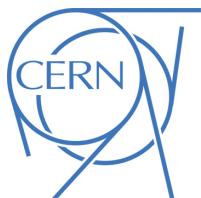




ATLAS Note

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

13th December 2016

7

This note presents the details of the Monte-Carlo studies on the subjet-assisted observables
8 for groomed large-radius jet. In particular the observables for the Energy Correlation Func-
9 tions and n-Subjettiness variables used by the ATLAS collaboration, C_2 , D_2 , τ_{21} and τ_{32} are
10 discussed using subjet-assisted tracks; the mass observable constructed with this technique,
11 m^{TAS} , is presented and discussed with a modified four-momentum prescription. In all the
12 variables studied, large improvement have been found using this novel techniques, the first
13 ones evaluating in terms of QCD event rejection in W/Z boson, top quark and Higgs boson
14 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	3
18	2 ATLAS detector	4
19	3 Monte Carlo Samples	5
20	4 Object Definition	5
21	4.1 Large-radius jet mass definitions	5
22	4.1.1 Substructure: Grooming Techniques	6
23	4.1.2 Trimming	6
24	4.1.3 Calorimeter Mass	7
25	4.1.4 Track Mass	7
26	4.1.5 Performance Figure of Merit (FoM)	8
27	4.1.6 Gaussian Fit	9
28	4.1.7 InterQuantile-Range	10
29	4.2 Energy Correlation Functions	10
30	4.3 n-Subjettiness	12
31	4.4 Receiver Operator Characteristics	13
32	5 Track-assisted subjet mass	13
33	5.1 Track-Assisted Mass (m^{TA})	13
34	5.2 Advantages and Limitation of m^{TA}	14
35	5.2.1 Performance in $W \rightarrow q'\bar{q}$ Decays	15
36	5.2.2 Performance in $t \rightarrow q'\bar{q}b$ Decays	15
37	5.2.3 Performance in $h \rightarrow b\bar{b}$ Decays	16
38	5.3 The Track-Assisted Sub-jet Mass (m^{TAS})	16
39	5.4 Observable Definition: Inputs	18
40	5.4.1 Tracks	18
41	5.4.2 Sub-jets	18
42	5.5 Observable Definition: Procedure	18
43	5.6 Performance in $W \rightarrow q'\bar{q}$ Decays	20
44	5.7 Performance in $t \rightarrow q'\bar{q}b$ Decays	20
45	5.8 Performance in $h \rightarrow b\bar{b}$ Decays	20
46	5.9 Performance in QCD Multijet Events	21
47	5.10 Performance in Massive $\tilde{W} \rightarrow q'\bar{q}$ Decays with $m_{\tilde{W}} = m_t$	22
48	5.11 Other Stability Quantifiers	22
49	5.12 Sub-jet Calibration	24
50	5.13 Preliminary Studies on Sub-jet Calibration	24
51	5.13.1 Perfect Calibration	24
52	5.13.2 Simple Sub-jet Calibration	24
53	5.14 Limitation of m^{TAS}	25
54	5.15 Alternative Observable Definitions	27
55	6 Combining the mass observables	28
56	6.1 Combination $m^{TA} - m^{calo}$	29
57	6.2 Combination $m^{TAS} - m^{calo}$	29

58	6.2.1	Procedure	30
59	6.3	Performance in $W \rightarrow q'\bar{q}$ Decays	30
60	6.4	Performance in $t \rightarrow q'\bar{q}b$ Decays	31
61	6.5	Performance in $h \rightarrow b\bar{b}$ Decays	31
62	7	Energy Correlation Functions and n-Subjettiness	31
63	7.1	Sub-jet track assistance	31
64	7.1.1	Event weighting and Mass-Cut	33
65	7.2	Track Selection	34
66	7.3	Performance with default β	35
67	7.3.1	Performance for W boson tagging	36
68	7.3.2	Un-assisted tracks and TAS at very high p_T	36
69	7.3.3	Correlation with p_T	38
70	7.4	Optimisation of β	38
71	8	Conclusions & Outlook	38
72	8.1	Jet mass observables	38
73	8.1.1	Outlook	40
74	8.2	Energy Correlation Functions and n-Subjettiness	40
75	Appendix		41
76	A	m^{TAS} distributions, boosted W/Z	50
77	B	m^{TAS} distributions, boosted tops	62
78	C	m^{TAS} distributions, boosted higgs	73
79	D	m_{TAS}^{comb} response distributions, boosted W/Z	84
80	E	m_{TAS}^{comb} response distributions, boosted tops	87
81	F	m_{TAS}^{comb} response distributions, Higgs	90
82	Auxiliary material		97

83 **1 Introduction**

84 Jets are collimated streams of particles resulting from quarks and gluons fragmentation and hadronization.
 85 The distribution of energy inside a jet contains information about the initiating particle. When a massive
 86 particle such as a top quark, Higgs boson or W/Z bosons is produced with significant Lorentz boost and
 87 decays into quarks, the entire hadronic decay may be captured inside a single jet. The mass of such jets
 88 (jet mass) is one of the most powerful tools for distinguishing massive particle decays from the continuum
 89 multijet background; the Energy Correlation Functions and n-Subjettiness C2, D2, τ_{21} and τ_{32} provide an
 90 ad-hoc tool pupusely developed for the multijet background and constitue a fundamental part of many for

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)

91 boson taggers. This note documents the so-called subjet-assisted techniques with the ATLAS detector
92 *insref*. The track-assisted subjet mass m^{TAS} definition is presented and confronted with the standard
93 development in ATLAS, m^{comb} and m^{TA} . Energy Correlation Functions and n-Subjettiness with the
94 modified subjet-assisted technique is presented and confronted with the standard one in ATLAS. The note
95 ends with conclusions and future outlook in *insref*.

96 2 ATLAS detector

97 ATLAS (A Toroidal ApparatuS) is a multi-purpose particle detector with nearly 4π coverage in solid angle.
98 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 ($|\eta| < 1.7$)
100 and two end-cap copper/liquid-argon sections extend to higher pseudo-rapidity ($1.5 < |\eta| < 3.2$). Finally,
101 the forward region ($3.1 < |\eta| < 4.9$) is covered by a liquid-argon calorimeter with Cu (W), absorber in the
102 electromagnetic (hadronic) section. Inside the calorimeters there is a 2 T solenoid that surrounds an inner
103 tracking detector which measures charged particle trajectories covering a pseudo-rapidity range $|\eta| < 2.5$
104 with pixel and silicon micro-strip detectors (SCT) and additionally which covers the region $|\eta| < 2.0$ with
105 a straw-tube transition radiation tracker (TRT). Outside the calorimeter there is a muon spectrometer: a
106 system of detectors for triggering up to $|\eta| < 2.4$ and precision tracking chambers up to $|\eta| < 2.7$ inside a
107 magnetic field supplied by three large superconducting toroid magnets.

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

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
$W' \rightarrow \tilde{W}\tilde{W}$ 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.

109 3 Monte Carlo Samples

110 *refrseMT* The samples used are divided into two main groups: SM background and beyond SM signal.
 111 The SM background includes the QCD multijet samples, produced with a falling p_T spectrum. The
 112 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
 113 ($t \rightarrow W(\rightarrow q\bar{q}')b$)) and RS-Graviton $\rightarrow hh \rightarrow b\bar{b}b\bar{b}$, i.e. final states have only jets in all the samples.
 114 The details of the samples are given in Table 1; the masses considered span from 0.5 to 5 TeV to improve
 115 and diversify the kinematic space covered.

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

123 4 Object Definition

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

127 4.1 Large-radius jet mass definitions

128 Large-radius jet, or arge- R jets are jets constructed with a radius parameter of the reclustering algorithm
 129 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
 130 (the area of C/A is $\sim 20\%$ smaller than anti- k_t).

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

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

136 **4.1.1 Substructure: Grooming Techniques**

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

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

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

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

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

151 **4.1.2 Trimming**

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

- 157 • k_t algorithm (but of course other choices are also possible) is used to create sub-jets with a smaller
 158 radius R_{sub} , aiming at separating the soft radiation from the hard one in different sub-jets. Typical
 159 choices are 0.2 and 0.3 (0.2 is used as standard);
- 160 • 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
 161 below a certain value, the sub-jet is removed. Standard choice is $f_{cut} = \frac{p_T}{p_T^{jet}} = 0.05$;
- 162 • the sub-jets which survived this procedure are the only one which compose the trimmed jet.

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

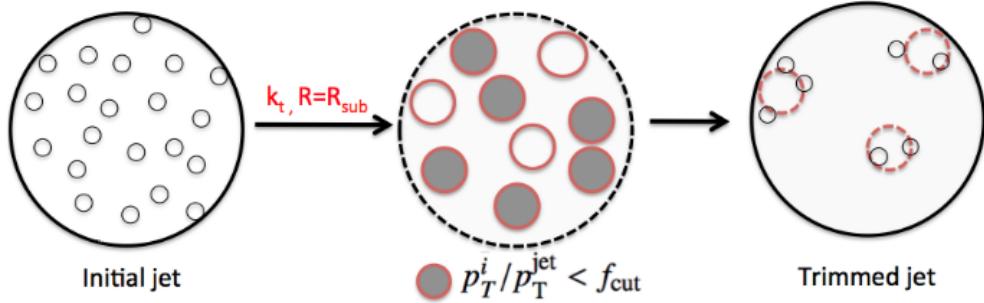


Figure 1: Schematic of the trimming algorithm.

165 **4.1.3 Calorimeter Mass**

166 Once the collection of constituents from the large- R jet is groomed, it is possible to use them for the
 167 measure of physical related properties such as the jet mass, since the possible sources of soft radiation
 168 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:

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

169 where J labels the Large- R jet.

170 **4.1.4 Track Mass**

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

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

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

- 181 • $p_T^{track} > 400$ MeV;
- 182 • $|\eta| < 2.5$;
- 183 • Maximum 7 hits in the Pixel and STC sub-detectors;

- 184 • Maximum 1 Pixel hole;
- 185 • Maximum 2 silicon holes;
- 186 • Less than 3 shared modules;
- 187 • Maximum 2 mm of displacement along beam axis (z_0) from the primary vertex;
- 188 • Maximum 2.5 mm of distance in x-y plane from the primary vertex and point of closest approach
189 (d_0).

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

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

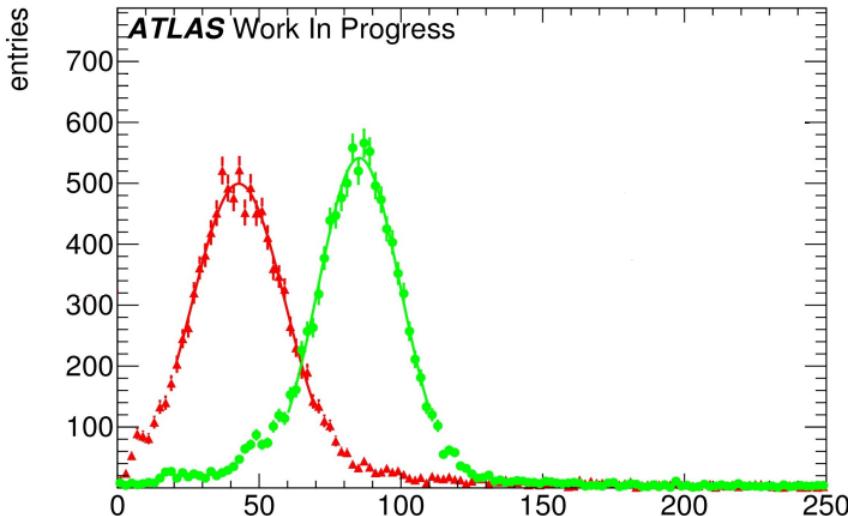


Figure 2: Mass distribution boosted W/Z : in green the m^{calo} and in red the m^{track} .

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

198 4.1.5 Performance Figure of Merit (FoM)

- 199 Since we already introduced the calorimeter and track mass, a concrete, quantitative feature has to be
200 defined in order to understand which observable is “better”, in the sense that we would prefer one or the
201 other according to this criterion. This is often referred to as *Figure of Merit* or simply FoM.
202 There are few ways to look at the FoM: one can e.g. naively think about the mean of the mass distribution,
203 since closer values of the mean to the e.g. W or Z mass (if we are speaking about W/Z decays), indicate a
204 more correct mass reconstruction. However, this does not take into account the width of this distribution,

as a large width spoils the reconstruction in terms of percentage of jets misreconstructed. Moreover, the mean is not as important since it can be rescaled to the desired value in a calibration procedure.

4.1.6 Gaussian Fit

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

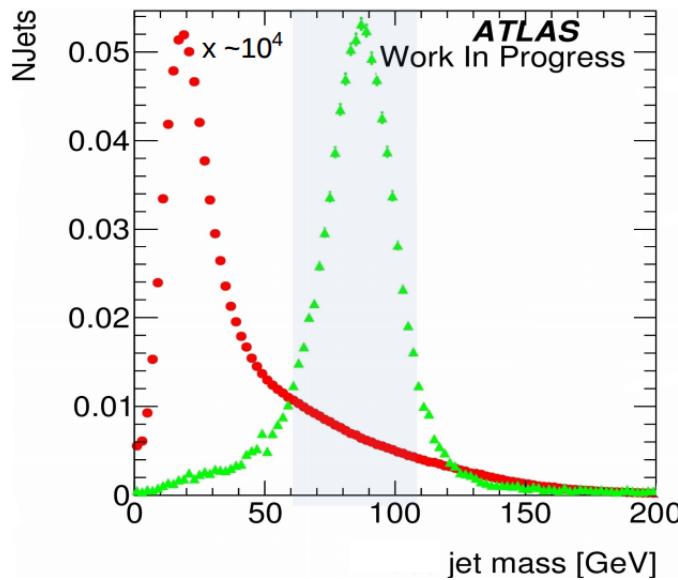


Figure 3: Mass distributions: in red the QCD dijet background rescaled, in green the W/Z from the W' sample. Highlighted the width of the W/Z distribution.

The width σ of the distribution, which can be obtained from a fit to the Gaussian core, is already a valid FoM, which has an underlying physical feature. Moreover, in order to be independent from the mean of the distribution, the width can be divided by the mean itself. This was in fact the FoM which was used 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.

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.

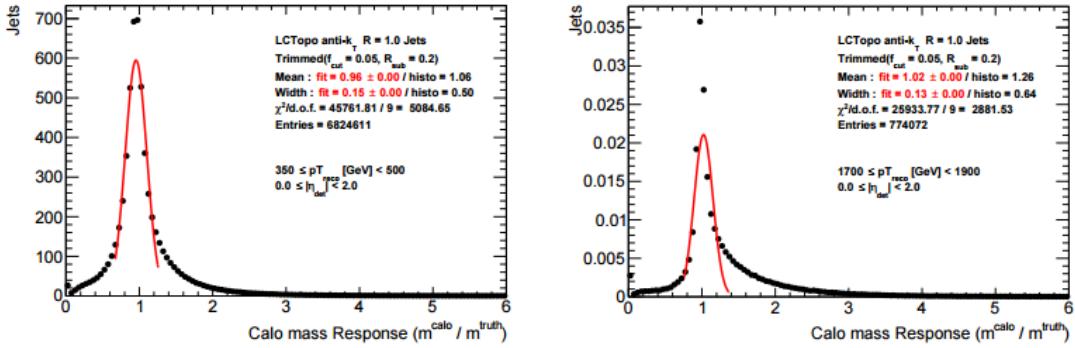


Figure 4: Mass Response distributions for the QCD multijet for various p_T ranges: on the right the failure of the Gaussian fit shows the limitation of this approach to evaluate the Figure of Merit. On the plot the fit parameters and transverse momentum ranges.

224 4.1.7 InterQuantile-Range

225 Another way to look at the mass FoM is half of the 68% of the InterQuantile range (IQnR) (here defined
 226 such as it corresponds to a sigma of a “perfect” Gaussian distribution: $q84\% - q16\%$ where $q84\%$ is the
 227 84th percentile and $q16\%$ is the 16th, not to be confused with the InterQuartile Range (IQR) which is the
 228 $q75\% - q25\%$ and does not correspond to the sigma) divided by the Median ($\frac{1}{2} \times 68\% \text{ IQnR}/\text{median}$). It
 229 provides stability and high sensitivity to left-hand-side and right-hand-side tails.

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

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

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

234 In Figure 5 a mass response for a single range of transverse momentum is shown, for the calorimeter mass.
 235 On the plot the contours of a standard deviation and of $q16\%$ and $q84\%$ are drawn with dashed and solid
 236 lines, respectively, showing the difference induced by the tail. This sort of plot is the key when looking
 237 quantitatively to the observable performance and can be found in the Appendix for each of the process
 238 studied in every p_T range considered. In this chapter will be shown, however, the quantity which describes
 239 this FOM, the IQnR, as a function of p_T , in order to get an understanding of the behavior in the entire
 240 spectrum and assure the exclusion of local sub-optimalities.

241 4.2 Energy Correlation Functions

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

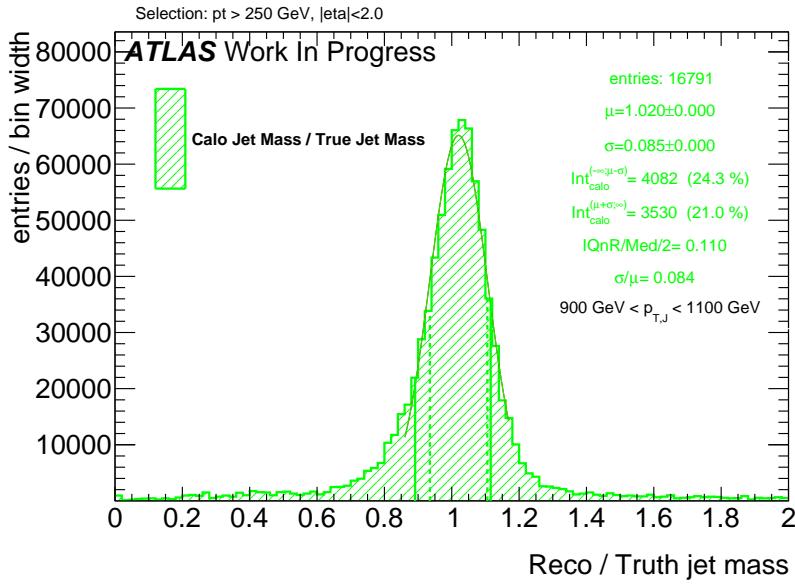


Figure 5: Calorimeter mass response plot for boosted W/Z . One the plot, right, are shown: the number of entries, the mean and the width of the fit to the Gaussian core, the integral from 0 to $\mu - \sigma$ and the one from $\mu + \sigma$ to $+\infty$, the values $\frac{1}{2} \times 68\%$ IQnR/median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

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

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

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

257 The ECF(N) variable tends to very small values for collinear or soft configurations of N constituents and
 258 is defined to be zero for jets with less than N constituents. For ECF(2), only pairs of constituents that
 259 are angular separated but not soft result in sum terms that are non-negligible, which directly leads to the
 260 picture of two hard substructures inside the jet. A similar conclusion can be made for ECF(3) and three

261 hard substructures. Resulting from this, a jet with N or more hard substructures features a high ECFN
 262 value while a jet with fewer than N substructures has a lower ECF(N) value. Consequently, one can define
 263 ratios of Energy Correlation Functions. Two of them, called C2 and D2 are found to be very powerful to
 264 distinguish between one- and two-prong like jets, see e.g. Reference [bib:power_counting].

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

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

272 4.3 n-Subjettiness

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

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

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

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

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

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

287 Small values of τ_N correspond to a jet with all constituents more or less aligned or near to the given
 288 N sub-jet axes, hence the jet is compatible with the assumption to be composed of N or fewer sub-jets.
 289 A higher value in contrast indicates a consistency with more than N sub-jets as a non negligible part is
 290 located apart of the N sub-jet axes. Consequently, W/Z or Higgs boson jets are likely to feature a small

291 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.
 292 While τ_1 and τ_2 alone provide only slight separation, the ratio

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

293 is an effective discrimination variable.

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

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

296

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

299 **4.4 Receiver Operator Characteristics**

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

311 **5 Track-assisted subjet mass**

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

314 **5.1 Track-Assisted Mass (m^{TA})**

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

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

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

321 One could then think about creating a variable which exploits the advantages of both and minimizes the
 322 disadvantages. As seen, the track mass is missing the neutral component, i.e. each measurement is missing
 323 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
 324 the *track-assisted mass* (m^{TA}):

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

325 It can be intuitively understood as follows: the term m^{track} has the superior angular resolution, but misses
 326 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*.

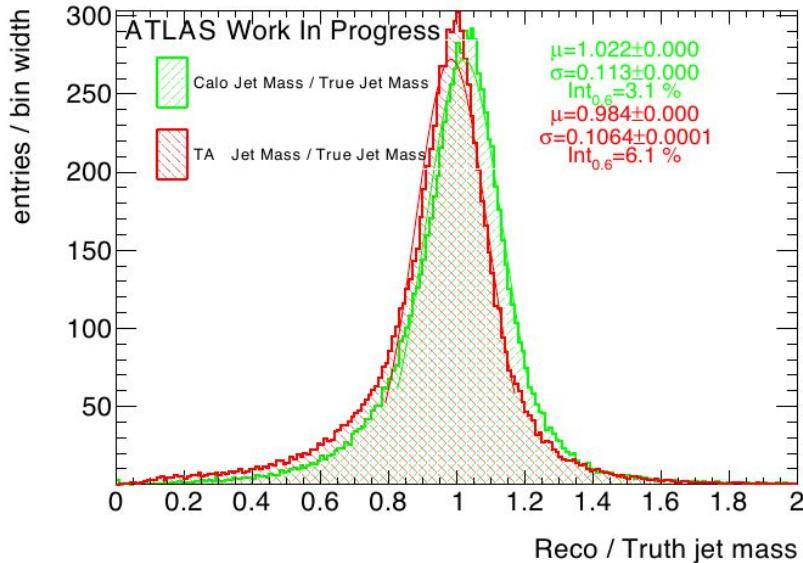


Figure 6: Track-assisted mass response plot for boosted W/Z : in green the calorimeter mass, in red the track-assisted mass. On the right are shown properties of the fit to the Gaussian core; it can be seen that the width of the m^{TA} distribution is smaller, and the mean is slightly below the calorimeter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

327

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

330 5.2 Advantages and Limitation of m^{TA}

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

333 Another big advantage which supports the use of the track-assisted mass is the relatively small uncertainties:
 334 in Figure 7 the comparison of m^{calo} (left) and m^{TA} (right) fractional uncertainties on the JMS, shows how
 335 the tracking uncertainties are much smaller because of the ratio $m^{\text{track}}/p_T^{\text{track}}$. On the right plot the black
 336 line indicates the JMS fractional uncertainty for the m^{calo} , and is always above the m^{TA} . Of course this
 337 introduces another argument in the development of new techniques, which is to look for a good balance

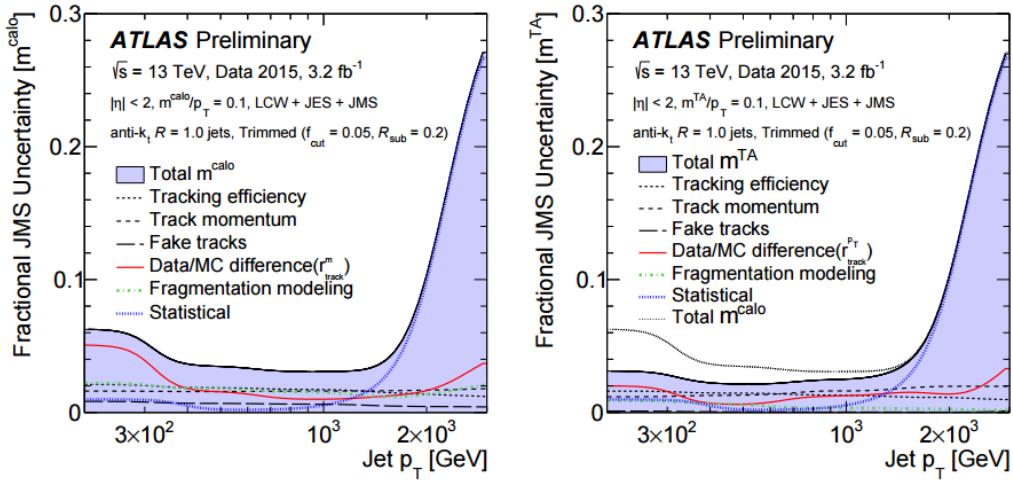


Figure 7: Comparison of the uncertainties for m^{calo} , on the left, and m^{TA} , on the right the rise on the high jet p_{T} is due to statistics. From the [art35].

338 between performance and small uncertainties: a perfect observable in terms of behavior which has very
 339 big uncertainties is not really useful.

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

345 5.2.1 Performance in $W \rightarrow q'\bar{q}$ Decays

346 The performance in all the bins of p_{T} can be studied looking at Figure 9; these plots have as horizontal
 347 axis the transverse momentum and as vertical one the value of the $\frac{1}{2} \times 68\%$ IQnR/median calculated
 348 from the correspondingly response. For W/Z jets, there is a crossing point around $p_{\text{T}} \sim 1$ TeV, which can
 349 be understood as the point in which the two sub-jet present start merging (sub-jet multiplicity shown in
 350 Figure 42 in Appendix).

351 5.2.2 Performance in $t \rightarrow q'\bar{q}b$ Decays

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

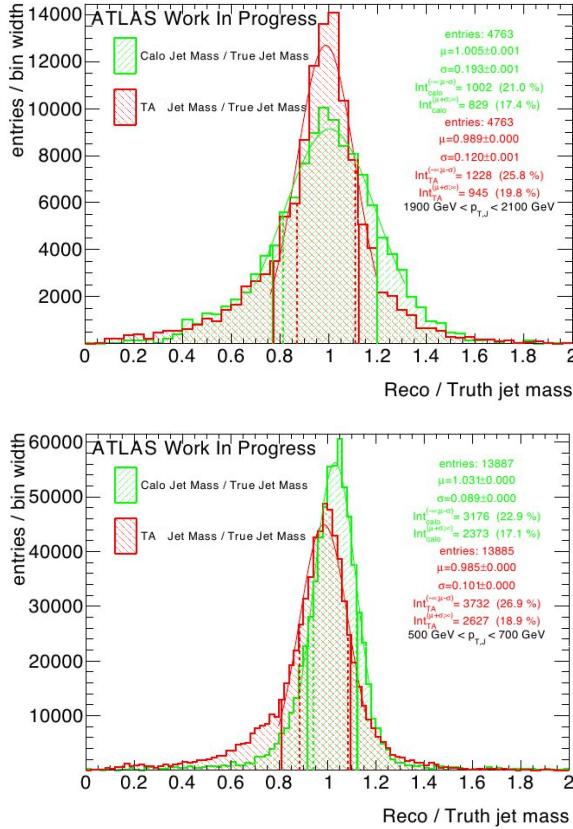


Figure 8: Mass response plots for selected ranges of p_T : on the bottom, a “low” range, $500 \text{ GeV} < p_T < 700 \text{ GeV}$, on the top an high p_T , $1900 \text{ GeV} < p_T < 2100 \text{ GeV}$. A difference in performance can be clearly seen.

358 5.2.3 Performance in $h \rightarrow b\bar{b}$ Decays

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

363 5.3 The Track-Assisted Sub-jet Mass (m^{TAS})

364 In this section the main outcome of the work of this thesis is presented: the *track-assisted sub-jet mass*
 365 (m^{TAS}). The main idea takes inspiration from the track-assisted mass: if one can use the tracks to exploit
 366 the better angular resolution and correct the missing neutral component jet-by-jet, there is an additional
 367 information that can be used. The neutral fraction, in fact, varies stochastically not only per-jet basis, but
 368 even per-sub-jet basis, since each sub-jet is originated from a different quark. Correcting the missed neutral
 369 component per-sub-jet, it should perform better already at an intuitive level, as it accesses information
 370 from the jet substructure. There are few question in the definition of this mass observable, whose answers
 371 are in the next section:

- 372 • Regarding the inputs:

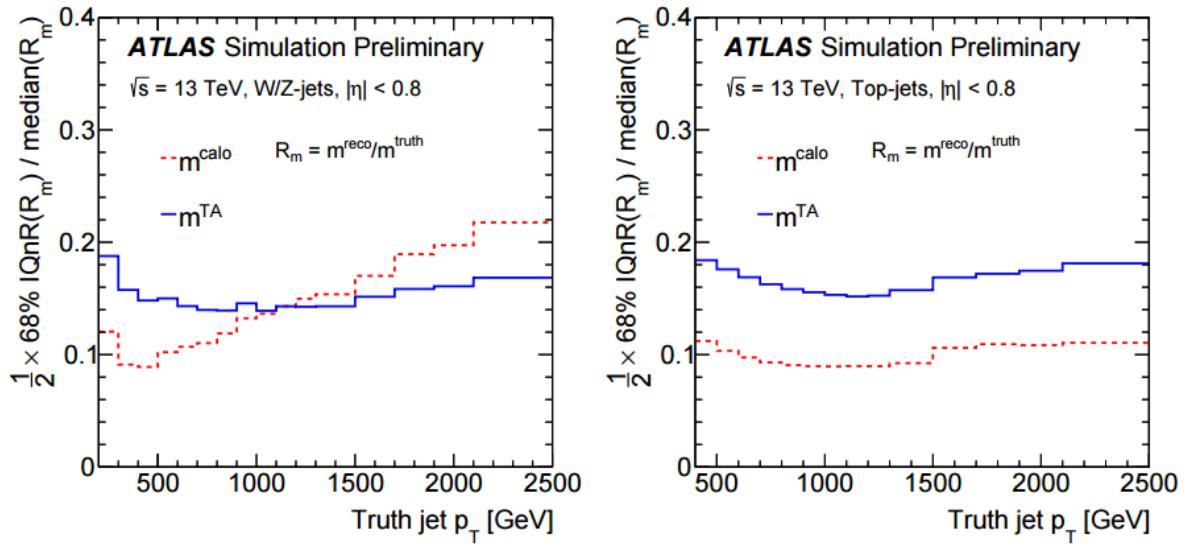


Figure 9: The comparison between the performance of m^{calo} and m^{TA} for W/Z jets (left) and top jets (right); on the x-axis the transverse momentum and on the y-axes the $\frac{1}{2} \times 68\%$ IQnR/median of the mass distribution, from [art35]. A better observable has lower values on the y-axis.

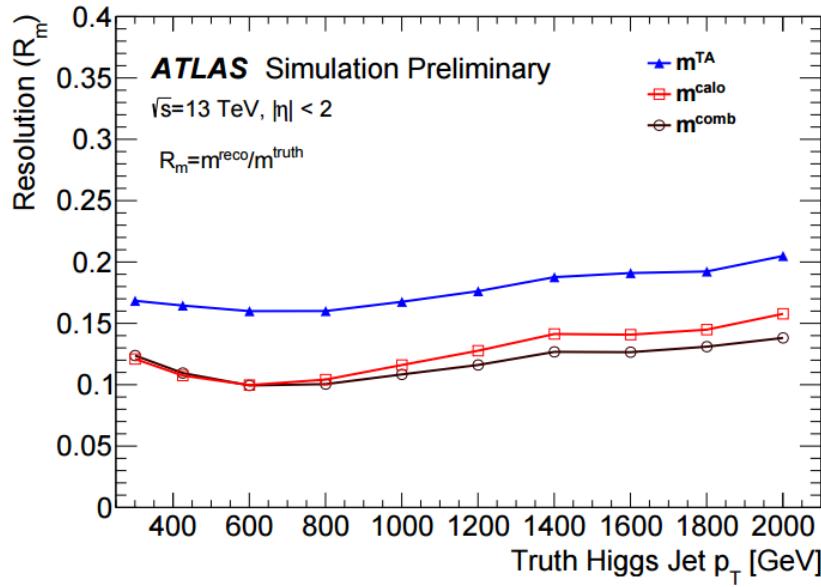


Figure 10: Performance of the m^{TA} with the boosted Higgs sample; the m^{TA} is the blue line, the m^{comb} will be described later in this chapter. From [art39]. The FoM here is the resolution of the Response.

- ³⁷³ – How to select the set of tracks to be used?
- ³⁷⁴ – Which kind of sub-jet should be used?
- ³⁷⁵ • Regarding the procedure
 - ³⁷⁶ – How to associate the tracks to a sub-jet?

-
- 377 – How to correct for the missed neutrals on a sub-jet basis?
 378 – How to add everything back together?

379 Those details are given in the next subsection.

380 **5.4 Observable Definition: Inputs**

381 There are two inputs to the m^{TAS} : the tracks and the sub-jets. The definition of the standard inputs are
 382 give here; alternative approaches are given in subsection 5.15.

