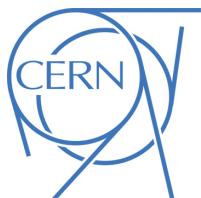




## ATLAS Note

KIP-2016



Draft version 0.2

1

# 2 Jet Observables using Subjet-assisted Tracks

3 Oleg Brandt<sup>a</sup>, Sascha Dreyer<sup>a</sup>, Fabrizio Napolitano<sup>a</sup>, Merve Sahinsoy<sup>a</sup>,  
4 Benjamin P. Nachman<sup>b</sup>

5 <sup>a</sup>*Heidelberg University*

6 <sup>b</sup>*Lawrence Berkeley National Laboratory*

7 27th December 2016

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

## 17 **Contents**

18	<b>1 Introduction</b>	4
19	<b>2 MC Samples and event weighting</b>	4
20	<b>3 Object Definition</b>	5
21	3.1 Standard Large-Radius jet	5
22	3.1.1 Grooming and Selection	5
23	3.1.2 Calorimeter Mass	6
24	3.1.3 Track Mass	6
25	3.1.4 Track-Assisted Mass ( $m^{TA}$ )	6
26	3.2 The Track-Assisted Sub-jet Mass ( $m^{TAS}$ )	7
27	3.2.1 Observable Definition: Inputs	7
28	3.2.2 Observable Definition: Procedure	8
29	3.3 The Combined Mass	10
30	3.3.1 Combination $m^{TA} - m^{calo}$	10
31	3.3.2 Combination $m^{TAS} - m^{calo}$	10
32	3.4 Jet Substructure observables with assisted tracks	12
33	3.4.1 Energy Correlation Functions	12
34	3.4.2 n-Subjettiness	13
35	<b>4 Figures of Merit for Performance Studies</b>	14
36	4.1 For jet mass	14
37	4.2 ROC Cursves and prior mass cut	14
38	<b>5 Performance of Track-assisted subjet mass</b>	15
39	5.1 Performance in $W \rightarrow q'\bar{q}$ Decays	15
40	5.2 Performance in $h \rightarrow b\bar{b}$ Decays	15
41	5.3 Performance in $t \rightarrow q'\bar{q}b$ Decays	16
42	5.4 Performance in QCD Multijet Events	16
43	5.5 Performance in Massive $\tilde{W} \rightarrow q'\bar{q}$ Decays with $m_{\tilde{W}} = m_t$	17
44	5.6 Stability of Mean of Response and Left-Hand-Side Integral	18
45	5.7 Potential Improvements from Sub-jet Calibration	18
46	5.7.1 Simple Sub-jet Calibration	19
47	5.8 Limitation of $m^{TAS}$ from tracking	19
48	5.9 Performance with Alternate Inputs to the $m^{TAS}$	21
49	<b>6 Performance of Combined Calorimeter and Track-Assisted Sub-Jet Mass</b>	23
50	6.1 Performance in $W \rightarrow q'\bar{q}$ Decays	23
51	6.2 Performance in $h \rightarrow b\bar{b}$ Decays	24
52	6.3 Performance in $t \rightarrow q'\bar{q}b$ Decays	24
53	<b>7 Performance of Jet Substructure observables with (assisted) tracks</b>	25
54	7.1 Event weighting and Mass-Cut	25
55	7.2 Track Selection	26
56	7.3 Performance with default $\beta$	28
57	7.3.1 Performance for $W$ boson tagging	28

---

58	7.3.2	Un-assisted tracks and TAS at very high $p_T$	29
59	7.3.3	Correlation with $p_T$	30
60	7.3.4	Performance for Higgs boson tagging	31
61	7.3.5	Performance for Top quark tagging	32
62	7.4	Optimisation of $\beta$	32
63	7.4.1	Optimisation for $W$ boson jets	33
64	7.4.2	Optimisation for Higgs boson jets	34
65	7.4.3	Optimisation for Top quark jets	38
66	<b>8</b>	<b>Uncertainties on observables with sub-jet-assisted tracks</b>	<b>41</b>
67	<b>9</b>	<b>Conclusions &amp; Outlook</b>	<b>42</b>
68	<b>Appendix</b>		<b>43</b>
69	<b>A</b>	<b>Monte Carlo Samples</b>	<b>43</b>
70	<b>B</b>	<b>Trimming</b>	<b>43</b>
71	<b>C</b>	<b>Tracks details</b>	<b>44</b>
72	<b>D</b>	<b>Alternative Performance Figure of Merit (FoM)</b>	<b>45</b>
73	D.1	Gaussian Fit	45
74	<b>E</b>	<b><math>m^{TAS}</math> distributions, boosted <math>W/Z</math></b>	<b>53</b>
75	<b>F</b>	<b><math>m^{TAS}</math> distributions, boosted tops</b>	<b>65</b>
76	<b>G</b>	<b><math>m^{TAS}</math> distributions, boosted higgs</b>	<b>76</b>
77	<b>H</b>	<b><math>m_{TAS}^{comb}</math> response distributions, boosted <math>W/Z</math></b>	<b>87</b>
78	<b>I</b>	<b><math>m_{TAS}^{comb}</math> response distributions, boosted tops</b>	<b>90</b>
79	<b>J</b>	<b><math>m_{TAS}^{comb}</math> response distributions, Higgs</b>	<b>93</b>
80	<b>K</b>	<b><math>W</math> boson Tagging</b>	<b>96</b>
81	<b>L</b>	<b>Higgs Boson Tagging</b>	<b>96</b>
82	<b>M</b>	<b>Top Quark Tagging</b>	<b>96</b>
83	<b>N</b>	<b><math>W</math> Distributions</b>	<b>134</b>
84	<b>O</b>	<b>Higgs Distributions</b>	<b>134</b>
85	<b>P</b>	<b>Top Distributions</b>	<b>134</b>

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.

## 86    **Auxiliary material**

**138**

## 87    **1 Introduction**

88 Jets are collimated streams of particles resulting from quarks and gluons fragmentation and hadronization.  
 89 The distribution of energy inside a jet contains information about the initiating particle. When a massive  
 90 particle such as a top quark, Higgs boson or  $W/Z$  bosons is produced with significant Lorentz boost and  
 91 decays into quarks, the entire hadronic decay may be captured inside a single jet. The mass of such jets  
 92 (jet mass) is one of the most powerful tools for distinguishing massive particle decays from the continuum  
 93 multijet background; the Energy Correlation Functions and n-Subjettiness  $C_2$ ,  $D_2$ ,  $\tau_{21}$  and  $\tau_{32}$  provide an  
 94 ad-hoc tool pupusely developed for the multijet background and constitue a fundamental part of many for  
 95 boson taggers. This note documents the so-called subjet-assisted techniques with the ATLAS detector.  
 96 The track-assisted subjet mass  $m^{TAS}$  definition is presented and confronted with the standard development  
 97 in ATLAS,  $m^{comb}$  and  $m^TA$ . Energy Correlation Functions and n-Subjettiness with the modified subjet-  
 98 assisted technique are presented and confronted with the standard one in ATLAS. The note ends with  
 99 conclusions for the jet observables using subjet-assisted tracks.

## 100    **2 MC Samples and event weighting**

101 The samples used are divided into two main groups: SM background and beyond SM signal. The  
 102 SM background includes the QCD multijet samples, produced with a falling  $p_T$  spectrum. The beyond  
 103 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  
 104 ( $t \rightarrow W(\rightarrow q\bar{q}')b$ )) and RS-Graviton  $\rightarrow hh \rightarrow b\bar{b}b\bar{b}$ , i.e. final states have only jets in all the samples.  
 105 The details of the samples are given in Table 9; the masses considered span from 0.5 to 5 TeV to improve  
 106 and diversify the kinematic space covered.

## <sup>107</sup> 3 Object Definition

<sup>108</sup> This section gives an overview of the objects used for the observables based on subjet-assisted tracks,  
<sup>109</sup> which are the large-radius jet mass, the Energy Correlation Functions and the n-Subjettiness, as they are  
<sup>110</sup> used within ATLAS.

### <sup>111</sup> 3.1 Standard Large-Radius jet

<sup>112</sup> Large-radius jet, or large- $R$  jets are jets constructed with a radius parameter of the reclustering algorithm  
<sup>113</sup> of 1.0 for those built using the anti- $k_t$  algorithm and 1.2 for the C/A algorithm. Since the active area of  
<sup>114</sup> this jets is typically six times bigger than their counterparts of radius 0.4 which is the usual choice of jet  
<sup>115</sup> radii within ATLAS, the necessity of further techniques is required to have control over the effect of soft  
<sup>116</sup> radiation contamination from Pile-Up (PU) and Underlying Event (UE).

#### <sup>117</sup> 3.1.1 Grooming and Selection

<sup>118</sup> **Grooming** In order to use large- $R$  jets, it is necessary to gain additional information on the interior of  
<sup>119</sup> these objects, i.e. using techniques that exploit its substructure allowing a jet-by-jet discrimination of the  
<sup>120</sup> energy deposit most likely coming from the hard-scattering to other soft radiation.

<sup>121</sup> A common feature in substructure is the use of *sub-jet*, i.e. jets obtained from a parent jet (e.g. the large- $R$   
<sup>122</sup> jet), using its constituent but running the jet reclustering algorithm with a smaller radius parameter; in one  
<sup>123</sup> large- $R$  jet, typically there are two or more sub-jets depending on the originating process and its  $p_T$ .

<sup>124</sup> Techniques have been developed, both using sub-jets or directly constituents of a jet, which are referred  
<sup>125</sup> to as *grooming* algorithms.

<sup>126</sup> Grooming algorithms are designed to retain the characteristic substructure within such a large- $R$  jet while  
<sup>127</sup> reducing the impact of the fluctuations of the parton shower and the UE, thereby improving the mass  
<sup>128</sup> resolution and mitigating the influence of pile-up.

<sup>129</sup> The grooming algorithms presented here are the most important ones in ATLAS: the *Trimming*; other  
<sup>130</sup> used as well, the *Split-Filtering* and the *Pruning* can be found in [substructure1]. Details on Trimming,  
<sup>131</sup> the most used within ATLAS and in this note, are given in the Appendix.

<sup>132</sup> **Selection** The selection applied is typical fro many Beyond the Standard Model searches:  $p_T > 250$   
<sup>133</sup> GeV and  $|\eta| < 2.0$  for the large-radius jet. No other requirements were made for the purpose of the  
<sup>134</sup> performance studies here shown if not stated differently, e.g. the mass cut selection.

---

<sup>135</sup> **3.1.2 Calorimeter Mass**

<sup>136</sup> Once the collection of topo-clusters from the large- $R$  jet is groomed, i.e. cleaned from PU contamination  
<sup>137</sup> through the trimming technique, it is possible to use them for the measure of physical related properties  
<sup>138</sup> such as the jet mass, since the possible sources of soft radiation from PU and UE have been reduced.

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

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

<sup>139</sup> where  $J$  labels the Large- $R$  jet and assuming the topo-clusters as massless.

<sup>140</sup> **3.1.3 Track Mass**

<sup>141</sup> There are significant advantages and few disadvantages of the use of tracks for large-radius jet mass  
<sup>142</sup> reconstruction, inherited both from the detector experimental properties and from the underlying physical  
<sup>143</sup> processes. Main advantages are: performance of angular separation and the association of the tracks to  
<sup>144</sup> the primary vertex for rejection of soft radiation background. Tracks can additionally required to be well  
<sup>145</sup> reconstructed from the detector and they are classified in LOOSE, MEDIUM and TIGHT for increasing  
<sup>146</sup> quality criteria. The mass  $m^{track}$  is then calculated summing up the 4-momenta of those tracks which  
<sup>147</sup> passed the selection and are ghost associated to the groomed jet.

<sup>148</sup> The important disadvantage comes from the complete blindness of the tracker system to the electrically  
<sup>149</sup> neutral component (mostly  $\pi^0$ ) of the jet. As seen in Figure 1, the track mass (red distribution) is not only  
<sup>150</sup> shifted towards lower values than the calorimeter mass (green distribution), but its width also degrades.

<sup>151</sup> Tracks could be used either for independent mass reconstruction or, most importantly, as an additional  
<sup>152</sup> information to the calorimeter measurement.

<sup>153</sup> **3.1.4 Track-Assisted Mass ( $m^{TA}$ )**

<sup>154</sup> The track-assisted mass,  $m^{TA}$ , was one of the first attempts to combine the information form the tracker  
<sup>155</sup> system and from the calorimeter. It is defined as  $m^{TA} = \frac{p_T^{calo}}{p_T^{track}} \times m^{track}$ , where the  $p_T^{track}$  and the  $m^{track}$   
<sup>156</sup> are calculated from the tracks which are associated to the large-radius jet, adding up their 4-momenta (hence  
<sup>157</sup> exploiting the superior angular resolution of the tracker system); the  $p_T^{calo}$  is the transverse momentum  
<sup>158</sup> as measured from the calorimeter system. The ratio  $p_T^{calo}/p_T^{track}$  restores the fraction of the missing  
<sup>159</sup> neutral component in the  $m^{track}$ . The  $m^{TA}$  has a better performance on the reconstruction of boosted  
<sup>160</sup> objects such as  $W/Z$  in the extreme kinematic regime ( $\sim 1$  TeV) and above in the transverse momentum  
<sup>161</sup> of the decaying electroweak object. Another advantage of this observable shows up as it comes to the  
<sup>162</sup> systematic uncertainties: in particular jet mass scale and jet mass resolution uncertainty on  $m^{TA}$  can  
<sup>163</sup> be estimated by propagating the track reconstruction uncertainties and calorimeter-jet  $p_T$  uncertainties  
<sup>164</sup> through the definition of the variable given above. The tracking uncertainties are smaller for  $m^{TA}$  rather  
<sup>165</sup> than  $m^{calo}$  because a larger extent of the uncertainty cancels in the ratio  $m^{track}/p_T^{track}$ . Apart all of this

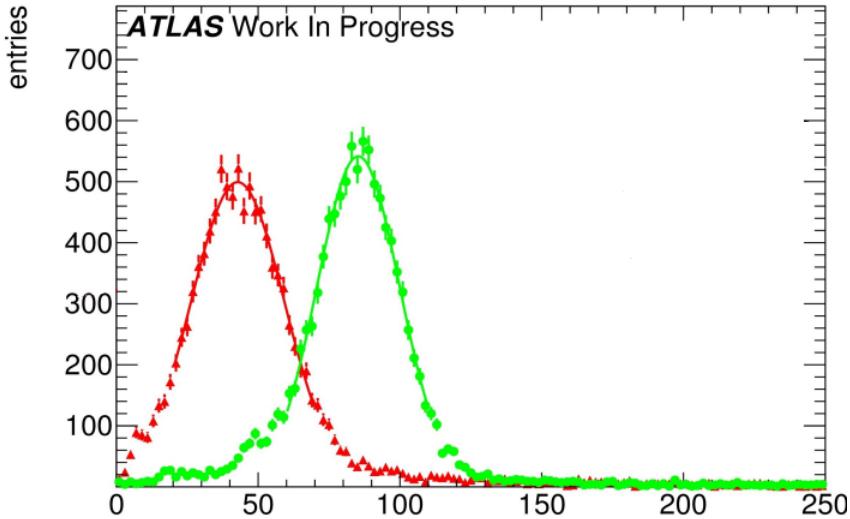


Figure 1: Mass distribution boosted  $W/Z$ : in green the  $m^{calo}$  and in red the  $m^{track}$ .

166 advantages, the track-assisted mass shows its limits when it comes to intermediate transverse momentum  
167 regimes and below ( $p_T < 1$  TeV) in  $W/Z$  and for Higgs and top quarks throughout the whole kinematic  
168 space. Full description of this variable is given in the ATLAS CONF Note [[art35](#)].

### 169 3.2 The Track-Assisted Sub-jet Mass ( $m^{TAS}$ )

170 In this section the main outcome of the optimization of the large-radius jet mass reconstruction is presented:  
171 the *track-assisted sub-jet mass* ( $m^{TAS}$ ). The main idea takes inspiration from the track-assisted mass: if  
172 one can use tracks to exploit the better angular resolution and correct the missing neutral component jet-  
173 by-jet, there is an additional information that can be used. The neutral fraction, in fact, varies stochastically  
174 not only per-jet basis, but even per-sub-jet basis, since each quark follows a different parton showering  
175 and hadronization process. Correcting the missed neutral component per-sub-jet, it should perform better  
176 already at an intuitive level, as it accesses information from jet substructure.

#### 177 3.2.1 Observable Definition: Inputs

178 There are two inputs to the  $m^{TAS}$ : tracks and sub-jets. The definition of the standard inputs are given here;  
179 alternative approaches are given in subsection [5.9](#).

180 **Tracks** Only the tracks that satisfy the quality criteria and primary vertex association, described in the  
181 appendix [3.1.3](#), are used. The tracks are additionally required to be ghost associated to the sub-jets of the  
182 groomed jet; namely only the sub-jets which survived the trimming procedure and are described in the  
183 next subsection. Ghost association provides a clear correspondence of tracks to the sub-jets set and was  
184 therefore chosen and preferred to other kind of assignments.

185 **Sub-jets** The choice of sub-jets must follow a simple requirement: of course we want to take those which  
 186 most likely come from the hard-scattering. This means that the choice of taking them after grooming is  
 187 strongly favored.

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

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

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

193 **3.2.2 Observable Definition: Procedure**

194 There are two ways of sub-jet assisting the tracks for the calculation of the  $m^{TAS}$ : assisting track-jets  
 195 changing the mass or assisting single tracks changing the transverse momentum. The first approach was  
 196 the first one also historically, adopted because of higher versatility and feasibility of implementation. The  
 197 second approach was also found to be equivalent to the first [presentation].

198 For the substructure variable, however, the first approach cannot be used because of simple cancellation  
 199 in the computation of the variable.

200 To generalize the scheme adopted for both, tracks should be assisted singularly. In this note will be shown  
 201 the  $m^{TAS}$  obtained assisting track-jets, since the differences are negligible.

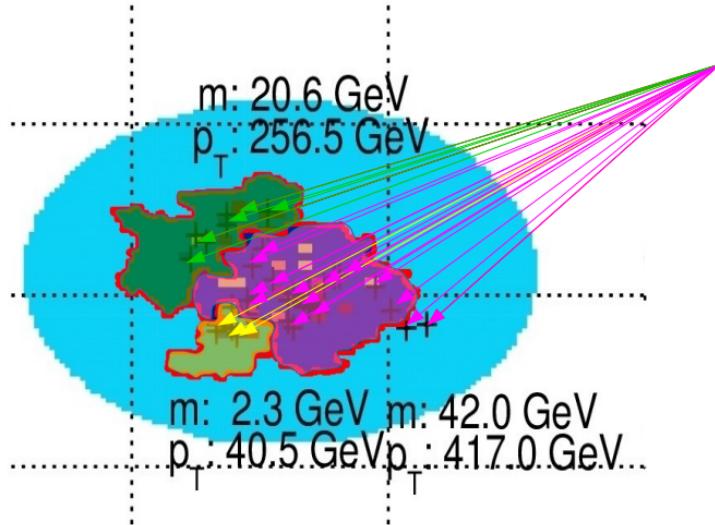


Figure 2: Pictorial event display showing the  $\eta \phi$  region of a large- $R$  anti- $k_t$  trimmed jet, (in blue the catchment area of the anti- $k_t$ ) showing the different  $k_t$  sub-jets: they are highlighted in green, fuchsia and yellow. The associated track-jets (here indicated as arrows pointing the calorimeter area) are colored with the same color of the correspondent sub-jet. Some tracks associated with  $\Delta R$  procedure can be seen in the fuchsia sub-jet. The transverse momenta and mass values are also shown for the sub-jets.

202 **Assisting Track-Jets** Having tracks and sub-jets now well defined, we can describe the recipe to produce  
 203 the  $m^{TAS}$ . For brevity we will call the sub-jets SJ in the formulae below.

204 As said, the tracks are the ones ghost-associated to the sub-jets; however, tracks which fall inside the area  
 205 of the large- $R$  jet, but not inside the sub-jets area, are still much probably coming from the hard-scattering.  
 206 They are then associated again to the closest sub-jets via  $\Delta R$  association.

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

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

211 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  
 212 level, i.e.

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

213 Where the summation symbol between quotation mark symbolize that the sum must be intended at 4-vector  
 214 level: since now we are working inside the sub-jets, in fact, we need to change the sub-jet’s 4-vector itself  
 215 and not only the mass. If we call  $p_\mu^{TJ}$  the Lorentz vector of the track-jet,

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

216 where  $p_\mu^{TA}$  is the track-assisted sub-jet’s 4-vector. If we label  $i$  the  $i$ -th track-jet of the  $N$  ones present in  
 217 the large- $R$  jet,

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

218 **Assisting Single Tracks** This correction is now applied on single track rather than the whole track-jet  
 219 and on the transverse momentum, not the mass. The TAS correction reads:

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

220 The corection factor  $\frac{p_T^{SJ}}{p_T^{TJ}}$  refers to the  $p_T$  of the sub-jet in which the track is associated and the  $p_T$  of the  
 221 track-jet associated to it.

222 As before, these four-momenta are then summed together to give this alternative definition:

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

223 where now the sum refers from the first to the M-th tracks associated to the large- $R$  jet.

224 An important remark is that, in the case of a large- $R$  jet with only one sub-jet, the  $m^{TAS}$  has exactly  
225 the same definition of the  $m^{TA}$ . This implies, since the angular separation of the decay product scales  
226 inversely with  $p_T$ , that the performance should approach the one of the  $m^{TA}$  at very high transverse  
227 momenta. However, the space for improvement is precisely in the low-intermediate  $p_T$  regime.

### 228 3.3 The Combined Mass

229 Since the calorimeter large- $R$  jet mass is not explicitly used in the track-assisted (sub-jet) mass, it may be  
230 possible to improve the performance creating a new observable which combines both mass definitions.  
231 This is discussed in great details in the BOOST 2016 Conf Note [[art35](#)].

#### 232 3.3.1 Combination $m^{TA} - m^{calo}$

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

$$\begin{aligned} m^{comb} &= a \times m^{calo} + b \times m^{TA}, \\ a &= \frac{\sigma_{calo}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} & b &= \frac{\sigma_{TA}^{-2}}{\sigma_{calo}^{-2} + \sigma_{TA}^{-2}} \end{aligned} \quad (1)$$

234 where  $\sigma_{calo}$  and  $\sigma_{TA}$  are the  $m^{calo}$ 's and  $m^{TA}$ 's resolution functions. The  $m^{comb}$  then is the  $m^{TA} - m^{calo}$   
235 combination. The weights are here and also afterwards computed from the mass response distribution;  
236 the sigma parameter corresponds to the width of the Gaussian distribution, which is estimated using the  
237 InterQuantile range.

#### 238 3.3.2 Combination $m^{TAS} - m^{calo}$

239 There is a main difference between the  $m^{TAS}$  and  $m^{TA}$  when it comes to combination: since the  $m^{TAS}$  is  
240 using sub-jet level information but  $m^{TA}$  not, the correlation with the  $m^{calo}$  is expected to be higher. This  
241 can be seen e.g. in the plots in Figure 3 (additional plots shown in Figure 42 in Appendix), where the  
242 correlation is not only higher for the simple W/Z and Higgs jets, but above 50% for tops. The assumption  
243 of independent variables here falls, since the observable are only approximately Gaussian. The Ansatz is  
244 to take into account the correlation via the formula:

$$\begin{aligned} m_{TAS}^{comb} &= w \times m^{calo} + (1 - w) \times m^{TAS}, \\ w &= \frac{\sigma_{TAS}^2 - \rho \sigma_{calo} \sigma_{TAS}}{\sigma_{calo}^2 + \sigma_{TAS}^2 - 2 \rho \sigma_{calo} \sigma_{TAS}} \end{aligned} \quad (2)$$

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

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

$$a = \frac{\sigma_{TAS}^2 - \rho \sigma_{calo} \sigma_{TAS}}{\sigma_{calo}^2 + \sigma_{TAS}^2 - 2\rho \sigma_{TAS} \sigma_{calo}} \quad b = \frac{\sigma_{calo}^2 - \rho \sigma_{calo} \sigma_{TAS}}{\sigma_{calo}^2 + \sigma_{TAS}^2 - 2\rho \sigma_{TAS} \sigma_{calo}} \quad (3)$$

246 which reduces to equation (1) after simple algebra for the case when  $\rho = 0$ . Of course, this value can be  
247 set to the value of the specific sample considered, or to an average of 0.3 if one wants to give a definition  
248 generally valid for all the cases considered; in this case, the performance would be slightly sub-optimal.

249 **Procedure** The procedure of producing the  $m_{TAS}^{comb}$  is defined as follows:

- 250 1. For the given sample, the  $m^{TAS}$  and  $m^{calo}$  are calculated;
- 251 2. The mass responses are also produced for the given ranges of  $p_T$ ;
- 252 3. For each of these responses, the value of the  $\frac{68\% \text{ IQnR}}{2}$  (identified as the  $\sigma$  in Eq 2) as defined  
253 previously is calculated and stored;
- 254 4. The average correlation factor of 0.3 (an average value for the samples considered) is assumed;
- 255 5. With the formula 2,  $m_{TAS}^{comb}$  is calculated using the  $m^{TAS}$ ,  $m^{calo}$  and the values stored in step 3.

256 In this note, the IQnR weights are produced for each sample specifically. In order to give a sample-  
257 independent definition of the  $m_{TAS}^{comb}$ , following also the procedure adopted for the  $m^{comb}$ , these weights  
258 could be taken from a QCD multijet sample and applied indiscriminately to the particular case. Here of  
259 course the performance would be again sub-optimal, being the variable not developed in an ad-hoc way  
260 for each signal sample, but from QCD multijet only.

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

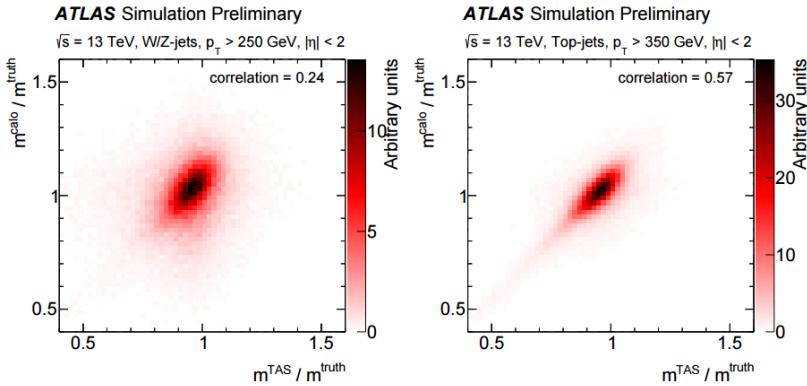


Figure 3: The calorimeter based jet mass mass response versus the track-assisted sub-jet mass response, on the left for boosted  $W/Z$  on the right for boosted tops.

265 **3.4 Jet Substructure observables with assisted tracks**

266 **3.4.1 Energy Correlation Functions**

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

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

$$\begin{aligned} \text{ECF1} &= \sum_{\text{constituents}} p_T \\ \text{ECF}(2, \beta) &= \sum_{i=1}^n \sum_{j=i+1}^n p_{T,i} p_{T,j} \Delta R_{ij}^\beta \\ (\text{ECF}(3, \beta)) &= \sum_{i=1}^n \sum_{j=i+1}^n \sum_{k=j+1}^n p_{T,i} p_{T,j} p_{T,k} (\Delta R_{ij} \Delta R_{ik} \Delta R_{jk})^\beta \end{aligned} \quad (4)$$

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

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

$$\begin{aligned} \text{C2} &= \frac{\text{ECF}(3) \cdot \text{ECF}(1)}{\text{ECF}(2)^2} \\ \text{D2} &= \frac{\text{ECF}(3) \cdot \text{ECF}(1)^3}{\text{ECF}(2)^3} \end{aligned} \quad (5)$$

290 E.g. a jet originated from a  $W$  boson features a small ECF(3) but a high ECF(2) value resulting in small  
 291 C2/D2, corresponding to a high agreement with the two-prong hypothesis. QCD jets feature a very small  
 292 ECF(3) and a small ECF(2) value. This results, considering the power of ECF(2) in the definitions, in a  
 293 higher C2/D2 value as for a  $W$  boson jet. These variables are IRC-safe for  $\beta > 0$  and theoretically very

294 well understood, see Reference [bib:analytic\_ECF]. D2 was found to perform slightly better for tagging  
 295  $W$  boson jets as C2 in Reference [bib:w\_tagging], most notably due to a more  $p_T$  robust cut value and a  
 296 somewhat higher background rejection.

297 **3.4.2 n-Subjettiness**

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

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

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})^\beta \quad (6)$$

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

$$d_0 = \sum_k p_{T,k} R_0 \quad (7)$$

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

312 Small values of  $\tau_N$  correspond to a jet with all constituents more or less aligned or near to the given  
 313  $N$  sub-jet axes, hence the jet is compatible with the assumption to be composed of  $N$  or fewer sub-jets.  
 314 A higher value in contrast indicates a consistency with more than  $N$  sub-jets as a non negligible part is  
 315 located apart of the  $N$  sub-jet axes. Consequently,  $W/Z$  or Higgs boson jets are likely to feature a small  
 316  $\tau_2$  and a high  $\tau_1$  value. QCD jets with their one-prong structure result in a high  $\tau_2$  and a small  $\tau_1$  value.  
 317 While  $\tau_1$  and  $\tau_2$  alone provide only slightly separation, the ratio

$$\tau_{21} = \frac{\tau_2}{\tau_1} \quad (8)$$

318 is an effective discrimination variable.

319 The extension to three-prong like jet identification and discrimination from one and two-prong structures  
 320 follows quite naturally by taking the ratio of  $\tau_3$  and  $\tau_2$ .

$$\tau_{32} = \frac{\tau_3}{\tau_2} \quad (9)$$

321  
 322 Consequently, the hadronic decay of top quarks via  $t \rightarrow Wb$  and the  $W$  decaying into two quarks can be  
 323 tagged using the  $\tau_{32}$  variable.

## 324 4 Figures of Merit for Performance Studies

### 325 4.1 For jet mass

326 The general idea of Figure of Merit (FoM) is given in the Appendix; here the InterQuantile range  
 327 is described since used in this note and identical to the one used in the conference BOOST 2016.  
 328 The InterQuantile range (IQnR) is here defined as it corresponds to a sigma of a “perfect” Gaussian  
 329 distribution:  $q84\% - q16\%$  where  $q84\%$  is the 84<sup>th</sup> percentile and  $q16\%$  is the 16<sup>th</sup>, not to be confused  
 330 with the InterQuartile Range (IQR) which is the  $q75\% - q25\%$  and does not correspond to the sigma.  
 331 The final descriptor is then divided by the Median ( $\frac{1}{2} \times 68\% \text{ IQnR}/\text{median}$ ). It provides stability and high  
 332 sensitivity to left-hand-side and right-hand-side tails.

333 The IQnR is then applied to the response distribution Figure of Merit: given the reconstructed mass  
 334 (calorimeter, track etc.) one can compare it to its *truth* mass ( $m^{truth}$ ), computed from the particle at MC  
 335 level before the interaction with the detector:

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

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

337 In Figure 4 a mass response for a single range of transverse momentum is shown, for the calorimeter mass.  
 338 On the plot the contours of a standard deviation and of  $q16\%$  and  $q84\%$  are drawn with dashed and solid  
 339 lines, respectively, showing the difference induced by the tail. This sort of plot is the key when looking  
 340 quantitatively to the observable performance and can be found in the Appendix for each of the process  
 341 studied in every  $p_T$  range considered.

### 342 4.2 ROC Cursves and prior mass cut

343 The separation power of discrimination variables can be studied quite intuitively by comparing the signal  
 344 and background distributions of a certain variable. Another used figure of merit for the performance,  
 345 especially for comparisons of different variables, is to use RECEIVER OPERATOR CHARACTERISTICS (ROCs)  
 346 which show the achieved background rejection for different values of signal efficiency (signal fraction left  
 347 after performing a cut). Each point is calculated from the underlying signal and background distributions  
 348 by integrating the background distribution from zero <sup>1</sup> to the point where the desired signal fraction is  
 349 achieved. The fraction of background events contained in this region are kept when cutting at this signal  
 350 efficiency, hence the inverse of this fraction,  $\frac{1}{\epsilon_{background}}$  is an estimate for the background rejection. The  
 351 lower the fraction of background events in the region, the better is the achieved exclusion. Accordingly, a  
 352 good discrimination variable is represented by a ROC with preferably high values of background rejection  
 353 up to high signal efficiencies.

354 \*\*\* add here mass cut \*\*\*

<sup>1</sup> If the signal distribution lies at lower values as the background.

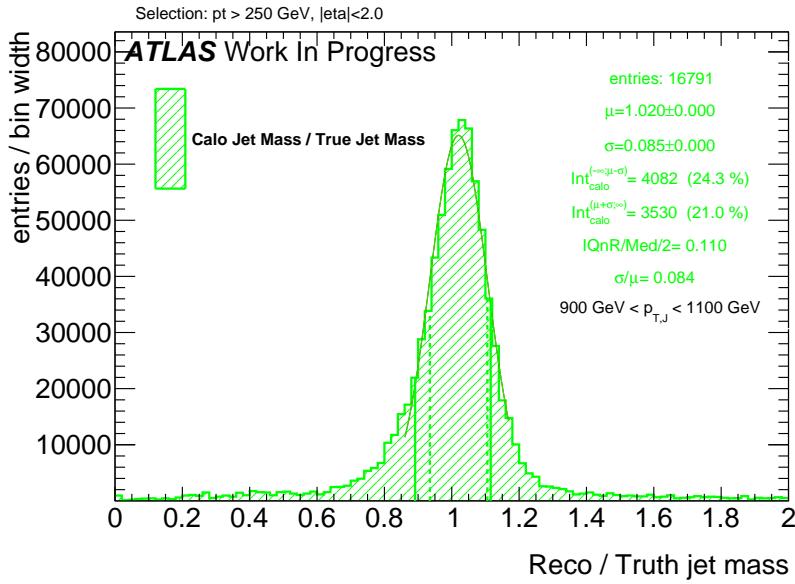


Figure 4: Calorimeter mass response plot for boosted  $W/Z$ . One the plot, right, are shown: the number of entries, the mean and the width of the fit to the Gaussian core, the integral from 0 to  $\mu - \sigma$  and the one from  $\mu + \sigma$  to  $+\infty$ , the values  $\frac{1}{2} \times 68\%$  IQnR/median and  $\sigma/\mu$ . On the distribution the dashed vertical lines represent the points  $\mu - \sigma$  and  $\mu + \sigma$  and the solid lines represent the  $q16\%$  and  $q84\%$ . These lines also explicitly show the asymmetry between the left-hand-side flank, in general more pronounced, and the right-hand-side one

## 355 5 Performance of Track-assisted subjet mass

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

### 358 5.1 Performance in $W \rightarrow q' \bar{q}$ Decays

359 The boosted  $W/Z$  was the first one looked at, and with which the  $m^{TAS}$  was designed. The  $m^{calo}$  shows a  
360 fast deterioration of the performance at high  $p_T$ , and, as shown in the previous section, the  $m^{TA}$  prevents  
361 this deterioration but suffers at low transverse momenta ( $p_T < 1 \text{ TeV}$ ). The  $m^{TAS}$  has a similar behavior in  
362 the extreme transverse momentum regime as the  $m^{TA}$ , since the sub-jet multiplicity peaks at one, where  
363 there are no differences between the two observables. In the low- $p_T$  regime, on the contrary, it exploits  
364 the difference in charged to neutral ratio for each sub-jet, achieving a better performance. This is shown  
365 in Figure 5(a) as a function of  $p_T$ : below  $\sim 1 \text{ TeV}$  achieves lower values of the IQnR converging from  
366 below to the  $m^{TA}$  as the number of sub-jets decreases to one.

### 367 5.2 Performance in $h \rightarrow b\bar{b}$ Decays

368 In the Randall-Sundrum graviton to di-Higgs to four b-quark, the performance is again problematic for the  
369  $m^{TA}$  with respect to  $m^{calo}$ , which is far beyond the latter, while the performance of the  $m^{TAS}$  is partially  
370 similar to the boosted top-quark sample, but degrades much more in the extreme  $p_T$  regime, following the  
371  $m^{TA}$ . Shown in Figure 5(b).

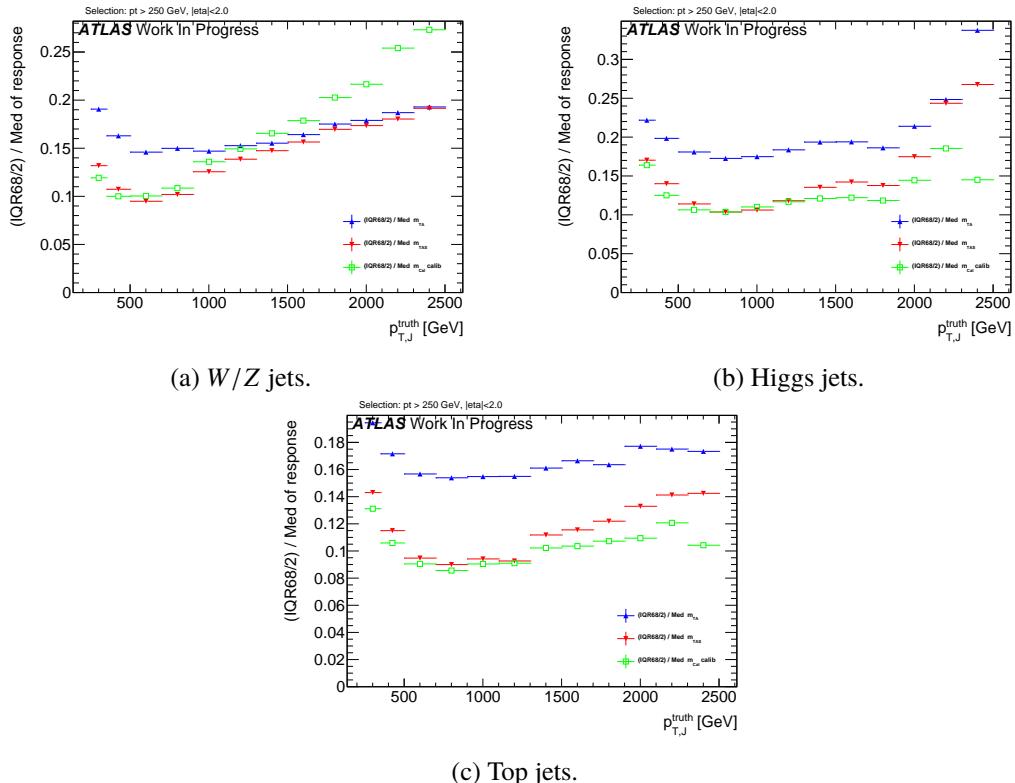


Figure 5: Performance of the  $m^{TAS}$  versus the  $m^{calo}$  and  $m^{TA}$  for  $W/Z$ , top left, where  $m^{TA}$  is not better than  $m^{calo}$  in the low  $p_T$  range but is outperformed by the  $m^{TAS}$ ; Higgs decay, where  $m^{calo}$  is everywhere better than  $m^{TA}$ , yet comparable with  $m^{TAS}$  and top decays where the more complex topology makes critical the high  $p_T$  regime

### **372 5.3 Performance in $t \rightarrow q'\bar{q}b$ Decays**

373 The boosted tops are shown on Figure 5(c); the  $m^{TAS}$  is comparable yet slightly worse than the  $m^{calo}$   
 374 in the low-middle  $p_T$  regime, while degrades at higher  $p_T$  approaching the  $m^{TA}$ , which is far beyond the  
 375 track-assisted sub-jet mass in performance. As already noted, the worse performance can be ascribed both  
 376 to the higher top-quark mass, and to its different and more complex decay topology.

## 377 5.4 Performance in QCD Multijet Events

<sup>378</sup> The behavior of the QCD multijet sample is similar to the boosted  $W/Z$  sample, where the  $m^{TA}$  exhibits  
<sup>379</sup> a crossing point in the middle-low regime  $p_T \simeq 900$  GeV and proceeds with a better performance at high  
<sup>380</sup> transverse momenta. Again the  $m^{TAS}$  follows this similarity showing no crossing point and an optimal  
<sup>381</sup> overall behavior, both with respect to calorimeter- and track-assisted-based mass definition. On Figure  
<sup>382</sup> 6.

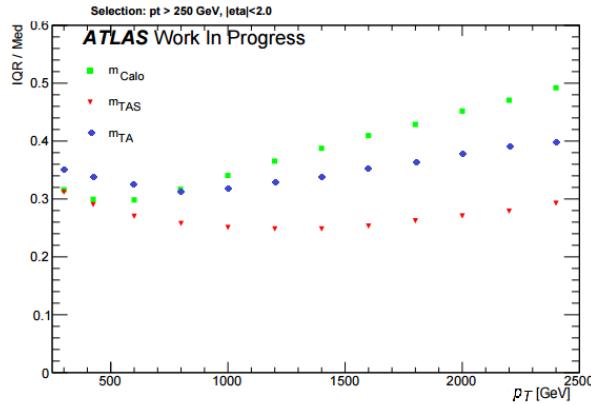


Figure 6: Performance of the  $m^{TAS}$  versus the  $m^{calo}$  and  $m^{TA}$  for the QCD multijet shows a much better behavior of the track-assisted sub-jet mass. Here shown 50% IQnR/median and not the  $\frac{1}{2} \times 68\%$  IQnR/median.

### 383 5.5 Performance in Massive $\tilde{W} \rightarrow q'\bar{q}$ Decays with $m_{\tilde{W}} = m_t$

384 The massive  $W$  sample is a special sample which was used to understand the behavior of the boosted tops,  
 385 whether its worse resolution was coming from the higher mass of the top quark or from the more complex  
 386 decay topology (three-pronged instead of two-pronged decay and  $b$ -quark presence). The sample is almost  
 387 identical to the boosted  $W/Z$  one ( $W' \rightarrow WZ$ ) but in this case the SM electroweak boson have the mass  
 388 of the top quark  $m_{\tilde{W}} = m_t$ . In fact, from the rule  $\Delta R = 2m/p_T$ , a bigger separation is expected between  
 389 quarks from the hadronic decay. The comparison with  $m^{calo}$  is shown in Figure 7, together with the  
 390 boosted top-quark for completeness. As seen here, the performance of the latter is clearly worse than the  
 391 former, the trend is yet very similar. This difference is interpreted in terms of different and more complex  
 392 topology and hence higher sub-jet multiplicity: in the three sub-jet structure, resolving accurately the  
 393 components is more challenging.

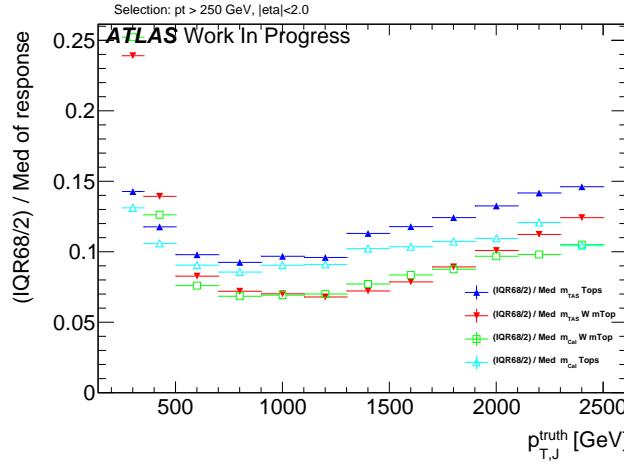


Figure 7: Performance of the  $m^{TAS}$  versus the  $m^{calo}$  for the massive  $W/Z$  (in red and green); shown on the same plot also the boosted top sample (in blue and light blue).

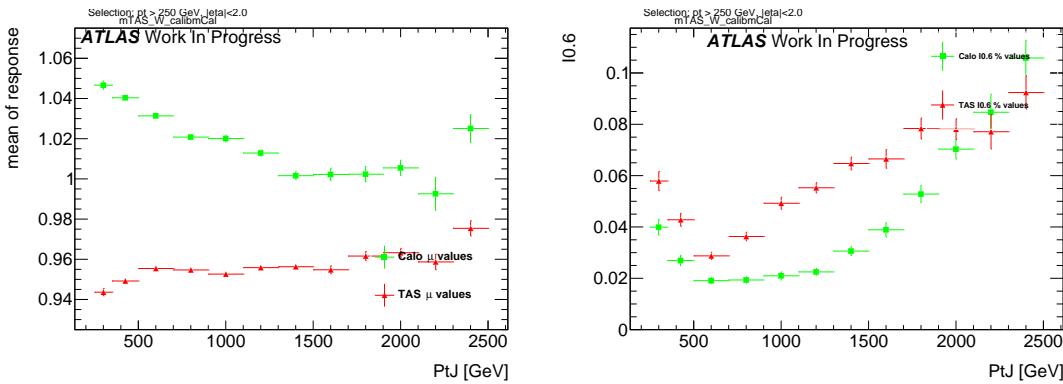


Figure 8: Stability quantifiers which were checked for the  $m^{TAS}$ : mean on the left and normalized left-hand side integral of the mass response distribution on the right. The mean is calculated from a Gaussian fit and the integral goes from 0 to 0.6.

### 394 5.6 Stability of Mean of Response and Left-Hand-Side Integral

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

400 The mean of the response distribution is shown for boosted  $W/Z$  decays in Figure 8, left; as seen here,  
 401 despite the mean being constantly below unity, its behavior is much more flat and independent of  $p_T$ ,  
 402 especially in the low-intermediate regime. This is surprising since the  $m^{calo}$  is already shown after the  
 403 calibration step, which is not taken instead for the  $m^{TAS}$ . Conversely the left-hand-side tail of the mass  
 404 response which is shown in the same figure, right, shows a more enhanced behavior than the  $m^{calo}$ , but  
 405 still never reaches the 10%. Of course an enhancement of the tail causes a loss of gaussianity and a  
 406 number of jets which are reconstructed with a lower mass than they should, but it is still comparable with  
 407 the calorimeter mass.

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

### 410 5.7 Potential Improvements from Sub-jet Calibration

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

414 The *perfect calibration* refers to the procedure of using  $m^{TAS}$  with truth-level information for calorimeter  
 415 and tracker system, i.e. looking at the best possible scenario with an ideal detector. The performance is of  
 416 course expected to be optimal, because of the use of the truth-level. This step was necessary as feasibility  
 417 study, to understand whether ulterior efforts in this direction were meaningful. Truth-level tracks are the  
 418 particles in the jet which have an electric charge and are stable, truth-level sub-jets are all the particles,

419 charged and not, which are ghost associated to the calorimeter sub-jets. There are few possibilities in  
 420 doing so, here some nomenclature for this study will be introduced:

- 421 •  $m^{TAS}$  using truth-level sub-jets and tracks; normal tracks (with all detector effects) are used to assist  
 422 the truth-level sub-jets;
- 423 •  $m^{TAS}$  using truth-level tracks and truth-level sub-jets; the truth-level tracks are used to assist the  
 424 truth-level sub-jets;
- 425 •  $m^{calo}$  truth, calculated using only the truth sub-jets.

#### 426 5.7.1 Simple Sub-jet Calibration

427 The perfect calibration using truth level sub-jets and tracks is shown in Figure 9 in blue dots; since the  
 428 performance exhibits room for big improvement below  $\sim 1$  TeV and moderate to small improvement above  
 429 this value, the second step of a simple calibration was tried.

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

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

439 This procedure was called *poor man's calibration* or PM calibration or *simple calibration*. A check on  
 440 the  $p_T$  response before and after calibration together with the mean of the entire Large- $R$  jet response is  
 441 shown in Figure 43 and 44 in Appendix.

442 The results are on Figure 9; there are only marginal improvements in few ranges of low transverse  
 443 momentum where the scale factors are further away from unity, and the overall observable is not performing  
 444 better than the standard  $m^{TAS}$ . This is interpreted both in terms of a missing calibration as a function of  
 445 the  $\eta$  variables (having hence a befit from the crack region) and because the correction done on average  
 446 does not provide the sufficient handle in a jet-by-jet basis, especially when all the sub-jets are rescaled by  
 447 similar factors (which translates into a similarity of  $p_{TS}$  of the sub-jets, often the case for e.g. boosted  
 448  $W/Z$ , less for boosted tops entirely contained in the large- $R$  jet).

#### 449 5.8 Limitation of $m^{TAS}$ from tracking

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

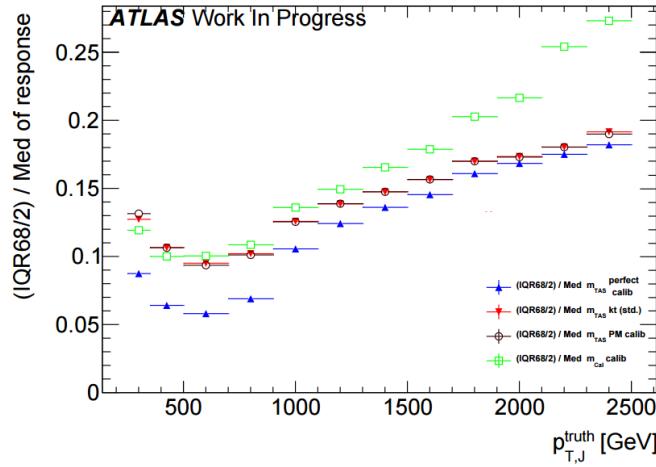


Figure 9: Performance of the poor man’s calibration. The improvement is marginal throughout the entire transverse momentum space.

454 The preliminary investigation in this direction was then the study on the track mass resolution: a response  
 455 of the mass of the tracks associated to the jet ( $m^{track}$ ), was constructed, using the truth-level tracks.

456 The result is shown on Figure 10: for the samples considered, it shows a linear degradation of the mass of the  
 457 tracks associated to the jet ( $m^{track}$ ), both for massive and SM  $W/Z$ .

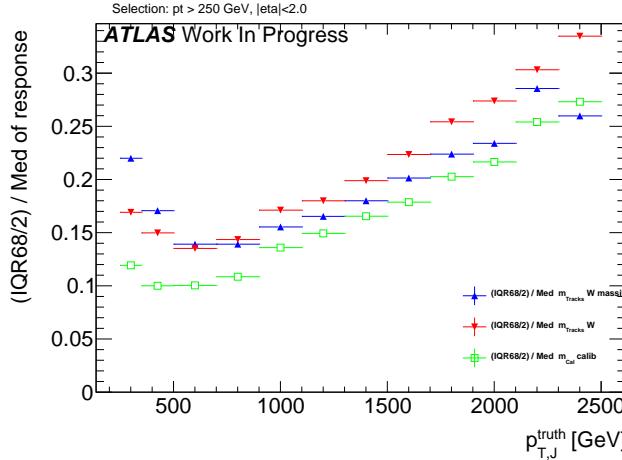


Figure 10: The performance of the track mass ( $m^{track}$ ) in blue and red for massive  $W$  sample and boosted  $W/Z$  respectively; for reference in green the calorimeter mass of the large- $R$  jet.

458 The hypothesis of the degradation of the  $m^{TAS}$  driven by the tracks is also supported by the Figure 45  
 459 in Appendix, where the truth-level tracks are used instead of real tracks to compute the variable; it can  
 460 be seen the flat behavior at high  $p_T$ , hence ascribing the worsening of the resolution to tracks at higher  
 461 transverse momenta.

462 A complete breakdown of the variable in terms of truth-level particles is given in Figure 11(a), where  
 463 all the different components are separated. In particular the black dots show the  $m^{TAS}$  using truth-level  
 464 sub-jets but real tracks for the track assistance procedure. Even combining this truth-level information, in  
 465 fact, it shows a large worsening of the performance (truth-level sub-jets only are shown as blue dots). On

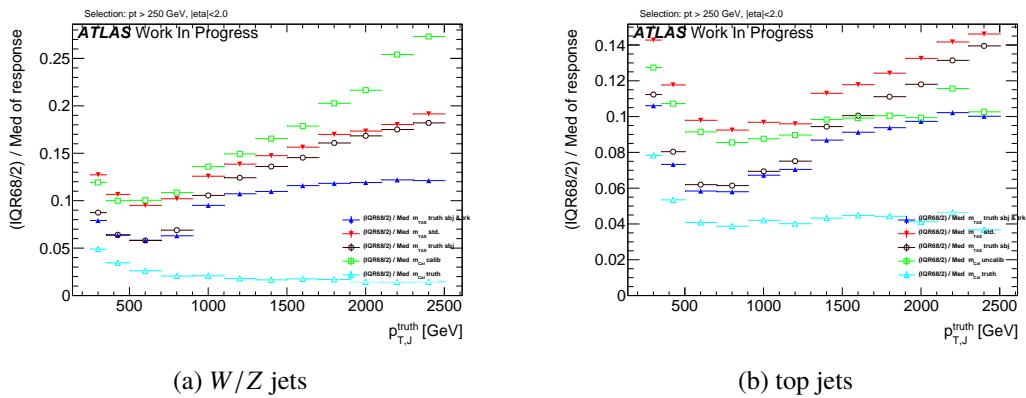


Figure 11: Breakdown of the  $m^{TAS}$  in its component using truth-level information for boosted  $W/Z$  decays, on the left. In blue the  $m^{TAS}$  using truth-level sub-jets and truth level tracks, in black  $m^{TAS}$  using truth level sub-jets but real tracks and in light blue for reference the mass of the truth level particles associated to the sub-jets. As usual, in red and green the standard  $m^{TAS}$  and the  $m^{calo}$ . On the right the same for top jets.

the other side using again truth-level tracks for the track assistance procedure of the truth-level sub-jet, shows a recovery of the loss in performance.

Additional studies on the limitation of the  $m^{TAS}$  based on MC studies without detector interactions are also presented. In particular, the truth study presented for boosted  $W/Z$  decay in were extended for boosted top quark decays.

As seen on Figure 11(b), the breakdown of the  $m^{TAS}$  shows that, in particular for the high transverse momenta regimes, the tracks are subjected to fast degradation which makes their combination with the calorimeter mass not anymore an advantage.

474 This is a limitation which was expected and understood from the detector performance point of view,  
 475 and here shows the impossibility, with the variables which are presented here  $m^{TA}$  and  $m^{TAS}$  to reach a  
 476 competitive standpoint with the  $m^{calo}$  in the extreme kinematic regime for the top quark decay.

477 In black, in fact, the performance of the  $m^{TAS}$  variable using tracks with detector effect and sub-jets  
 478 without those effects, shows this intrinsic limit which takes place already at 1.5 TeV.

479 The crossing point is, as already pointed out for the boosted tops, present because of the optimal perform-  
480 ance of the calorimeter system caused by the higher mass of the top quark, and partially also because of  
481 its more complex decay structure and difficulty to be resolved in sub-jets.

482 5.9 Performance with Alternate Inputs to the  $m^{TAS}$

483 There are quite a few ways to modify the track-assisted sub-jet mass; however, all the alternative approaches  
 484 showed worse performance, and they are mentioned here for completeness only. The per-track four  
 485 momentum correction scheme which is used for the ECF and the n-Subjettiness and also explored with  
 486 the  $m^{TAS}$  with no significant difference was described in [3.2.2](#).

<sup>487</sup> The other alternatives considered were:

- 488 • for the tracks:

- 489     – use of tracks not as input directly, but only taking those belonging to anti- $k_t$  reclustered  
 490       track-jet with radius of 0.3 or 0.2;  
 491     – tighter or looser quality conditions were explored;  
 492     – tighter or looser primary vertex association requirement were explored.  
 493     • for the sub-jets:  
 494       – the trimming procedure was modified: various radii  $R_{sub}$  of the sub-jets were tested;  
 495       – the sub-jets were reclustered using not only the standard  $k_t$ , but also anti- $k_t$  and C/A.  
 496     • for the procedure: different 4-momentum correction scheme was also studied in more details, see  
 497       [3.2.2](#).

