1980).
- [16] Ahmed Ali (DESY), Gustav Kramer (Univ. Hamburg), *Jets and QCD: A Historical Review of the Discovery of the Quark and Gluon Jets and its Impact on QCD*, arXiv:1012.2288v2 [hep-ph].
- [17] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, *The anti- k_t jet clustering algorithm*, arXiv:0802.1189v2 [hep-ph].
- [18] F. Halzen, A.D. Martin and D.M. Scott, Phys. Rev. D25 (1982) 754, Phys. Lett. B112 (1982) 160.
- [19] Torbjorn Sjostrand, *Monte Carlo Generators*, hep-ph/0611247, CERN-LCGAPP-2006-06.
- [20] W. Lampl *et al.*, *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002.
- [21] The ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector*, CERN-PH-EP-2012-191 arXiv:1210.6210v1 [hep-ex].
- [22] GEANT4 Collaboration (S. Agostinelli *et al.*), *GEANT4: A Simulation toolkit*, Nucl.Instrum.Meth. A506 (2003) 250-303 DOI: 10.1016/S0168-9002(03)01368-8 SLAC-PUB-9350, FERMILAB-PUB-03-339.
- [23] Ariel Schwartzman, *Jet energy calibration at the LHC*, SLAC National Accelerator Laboratory, 2575 Sand Hill Road Menlo Park, California 94025, USA, arXiv:1509.05459v1 [hep-ex] 17 Sep 2015.

- 981 [24] The ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*,
982 ATLAS-CONF-2016-035 19 July 2016.
- 983 [25] The ATLAS Collaboration, *Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at*
984 $\sqrt{s} = 13 \text{ TeV}$, ATLAS-CONF-2016-039 30th July 2016.
- 985 [26] The ATLAS Collaboration, *Search for high-mass diboson resonances with boson-tagged jets in*
986 *proton-proton collisions at $\sqrt{s} = 8 \text{ TeV}$ with the ATLAS detector*, arXiv:1506.00962v3 [hep-ex].
- 987 [27] The CMS Collaboration, *Combination of searches for WW, WZ, ZZ, WH, and ZH resonances at $\sqrt{s}=$*
988 *8 and 13 TeV*, CMS PAS B2G-16-007.
- 989 [28] The ATLAS Collaboration, *Search for heavy particles decaying to pairs of highly-boosted top quarks*
990 *using lepton-plus-jets events in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*,
991 ATLAS-CONF-2016-014.
- 992 [29] The ATLAS Collaboration, *Search for resonances with boson-tagged jets in 15.5 fb^{-1} of p-p*
993 *collisions at $\sqrt{s} = 13 \text{ TeV}$ collected with the ATLAS detector*, ATLAS-CONF-2016-055.
- 994 [30] The ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}\bar{b}\bar{b}$ final state using*
995 *proton proton collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, ATLAS-CONF-2016-049.
- 996 [31] The ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in*
997 *proton-proton collisions at $\sqrt{s} = 7 \text{ TeV}$ using the ATLAS detector*, CERN-PH-EP-2013-069
998 arXiv:1306.4945 [hep-ex].
- 999 [32] Oleg Zenin, on behalf of the ATLAS Collaboration, *Soft QCD and underlying event measurements*
1000 *at ATLAS*, Nuclear and Particle Physics Proceedings Volumes 273275, April-June 2016, Pages 2053-
1001 2058 37th International Conference on High Energy Physics (ICHEP).
- 1002 [33] Page title: JetSubstructureECFA2014 AtlasPublic TWiki Website name: Twiki.cern.ch
1003 URL:<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetSubstructureECFA2014> Access date: 27th
1004 October 2016.
- 1005 [34] E. Lehmann and G. Casella, *Theory of Point Estimation*, Springer Verlag, 1998, isbn: 0387985026.
- 1006 [35] Andrew J. Larkoski, Ian Moult, and Duff Neill. *Power Counting to Better Jet Observables*. *JHEP*,
1007 12:009, 2014.
- 1008 [36] Andrew J. Larkoski, Ian Moult, and Duff Neill. *Analytic Boosted Boson Discrimination*. *JHEP*,
1009 05:117, 2016.
- 1010 [37] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. *Energy Correlation Functions for Jet Sub-*
1011 *structure*. *JHEP*, 06:108, 2013.
- 1012 [38] Jesse Thaler and Ken Van Tilburg. *Identifying Boosted Objects with N-subjettiness*. *JHEP*, 03:015,
1013 2011.
- 1014 [39] Georges Aad et al. *Identification of boosted, hadronically decaying W bosons and comparisons with*
1015 *ATLAS data taken at $\sqrt{s} = 8 \text{ TeV}$* . *Eur. Phys. J.*, C76(3):154, 2016.

1016 List of contributions

1017

Auxiliary material

1018 In an ATLAS paper, auxiliary plots and tables that are supposed to be made public should be collected in
1020 an appendix that has the title ‘Auxiliary material’. This appendix should be printed after the Bibliography.
1021 At the end of the paper approval procedure, this information can be split into a separate document – see
1022 `atlas-auxmat.tex`.

1023 In an ATLAS note, use the appendices to include all the technical details of your work that are relevant
1024 for the ATLAS Collaboration only (e.g. dataset details, software release used). This information should
1025 be printed after the Bibliography.