383 **5.4.1 Tracks**

384 Only the tracks that satisfy the quality criteria and primary vertex association, described in the previous
 385 section 4.1.4, are used. The tracks taken additionally are required to be ghost associated to the sub-jets
 386 of the groomed jet; namely only the sub-jets which survived the trimming procedure and are described in
 387 the next subsection. Ghost association provides a one-to-one correspondence to the sub-jets set, and was
 388 therefore chosen and preferred to other kind of assignments.

389 **5.4.2 Sub-jets**

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

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

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

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

397 **5.5 Observable Definition: Procedure**

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

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

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

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

407 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
 408 level, i.e.

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

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

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

411 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
 412 the large- R jet,

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

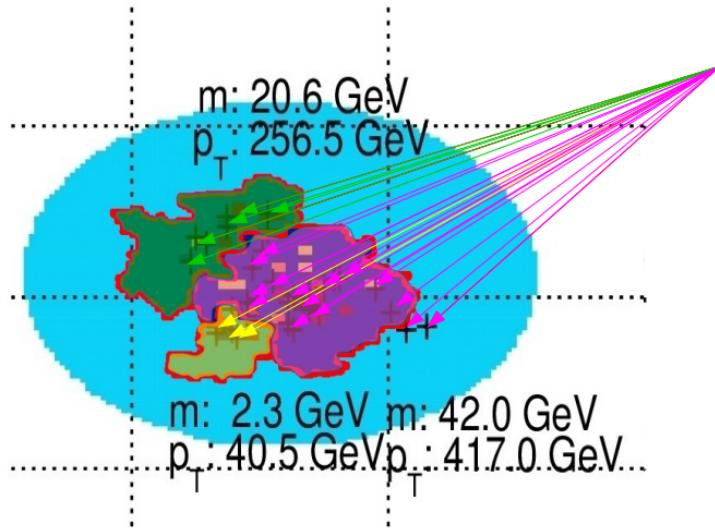


Figure 11: Pictorial event display showing the $\eta \phi$ region of a large- R jet, (in blue the catchment area of the anti- k_t) showing the different k_t sub-jets: they are highlighted in green, fuchsia and yellow. The associated track-jets (here as arrows pointing the calorimeter area) are colored 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.

413 An important remark is that, in the case of a large- R jet with only one sub-jet, the m^{TAS} has exactly
 414 the same definition of the m^{TA} . This implies, since the angular separation of the decay product scales
 415 inversely with p_T , that the performance should approach the one of the m^{TA} in the extreme kinematic

416 regime. However, the space for improvement is precisely in the low-middle p_T regime, as seen in the m^{TA}
 417 section.

418 5.6 Performance in $W \rightarrow q' \bar{q}$ Decays

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

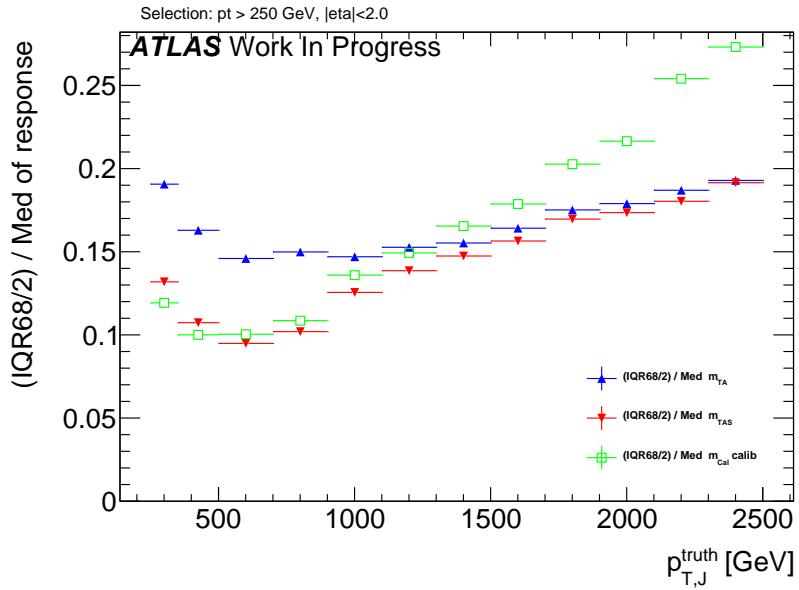


Figure 12: Performance of the m^{TAS} versus the m^{calo} and m^{TA} for the boosted W/Z sample.

427 5.7 Performance in $t \rightarrow q' \bar{q} b$ Decays

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

432 5.8 Performance in $h \rightarrow b\bar{b}$ Decays

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

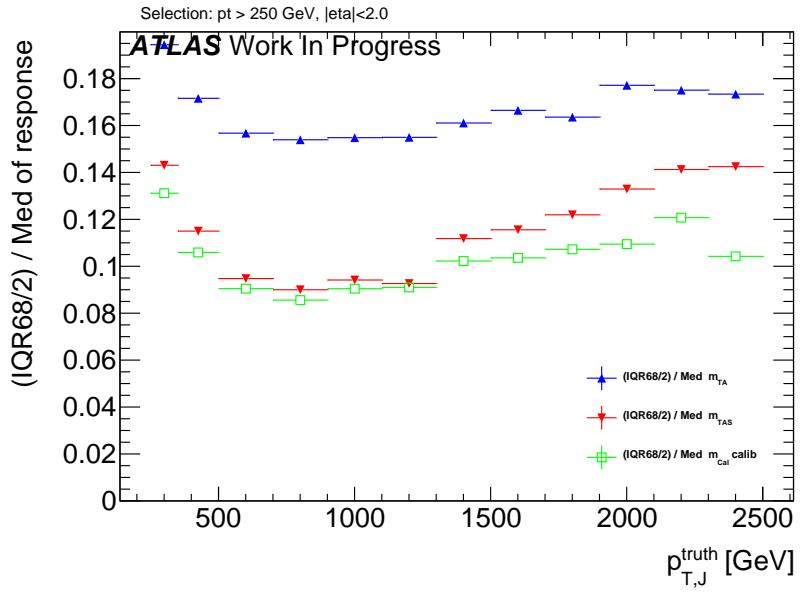


Figure 13: Performance of the m^{TAS} versus the m^{calo} and m^{TA} for the boosted top sample.

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.

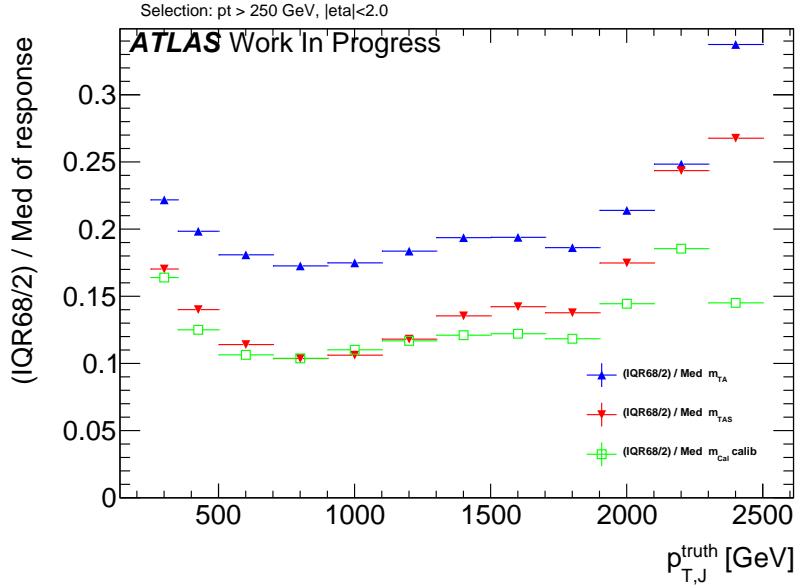


Figure 14: Performance of the m^{TAS} versus the m^{calo} and m^{TA} for the boosted Higgs sample.

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

440 transverse momenta. Again the m^{TAS} follows this similarity showing no crossing point and an optimal
 441 overall behavior, both with respect to calorimeter- and track-assisted-based mass definition. On Figure
 442 15.

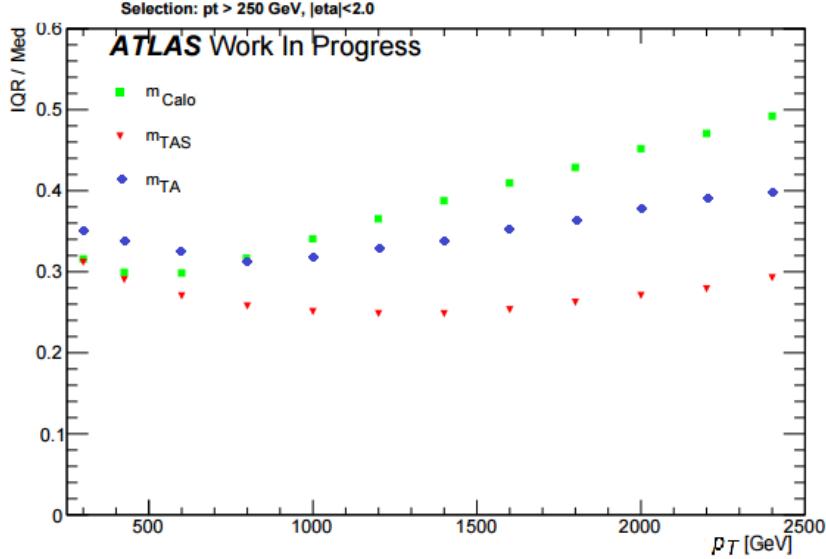


Figure 15: Performance of the m^{TAS} versus the m^{calo} and m^T for the QCD multijet. Here shown IQR/Med not $\frac{1}{2} \times 68\%$ IQnR/median.

443 5.10 Performance in Massive $\tilde{W} \rightarrow q'\bar{q}$ Decays with $m_{\tilde{W}} = m_t$

444 The massive W sample is a special sample which was used to understand the behavior of the boosted
 445 tops, whether its worse resolution was coming from the higher mass of the top quark or from the more
 446 complex decay topology (three-pronged instead of two-pronged decay and b-quark presence). The sample
 447 is almost identical to the boosted W/Z one ($W' \rightarrow WZ$) but in this case the SM electroweak boson are
 448 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
 449 expected between the quark from the hadronic decay. The comparison with m^{calo} is shown in Figure 16,
 450 together with the boosted top-quark for comparison. As seen here, the performance of the latter is clearly
 451 worse than the former, the trend is yet very similar. This difference is interpreted in terms of different
 452 and more complex topology and hence higher sub-jet multiplicity: in the three sub-jet structure, resolving
 453 accurately the components is more challenging.

454 5.11 Other Stability Quantifiers

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

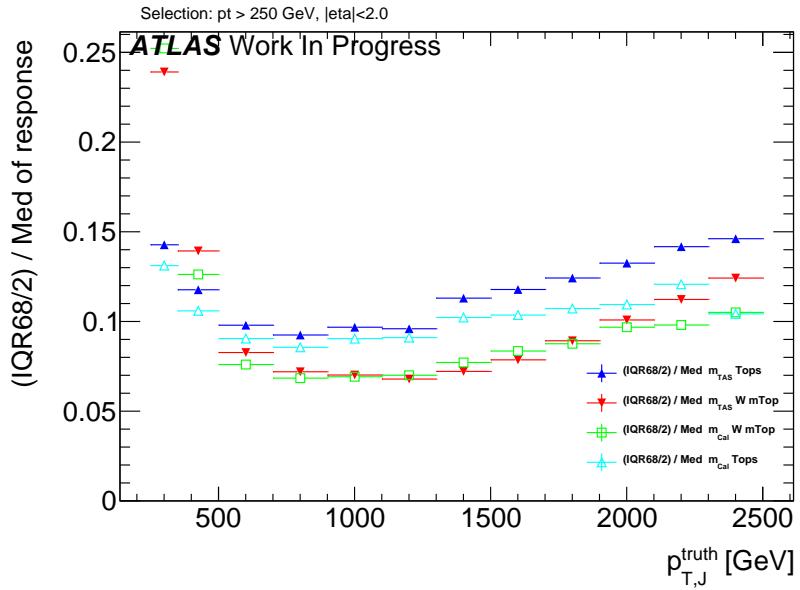


Figure 16: Performance of the m^{TAS} versus the m^{calo} for the massive W/Z (in red and green); shown on the same plot also the boosted top sample (in blue and light blue).

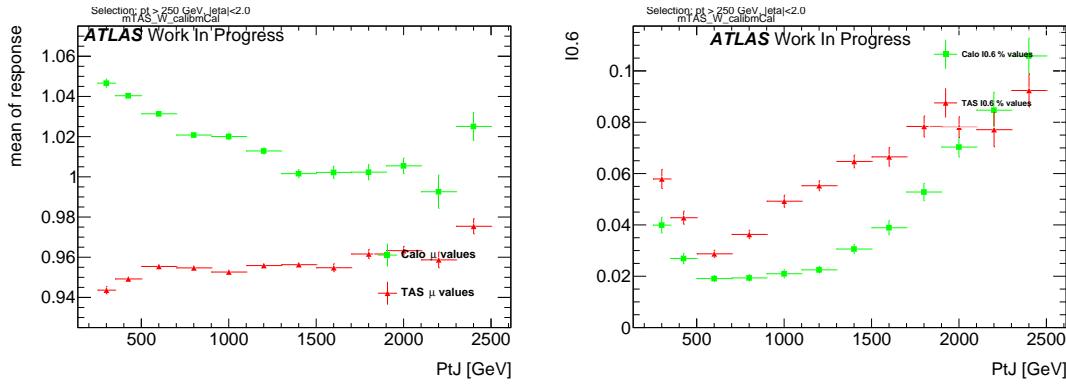


Figure 17: Stability quantifiers which were checked for the m^{TAS} : mean, on the left, and normalized left-hand side integral, on the right, of the mass response distribution. The mean is calculated from a Gaussian fit and the integral goes from 0 to 0.6.

- The mean of the response distribution is shown for boosted W/Z decays in Figure 17, left; as seen here, despite being the mean constantly below the unity, its behavior is much more flat and independent of p_T , especially in the low-middle regime. This is surprising since the m^{calo} is already shown after the calibration step, which is not taken instead for the m^{TAS} . Conversely the left-hand-side tail of the mass response which is shown in the same figure, right, shows a more enhanced behavior than the m^{calo} , but still never reaches the 10%. Of course an enhancement of the tail causes a loss of gaussianity and a number of jets which are reconstructed with a lower mass than they should, but it is still comparable with the calorimeter mass.
- Those quantifiers show analogous behavior for the other samples considered and those figures can be found in the Appendix.

470 5.12 Sub-jet Calibration

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

474 5.13 Preliminary Studies on Sub-jet Calibration

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

- 480 • m^{TAS} using truth-level sub-jets and tracks; normal tracks (with all detector effects) are used to assist
 481 the truth-level sub-jets;
- 482 • m^{TAS} using truth-level tracks and truth-level sub-jets; the truth-level tracks are used to assist the
 483 truth-level sub-jets;
- 484 • m^{calo} truth, calculated using only the truth sub-jets.

485 5.13.1 Perfect Calibration

486 The *perfect calibration* refers to the procedure of using m^{TAS} with truth-level sub-jets and track, i.e.
 487 looking at the best possible scenario with an ideal detector. The performance is of course expected to be
 488 optimal, because of the use of the truth-level. This step was necessary as feasibility study, to understand
 489 whether ulterior efforts in this direction were meaningful. The perfect calibration is shown in Figure
 490 18; since the performance exhibits room for big improvement below ~ 1 TeV and moderate to small
 491 improvement above this value, the second step of a simple calibration was tried.

492 5.13.2 Simple Sub-jet Calibration

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

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

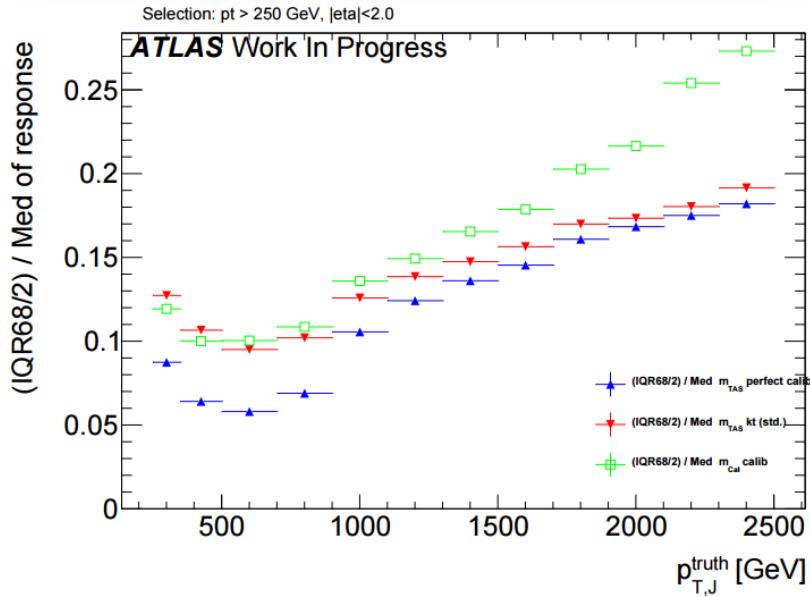


Figure 18: Performance of the perfect calibration. It shows room for improvement especially at low-middle p_T .

502 This procedure was called *poor man's calibration* or PM calibration or *simple calibration*. A check on
 503 the p_T response before and after calibration together with the mean of the entire Large- R jet response is
 504 shown in Figure 48 and 49 in Appendix.

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

512 5.14 Limitation of m^{TAS}

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

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

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

522 The hypothesis of the degradation of the m^{TAS} driven by the tracks is also supported by the Figure 50
 523 in Appendix, where the truth-level tracks are used instead of real tracks to compute the variable; it can

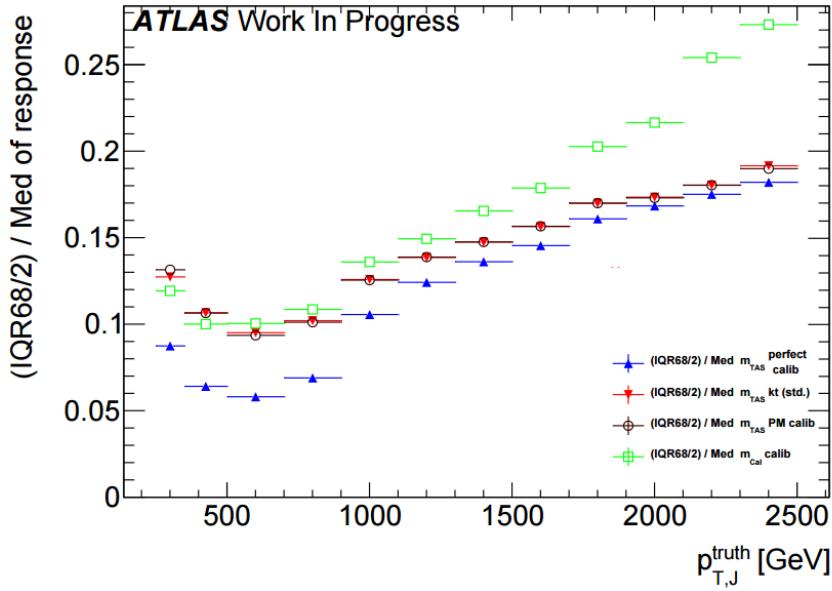


Figure 19: Performance of the poor man’s calibration. The improvement is marginal throughout the entire transverse momentum space.

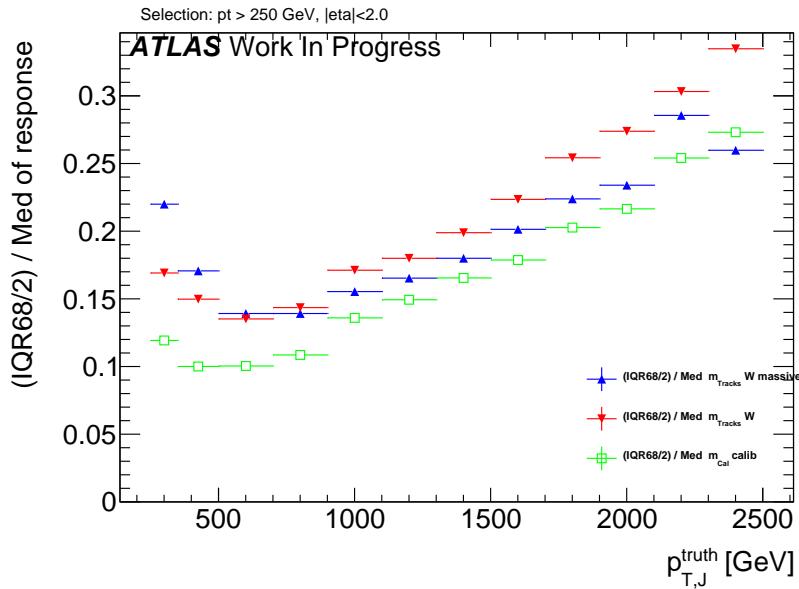


Figure 20: The performance of the track mass in blue and red for massive W sample and boosted W/Z respectively; for reference in green the calorimeter mass of the large- R jet.

524 be seen the flat behavior at high p_T , hence ascribing the worsening of the resolution to tracks at higher
 525 transverse momenta.

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

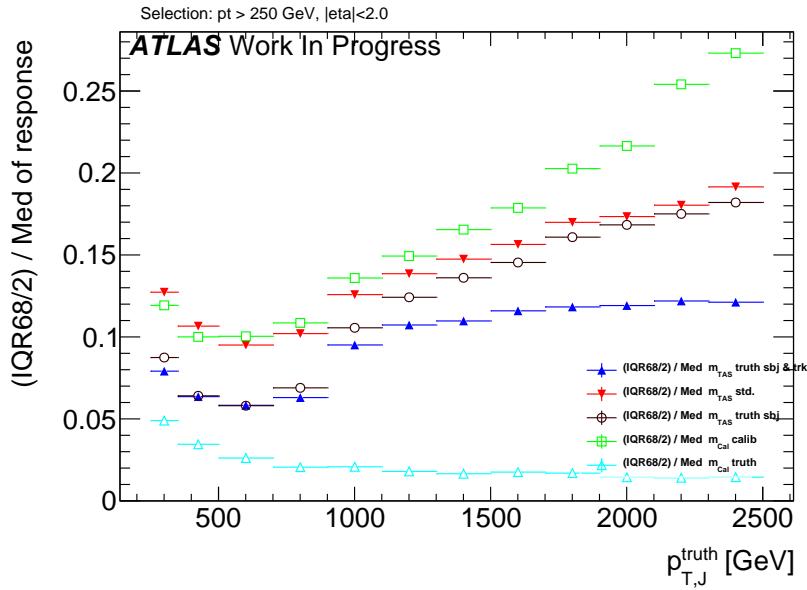


Figure 21: Breakdown of the m^{TAS} in its component using truth-level information for boosted W/Z decays.

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

531 5.15 Alternative Observable Definitions

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

534 Alternatives considered were:

- 535 • for the tracks:
 - 536 – use of tracks not as input directly, but only taking those belonging to anti- k_t reclustered
537 track-jet with radius of 0.3 or 0.2;
 - 538 – tighter or looser quality conditions were explored;
 - 539 – tighter or looser primary vertex association requirement were explored.

- 540 • for the sub-jets:
 - 541 – the trimming procedure was modified: various radii R_{sub} of the sub-jets were tested;
 - 542 – the sub-jets were reclustered using not only the standard k_t , but also anti- k_t and C/A.
- 543 • for the procedure: different 4-momentum correction scheme was also explored.

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

548 In the Appendix, figure 43 44 45 the performance for boosted W/Z , tops and Higgs are shown, respectively.
 549 It can be seen that the k_t algorithm provides the best observable definition, in all the samples considered.
 550 However, the anti- k_t algorithm provides similar performances; this was an important check as the jet
 551 calibration procedure currently going on in ATLAS, the *R-Scan* procedures includes the anti- k_t algorithm
 552 with radius of $R=0.2$ and aims at providing the calibration and uncertainties that could be used directly in
 553 the computation of the m^{TAS} .

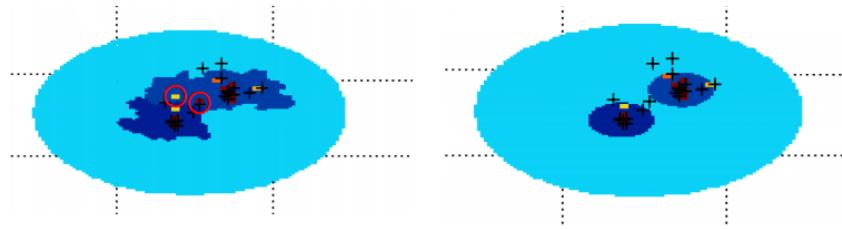


Figure 22: An example of event-display shows the differences in the reclustering algorithm used for the sub-jets: on the right k_t and on the left anti- k_t . Highlighted some constituents trimmed away with the second choice.

554 6 Combining the mass observables

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

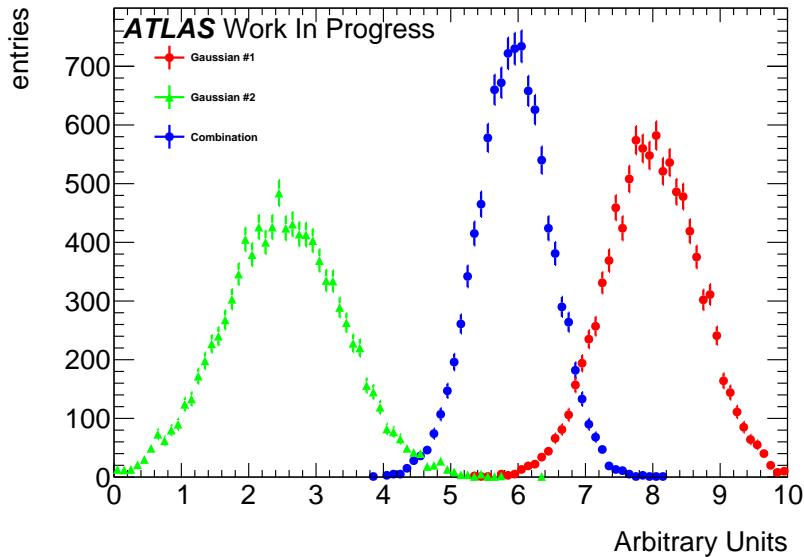


Figure 23: A toy example of the combination of two independent Gaussian observables, in red and green, and their combination, in blue. It can be seen that the combination has a smaller width.

557 This is true for both the m^{TA} and the m^{TAS} ; they are introduced in the next subsections. Provided that the
 558 two observables are nearly independent (correlation coefficient are $\sim 10\%$, see Figure 46 in the Appendix),
 559 due to the Gaussian nature of the p_T and mass response, the optimal combination of the two is linear¹. An
 560 example is provided in Figure 23.

561 **6.1 Combination $m^{TA} - m^{calo}$**

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

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

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

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

565 **6.2 Combination $m^{TAS} - m^{calo}$**

566 There is a main difference between the m^{TAS} and m^{TA} when it comes to combination: since the m^{TAS} is
 567 using sub-jet level information but m^{TA} not, the correlation with the m^{calo} is expected to be higher. This
 568 can be seen e.g. in the plots in Figure 24 (additional plots shown in Figure 47 in Appendix), where the
 569 correlation is not only higher for the simple W/Z and Higgs jets, but above 50% for tops. The assumption
 570 of independent variables here falls, forcing a more complete approach. The Ansatz is to take into account
 571 the correlation via the formula:

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

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

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

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

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

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

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 **6.2.1 Procedure**

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

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

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

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

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

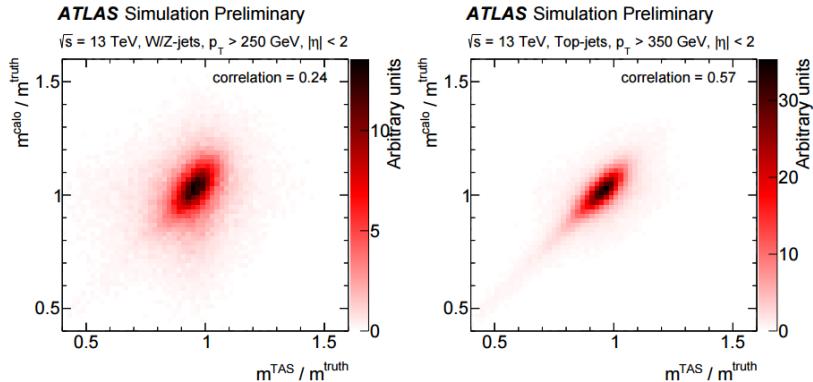


Figure 24: The calorimeter based jet mass response versus the track-assisted sub-jet mass response, on the left for boosted W/Z on the right for boosted tops.

595 **6.3 Performance in $W \rightarrow q' \bar{q}$ Decays**

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

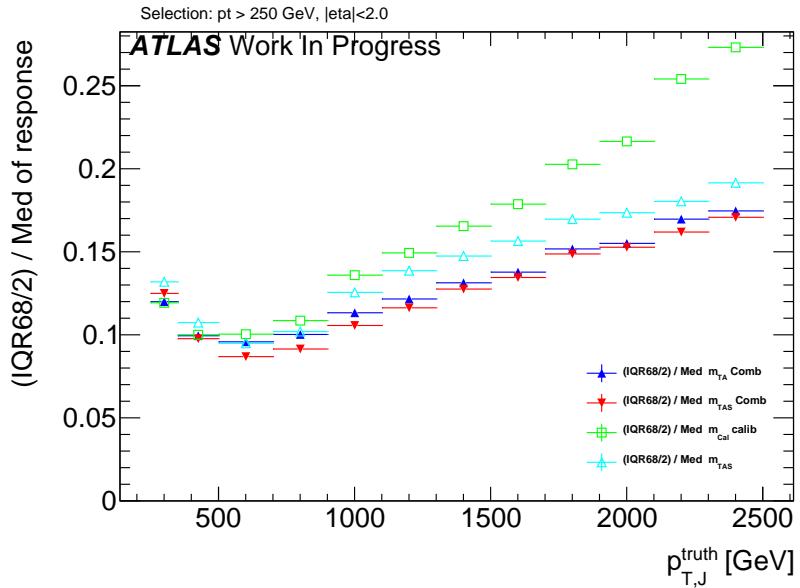


Figure 25: Performance of the combined mass on W/Z samples; here shown the two definitions of the combined mass, m^{comb} and m_{TAS}^{comb} , together with the calorimeter mass and the track-assisted sub-jet mass.

600 6.4 Performance in $t \rightarrow q'\bar{q}b$ Decays

601 The boosted top sample remains the most challenging one also with the combined mass; as seen on Figure
 602 26, the m^{comb} performs quite similarly to the calorimeter based mass definition, yet behaving considerably
 603 better than the m^{TAS} especially at high transverse momentum. The m_{TAS}^{comb} , however, outperforms all the
 604 other definitions, and shows its optimal observable strength at middle p_T i.e. in the range $1 < p_T < 1.6$
 605 TeV.

606 6.5 Performance in $h \rightarrow b\bar{b}$ Decays

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

612 7 Energy Correlation Functions and n-Subjettiness

613 7.1 Sub-jet track assistance

614 Tracks and their angular resolution could not only improve the jet mass definition but also the performance
 615 of tagging variables such as the Energy Correlation Functions or n-Subjettiness. These variables are
 616 usually calculated with calorimeter clusters as input, studied here are tracks and assisted tracks as input in
 617 comparison with the default method using clusters. In contrast to the m^{TA} variable introduced in Section

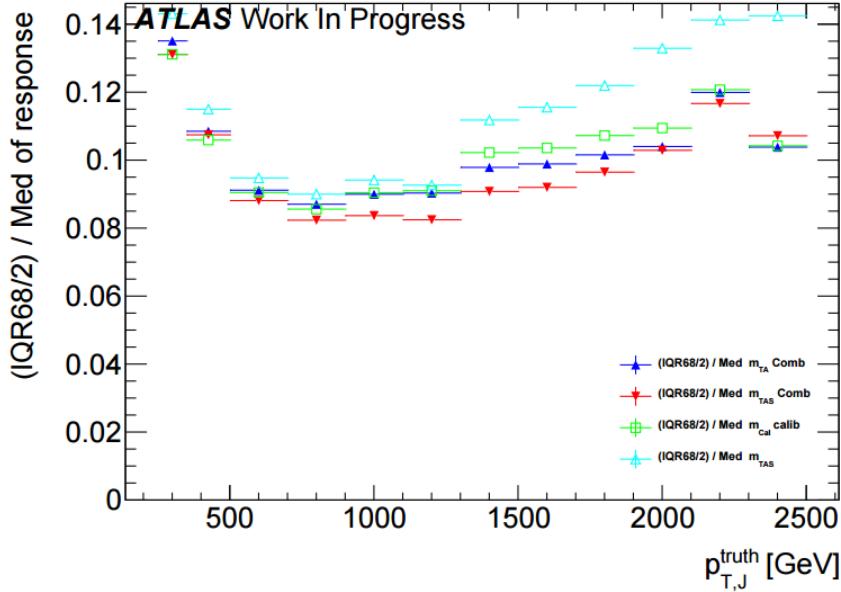


Figure 26: Performance of the combined mass on the top sample; here shown the two definitions of the combined mass, m^{comb} and m_{TAS}^{comb} , together with the calorimeter mass and the track-assisted sub-jet mass.