498     The different reclustering algorithm choice has a deep impact and was studied in details, since it changes  
 499       the topo-cluster added to the sub-jets and the tracks associated to them. The situation is depicted in the  
 500       event-display in Figure 12; the display on the left shows the standard choice of  $k_t$ , the one on the right  
 501       shows the modified approach anti- $k_t$ .

502     In Figure 13(a) 13(c) 13(b) the performance for boosted  $W/Z$ , tops and Higgs are shown, respectively. It  
 503       can be seen that the  $k_t$  algorithm provides the best observable definition, in all the samples considered.  
 504     However, the anti- $k_t$  algorithm provides similar performances; this was an important check as the jet  
 505       calibration procedure currently going on in ATLAS, the  $R$ -Scan procedures includes the anti- $k_t$  algorithm  
 506       with radius of  $R=0.2$  and aims at providing the calibration and uncertainties that could be used directly in  
 507       the computation of the  $m^{TAS}$ .

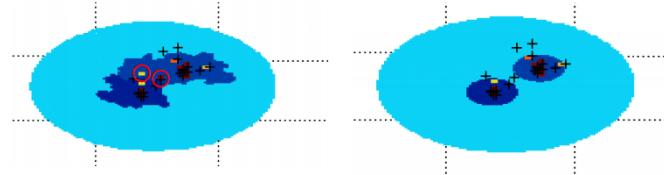


Figure 12: An example of event-display shows the differences in the reclustering algorithm used for the sub-jets: on the right  $k_t$  and on the left anti- $k_t$ . Highlighted some constituents trimmed away with the second choice.

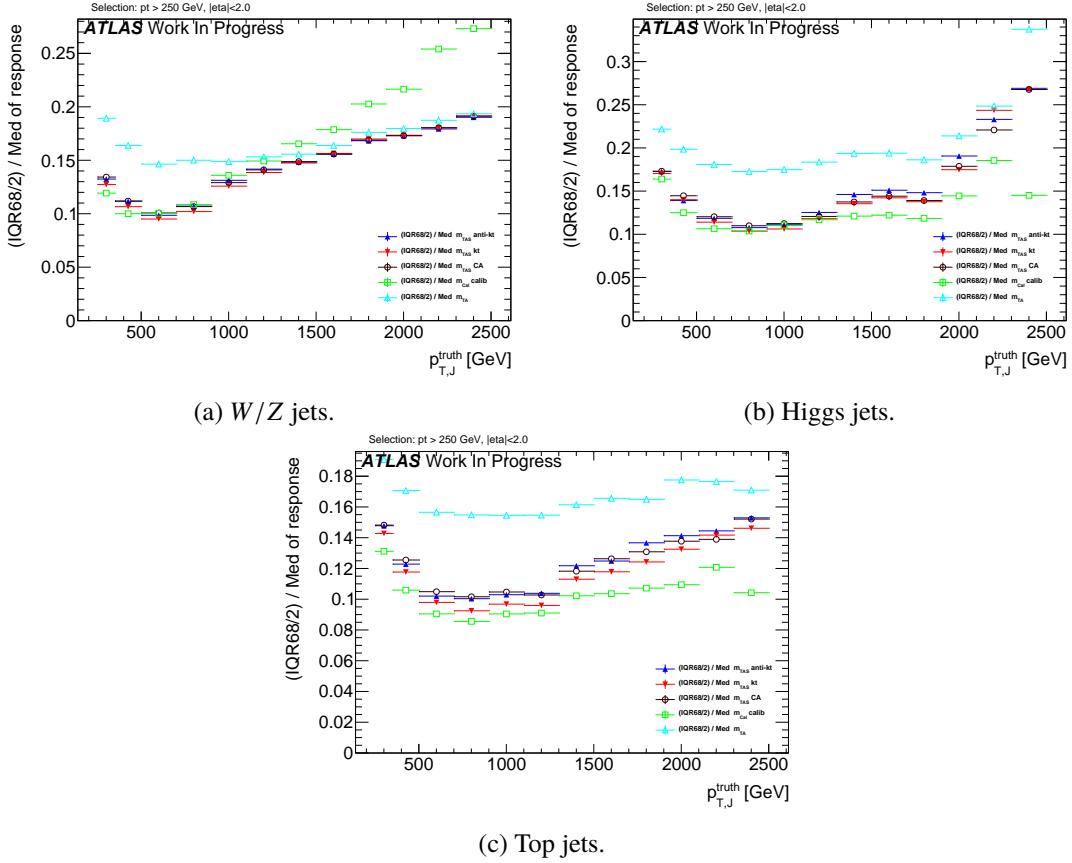


Figure 13: Performance of  $m^{TAS}$  with different reclustering algorithms for the sub-jets: anti- $k_t$ ,  $k_t$  and C/A and for  $W/Z$  jets, top left, Higgs jets, top right and top jets, bottom. In all the cases shown, the  $k_t$  is producing the better results, but all the three have a very similar performance.

## 508 6 Performance of Combined Calorimeter and Track-Assisted Sub-Jet 509 Mass

510 This section presents the achievement of the variable obtained combining the  $m^{TAS}$  and the  $m^{calo}$ , the  
511  $m_{TAS}^{comb}$  with respect to the combination of the  $m^{TA}$  and the  $m^{calo}$ , the  $m^{comb}$ . Both these variables were  
512 defined in 3.3

### 513 6.1 Performance in $W \rightarrow q' \bar{q}$ Decays

514 On the boosted  $W/Z$ s sample, the  $m_{TAS}^{comb}$  outperforms all the other definitions throughout all the transverse  
515 momentum space; on Figure 14(a) they are shown for reference together with the  $m^{TAS}$ . It can be noted  
516 here that the track-assisted sub-jet mass, although being sub-optimal, has comparable performance, yet  
517 presenting fewer complications due to the combination procedure.

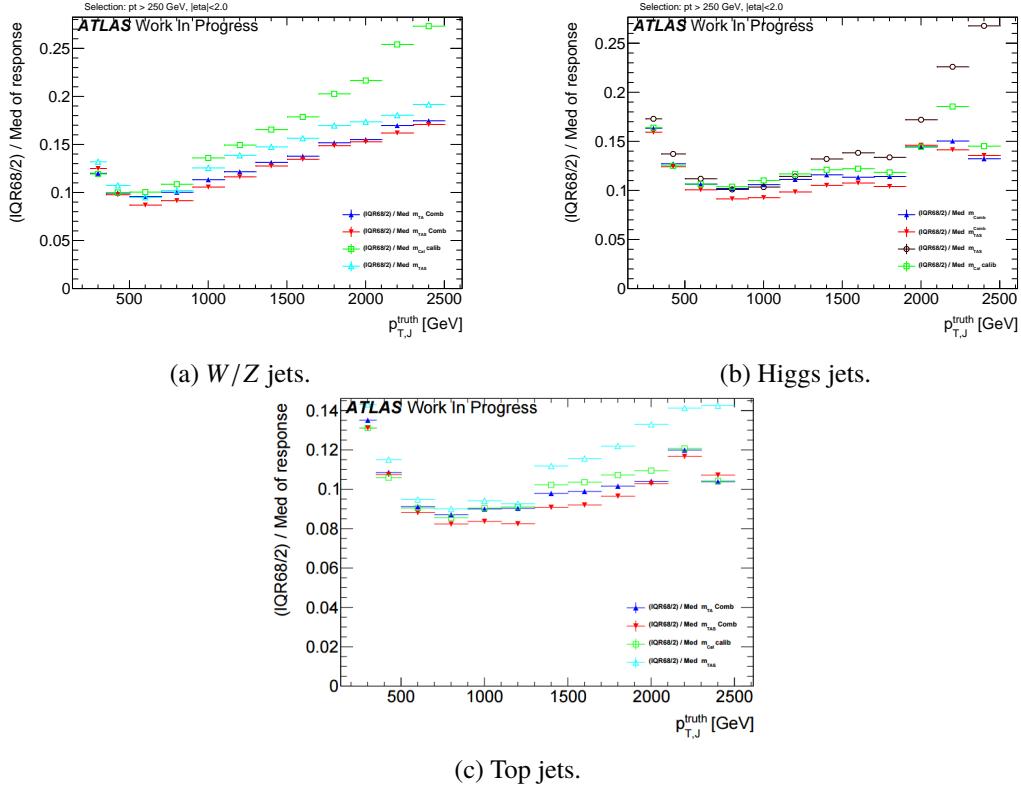


Figure 14: Performance of  $m^{comb}$  and  $m_{TAS}^{comb}$  for different samples: the  $W/Z$  jets, top left, the Higgs jets, top right and the top jets, bottom. The  $m_{TAS}^{comb}$  outperforms the other definitions throughout the whole spectrum of transverse momentum. The  $m^{TAS}$ , although being sub-optimal follows with similar performance the  $m^{comb}$ . The Higgs and top jets presents the same properties as shown before, and the combined mass reflects these properties.

## 518 6.2 Performance in $h \rightarrow b\bar{b}$ Decays

519 Again, for the Higgs decay there are similarities as for the top sample; on Figure 14(b) the two definitions  
 520 of the combined mass, together with the simpler  $m^{TAS}$ . Although this variable is slightly sub-optimal yet  
 521 still comparable in the low to intermediate range in transverse momenta, where the tracks are driving a  
 522 decrease in performance for the high to very-high  $p_T$ . The  $m_{TAS}^{comb}$  uses this advantage to achieve optimal  
 523 behavior in the entire transverse momentum spectrum, outperforming both  $m^{calo}$  and  $m^{comb}$  almost  
 524 everywhere.

## 525 6.3 Performance in $t \rightarrow q'\bar{q}b$ Decays

526 The boosted top sample remains the most challenging one also with the combined mass; as seen on  
 527 Figure 14(c), the  $m^{comb}$  performs quite similarly to the calorimeter based mass definition, yet behaving  
 528 considerably better than the  $m^{TAS}$  especially at high transverse momentum. The  $m_{TAS}^{comb}$ , however,  
 529 outperforms all the other definitions, and shows its optimal observable strength at intermediate  $p_T$  i.e. in  
 530 the range  $0.8 < p_T < 1.6$  TeV.

## 531 7 Performance of Jet Substructure observables with (assisted) tracks

532 Tracks and their angular resolution could not only improve the jet mass definition but also the performance  
 533 of tagging variables such as the Energy Correlation Functions or n-Subjettiness. These variables are  
 534 usually calculated with calorimeter clusters as input, studied here are tracks and assisted tracks as input in  
 535 comparison with the default method using clusters. In contrast to the  $m^{TA}$  variable introduced in Section  
 536 ??, not the mass but the  $p_T$  of each track is scaled, since C2, D2,  $\tau_{21}$  and  $\tau_{32}$  are calculated with the  
 537 constituents  $p_T$ .

538 The concept of track assisting with the  $p_T$  ratio of the whole jet is without effect for the studied substructure  
 539 variables. This can be understood from the definitions of the weighted  $p_T$  sums. If corrected with only  
 540 one ratio, all tracks are scaled by the same factor  $c$ , which then can be put in front of the sum and cancels  
 541 as soon as the ratios  $\tau_{21}$  and  $\tau_{32}$ , respectively C2 and D2 are formed.

$$\begin{aligned}\tau_N &= \frac{1}{d_0} \sum_k p_{T,k} c \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})^\beta \\ &= \frac{c}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k})^\beta\end{aligned}\quad (10)$$

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

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

544 This leads to the following adaption of the TAS procedure:

$$\begin{pmatrix} m_{track} \\ p_{T,track} \\ \eta_{track} \\ \phi_{track} \end{pmatrix} \rightarrow \begin{pmatrix} m_{track} \\ p_{T,track} \frac{p_{T,sub-jet}}{\sum_{ga\,tracks} p_{T,track}} \\ \eta_{track} \\ \phi_{track} \end{pmatrix} \quad (12)$$

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

### 546 7.1 Event weighting and Mass-Cut

547 The substructure variables are compared via their QCD (multi-jet) rejection performance. While the  $p_T$   
 548 distribution of the multi-jet sample falls exponentially, the  $p_T$  of the signal samples features characteristic  
 549 peaks related to the different resonance masses, see Figure 15. To avoid bias in the comparison, the  
 550 signal sample is given weights such that the truth  $p_T$  distribution of the leading jet matches the one of the  
 551 background sample. Furthermore, the spectrum is split into six different  $p_T$  regions to study the behavior  
 552 with rising energy.

553 Tagging variables such as C2, D2,  $\tau_{21}$  and  $\tau_{32}$  are usually used after applying a mass cut around the interval  
 554 that contains 68% of the signal events. Therefore, a cut is applied on the calibrated mass of the large-R  
 555 calorimeter jet which is calculated to cover the smallest interval around the peak mass that contains 68% of  
 556 the signal events. The comparison is performed in six different  $p_T$  regions to study the behavior connected

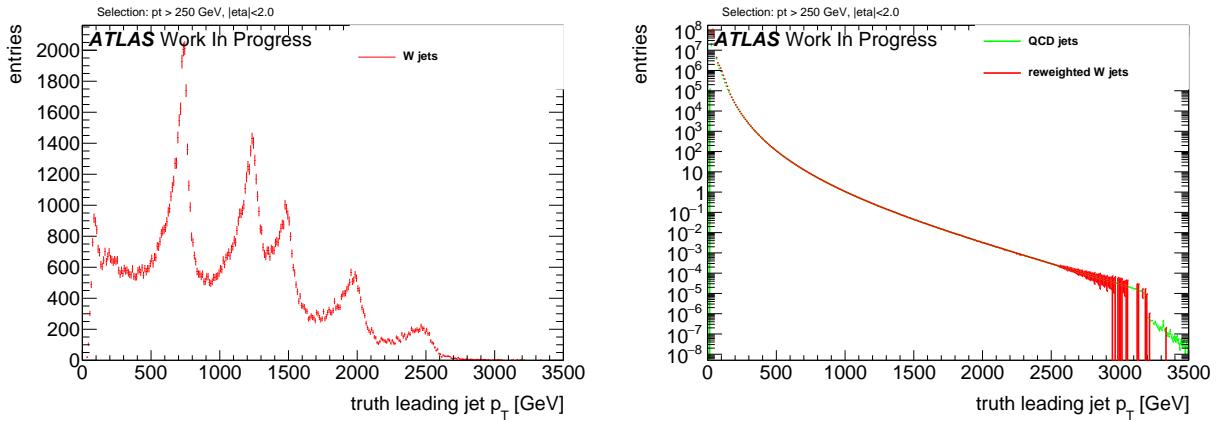


Figure 15: Exemplary  $p_T$  distributions of (left)  $W$  boson jets and (right) QCD jets from multi-jet events with reweighted  $W$  boson events

$p_T$ [GeV]	W boson		Higgs boson		Top quark	
	Mass [GeV]	$\frac{1}{\epsilon_{b,gr}}$	Mass [GeV]	$\frac{1}{\epsilon_{b,gr}}$	Mass [GeV]	$\frac{1}{\epsilon_{b,gr}}$
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.

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

## 7.2 Track Selection

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

The distributions showing the number of tracks associated to a calorimeter jet, see the left side of Figure 16, indicate, that on average around four tracks less are associated to the sub-jets compared to the ungroomed jet. The right side of Figure 16 shows the angular distance  $\Delta R$  between the single tracks and

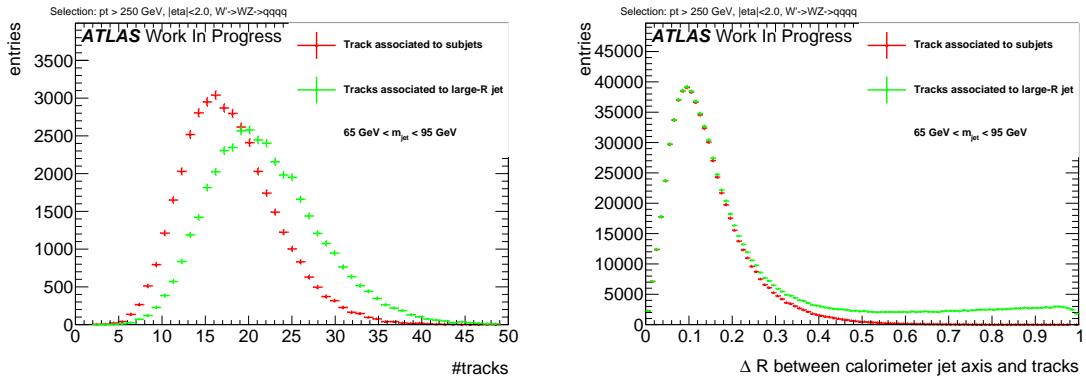


Figure 16: The number of tracks ghost associated to the large-R jet and to the sub-jets (left) and angular distance of associated tracks to the large-R calorimeter jet axis (right). Signal events were not reweighted at this step.

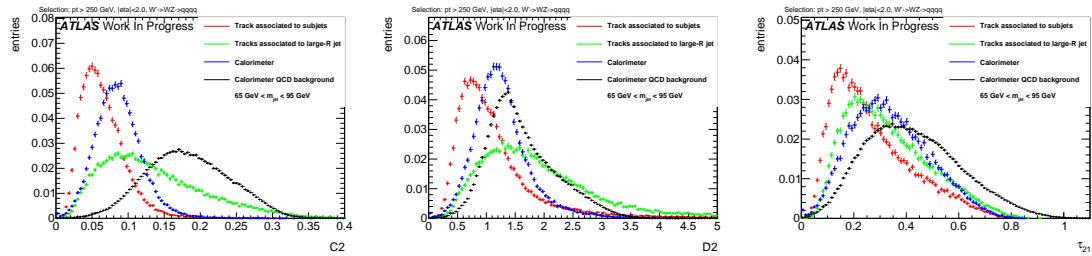


Figure 17: Substructure variables (left) C2, (right) D2 and (below)  $\tau_{21}$  calculated with calorimeter clusters as well as tracks associated to sub-jets and to the large-R jet. Signal events were not reweighted at this step.

the axis of the large-R calorimeter jet. Both distributions are aligned in the lower  $\Delta R$  region while the histogram representing the tracks associated to the ungroomed jet shows an enhancement towards larger  $\Delta R$ . Accordingly, these additional tracks feature an angular separation from the jet axis of more than 0.3, and are in consequence distributed primarily around the outer regions of the large-R jet. Given the required primary vertex association, it is unlikely that these tracks originate from pile-up. Instead, the origin might be found in final- or initial state radiation.

Figure 17 shows the signal distributions of the C2/D2, and  $\tau_{21}$ , calculated with both selections of tracks for  $W$  boson jets. The large  $\Delta R$  to the jet axis of the differing tracks push the substructure variables to higher, more background like values. The broader distributions are a result of the variating nature of these tracks. C2 and D2 are more sensitive to tracks with a large  $\Delta R$  to the jet axis, because the angular distance between all pairs and triples of tracks is considered, among tracks on possibly opposite ends of the large-R jet, whereas  $\tau_{21}$  uses distances to  $k_T$ -WTA axes. For comparison, the signal and background distributions for the variables calculated with calorimeter clusters are shown as well. It is possible to anticipate that the performance of variables calculated with tracks and assisted tracks is not worse than cluster base variables. In contrast to the previously studied jet mass variable, ratios of ECF(N) and  $\tau_N$  are rather energy scale independent and are found to not be as sensitive to the missing neutral fraction with un-assisted tracks. Starting from this observations, the performance of substructure techniques is compared with the following objects as input:

- Calorimeter clusters, labeled 'calo'.
- Tracks selected as described in Section ??, labeled 'tracks'.

- 593 • The same collection of tracks, assisted as defined in Section 7, labeled 'TAS'.

### 594 7.3 Performance with default $\beta$

595 The performance of track and TAS based tagging variables with the default angular weighting of  $\beta =$   
 596 1 is compared to the corresponding calorimeter variables for  $W$  boson, Higgs boson and Top quark  
 597 identification. The stated signal efficiencies are calculated after the mass cut plus tagging with n-  
 598 Subjettiness or C2/D2. Therefore, the endpoint of the ROCs is at 68% signal efficiency, the fraction kept  
 599 after the mass cut. Consequently, it is required to achieve a tagging only signal efficiency of  $\frac{0.5}{0.68} \sim 0.74$   
 600 for a signal efficiency of 0.5 after mass cut and tagging. Similarly, the stated and compared background  
 601 rejections result from the multiplication of both, thus representing the QCD rejection of the combined  
 602 mass- and tagging variable-cut. The complete set of signal and background distributions for different  
 603 inputs can be found in the Appendix ??.

#### 604 7.3.1 Performance for $W$ boson tagging

605 Shown in Figure 18 are exemplary signal and background distributions in intermediate  $p_T$  bins compared  
 606 for different inputs. This shows throughout narrower signal distributions rising slightly sharper for tracks  
 607 and assisted tracks compared to calorimeter clusters which can be explained by the high angular resolution.  
 608 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.

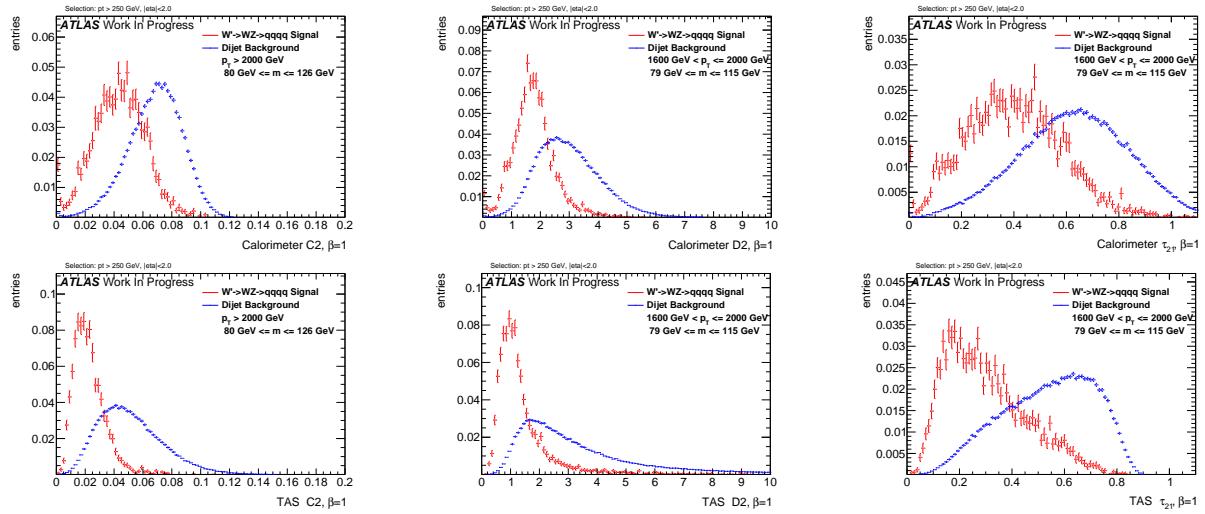


Figure 18:  $W$  boson signal and QCD background distributions for calorimeter (left) and TAS (right) at  $\beta = 1$  with C2 (top) for more than 2000 GeV and D2 (middle) and  $\tau_{21}$  (bottom) for 1200-1600 GeV

609

610 The ROCs in Figure 19, 20 and 21 show the actual achieved background rejection at different  $p_T$  values.  
 611 For lower  $p_T$  values, TAS perform comparably to calorimeter clusters. Tracks without assisting achieve  
 612 a considerably lower background rejection with D2 and  $\tau_{21}$  for lower energies. Tracks and TAS perform  
 613 equally well at high energies for D2 and  $\tau_{21}$  and for C2 over the whole studied range. At higher boosts,

614 the angular resolution of the tracks becomes more and more relevant as the separation between jet  
 615 constituents shrinks. Consequently, tracks and TAS start to outperform calorimeter based variables and  
 become increasingly effective with rising energy.

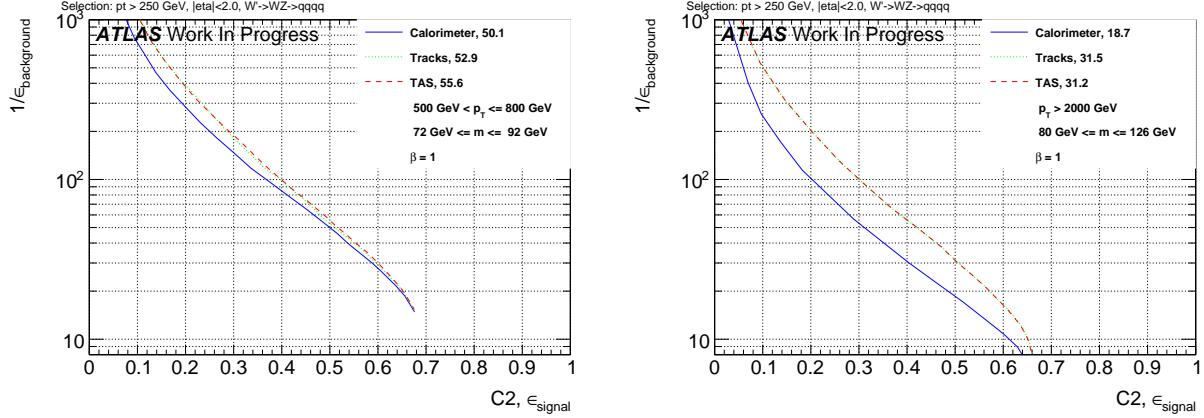


Figure 19: ROCs showing QCD rejection against  $W$  boson efficiency for tracks, TAS and colorimeter  $C_2$  at  $\beta = 1$  for 500-800 GeV (left) and  $> 2000$  GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

616

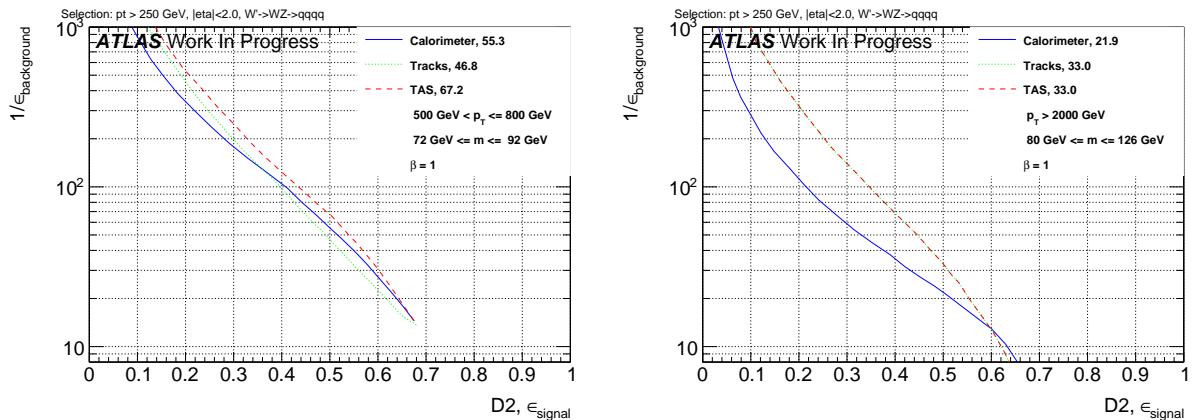


Figure 20: ROCs showing QCD rejection against  $W$  boson efficiency for tracks, TAS and colorimeter  $D_2$  at  $\beta = 1$  for 500-800 GeV (left) and  $> 2000$  GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

### 617 7.3.2 Un-assisted tracks and TAS at very high $p_T$

618 The  $C_2$  variable was found to perform equally well with tracks and TAS as input. This variable seems to  
 619 be relative insensitive to the track assisting and tracks alone already perform well.  $D_2$  and  $\tau_{21}$  in contrast,  
 620 feature a visibly worse separation with tracks than with assisted tracks. In these cases, the scale difference  
 621 due to the missing neutral fraction seems to have a greater influence.

622 For very high  $p_T$  values however, it is often the case that the large-R calorimeter jet features only one  
 623  $R = 0.2$  sub-jet after trimming due to the now small separation of constituents. A single sub-jet results in  
 624 the TAS procedure to fall back to TA. As stated in Section 7, TA has no impact on the ratios. Therefore,  
 625  $C_2/D_2$  and  $\tau_{21}$  perform equally well when calculated with tracks or TAS for events with only one sub-jet  
 626 and thereby the difference between both decreases for very high energies.

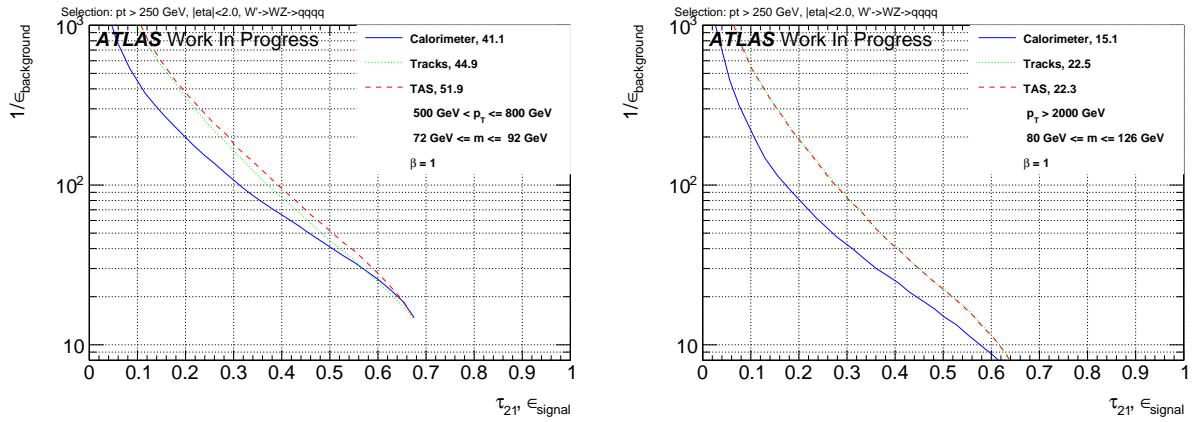


Figure 21: ROCs showing QCD rejection against  $W$  boson efficiency for tracks, TAS and calorimeter  $\tau_{21}$  at  $\beta = 1$  for 500-800 GeV (left) and  $>2000$  GeV (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

### 627 7.3.3 Correlation with $p_T$

628 Due to the rapidly falling  $p_T$  spectrum and hence low weights for high  $p_T$  are the correlation plots divided  
 629 into the six different  $p_T$  regions. For C2, see Figure 22, one can observe a strong trend to lower values  
 630 for signal and background with calorimeter clusters as well as TAS. Furthermore, it is possible to observe  
 631 that the TAS distributions concentrate at lower values compared to calorimeter counterparts.

632 In the cases of D2, Figure 23, and  $\tau_{21}$ , Figure 24, there is a small upward trend of the calorimeter variables  
 633 visible in the lower  $p_T$  regions which, with rising boost, slows down for D2 and  $\tau_{21}$  and ends for  $\tau_{21}$  in  
 634 a broader distribution. This verifies the higher  $p_T$  dependence of the C2 variable in comparison to D2  
 635 and  $\tau_{21}$ . The TAS counterparts feature an even more robust signal with the background moving to higher  
 636 values, hence improving separation. The  $p_T$  dependence of variables calculated with tracks is very similar  
 637 to the ones with TAS, therefore they are omitted.

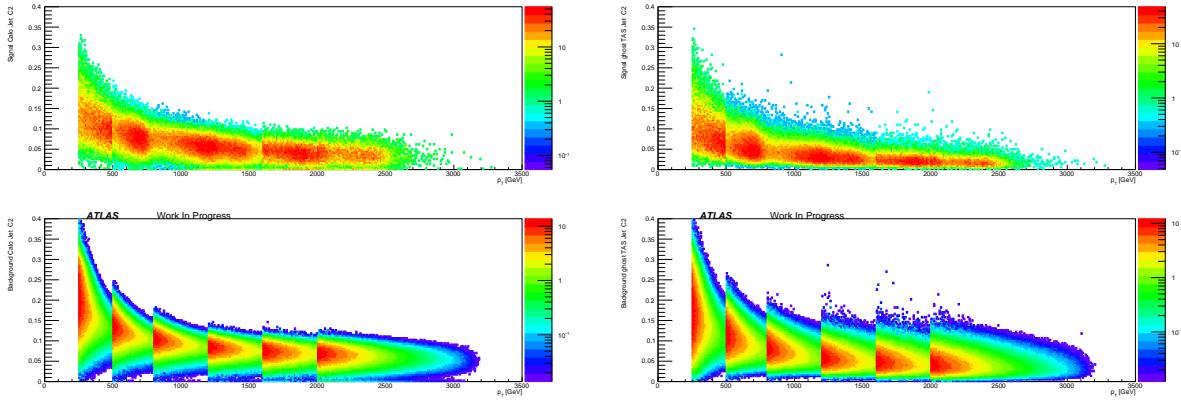


Figure 22: Correlation between C2 at  $\beta = 1$  and  $p_T$  applied on  $W$  boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

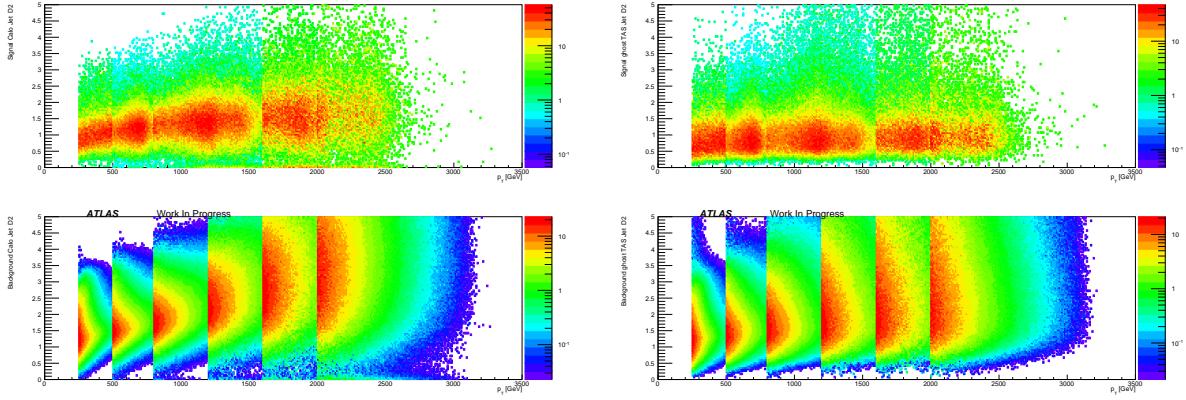


Figure 23: Correlation between D2 at  $\beta = 1$  and  $p_T$  applied on  $W$  boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

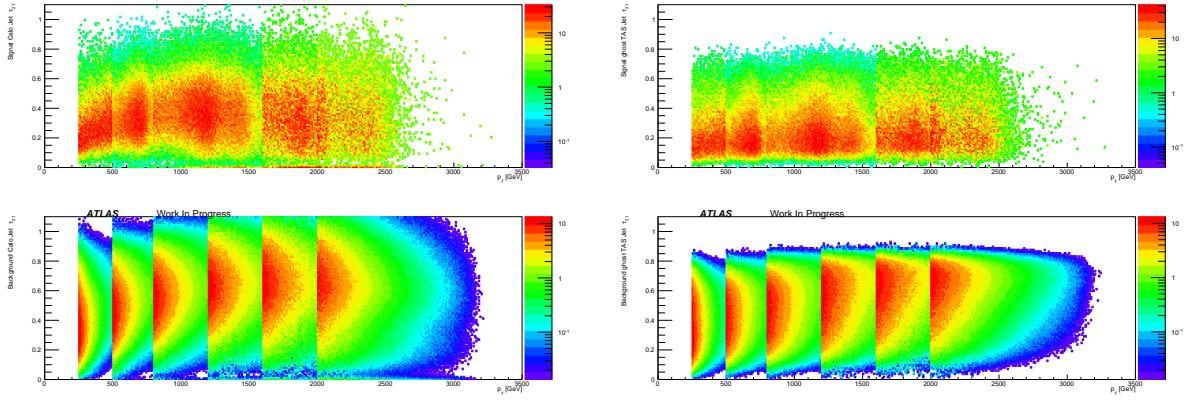


Figure 24: Correlation between  $\tau_{21}$  at  $\beta = 1$  and  $p_T$  applied on  $W$  boson signal (above) and QCD background (below) for calorimeter (left) and TAS (right).

### 638 7.3.4 Performance for Higgs boson tagging

639 The Higgs boson is heavier than the  $W$  or  $Z$  boson, resulting in a higher angular separation of the jet  
 640 constituents considering the rule of thumb  $\delta R \sim \frac{2m}{p_T}$  for decay products. As a result, angular resolution  
 641 effects won't have the same impact as for the  $W$  boson. This can be verified by the performance of  
 642 track-based variables in the ROCs found in Figure 25.

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

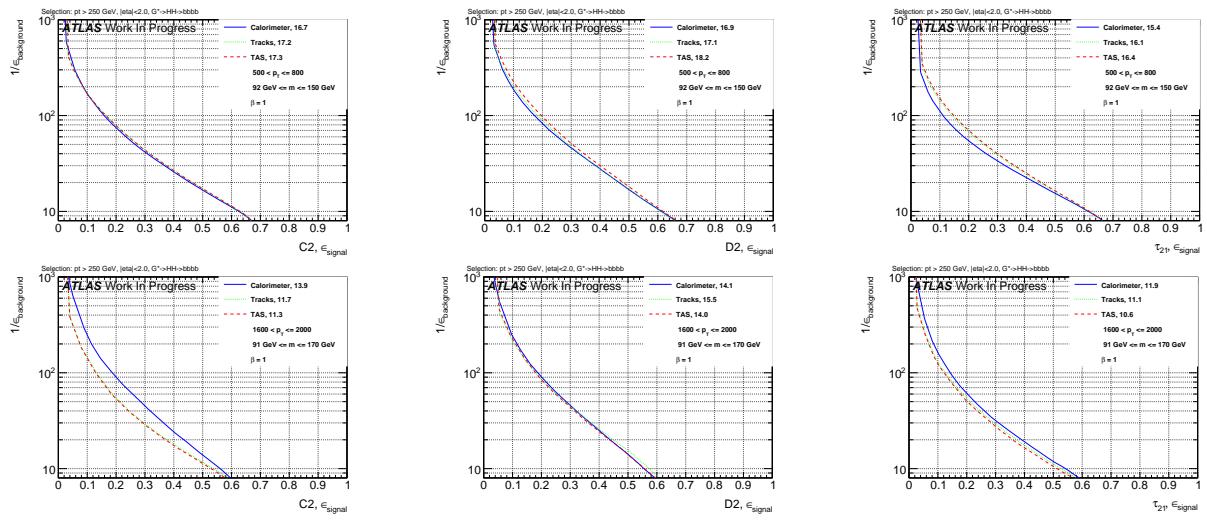


Figure 25: ROCs showing QCD rejection against Higgs boson efficiency for tracks, TAS and calorimeter. C2 (left), D2 (middle) and  $\tau_{21}$  (right) at  $\beta = 1$ . Shown is the energy range between 500-800 GeV (top) and 1600-2000 GeV (bottom). The numbers in the legend second  $p_T$  bin (left) and highest bin (right). The numbers in the legend indicate the achieved background rejection at 50% signal efficiency.

### 648 7.3.5 Performance for Top quark tagging

649 The top quark features a characteristic three body decay and a very high mass around 173 GeV. Studied  
650 here is the n-Subjettiness ratio  $\tau_{32}$  to distinguish the three prong like top quark jets and QCD background  
651 jets.

652 The ROCs in Figure 26 show the accompanying improvements in the separation power of  $\tau_{32}$  possible  
653 with TAS. Tagging tops quark events with  $\tau_{32}$  is found to greatly benefit from the excellent angular  
654 resolution of tracks. This is especially the case for high  $p_T$  where the limitation of the calorimeter cell  
655 size clearly diminishes the possible identification of three distinct substructures inside a large-R jet. The  
656 enhancements are not as articulated for the low  $p_T$  regions, nevertheless TAS  $\tau_{32}$  performs here at least  
657 equally well as calorimeter  $\tau_{32}$ . Furthermore, tracks are observed to perform slightly worse in comparison  
658 with TAS for the lower  $p_T$  regions, but match the TAS performance for very large boosts as expected.

### 659 7.4 Optimisation of $\beta$

660 The observed background rejection of variables calculated with TAS and tracks is at least as high as of  
661 calorimeter cluster based variables due to the high angular resolution of tracks. Therefore, studied are the  
662 effects of a higher weighting of the angular part of the substructure variables. For completeness, considered  
663 as well is a lower weighting. Previous studies of default calorimeter variables for W boson tagging, see e.g.  
664 Reference [bib:w\_tagging], found  $\beta = 1$  to maximize the separation power of calorimeter variables.

665 A scan over the values  $\beta = 0.5, 1, 1.7, 2, 3$  is performed in order to identify the best variables for the specific  
666 scenarios of tagging W boson, Higgs boson or top quark jets. The background rejections, achieved at the  
667 50 % working point after mass cut and tagging are summarized in tables. The corresponding ROCs can

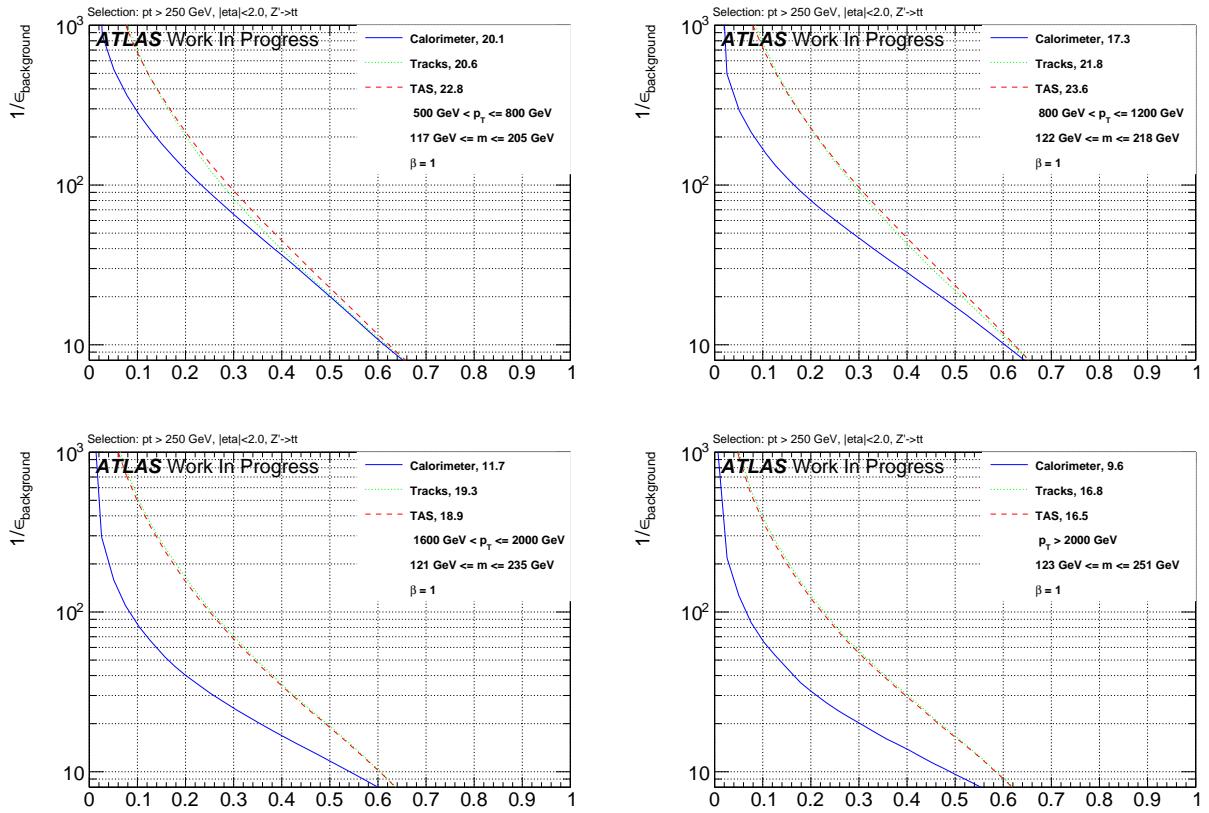


Figure 26: ROCs showing QCD rejection against top quark efficiency for tracks, TAS and calorimeter  $\tau_{32}$  at  $\beta = 1$ ,  $p_T$  ordering from upper left to lower right.

be found in Appendix J. Pseudo-experiments were used to propagate the uncertainties on the signal and background distributions due to the finite size of the MC samples to the background rejections.

#### 7.4.1 Optimisation for $W$ boson jets

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

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

C2 is found to be relatively insensitive to the track assisting, whereas separation with  $\tau_{21}$  and D2 (for lower energies) is visibly poorer with tracks compared to TAS. Starting around 1200 GeV, tracks perform comparably and for even higher energies equally well as TAS due to the rising amount of large-R jets with

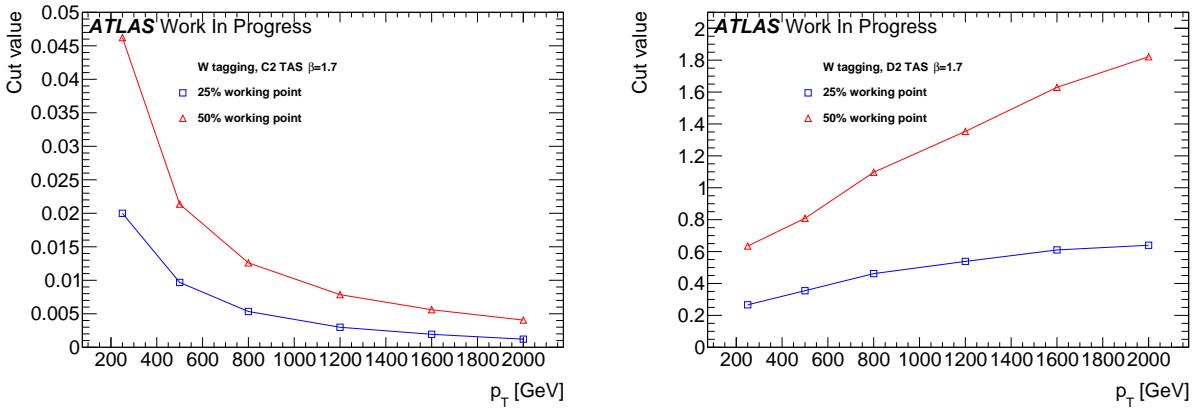


Figure 27: Cut values for  $C2_{TAS}^{(\beta=1.7)}$  (left) and  $D2_{TAS}^{(\beta=1.7)}$  (right) to achieve 50 % and 25 %  $W$  boson efficiency.

683 only one sub-jet. Comparing variables independently for the different inputs,  $\tau_{21}$  performs worse than  $C2$   
684 and  $D2$ . However, e.g.  $\tau_{21}$ , TAS can outperform  $C2_{calo}$  and  $D2_{calo}^{(1)}$ .

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

690 Shown in Figure 27 are the cut values for 50 % and 25 % signal efficiency for  $D2_{TAS}^{(\beta=1.7)}$  and  $C2_{TAS}^{(\beta=1.7)}$ .  
691  $D2_{TAS}^{(\beta=1.7)}$ . As for the default cluster variables, the  $D2_{TAS}^{(\beta=1.7)}$  cut is more  $p_T$  robust than the cut on  $C2_{TAS}^{(\beta=1.7)}$ .  
692

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

#### 702 7.4.2 Optimisation for Higgs boson jets

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

707 No improvement of  $\tau_{21}$  is observed with tracks or TAS, clusters perform equally well for lower  $p_T$  and  
708 slightly better at high energies. Again, the QCD rejection achieved with  $\tau_{21}$  is exceeded by  $C2$  and  $D2$ .

$p_T$ [GeV]	Calorimeter						$\tau_{21}$								
	C2			D2			C2			D2					
	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	29.7(1.5)	31.7(1.9)	31.4(1.6)	30.7(1.9)	28.5(1.4)	27.2(2.0)	35.0(2.0)	33.0(1.8)	25.7(1.2)	33.1(1.8)	27.6(1.3)	26.2(1.4)	25.1(1.2)	22.4(0.8)	
500 - 800	44.2(1.8)	50.1(2.0)	49.6(1.9)	48.6(1.8)	42.6(1.9)	40.3(2.2)	55.3(2.6)	56.3(2.4)	52.5(2.1)	39.3(1.3)	49.4(2.0)	41.1(1.4)	43.3(1.7)	41.3(1.6)	
800 - 1200	32.0(1.5)	37.5(1.7)	35.4(1.5)	33.4(1.5)	26.8(0.9)	34.0(2.1)	41.1(2.0)	38.5(1.6)	34.9(1.3)	25.4(0.7)	30.5(1.2)	30.9(1.2)	32.5(1.3)	28.1(0.9)	
1200 - 1600	30.1(1.3)	34.4(1.8)	29.4(1.3)	26.8(1.0)	20.7(0.8)	34.1(1.8)	38.1(1.9)	31.4(1.4)	27.6(1.2)	19.3(0.5)	23.1(0.9)	27.3(1.1)	31.1(1.2)	29.9(1.3)	
1600 - 2000	20.9(1.3)	22.4(1.5)	18.2(1.2)	16.5(0.9)	12.9(0.6)	26.4(1.7)	25.4(1.3)	19.3(1.1)	16.9(0.9)	11.9(0.5)	16.4(1.0)	19.1(1.1)	21.1(1.1)	19.9(1.0)	
> 2000	16.9(1.4)	18.7(1.4)	14.1(0.9)	12.6(0.8)	9.9(0.7)	23.3(1.9)	21.9(1.7)	15.7(1.1)	13.5(0.9)	9.2(0.4)	12.3(1.1)	15.5(1.1)	17.2(1.2)	15.7(1.1)	
TAS	C2						D2						$\tau_{21}$		
$p_T$ [GeV]	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	29.4(1.9)	30.1(1.9)	28.9(1.5)	28.5(1.3)	27.7(1.3)	28.6(2.0)	37.7(2.1)	35.4(2.3)	29.4(1.2)	36.2(2.2)	31.5(1.6)	26.8(1.3)	25.4(1.4)	24.0(1.0)	
500 - 800	48.2(2.0)	55.5(2.7)	58.6(2.6)	59.1(2.7)	56.8(2.0)	42.8(2.3)	67.2(3.1)	67.6(3.2)	63.7(3.0)	52.6(2.3)	55.7(2.6)	51.9(2.1)	45.5(2.0)	44.0(1.9)	
800 - 1200	31.0(1.2)	44.6(1.9)	54.6(2.8)	55.2(2.8)	53.0(3.2)	26.1(1.3)	47.6(2.3)	54.9(2.4)	52.6(2.8)	43.1(1.5)	36.4(1.8)	37.3(1.7)	36.2(1.8)	41.3(1.5)	
1200 - 1600	20.9(0.7)	39.1(1.9)	53.8(2.6)	55.1(3.0)	50.1(1.6)	22.7(1.4)	42.1(2.4)	50.8(1.8)	49.6(2.3)	41.1(1.2)	27.9(1.3)	31.4(1.5)	33.4(1.6)	35.5(1.6)	
1600 - 2000	16.7(0.7)	36.9(2.9)	50.9(4.3)	50.3(4.4)	42.2(2.4)	18.7(1.7)	32.7(3.3)	37.8(2.0)	36.1(2.4)	28.7(1.2)	20.5(1.2)	24.8(1.6)	26.1(2.0)	33.0(1.8)	
> 2000	11.6(0.6)	31.2(3.2)	46.1(4.7)	45.5(5.2)	35.5(3.8)	17.8(2.0)	33.0(4.0)	36.3(2.0)	34.0(2.5)	27.4(1.3)	16.4(1.3)	22.3(2.0)	24.2(2.2)	24.4(2.5)	
Tracks	C2						D2						$\tau_{21}$		
$p_T$ [GeV]	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3	$\beta = 0.5$	1	1.7	2	3
250 - 500	27.1(1.2)	28.1(1.5)	28.7(1.9)	28.2(1.7)	21.6(1.2)	28.9(2.0)	29.5(1.8)	29.1(1.6)	28.1(1.3)	28.7(1.8)	28.0(1.7)	25.6(1.3)	25.1(1.3)	24.2(0.9)	
500 - 800	46.5(1.9)	52.9(2.4)	57.7(2.6)	58.1(2.7)	55.8(2.5)	30.1(1.8)	46.8(2.4)	53.4(2.2)	52.1(2.3)	46.6(1.7)	46.1(2.3)	44.9(1.8)	41.7(2.1)	40.6(1.8)	
800 - 1200	30.3(1.1)	44.5(2.2)	54.8(2.8)	56.4(3.0)	53.7(3.6)	24.5(1.5)	42.3(2.3)	48.6(2.5)	47.5(1.2)	42.4(1.2)	34.5(1.6)	36.2(1.8)	36.0(1.8)	39.2(1.5)	
1200 - 1600	20.7(0.6)	39.0(1.9)	54.2(2.7)	55.5(3.3)	50.9(1.7)	22.7(1.3)	41.0(2.2)	50.0(1.6)	47.6(2.2)	41.4(1.2)	27.7(1.2)	31.3(1.4)	33.3(1.6)	35.7(1.5)	
1600 - 2000	16.6(0.7)	36.7(2.3)	51.7(5.2)	51.6(4.0)	43.1(2.3)	18.5(1.7)	32.1(3.0)	37.0(1.9)	35.9(2.3)	29.3(1.2)	20.5(1.3)	24.6(1.7)	26.2(1.8)	33.2(1.8)	
> 2000	11.6(0.5)	31.5(3.0)	46.8(5.7)	46.0(4.2)	36.1(4.3)	17.8(2.2)	33.0(3.3)	35.9(2.1)	34.2(2.6)	28.1(1.0)	16.4(1.4)	22.5(1.8)	24.5(2.4)	21.8(2.6)	

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

<b>50% <math>\epsilon_{signal}</math></b>		<b>W tagging</b>		
$p_T$ [GeV]		D2 <sub>calo</sub> <sup>(<math>\beta=1</math>)</sup>	D2 <sub>TAS</sub> <sup>(<math>\beta=1.7</math>)</sup>	C2 <sub>TAS</sub> <sup>(<math>\beta=1.7</math>)</sup>
250 - 500		$35.0 \pm 2.0$	$35.4 \pm 2.3 (+1 \pm 9\%)$	$28.9 \pm 1.5 (-17 \pm 6\%)$
500 - 800		$55.3 \pm 2.6$	$67.6 \pm 3.2 (+22 \pm 8\%)$	$58.6 \pm 2.6 (+6 \pm 7\%)$
800 - 1200		$41.1 \pm 2.0$	$54.9 \pm 2.4 (+34 \pm 9\%)$	$54.6 \pm 2.8 (+33 \pm 9\%)$
1200 - 1600		$38.1 \pm 1.9$	$50.8 \pm 1.8 (+33 \pm 8\%)$	$53.8 \pm 2.7 (+41 \pm 10\%)$
1600 - 2000		$25.4 \pm 1.3$	$37.8 \pm 2.0 (+49 \pm 11\%)$	$50.9 \pm 4.3 (+100 \pm 20\%)$
> 2000		$21.9 \pm 1.7$	$36.3 \pm 2.0 (+66 \pm 16\%)$	$46.1 \pm 4.7 (+111 \pm 27\%)$

<b>25% <math>\epsilon_{signal}</math></b>		<b>W tagging</b>		
$p_T$ [GeV]		D2 <sub>calo</sub> <sup>(<math>\beta=1</math>)</sup>	D2 <sub>TAS</sub> <sup>(<math>\beta=1.7</math>)</sup>	C2 <sub>TAS</sub> <sup>(<math>\beta=1.7</math>)</sup>
250 - 500		$139.6 \pm 9.8$	$146.0 \pm 12.4 (+5 \pm 12\%)$	$108.2 \pm 7.5 (-22 \pm 8\%)$
500 - 800		$243.7 \pm 13.2$	$360.1 \pm 21.1 (+48 \pm 12\%)$	$298.4 \pm 15.9 (+22 \pm 9\%)$
800 - 1200		$181.0 \pm 8.8$	$308.5 \pm 19.3 (+70 \pm 14\%)$	$313.2 \pm 24.4 (+78 \pm 16\%)$
1200 - 1600		$156.9 \pm 8.3$	$295.4 \pm 17.8 (+88 \pm 15\%)$	$354.6 \pm 25.6 (+126 \pm 20\%)$
1600 - 2000		$84.6 \pm 5.7$	$219.6 \pm 10.9 (+160 \pm 22\%)$	$320.5 \pm 31.4 (+279 \pm 45\%)$
> 2000		$78.9 \pm 7.6$	$233.5 \pm 14.7 (+196 \pm 34\%)$	$288.4 \pm 33.3 (+266 \pm 55\%)$

Table 4: Listing of the background rejections after the jet mass cut and tagging at 50% and 25%  $W$  boson efficiency for the identified best variables D2<sub>TAS</sub><sup>( $\beta=1.7$ )</sup> & C2<sub>TAS</sub><sup>( $\beta=1.7$ )</sup> together with the improvements over the standard choice D2<sub>calo</sub><sup>( $\beta=1$ )</sup>

The discrimination with clusters profits from a slightly higher angular weighting, although the gain is not as significant as for tracks and TAS. This consistently shows the lower sensitivity to a variation of the angular weight. The small gain is connected to the higher separation of the Higgs decay products compared to the  $W$  boson case.