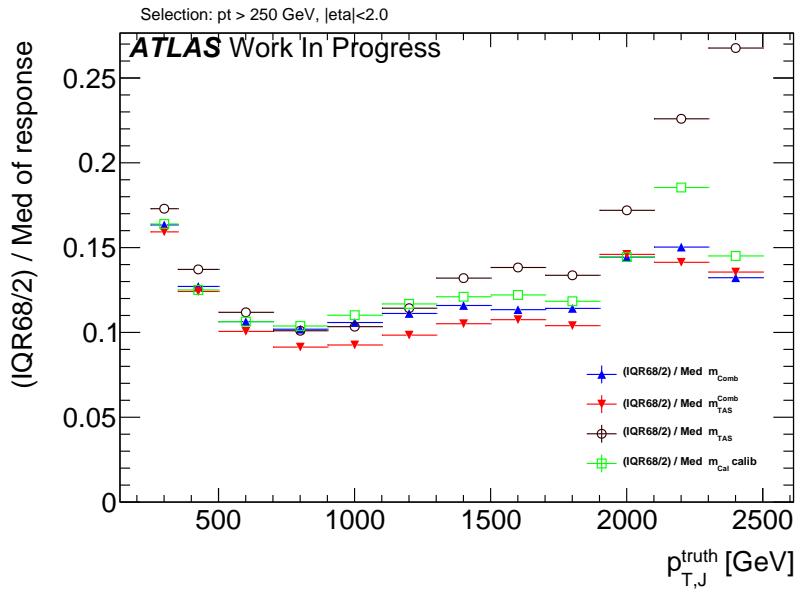


Figure 27: Performance of the combined mass on the Higgs decay; here shown the two definitions of the combined mass, m^{comb} and m_{TAS}^{comb} , together with the calorimeter mass and the track-assisted sub-jet mass.

618 5.1, not the mass but the p_T of each track is scaled, since C2, D2, τ_{21} and τ_{32} are calculated with the
 619 constituents p_T . The concept of track assisting with the p_T ratio of the whole jet is without effect for the
 620 studied substructure variables. This can be understood from the definitions of the weighted p_T sums. If
 621 corrected with only one ratio, all tracks are scaled by the same factor c , which then can be put in front of

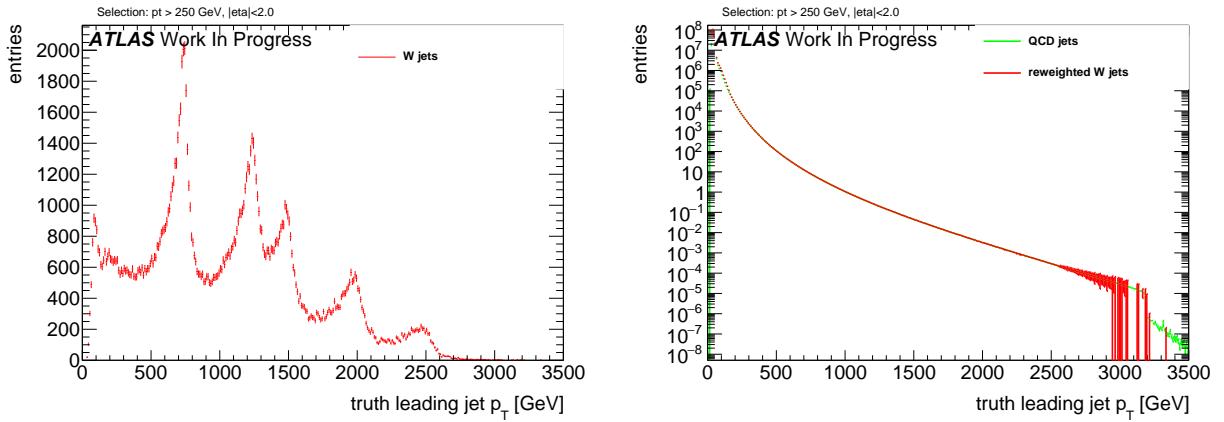


Figure 28: Exemplary p_T distributions of (left) W boson jets and (right) QCD jets from multi-jet events with reweighted W boson events

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 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.

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.

634 Tagging variables such as C2, D2, τ_{21} and τ_{32} are usually used after applying a mass cut around the interval
 635 that contains 68% of the signal events. Therefore, a cut is applied on the calibrated mass of the large-R
 636 calorimeter jet which is calculated to cover the smallest interval around the peak mass that contains 68% of
 637 the signal events. The comparison is performed in six different p_T regions to study the behavior connected
 638 with rising energy of the decaying particle. These regions are presented in the left part of Table 2. In case
 639 of the Higgs boson study, there is not enough statistics to derive a conclusive result for $p_T > 2000$ GeV,
 640 since the highest resonance mass of the $G^* \rightarrow HH$ samples is 3000 GeV in contrast to 5000 GeV for the
 641 $Z' \rightarrow tt$ and $W' \rightarrow WZ$ samples. Hence this study is restricted to the five lower p_T bins. Prior to tagging
 642 with the n-Subjettiness or C2/D2 variables, a cut on the calibrated calorimeter jet mass is applied, given
 643 that the mass is the main discriminant in QCD jet rejection. This cut is defined to choose the smallest
 644 interval around the peak mass containing 68% of the signal. However, the reconstructed mass depends on
 645 the p_T region, therefore a different cut was calculated for every region to meet the requirements.

646 7.2 Track Selection

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

651 The distributions showing the number of tracks associated to a calorimeter jet, see the left side of Figure
 652 29, indicate, that on average around four tracks less are associated to the sub-jets compared to the
 653 ungroomed jet. The right side of Figure 29 shows the angular distance ΔR between the single tracks and
 654 the axis of the large-R calorimeter jet. Both distributions are aligned in the lower ΔR region while the
 655 histogram representing the tracks associated to the ungroomed jet shows an enhancement towards larger
 656 ΔR . Accordingly, these additional tracks feature an angular separation from the jet axis of more than
 657 0.3, and are in consequence distributed primarily around the outer regions of the large-R jet. Given the
 658 required primary vertex association, it is unlikely that these tracks originate from pile-up. Instead, the
 659 origin might be found in final- or initial state radiation.

660 Figure 30 shows the signal distributions of the C2/D2, and τ_{21} , calculated with both selections of tracks
 661 for W boson jets. The large ΔR to the jet axis of the differing tracks push the substructure variables to
 662 higher, more background like values. The broader distributions are a result of the variating nature of
 663 these tracks. C2 and D2 are more sensitive to tracks with a large ΔR to the jet axis, because the angular

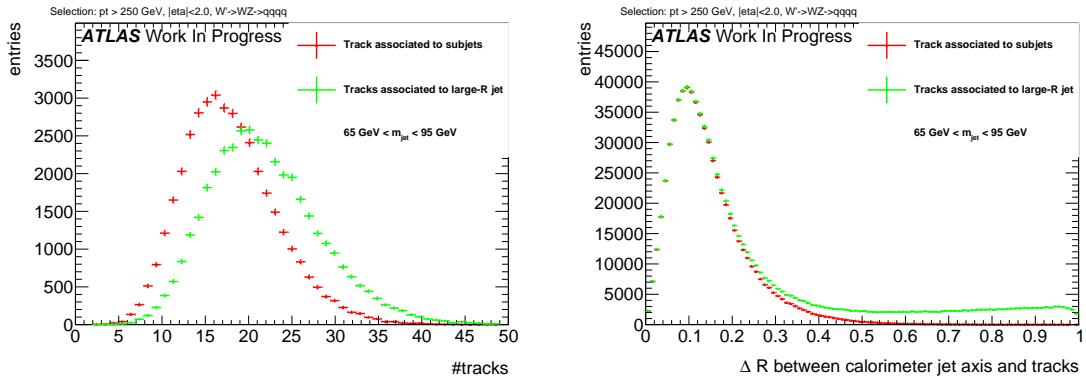


Figure 29: The number of tracks ghost associated to the large-R jet and to the sub-jets (left) and angular distance of associated tracks to the large-R calorimeter jet axis (right). Signal events were not reweighted at this step.

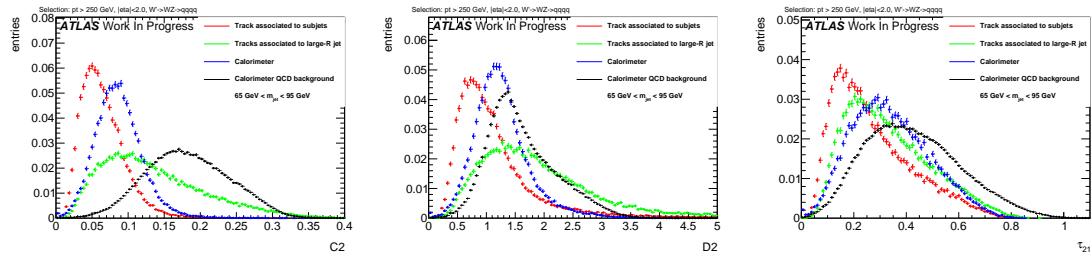


Figure 30: Substructure variables (left) C2, (right) D2 and (below) τ_{21} calculated with calorimeter clusters as well as tracks associated to sub-jets and to the large-R jet. Signal events were not reweighted at this step.

664 distance between all pairs and triples of tracks is considered, among tracks on possibly opposite ends of
 665 the large-R jet, whereas τ_{21} uses distances to k_T -WTA axes. For comparison, the signal and background
 666 distributions for the variables calculated with calorimeter clusters are shown as well. It is possible to
 667 anticipate that the performance of variables calculated with tracks and assisted tracks is not worse than
 668 cluster base variables. In contrast to the previously studied jet mass variable, ratios of ECF(N) and τ_N
 669 are rather energy scale independent and are found to not be as sensitive to the missing neutral fraction
 670 with un-assisted tracks. Starting from this observations, the performance of substructure techniques is
 671 compared with the following objects as input:

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

675 7.3 Performance with default β

676 The performance of track and TAS based tagging variables with the default angular weighting of $\beta =$
 677 1 is compared to the corresponding calorimeter variables for W boson, Higgs boson and Top quark
 678 identification. The stated signal efficiencies are calculated after the mass cut plus tagging with n-
 679 Subjettiness or C2/D2. Therefore, the endpoint of the ROCs is at 68% signal efficiency, the fraction kept
 680 after the mass cut. Consequently, it is required to achieve a tagging only signal efficiency of $\frac{0.5}{0.68} \sim 0.74$
 681 for a signal efficiency of 0.5 after mass cut and tagging. Similarly, the stated and compared background

rejections result from the multiplication of both, thus representing the QCD rejection of the combined mass- and tagging variable-cut. The complete set of signal and background distributions for different inputs can be found in the Appendix ??.

7.3.1 Performance for W boson tagging

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

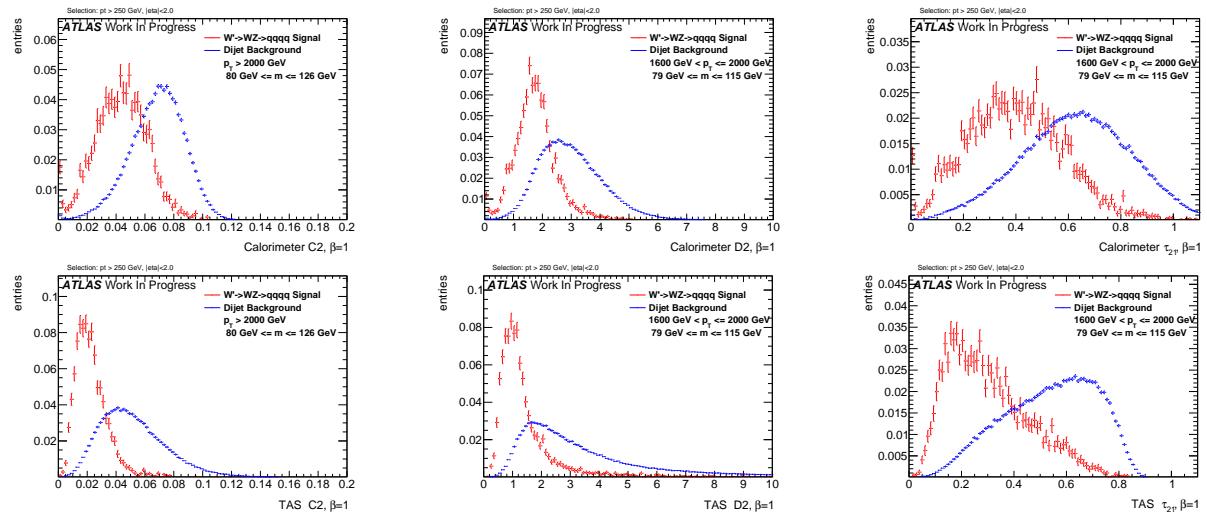


Figure 31: W boson signal and QCD background distributions for calorimeter (left) and TAS (right) at $\beta = 1$ with C2 (top) for more than 2000 GeV and D2 (middle) and τ_{21} (bottom) for 1200-1600 GeV

690

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

7.3.2 Un-assisted tracks and TAS at very high p_T

The C2 variable was found to perform equally well with tracks and TAS as input. This variable seems to be relative insensitive to the track assisting and tracks alone already perform well. D2 and τ_{21} in contrast, feature a visibly worse separation with tracks than with assisted tracks. In these cases, the scale difference due to the missing neutral fraction seems to have a greater influence.

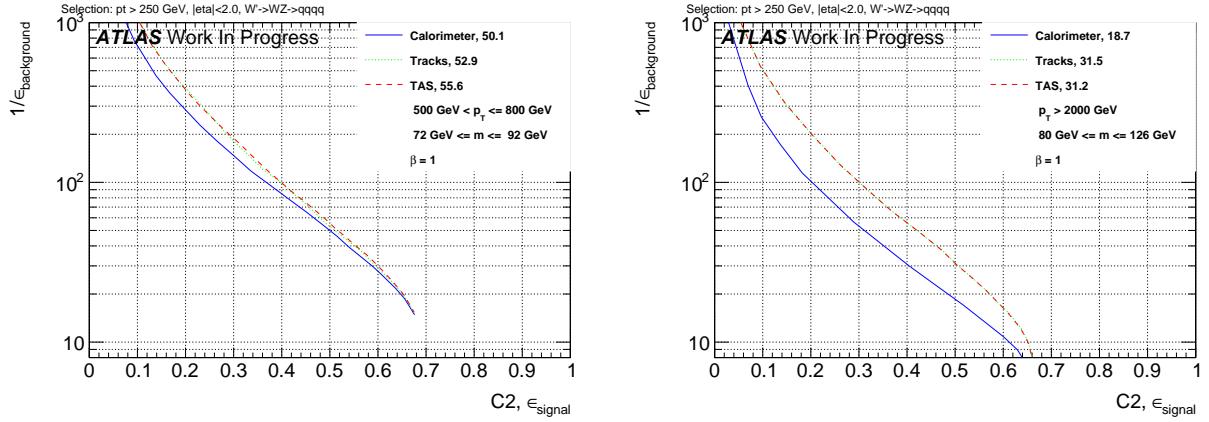


Figure 32: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and colorimeter $C2$ at $\beta = 1$ for 500 - 800 GeV (left) and >2000 GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

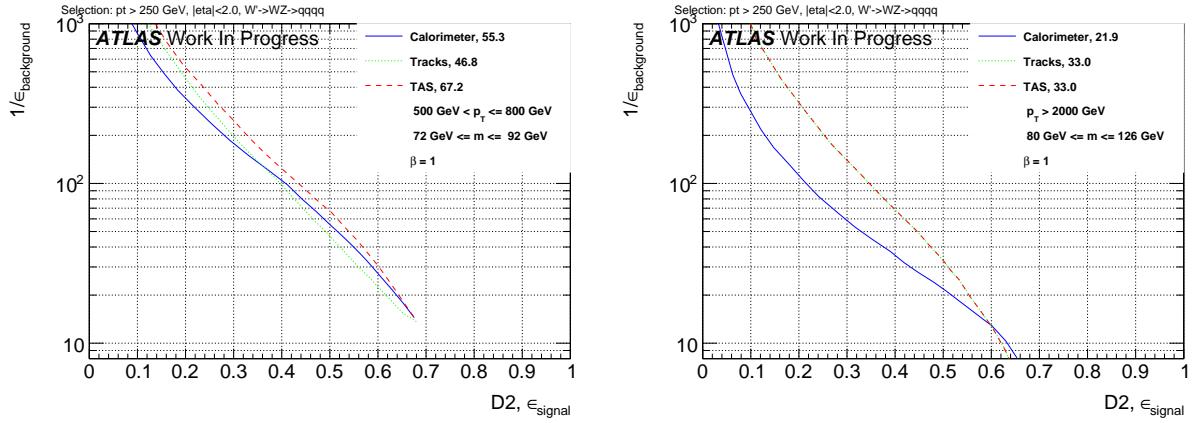


Figure 33: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and colorimeter $D2$ at $\beta = 1$ for 500 - 800 GeV (left) and >2000 GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

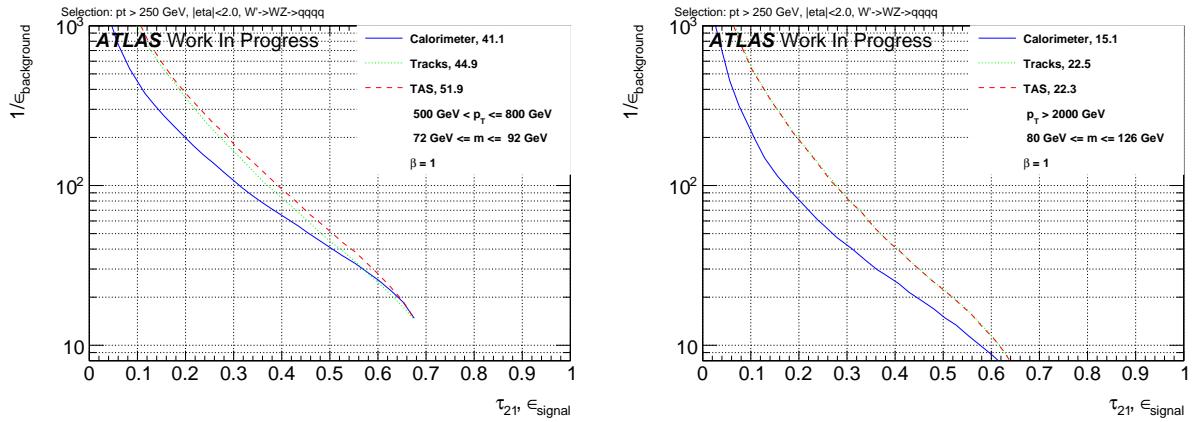


Figure 34: ROCs showing QCD rejection against W boson efficiency for tracks, TAS and colorimeter τ_{21} at $\beta = 1$ for 500 - 800 GeV (left) and >2000 GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

703 For very high p_T values however, it is often the case that the large- R calorimeter jet features only one
 704 $R = 0.2$ sub-jet after trimming due to the now small separation of constituents. A single sub-jet results in
 705 the TAS procedure to fall back to TA. As stated in Section 7.1, TA has no impact on the ratios. Therefore,
 706 C2/D2 and τ_{21} perform equally well when calculated with tracks or TAS for events with only one sub-jet
 707 and thereby the difference between both decreases for very high energies.

708 7.3.3 Correlation with p_T

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

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

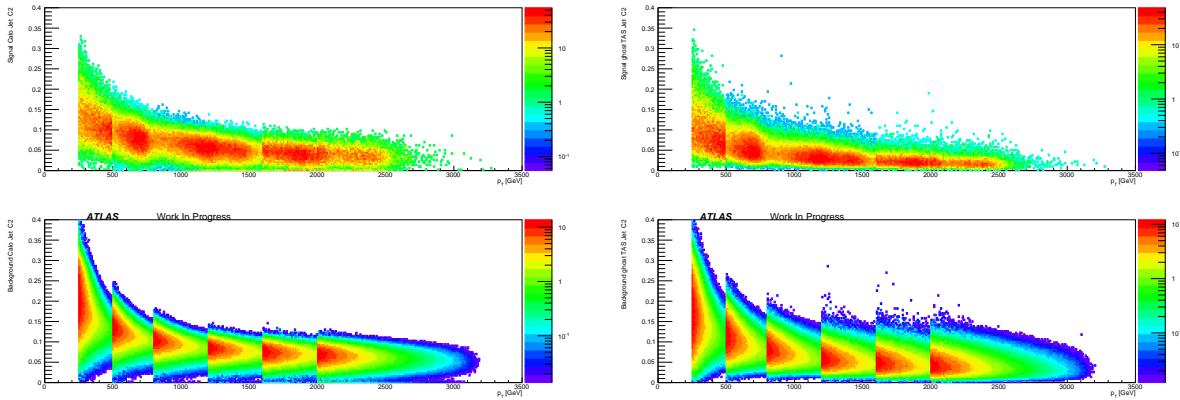


Figure 35: Correlation between C2 at $\beta = 1$ and p_T applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

719 7.4 Optimisation of β

720 8 Conclusions & Outlook

721 8.1 Jet mass observables

722 The m^{TAS} variable was developed for the large- R jet mass; it combines the information of the tracker- and
 723 calorimeter-system to achieve an higher precision in the jet mass reconstruction, correcting the missed
 724 neutral fraction which is absent in the tracker but not in the calorimeter. With respect to the m^{TA} , it

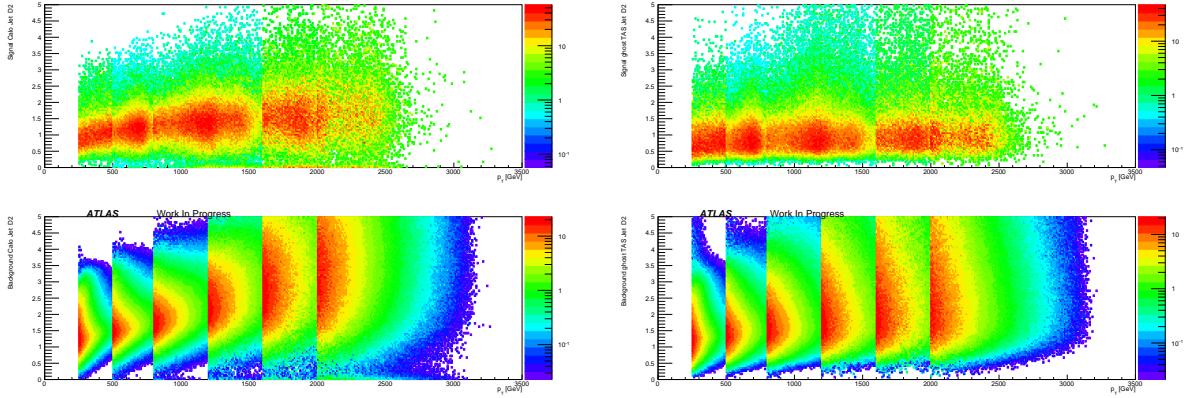


Figure 36: Correlation between D2 at $\beta = 1$ and p_T applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

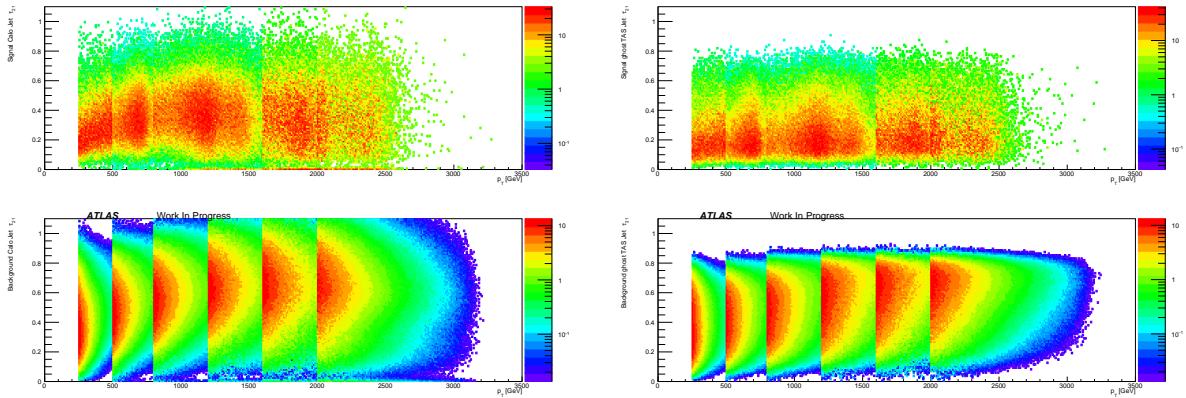


Figure 37: Correlation between τ_{21} at $\beta = 1$ and p_T applied on W boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

725 applies this correction at sub-jet by sub-jet level and not at jet by jet level, therefore providing a more
 726 accurate reconstruction. It was shown in Monte Carlo simulation to be a very good observable confronting
 727 quantitatively with the other definitions which are either standard or in preparation, m^{calo} , m^{TA} and m^{comb} .
 728 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
 729 merit, the mass response, for the boosted W/Z and QCD sample; is always better than the m^{TA} and similar
 730 to the m^{calo} for the boosted tops and Higgs. Moreover, it is a slightly worse observable than the m^{comb} , yet
 731 being comparable, and avoiding the development of ad-hoc weights. The optimal configuration of m^{TAS}
 732 is shown and confronted with different approaches, in particular in terms of different trimming procedure
 733 of the large- R jet to be used as an input. All the components of the observable have been studied with the
 734 use of truth Monte Carlo information without detector effect, in order to evaluate quantitatively its limits
 735 and strengths; the track p_T measure degradation was found to be the cause of the variable decreasing
 736 performance at higher transverse momenta.

737 The m^{comb}_{TAS} is the logical extension of the m^{TAS} , which improves by construction the results beyond the
 738 m^{calo} and the m^{TAS} , combining these two variables on the same way of the m^{comb} , but taking into account

739 the higher correlation factor which is inherited from the sub-jet usage. Weights for its construction can
 740 be in both cases either derived specifically for the sample considered, or constructed on average with the
 741 QCD sample, in this case getting a sub-optimal performance. In all the cases studied, it has a better
 742 behavior than the m^{comb} , m^{calo} and m^{TA} .

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

747 8.1.1 Outlook

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

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

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

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

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

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

767 8.2 Energy Correlation Functions and n-Subjettiness

768 ***here sascha conclusions***

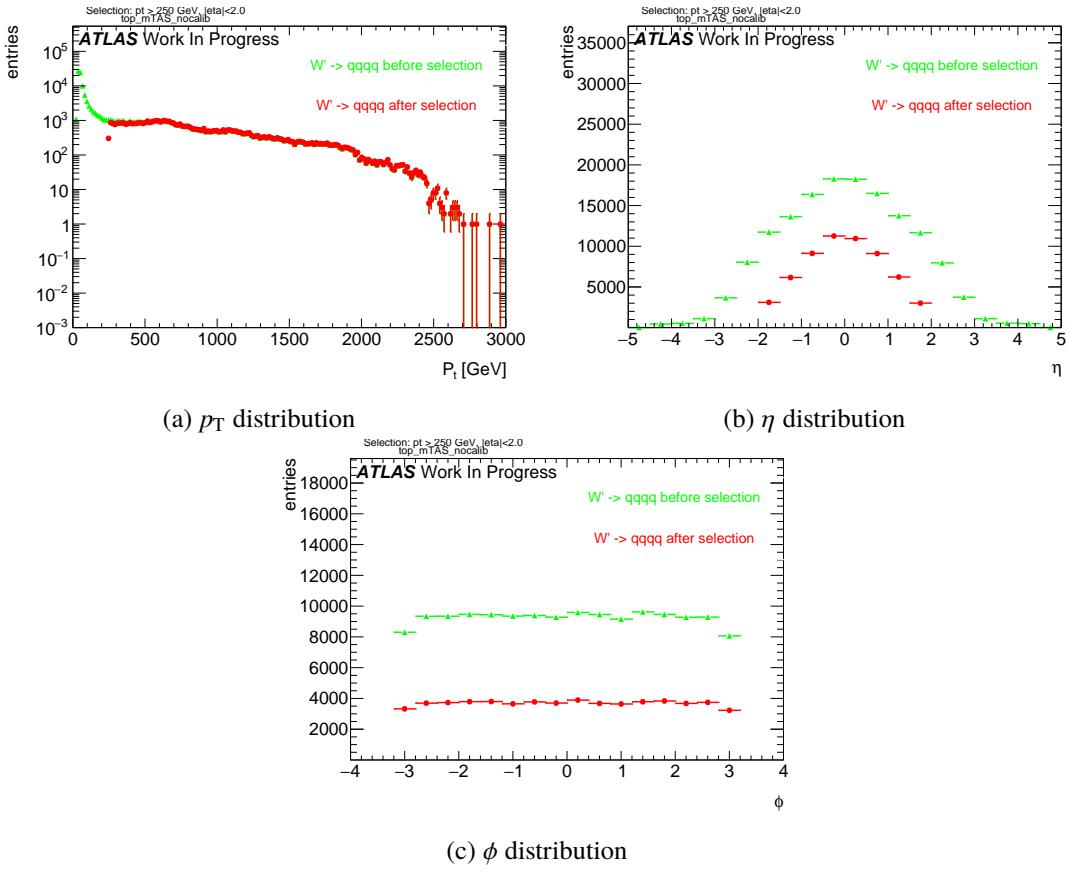


Figure 38: Boosted tops kinematic distribution.

769 Appendix

770 inputappendixA1.tex

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

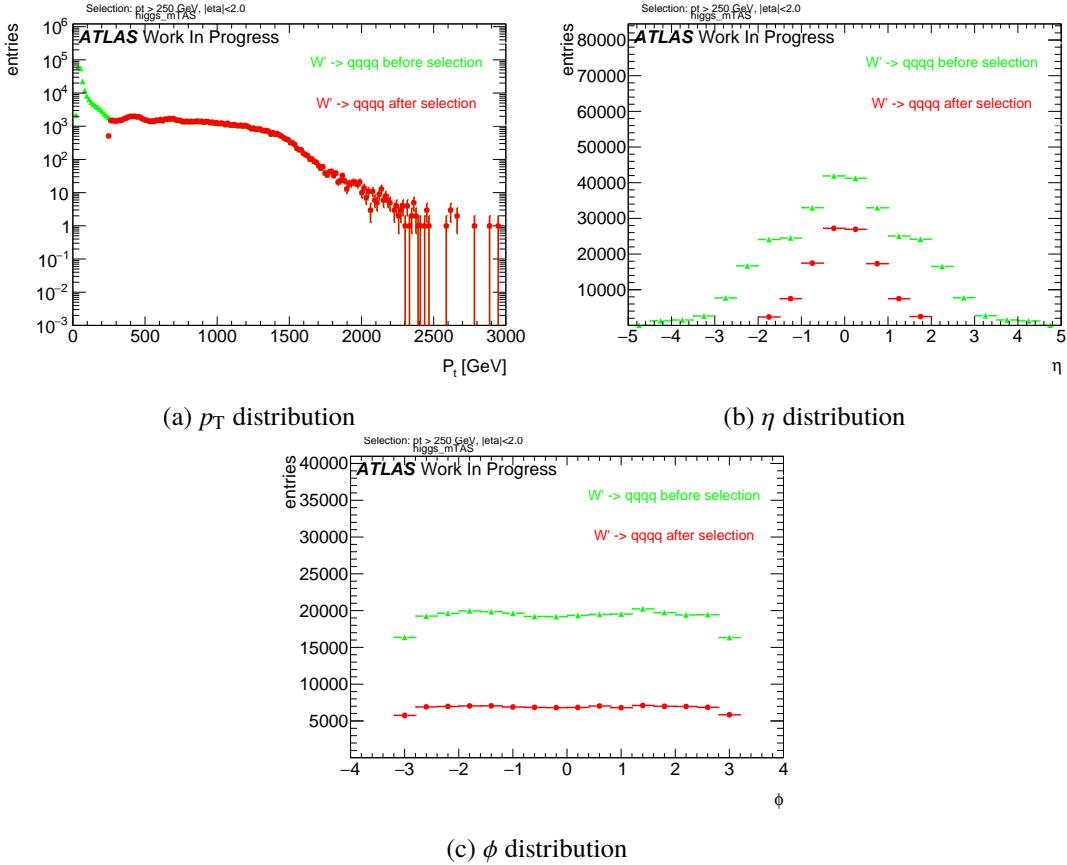


Figure 39: RS-Graviton kinematic distribution.

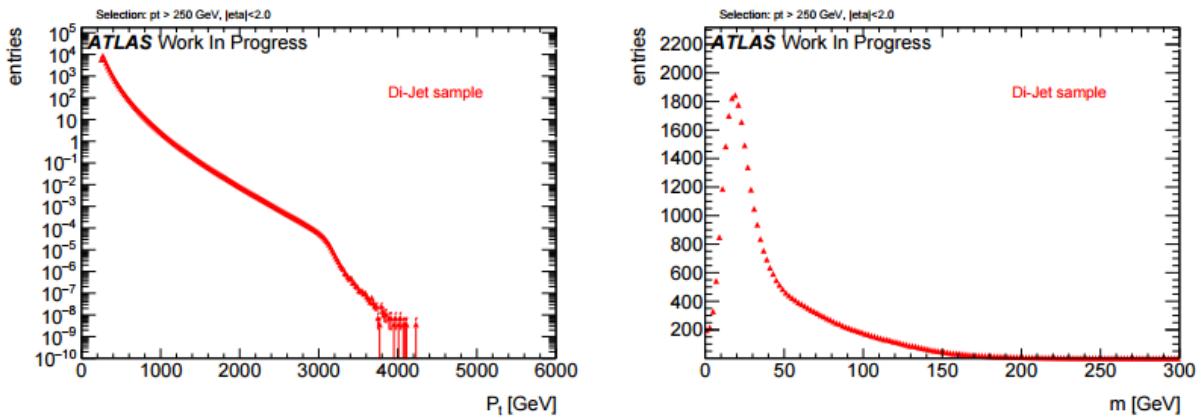


Figure 40: QCD dijet transverse momentum and mass distributions.

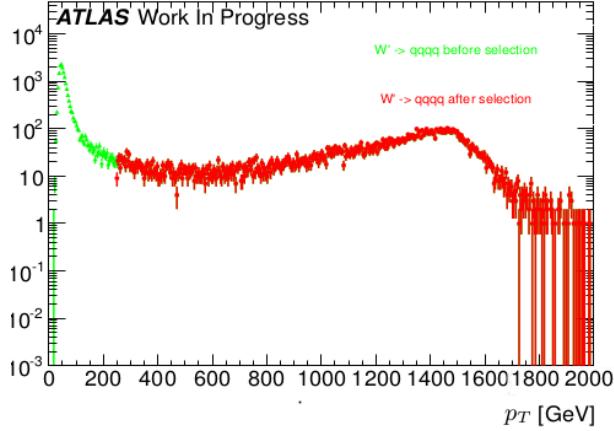


Figure 41: The p_T distribution of a 3 TeV resonance from the hadronically decaying W or Z , in logarithmic plot. As can be seen, the jacobian peak is around $p_T \simeq m_{W'}/2 \simeq 1.5$ TeV.

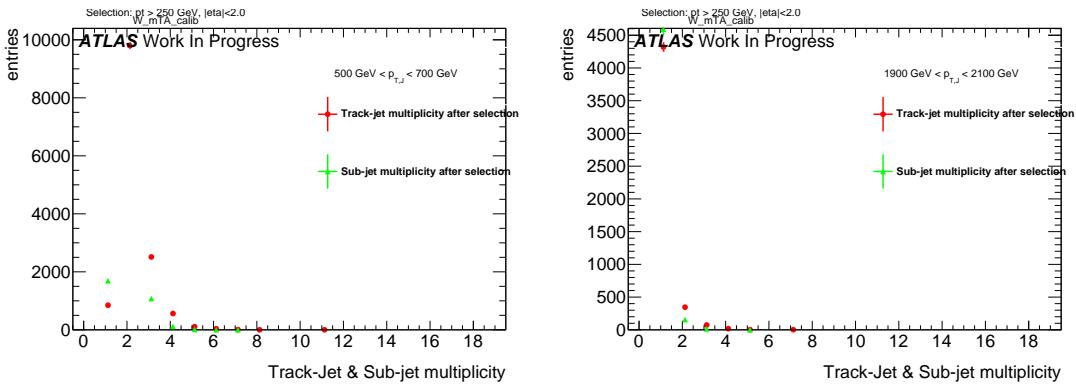


Figure 42: Sub-jet and Track-jet (jets created having tracks as input) multiplicity, for selected bins of transverse momentum.

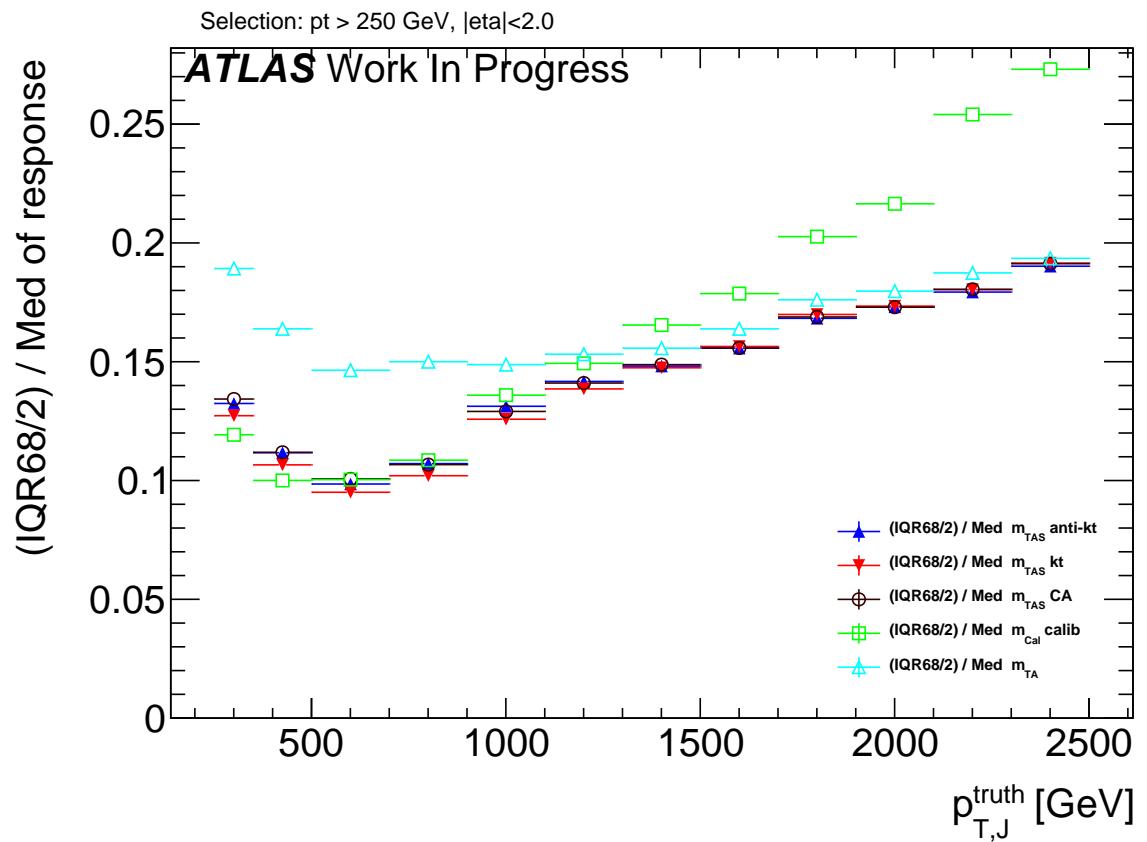


Figure 43: Performance of m^{TAS} with different reclustering algorithm for the sub-jets: anti- k_t , k_t and C/A. Boosted W/Z sample.

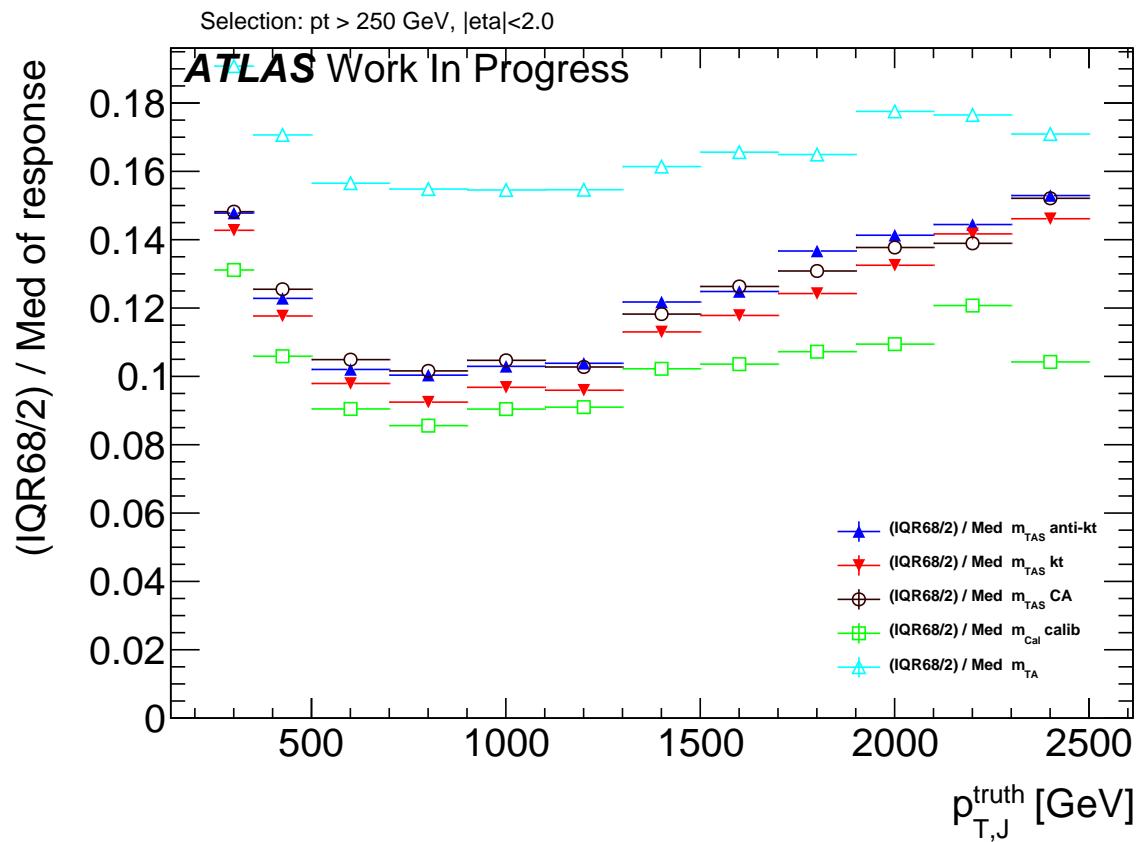


Figure 44: Performance of m^{TAS} with different reclustering algorithm for the sub-jets: anti- k_t , k_t and C/A. Boosted top sample.

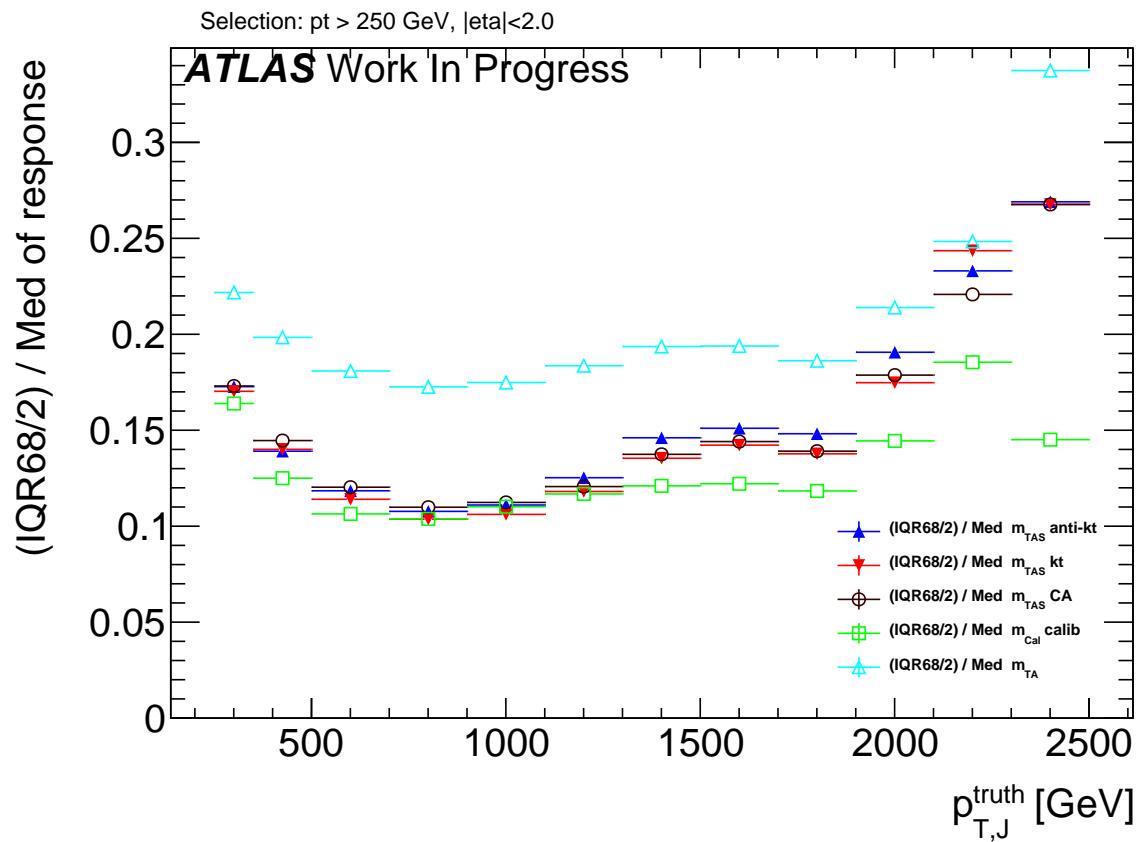


Figure 45: Performance of m^{TAS} with different reclustering algorithm for the sub-jets: anti- k_t , k_t and C/A. Boosted higgs sample.

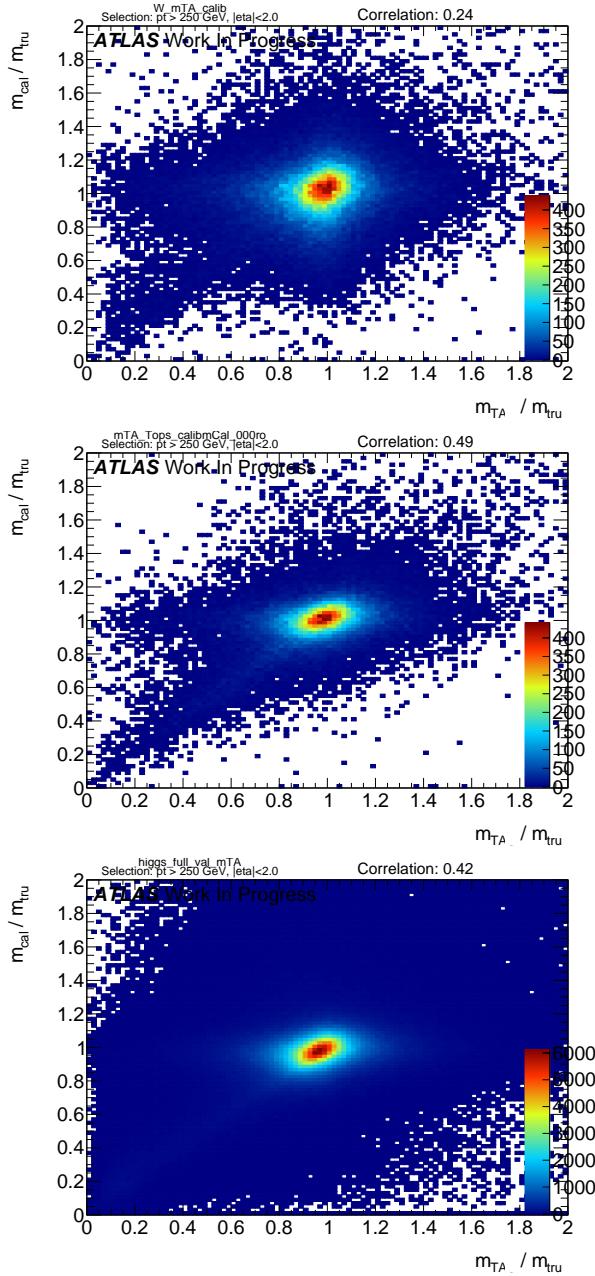


Figure 46: Calorimeter based jet mass response vs the track-assisted mass response for the three signal samples. Correlation coefficient is indicated on the top right.

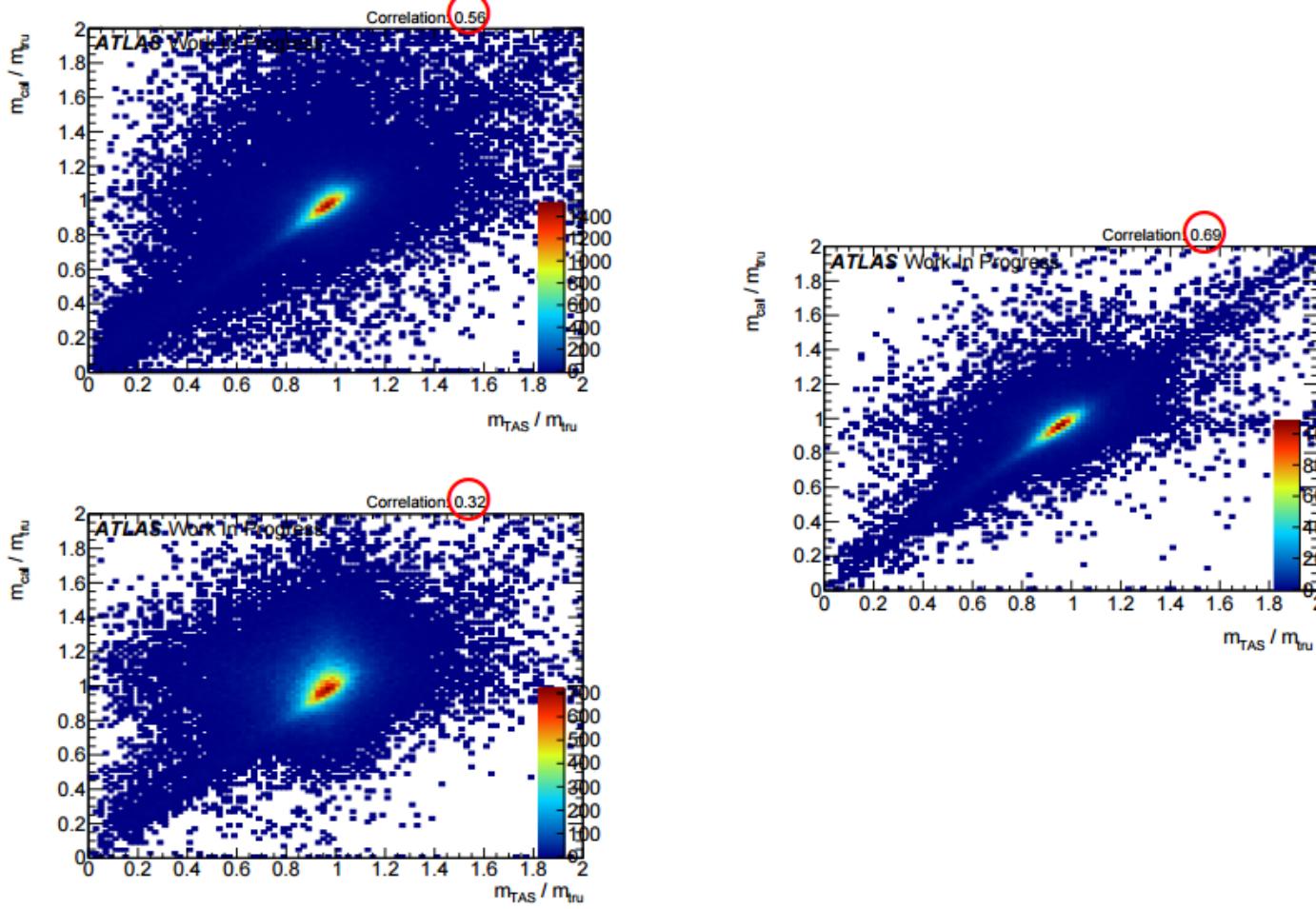


Figure 47: Calorimeter based jet mass response vs the track-assisted sub-jet mass response for the three signal samples. Correlation coefficient is indicated on the top right and highlighted. On the left, top, the higgs sample, bottom, the W/Z ; on the right the top-quark sample.

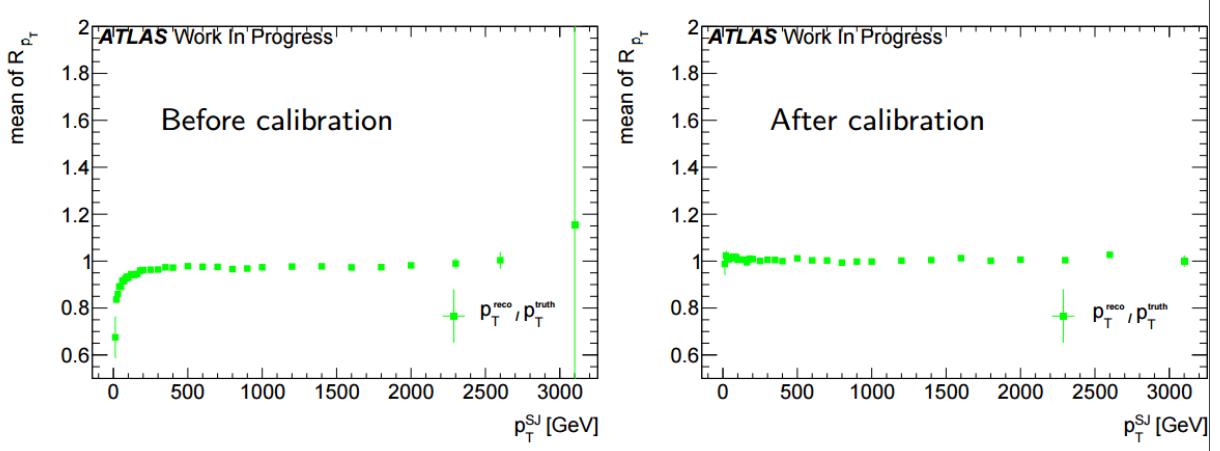


Figure 48: Poor's man calibration effect on mean of transverse momentum's response of the sub-jet, before, left, and after, right, the procedure.

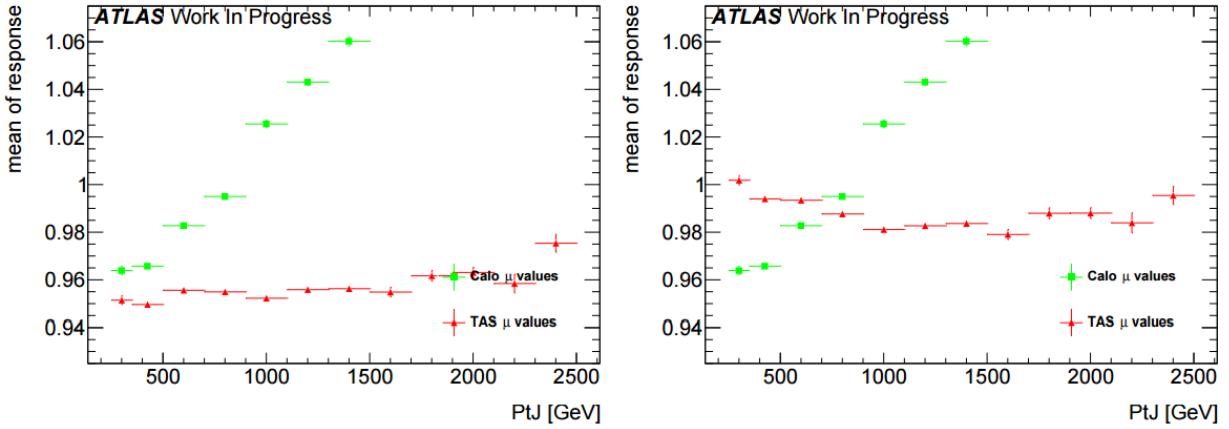


Figure 49: Poor's man calibration effect on the mean of the mass response of the large-R jet, before, left, and after, right, the procedure.

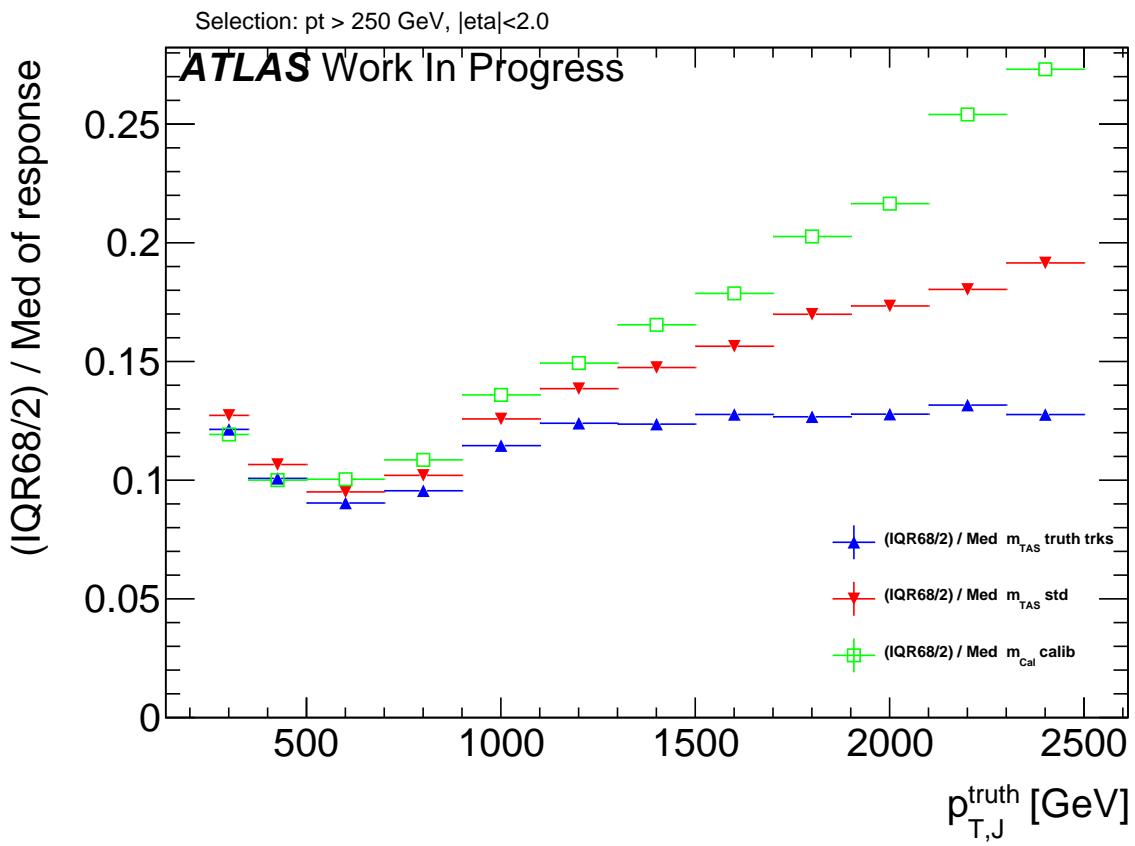
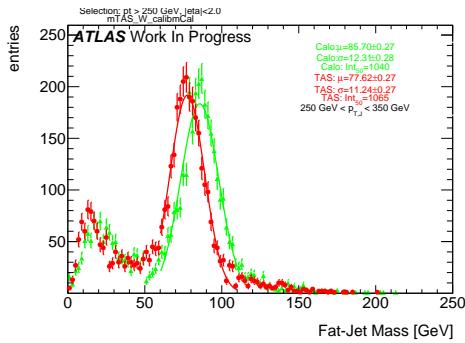
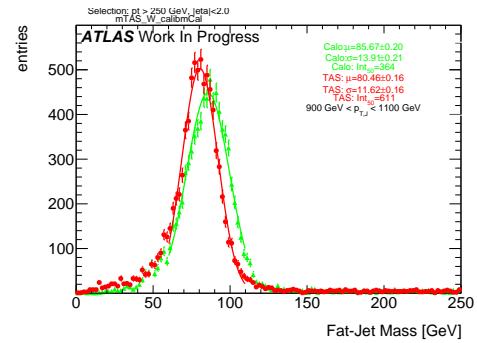
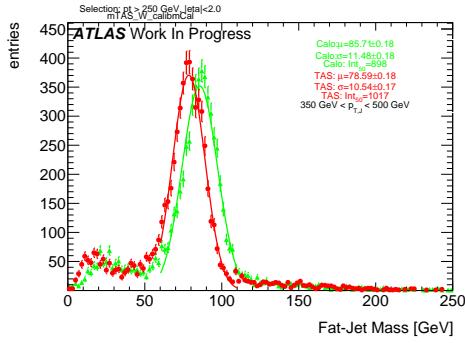
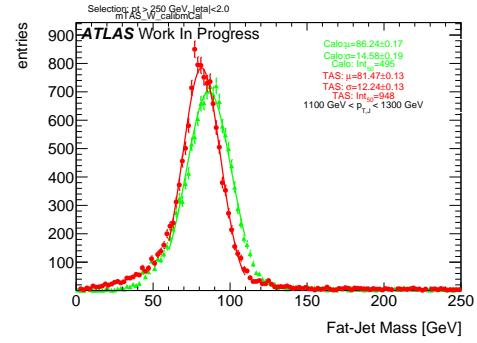
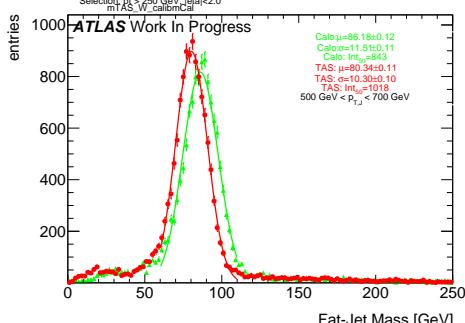
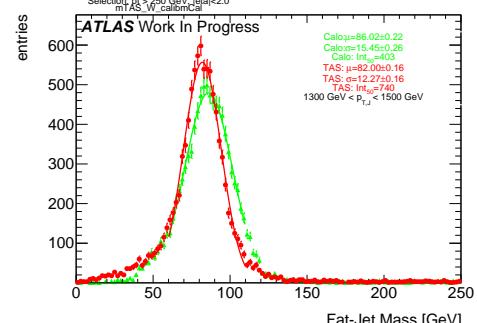
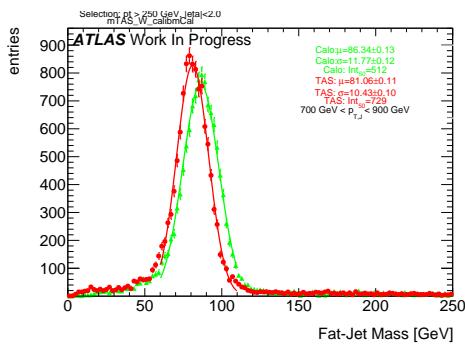
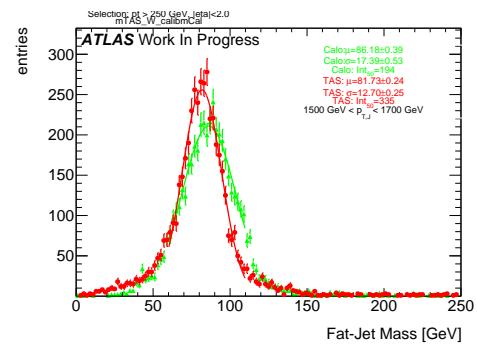


Figure 50: Comparison of the m^{TAS} and the same variable using truth-level information for the tracks.