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

The differences between  $\beta = 1.7$  and  $\beta = 2$  are inconclusive with minor advantages at high and slight inferiorities at low  $p_T$  for  $\beta = 2$ . Tracks perform slightly worse than TAS for lower energies but similarly better in the two highest studied  $p_T$  regions. Chosen for further examination are D2<sub>TAS</sub><sup>( $\beta=1.7$ )</sup> and D2<sub>track</sub><sup>( $\beta=1.7$ )</sup>.

Shown in Figure 28 are the cut values for 50 % and 25 % signal efficiency for D2<sub>TAS</sub><sup>( $\beta=1.7$ )</sup> and D2<sub>track</sub><sup>( $\beta=1.7$ )</sup>. The cut value shows a slight upward trend for rising  $p_T$ . Moreover, cut values for the first bin are higher as for the second, in contrast to the overall upward trend of D2. This is the result of the low boost in the lowest  $p_T$  region resulting in a left shoulder of the mass distributions representing large-R jets containing only part of the Higgs boson decay. These jets feature one-prong structure and result in background-like D2 values. The TAS D2 cut is marginally higher than the corresponding track D2 cut since the assisted tracks have a higher  $p_T$  and the D2 cut features a rising tendency with  $p_T$ .

Listed in Table 6 are the background rejections for D2<sub>TAS</sub><sup>( $\beta=1.7$ )</sup>, D2<sub>track</sub><sup>( $\beta=1.7$ )</sup>, and for the best calorimeter variable, which is D2<sub>calo</sub><sup>( $\beta=1$ )</sup>, with the corresponding improvements due to the use of TAS respectively

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

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

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

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

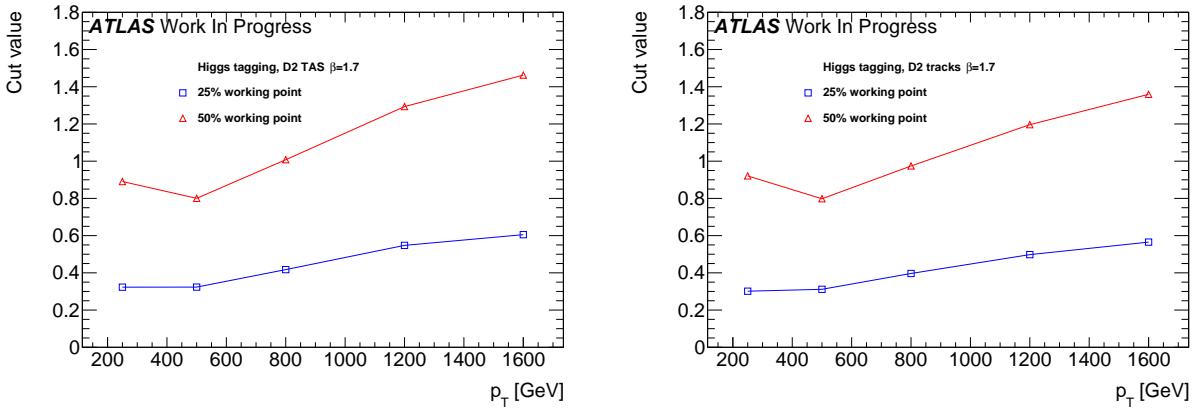


Figure 28: Cut values for  $D2_{\text{TAS}}^{(\beta=1.7)}$  (left) and  $D2_{\text{track}}^{(\beta=1.7)}$  (right) to achieve 50% and 25% Higgs boson efficiency.

50% $\epsilon_{\text{signal}}$		Higgs tagging		
$p_T$ [GeV]	$D2_{\text{calo}}^{(\beta=1)}$	$D2_{\text{TAS}}^{(\beta=1.7)}$	$D2_{\text{track}}^{(\beta=1.7)}$	
250 - 500	$8.4 \pm 0.2$	$8.5 \pm 0.2 (+1 \pm 4\%)$	$8.3 \pm 0.2 (-1 \pm 3\%)$	
500 - 800	$17.7 \pm 0.4$	$18.7 \pm 0.4 (+6 \pm 3\%)$	$17.9 \pm 0.4 (+1 \pm 3\%)$	
800 - 1200	$26.3 \pm 0.6$	$28.5 \pm 0.7 (+8 \pm 4\%)$	$28.0 \pm 0.7 (+6 \pm 4\%)$	
1200 - 1600	$27.0 \pm 0.8$	$28.9 \pm 0.7 (+7 \pm 4\%)$	$30.0 \pm 0.8 (+11 \pm 4\%)$	
1600 - 2000	$14.9 \pm 0.8$	$17.7 \pm 0.8 (+19 \pm 8\%)$	$18.5 \pm 0.8 (+24 \pm 9\%)$	

25% $\epsilon_{\text{signal}}$		Higgs tagging		
$p_T$ [GeV]	$D2_{\text{calo}}^{(\beta=1)}$	$D2_{\text{TAS}}^{(\beta=1.7)}$	$D2_{\text{track}}^{(\beta=1.7)}$	
250 - 500	$25.1 \pm 0.6$	$28.9 \pm 0.7 (+15 \pm 4\%)$	$30.5 \pm 0.8 (+22 \pm 4\%)$	
500 - 800	$54.1 \pm 1.4$	$69.6 \pm 1.9 (+29 \pm 5\%)$	$64.9 \pm 1.8 (+20 \pm 5\%)$	
800 - 1200	$90.8 \pm 2.5$	$121.3 \pm 3.4 (+34 \pm 5\%)$	$117.9 \pm 3.2 (+30 \pm 5\%)$	
1200 - 1600	$97.6 \pm 3.1$	$117.7 \pm 3.8 (+21 \pm 5\%)$	$122.4 \pm 4.2 (+25 \pm 6\%)$	
1600 - 2000	$54.6 \pm 3.5$	$74.0 \pm 5.7 (+36 \pm 14\%)$	$75.0 \pm 5.1 (+37 \pm 13\%)$	

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

730 tracks instead of clusters. At very high energies, the angle between the  $b\bar{b}$  pair is small despite the high  
 731 Higgs boson mass and the effect of the calorimeter cell size becomes significant. The improvements for  
 732 D2 calculated with TAS instead of clusters are single-digit percentages for low  $p_T$  and up to 20 % for the  
 733 highest studied  $p_T$  bin at 50 % Higgs boson efficiency. For the lower working point, they reach around  
 734 30 % of the QCD rejection achieved with cluster based D2.

#### 735 7.4.3 Optimisation for Top quark jets

736 The results of the optimisation for Top quark jets are shown in Table 7. Studied was  $\tau_{32}$  with values of  
 737  $\beta \geq 1$ , since the W boson and Higgs boson parts affirmed the expected lower performance of track and  
 738 TAS based variables with an angular weighting of  $\beta \leq 1$ . The calorimeter  $\tau_{32}$  variable profits from a

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

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

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

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

higher angular weighting up to around  $\beta = 2$ , but degrades in performance for  $\beta = 3$ . Since the involved three prong structure of the top quark decay requires a good angular separation of the jet constituents to be resolved, tracks and TAS perform superior to clusters. A higher angular weighting does not improve the separation power of track and TAS variables,  $\beta = 2$  already diminishes the performance. The best discrimination is achieved with TAS and  $\beta = 1, 1.7$ . The marginal differences between both values of  $\beta$  depend on the considered  $p_T$  region. Track  $\tau_{32, \text{track}}$  achieves lower separation as  $\tau_{32, \text{TAS}}$ , except for regions with very high boosts, but as well outperforms the cluster variable.

Shown in Figure 29 are the cut values for 50 % and 25 % signal efficiency for  $\tau_{32, \text{TAS}}^{(\beta=1.7)}$  and  $\tau_{32, \text{track}}^{(\beta=1.7)}$ . The crack between the first and second  $p_T$  bin is more evident since the top quark with its much higher mass is here very unlikely to be reconstructed into a single large-R jet, resulting in background like signal events. Furthermore,  $\tau_{32} (\beta = 1.7)$  needs to be cut at lower values as  $\tau_{32} (\beta = 1)$  to achieve a certain signal efficiency. This is the result of the higher angular weighting that shifts the overall distributions to lower values, because the angular distance between two constituents inside a (highly) boosted large-R jet is in the majority of cases lower than one. Thus, the angular part of  $\tau_{32}$  decreases with  $\beta > 1$ . The TAS  $\tau_{32}$  cut value is observed to be robust against variations of  $p_T$ , in accordance to the results of the  $p_T$  correlation

<b>50% <math>\epsilon_{signal}</math></b>		<b>Top Tagging</b>		
$p_T$ [GeV]	D2 <sub>calo</sub> <sup>(<math>\beta=2</math>)</sup>	$\tau_{32, TAS}^{(\beta=1)}$	$\tau_{32, TAS}^{(\beta=1.7)}$	
250 - 500	$9.5 \pm 0.2$	$10.7 \pm 0.2 (+13 \pm 3 \%)$	$10.1 \pm 0.2 (+6 \pm 3 \%)$	
500 - 800	$22.4 \pm 0.6$	$22.8 \pm 0.6 (+2 \pm 4 \%)$	$22.8 \pm 0.6 (+2 \pm 4 \%)$	
800 - 1200	$20.6 \pm 0.5$	$23.6 \pm 0.6 (+15 \pm 4 \%)$	$24.1 \pm 0.6 (+17 \pm 4 \%)$	
1200 - 1600	$16.6 \pm 0.4$	$22.0 \pm 0.6 (+33 \pm 5 \%)$	$22.3 \pm 0.6 (+34 \pm 5 \%)$	
1600 - 2000	$13.3 \pm 0.4$	$18.9 \pm 0.6 (+42 \pm 6 \%)$	$18.8 \pm 0.6 (+41 \pm 6 \%)$	
> 2000	$10.9 \pm 0.4$	$16.5 \pm 0.7 (+51 \pm 8 \%)$	$15.7 \pm 0.7 (+44 \pm 8 \%)$	

<b>25% <math>\epsilon_{signal}</math></b>		<b>Top Tagging</b>		
$p_T$ [GeV]	D2 <sub>calo</sub> <sup>(<math>\beta=2</math>)</sup>	$\tau_{32, TAS}^{(\beta=1)}$	$\tau_{32, TAS}^{(\beta=1.7)}$	
250 - 500	$33.7 \pm 1.0$	$37.6 \pm 1.4 (+12 \pm 5 \%)$	$36.7 \pm 1.2 (+9 \pm 5 \%)$	
500 - 800	$114.7 \pm 3.3$	$138.0 \pm 4.3 (+20 \pm 5 \%)$	$139.1 \pm 4.2 (+21 \pm 5 \%)$	
800 - 1200	$97.0 \pm 2.7$	$144.6 \pm 4.9 (+49 \pm 7 \%)$	$149.6 \pm 5.2 (+54 \pm 7 \%)$	
1200 - 1600	$68.6 \pm 2.1$	$133.2 \pm 4.6 (+94 \pm 9 \%)$	$134.7 \pm 5.1 (+96 \pm 10 \%)$	
1600 - 2000	$47.5 \pm 1.6$	$100.3 \pm 4.2 (+111 \pm 11 \%)$	$99.9 \pm 4.4 (+110 \pm 12 \%)$	
> 2000	$36.3 \pm 1.6$	$80.2 \pm 5.0 (+121 \pm 17 \%)$	$75.5 \pm 4.9 (+108 \pm 16 \%)$	

Table 8: Listing of the background rejections after the jet mass cut and tagging at 50% and 25% top signal efficiency for the identified best variables  $\tau_{32, TAS}^{(\beta=1,1.7)}$  together with the improvements over the best variable with clusters which is D2<sub>calo</sub><sup>( $\beta=2$ )</sup>.

plots, see 24.

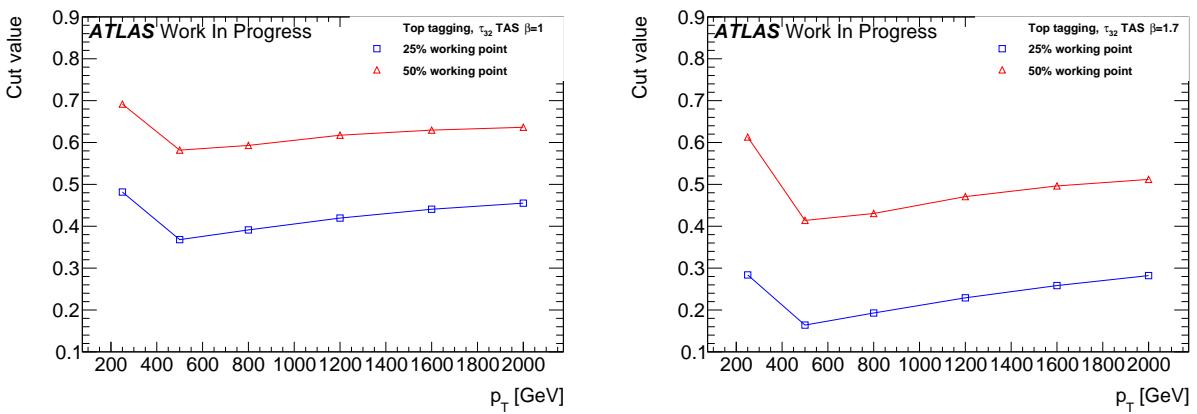


Figure 29: Cut values for  $\tau_{32, TAS}^{(\beta=1)}$  (left) and  $\tau_{32, TAS}^{(\beta=1.7)}$  (right) to achieve 50% and 25% Top quark efficiency

754

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

## 760 8 Uncertainties on observables with sub-jet-assisted tracks

761 This chapter gives a brief overview of the uncertainties on the track-assisted (sub-jet) mass variable.  
 762 For  $m^{TA}$  the uncertainties are smaller than calorimeter-based jet mass variables because of the way it  
 763 is constructed,  $m^{TA} = m^{trk} \times p_T^{calo}/p_T^{trk}$ : the ratio  $m^{trk}/p_T^{trk}$  causes a cancellation of the tracking  
 764 uncertainties to a large extent, which are smaller than  $m^{calo}$ . The remaining term  $p_T^{calo}$  is the additional  
 765 one where uncertainties on this variable need to be evaluated with special care.

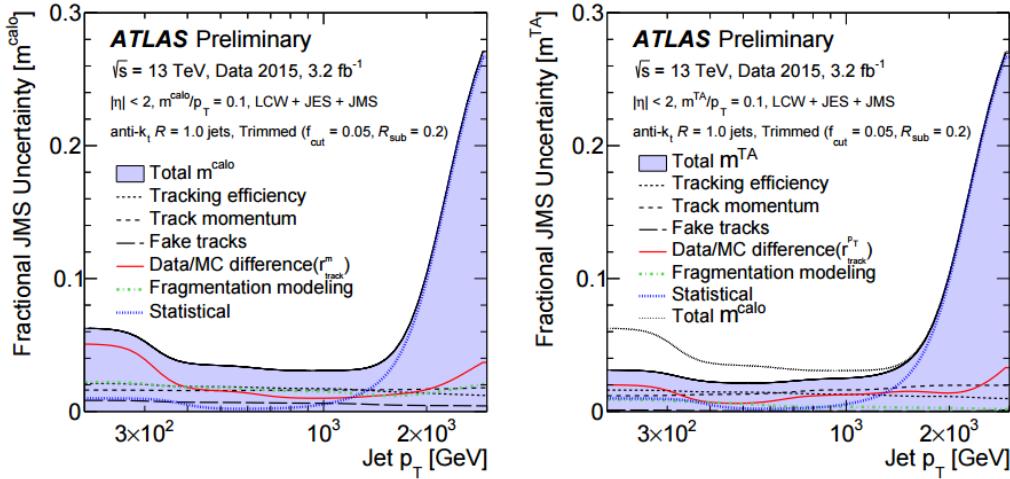


Figure 30: Comparison of the uncertainties for  $m^{calo}$ , on the left, and  $m^{TA}$ , on the right the rise on the high jet  $p_T$  is due to statistics. From the [art35].

766 For what concerns the  $m^{TAS}$ , the tracking uncertainties are expected to be identical to the  $m^{TA}$ , because  
 767 of the identical use of tracks in both variables, as also discussed in the BOOST Conference Note [art35].  
 768 The only significant difference in this regard of  $m^{TA}$  with respect to  $m^{TAS}$  is the  $p_T^{jet}$  instead of the  $p_T^{subjet}$ :  
 769 the uncertainties in the first one are calculated in-situ using  $p_T$  balance methods, and they are generally  
 770 well-behaved; for the second one, the uncertainties are also expected to lay in the same order of magnitude.  
 771 In Figure 30 the comparison of the uncertainties for the  $m^{calo}$  on the left and  $m^{TA}$  on the right shows the  
 772 smaller fractional JMS uncertainties in the use of the track-assisted method.  
 773 The path to bring  $m^{TAS}$  ready to usage needs of course to evaluate the uncertainties on the  $p_T^{subjet}$  which  
 774 can be provided with the R-Scan procedure which includes the anti- $k_t$  sub-jets of radius of 0.2; those were  
 775 already showed to have a similar performance of the  $k_t$  sub-jets used as standard.  
 776 The  $m^{TAS}$  moreover is expected to have little to none benefit from the calibration procedure which could  
 777 be provided from the R-Scan as well, as already shown and discussed in the previous chapter.

## 778 9 Conclusions & Outlook

779 The  $m^{TAS}$  variable was developed for the large- $R$  jet mass; it combines the information of the tracker- and  
 780 calorimeter-system to achieve an higher precision in the jet mass reconstruction, correcting the missed  
 781 neutral fraction which is absent in the tracker but not in the calorimeter. With respect to the  $m^{TA}$ , it  
 782 applies this correction at sub-jet by sub-jet level and not at jet by jet level, therefore providing a more  
 783 accurate reconstruction. It was shown in Monte Carlo simulation to be a very good observable confronting  
 784 quantitatively with the other definitions which are either standard or in preparation,  $m^{calo}$ ,  $m^{TA}$  and  $m^{comb}$ .

785 The optimal configuration of  $m^{TAS}$  is shown and confronted with different approaches, in particular in  
 786 terms of different trimming procedure of the large- $R$  jet to be used as an input. All the components of  
 787 the observable have been studied with the use of truth Monte Carlo information without detector effect, in  
 788 order to evaluate quantitatively its limits and strengths; the track  $p_T$  measure degradation was found to be  
 789 the cause of the variable decreasing performance at higher transverse momenta.

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

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

800 \*\*\*here sascha conclusions\*\*\*

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 9: 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.

## 801 Appendix

### 802 A Monte Carlo Samples

803 The samples used are divided into two main groups: SM background and beyond SM signal. The  
 804 SM background includes the QCD multijet samples, produced with a falling  $p_T$  spectrum. The beyond  
 805 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  
 806 ( $t \rightarrow W(\rightarrow q\bar{q}')b$ )) and RS-Graviton  $\rightarrow hh \rightarrow b\bar{b}b\bar{b}$ , i.e. final states have only jets in all the samples.  
 807 The details of the samples are given in Table 9; the masses considered span from 0.5 to 5 TeV to improve  
 808 and diversify the kinematic space covered.

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

## 816 B Trimming

817 The trimming algorithm is the most important in ATLAS and the one mainly used in this note. It takes  
 818 advantage of the fact that contamination from soft radiation has a much lower  $p_T$  with respect to the  
 819 hard-scattering component. Therefore uses a transverse momentum ratio to distinguish among those. The  
 820 algorithm works on a two-dimensional parameter space:  $R_{sub}$  and  $f_{cut}$ . The steps are as follows:

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

- 826 • the sub-jets which survived this procedure are the only one which compose the trimmed jet.

827 The trimming procedure is also explained in Figure 31, an example of performance in simulation with  
 828 standard parameters is shown in Appendix (Figure 33).

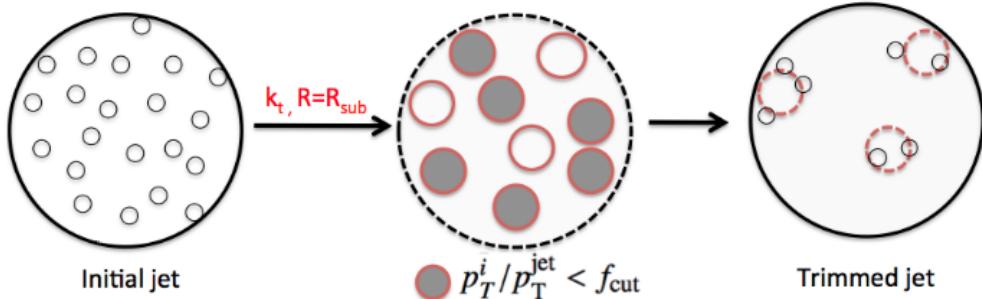


Figure 31: Schematic of the trimming algorithm.

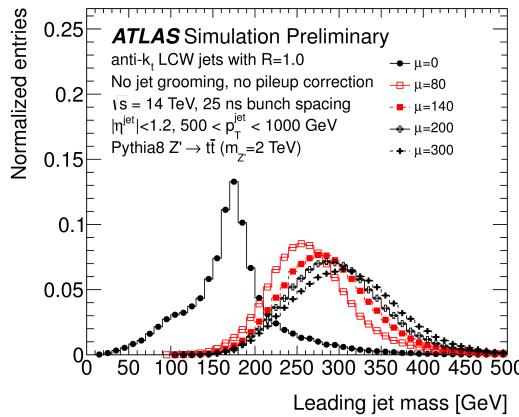


Figure 32: Effect of pile-up contamination in large- $R$  jets: here shown different PU conditions parametrized by  $\langle \mu \rangle$ . From [highlumi].

## 829 C Tracks details

830 The requirements applied on the track used in the work presented in this note are given here:

- 831 •  $p_T^{track} > 400$  MeV;
- 832 •  $|\eta| < 2.5$ ;
- 833 • Maximum 7 hits in the Pixel and STC sub-detectors;
- 834 • Maximum 1 Pixel hole;
- 835 • Maximum 2 silicon holes;
- 836 • Less than 3 shared modules;

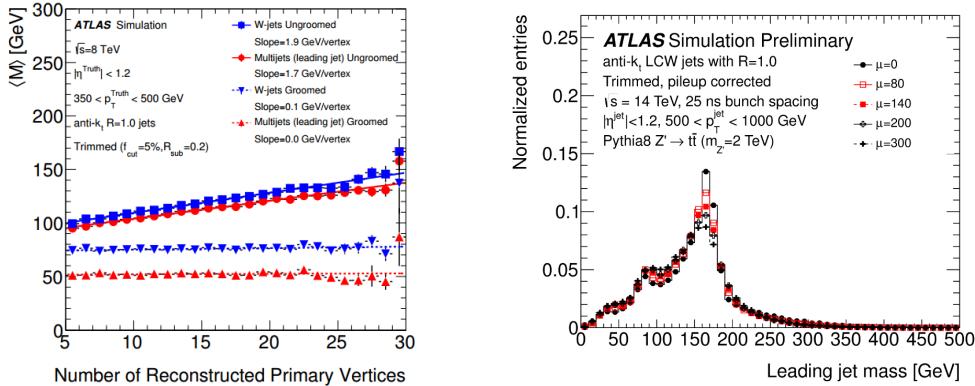


Figure 33: Left: mass reconstructed as a function of the number of primary vertices (parameterizing PU) for different samples; after trimming procedure the mass is pretty much independent of PU for all the samples. Right: mass distributions for different PU conditions: after trimming the reconstruction is not degraded as much as Figure 32.

- Maximum 2 mm of displacement along beam axis ( $z_0$ ) from the primary vertex;
- Maximum 2.5 mm of distance in x-y plane from the primary vertex and point of closest approach ( $d_0$ ).

## 840 D Alternative Performance Figure of Merit (FoM)

841 A concrete, quantitative feature has to be defined in order to understand which observable is “better”, in  
 842 the sense that we would prefer one or the other according to this criterion. This is often referred to as  
 843 *Figure of Merit* or simply FoM.

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

### 849 D.1 Gaussian Fit

850 The important feature to keep in mind, in fact, is the underlying physics which brings us to calculate the  
 851 mass of a jet. In figure 34 this is made clear: if the width of the invariant mass distribution of the jet is  
 852 smaller (highlighted), it allows a bigger background rejection, here shown as the QCD dijet, for the same  
 853 signal efficiency, by means of a simple mass requirement.

854 The width  $\sigma$  of the distribution, which can be obtained from a fit to the Gaussian core, is already a valid  
 855 FoM, which has an underlying physical feature. Moreover, in order to be independent from the mean of  
 856 the distribution, the width can be divided by the mean itself. This was in fact the FoM which was used  
 857 at the beginning of the work for this thesis, since it provided a simple and fast solution. However, special  
 858 care must be used both in the procedure of fitting Gaussian cores of responses, since they are asymmetric,  
 859 and to how the tails are treated.

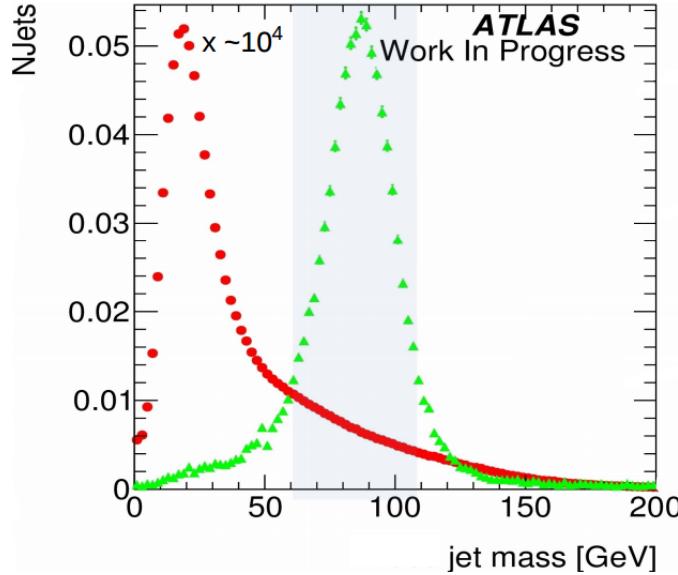


Figure 34: Mass distributions: in red the QCD dijet background rescaled, in green the  $W/Z$  from the  $W'$  sample. Highlighted the width of the 68% of the  $W/Z$  distribution.

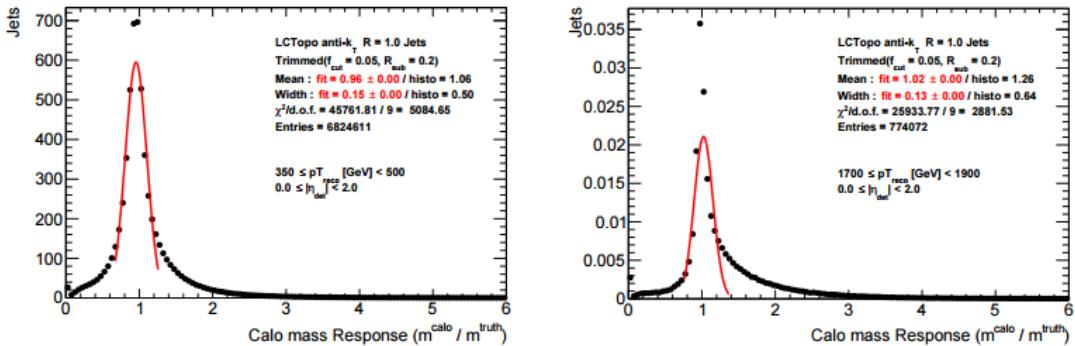


Figure 35: Mass Response distributions for the QCD multijet for various  $p_T$  ranges: on the right the failure of the Gaussian fit shows the limitation of this approach to serve as the Figure of Merit. On the plot the fit parameters and transverse momentum ranges.

The situation is depicted e.g. in Figure 35, 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 consider 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 is presented in the body of this note.

Jet Mass Observable Distribution Kinematic distribution for all the samples,  $p_T \eta$  and  $\phi$  is shown.

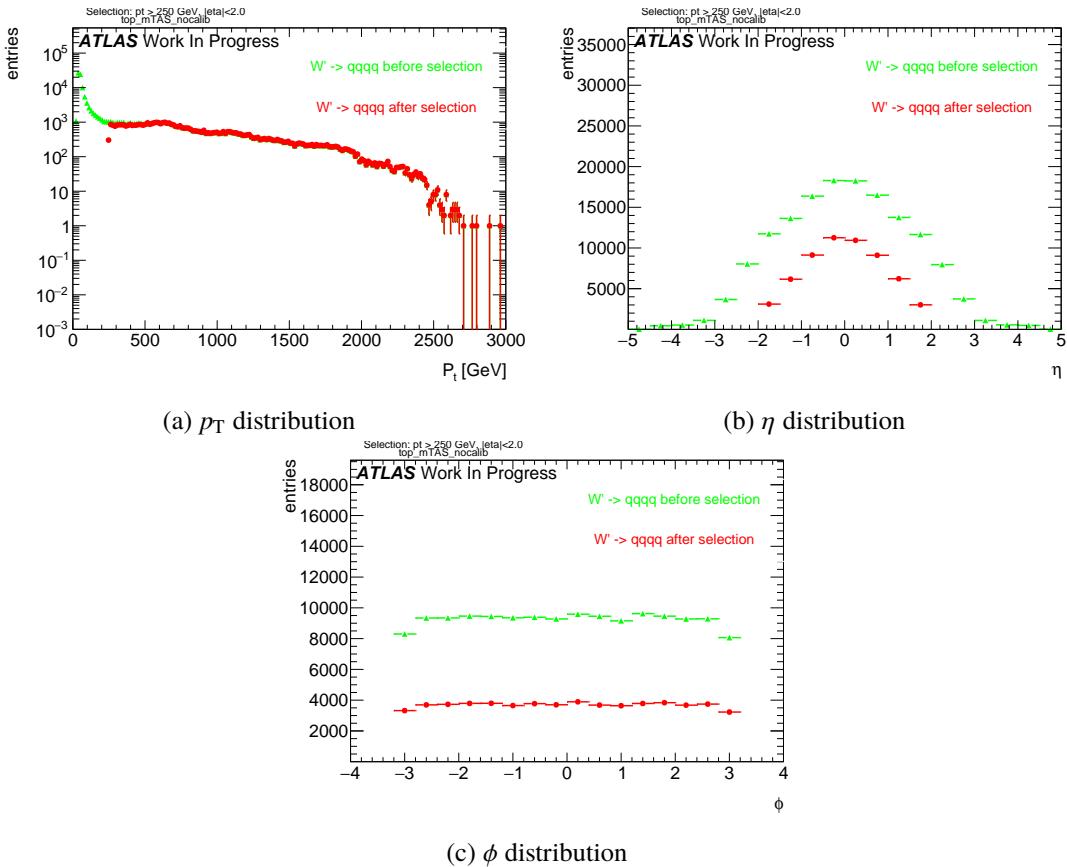


Figure 36: Boosted tops kinematic distribution.

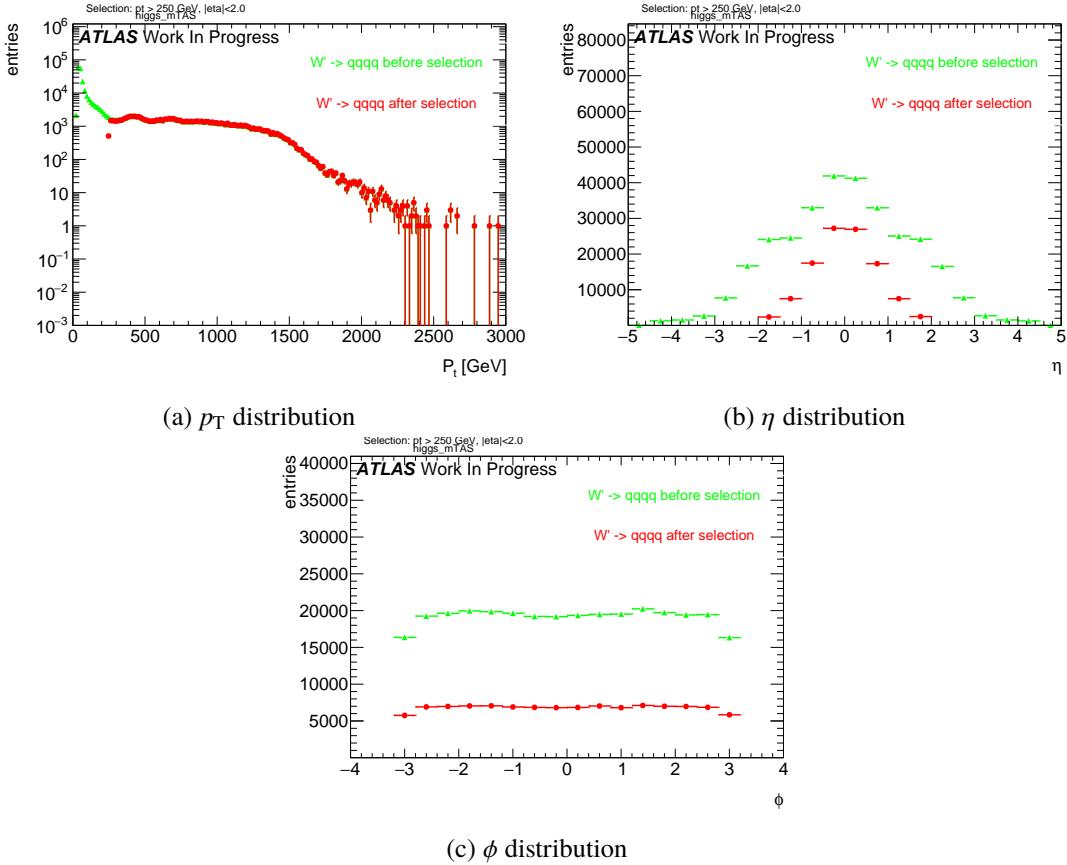


Figure 37: RS-Graviton kinematic distribution.

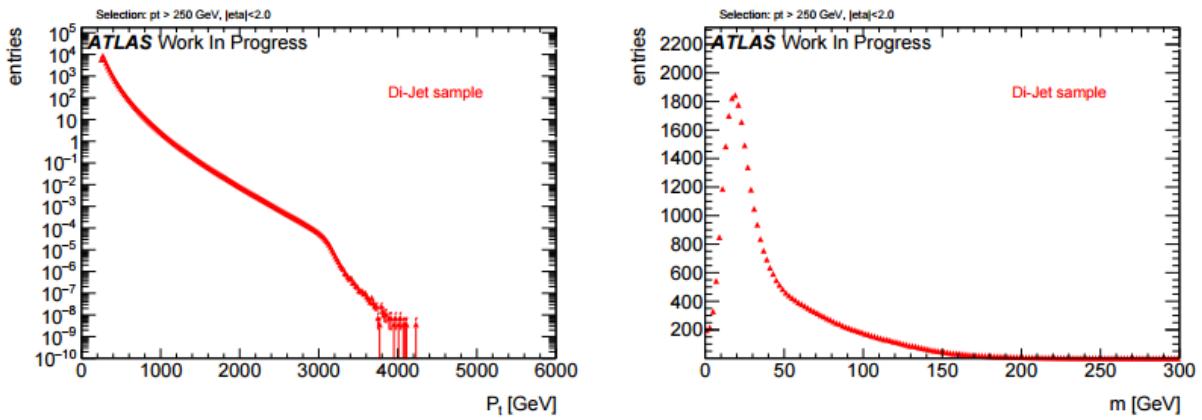


Figure 38: QCD dijet transverse momentum and mass distributions.

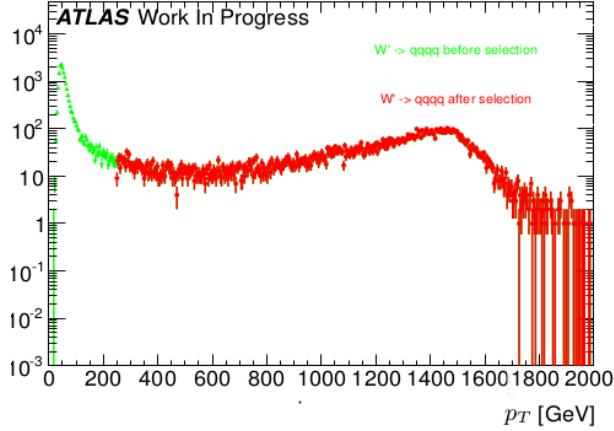


Figure 39: The  $p_T$  distribution of a 3 TeV resonance from the hadronically decaying  $W$  or  $Z$ , in logarithmic plot. As can be seen, the jacobian peak is around  $p_T \simeq m_{W'}/2 \simeq 1.5$  TeV.

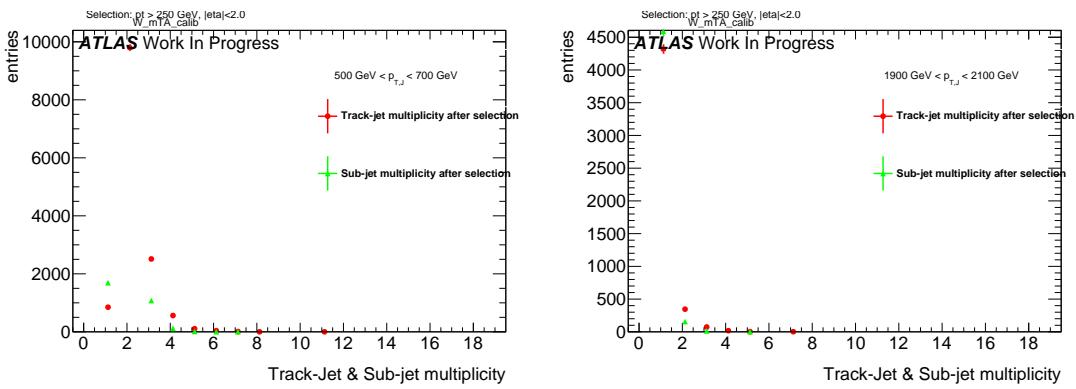


Figure 40: Sub-jet and Track-jet (jets created having tracks as input) multiplicity, for selected bins of transverse momentum.

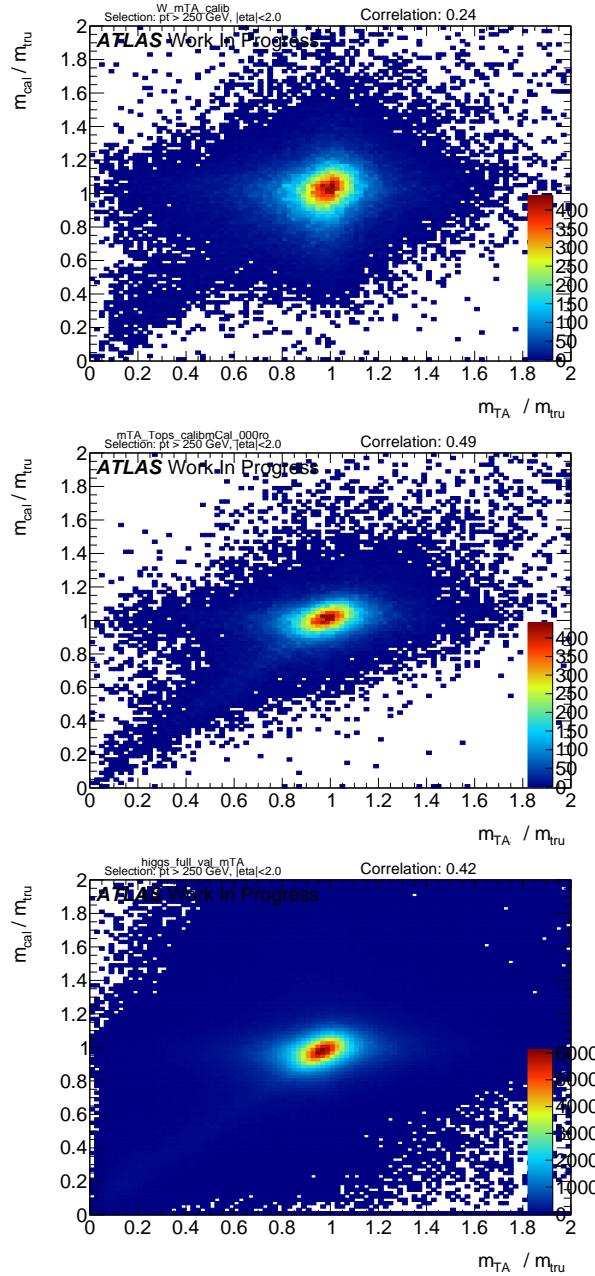


Figure 41: Calorimeter based jet mass response vs the track-assisted mass response for the three signal samples. Correlation coefficient is indicated on the top right.

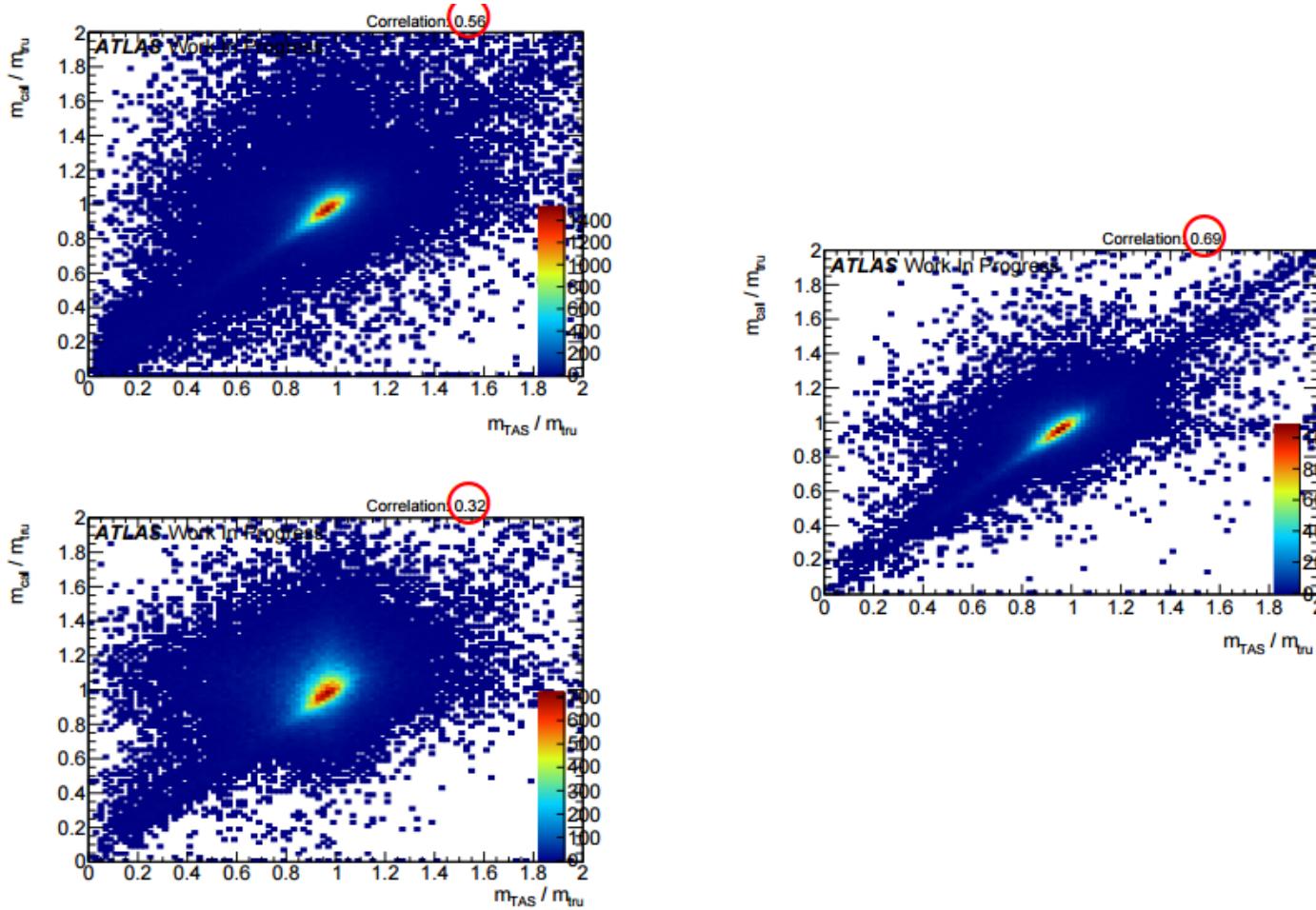


Figure 42: Calorimeter based jet mass response vs the track-assisted sub-jet mass response for the three signal samples. Correlation coefficient is indicated on the top right and highlighted. On the left, top, the higgs sample, bottom, the  $W/Z$ ; on the right the top-quark sample.

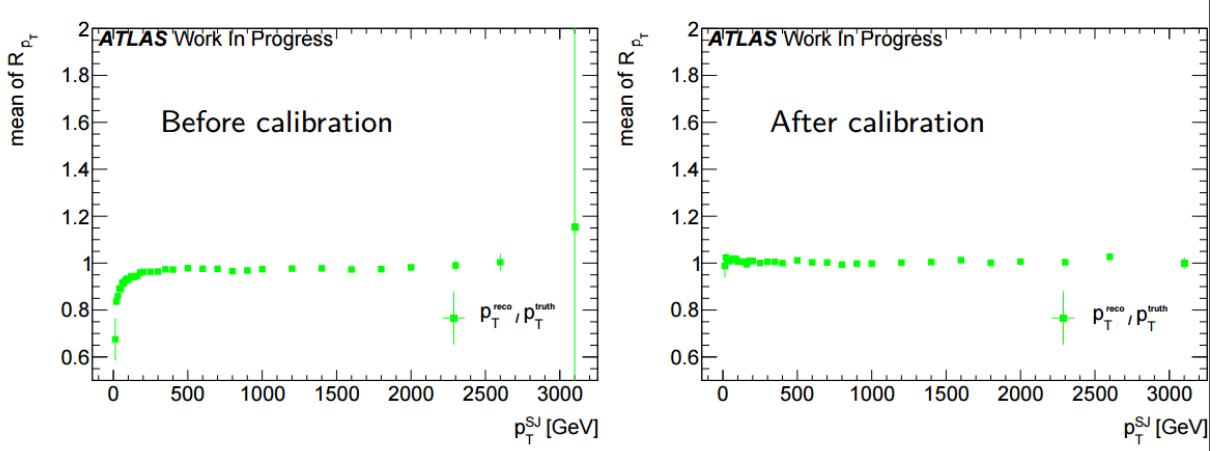


Figure 43: Poor's man calibration effect on mean of transverse momentum's response of the sub-jet, before, left, and after, right, the procedure.

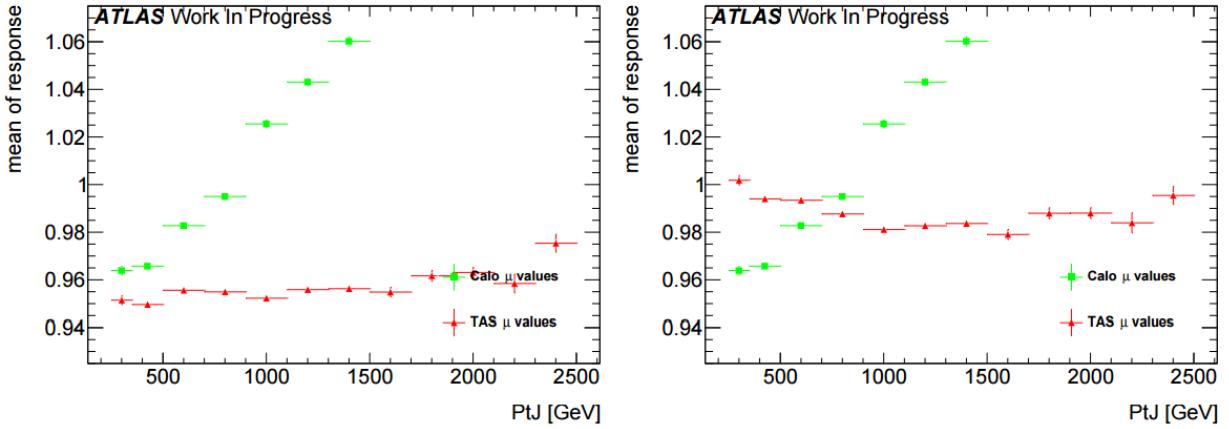


Figure 44: Poor's man calibration effect on the mean of the mass response of the large-R jet, before, left, and after, right, the procedure.

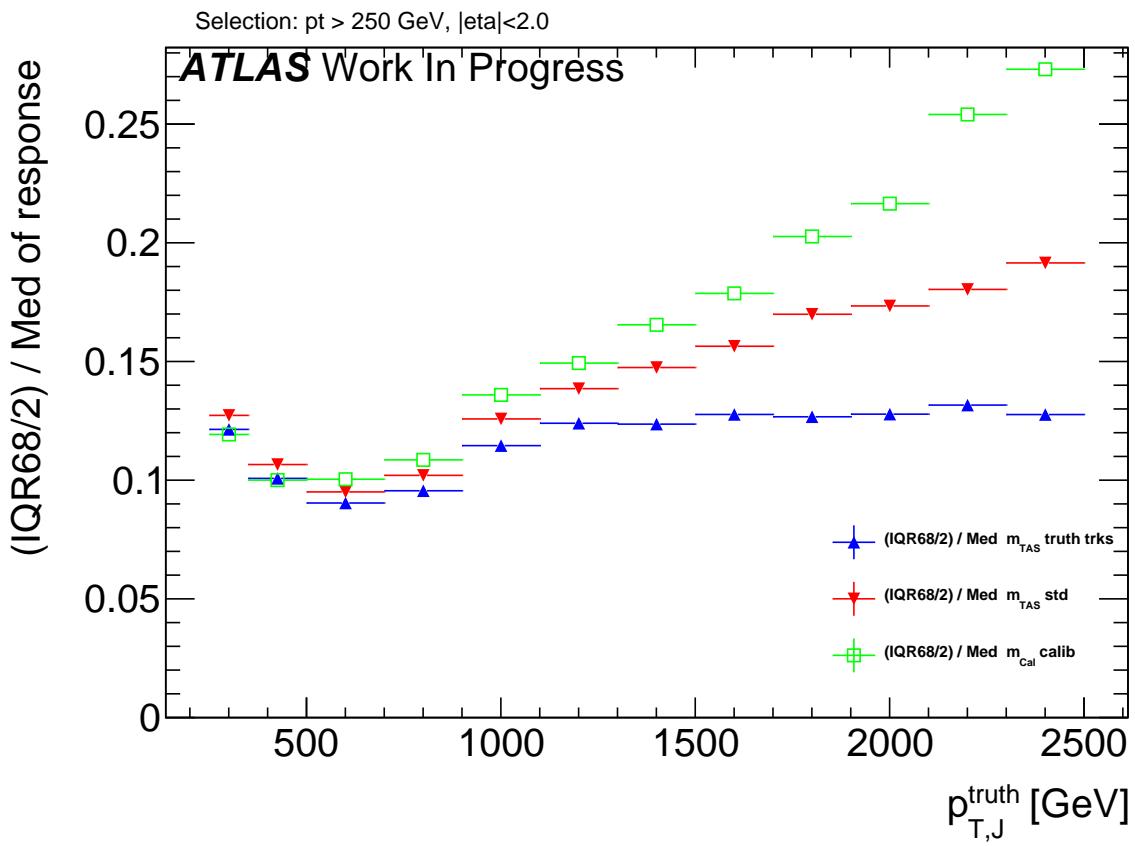
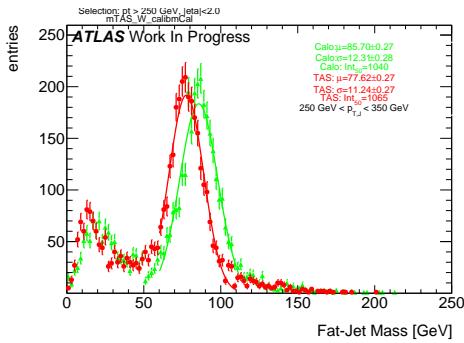
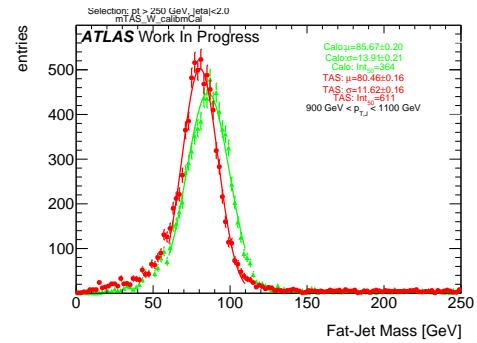
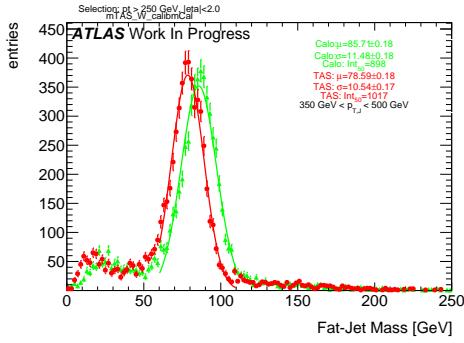
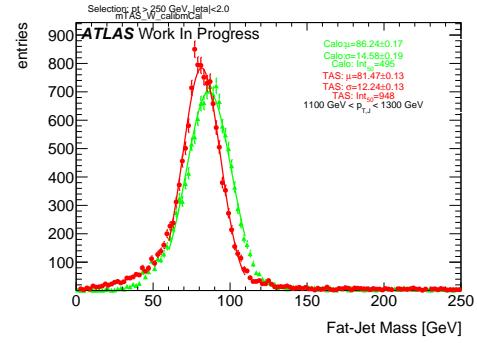
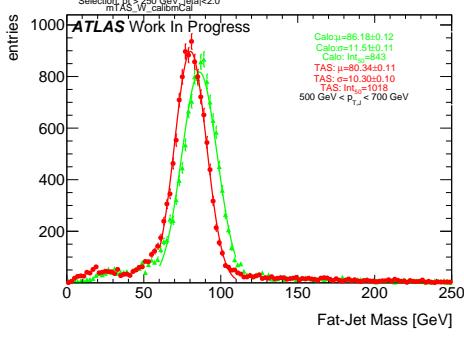
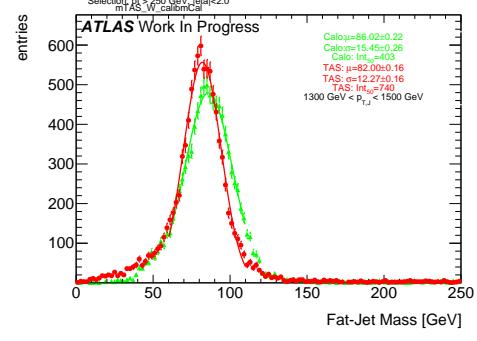
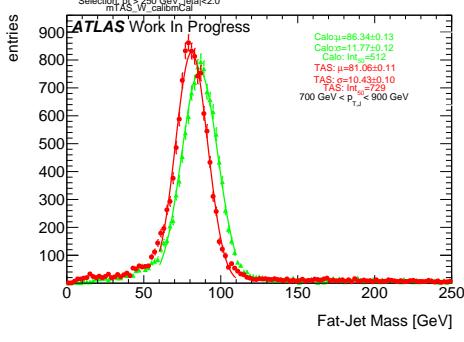
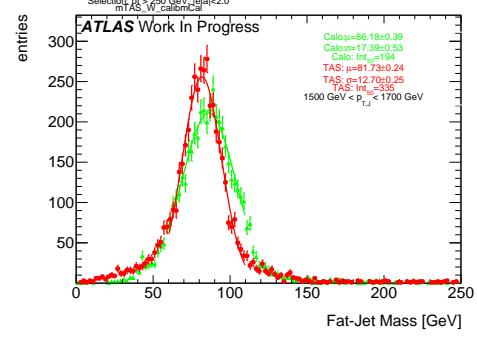


Figure 45: Comparison of the  $m^{TAS}$  and the same variable using truth-level information for the tracks.