₇₇₂ A m^{TAS} distributions, boosted W/Z

Figure 51: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 55: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 52: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 56: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)Figure 53: 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 54: 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)

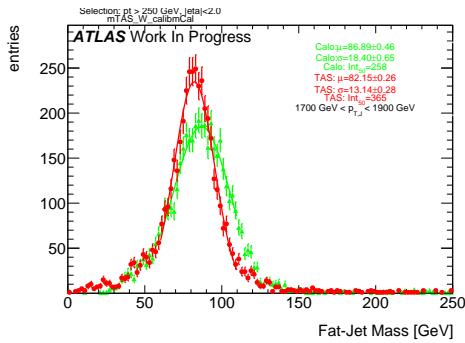


Figure 59: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

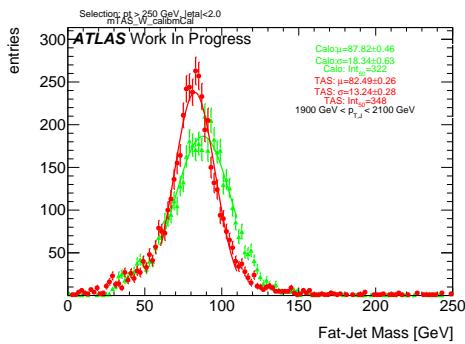


Figure 60: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

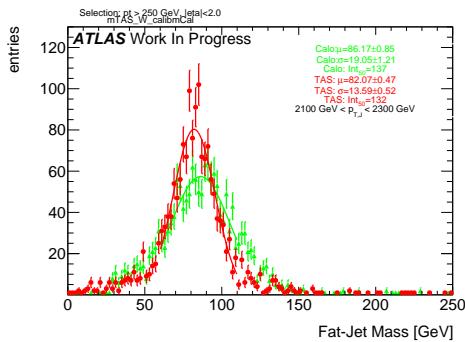


Figure 61: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

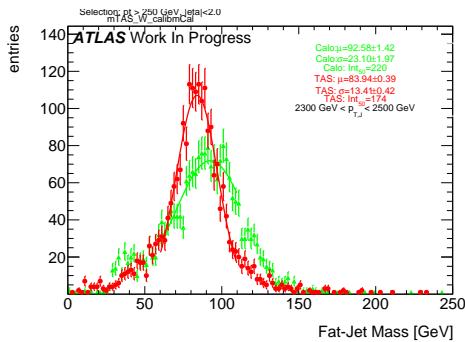


Figure 62: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

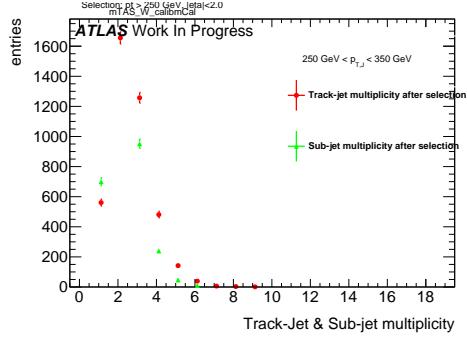


Figure 63: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

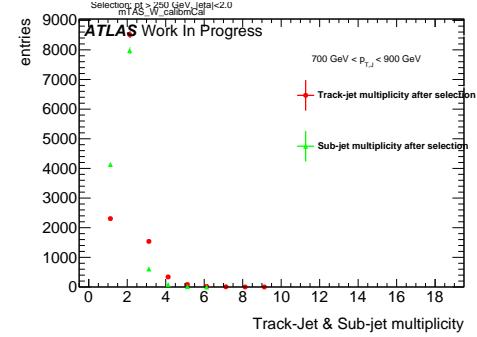


Figure 66: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

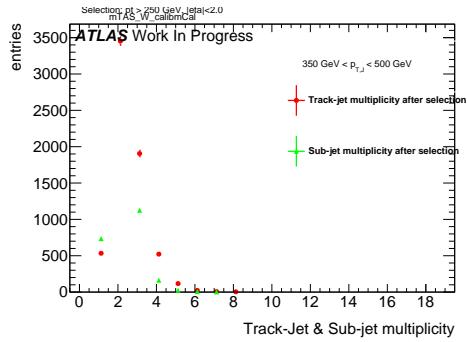


Figure 64: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

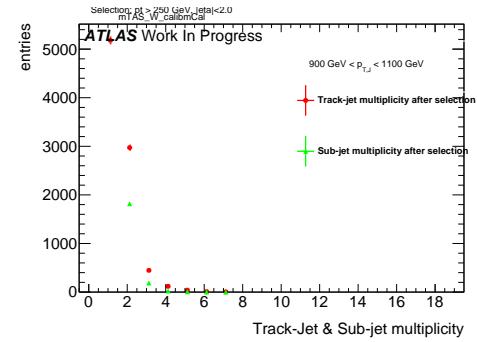


Figure 67: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

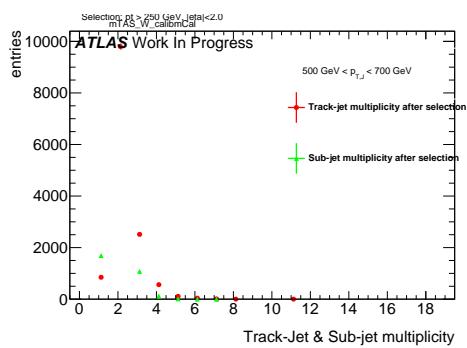


Figure 65: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

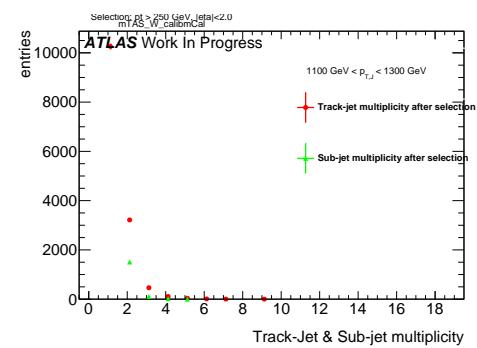


Figure 68: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

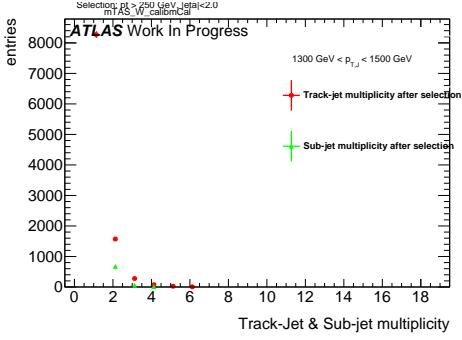


Figure 69: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

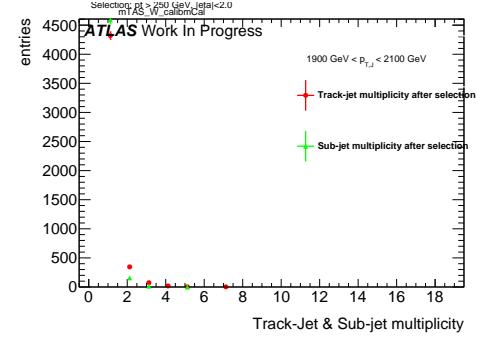


Figure 72: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

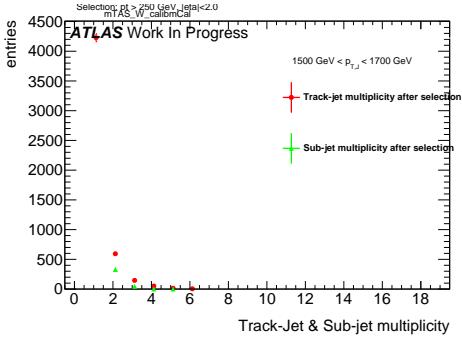


Figure 70: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

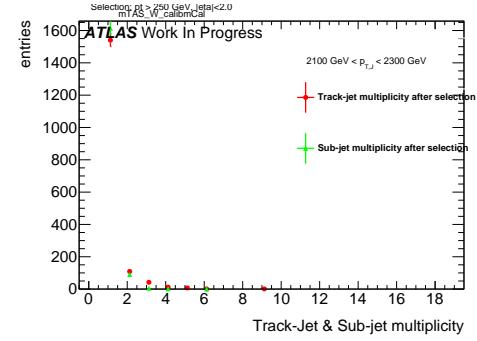


Figure 73: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

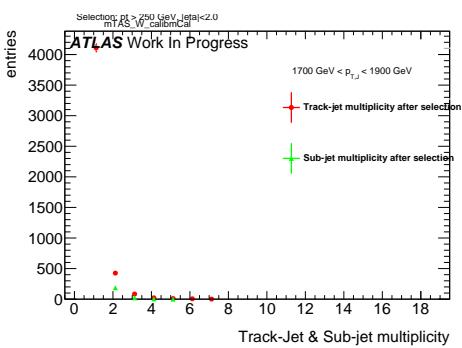


Figure 71: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

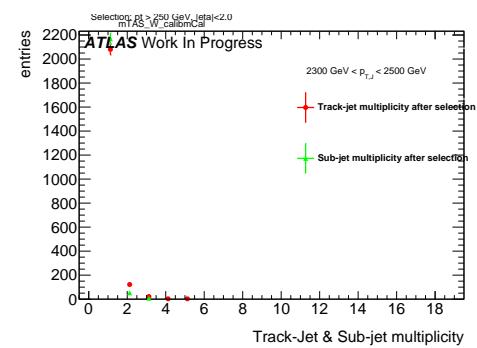


Figure 74: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

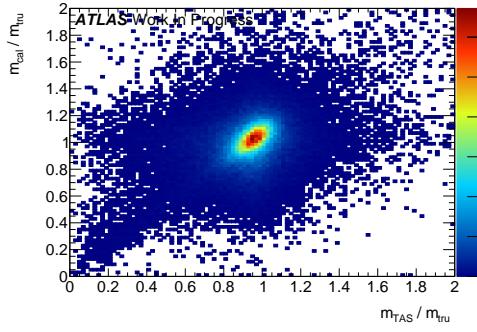
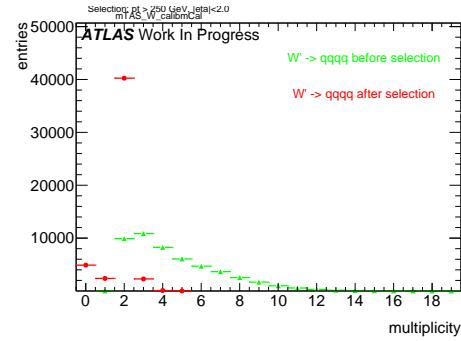
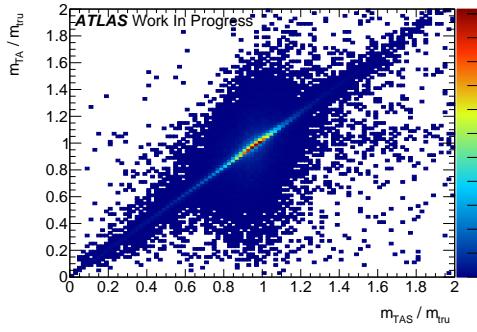
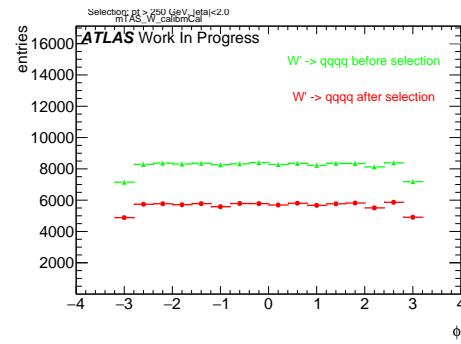
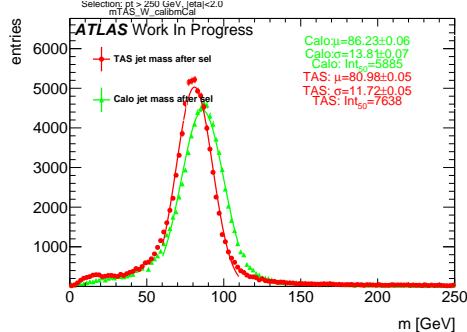
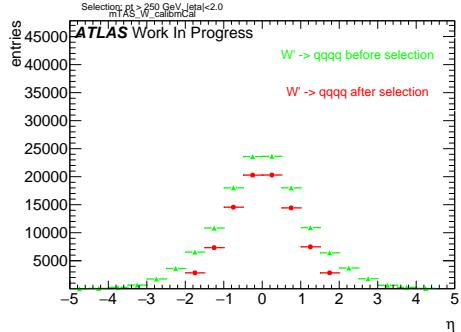
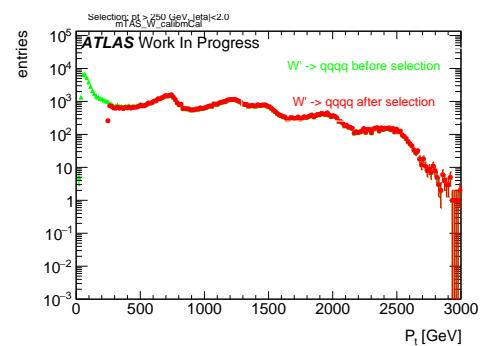
Figure 75: Scatter plot m^{TAS} versus m^{calo} responses

Figure 79: large-R jet Multiplicity, before and after selection

Figure 76: Scatter plot m^{TAS} versus m^{TA} responsesFigure 80: ϕ distribution of the large-R jet, before and after selectionFigure 77: m^{TAS} distribution in all the p_T binsFigure 78: η distribution of the large-R jet, before and after selectionFigure 81: p_T distribution of the large-R jet, before and after selection

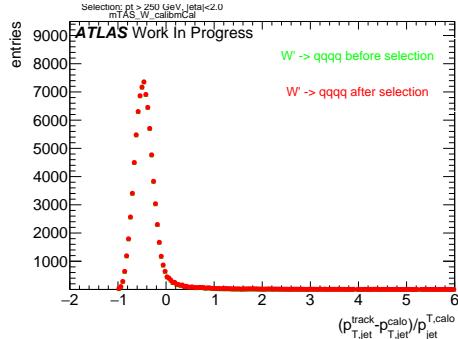


Figure 82: p_T resolution: $\frac{p_{T,\text{jet}}^{\text{track}} - p_{T,\text{jet}}^{\text{fat}}}{p_{T,\text{jet}}^{\text{fat}}}$, before and after selection

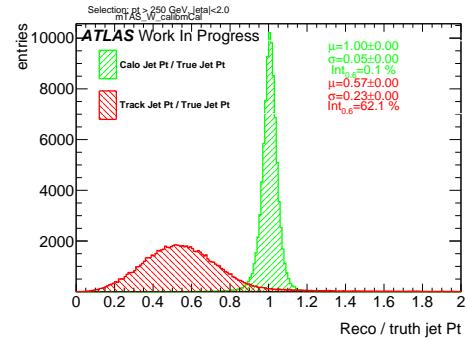


Figure 85: Transverse momentum response $p_T^{\text{Reco}} / p_T^{\text{Truth}}$ for calorimeter and tracks

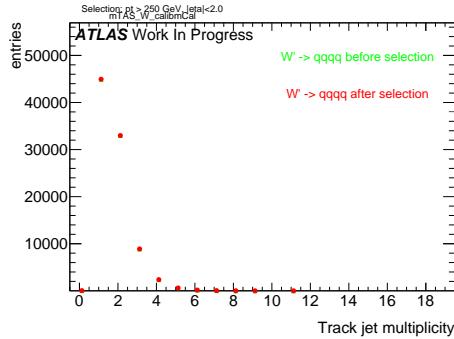


Figure 83: Multiplicity of track-jets $R=0.2$ per large-R jet

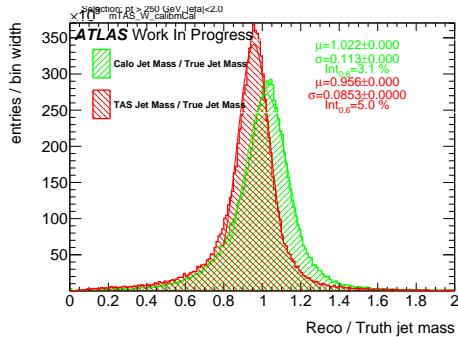


Figure 84: Response $m^{\text{Reco}} / m^{\text{Truth}}$ for all the p_T bins

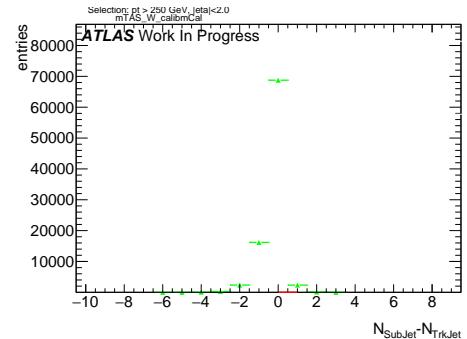


Figure 86: sub-jet - track-jet Multiplicity

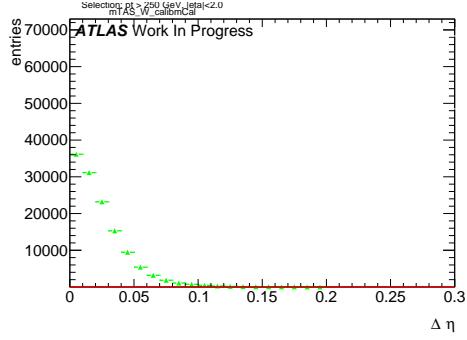


Figure 87: $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

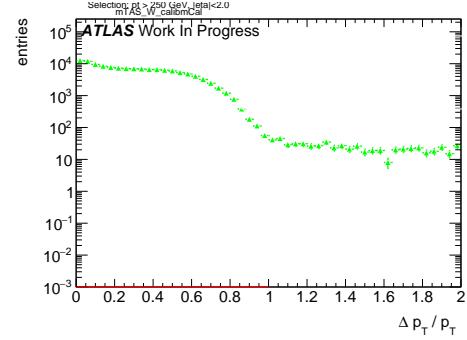


Figure 90: $|p_{T,\text{sub-jet}} - p_{T,\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

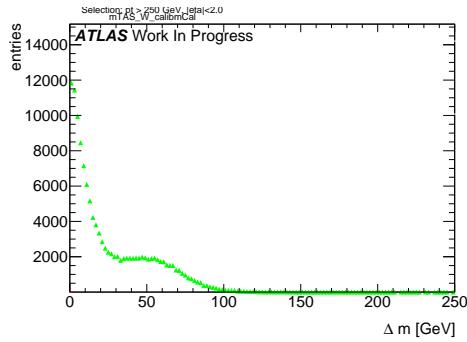


Figure 88: $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

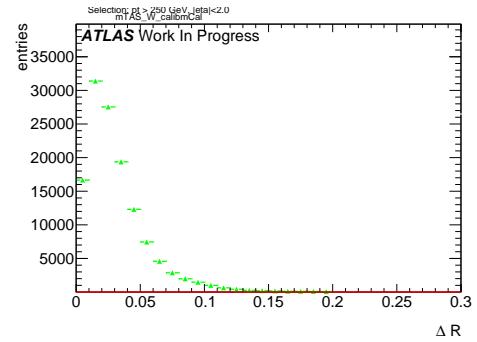


Figure 91: $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

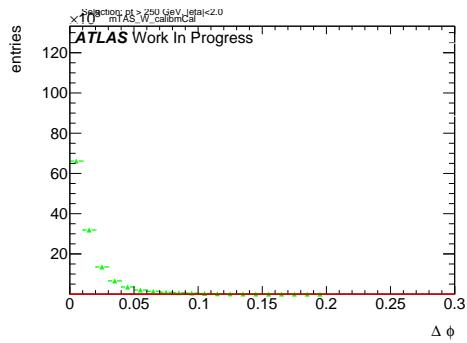


Figure 89: $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

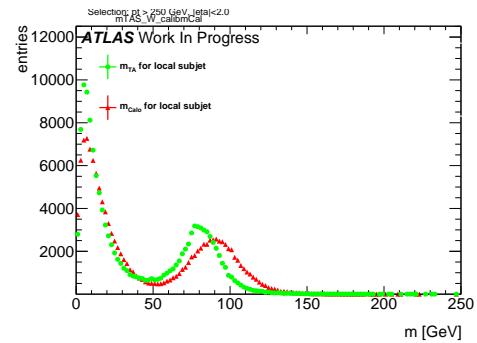


Figure 92: Mass distribution of the sub-jet, calorimeter and track-assisted

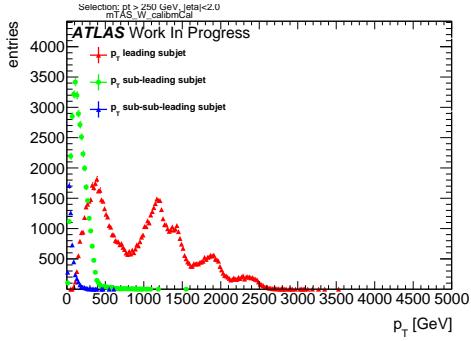


Figure 93: p_T distribution for leading, sub-leading and sub-sub-leading sub-jets

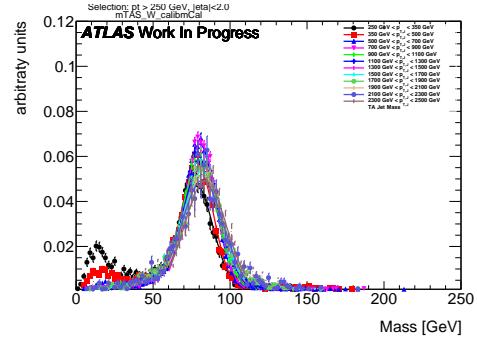


Figure 96: m^{TAS} for p_T^J bin, superimposed

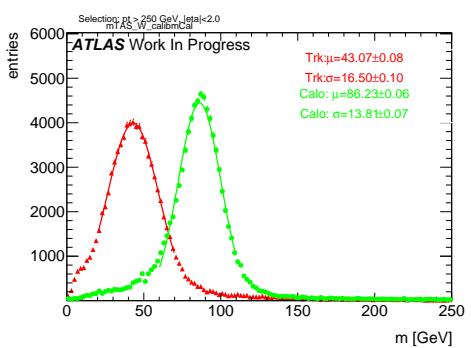


Figure 94: Mass distribution for calorimeter and tracks associated to the large- R jet

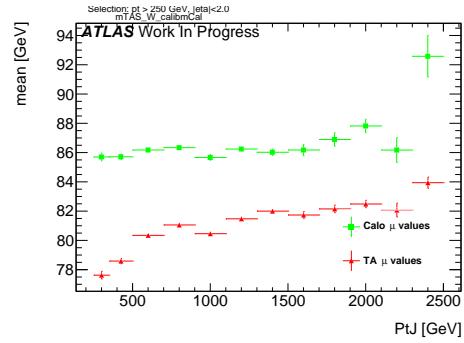


Figure 97: μ from fit of the mass distribution vs bin of p_T^J

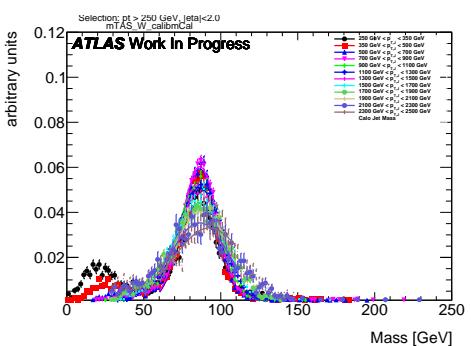


Figure 95: m^{calo} for p_T^J bin, superimposed

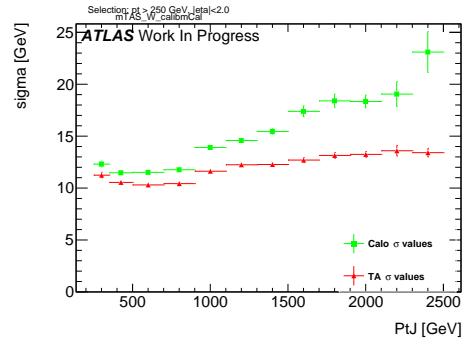


Figure 98: σ from fit of the mass distribution vs bin of p_T^J

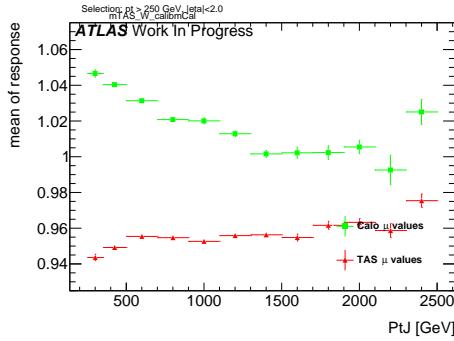


Figure 99: μ from fit of the mass Response vs bin of p_T^J

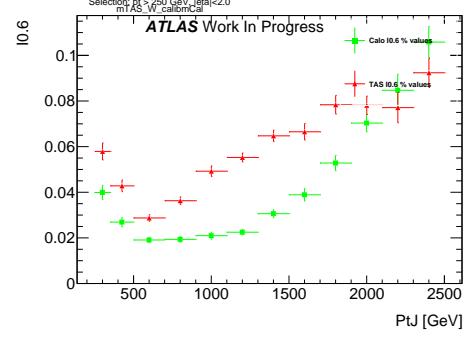


Figure 102: Left integral normalized, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

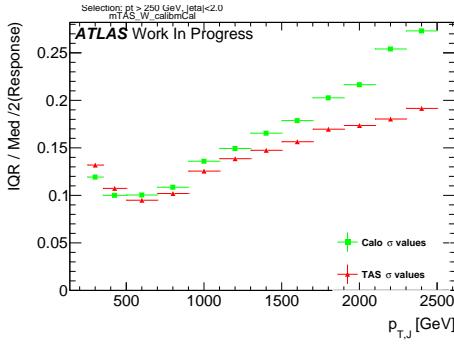


Figure 100: σ from fit of the mass Response vs bin of p_T^J

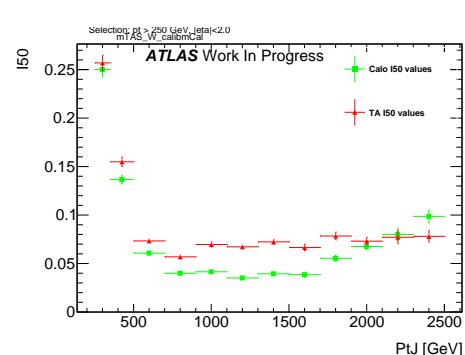


Figure 103: $\int_0^{50\text{GeV}}$ from fit of the mass distribution vs bin of p_T^J (normalized)

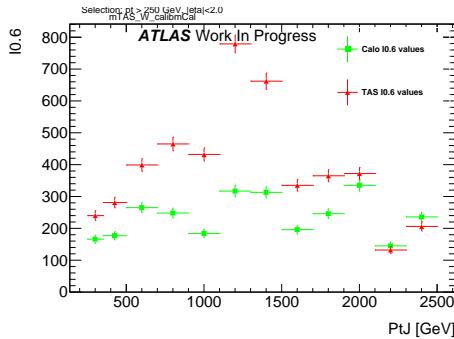


Figure 101: Left integral, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

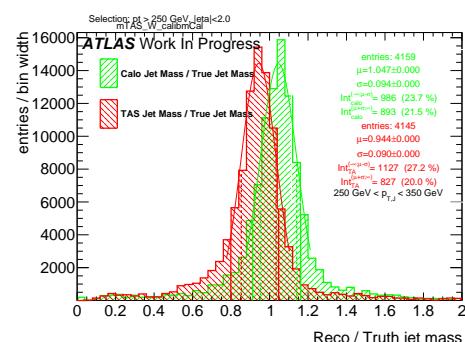
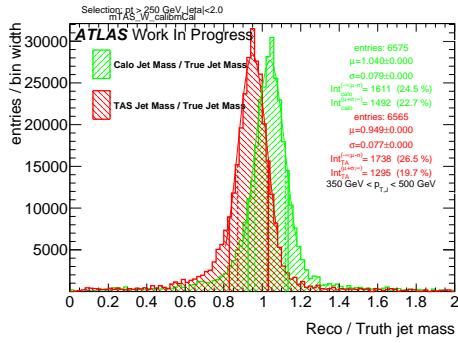
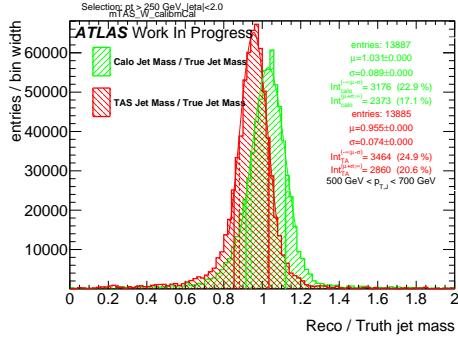
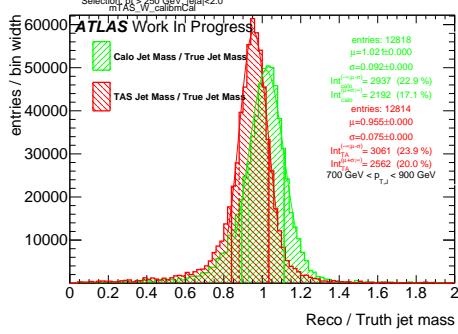
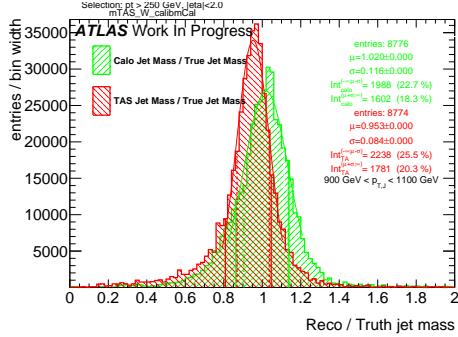
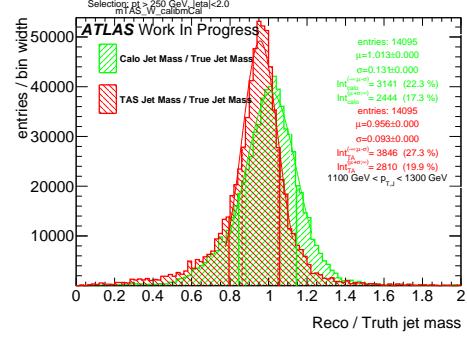
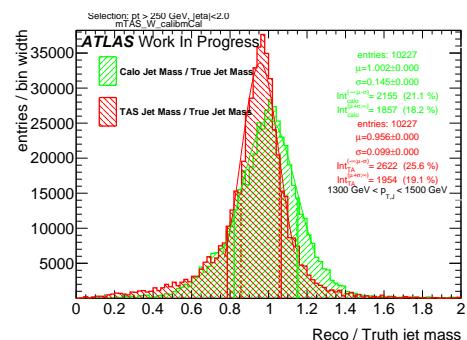
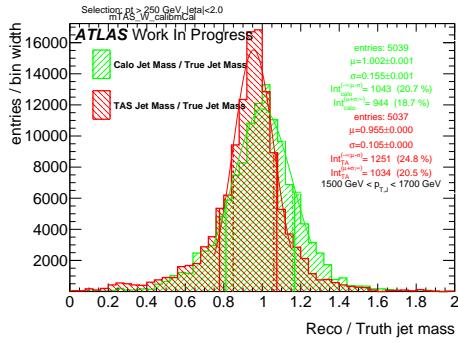
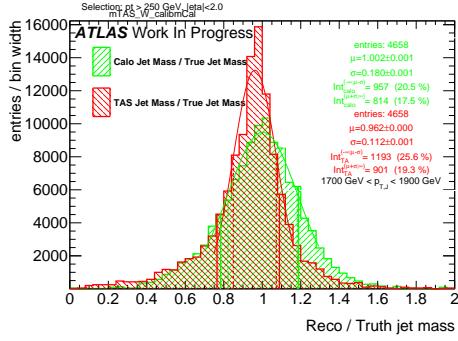
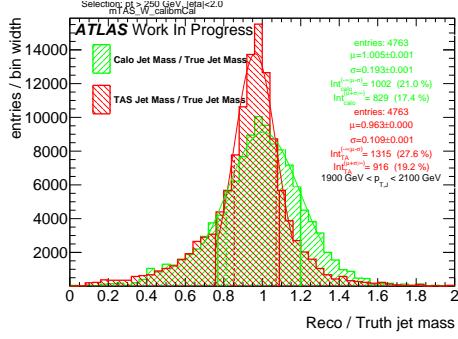
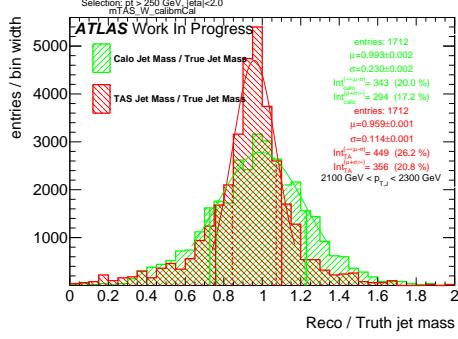
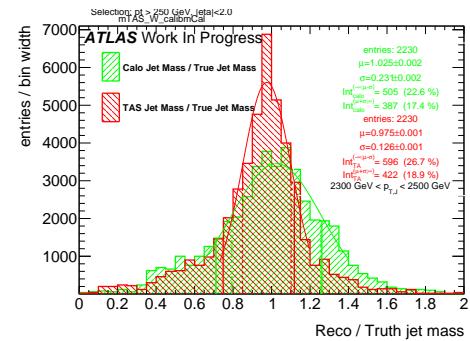


Figure 104: Response in bin of p_T^J (indicated on plot)

Figure 105: Response in bin of p_T^J (indicated on plot)Figure 106: Response in bin of p_T^J (indicated on plot)Figure 107: Response in bin of p_T^J (indicated on plot)Figure 108: Response in bin of p_T^J (indicated on plot)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)

₇₇₃ **B m^{TAS} distributions, boosted tops**

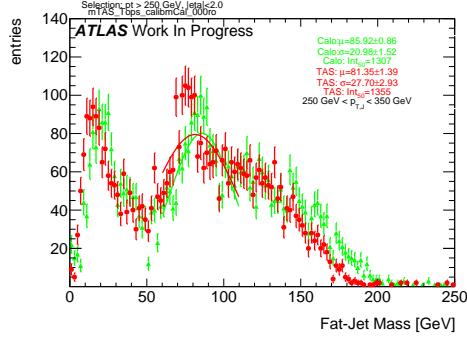


Figure 116: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

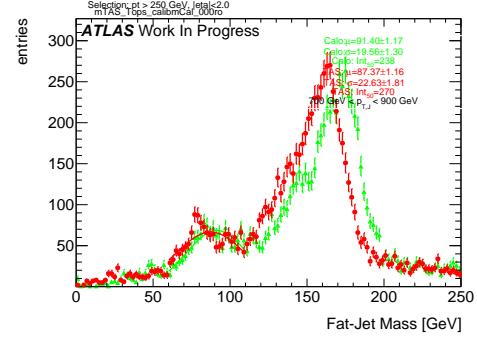


Figure 119: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

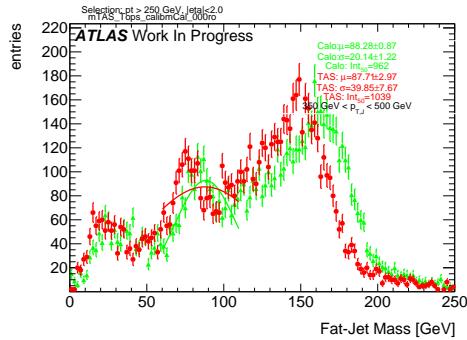


Figure 117: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

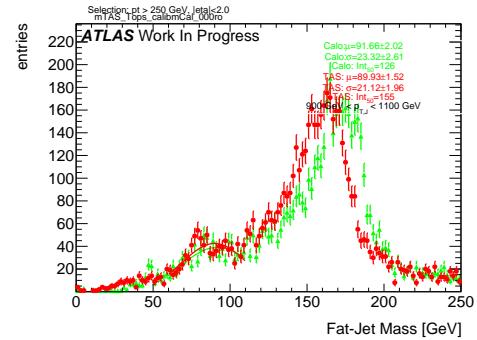


Figure 120: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

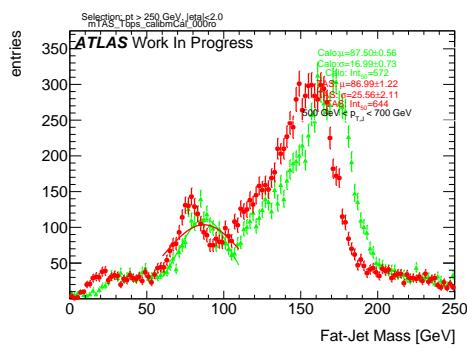


Figure 118: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

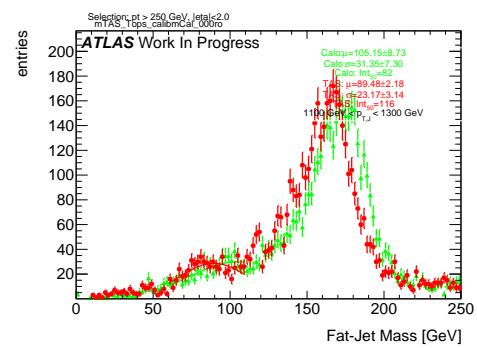


Figure 121: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

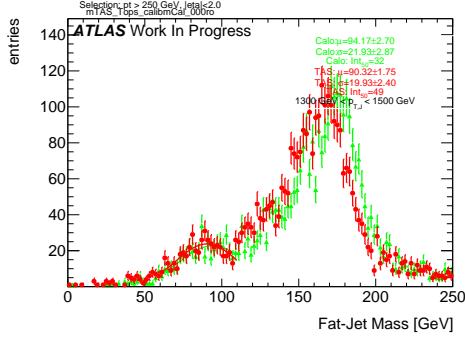


Figure 122: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

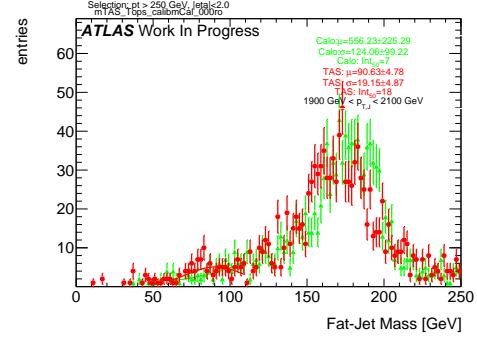


Figure 125: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

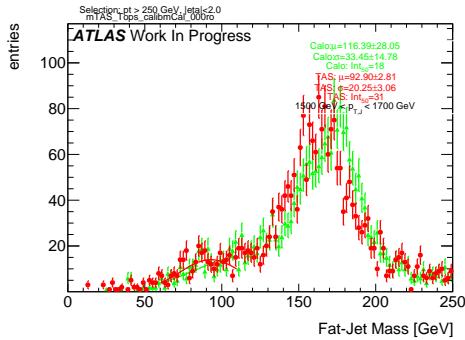


Figure 123: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

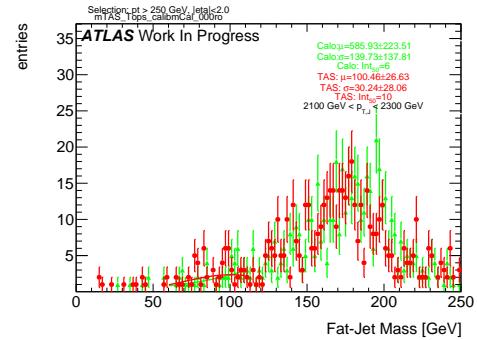


Figure 126: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

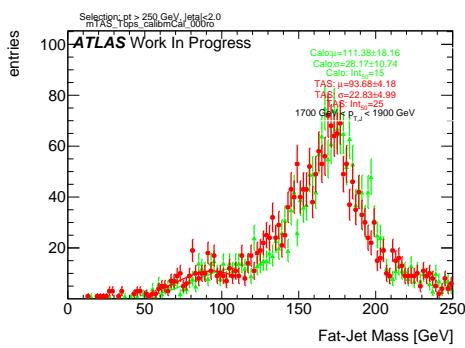


Figure 124: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

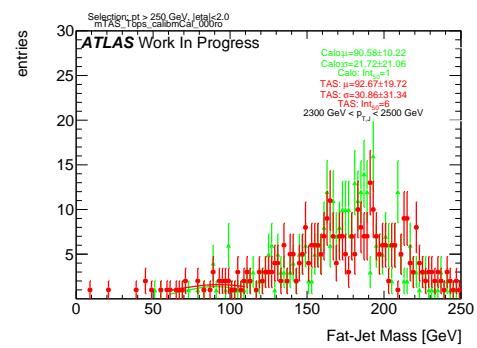


Figure 127: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

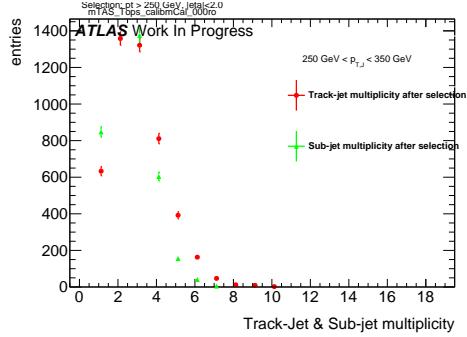


Figure 128: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

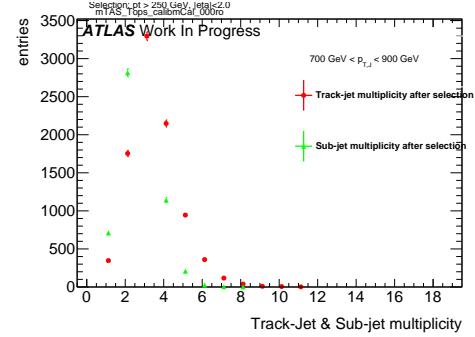


Figure 131: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

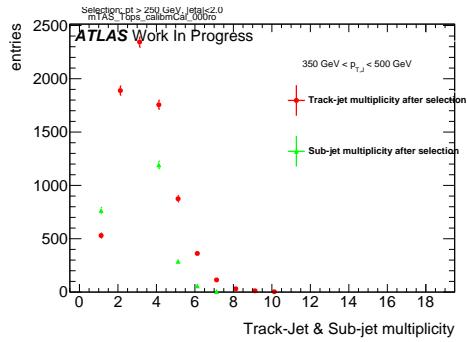


Figure 129: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

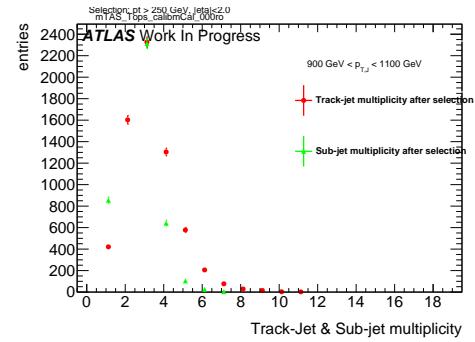


Figure 132: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

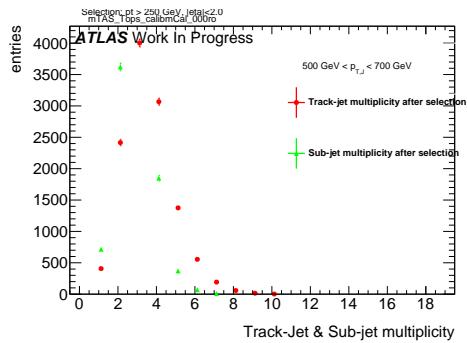


Figure 130: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

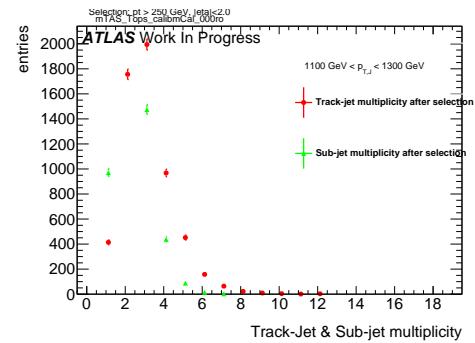


Figure 133: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

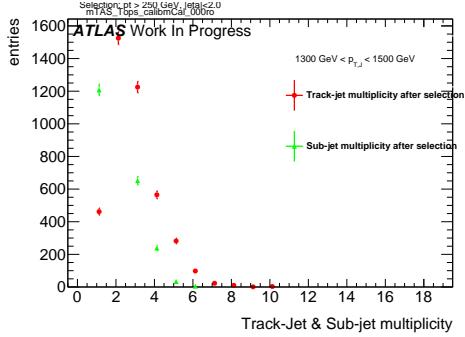


Figure 134: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

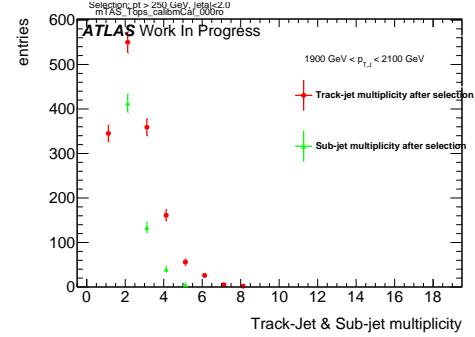


Figure 137: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

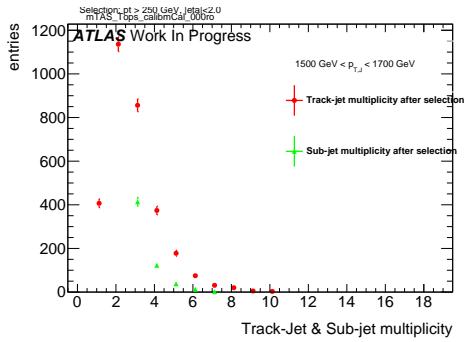


Figure 135: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

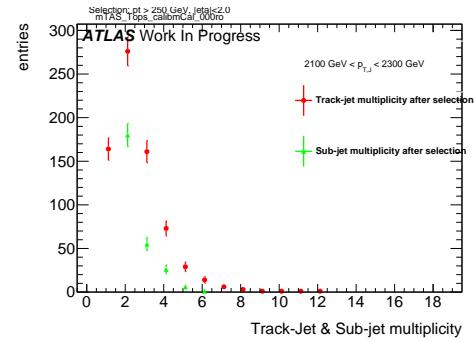


Figure 138: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

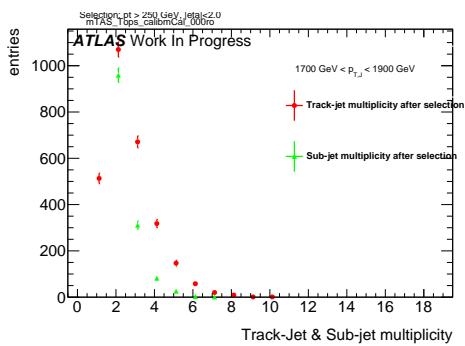


Figure 136: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

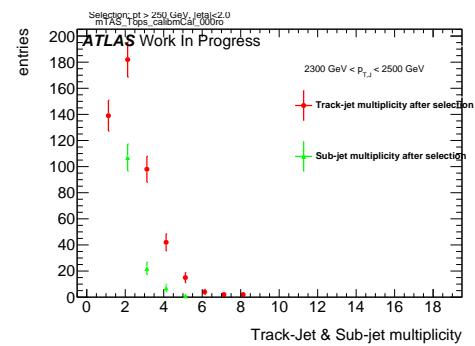


Figure 139: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

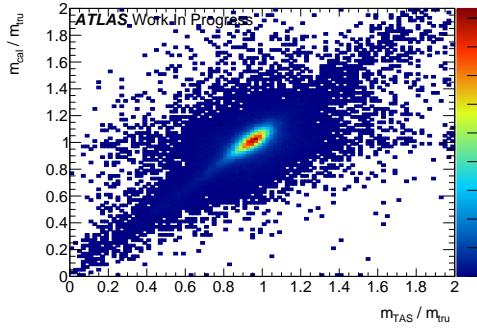
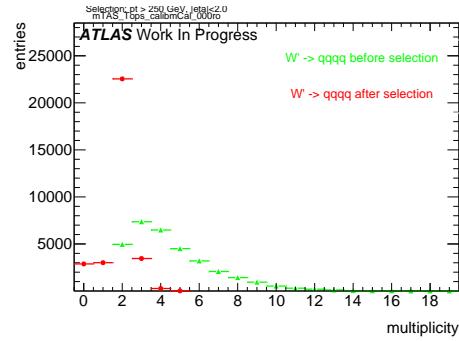
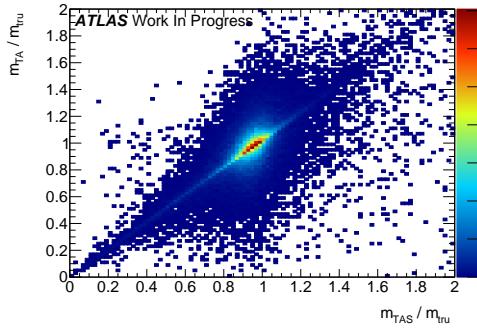
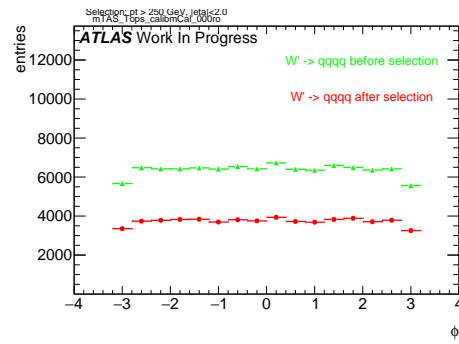
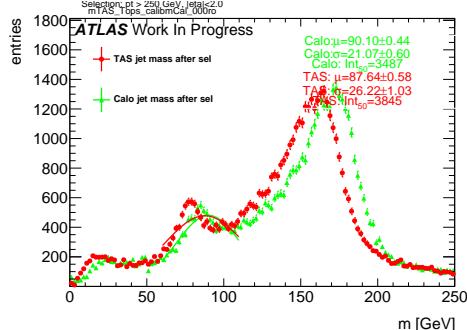
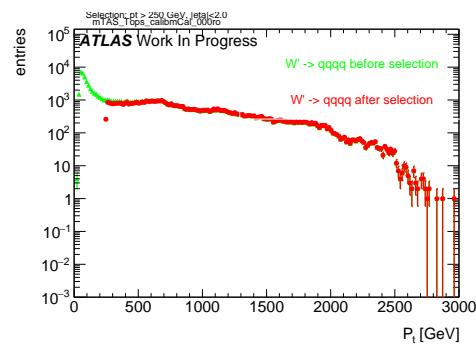
Figure 140: Scatter plot m^{TAS} versus m^{calo} responses

Figure 144: large-R jet Multiplicity, before and after selection

Figure 141: Scatter plot m^{TAS} versus m^A responsesFigure 145: ϕ distribution of the large-R jet, before and after selectionFigure 142: m^{TAS} distribution in all the p_T binsFigure 146: p_T distribution of the large-R jet, before and after selectionFigure 143: η distribution of the large-R jet, before and after selection

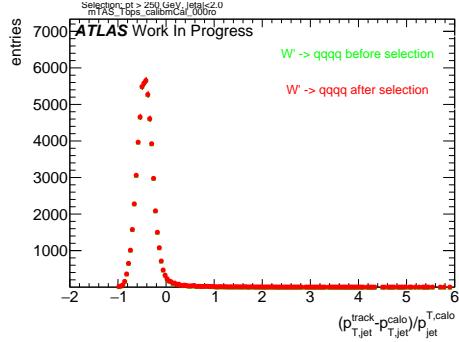


Figure 147: p_T resolution: $\frac{p_{T,jet}^{track} - p_{T,jet}^{calo}}{p_{T,jet}^{calo}}$, before and after selection

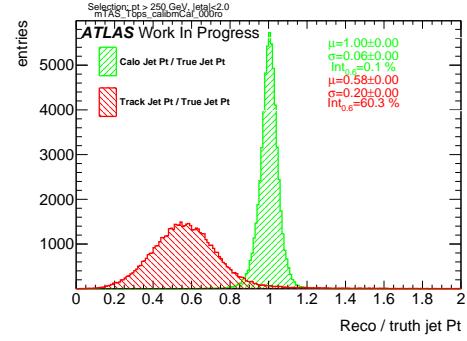


Figure 150: Transverse momentum response $p_T^{\text{Reco}} / p_T^{\text{Truth}}$ for calorimeter and tracks

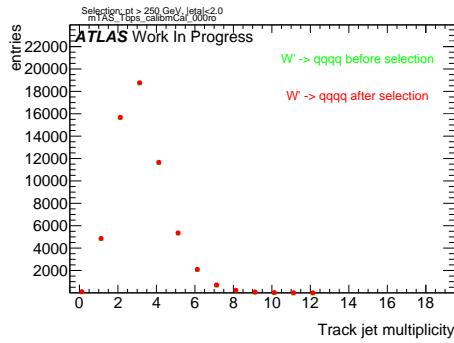


Figure 148: Multiplicity of track-jets R=0.2 per large-R jet

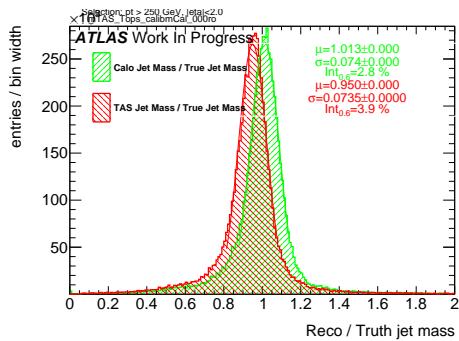


Figure 149: Response $m^{\text{Reco}} / m^{\text{Truth}}$ for all the p_T bins

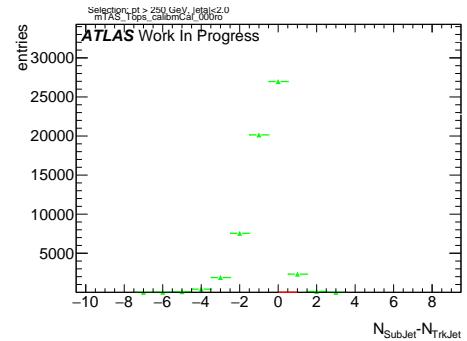


Figure 151: sub-jet - track-jet Multiplicity

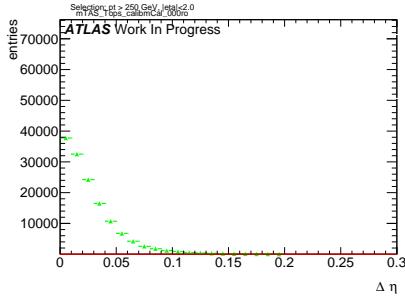


Figure 152: $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

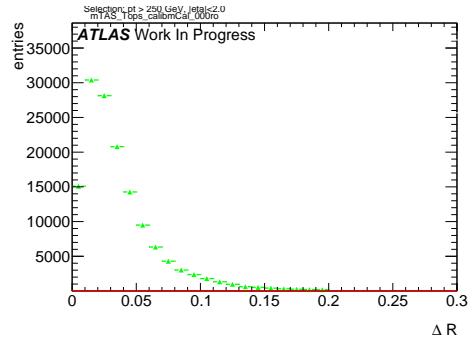


Figure 156: $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

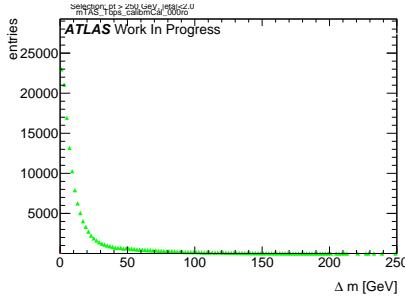


Figure 153: $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

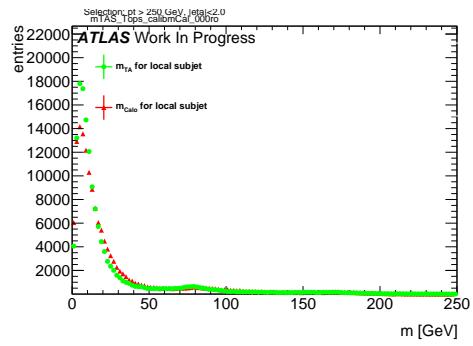


Figure 157: Mass distribution of the sub-jet, calorimeter and track-assisted

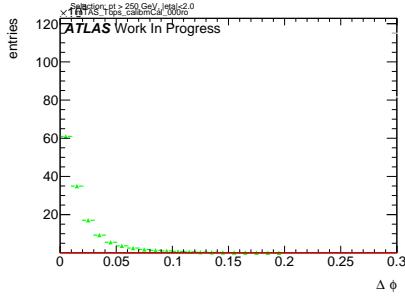


Figure 154: $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

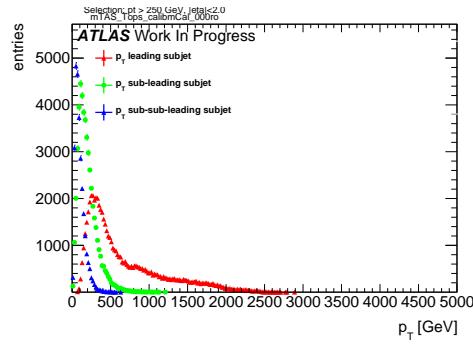


Figure 158: p_T distribution for leading, sub-leading and sub-sub-leading sub-jets

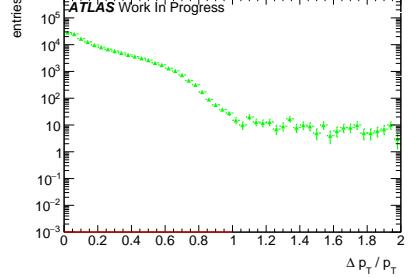


Figure 155: $|p_{T,\text{sub-jet}} - p_{T,\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

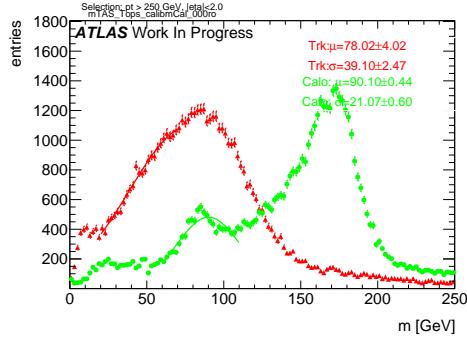


Figure 159: Mass distribution for calorimeter and tracks associated to the large-R jet

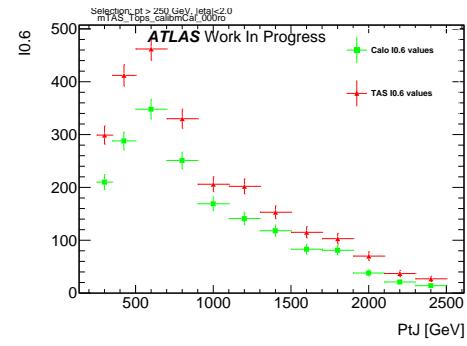


Figure 162: Left integral, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

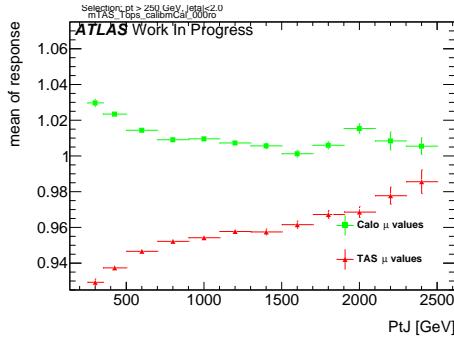


Figure 160: μ from fit of the mass Response vs bin of p_T^J

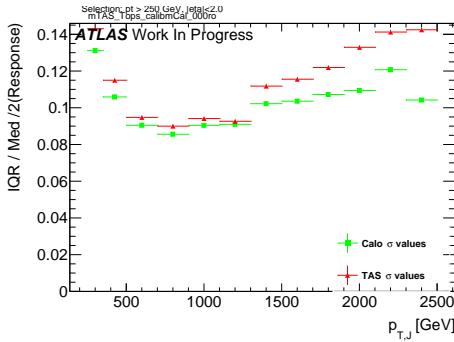


Figure 161: σ from fit of the mass Response vs bin of p_T^J

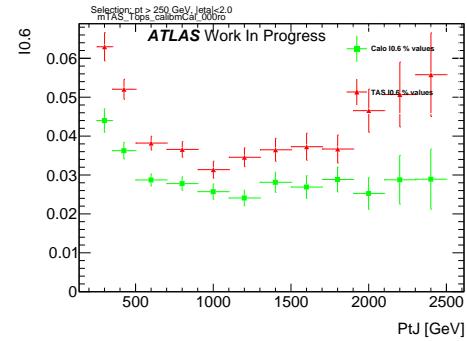
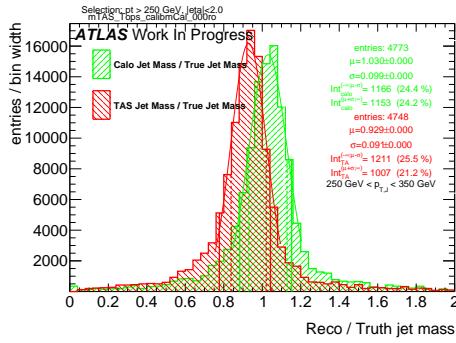
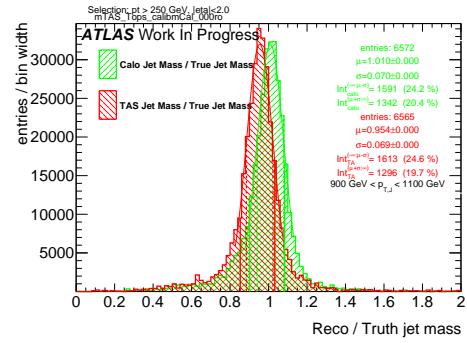
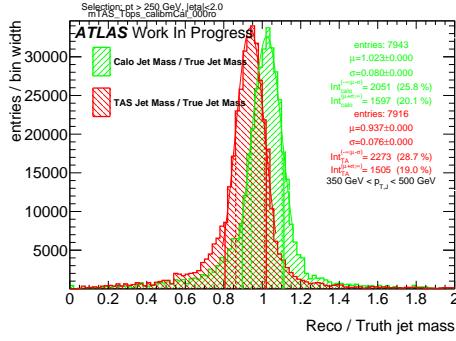
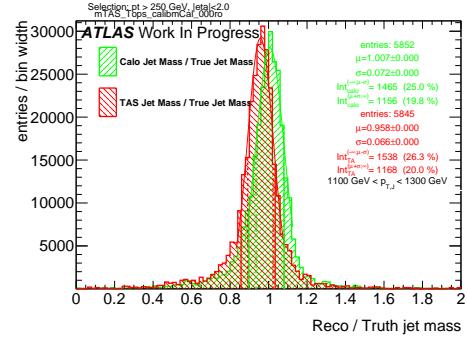
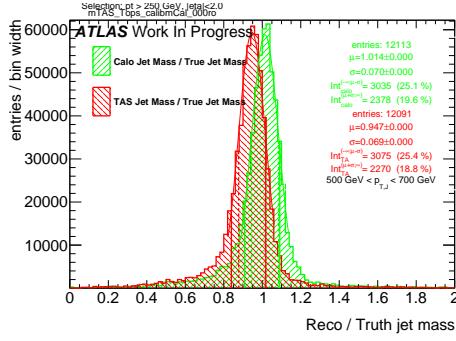
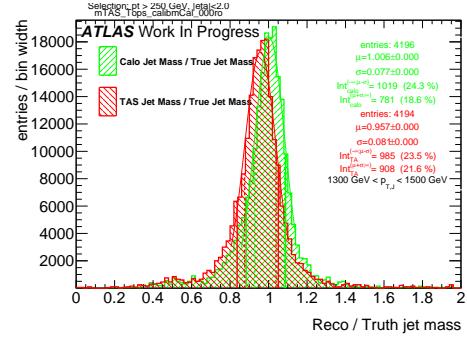
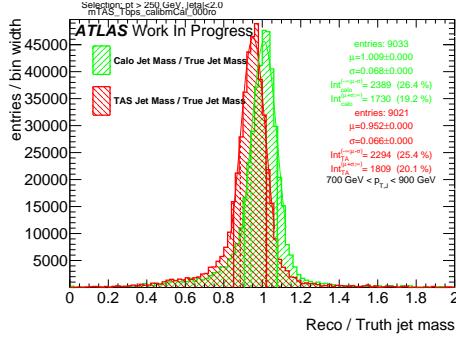
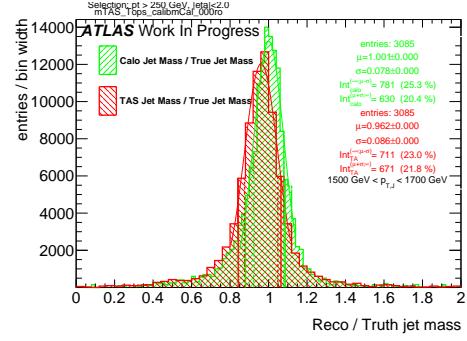
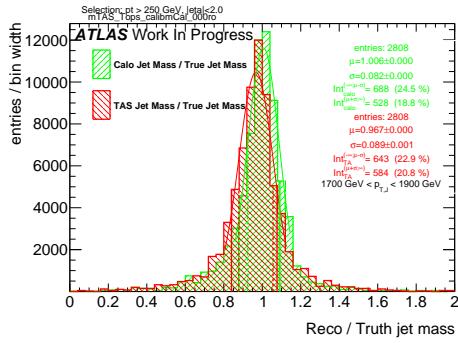
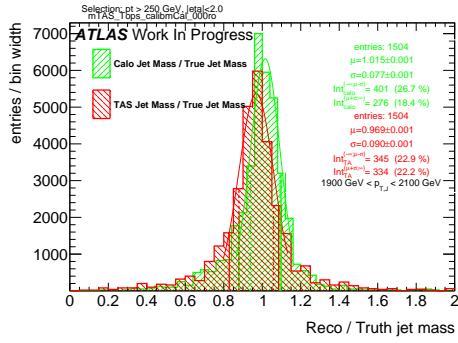
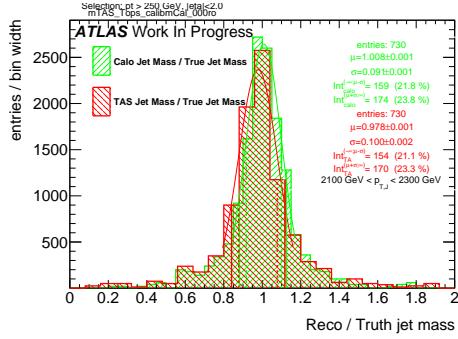
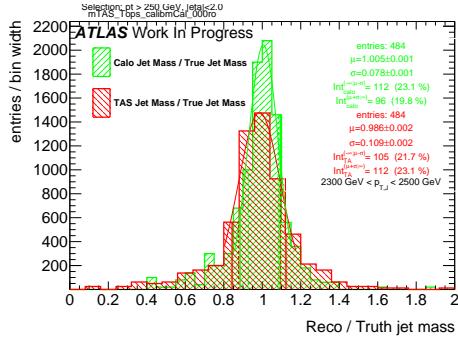