<sub>867</sub> E  $m^{TAS}$  distributions, boosted W/Z

Figure 46:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)Figure 50:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)Figure 47:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)Figure 51:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)Figure 48:  $m^{TAS}$  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 49:  $m^{TAS}$  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)

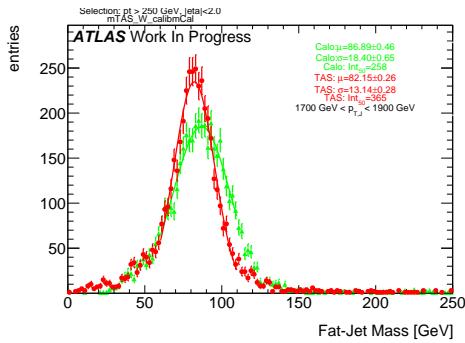


Figure 54:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

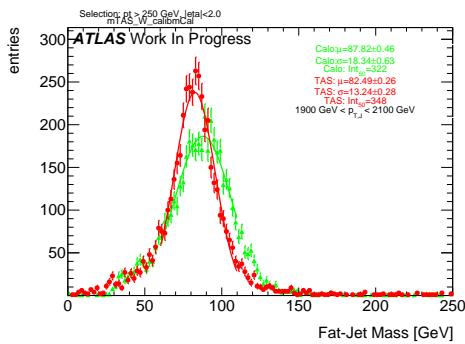


Figure 55:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

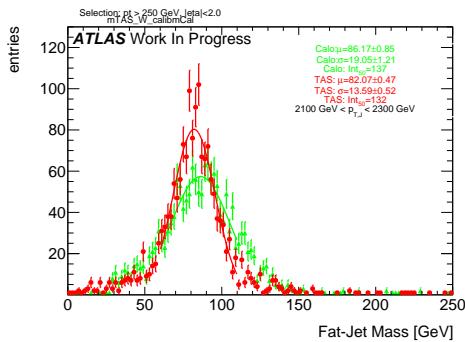


Figure 56:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

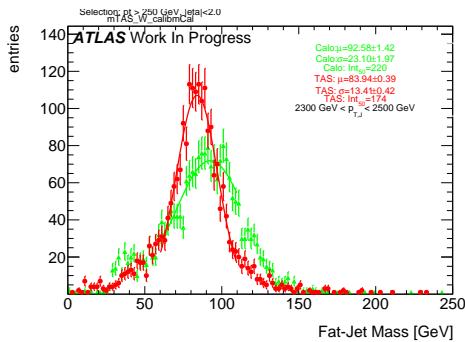


Figure 57:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

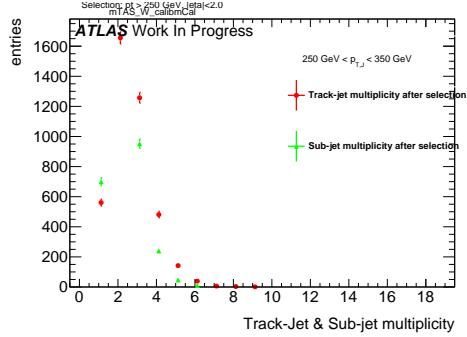


Figure 58: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

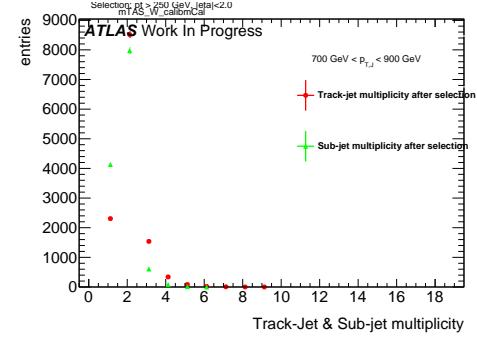


Figure 61: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

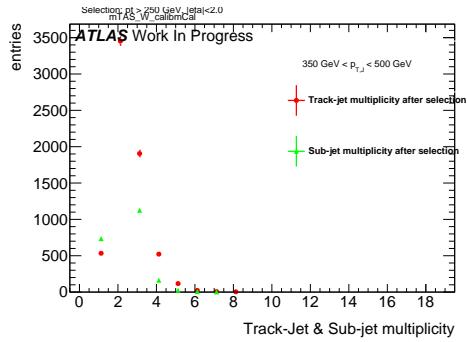


Figure 59: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

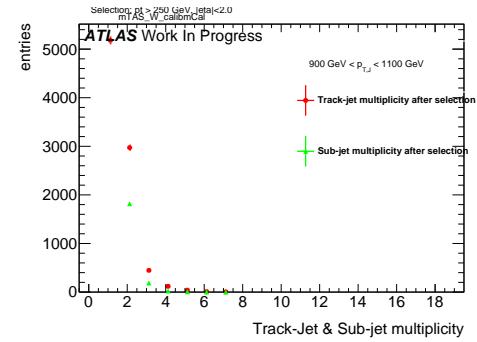


Figure 62: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

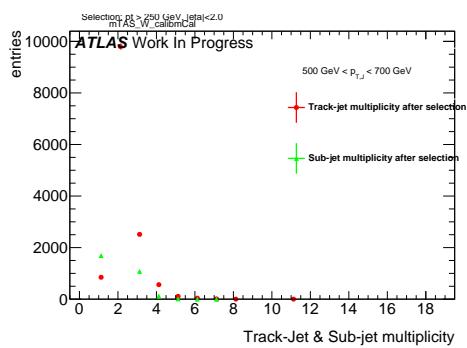


Figure 60: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

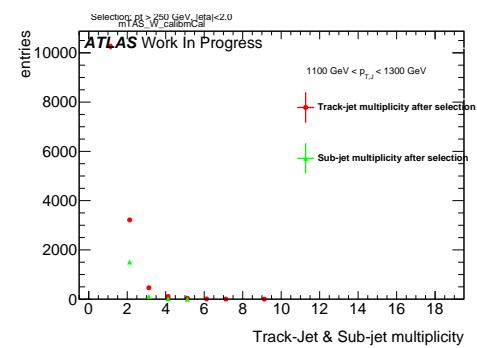


Figure 63: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

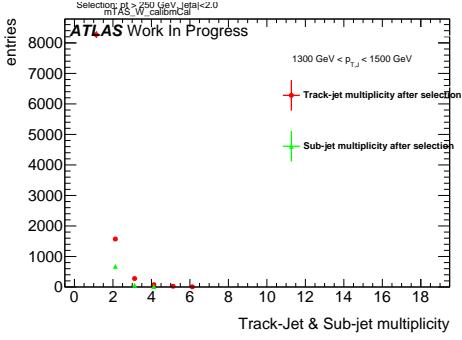


Figure 64: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

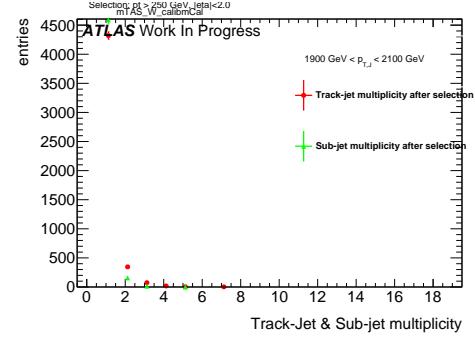


Figure 67: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

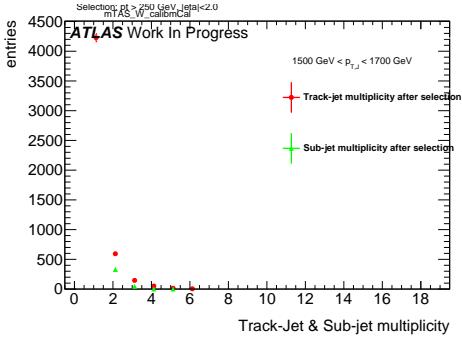


Figure 65: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

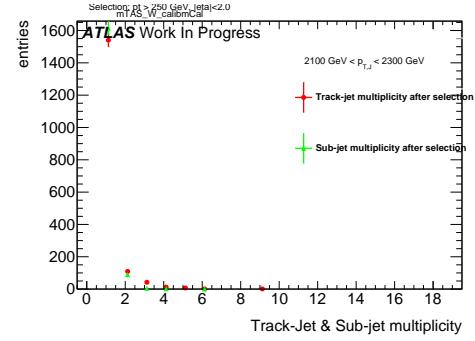


Figure 68: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

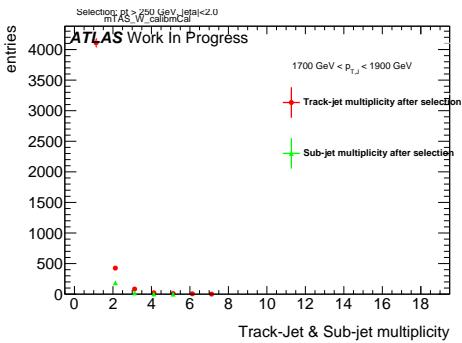


Figure 66: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

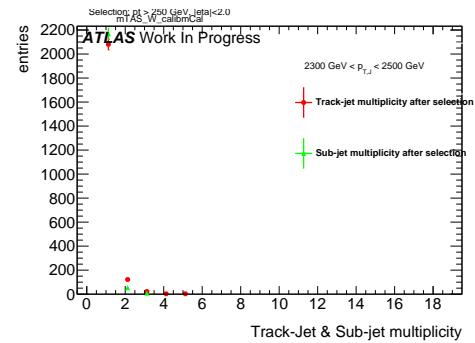


Figure 69: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

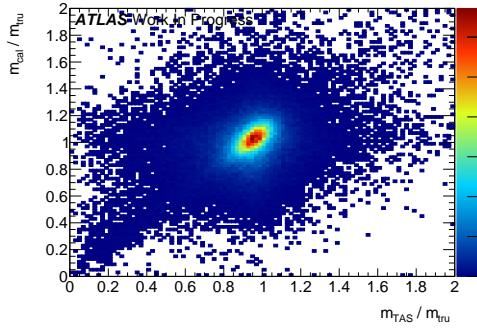
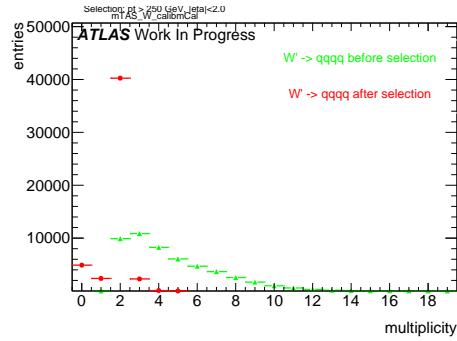
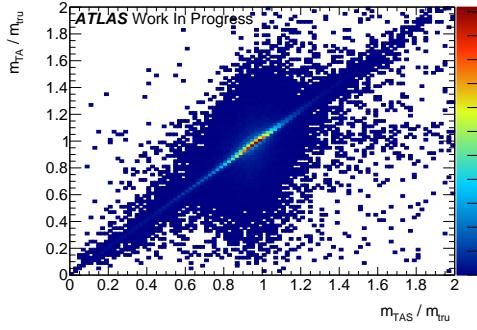
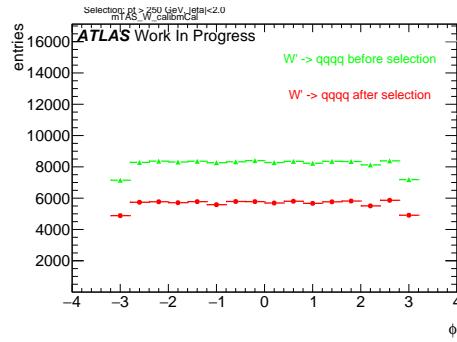
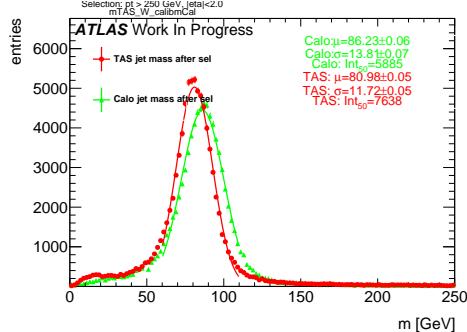
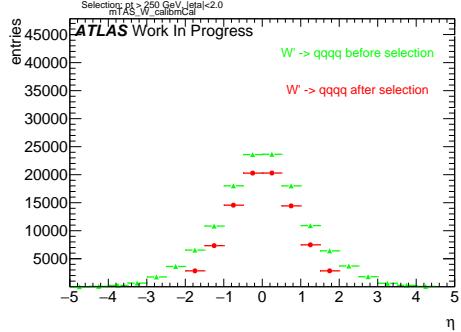
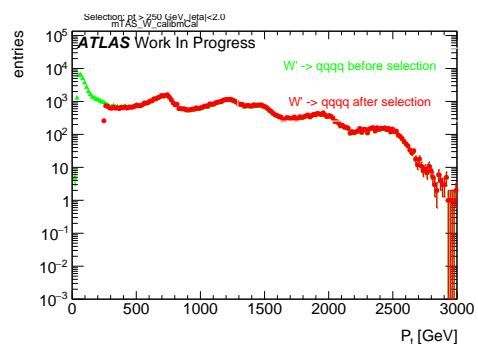
Figure 70: Scatter plot  $m^{TAS}$  versus  $m^{calo}$  responses

Figure 74: large-R jet Multiplicity, before and after selection

Figure 71: Scatter plot  $m^{TAS}$  versus  $m^{TA}$  responsesFigure 75:  $\phi$  distribution of the large-R jet, before and after selectionFigure 72:  $m^{TAS}$  distribution in all the  $p_T$  binsFigure 73:  $\eta$  distribution of the large-R jet, before and after selectionFigure 76:  $p_T$  distribution of the large-R jet, before and after selection

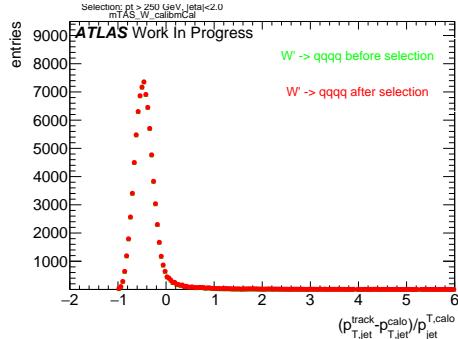


Figure 77:  $p_T$  resolution:  $\frac{p_{T,jet}^{track} - p_{T,jet}^{calo}}{p_{T,jet}^{calo}}$ , before and after selection

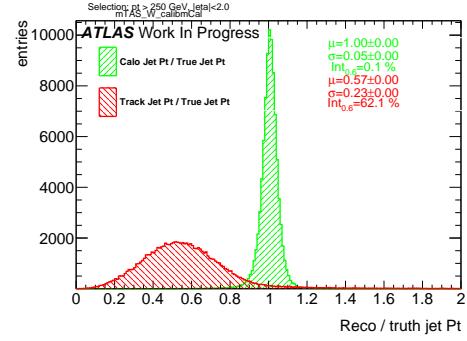


Figure 80: Transverse momentum response  $p_T^{Reco} / p_T^{Truth}$  for calorimeter and tracks

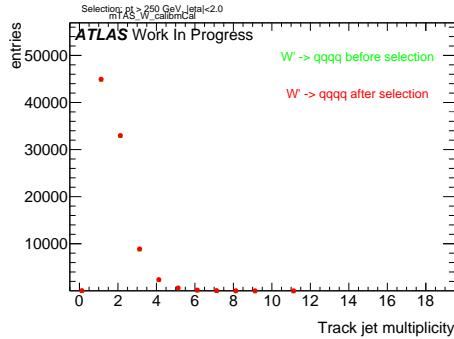


Figure 78: Multiplicity of track-jets R=0.2 per large-R jet

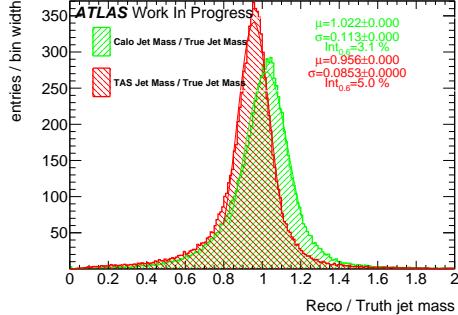


Figure 79: Response  $m^{Reco} / m^{Truth}$  for all the  $p_T$  bins

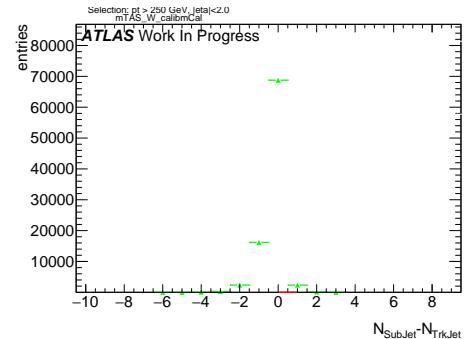


Figure 81: sub-jet - track-jet Multiplicity

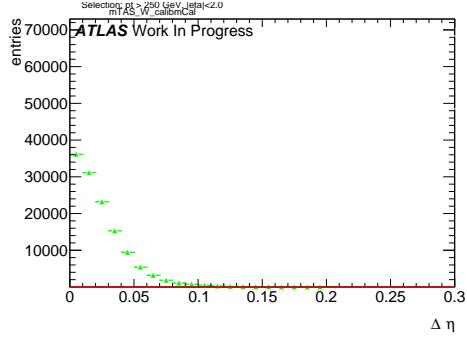


Figure 82:  $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

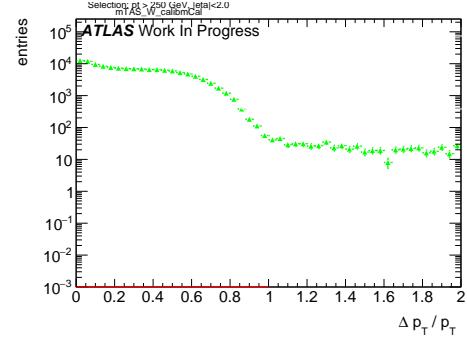


Figure 85:  $|p_{T,\text{sub-jet}} - p_{T,\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

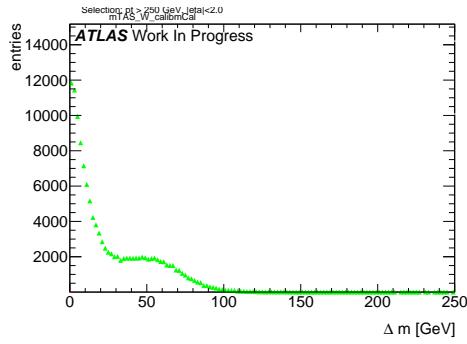


Figure 83:  $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

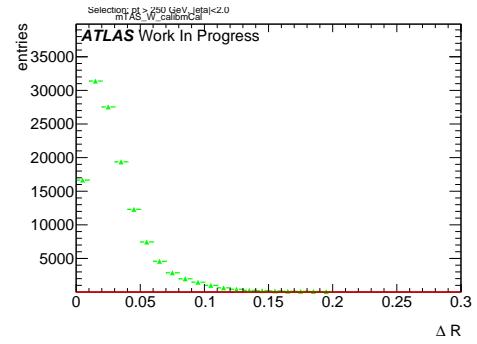


Figure 86:  $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

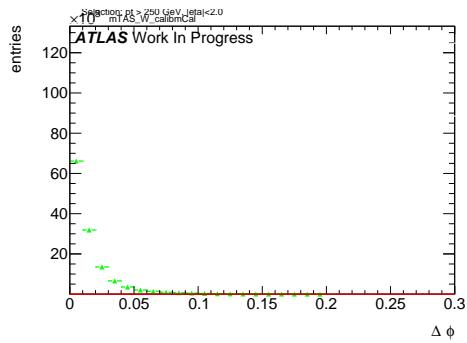


Figure 84:  $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

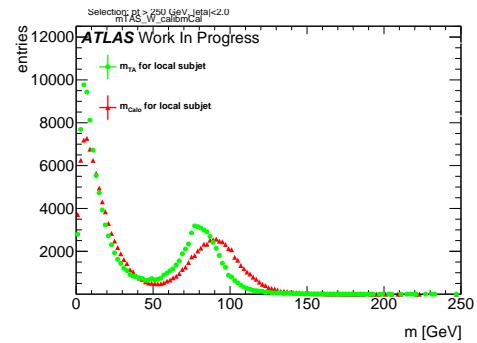


Figure 87: Mass distribution of the sub-jet, calorimeter and track-assisted

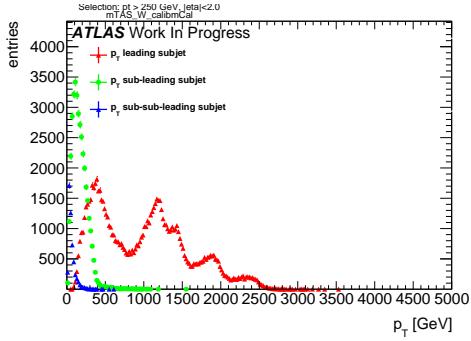


Figure 88:  $p_T$  distribution for leading, sub-leading and sub-sub-leading sub-jets

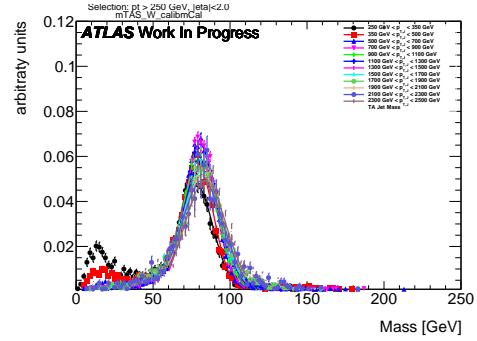


Figure 91:  $m^{TAS}$  for  $p_T^J$  bin, superimposed

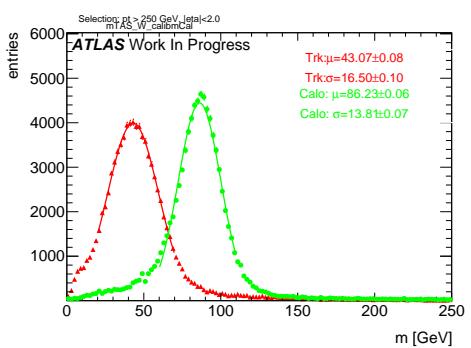


Figure 89: Mass distribution for calorimeter and tracks associated to the large- $R$  jet

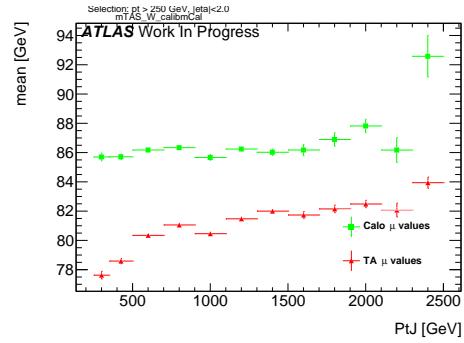


Figure 92:  $\mu$  from fit of the mass distribution vs bin of  $p_T^J$

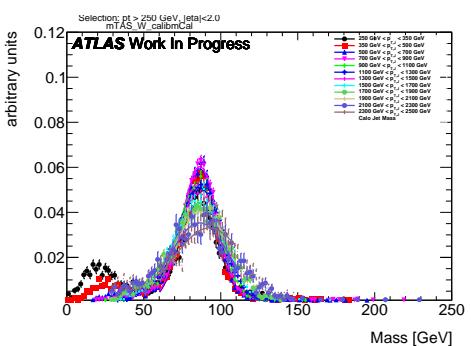


Figure 90:  $m^{calo}$  for  $p_T^J$  bin, superimposed

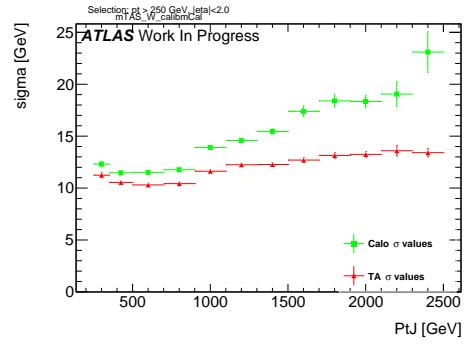


Figure 93:  $\sigma$  from fit of the mass distribution vs bin of  $p_T^J$

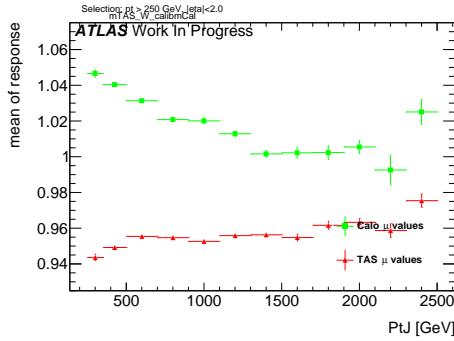


Figure 94:  $\mu$  from fit of the mass Response vs bin of  $p_T^J$

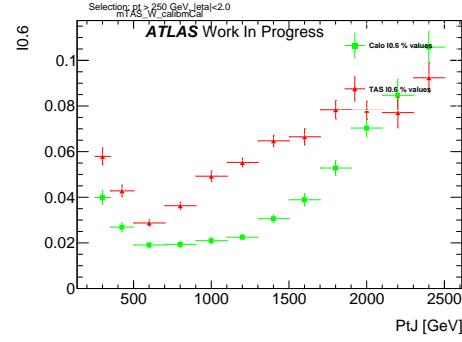


Figure 97: Left integral normalized,  $\int_0^{0.6}$  of the mass response, vs bin of  $p_T^J$

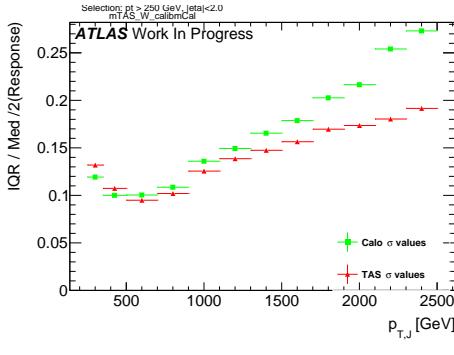


Figure 95:  $\sigma$  from fit of the mass Response vs bin of  $p_T^J$

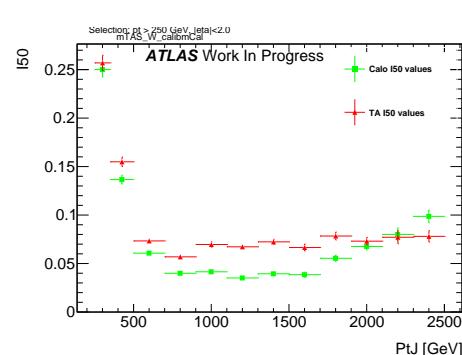


Figure 98:  $\int_0^{50GeV}$  from fit of the mass distribution vs bin of  $p_T^J$  (normalized)

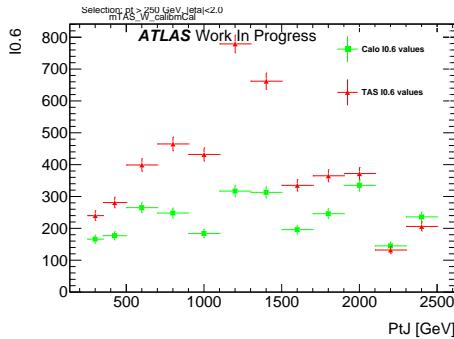


Figure 96: Left integral,  $\int_0^{0.6}$  of the mass response, vs bin of  $p_T^J$

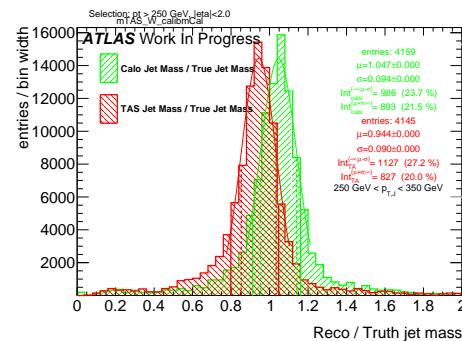
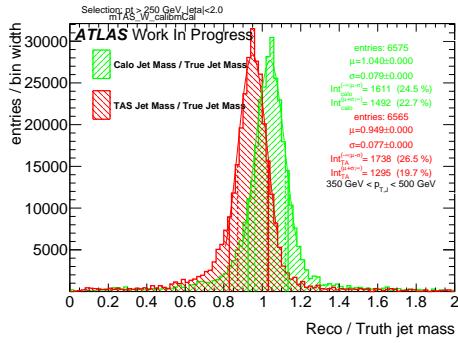
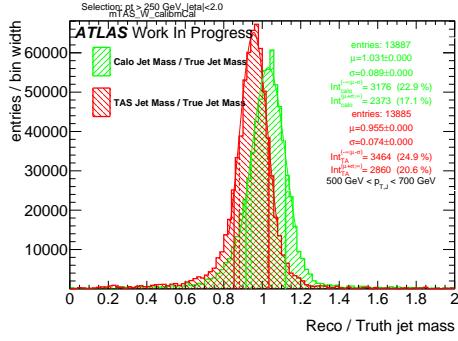
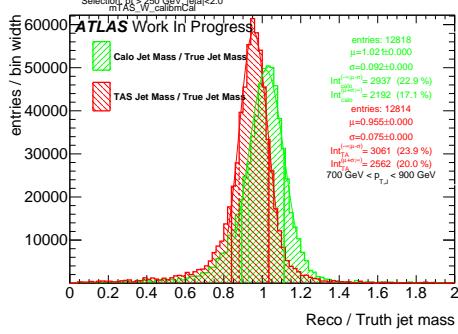
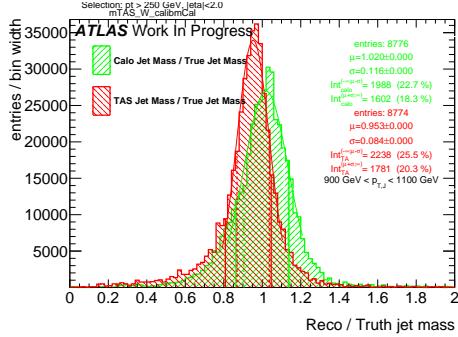
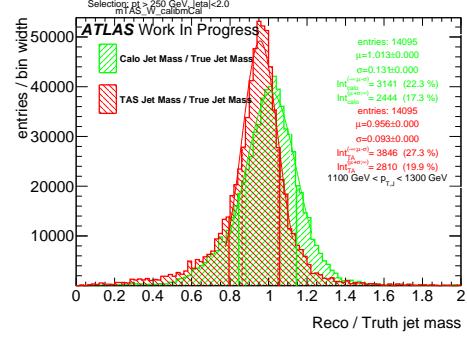
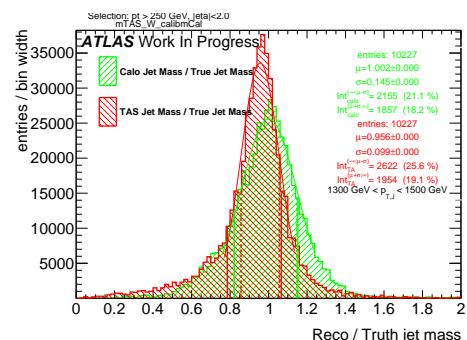
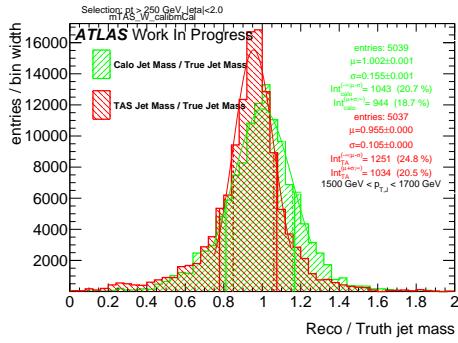
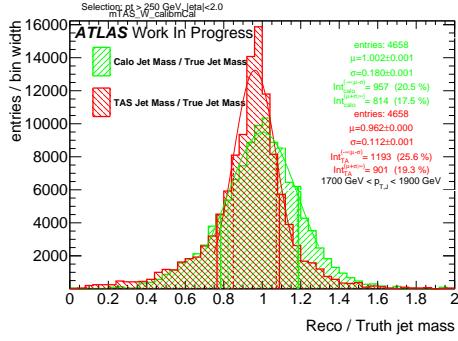
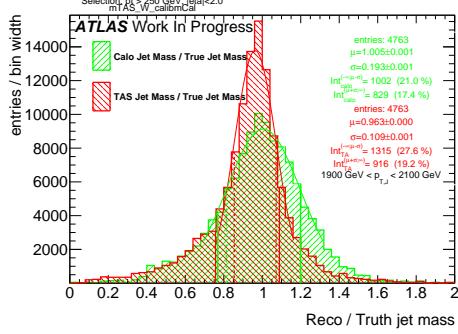
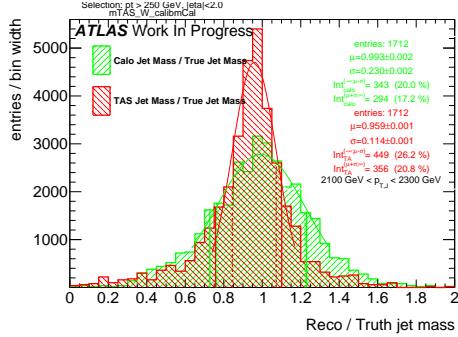
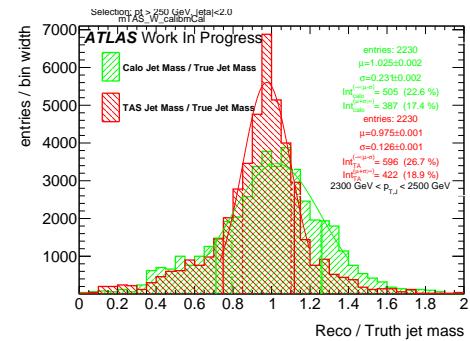


Figure 99: Response in bin of  $p_T^J$  (indicated on plot)

Figure 100: Response in bin of  $p_T^J$  (indicated on plot)Figure 101: Response in bin of  $p_T^J$  (indicated on plot)Figure 102: Response in bin of  $p_T^J$  (indicated on plot)Figure 103: Response in bin of  $p_T^J$  (indicated on plot)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)

<sub>868</sub> **F  $m^{TAS}$  distributions, boosted tops**

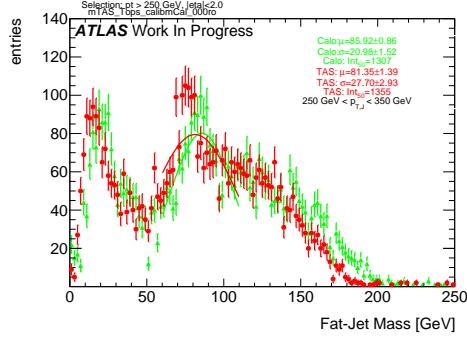


Figure 111:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

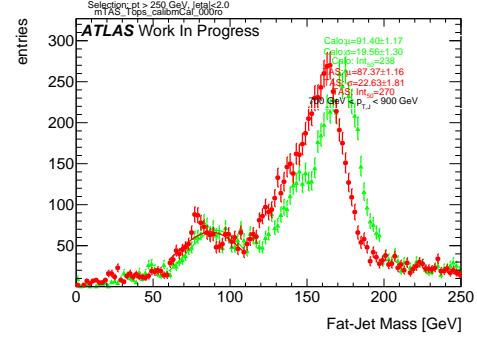


Figure 114:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

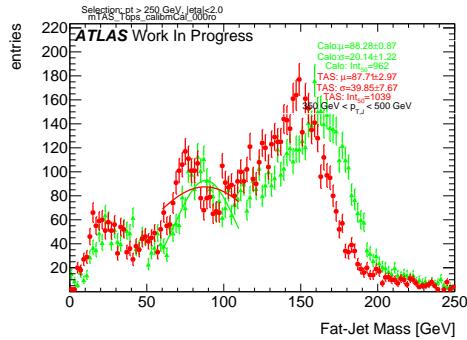


Figure 112:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

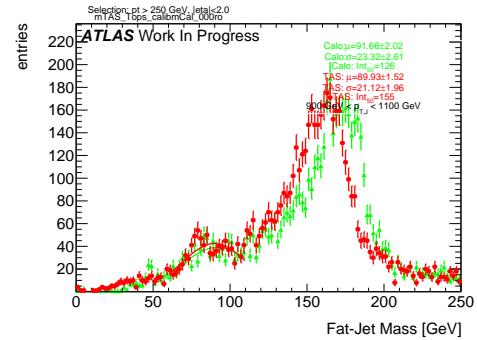


Figure 115:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

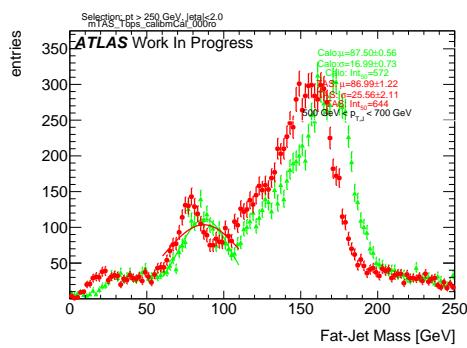


Figure 113:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

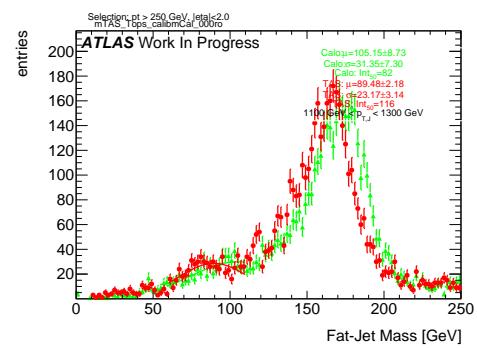


Figure 116:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

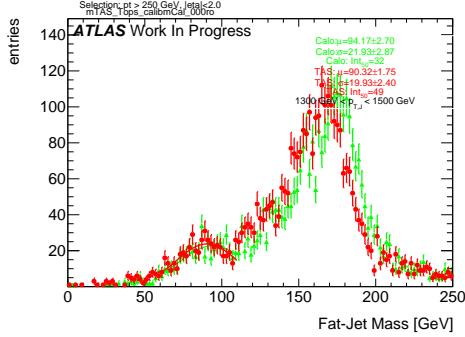


Figure 117:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

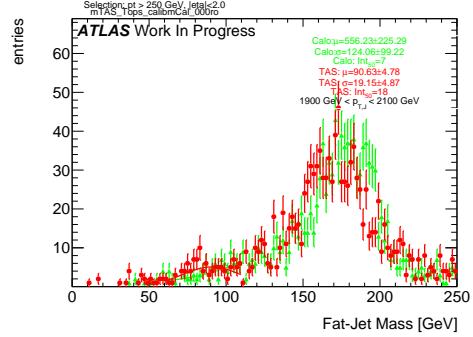


Figure 120:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

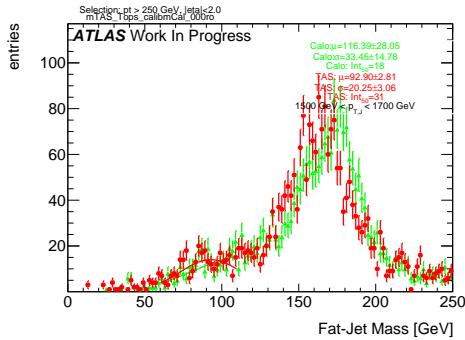


Figure 118:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

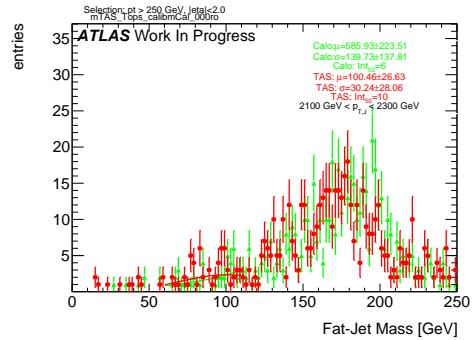


Figure 121:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

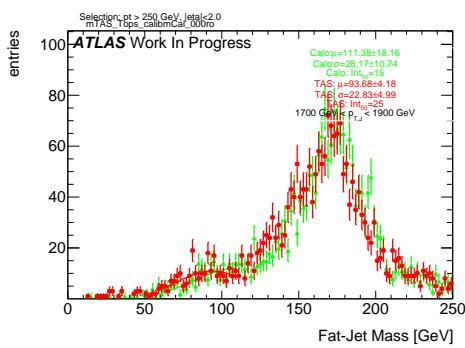


Figure 119:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

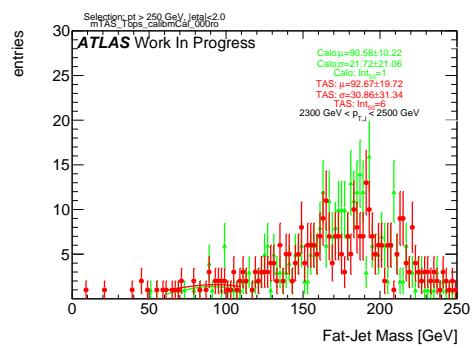


Figure 122:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

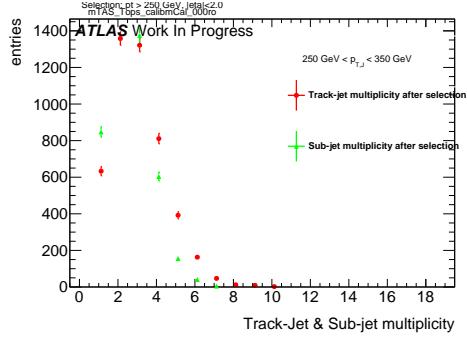


Figure 123: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

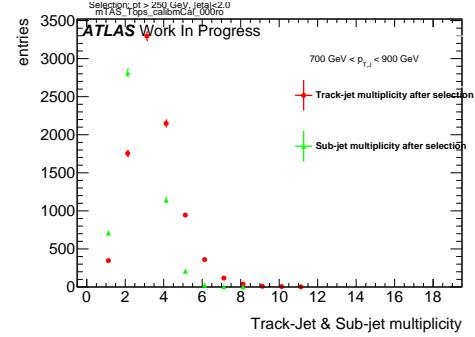


Figure 126: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

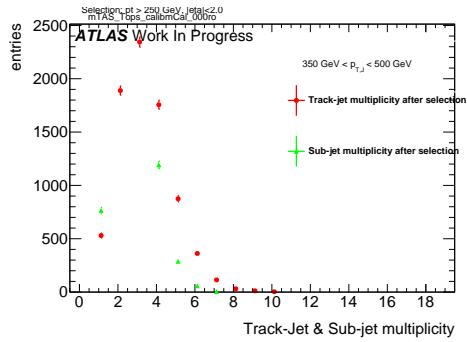


Figure 124: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

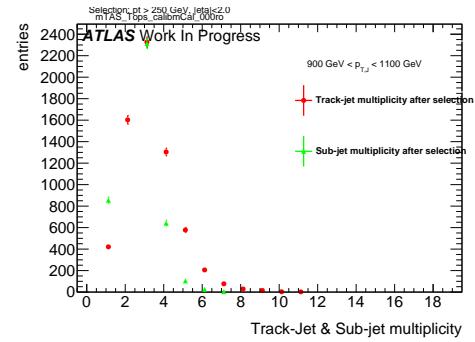


Figure 127: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

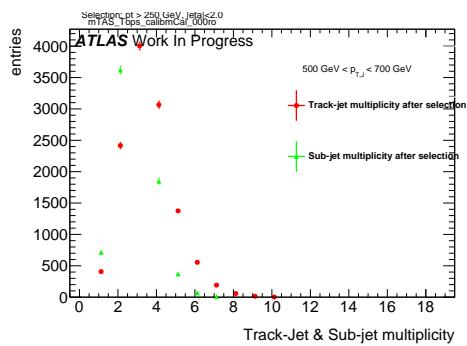


Figure 125: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

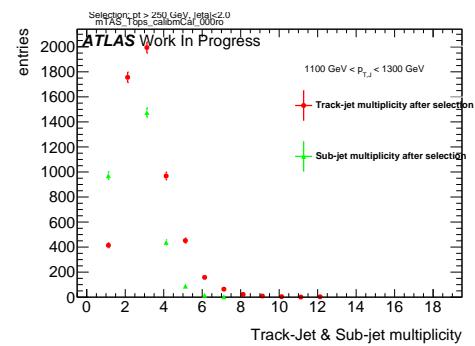


Figure 128: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

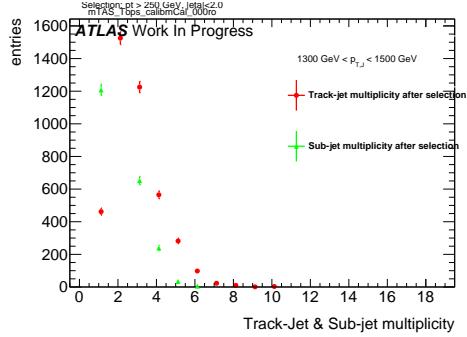


Figure 129: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

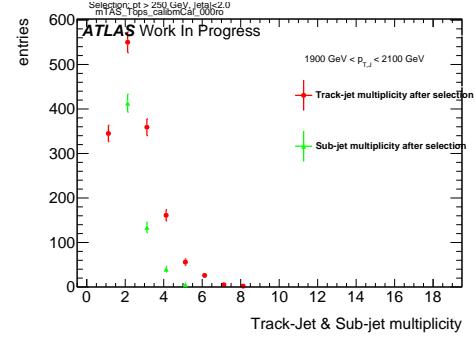


Figure 132: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

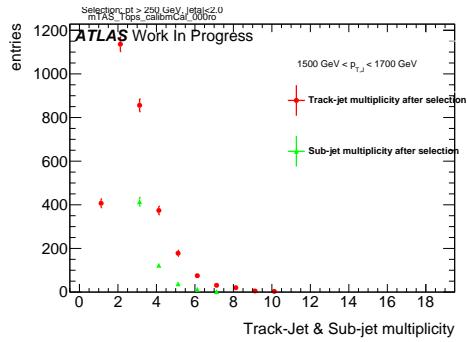


Figure 130: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

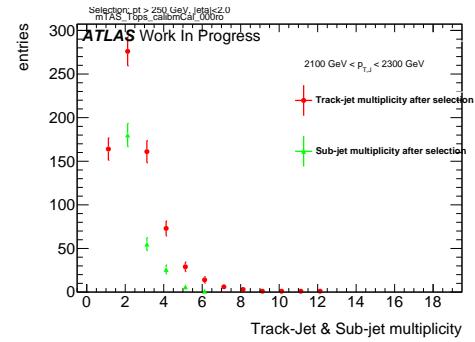


Figure 133: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

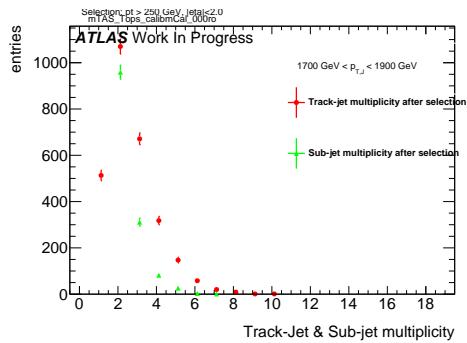


Figure 131: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

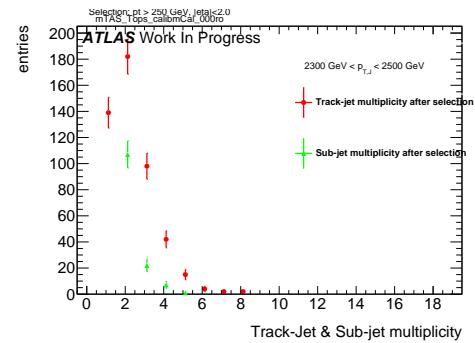


Figure 134: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

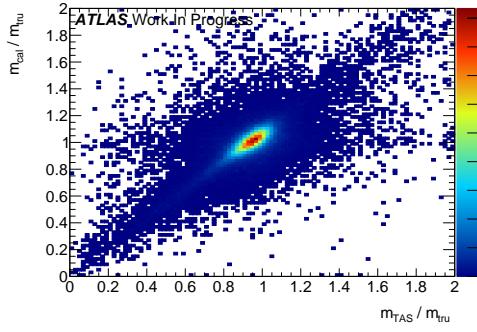
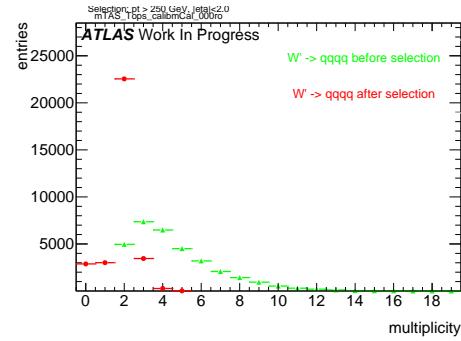
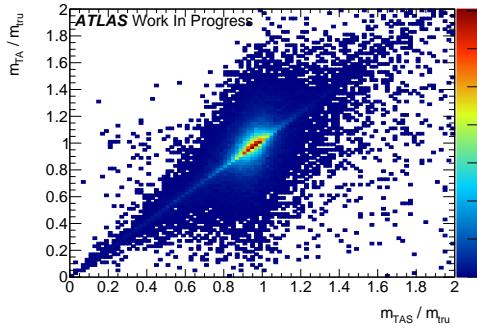
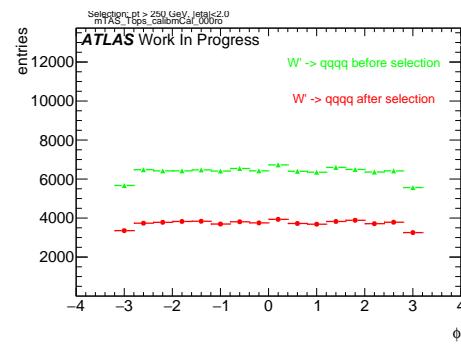
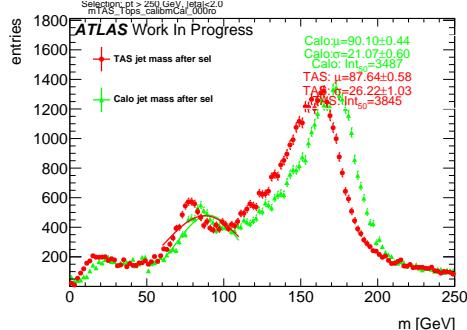
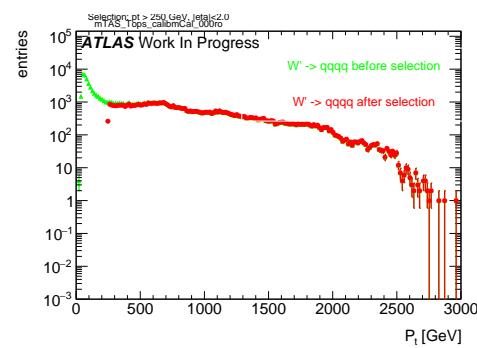
Figure 135: Scatter plot  $m^{TAS}$  versus  $m^{calo}$  responses

Figure 139: large-R jet Multiplicity, before and after selection

Figure 136: Scatter plot  $m^{TAS}$  versus  $m^A$  responsesFigure 140:  $\phi$  distribution of the large-R jet, before and after selectionFigure 137:  $m^{TAS}$  distribution in all the  $p_T$  binsFigure 141:  $p_T$  distribution of the large-R jet, before and after selectionFigure 138:  $\eta$  distribution of the large-R jet, before and after selection

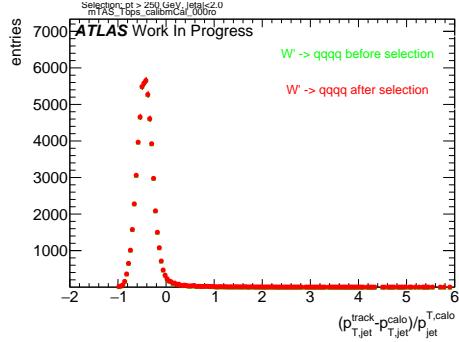


Figure 142:  $p_T$  resolution:  $\frac{p_{T,jet}^{track} - p_{T,jet}^{fat}}{p_{T,jet}^{fat}}$ , before and after selection

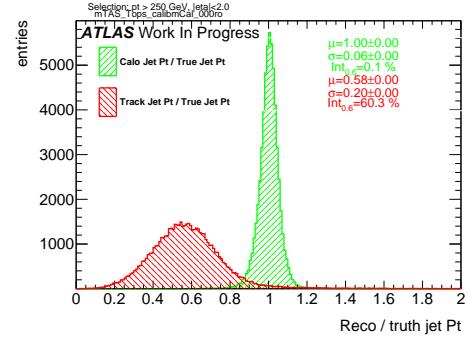


Figure 145: Transverse momentum response  $p_T^{Reco}/p_T^{Truth}$  for calorimeter and tracks

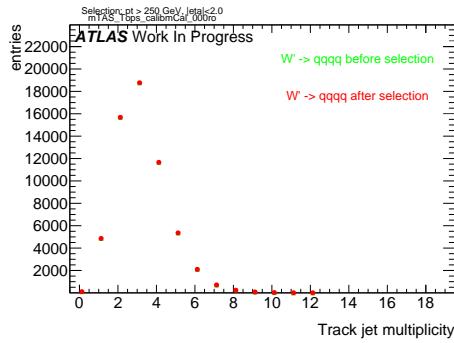


Figure 143: Multiplicity of track-jets R=0.2 per large-R jet

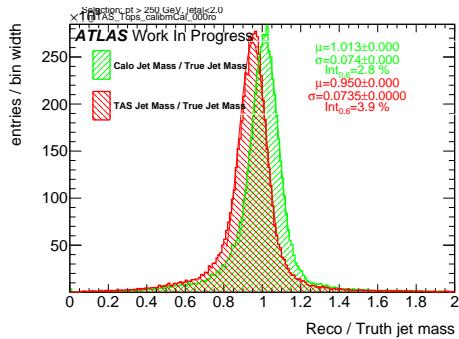


Figure 144: Response  $m^{Reco}/m^{Truth}$  for all the  $p_T$  bins

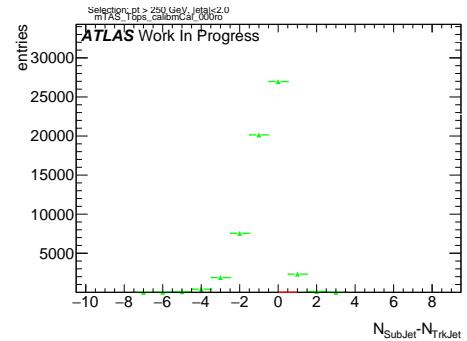


Figure 146: sub-jet - track-jet Multiplicity

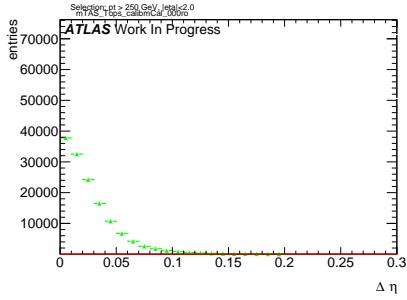


Figure 147:  $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

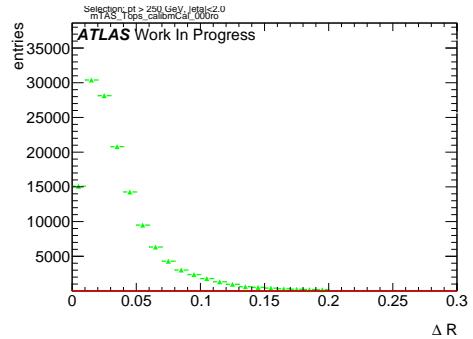


Figure 151:  $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

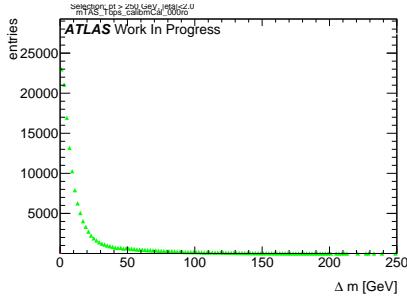


Figure 148:  $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

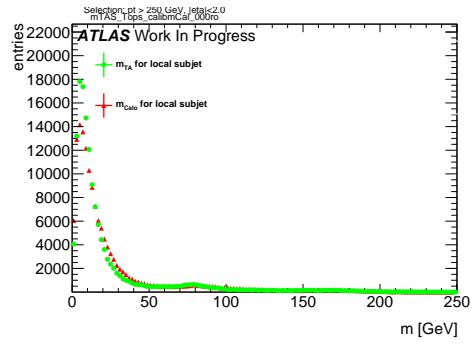


Figure 152: Mass distribution of the sub-jet, calorimeter and track-assisted

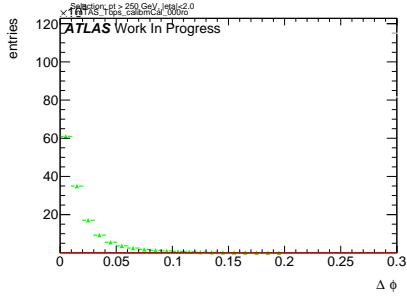


Figure 149:  $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

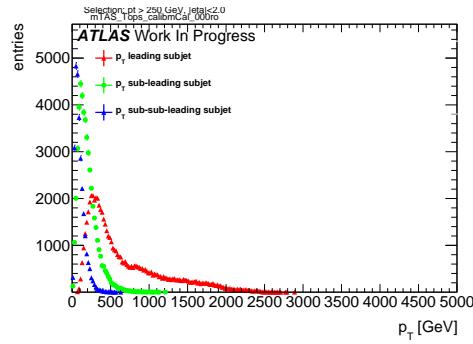


Figure 153:  $p_T$  distribution for leading, sub-leading and sub-sub-leading sub-jets

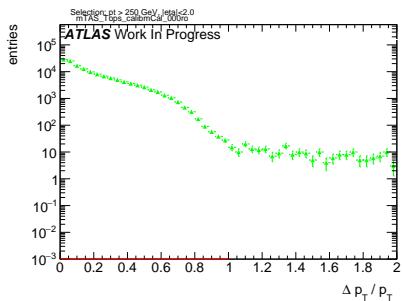


Figure 150:  $|p_{T,\text{sub-jet}} - p_{T,\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

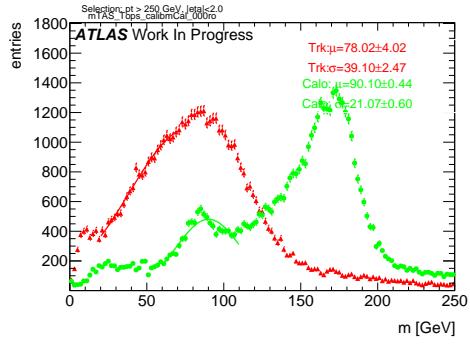


Figure 154: Mass distribution for calorimeter and tracks associated to the large-R jet

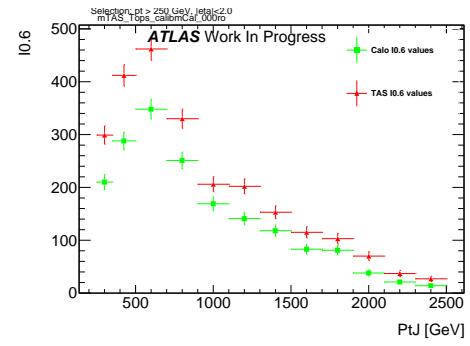


Figure 157: Left integral,  $\int_0^{0.6}$  of the mass response, vs bin of  $p_T^J$

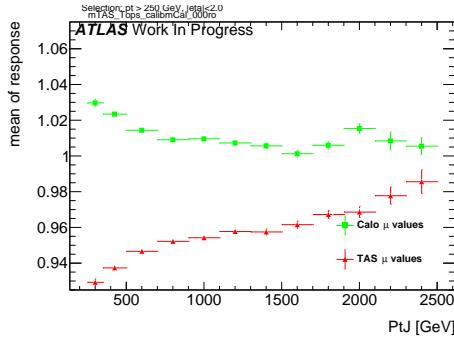


Figure 155:  $\mu$  from fit of the mass Response vs bin of  $p_T^J$

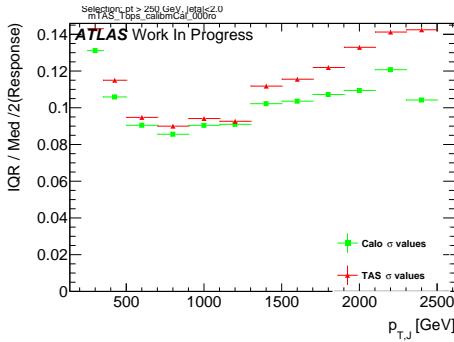


Figure 156:  $\sigma$  from fit of the mass Response vs bin of  $p_T^J$

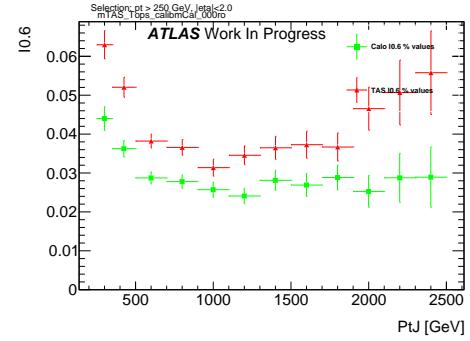
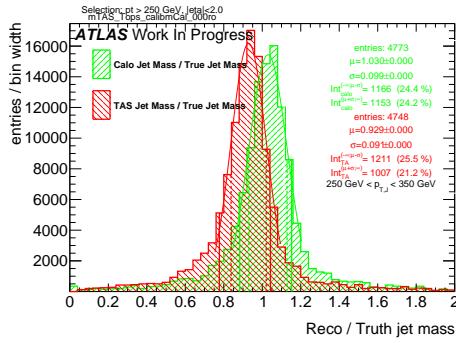
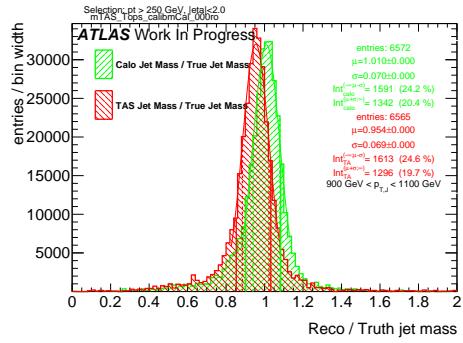
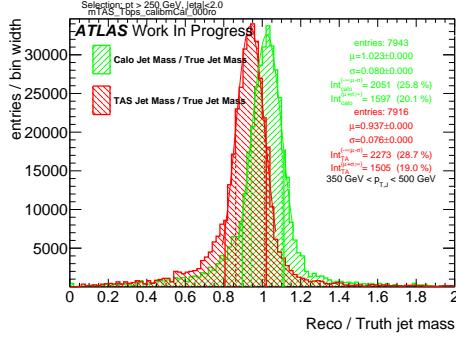
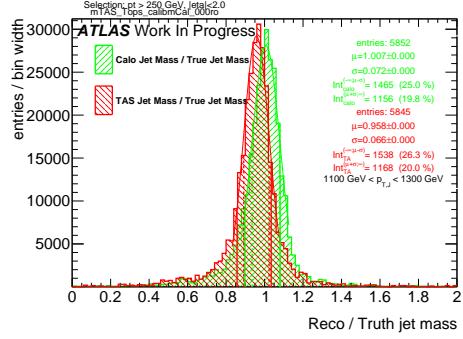
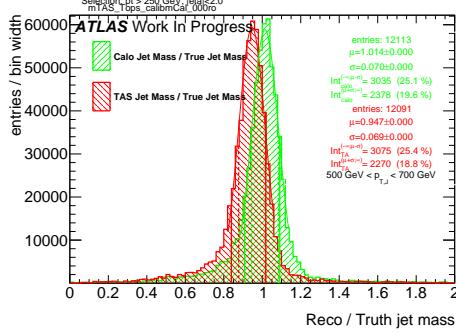
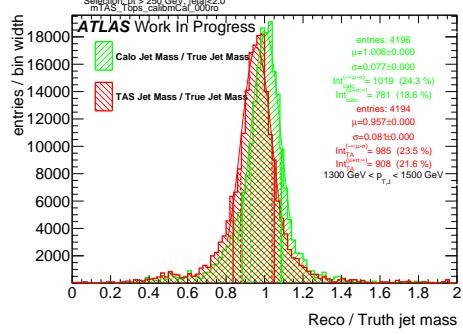
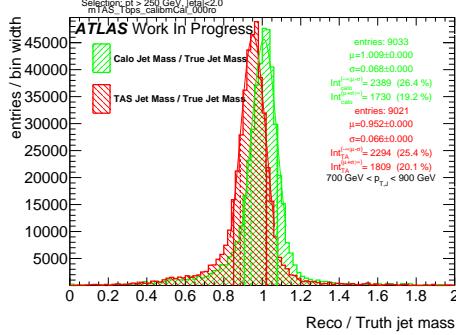
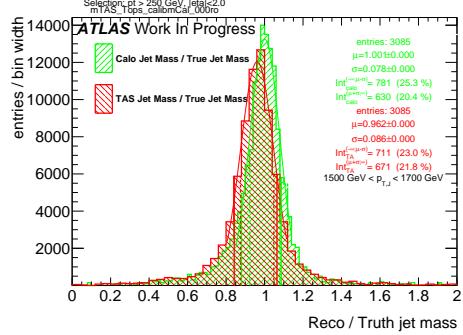
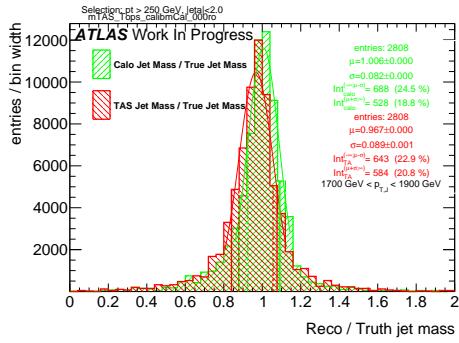
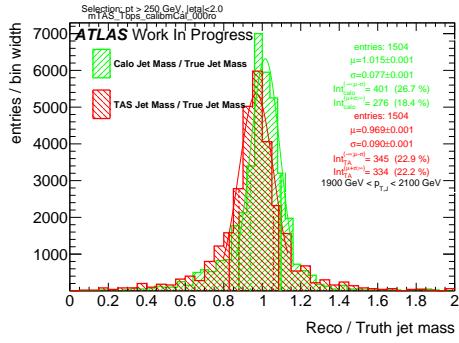
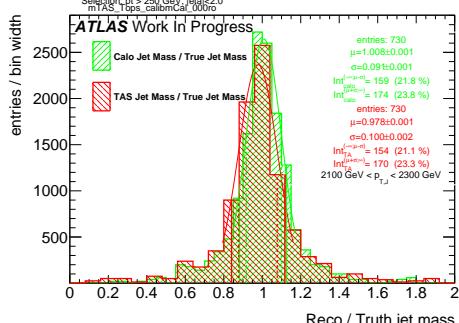
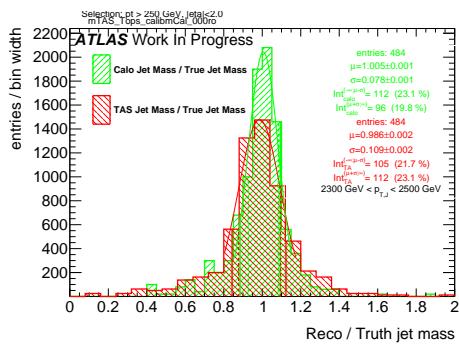


Figure 158: Left integral normalized,  $\int_0^{0.6}$  of the mass response, vs bin of  $p_T^J$

Figure 159: Response in bin of  $p_T^J$  (indicated on plot)Figure 163: Response in bin of  $p_T^J$  (indicated on plot)Figure 160: Response in bin of  $p_T^J$  (indicated on plot)Figure 164: Response in bin of  $p_T^J$  (indicated on plot)Figure 161: Response in bin of  $p_T^J$  (indicated on plot)Figure 165: Response in bin of  $p_T^J$  (indicated on plot)Figure 162: Response in bin of  $p_T^J$  (indicated on plot)Figure 166: Response in bin of  $p_T^J$  (indicated on plot)

Figure 167: Response in bin of  $p_T^J$  (indicated on plot)Figure 168: Response in bin of  $p_T^J$  (indicated on plot)Figure 169: Response in bin of  $p_T^J$  (indicated on plot)Figure 170: Response in bin of  $p_T^J$  (indicated on plot)

<sub>869</sub> **G  $m^{TAS}$  distributions, boosted higgs**

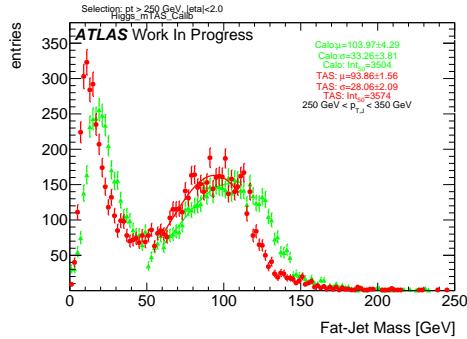


Figure 171:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

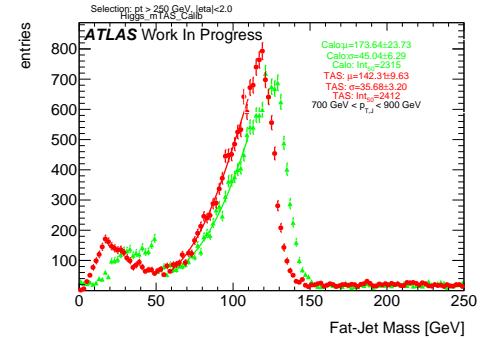


Figure 174:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

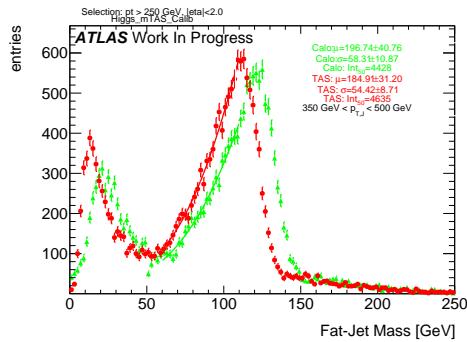


Figure 172:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

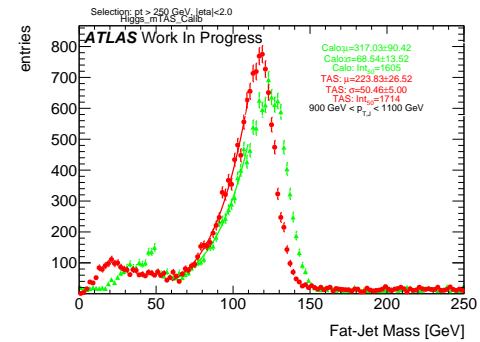


Figure 175:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

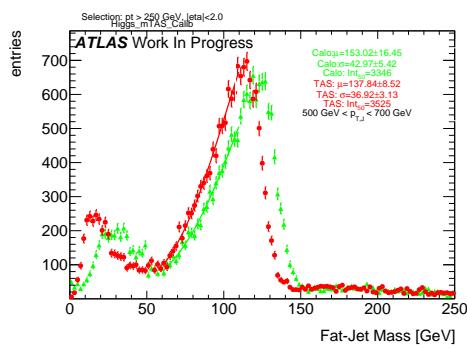


Figure 173:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

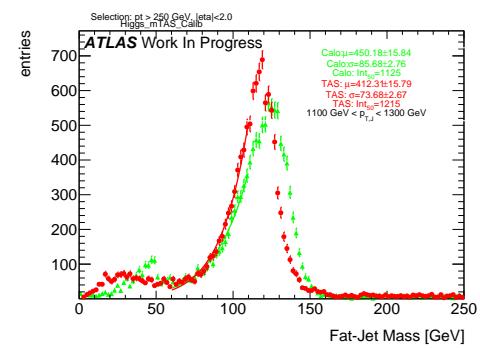


Figure 176:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

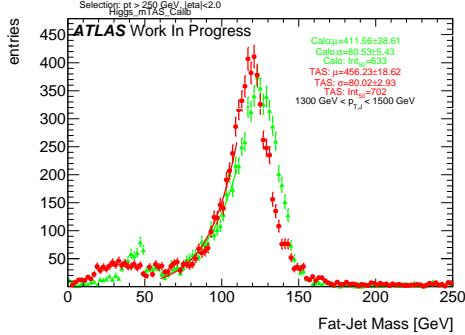


Figure 177:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

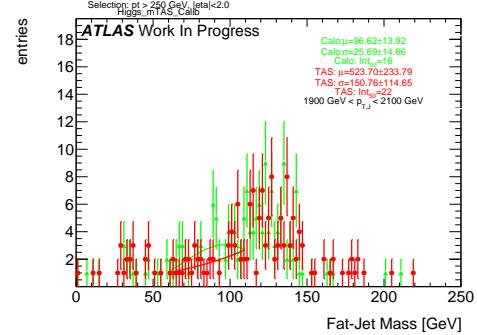


Figure 180:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

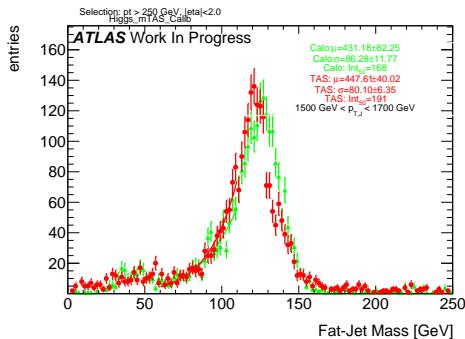


Figure 178:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

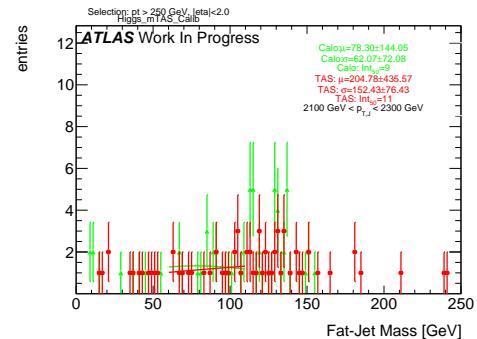


Figure 181:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

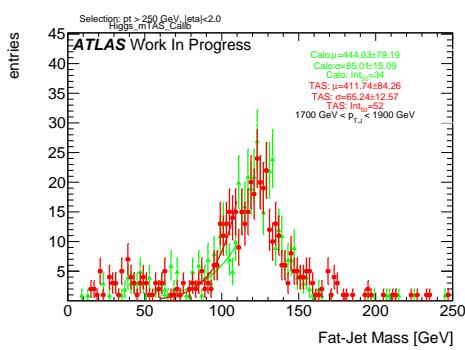


Figure 179:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

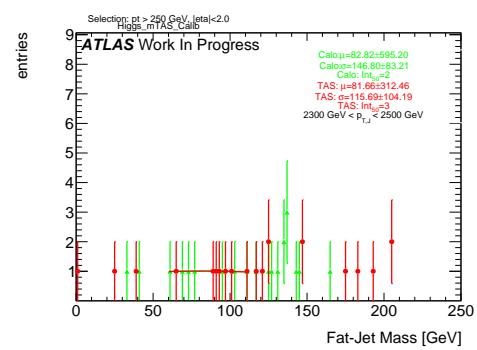


Figure 182:  $m^{TAS}$  and  $m^{calo}$  for  $p_T^J$  bin (indicated on plot)

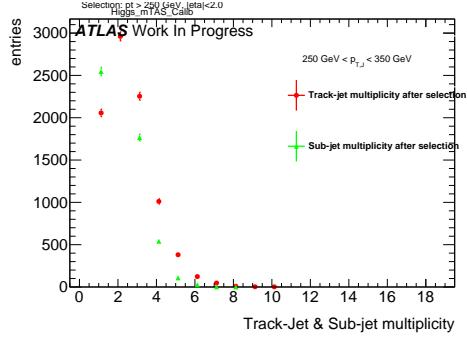


Figure 183: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

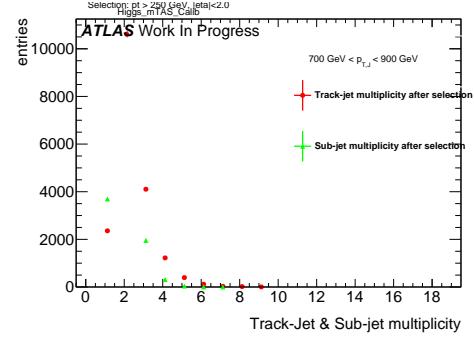


Figure 186: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

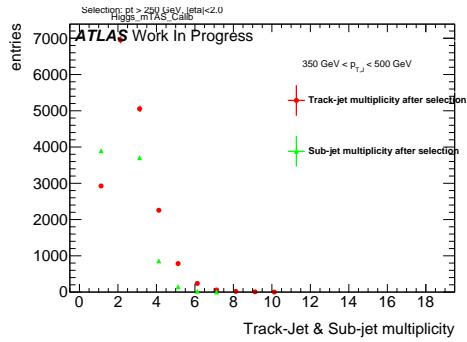


Figure 184: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

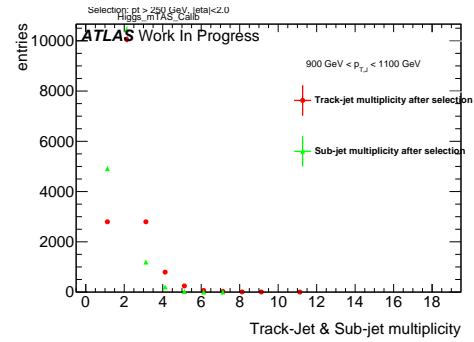


Figure 187: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

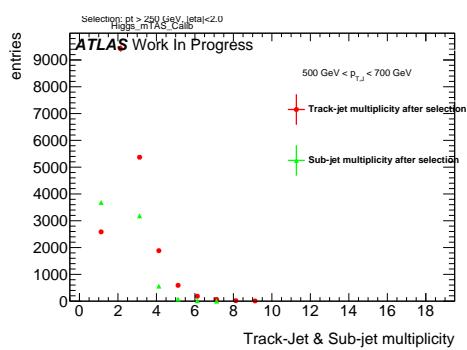


Figure 185: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

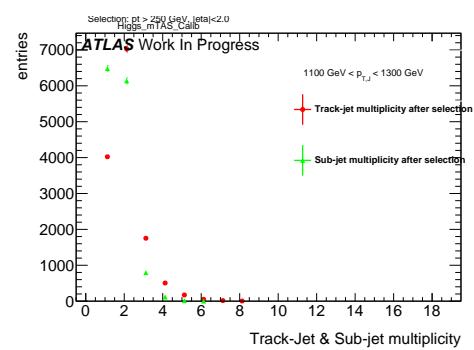


Figure 188: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

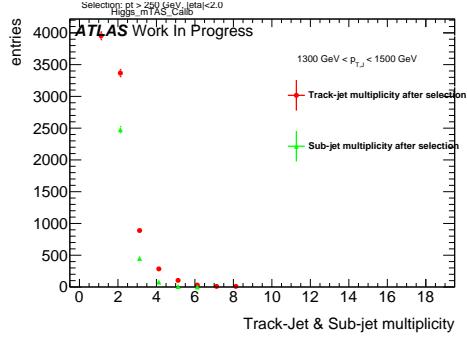


Figure 189: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

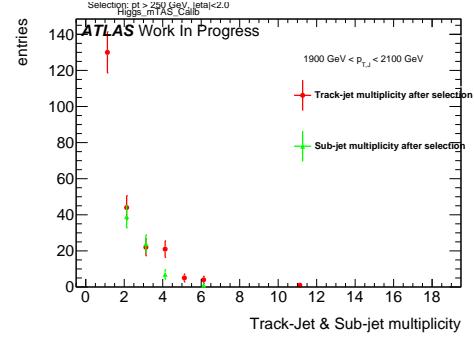


Figure 192: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

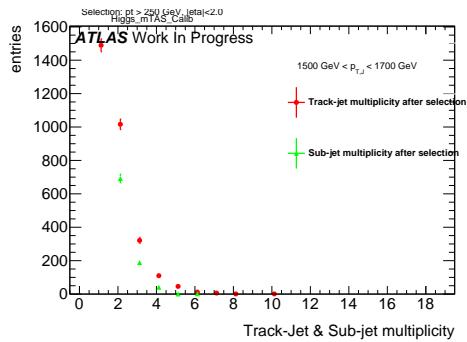


Figure 190: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

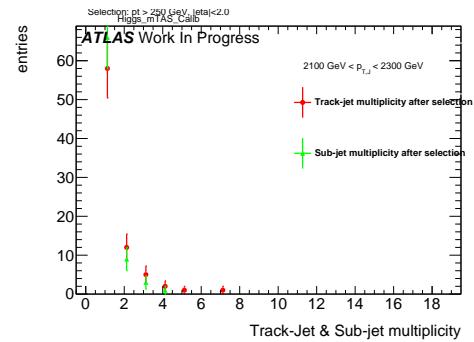


Figure 193: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

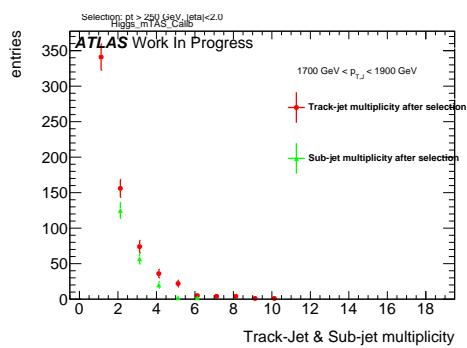


Figure 191: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

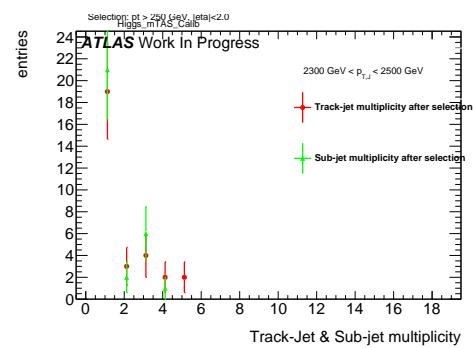


Figure 194: Track-jet R=0.2 and sub-jet multiplicity for  $p_T^J$  bin (indicated on plot)

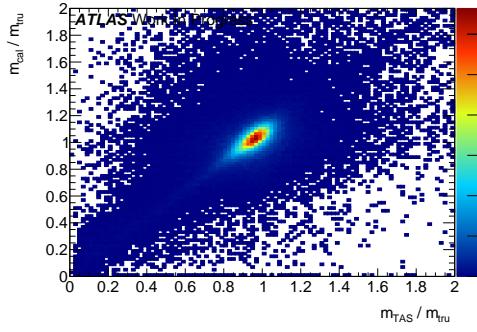
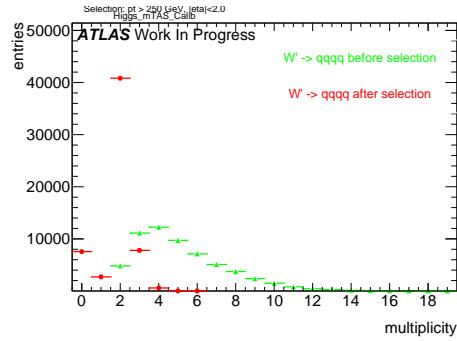
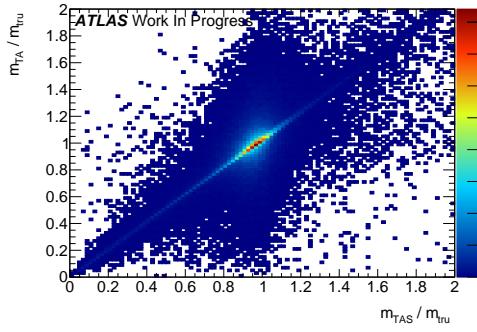
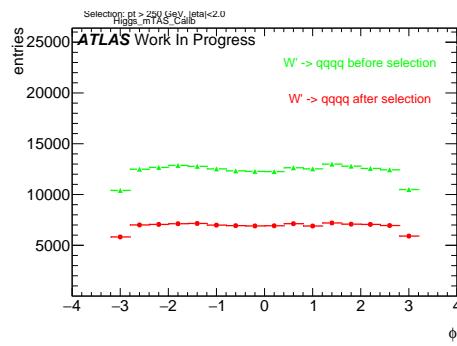
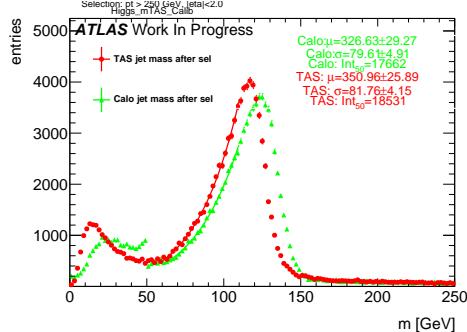
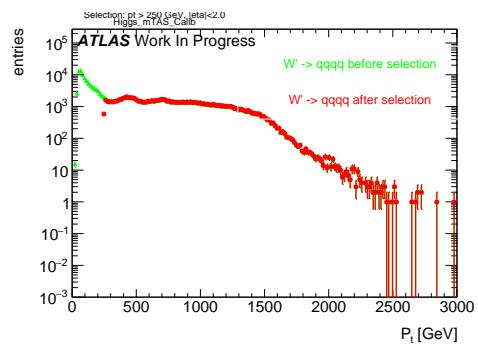
Figure 195: Scatter plot  $m^{TAS}$  versus  $m^{calo}$  responses

Figure 199: large-R jet Multiplicity, before and after selection

Figure 196: Scatter plot  $m^{TAS}$  versus  $m^A$  responsesFigure 200:  $\phi$  distribution of the large-R jet, before and after selectionFigure 197:  $m^{TAS}$  distribution in all the  $p_T$  binsFigure 201:  $p_T$  distribution of the large-R jet, before and after selectionFigure 198:  $\eta$  distribution of the large-R jet, before and after selection

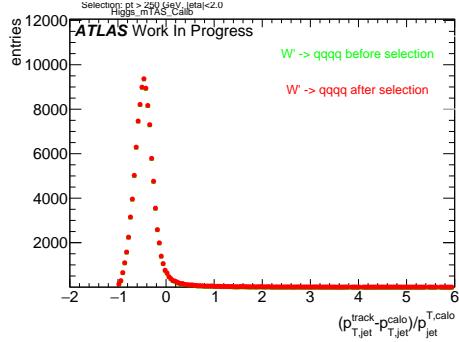


Figure 202:  $p_T$  resolution:  $\frac{p_{T,jet}^{track} - p_{T,jet}^{fat}}{p_{T,jet}^{fat}}$ , before and after selection

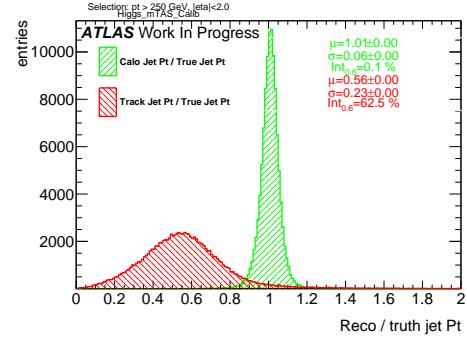


Figure 205: Transverse momentum response  $p_T^{Reco} / p_T^{Truth}$  for calorimeter and tracks

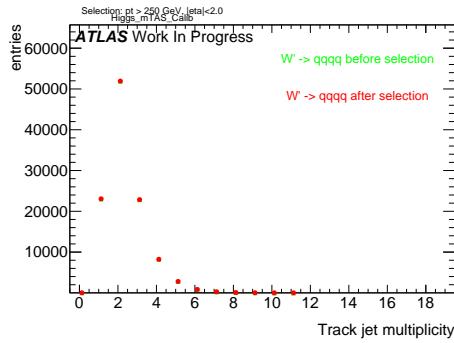


Figure 203: Multiplicity of track-jets R=0.2 per large-R jet

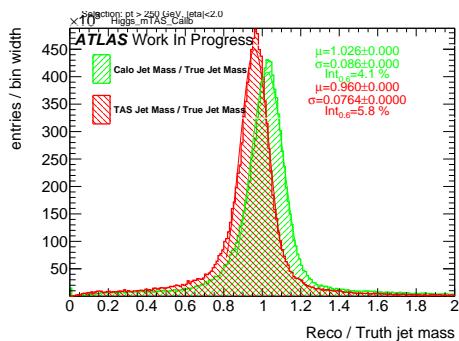


Figure 204: Response  $m^{Reco} / m^{Truth}$  for all the  $p_T$  bins

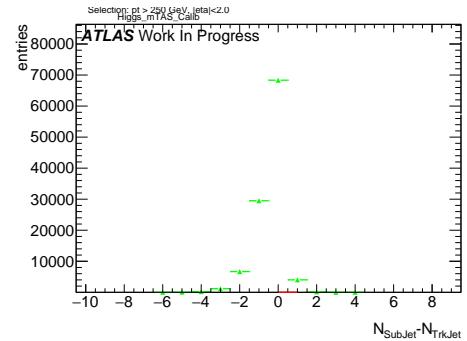


Figure 206: sub-jet - track-jet Multiplicity

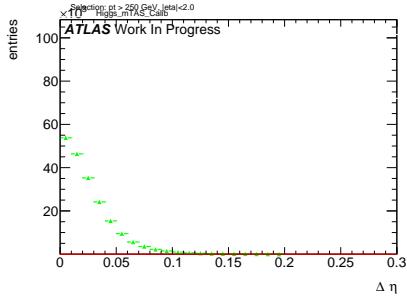


Figure 207:  $|\eta_{\text{sub-jet}} - \eta_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

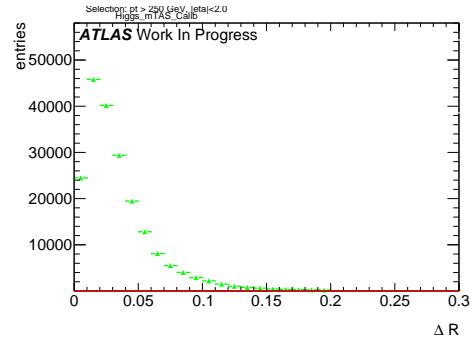


Figure 211:  $|R_{\text{sub-jet}} - R_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

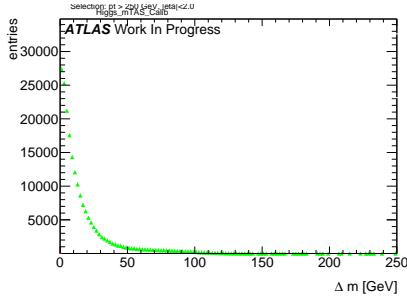


Figure 208:  $|m_{\text{sub-jet}} - m_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

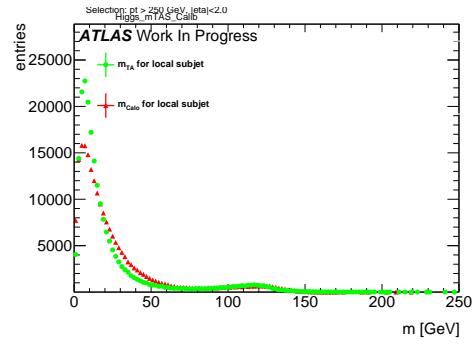


Figure 212: Mass distribution of the sub-jet, calorimeter and track-assisted

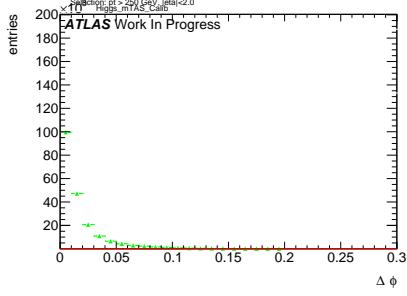


Figure 209:  $|\phi_{\text{sub-jet}} - \phi_{\text{track-jet}}|$  distribution, where sub-jet and track-jet are the closest

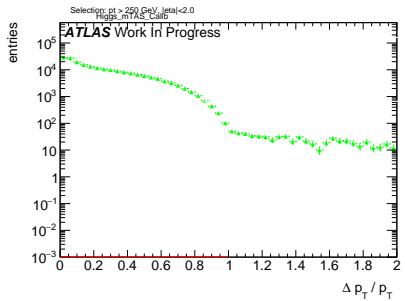


Figure 210:  $|p_{\text{T,sub-jet}} - p_{\text{T,track-jet}}|$  distribution, where sub-jet and track-jet are the closest

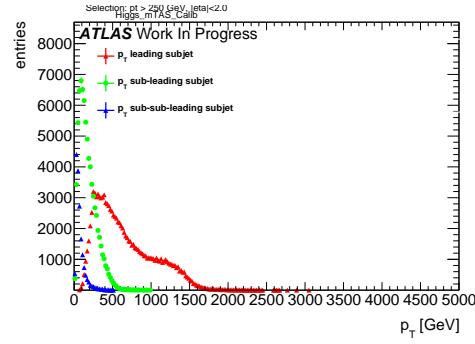


Figure 213:  $p_{\text{T}}$  distribution for leading, sub-leading and sub-sub-leading sub-jets

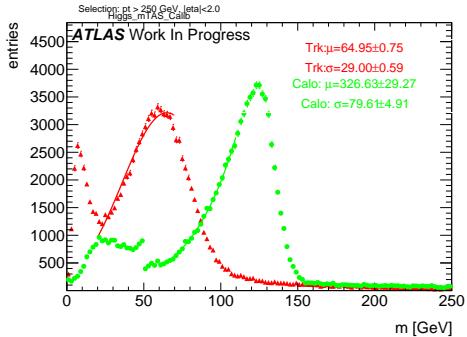


Figure 214: Mass distribution for calorimeter and tracks associated to the large-R jet

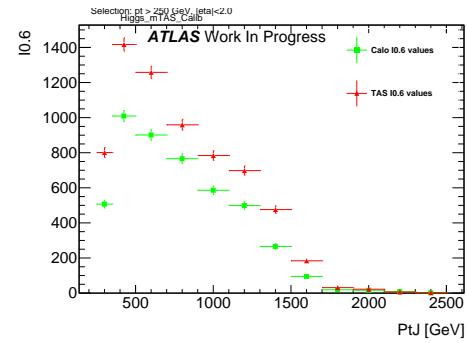


Figure 217: Left integral,  $\int_0^{0.6}$  of the mass response, vs bin of  $p_T^J$

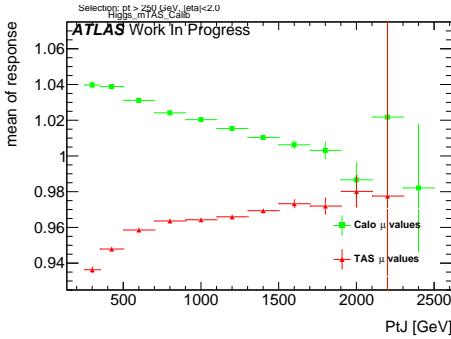


Figure 215:  $\mu$  from fit of the mass Response vs bin of  $p_T^J$

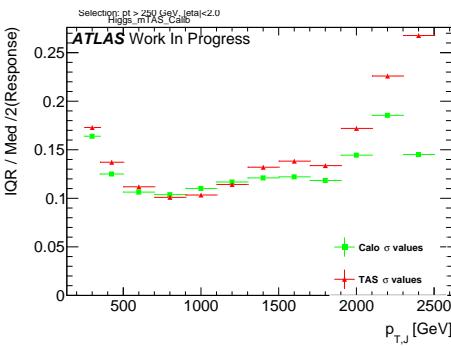


Figure 216:  $\sigma$  from fit of the mass Response vs bin of  $p_T^J$

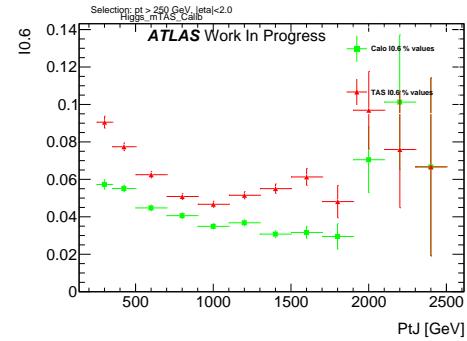
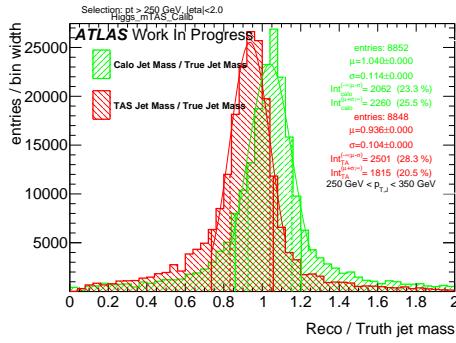
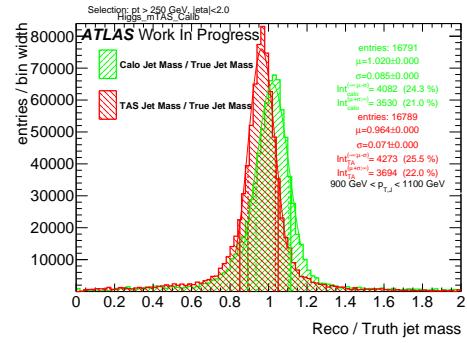
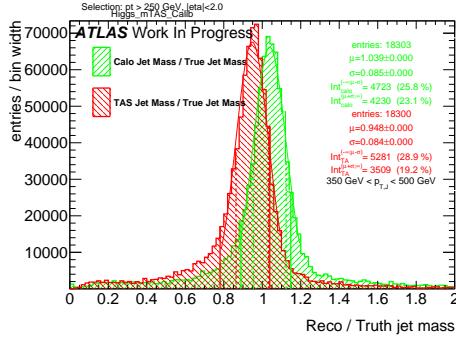
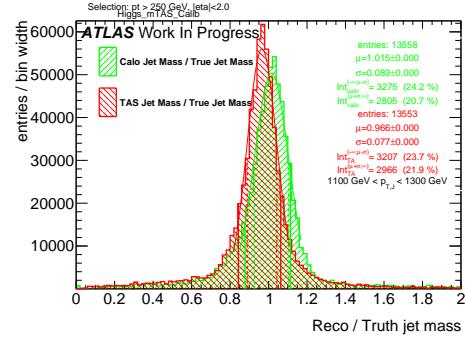
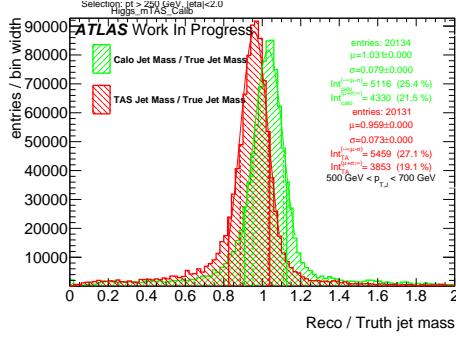
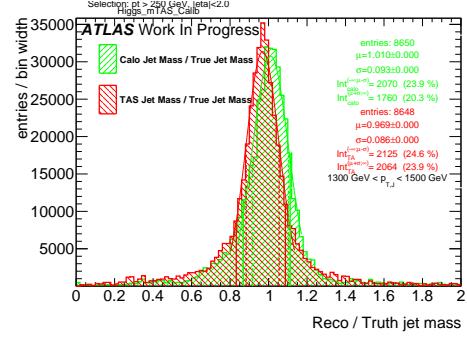
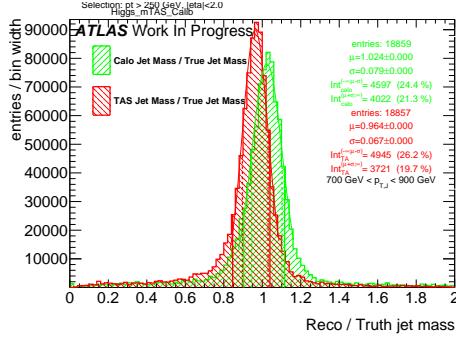
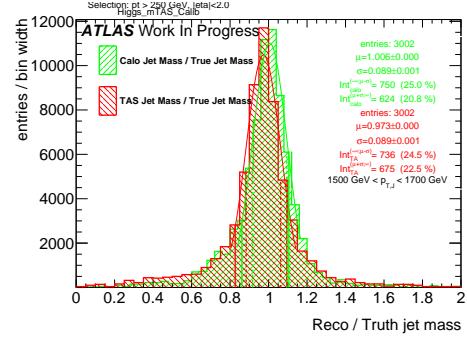


Figure 218: Left integral normalized,  $\int_0^{0.6}$  of the mass response, vs bin of  $p_T^J$

Figure 219: Response in bin of  $p_T^J$  (indicated on plot)Figure 223: Response in bin of  $p_T^J$  (indicated on plot)Figure 220: Response in bin of  $p_T^J$  (indicated on plot)Figure 224: Response in bin of  $p_T^J$  (indicated on plot)Figure 221: Response in bin of  $p_T^J$  (indicated on plot)Figure 225: Response in bin of  $p_T^J$  (indicated on plot)Figure 222: Response in bin of  $p_T^J$  (indicated on plot)Figure 226: Response in bin of  $p_T^J$  (indicated on plot)

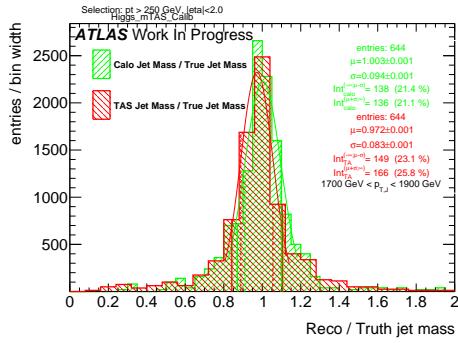


Figure 227: Response in bin of  $p_T^J$  (indicated on plot)

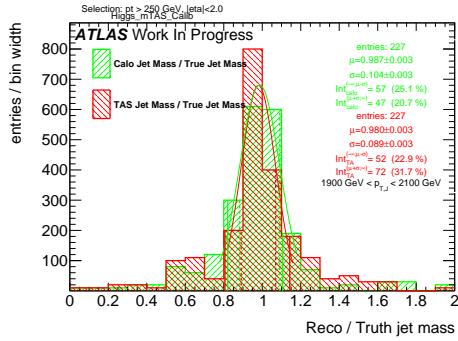


Figure 228: Response in bin of  $p_T^J$  (indicated on plot)

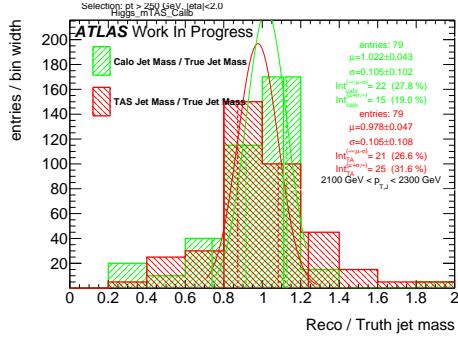


Figure 229: Response in bin of  $p_T^J$  (indicated on plot)

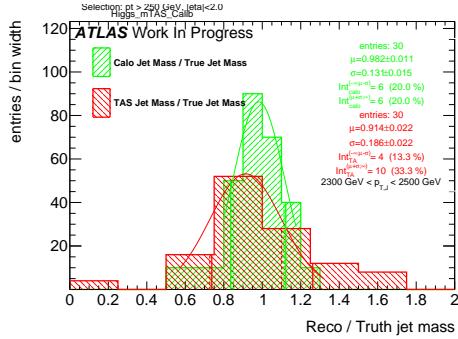
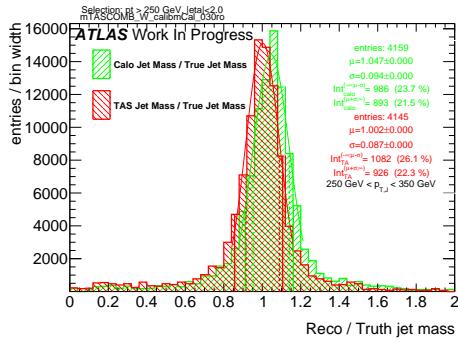
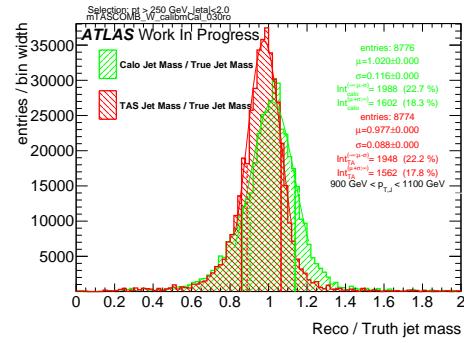
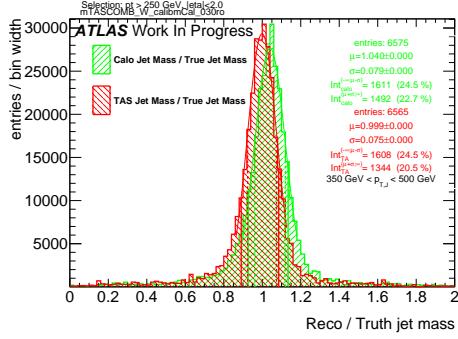
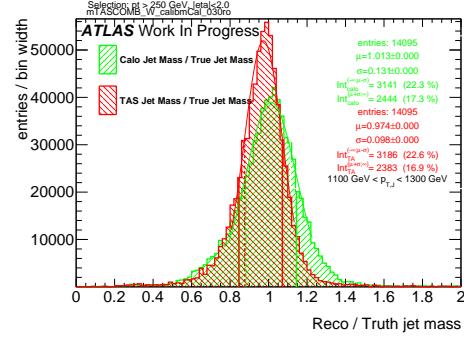
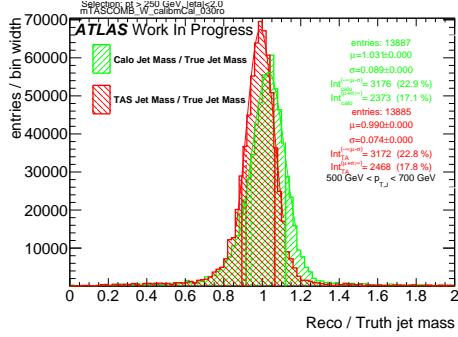
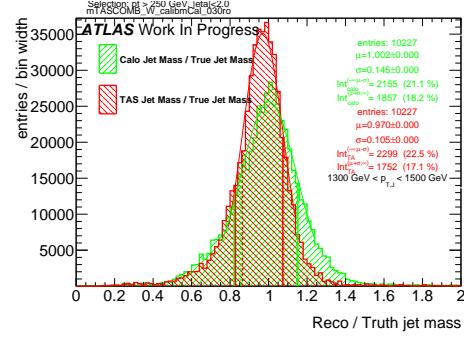
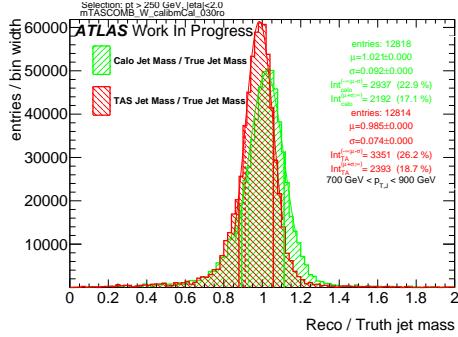
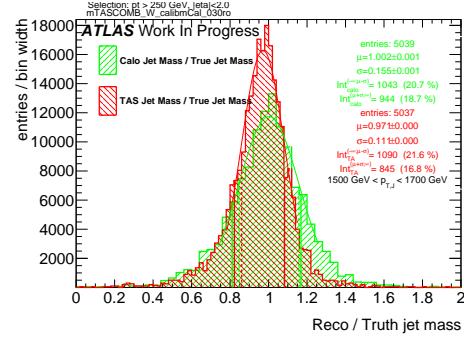
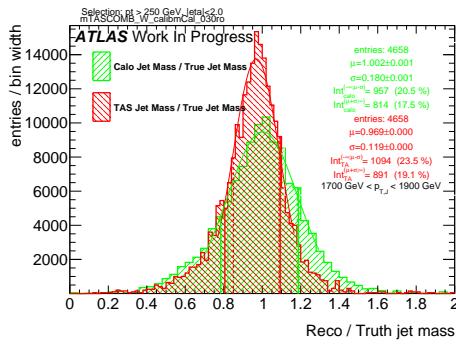
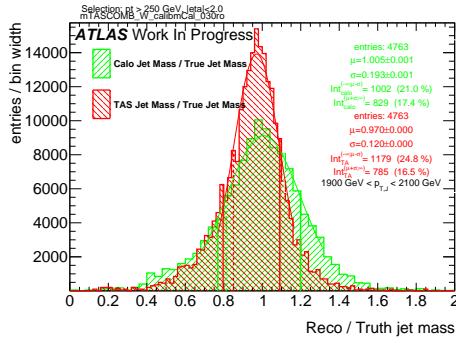
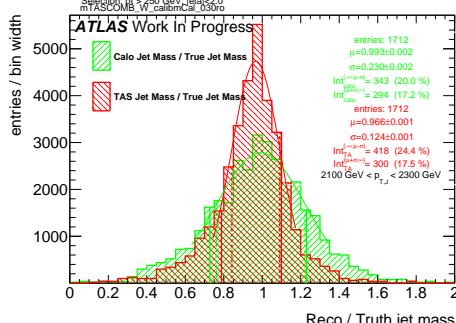
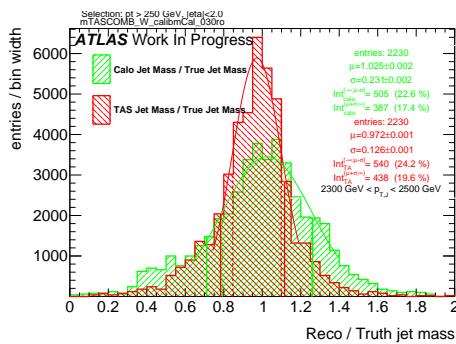