Figure 163: Left integral normalized, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

Figure 164: Response in bin of p_T^J (indicated on plot)Figure 168: Response in bin of p_T^J (indicated on plot)Figure 165: Response in bin of p_T^J (indicated on plot)Figure 169: Response in bin of p_T^J (indicated on plot)Figure 166: Response in bin of p_T^J (indicated on plot)Figure 170: Response in bin of p_T^J (indicated on plot)Figure 167: Response in bin of p_T^J (indicated on plot)Figure 171: Response in bin of p_T^J (indicated on plot)

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

⁷⁷⁴ **C m^{TAS} distributions, boosted higgs**

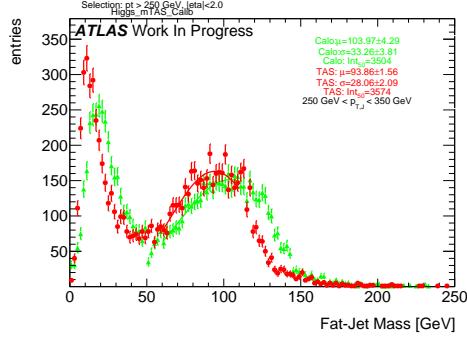


Figure 176: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

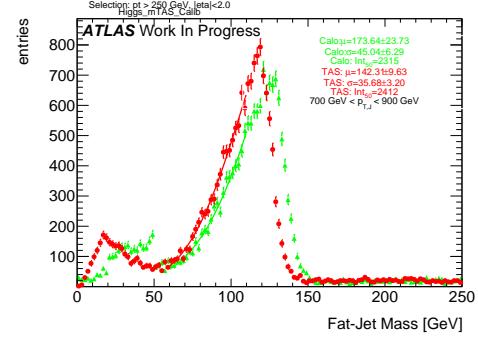


Figure 179: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

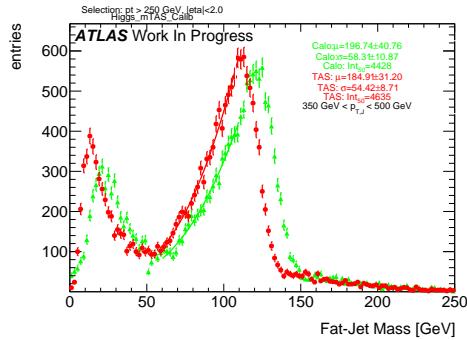


Figure 177: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

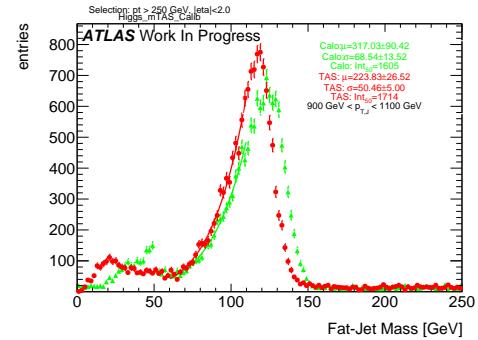


Figure 180: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

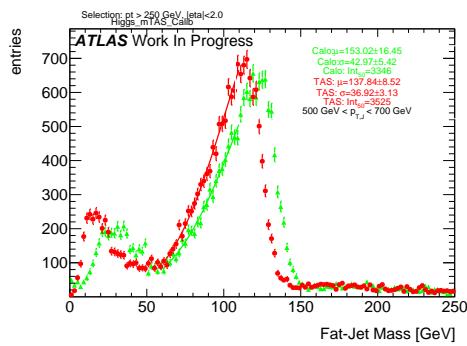


Figure 178: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

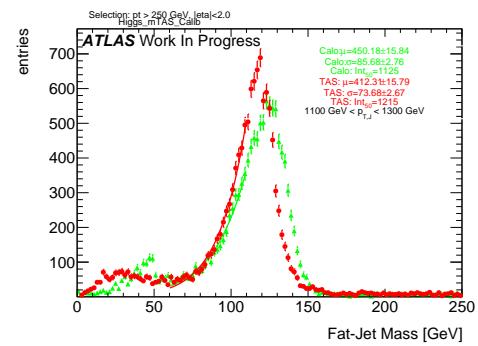


Figure 181: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

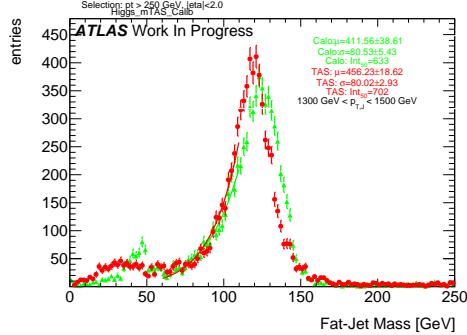


Figure 182: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

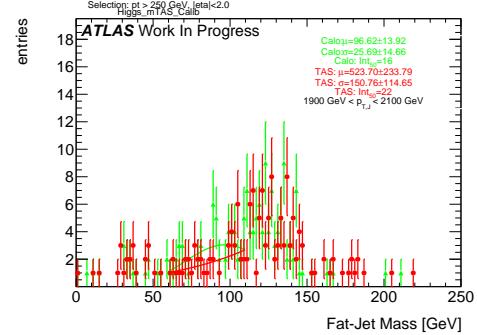


Figure 185: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

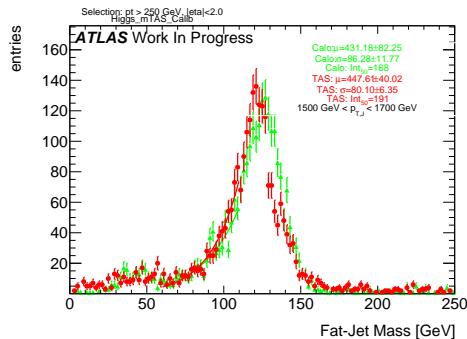


Figure 183: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

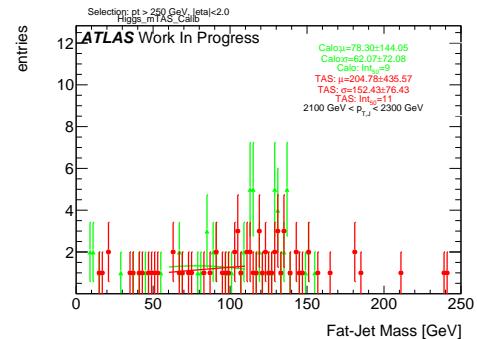


Figure 186: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

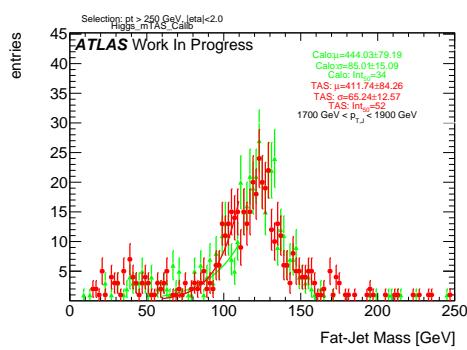


Figure 184: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

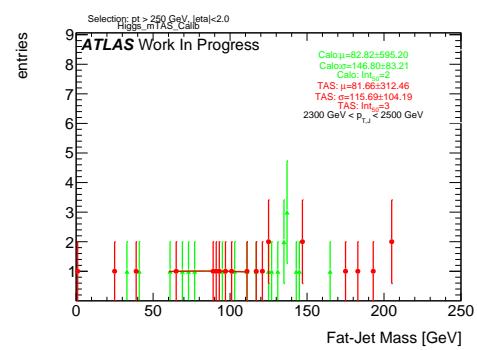


Figure 187: m^{TAS} and m^{calo} for p_T^J bin (indicated on plot)

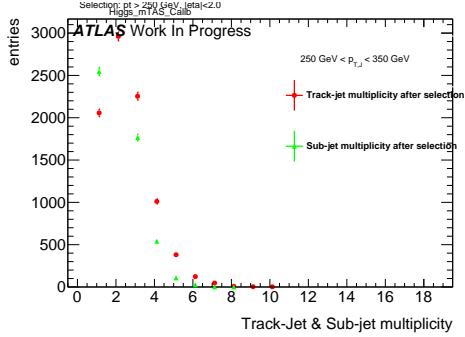


Figure 188: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

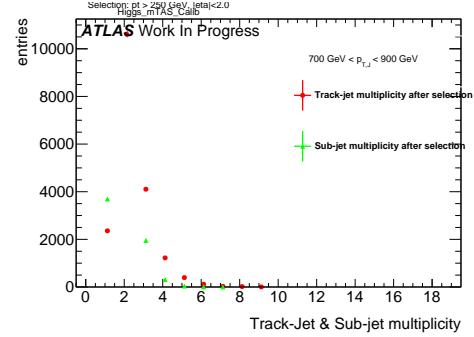


Figure 191: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

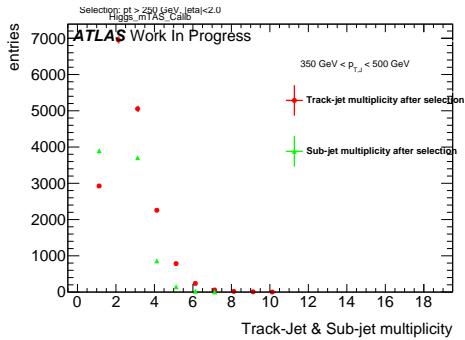


Figure 189: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

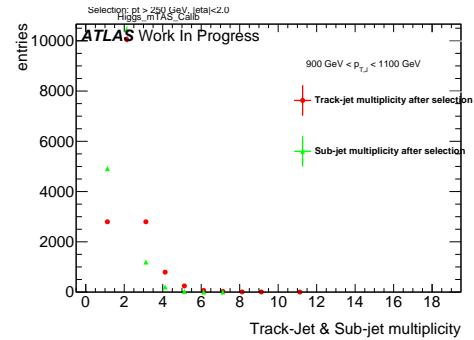


Figure 192: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

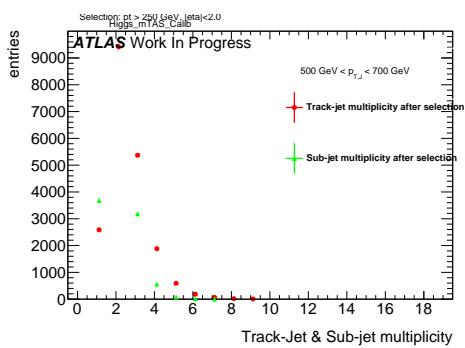


Figure 190: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

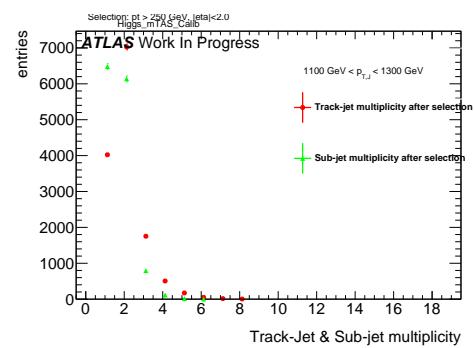


Figure 193: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

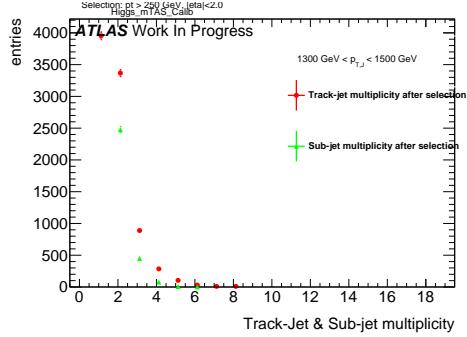


Figure 194: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

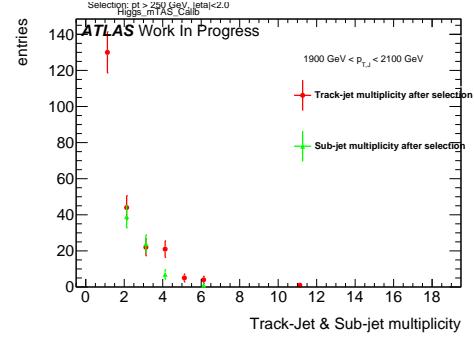


Figure 197: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

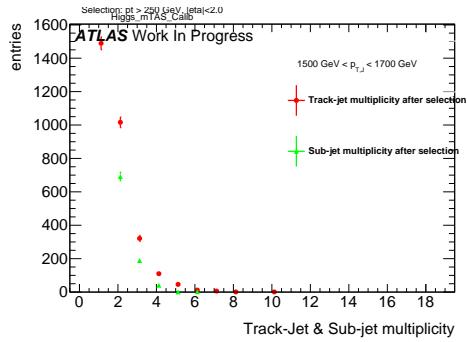


Figure 195: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

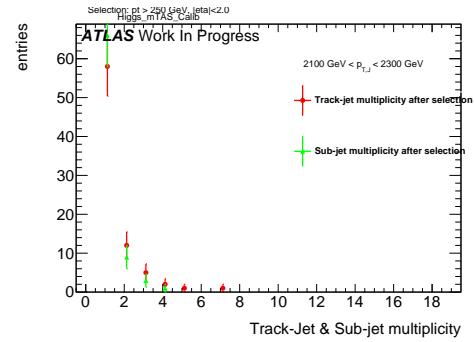


Figure 198: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

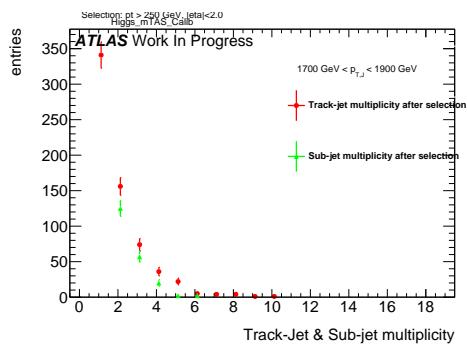


Figure 196: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

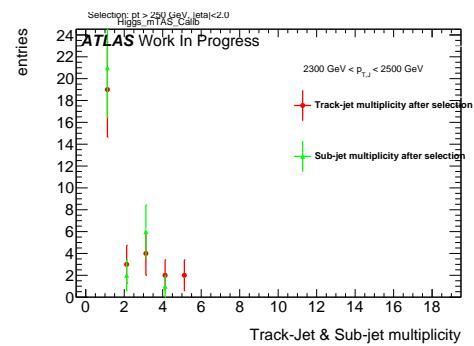


Figure 199: Track-jet R=0.2 and sub-jet multiplicity for p_T^J bin (indicated on plot)

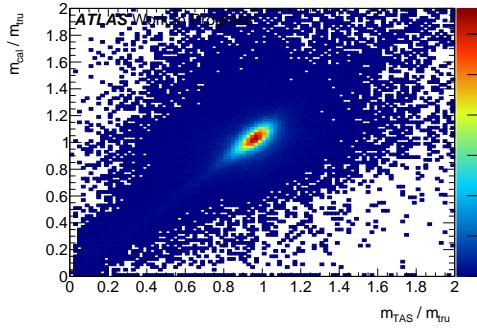
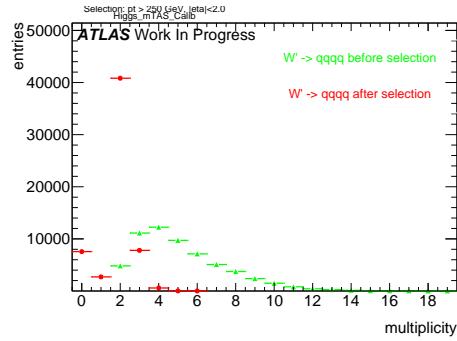
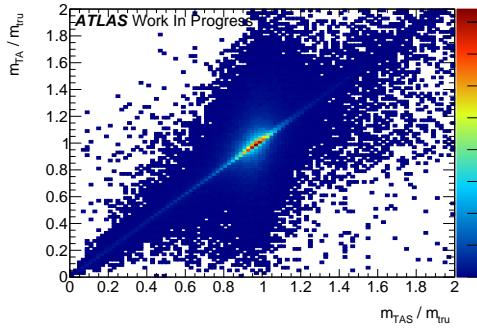
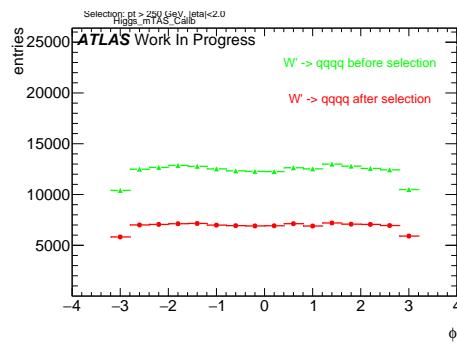
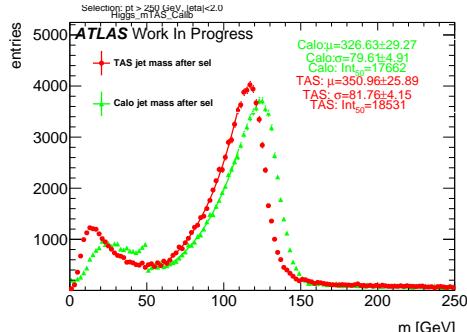
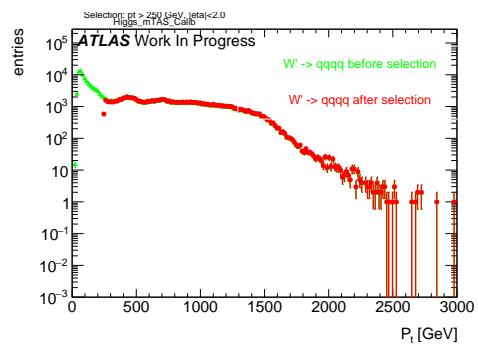
Figure 200: Scatter plot m^{TAS} versus m^{calo} responses

Figure 204: large-R jet Multiplicity, before and after selection

Figure 201: Scatter plot m^{TAS} versus m^A responsesFigure 205: ϕ distribution of the large-R jet, before and after selectionFigure 202: m^{TAS} distribution in all the p_T binsFigure 206: p_T distribution of the large-R jet, before and after selectionFigure 203: η distribution of the large-R jet, before and after selection

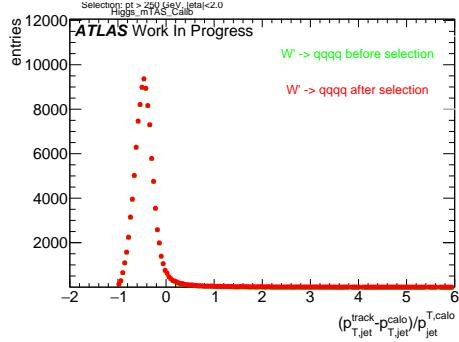


Figure 207: p_T resolution: $\frac{p_{T,jet}^{track} - p_{T,jet}^{fat}}{p_{T,jet}^{fat}}$, before and after selection

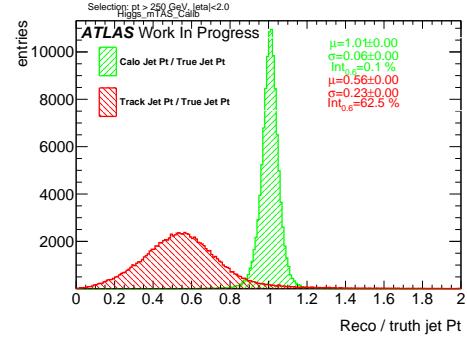


Figure 210: Transverse momentum response p_T^{Reco} / p_T^{Truth} for calorimeter and tracks

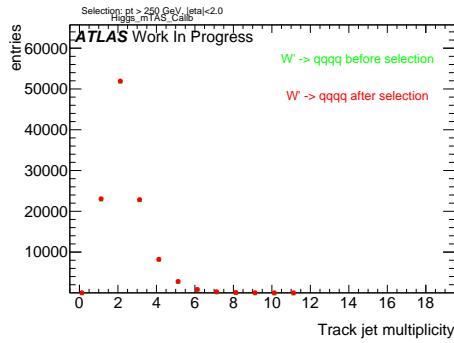


Figure 208: Multiplicity of track-jets R=0.2 per large-R jet

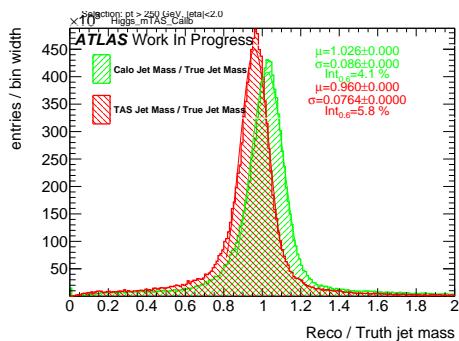


Figure 209: Response m^{Reco} / m^{Truth} for all the p_T bins

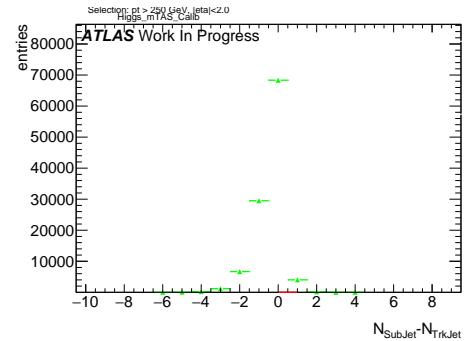


Figure 211: sub-jet - track-jet Multiplicity

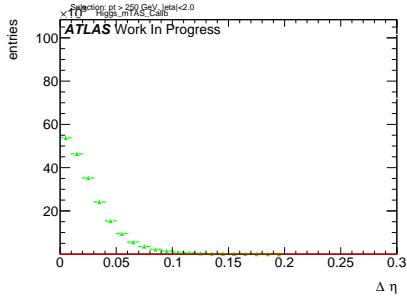


Figure 212: $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

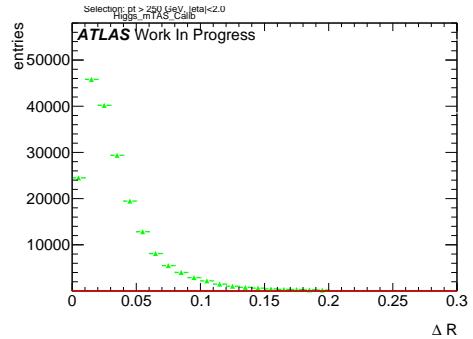


Figure 216: $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

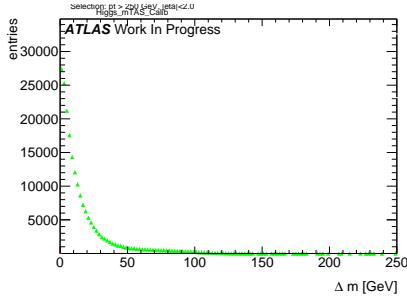


Figure 213: $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

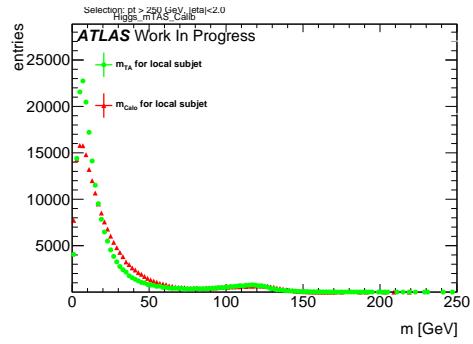


Figure 217: Mass distribution of the sub-jet, calorimeter and track-assisted

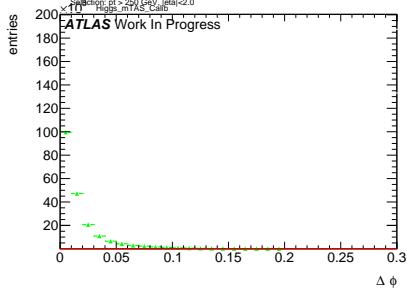


Figure 214: $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$ distribution, where sub-jet and track-jet are the closest

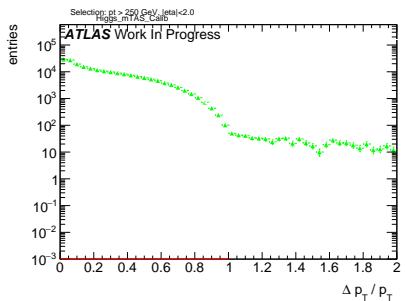


Figure 215: $|p_{\text{T,sub-jet}} - p_{\text{T,track-jet}}|$ distribution, where sub-jet and track-jet are the closest

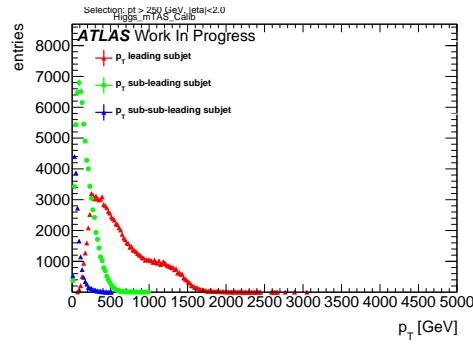


Figure 218: p_{T} distribution for leading, sub-leading and sub-sub-leading sub-jets

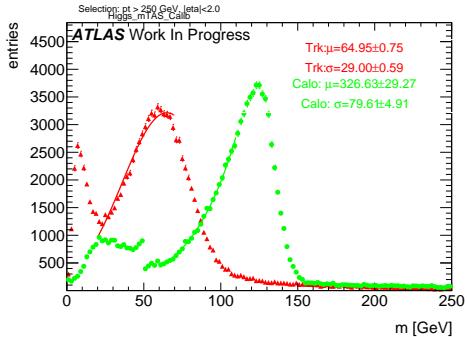


Figure 219: Mass distribution for calorimeter and tracks associated to the large-R jet

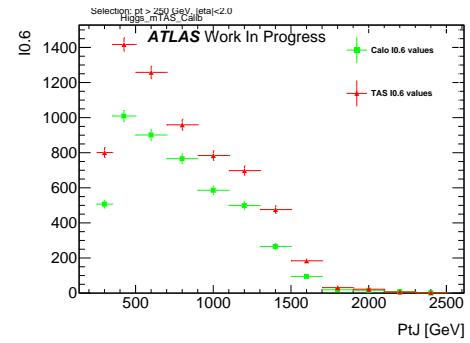


Figure 222: Left integral, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

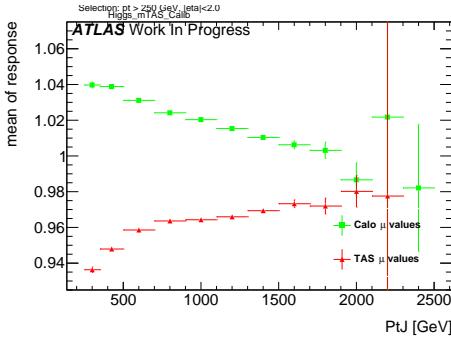


Figure 220: μ from fit of the mass Response vs bin of p_T^J

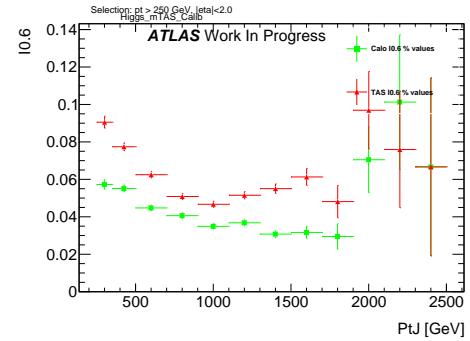


Figure 223: Left integral normalized, $\int_0^{0.6}$ of the mass response, vs bin of p_T^J

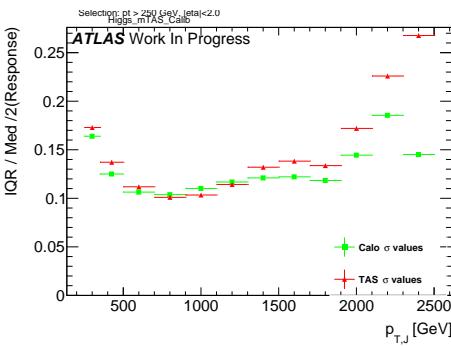
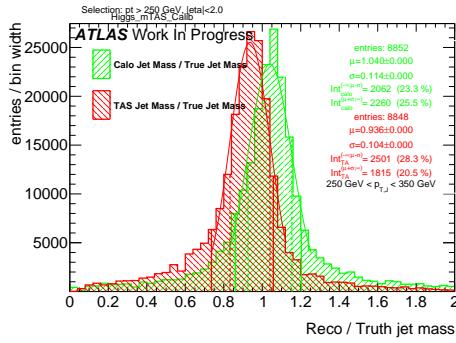
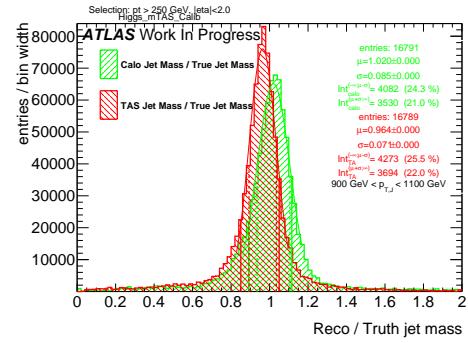
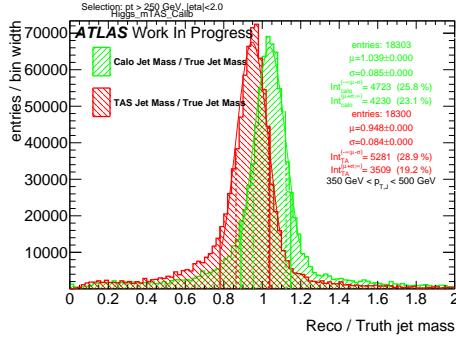
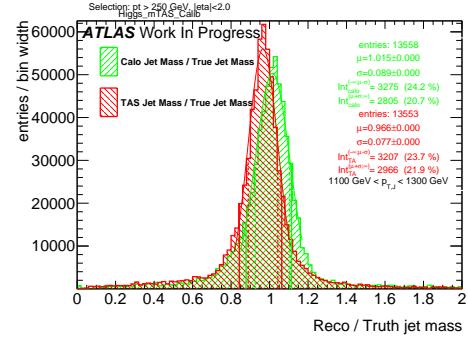
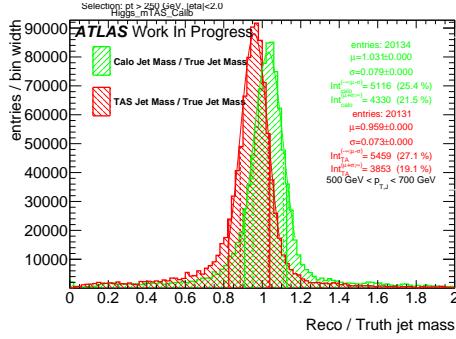
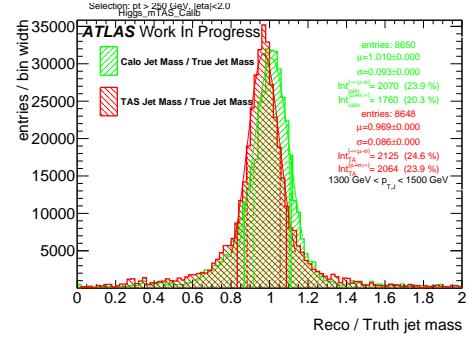
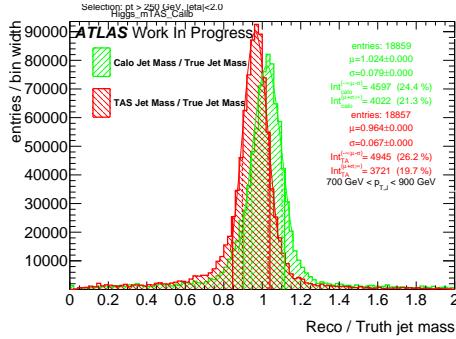
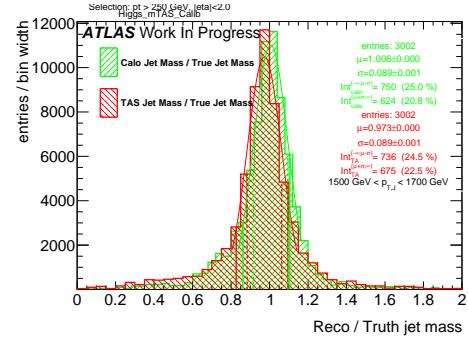