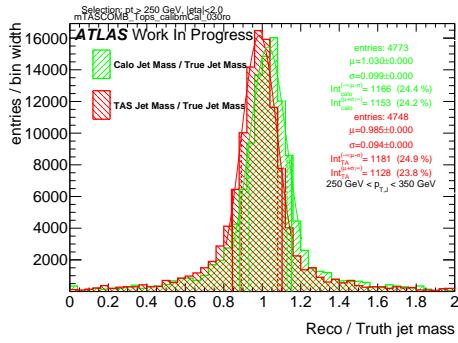
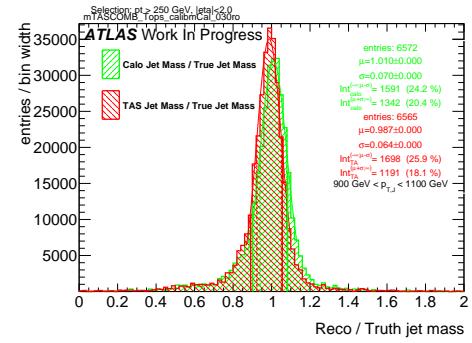
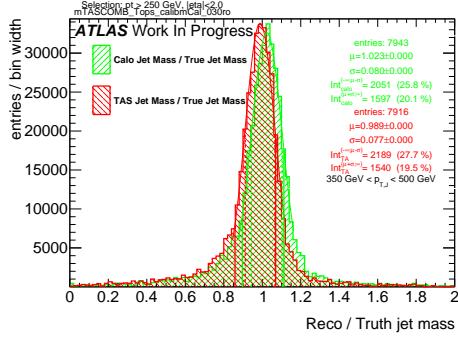
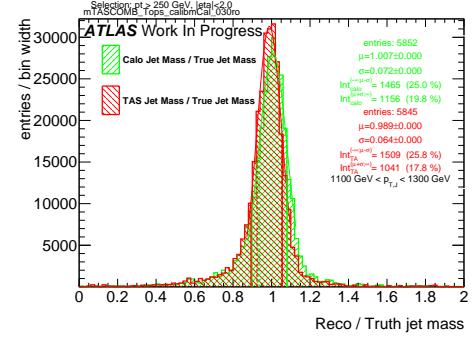
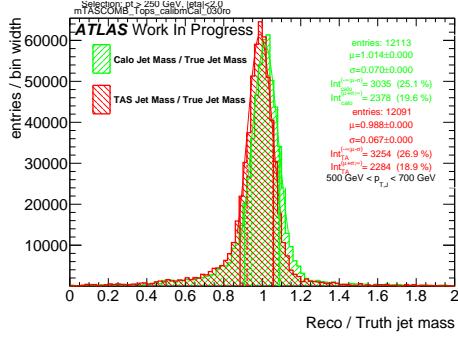
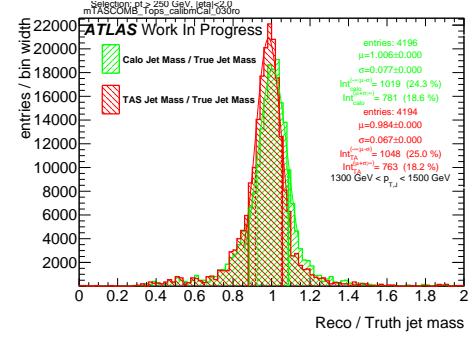
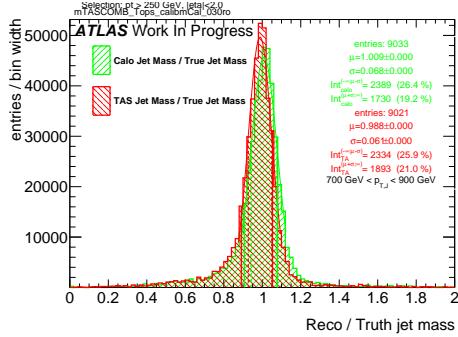
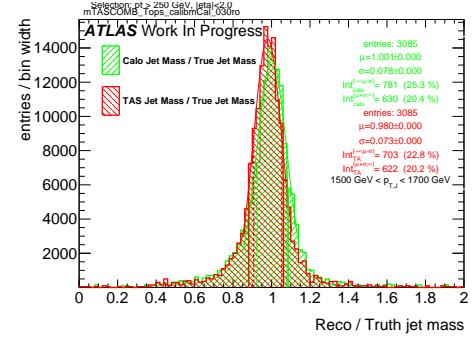
Figure 230: Response in bin of  $p_T^J$  (indicated on plot)

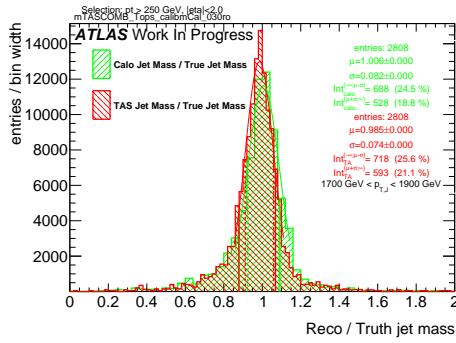
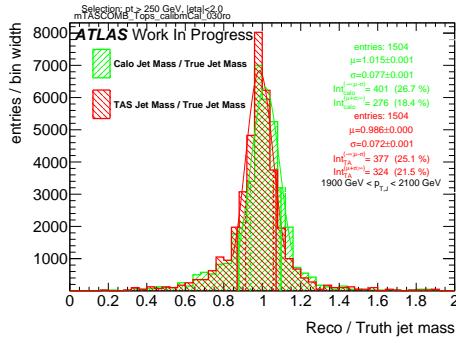
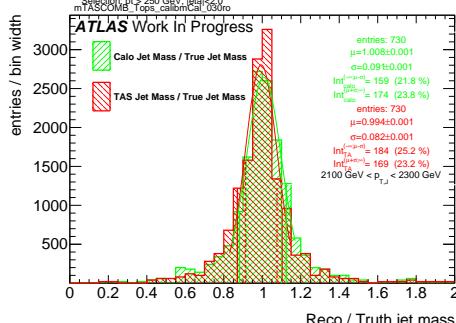
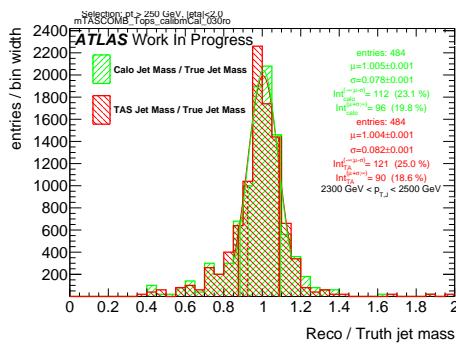
870 **H  $m_{TAS}^{comb}$  response distributions, boosted W/Z**

Figure 231: Response in bin of  $p_T^J$  (indicated on plot)Figure 235: Response in bin of  $p_T^J$  (indicated on plot)Figure 232: Response in bin of  $p_T^J$  (indicated on plot)Figure 236: Response in bin of  $p_T^J$  (indicated on plot)Figure 233: Response in bin of  $p_T^J$  (indicated on plot)Figure 237: Response in bin of  $p_T^J$  (indicated on plot)Figure 234: Response in bin of  $p_T^J$  (indicated on plot)Figure 238: Response in bin of  $p_T^J$  (indicated on plot)

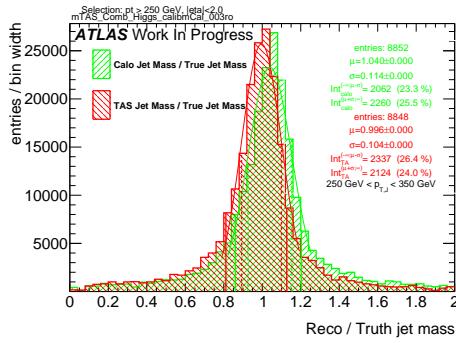
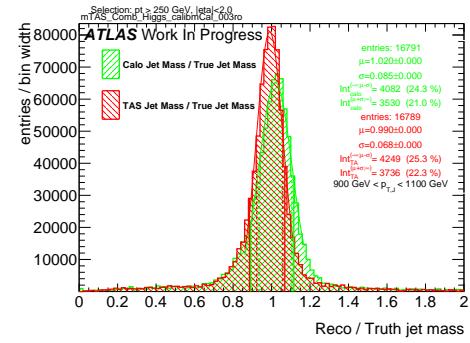
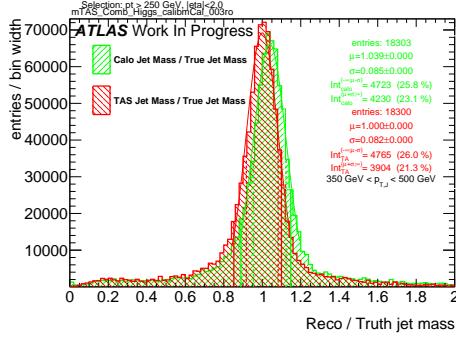
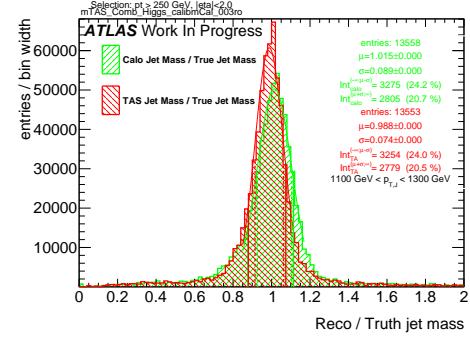
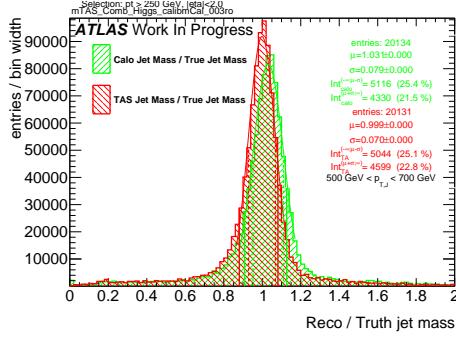
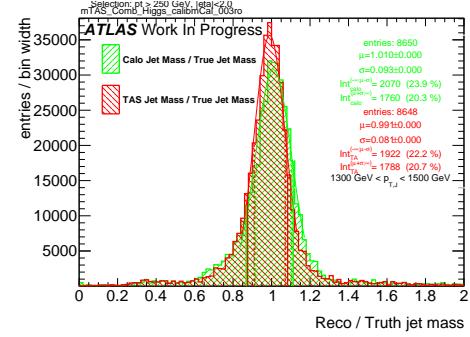
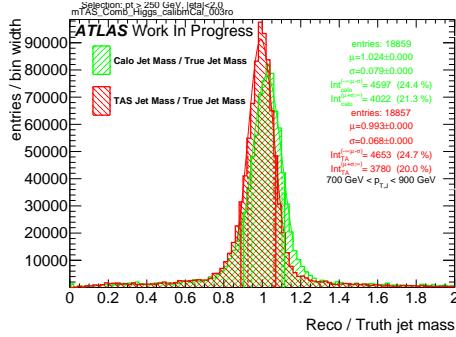
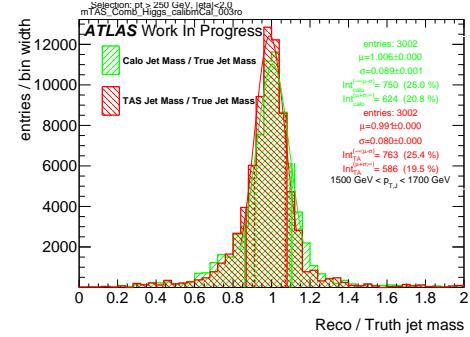
Figure 239: Response in bin of  $p_T^J$  (indicated on plot)Figure 240: Response in bin of  $p_T^J$  (indicated on plot)Figure 241: Response in bin of  $p_T^J$  (indicated on plot)Figure 242: Response in bin of  $p_T^J$  (indicated on plot)

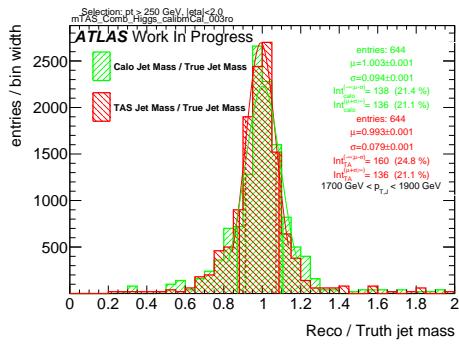
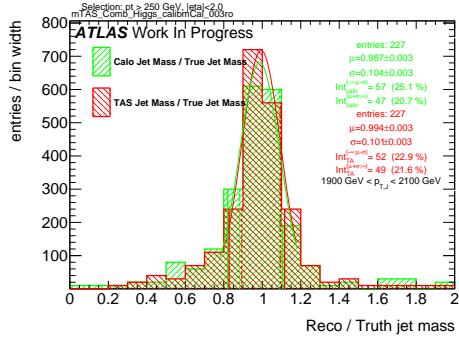
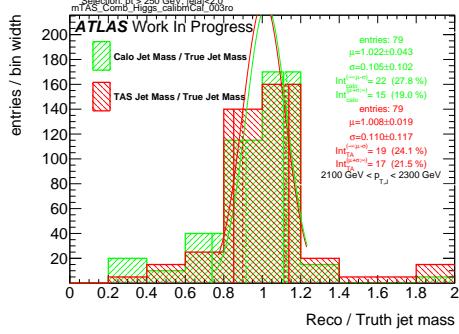
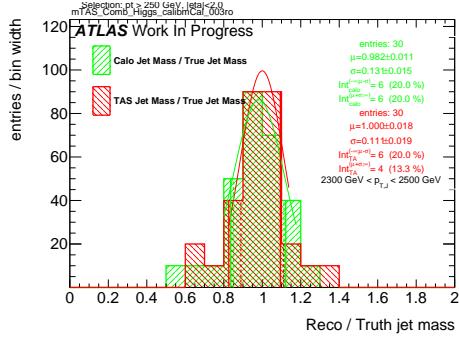
<sup>871</sup> I  $m_{TAS}^{comb}$  response distributions, boosted tops

Figure 243: Response in bin of  $p_T^J$  (indicated on plot)Figure 247: Response in bin of  $p_T^J$  (indicated on plot)Figure 244: Response in bin of  $p_T^J$  (indicated on plot)Figure 248: Response in bin of  $p_T^J$  (indicated on plot)Figure 245: Response in bin of  $p_T^J$  (indicated on plot)Figure 249: Response in bin of  $p_T^J$  (indicated on plot)Figure 246: Response in bin of  $p_T^J$  (indicated on plot)Figure 250: Response in bin of  $p_T^J$  (indicated on plot)

Figure 251: Response in bin of  $p_T^J$  (indicated on plot)Figure 252: Response in bin of  $p_T^J$  (indicated on plot)Figure 253: Response in bin of  $p_T^J$  (indicated on plot)Figure 254: Response in bin of  $p_T^J$  (indicated on plot)

<sup>872</sup> **J  $m_{TAS}^{comb}$  response distributions, Higgs**

Figure 255: Response in bin of  $p_T^J$  (indicated on plot)Figure 259: Response in bin of  $p_T^J$  (indicated on plot)Figure 256: Response in bin of  $p_T^J$  (indicated on plot)Figure 260: Response in bin of  $p_T^J$  (indicated on plot)Figure 257: Response in bin of  $p_T^J$  (indicated on plot)Figure 261: Response in bin of  $p_T^J$  (indicated on plot)Figure 258: Response in bin of  $p_T^J$  (indicated on plot)Figure 262: Response in bin of  $p_T^J$  (indicated on plot)

Figure 263: Response in bin of  $p_T^J$  (indicated on plot)Figure 264: Response in bin of  $p_T^J$  (indicated on plot)Figure 265: Response in bin of  $p_T^J$  (indicated on plot)Figure 266: Response in bin of  $p_T^J$  (indicated on plot)

<sup>873</sup> ROCs for the Best Variables

<sup>874</sup> **K W boson Tagging**

<sup>875</sup> **L Higgs Boson Tagging**

<sup>876</sup> **M Top Quark Tagging**

<sup>877</sup> Signal and Background Distributions

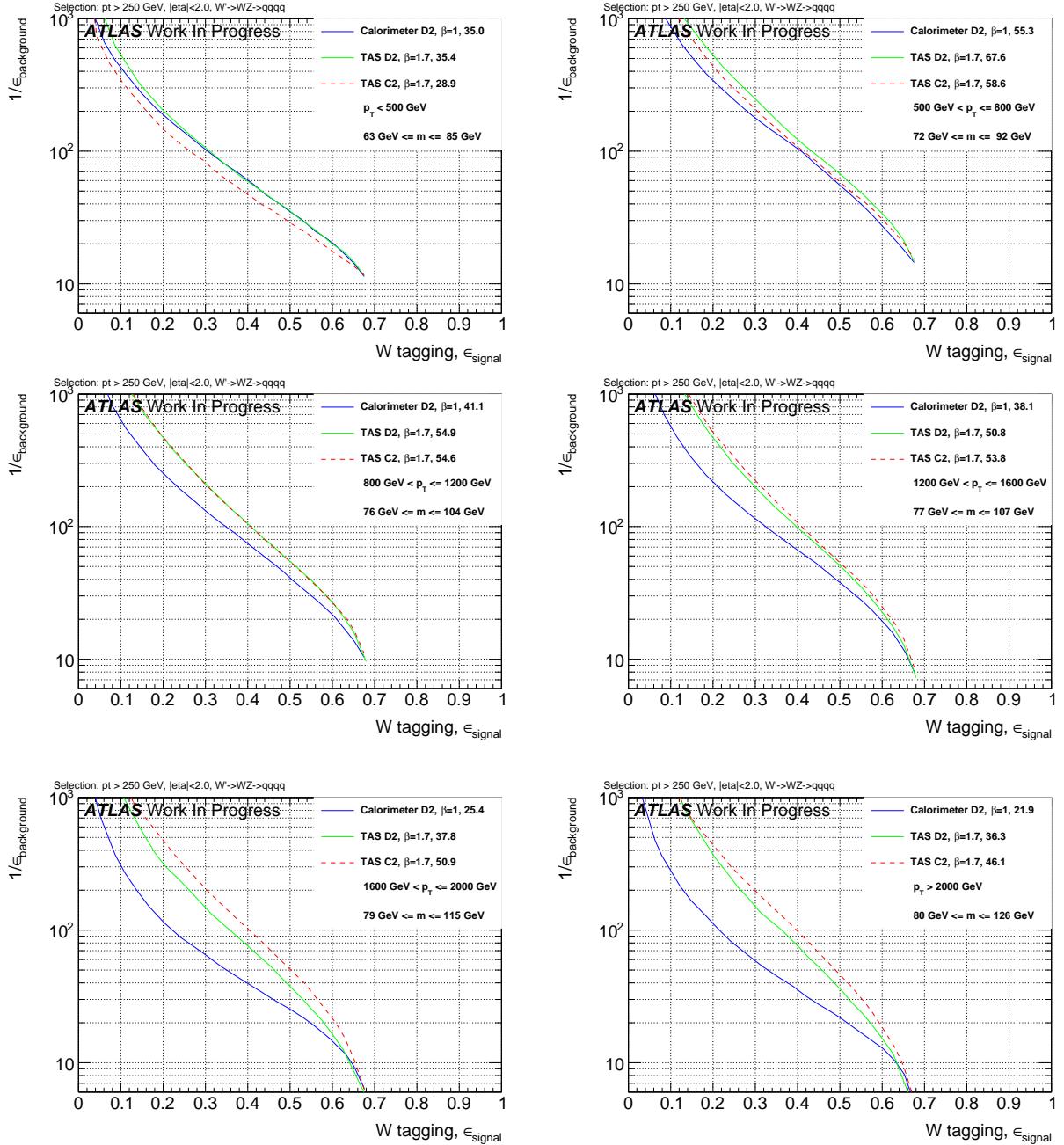


Figure 267: ROCs showing QCD rejection against  $W$  jet efficiency for  $D2_{\text{TAS}}^{(\beta=1.7)}$  &  $C2_{\text{TAS}}^{(\beta=1.7)}$  compared to  $D2_{\text{calo}}^{(\beta=1)}$ .

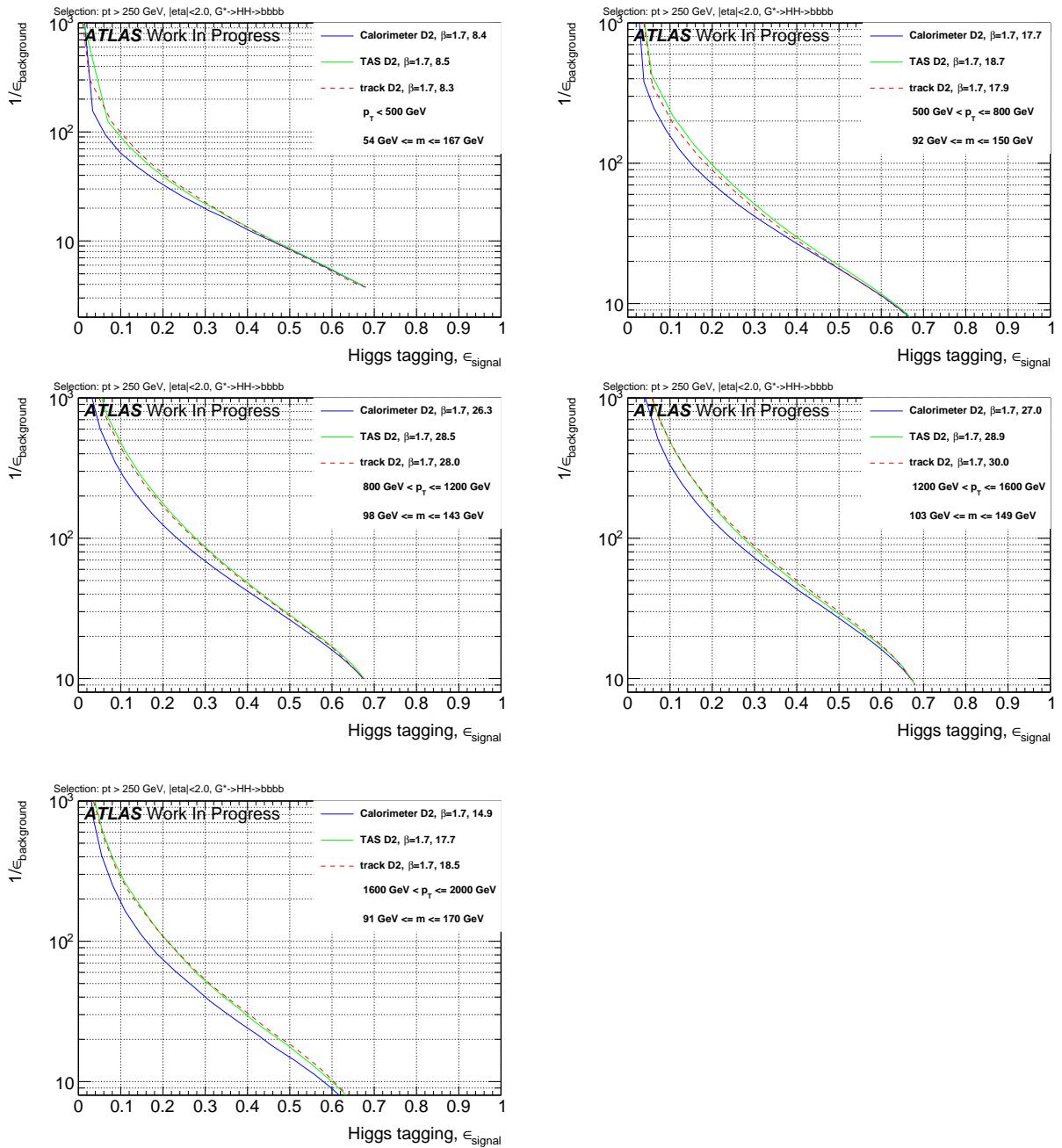


Figure 268: ROCs showing QCD rejection against Higgs jet efficiency for  $D2_{\text{TAS}}^{(\beta=1.7)}$  &  $D2_{\text{track}}^{(\beta=1.7)}$  compared to  $D2_{\text{calo}}^{(\beta=1)}$ .

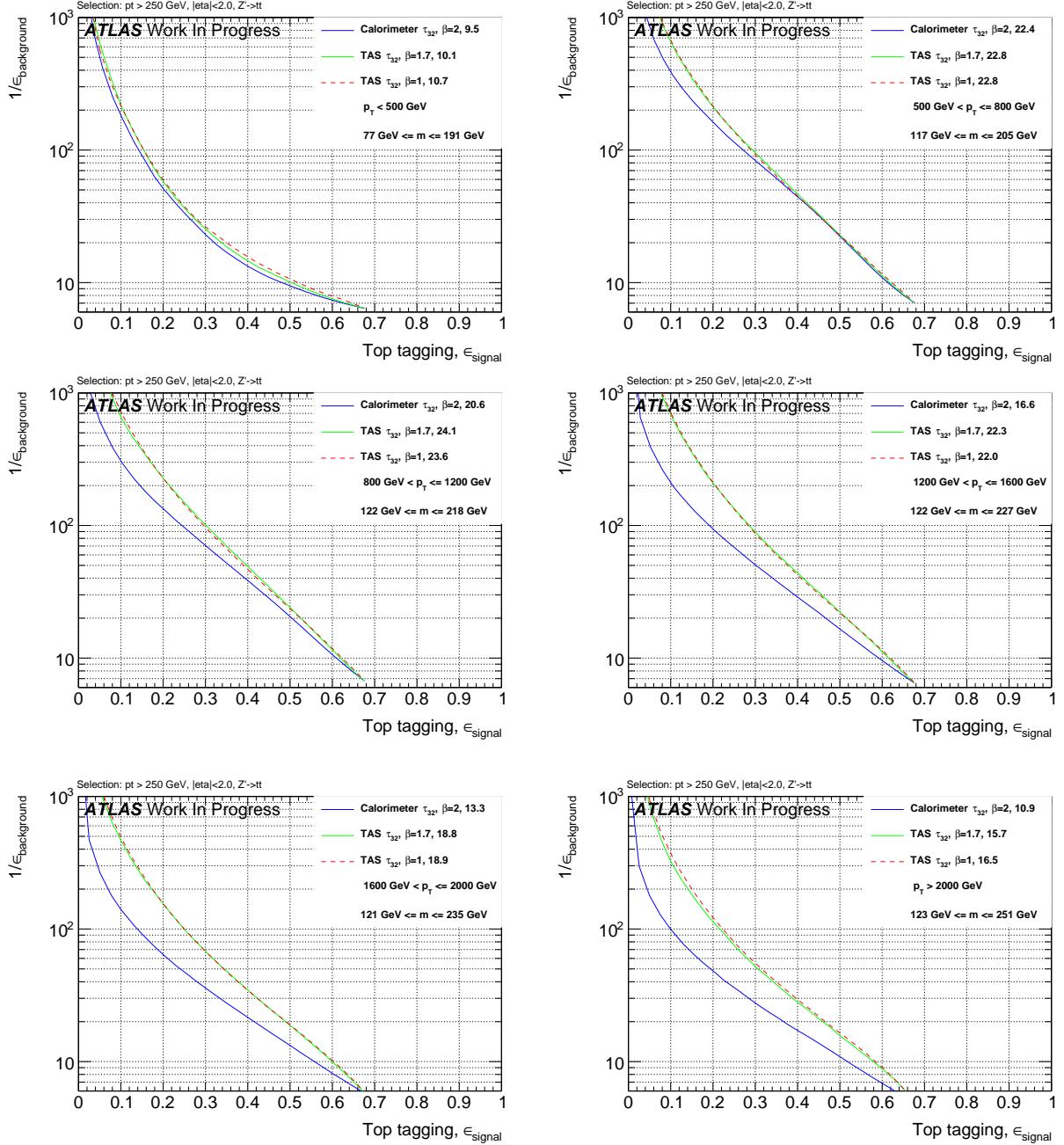
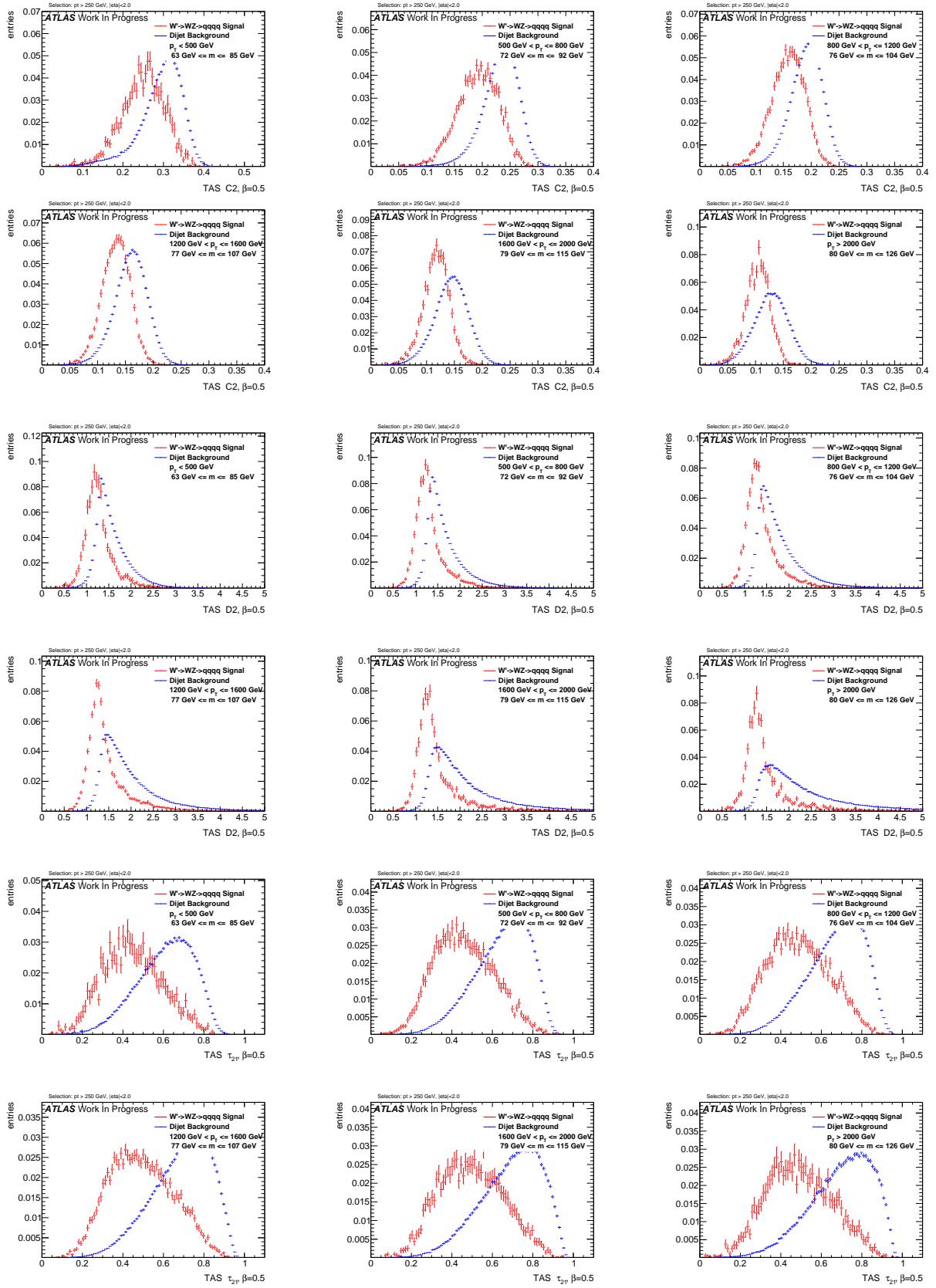
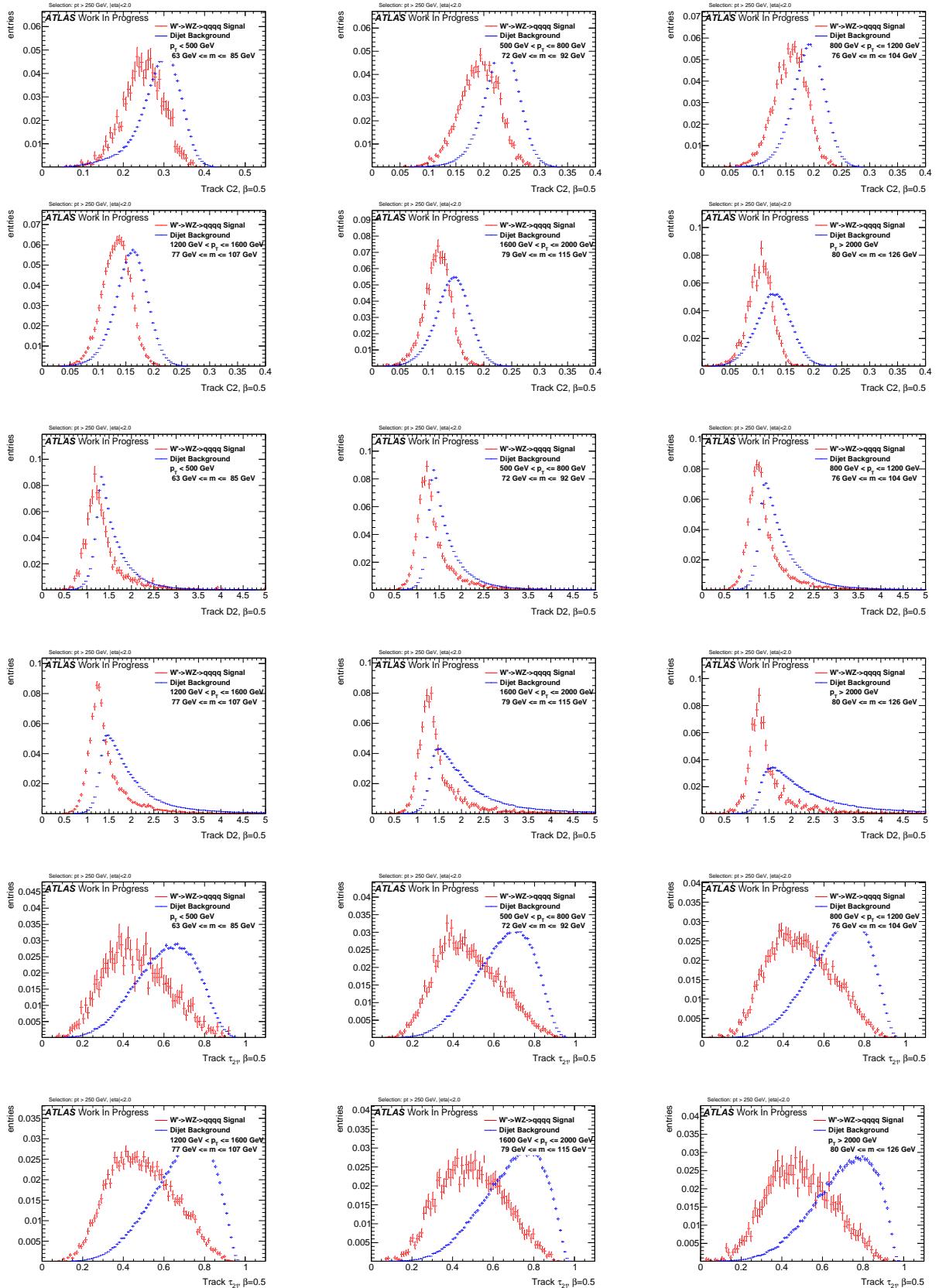
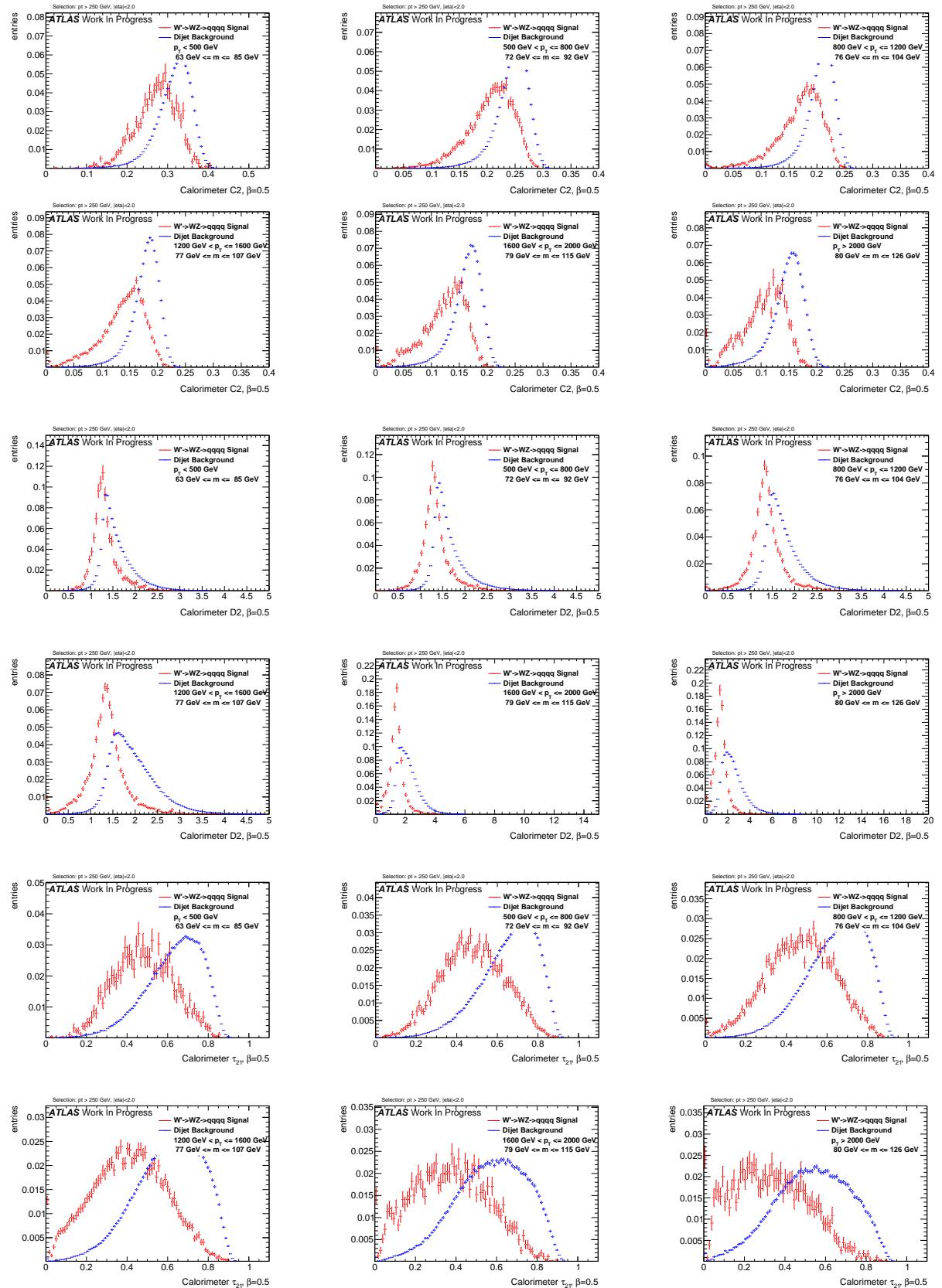
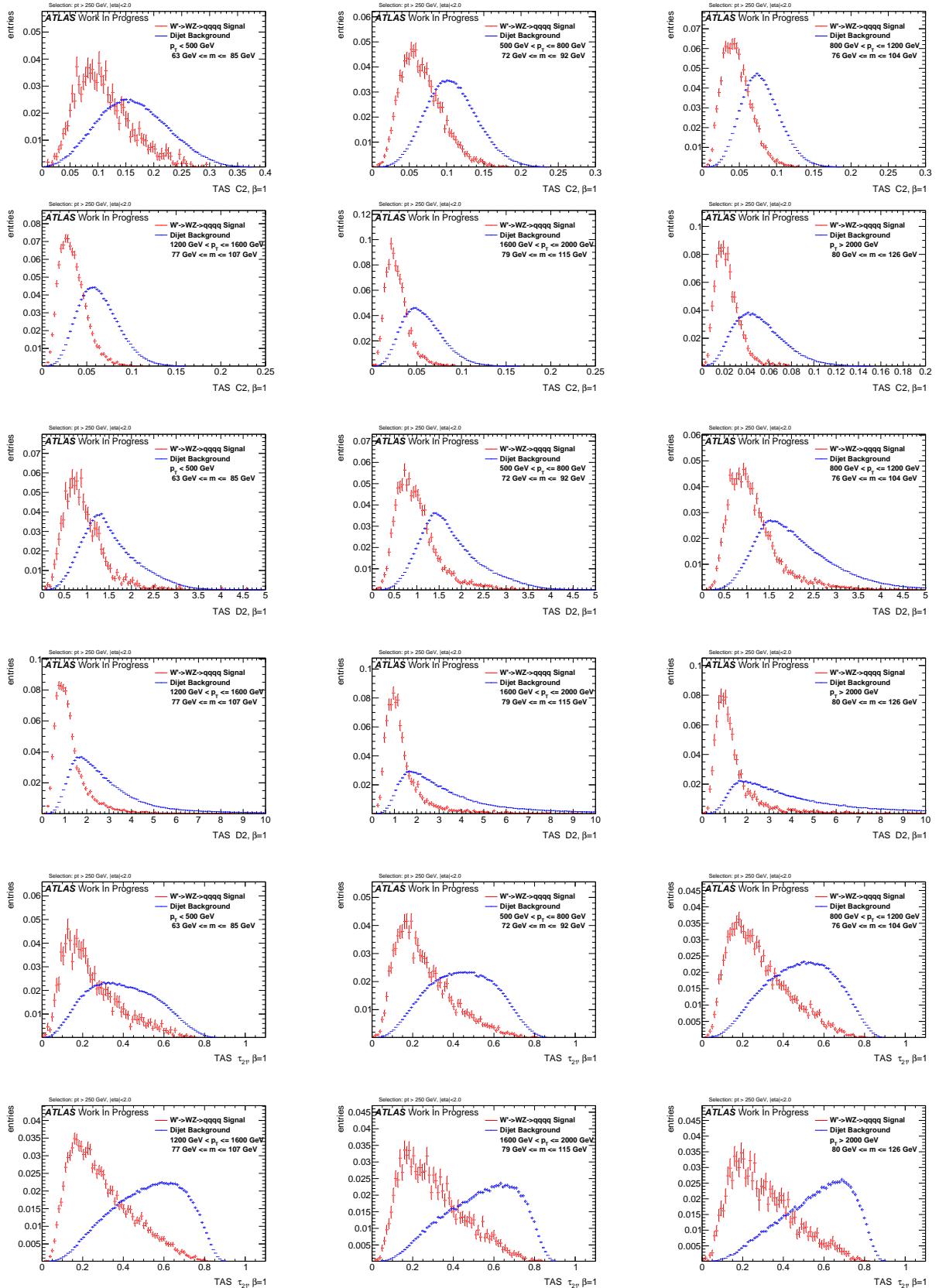


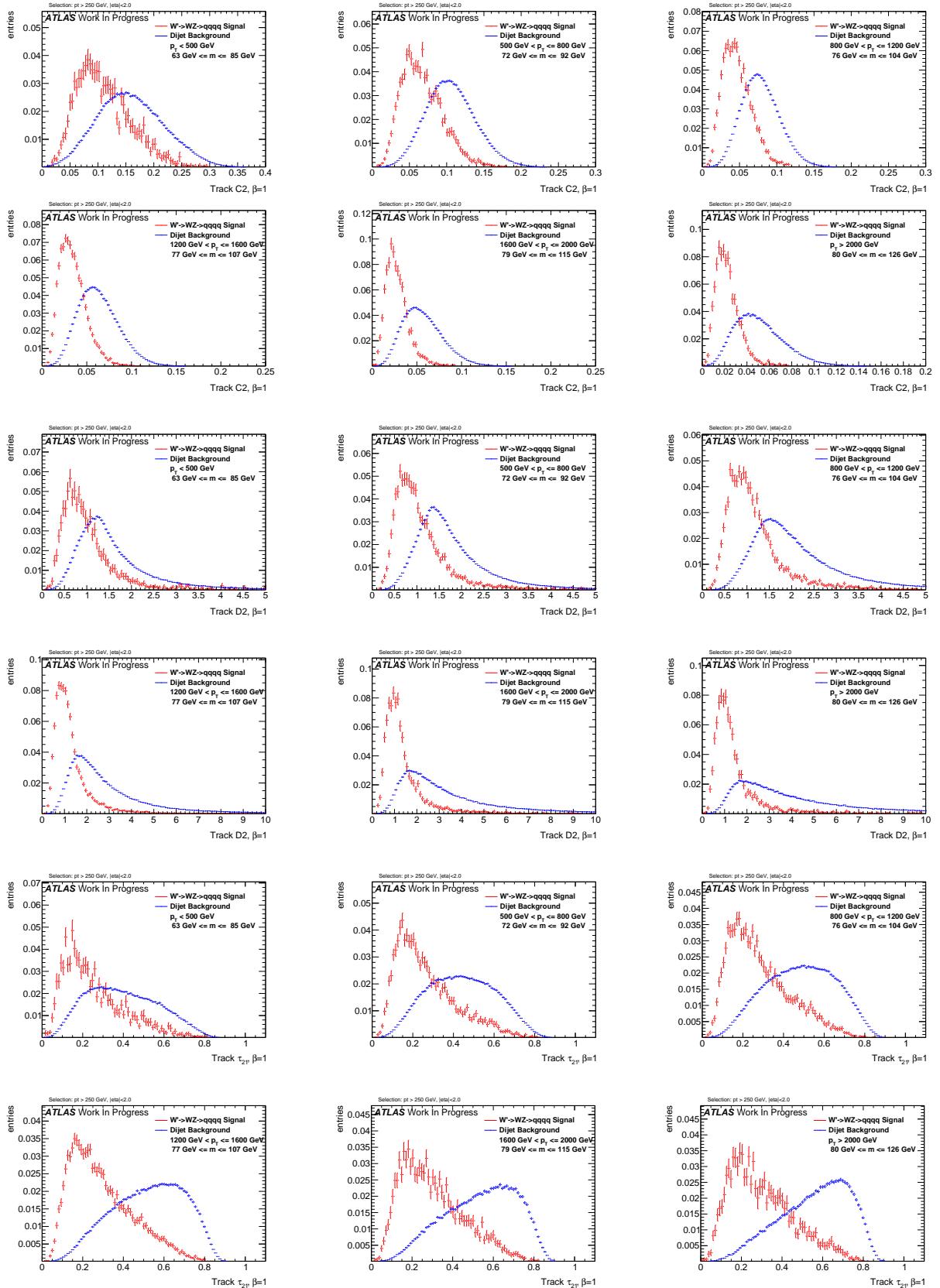
Figure 269: ROCs showing QCD rejection against Top jet efficiency for  $\tau_{32, \text{TAS}}^{(\beta=1)}$  &  $\tau_{32, \text{TAS}}^{(\beta=1.7)}$  compared to  $\tau_{32, \text{TAS}}^{(\beta=2)}$

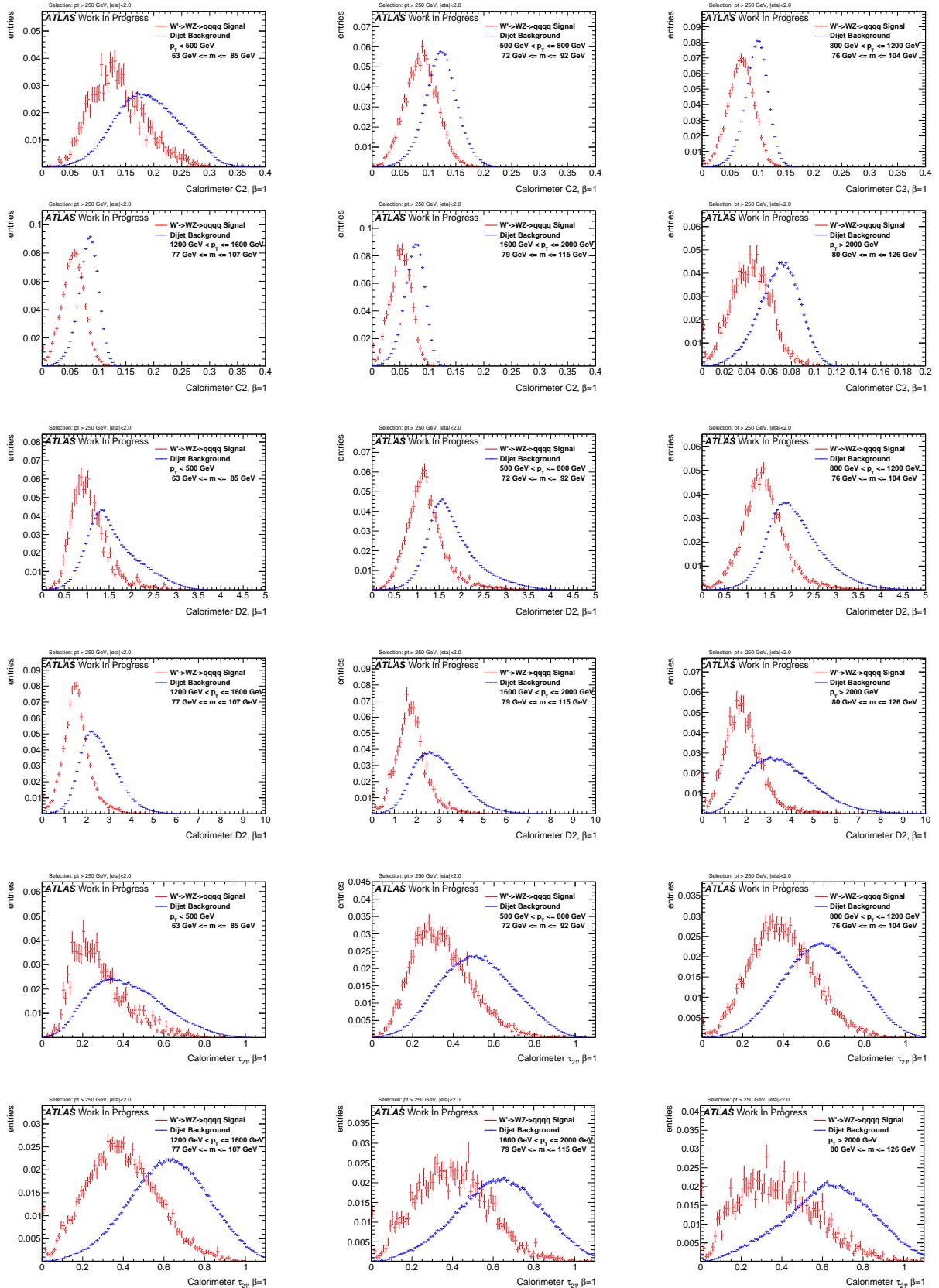
Figure 270: Distributions for  $W$  boson tagging using TAS  $\beta = 0.5$ . C2, D2,  $\tau_{21}$  top down.

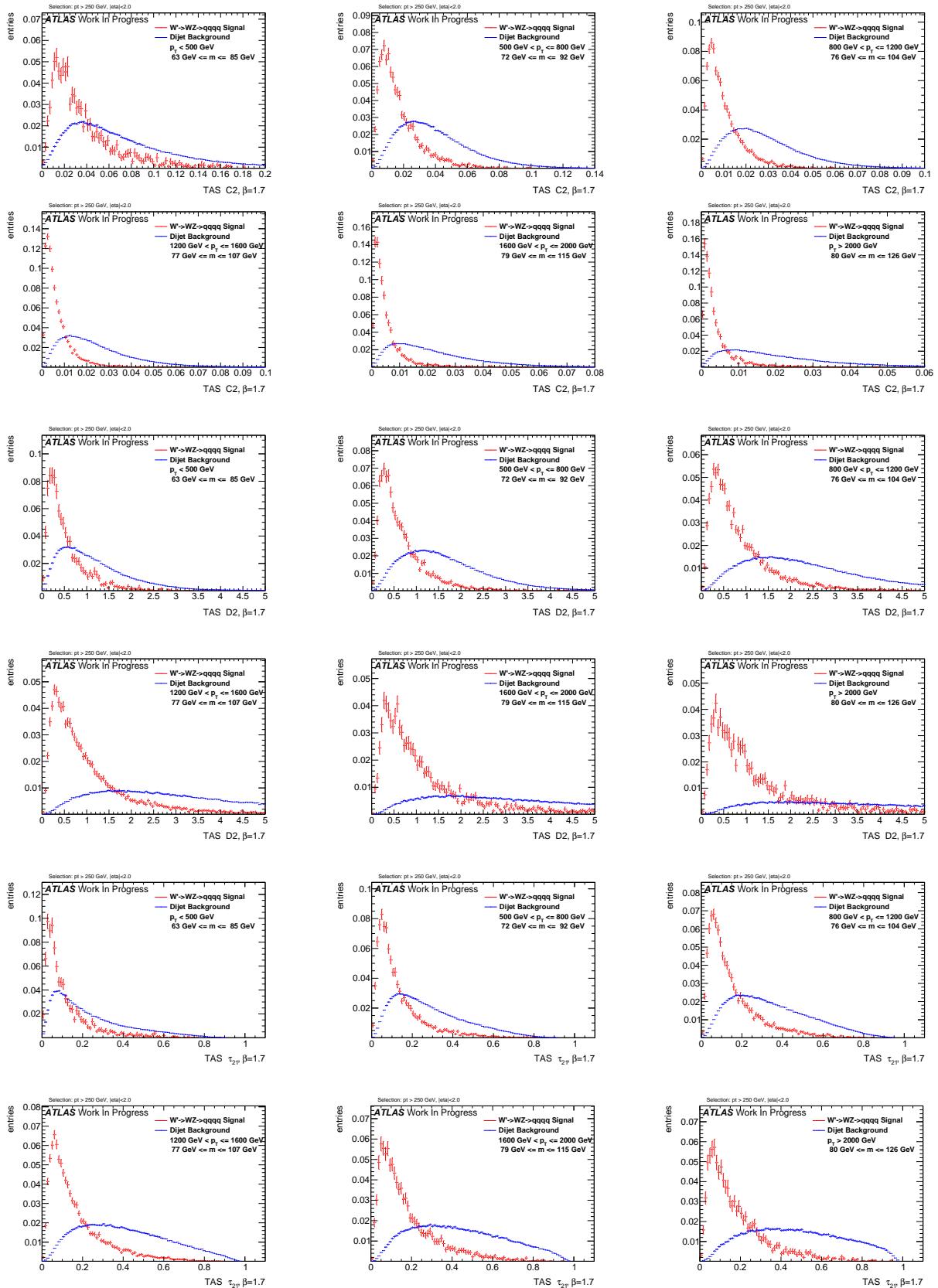
Figure 271: Distributions for  $W$  boson tagging using tracks  $\beta = 0.5$ . C2, D2,  $\tau_{21}$  top down.

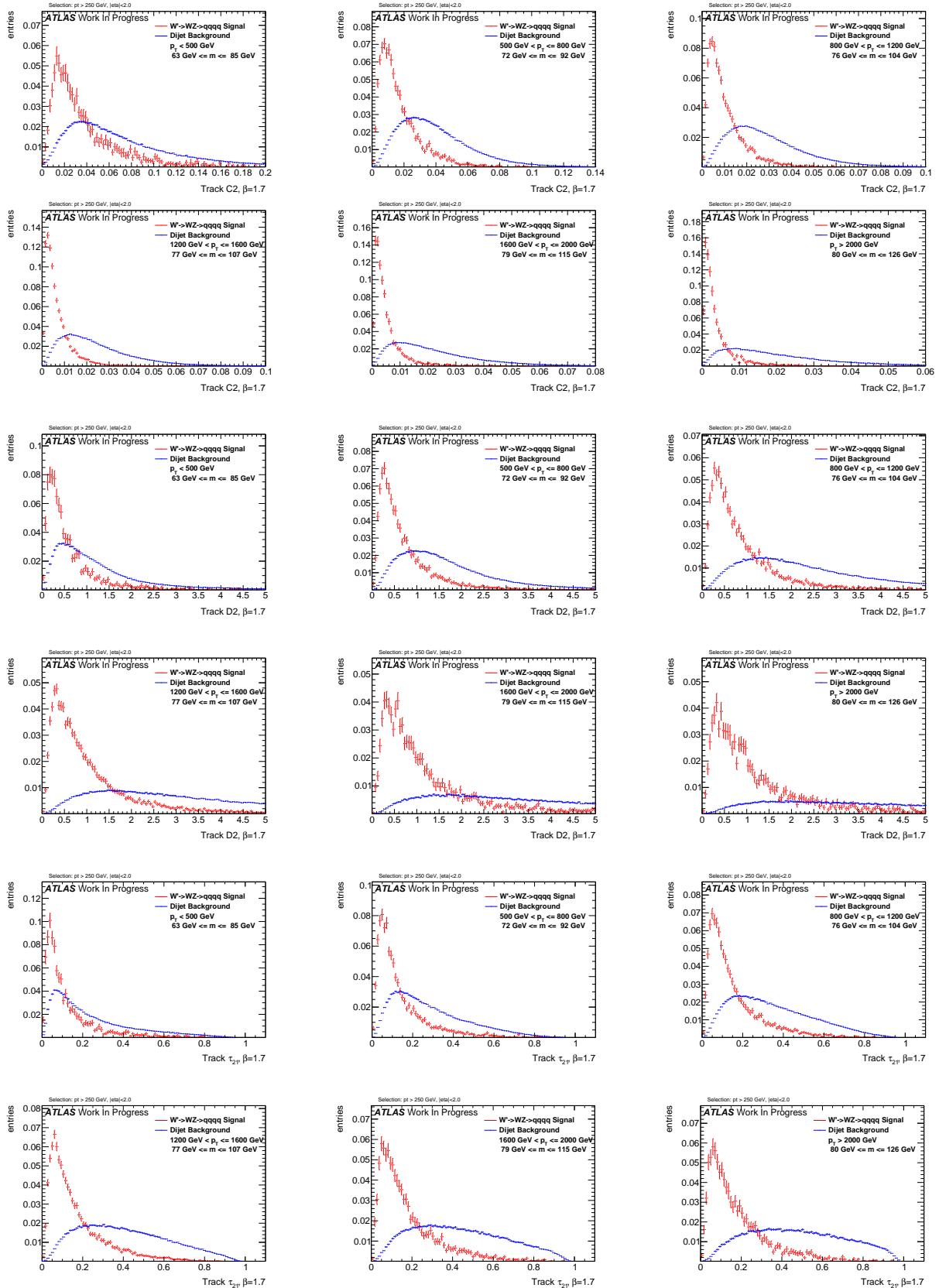
Figure 272: Distributions for  $W$  boson tagging using calorimeter clusters  $\beta = 0.5$ . C2, D2,  $\tau_{21}$  top down.

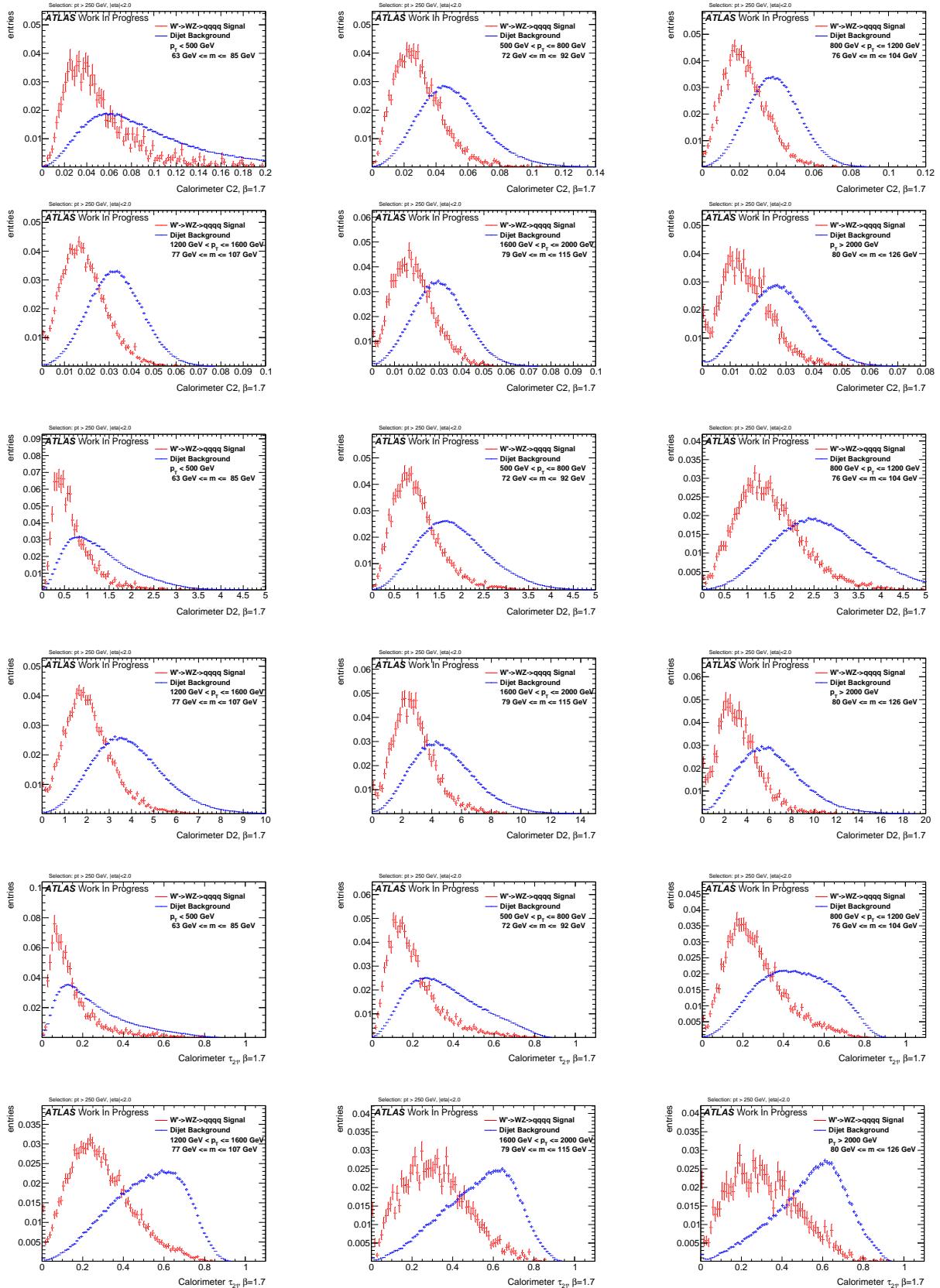
Figure 273: Distributions for  $W$  boson tagging using TAS  $\beta = 1$ . C2, D2,  $\tau_{21}$  top down.

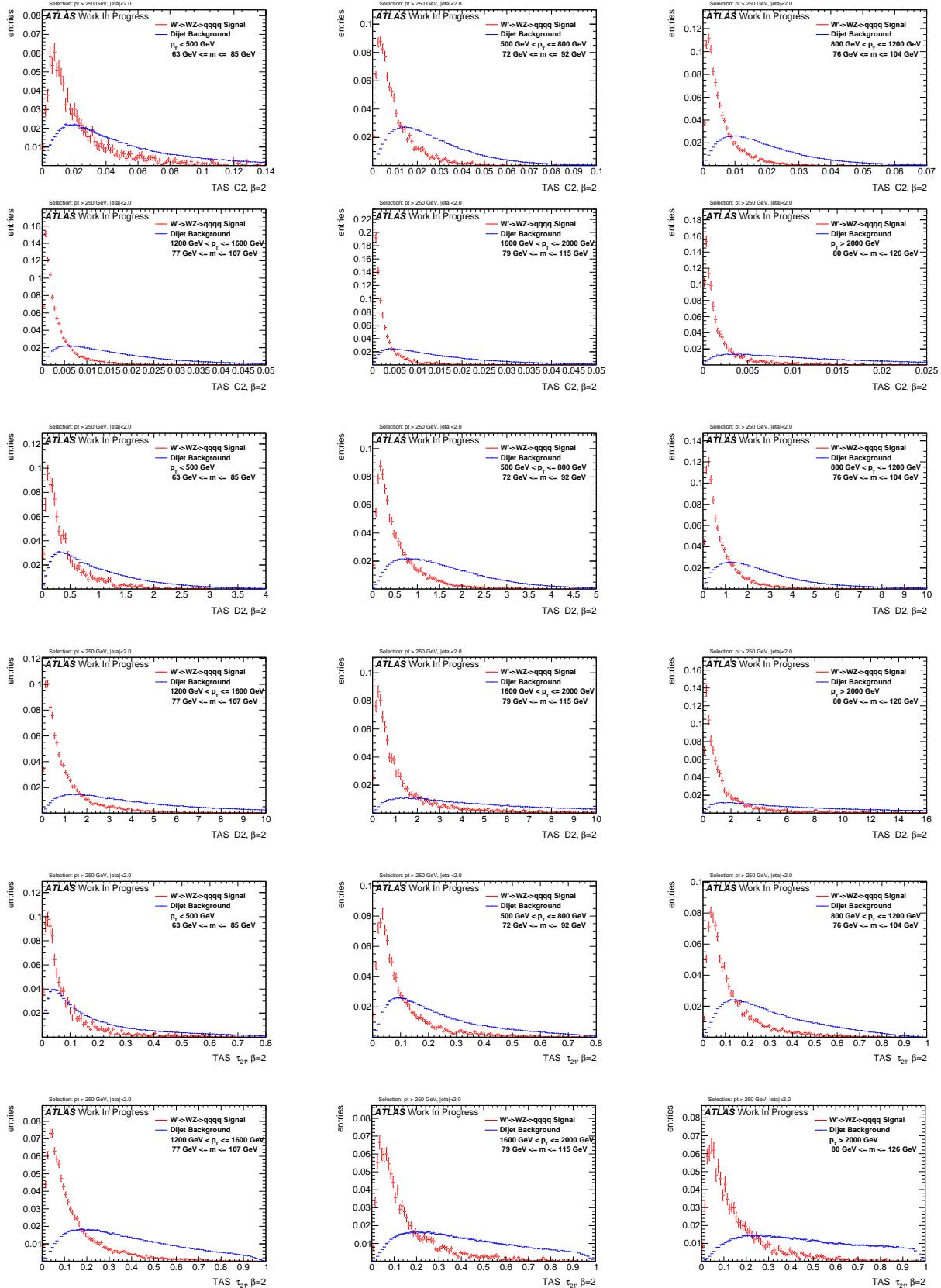
Figure 274: Distributions for W boson tagging using tracks  $\beta = 1$ . C2, D2,  $\tau_{21}$  top down.

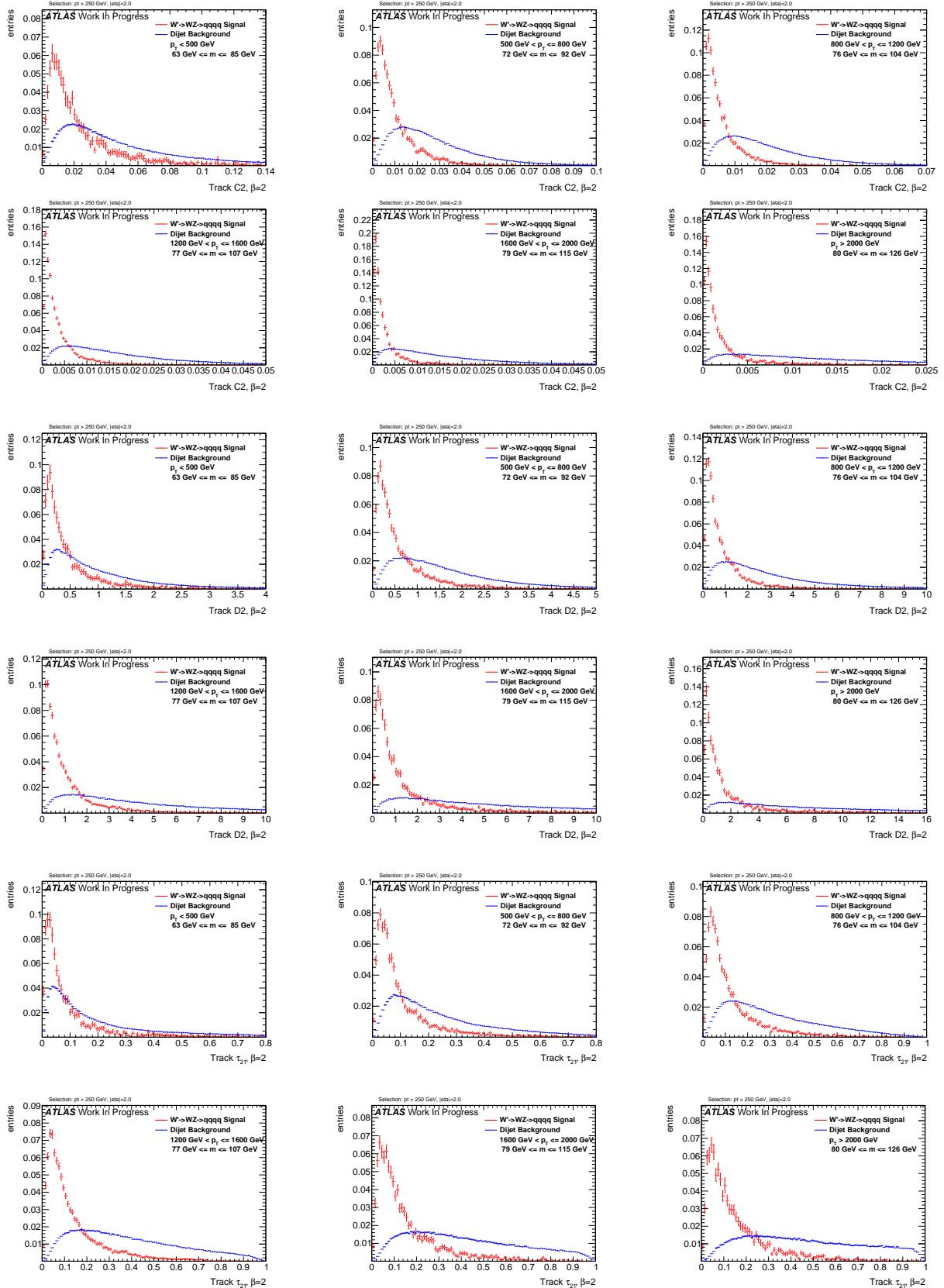
Figure 275: Distributions for  $W$  boson tagging using calorimeter clusters  $\beta = 1$ . C2, D2,  $\tau_{21}$  top down.

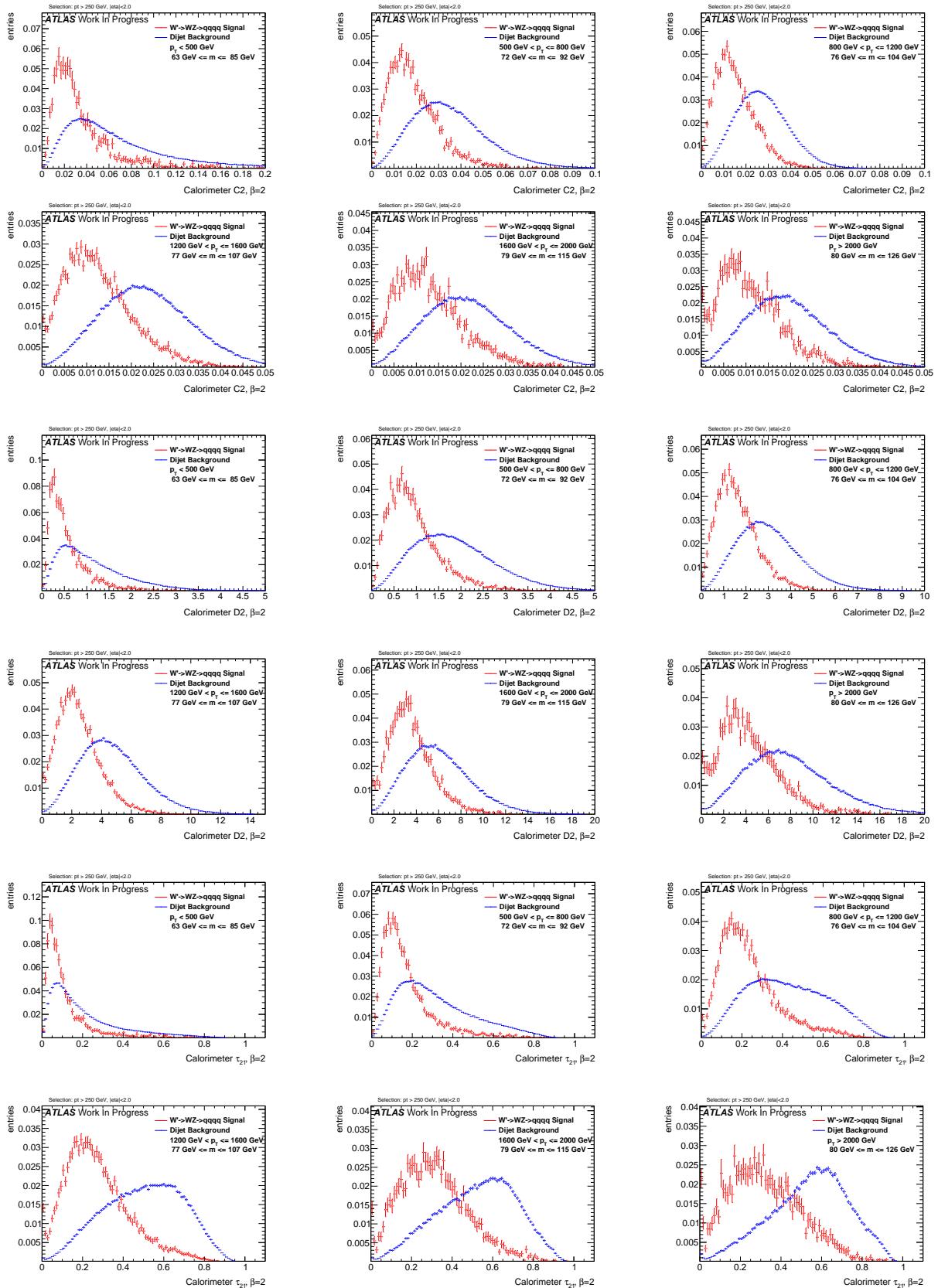
Figure 276: Distributions for  $W$  boson tagging using TAS  $\beta = 1.7$ . C2, D2,  $\tau_{21}$  top down.

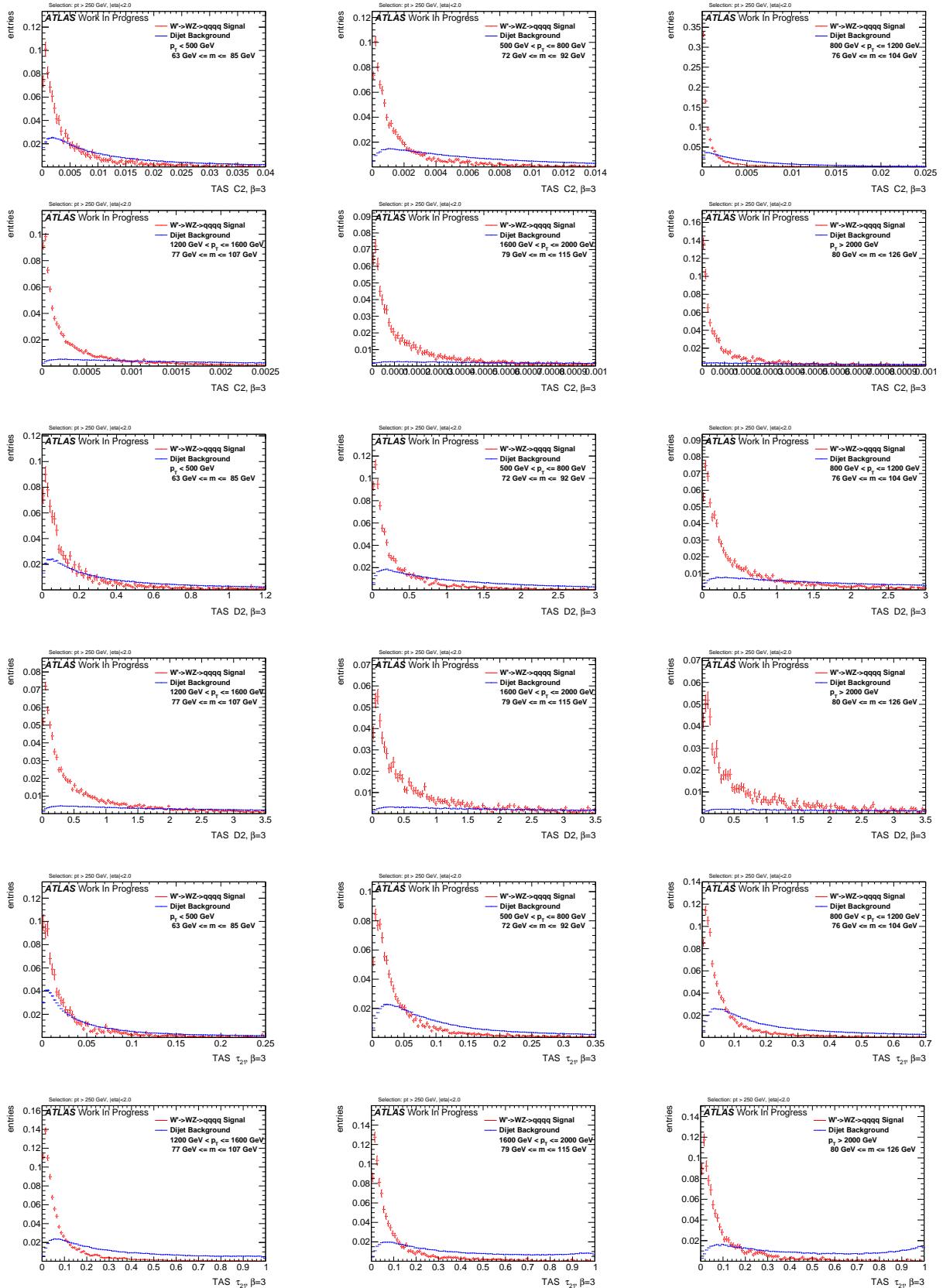
Figure 277: Distributions for  $W$  boson tagging using tracks  $\beta = 1.7$ . C2, D2,  $\tau_{21}$  top down.

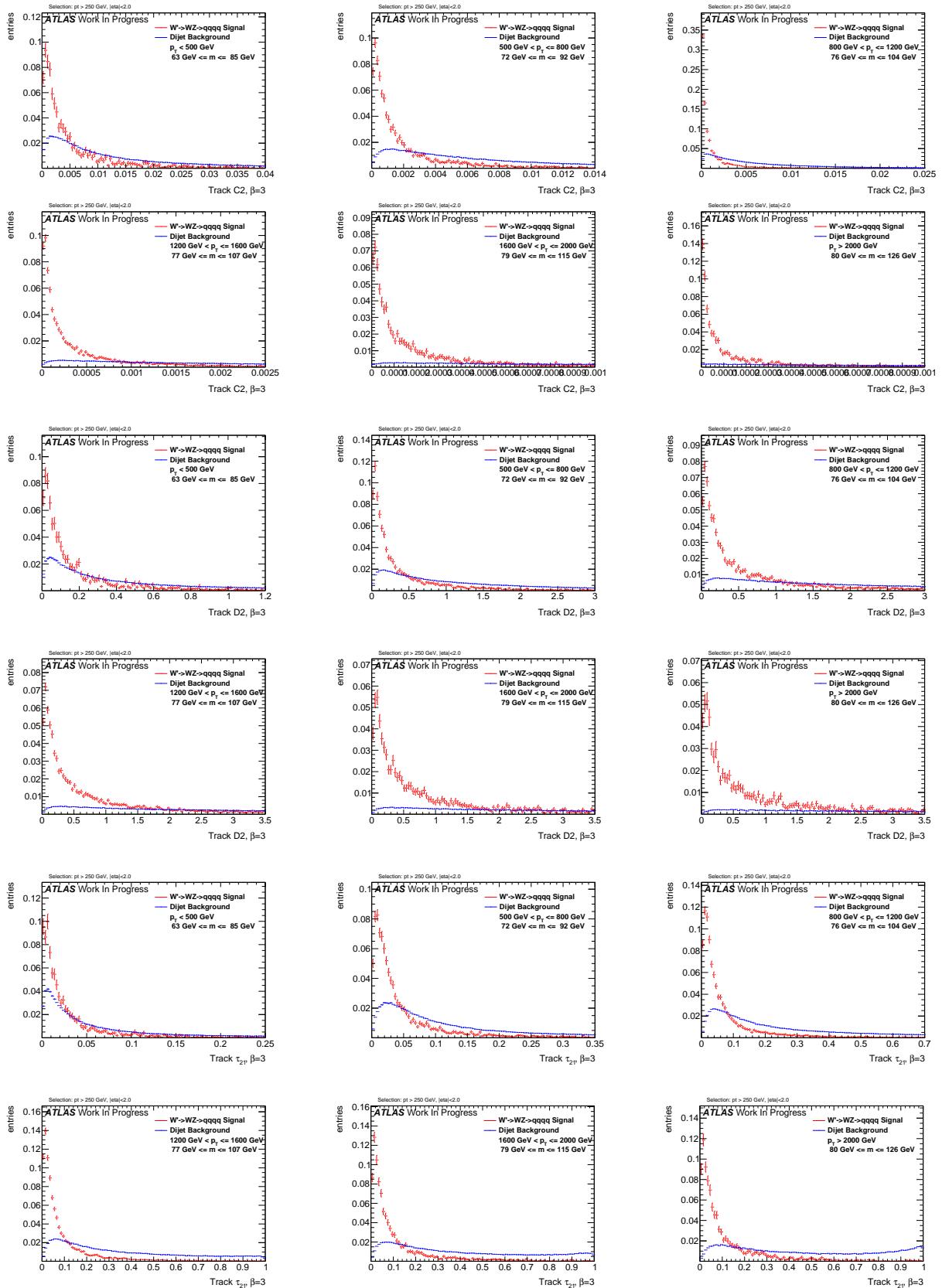
Figure 278: Distributions for  $W$  boson tagging using calorimeter clusters  $\beta = 1.7$ . C2, D2,  $\tau_{21}$  top down.

Figure 279: Distributions for  $W$  boson tagging using TAS  $\beta = 2$ . C<sub>2</sub>, D<sub>2</sub>,  $\tau_{21}$  top down.

Figure 280: Distributions for W boson tagging using tracks  $\beta = 2$ . C<sub>2</sub>, D<sub>2</sub>,  $\tau_{21}$  top down.

Figure 281: Distributions for  $W$  boson tagging using calorimeter clusters  $\beta = 2$ . C2, D2,  $\tau_{21}$  top down.

Figure 282: Distributions for  $W$  boson tagging using TAS  $\beta = 3$ . C2, D2,  $\tau_{21}$  top down.

Figure 283: Distributions for  $W$  boson tagging using tracks  $\beta = 3$ . C2, D2,  $\tau_{21}$  top down.

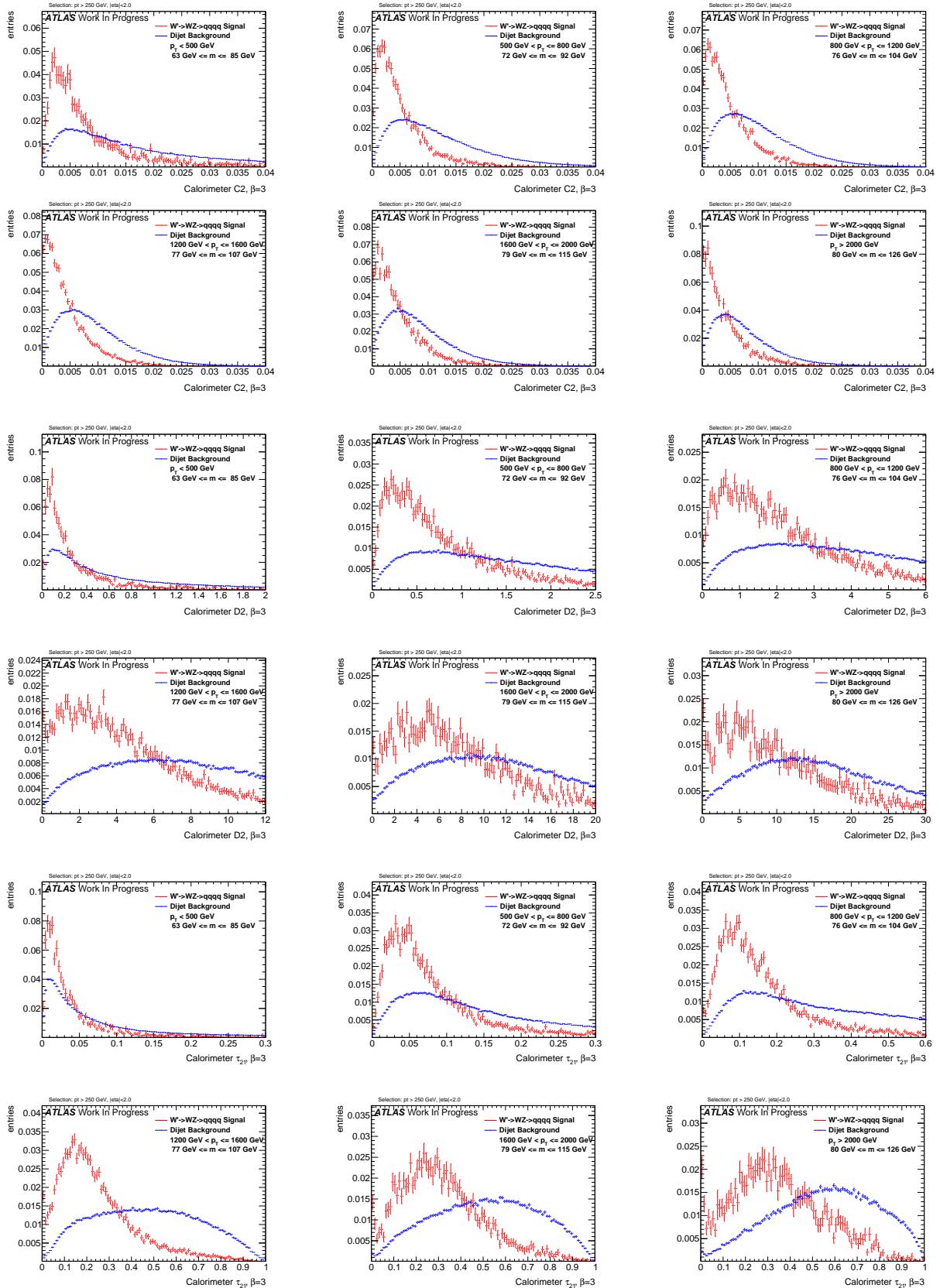
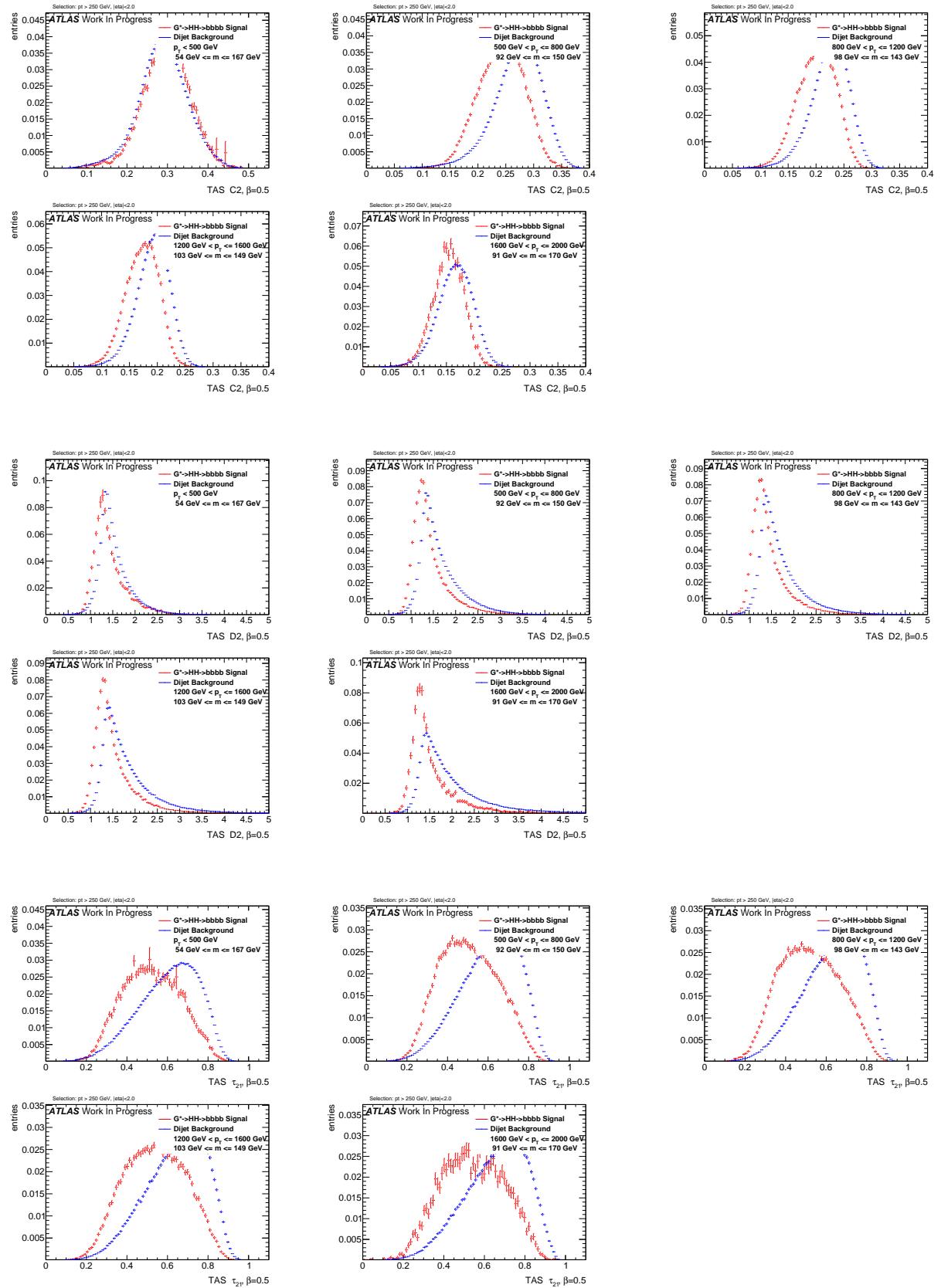
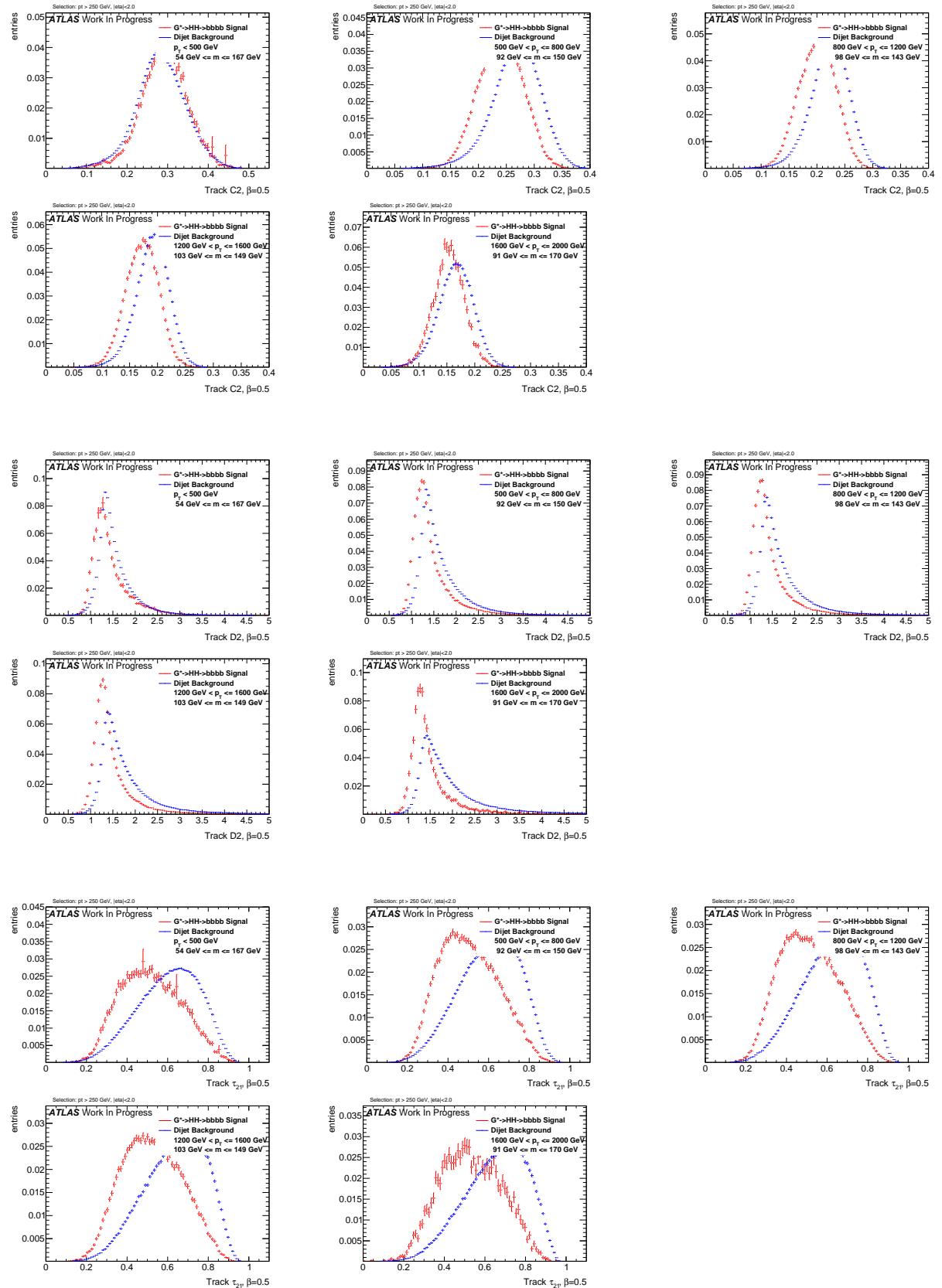
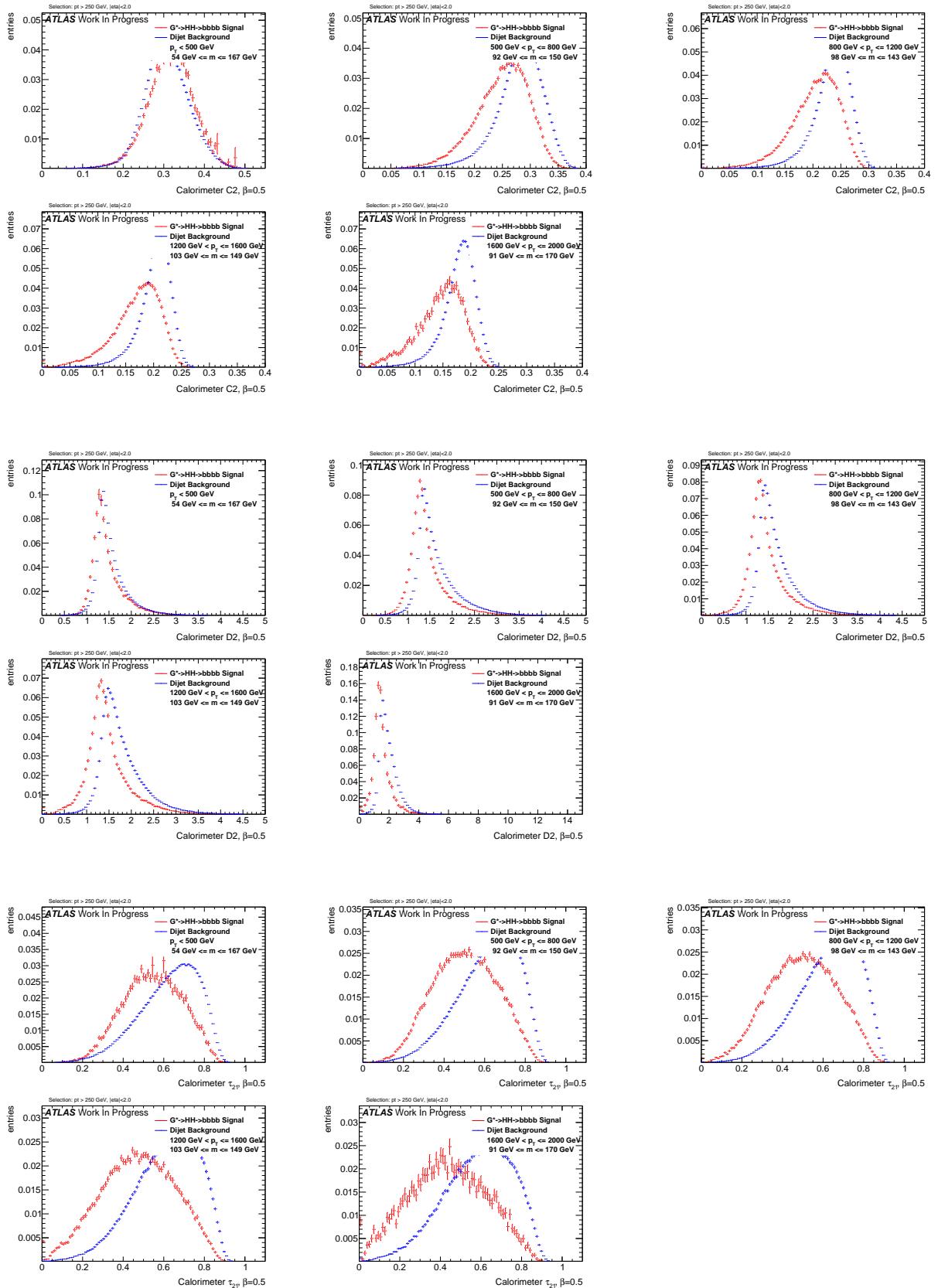


Figure 284: Distributions for  $W$  boson tagging using calorimeter clusters  $\beta = 3$ . C2, D2,  $\tau_{21}$  top down.

Figure 285: Distributions for Higgs boson tagging using TAS  $\beta = 0.5$ . C2, D2,  $\tau_{21}$  top down.

Figure 286: Distributions for Higgs boson tagging using tracks  $\beta = 0.5$ . C2, D2,  $\tau_{21}$  top down.

Figure 287: Distributions for Higgs boson tagging using calorimeter clusters  $\beta = 0.5$ . C2, D2,  $\tau_{21}$  top down.

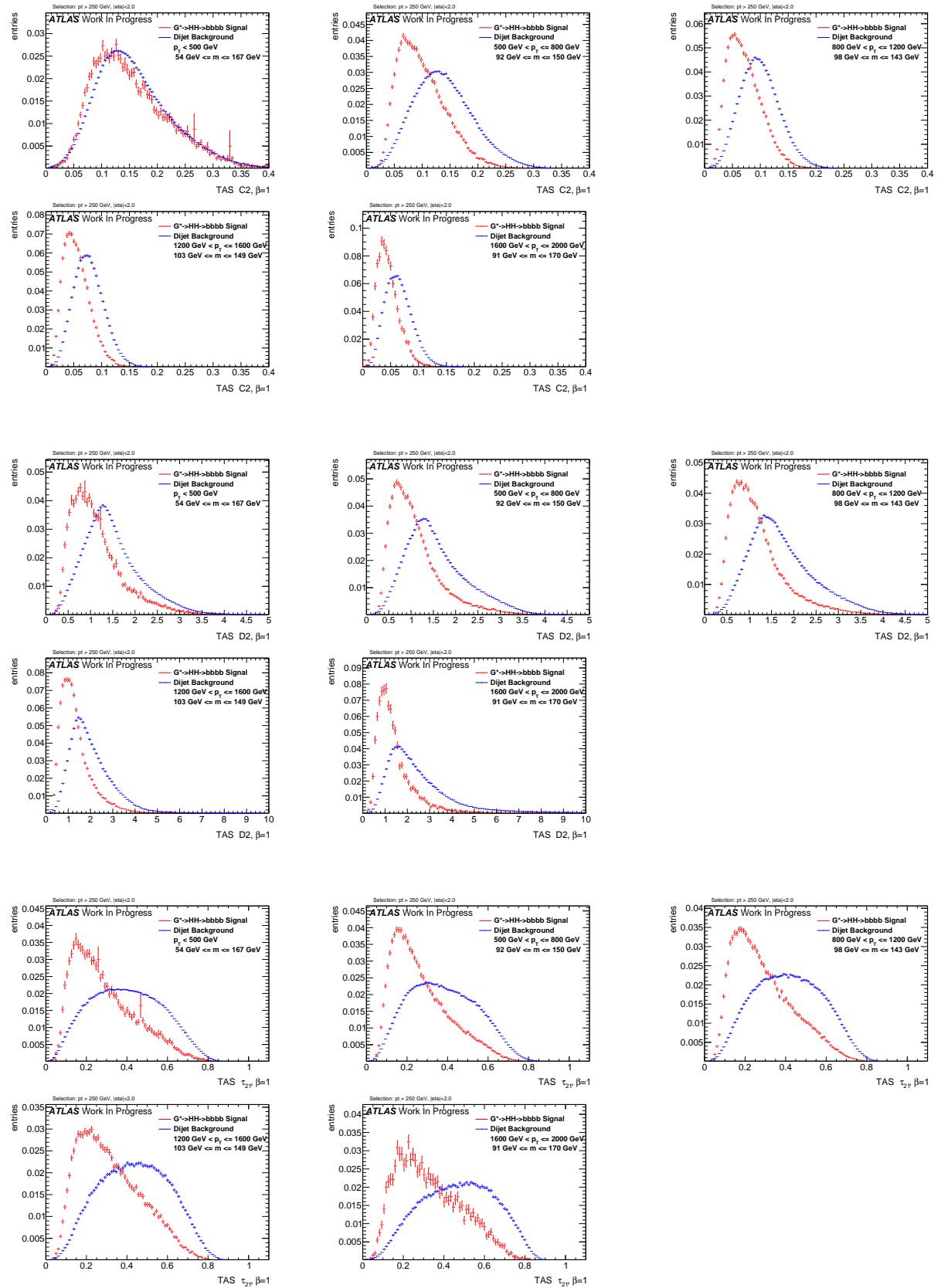
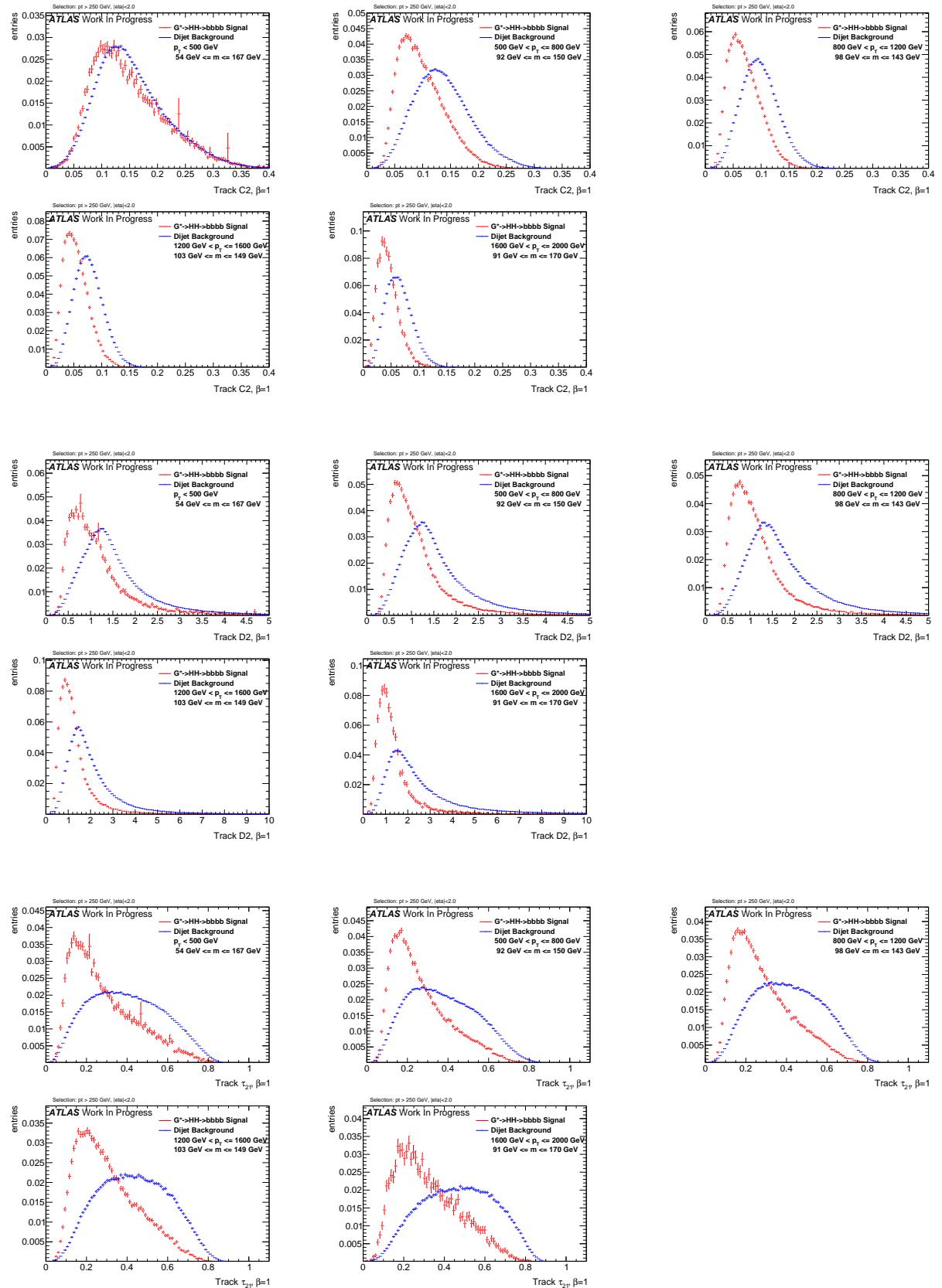
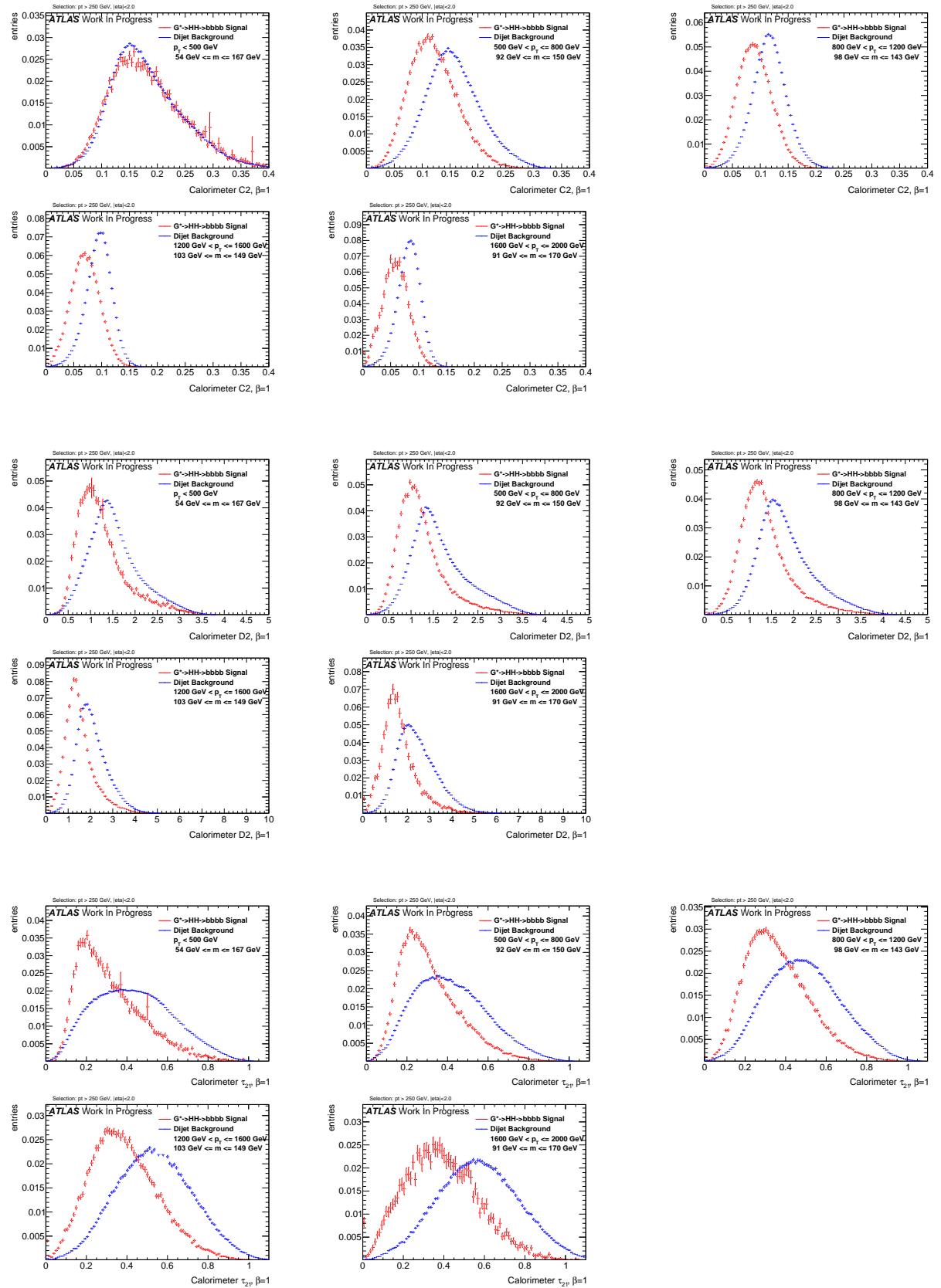
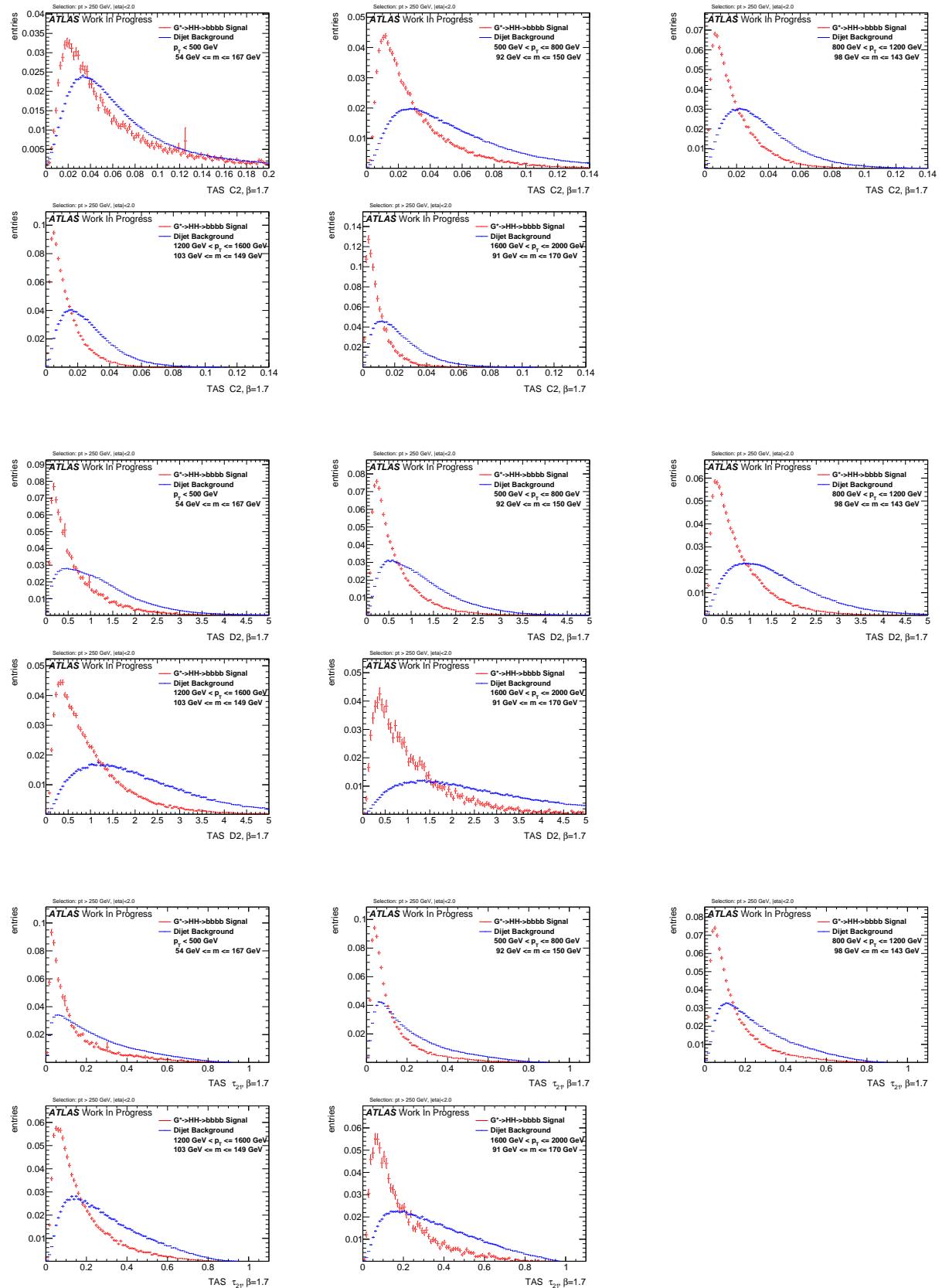
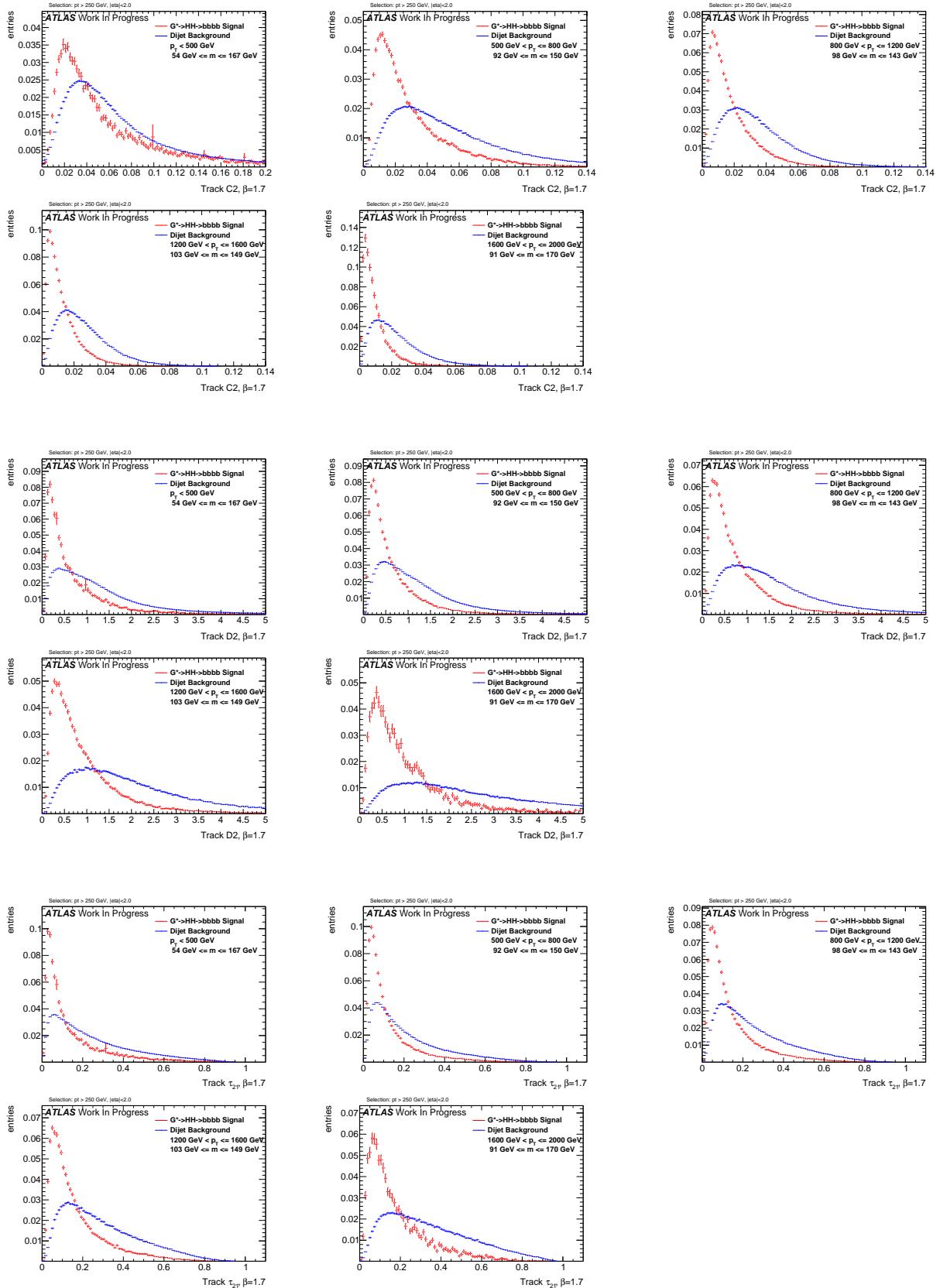


Figure 288: Distributions for Higgs boson tagging using TAS  $\beta = 1$ . C2, D2,  $\tau_{21}$  top down.