Figure 221: σ from fit of the mass Response vs bin of p_T^J

Figure 224: Response in bin of p_T^J (indicated on plot)Figure 228: Response in bin of p_T^J (indicated on plot)Figure 225: Response in bin of p_T^J (indicated on plot)Figure 229: Response in bin of p_T^J (indicated on plot)Figure 226: Response in bin of p_T^J (indicated on plot)Figure 230: Response in bin of p_T^J (indicated on plot)Figure 227: Response in bin of p_T^J (indicated on plot)Figure 231: Response in bin of p_T^J (indicated on plot)

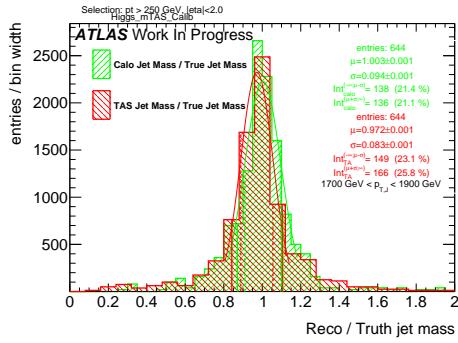


Figure 232: Response in bin of p_T^J (indicated on plot)

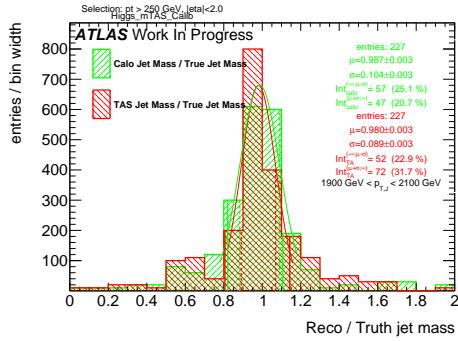


Figure 233: Response in bin of p_T^J (indicated on plot)

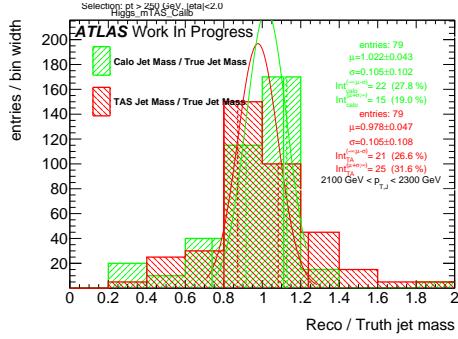


Figure 234: Response in bin of p_T^J (indicated on plot)

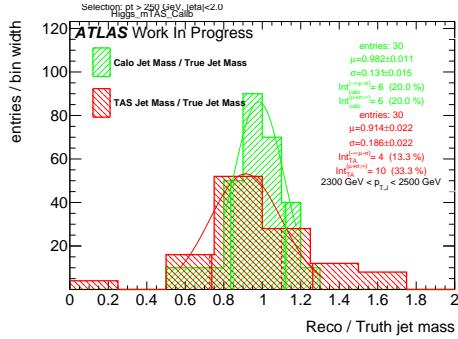
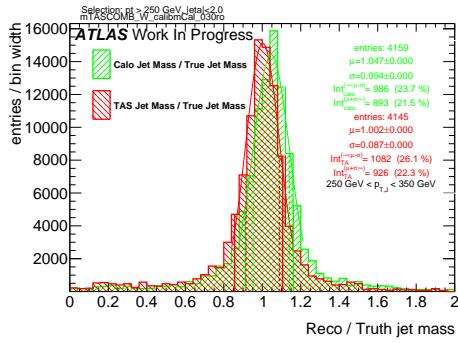
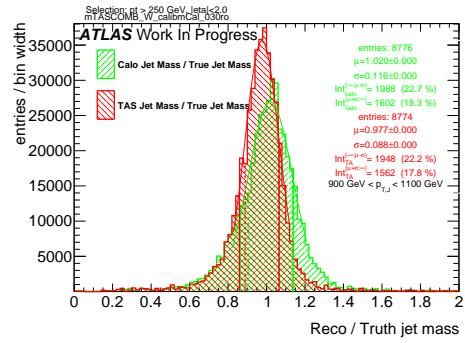
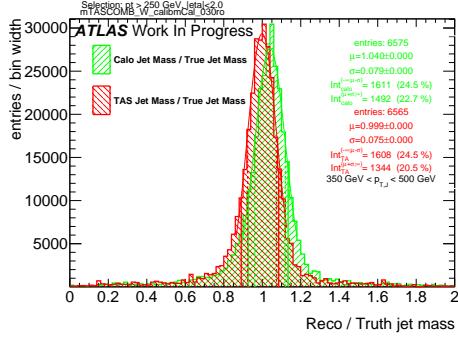
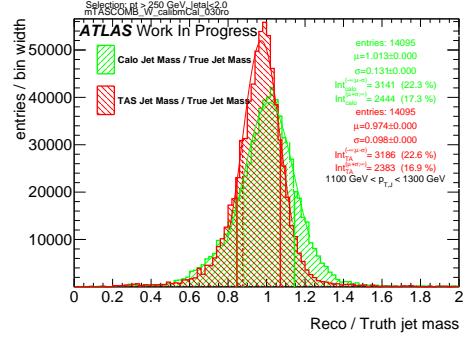
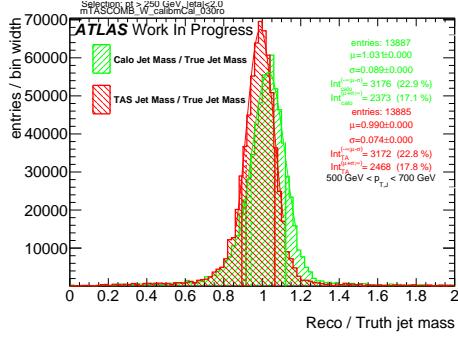
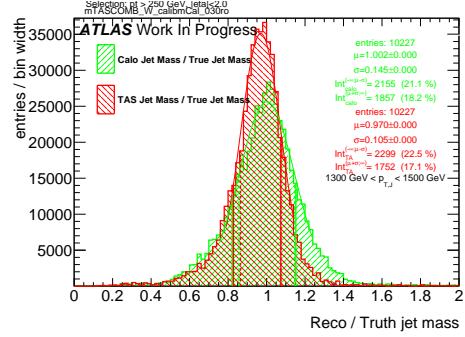
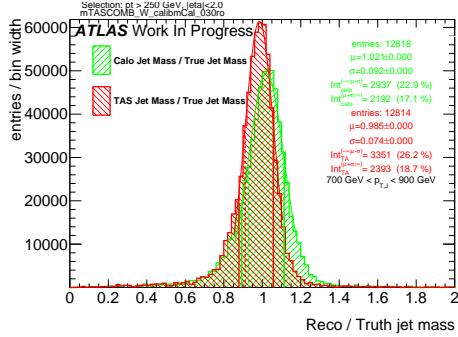
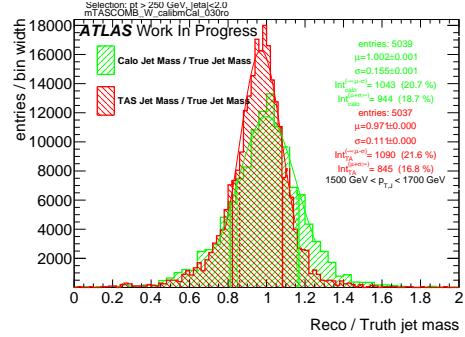
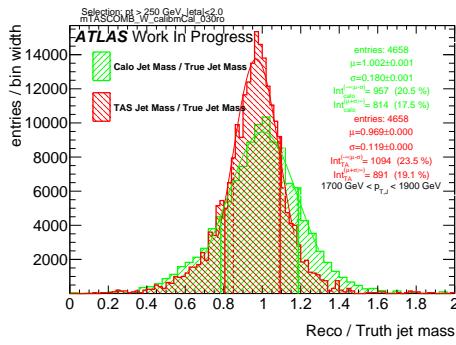
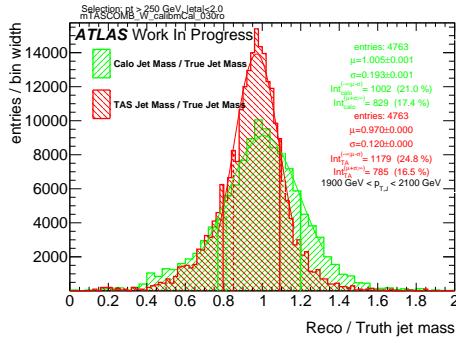
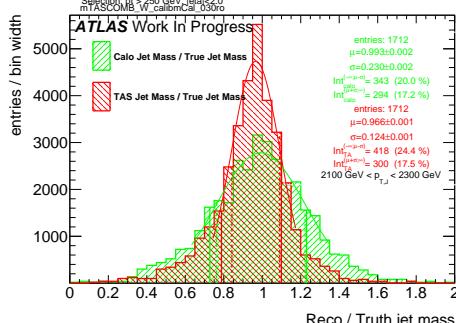
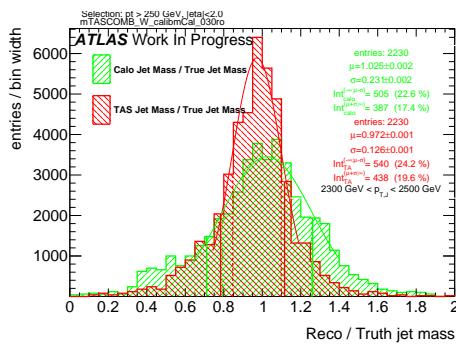


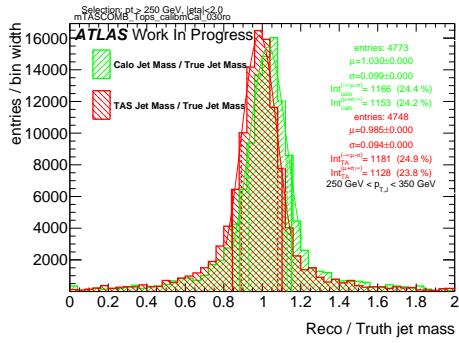
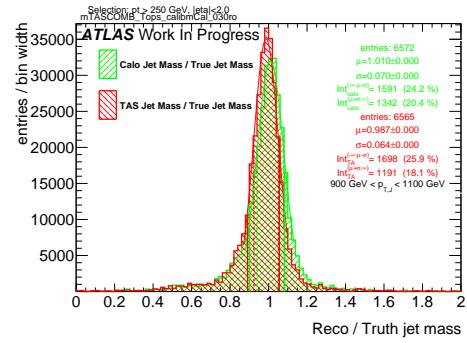
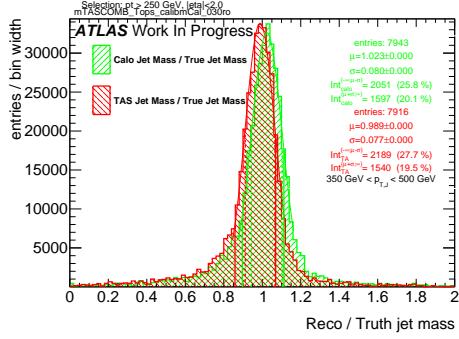
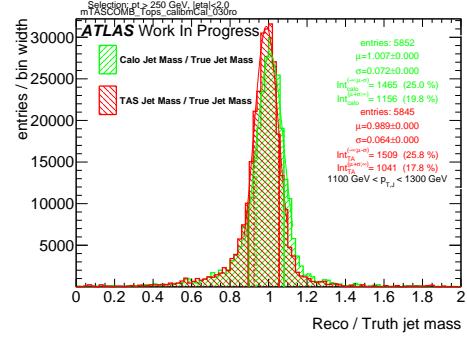
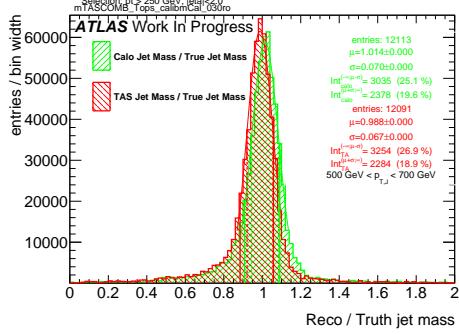
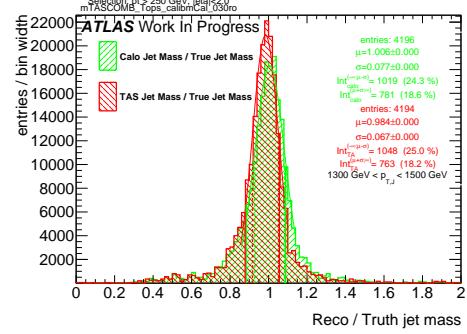
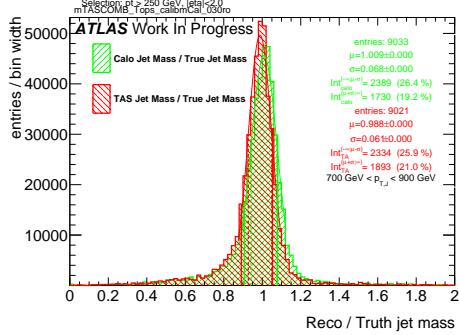
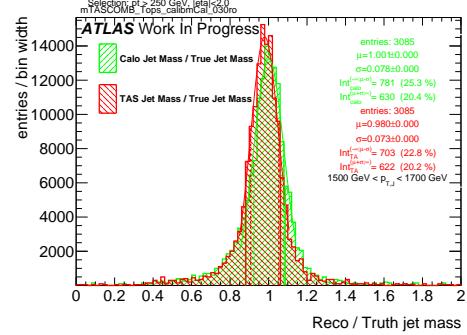
Figure 235: Response in bin of p_T^J (indicated on plot)

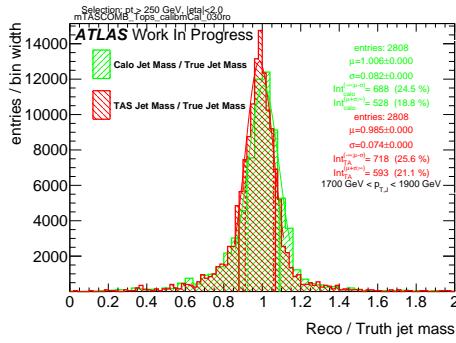
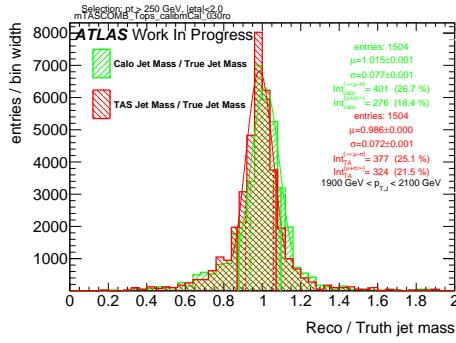
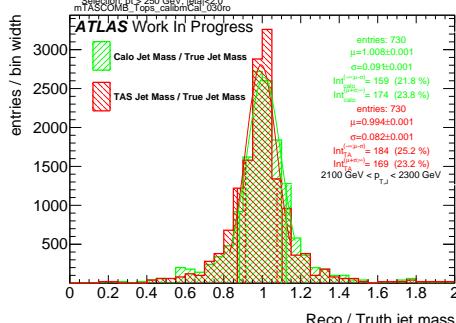
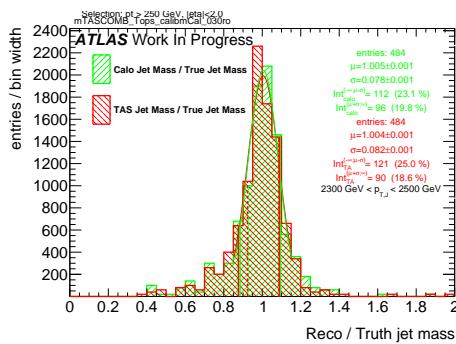
⁷⁷⁵ **D m_{TAS}^{comb} response distributions, boosted W/Z**

Figure 236: Response in bin of p_T^J (indicated on plot)Figure 240: Response in bin of p_T^J (indicated on plot)Figure 237: Response in bin of p_T^J (indicated on plot)Figure 241: Response in bin of p_T^J (indicated on plot)Figure 238: Response in bin of p_T^J (indicated on plot)Figure 242: Response in bin of p_T^J (indicated on plot)Figure 239: Response in bin of p_T^J (indicated on plot)Figure 243: Response in bin of p_T^J (indicated on plot)

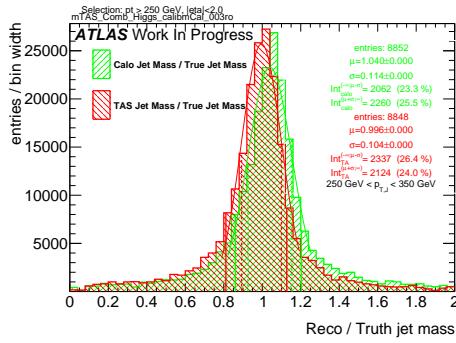
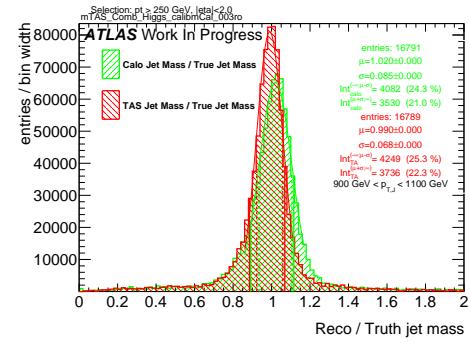
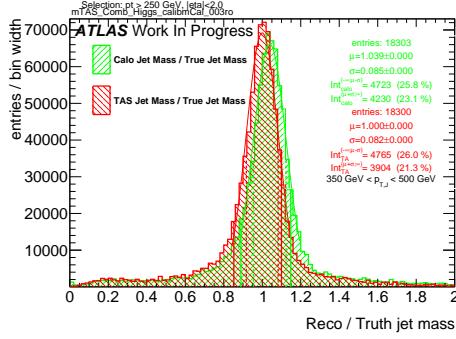
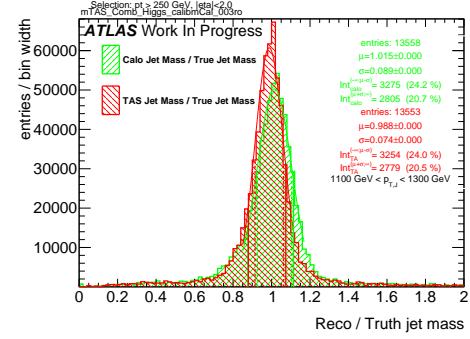
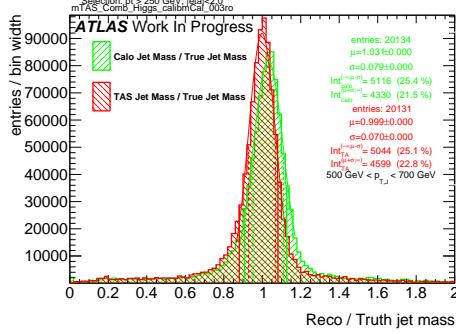
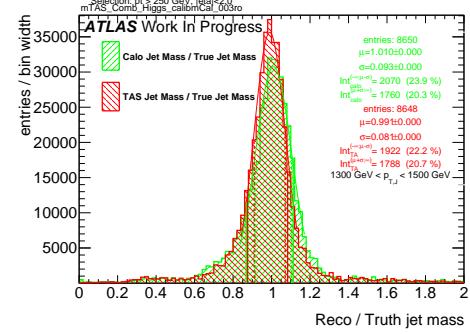
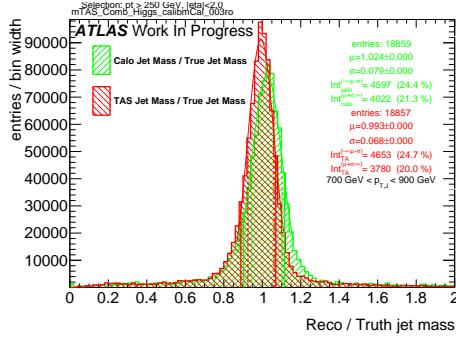
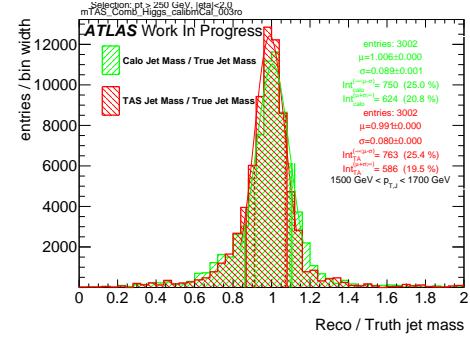
Figure 244: Response in bin of p_T^J (indicated on plot)Figure 245: Response in bin of p_T^J (indicated on plot)Figure 246: Response in bin of p_T^J (indicated on plot)Figure 247: Response in bin of p_T^J (indicated on plot)

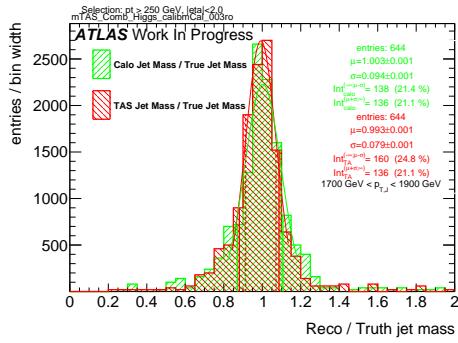
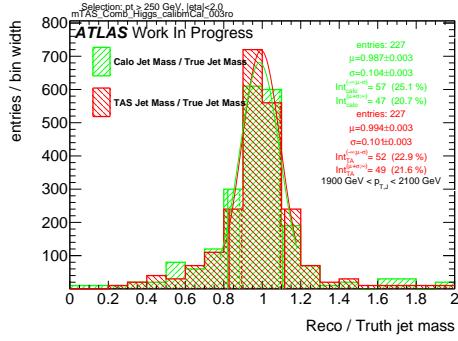
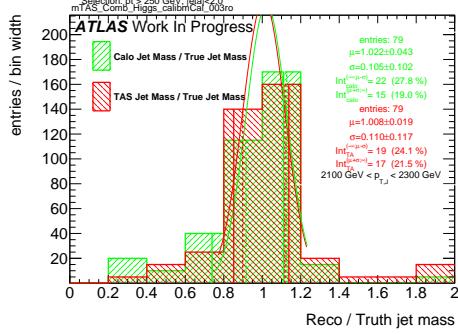
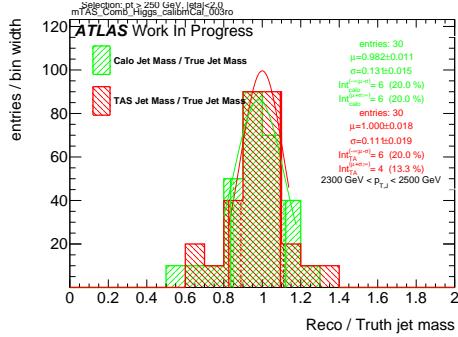
⁷⁷⁶ E m_{TAS}^{comb} response distributions, boosted tops

Figure 248: Response in bin of p_T^J (indicated on plot)Figure 252: Response in bin of p_T^J (indicated on plot)Figure 249: Response in bin of p_T^J (indicated on plot)Figure 253: Response in bin of p_T^J (indicated on plot)Figure 250: Response in bin of p_T^J (indicated on plot)Figure 254: Response in bin of p_T^J (indicated on plot)Figure 251: Response in bin of p_T^J (indicated on plot)Figure 255: Response in bin of p_T^J (indicated on plot)

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

⁷⁷⁷ F m_{TAS}^{comb} response distributions, Higgs

Figure 260: Response in bin of p_T^J (indicated on plot)Figure 264: Response in bin of p_T^J (indicated on plot)Figure 261: Response in bin of p_T^J (indicated on plot)Figure 265: Response in bin of p_T^J (indicated on plot)Figure 262: Response in bin of p_T^J (indicated on plot)Figure 266: Response in bin of p_T^J (indicated on plot)Figure 263: Response in bin of p_T^J (indicated on plot)Figure 267: Response in bin of p_T^J (indicated on plot)

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

778 References

- 779 [1] Particle Data Group Collaboration, K. Olive *et al.*, *Review of Particle Physics*, Chinese Physics C 38
780 (2014) 090001.
- 781 [2] Paul Langacker, *The Physics of Heavy Z Gauge Bosons*. In: Rev. Mod. Phys. 81 (2009), pp. 1199–1228.
782 doi: 10.1103/RevModPhys.81.1199. arXiv: 0801.1345 [hep-ph].
- 783 [3] Martin Schmaltz and Christian Spethmann, *Two Simple W' Models for the Early LHC*. In: JHEP 07
784 (2011), p. 046. doi: 10.1007/JHEP07(2011)046. arXiv: 1011. 5918 [hep-ph].
- 785 [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:
786 10.1016/j.physletb.2016.02.015. arXiv: 1512.05099 [hep-ex].
- 787 [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
788 [hep-ex].
- 789 [6] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016).
790 <http://pdg.lbl.gov/2015/reviews/rpp2015-rev-wprime-searches.pdf>
- 791 [7] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016).
792 <http://pdg.lbl.gov/2015/reviews/rpp2015-rev-zprime-searches.pdf>
- 793 [8] L. Randall and R. Sundrum, *A Large Mass Hierarchy from a Small Extra Dimension*, Phys. Rev. Lett.
794 83 (1999) 3370, arXiv: hep-ph/9905221.
- 795 [9] H. Davoudiasl, J.L. Hewett and T.G. Rizzo, *Warped Phenomenology*, Stanford Linear Accelerator
796 Center Stanford CA 94309, USA, arXiv:hep-ph/9909255v1 6 Sep 1999.
- 797 [10] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1 & 2*
798 , Tech. Rep. ATLAS TDR 14, CERN/LHCC 99-14, Geneva, 1999.
- 801 [11] L. Evans and P. Bryant, *LHC Machine*, JINST 3 (2008) S08001.
- 802 [12] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016).
803 <http://pdg.lbl.gov/2010/reviews/rpp2010-rev-qcd.pdf>
- 804 [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*,
805 CERN-PH-EP/2013-037 2015/05/04 arXiv:hep-ph/14121633.
- 807 [14] The ATLAS Collaboration, *Improved luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using
808 the ATLAS detector at the LHC*, Eur. Phys. J. C (2013) 73: 2518. doi:10.1140/epjc/s10052-013-2518-3.
- 809 [15] W. Bartel *et al.* [JADE Collaboration], *Observation of Planar Three Jet Events in e + e - Annihilation
810 and Evidence for Gluon Bremsstrahlung*, Phys. Lett. B 91, 142 (1980).
- 811 [16] Ahmed Ali (DESY), Gustav Kramer (Univ. Hamburg), *Jets and QCD: A Historical Review of the
812 Discovery of the Quark and Gluon Jets and its Impact on QCD*, arXiv:1012.2288v2 [hep-ph].
- 813 [17] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, *The anti- k_t jet clustering algorithm*,
814 arXiv:0802.1189v2 [hep-ph].

- 815 [18] F. Halzen, A.D. Martin and D.M. Scott, Phys. Rev. D25 (1982) 754, Phys. Lett. B112 (1982) 160.
- 816 [19] Torbjorn Sjostrand, *Monte Carlo Generators*, hep-ph/0611247, CERN-LCGAPP-2006-06.
- 817 [20] W. Lampl *et al.*, *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-
818 PUB-2008-002.
- 819 [21] The ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded*
820 *in 2010 with the ATLAS detector*, CERN-PH-EP-2012-191 arXiv:1210.6210v1 [hep-ex].
- 821 [22] GEANT4 Collaboration (S. Agostinelli *et al.*), *GEANT4: A Simulation toolkit*, Nucl.Instrum.Meth.
822 A506 (2003) 250-303 DOI: 10.1016/S0168-9002(03)01368-8 SLAC-PUB-9350, FERMILAB-PUB-
823 03-339.
- 824 [23] Ariel Schwartzman, *Jet energy calibration at the LHC*, SLAC National Accelerator Laboratory, 2575
825 Sand Hill Road Menlo Park, California 94025, USA, arXiv:1509.05459v1 [hep-ex] 17 Sep 2015.
- 826 [24] The ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*,
827 ATLAS-CONF-2016-035 19 July 2016.
- 828 [25] The ATLAS Collaboration, *Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at*
829 *$\sqrt{s} = 13$ TeV*, ATLAS-CONF-2016-039 30th July 2016.
- 830 [26] The ATLAS Collaboration, *Search for high-mass diboson resonances with boson-tagged jets in*
831 *proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, arXiv:1506.00962v3 [hep-ex].
- 832 [27] The CMS Collaboration, *Combination of searches for WW, WZ, ZZ, WH, and ZH resonances at $\sqrt{s} =$*
833 *8 and 13 TeV*, CMS PAS B2G-16-007.
- 834 [28] The ATLAS Collaboration, *Search for heavy particles decaying to pairs of highly-boosted top quarks*
835 *using lepton-plus-jets events in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
836 ATLAS-CONF-2016-014.
- 837 [29] The ATLAS Collaboration, *Search for resonances with boson-tagged jets in 15.5 fb^{-1} of p-p*
838 *collisions at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, ATLAS-CONF-2016-055.
- 839 [30] The ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}\bar{b}\bar{b}$ final state using*
840 *proton proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, ATLAS-CONF-2016-049.
- 841 [31] The ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in*
842 *proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector*, CERN-PH-EP-2013-069
843 arXiv:1306.4945 [hep-ex].
- 844 [32] Oleg Zenin, on behalf of the ATLAS Collaboration, *Soft QCD and underlying event measurements*
845 *at ATLAS*, Nuclear and Particle Physics Proceedings Volumes 273275, April-June 2016, Pages 2053-
846 2058 37th International Conference on High Energy Physics (ICHEP).
- 847 [33] Page title: JetSubstructureECFA2014 AtlasPublic TWiki Website name: Twiki.cern.ch
848 URL:<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetSubstructureECFA2014> Access date: 27th
849 October 2016.
- 850 [34] E. Lehmann and G. Casella, *Theory of Point Estimation*, Springer Verlag, 1998, isbn: 0387985026.
- 851 [35] Andrew J. Larkoski, Ian Moult, and Duff Neill. *Power Counting to Better Jet Observables*. *JHEP*,
852 12:009, 2014.

- 853 [36] Andrew J. Larkoski, Ian Moult, and Duff Neill. *Analytic Boosted Boson Discrimination*. *JHEP*,
854 05:117, 2016.
- 855 [37] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. *Energy Correlation Functions for Jet Sub-*
856 *structure*. *JHEP*, 06:108, 2013.
- 857 [38] Jesse Thaler and Ken Van Tilburg. *Identifying Boosted Objects with N-subjettiness*. *JHEP*, 03:015,
858 2011.
- 859 [39] Georges Aad et al. *Identification of boosted, hadronically decaying W bosons and comparisons with*
860 *ATLAS data taken at $\sqrt{s} = 8 \text{ TeV}$* . *Eur. Phys. J.*, C76(3):154, 2016.

861 **List of contributions**

862

863 Auxiliary material

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

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