Figure 289: Distributions for Higgs boson tagging using tracks  $\beta = 1$ . C2, D2,  $\tau_{21}$  top down.

Figure 290: Distributions for Higgs boson tagging using calorimeter clusters  $\beta = 1$ . C2, D2,  $\tau_{21}$  top down.

Figure 291: Distributions for Higgs boson tagging using TAS  $\beta = 1.7$ . C2, D2,  $\tau_{21}$  top down.

Figure 292: Distributions for Higgs boson tagging using tracks  $\beta = 1.7$ . C2, D2,  $\tau_{21}$  top down.

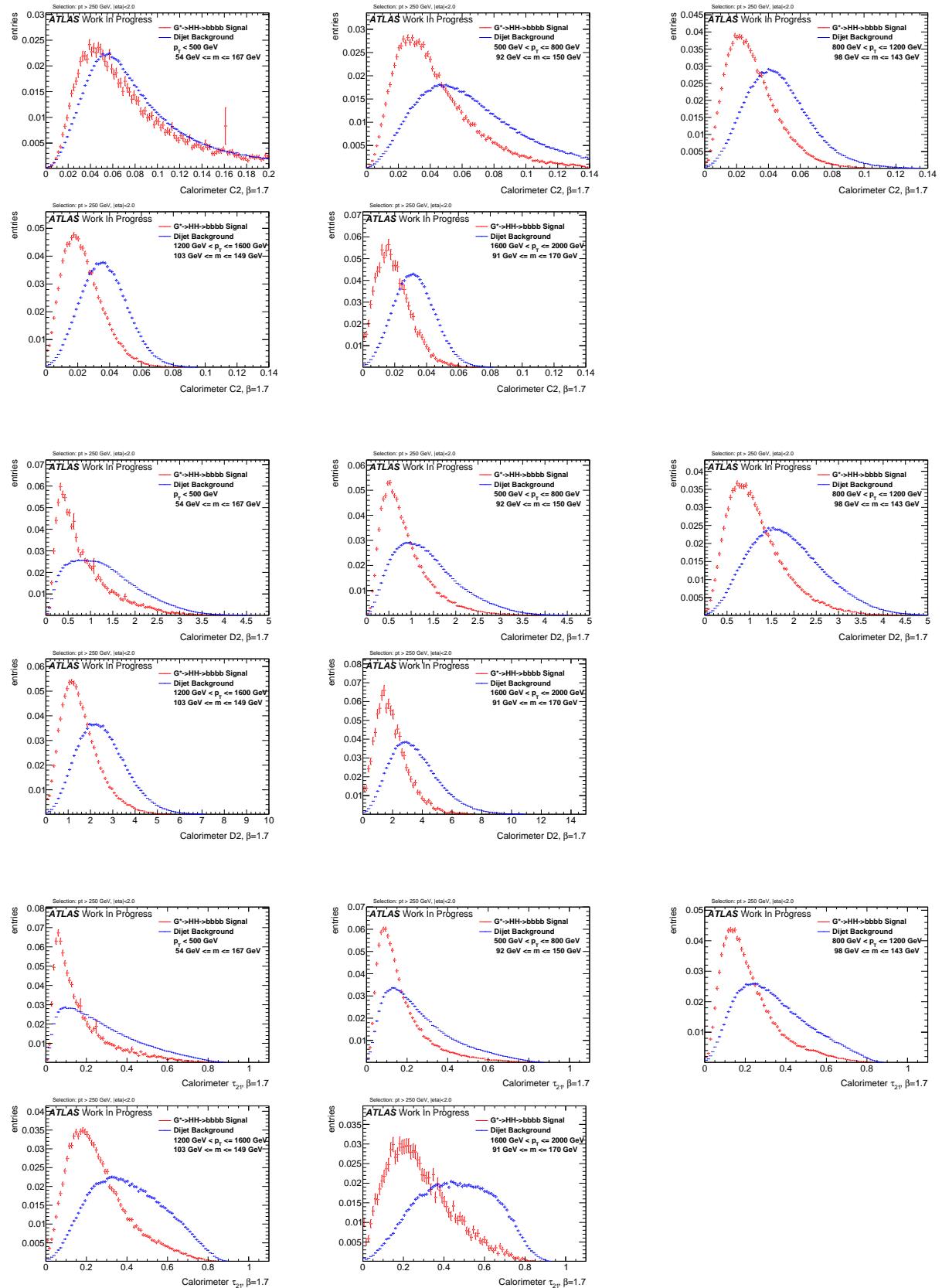
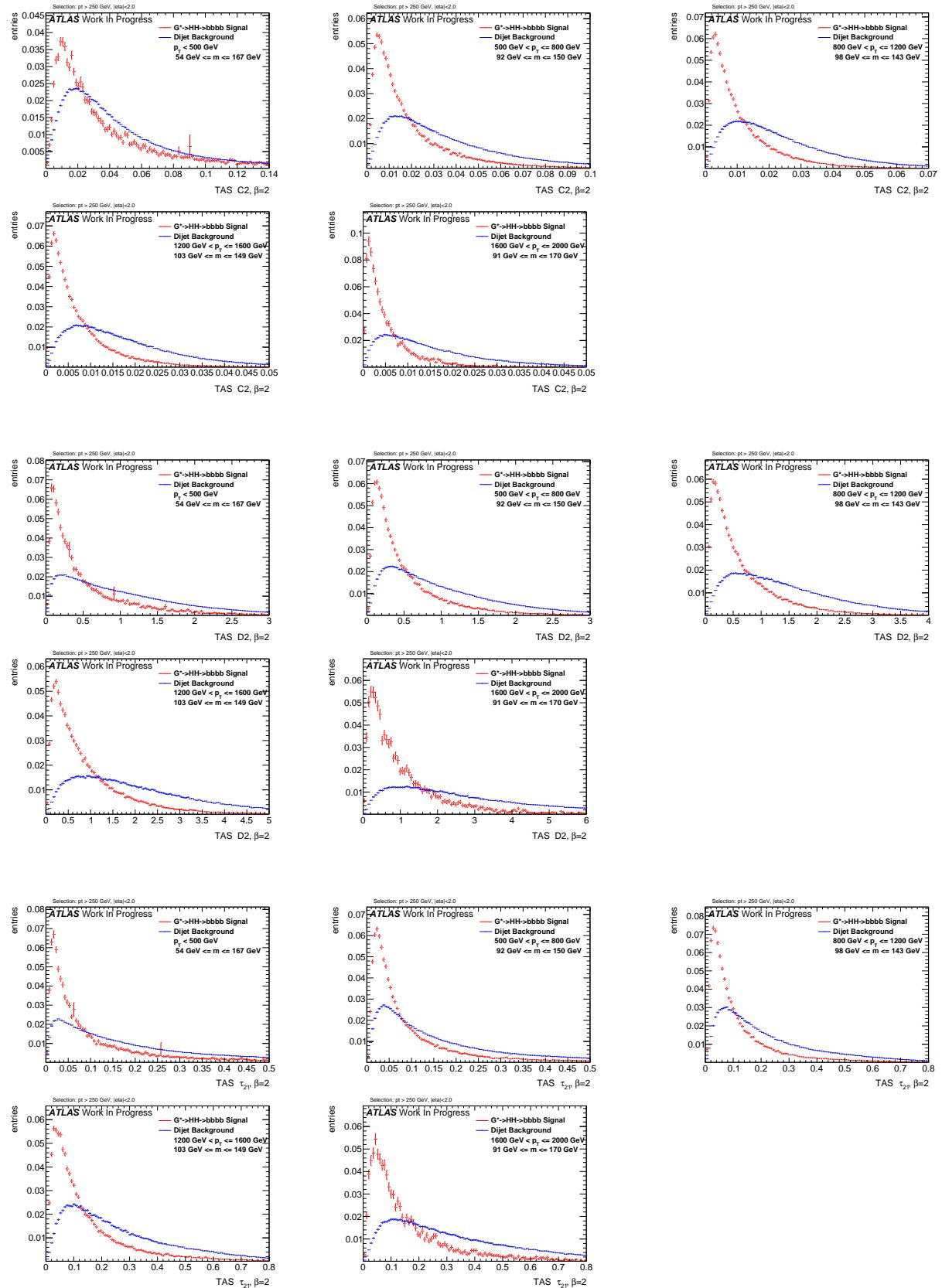
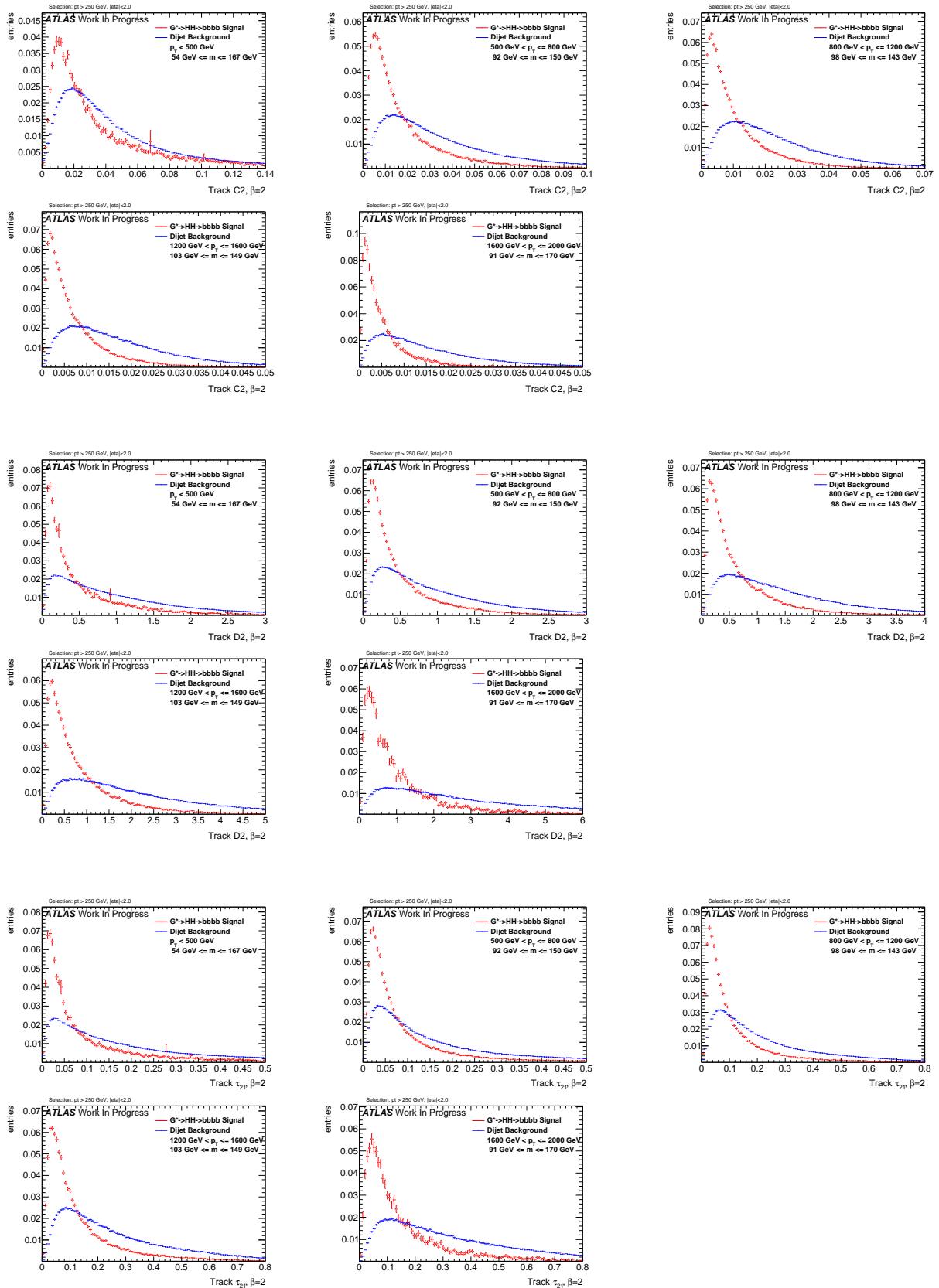
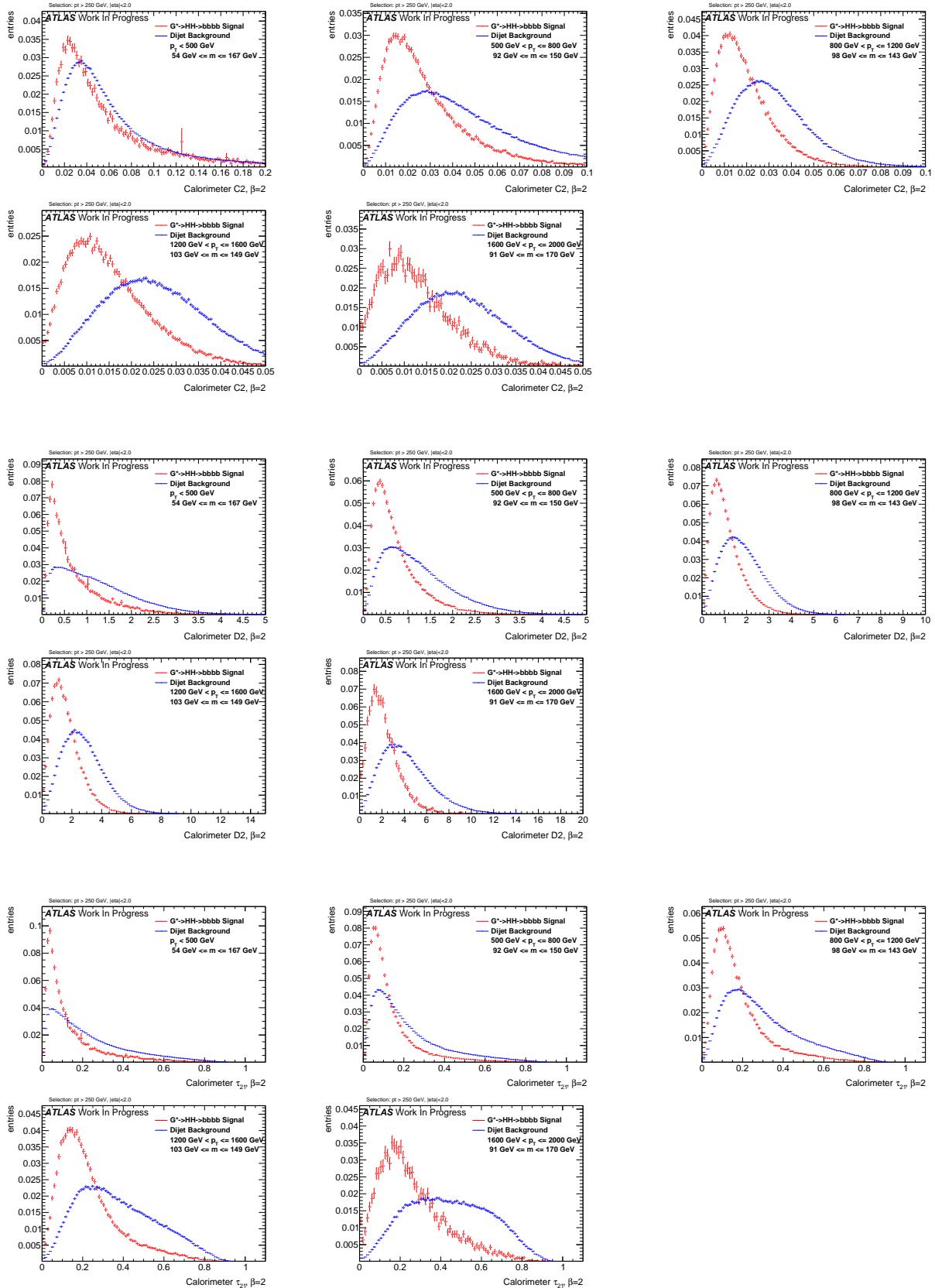
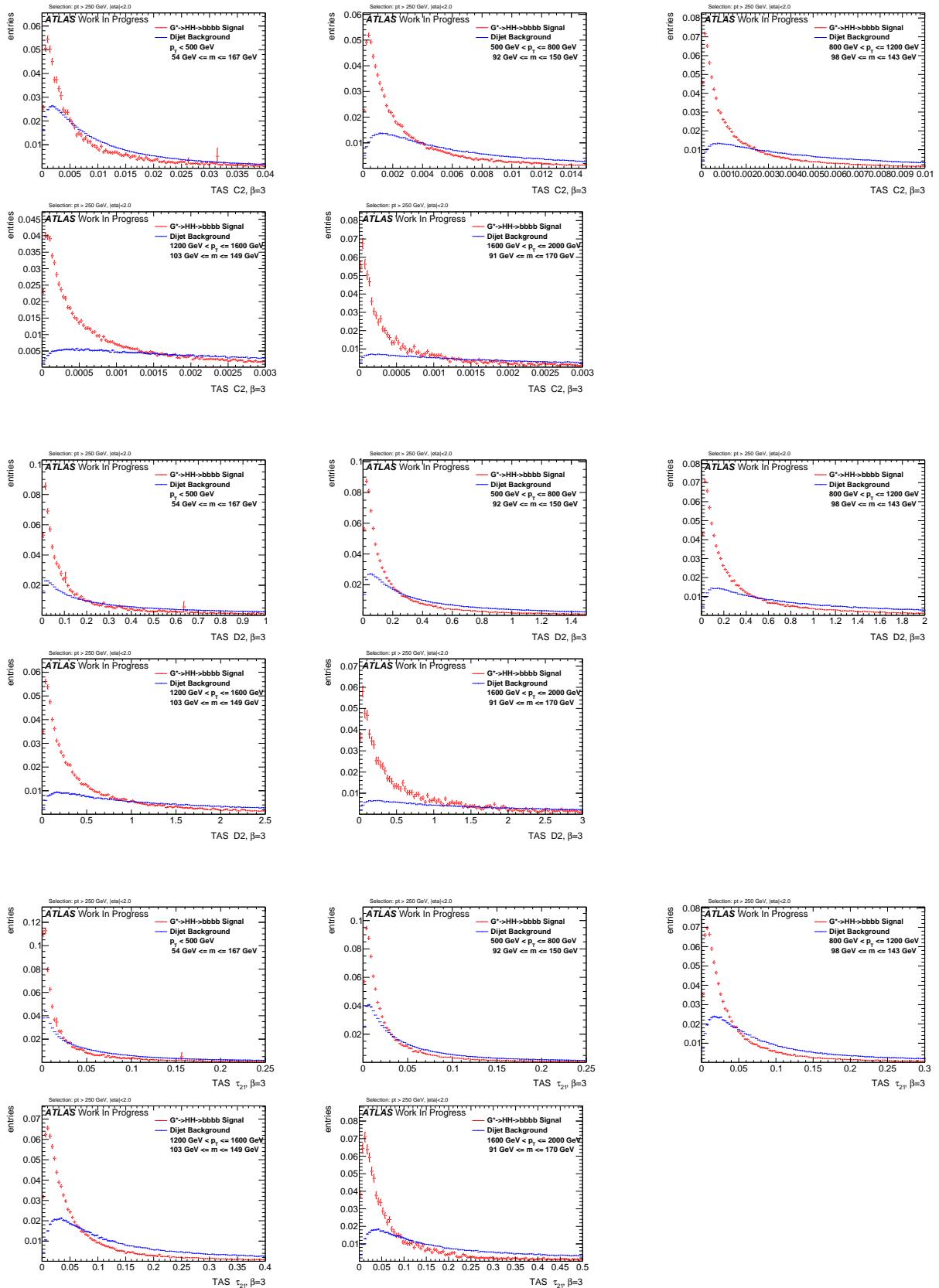


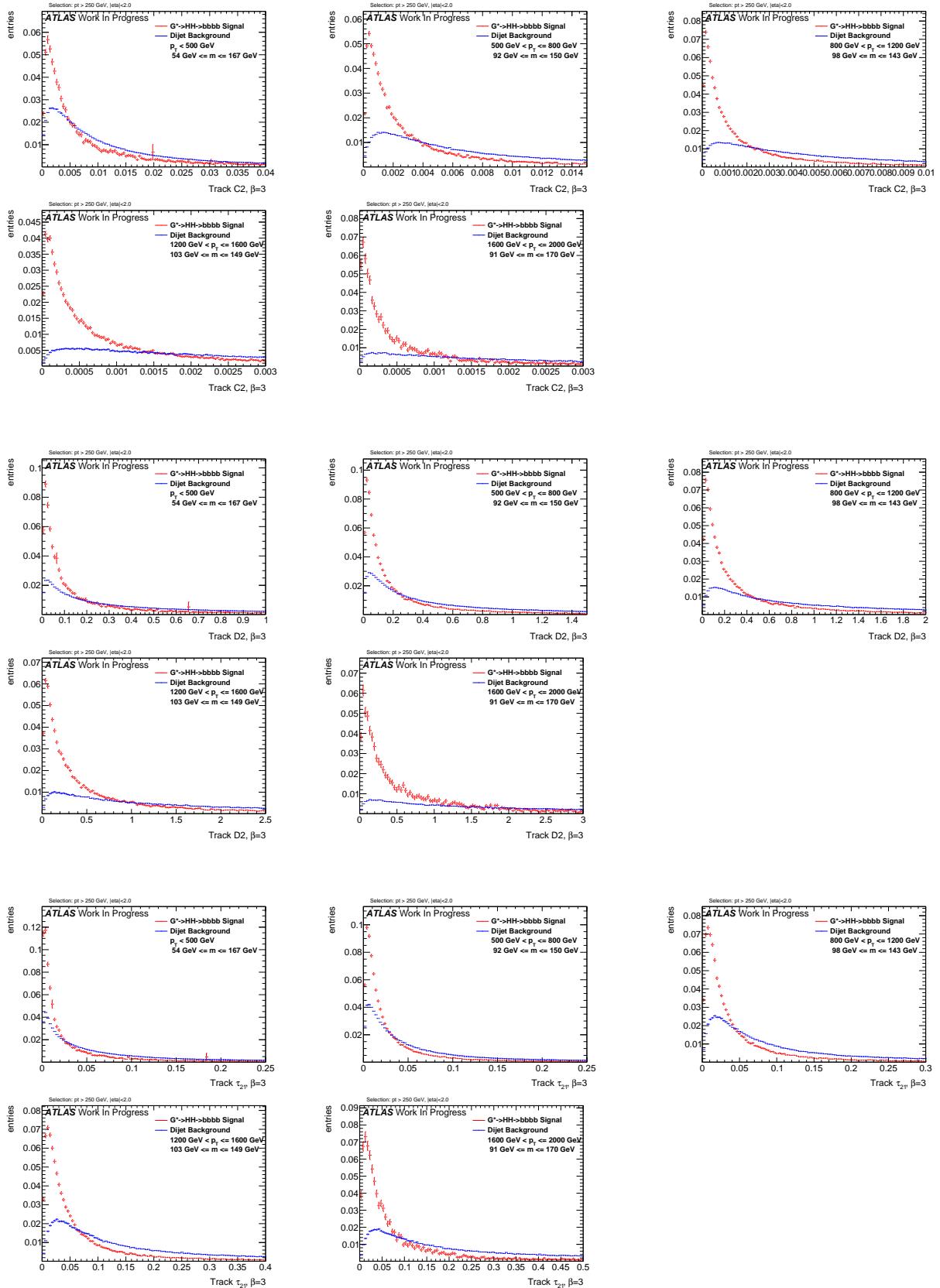
Figure 293: Distributions for Higgs boson tagging using calorimeter clusters  $\beta = 1.7$ . C2, D2,  $\tau_{21}$  top down.

Figure 294: Distributions for Higgs boson tagging using TAS  $\beta = 2$ . C2, D2,  $\tau_{21}$  top down.

Figure 295: Distributions for Higgs boson tagging using tracks  $\beta = 2$ . C2, D2,  $\tau_{21}$  top down.

Figure 296: Distributions for Higgs boson tagging using calorimeter clusters  $\beta = 2$ . C2, D2,  $\tau_{21}$  top down.

Figure 297: Distributions for Higgs boson tagging using TAS  $\beta = 3$ . C2, D2,  $t_{21}$  top down.

Figure 298: Distributions for Higgs boson tagging using tracks  $\beta = 3$ . C2, D2,  $\tau_{21}$  top down.

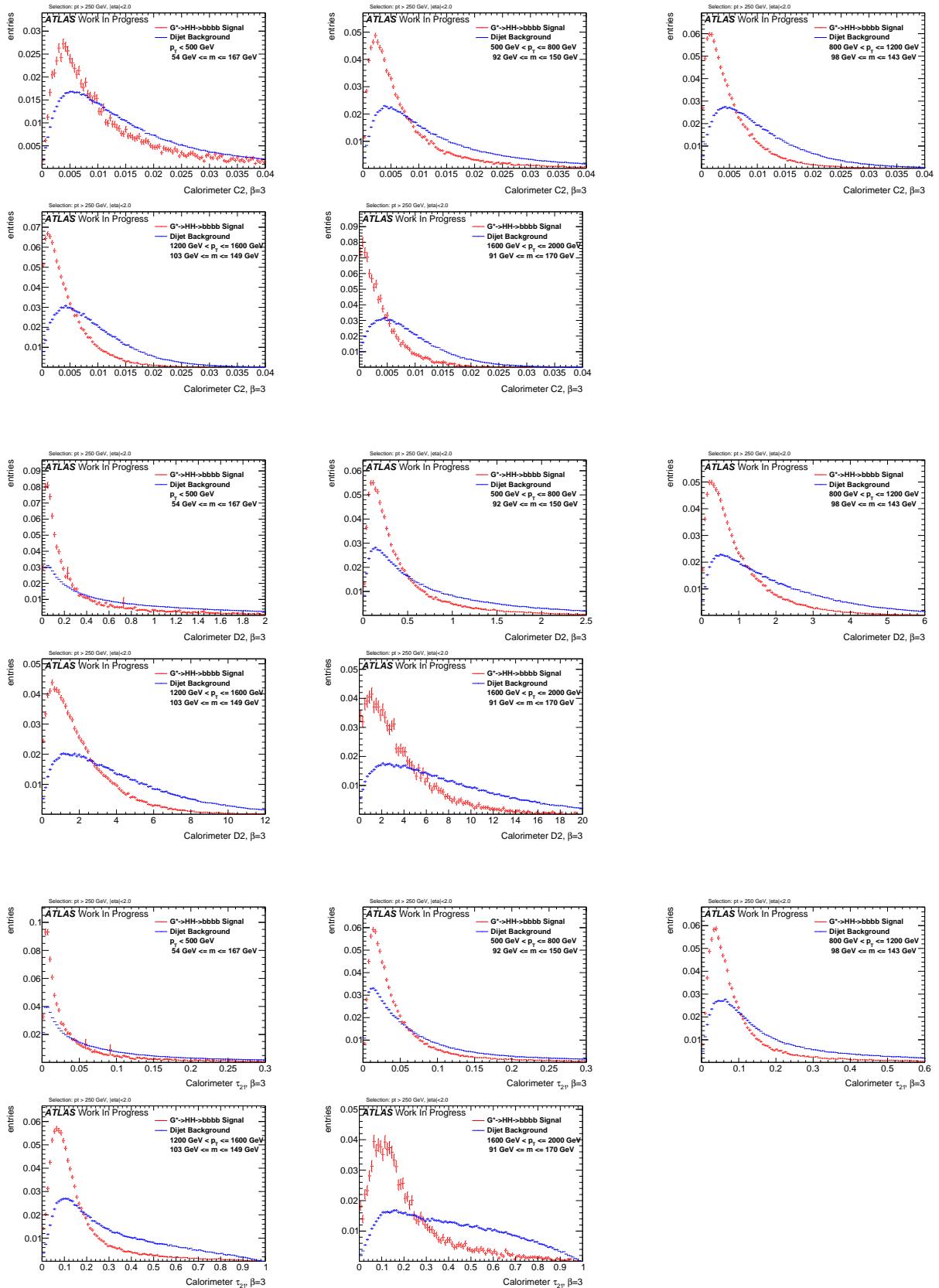
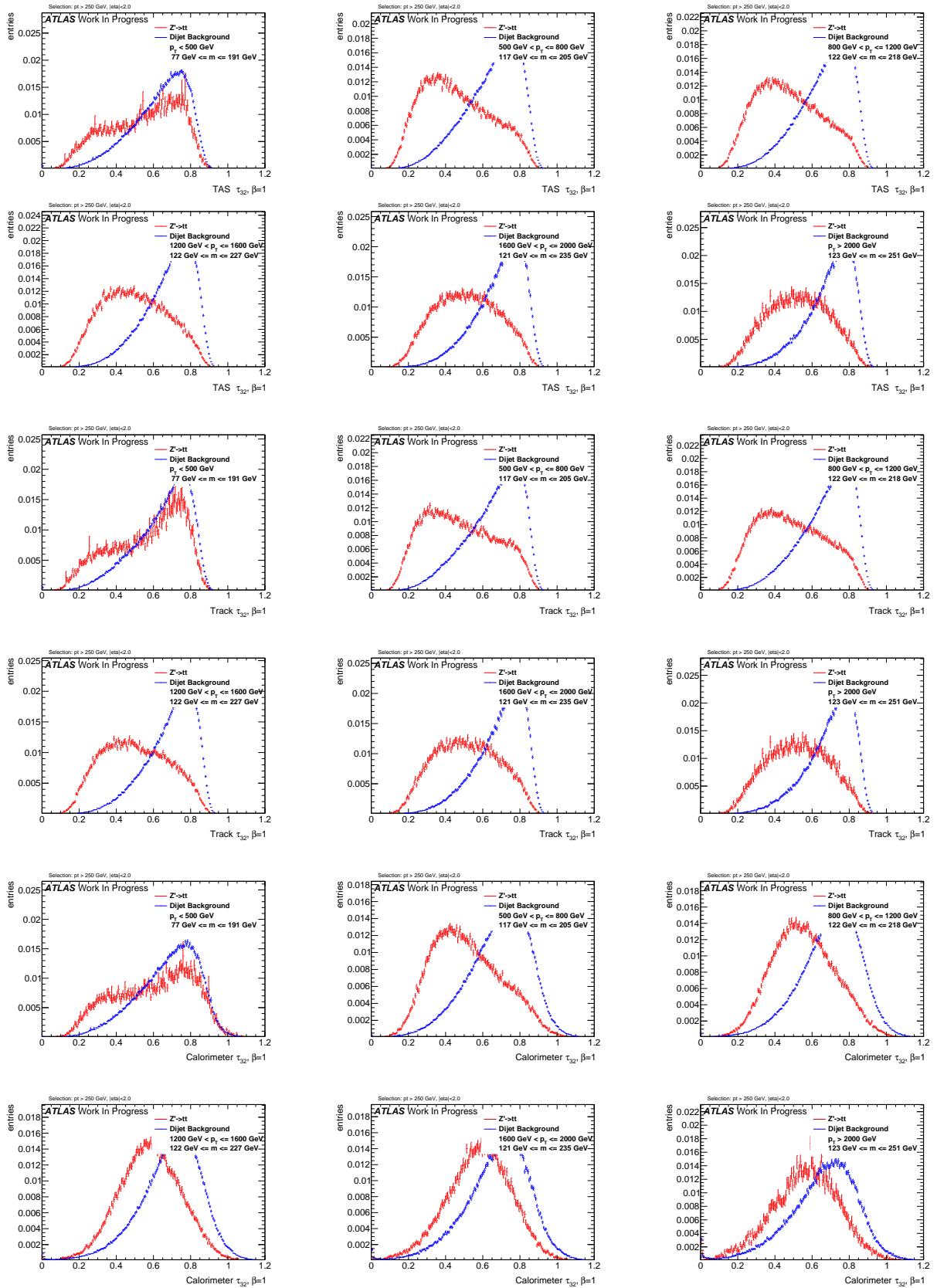
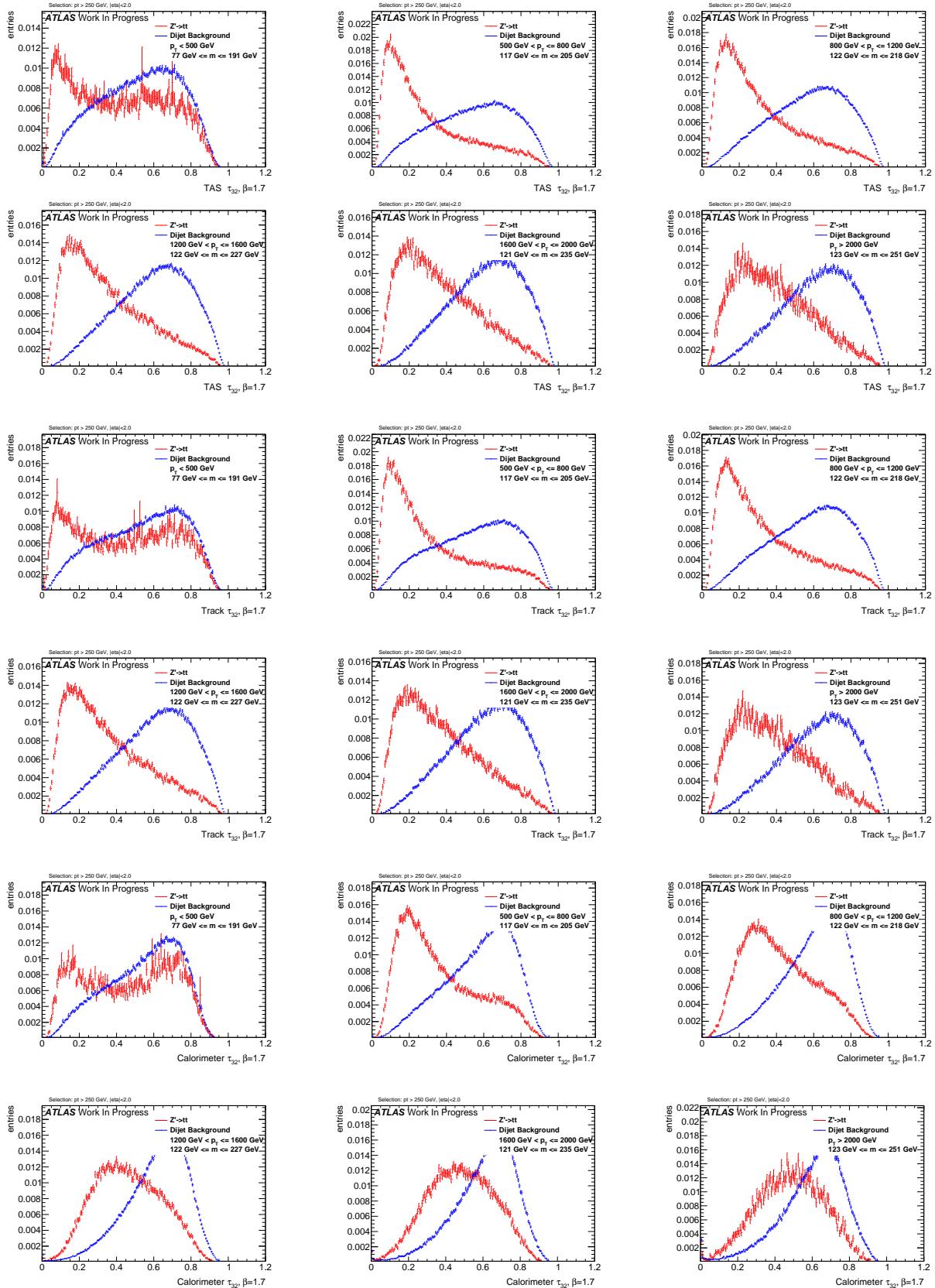
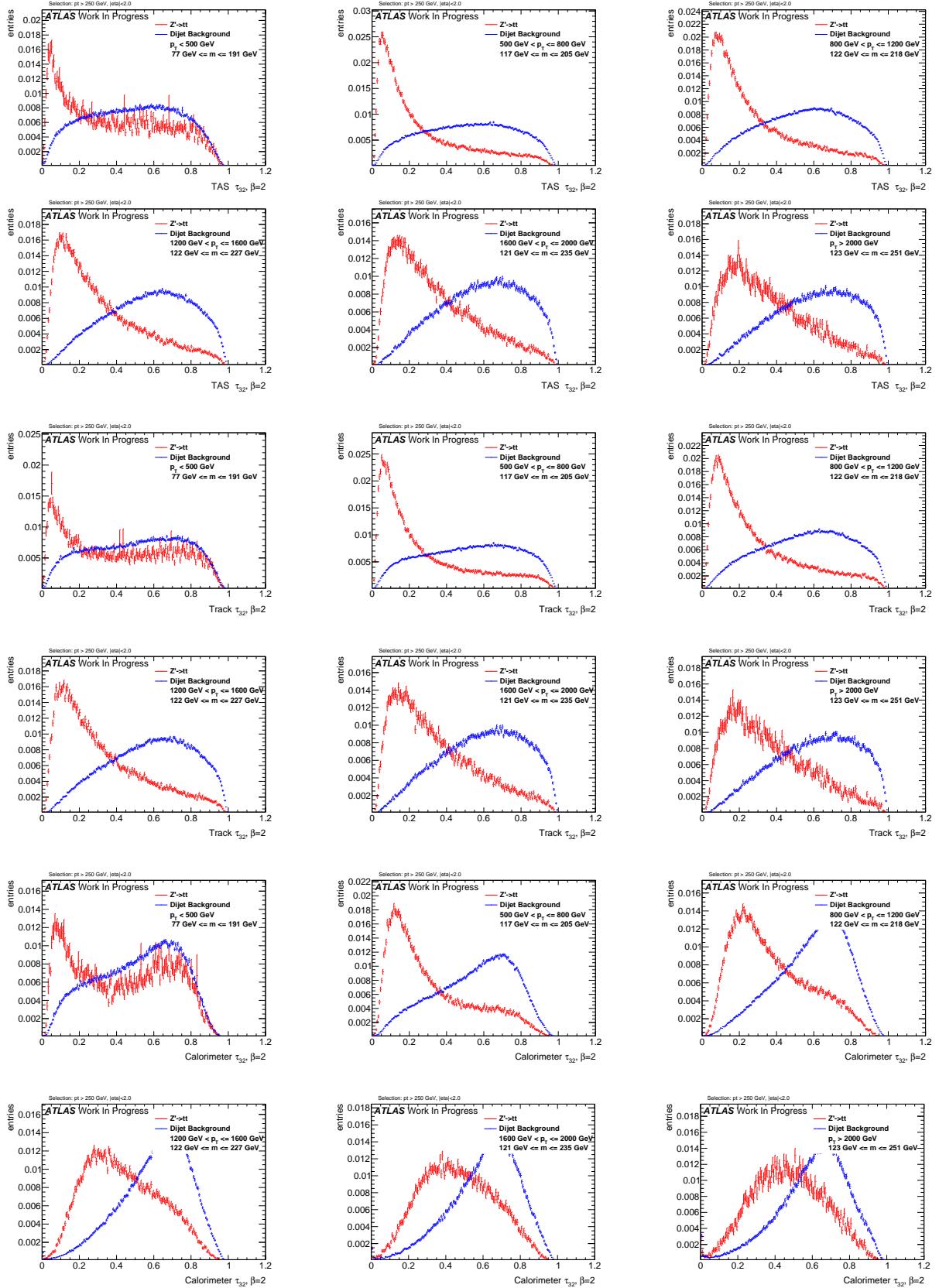
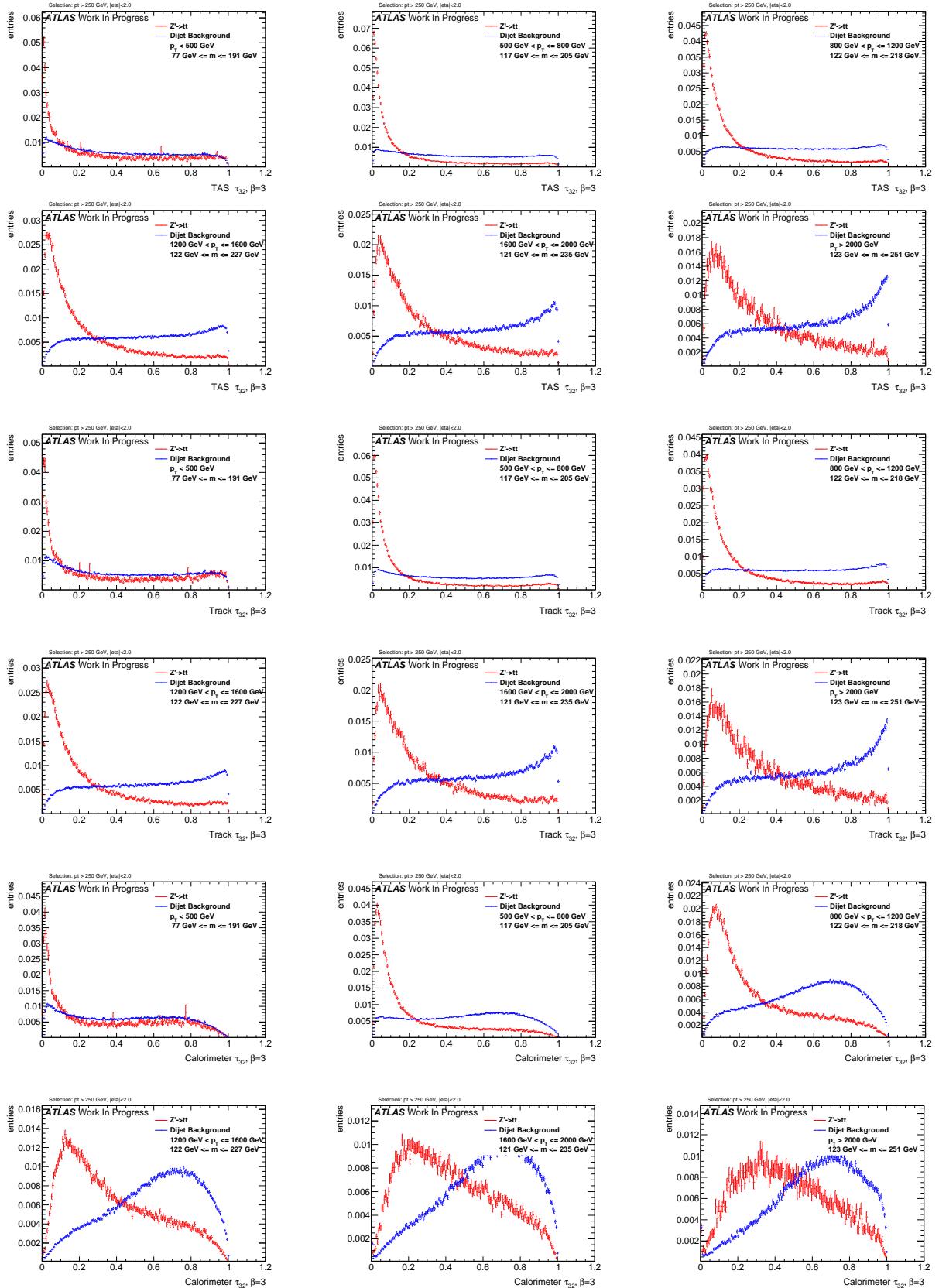


Figure 299: Distributions for Higgs boson tagging using calorimeter clusters  $\beta = 3$ . C2, D2,  $\tau_{21}$  top down.

Figure 300: Distributions for Top tagging using  $\tau_{32}$  ( $\beta = 1$ ) using TAS, tracks and calorimeter clusters top down.

Figure 301: Distributions for Top tagging using  $\tau_{32}$  ( $\beta = 1.7$ ) using TAS, tracks and calorimeter clusters top down.

Figure 302: Distributions for Top tagging using  $\tau_{32}$  ( $\beta = 2$ ) using TAS, tracks and calorimeter clusters top down.

Figure 303: Distributions for Top tagging using  $\tau_{32}$  ( $\beta = 3$ ) using TAS, tracks and calorimeter clusters top down.

---

## 878 N W Distributions

879  $\beta = 0.5$

880  $\beta = 1$

881  $\beta = 1.7$

882  $\beta = 2$

883  $\beta = 3$

## 884 O Higgs Distributions

885  $\beta = 0.5$

886  $\beta = 1$

887  $\beta = 1.7$

888  $\beta = 2$

889  $\beta = 3$

## 890 P Top Distributions

891  $\beta = 1$

892  $\beta = 1.7$

893  $\beta = 2$

894  $\beta = 3$

## 895 References

- 896 [1] Particle Data Group Collaboration, K. Olive *et al.*, *Review of Particle Physics*, Chinese Physics C 38  
897 (2014) 090001.
- 898 [2] Paul Langacker, *The Physics of Heavy Z Gauge Bosons*. In: Rev. Mod. Phys. 81 (2009), pp. 1199–1228.  
899 doi: 10.1103/RevModPhys.81.1199. arXiv: 0801.1345 [hep-ph].
- 900 [3] Martin Schmaltz and Christian Spethmann, *Two Simple W' Models for the Early LHC*. In: JHEP 07  
901 (2011), p. 046. doi: 10.1007/JHEP07(2011)046. arXiv: 1011. 5918 [hep-ph].
- 902 [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:  
903 10.1016/j.physletb.2016.02.015. arXiv: 1512.05099 [hep-ex].

- 905 [5] The CMS Collaboration, *Search for massive resonances decaying into pairs of boosted bosons in*  
 906 *semi-leptonic final states at  $\sqrt{s} = 8$  TeV*, CERN-PH-EP/2013-037 2014/09/03 arXiv:1405.3447v2  
 907 [hep-ex].
- 908 [6] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016).  
 909 <http://pdg.lbl.gov/2015/reviews/rpp2015-rev-wprime-searches.pdf>
- 910 [7] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016).  
 911 <http://pdg.lbl.gov/2015/reviews/rpp2015-rev-zprime-searches.pdf>
- 912 [8] L. Randall and R. Sundrum, *A Large Mass Hierarchy from a Small Extra Dimension*, Phys. Rev. Lett.  
 913 83 (1999) 3370, arXiv: hep-ph/9905221.
- 914 [9] H. Davoudiasl, J.L. Hewett and T.G. Rizzo, *Warped Phenomenology*, Stanford Linear Accelerator  
 915 Center Stanford CA 94309, USA, arXiv:hep-ph/9909255v1 6 Sep 1999.
- 916 [10] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report, 1 & 2*  
 917 , Tech. Rep. ATLAS TDR 14, CERN/LHCC 99-14, Geneva, 1999.
- 918 [11] L. Evans and P. Bryant, *LHC Machine*, JINST 3 (2008) S08001.
- 919 [12] C. Patrignani *et al.*, Particle Data Group, Chin. Phys. C, 40, 100001 (2016).  
 920 <http://pdg.lbl.gov/2010/reviews/rpp2010-rev-qcd.pdf>
- 921 [13] The CMS Collaboration, *Measurement of the inclusive 3-jet production differential cross section in*  
 922 *proton-proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range*,  
 923 CERN-PH-EP/2013-037 2015/05/04 arXiv:hep-ph/14121633.
- 924 [14] The ATLAS Collaboration, *Improved luminosity determination in pp collisions at  $\sqrt{s} = 7$  TeV using*  
 925 *the ATLAS detector at the LHC*, Eur. Phys. J. C (2013) 73: 2518. doi:10.1140/epjc/s10052-013-2518-3.
- 926 [15] W. Bartel *et al.* [JADE Collaboration], *Observation of Planar Three Jet Events in  $e + e^-$  Annihilation*  
 927 *and Evidence for Gluon Bremsstrahlung*, Phys. Lett. B 91, 142 (1980).
- 928 [16] Ahmed Ali (DESY), Gustav Kramer (Univ. Hamburg), *Jets and QCD: A Historical Review of the*  
 929 *Discovery of the Quark and Gluon Jets and its Impact on QCD*, arXiv:1012.2288v2 [hep-ph].
- 930 [17] Matteo Cacciari, Gavin P. Salam, Gregory Soyez, *The anti- $k_t$  jet clustering algorithm*,  
 931 arXiv:0802.1189v2 [hep-ph].
- 932 [18] F. Halzen, A.D. Martin and D.M. Scott, Phys. Rev. D25 (1982) 754, Phys. Lett. B112 (1982) 160.
- 933 [19] Torbjorn Sjostrand, *Monte Carlo Generators*, hep-ph/0611247, CERN-LCGAPP-2006-06.
- 934 [20] W. Lampl *et al.*, *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-  
 935 PUB-2008-002.
- 936 [21] The ATLAS Collaboration, *Jet energy resolution in proton-proton collisions at  $\sqrt{s} = 7$  TeV recorded*  
 937 *in 2010 with the ATLAS detector*, CERN-PH-EP-2012-191 arXiv:1210.6210v1 [hep-ex].
- 938 [22] GEANT4 Collaboration (S. Agostinelli *et al.*), *GEANT4: A Simulation toolkit*, Nucl.Instrum.Meth.  
 939 A506 (2003) 250-303 DOI: 10.1016/S0168-9002(03)01368-8 SLAC-PUB-9350, FERMILAB-PUB-  
 940 03-339.
- 941 [23] Ariel Schwartzman, *Jet energy calibration at the LHC*, SLAC National Accelerator Laboratory, 2575  
 942 Sand Hill Road Menlo Park, California 94025, USA, arXiv:1509.05459v1 [hep-ex] 17 Sep 2015.

- 943 [24] The ATLAS Collaboration, *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*,  
944 ATLAS-CONF-2016-035 19 July 2016.
- 945 [25] The ATLAS Collaboration, *Boosted Higgs ( $\rightarrow b\bar{b}$ ) Boson Identification with the ATLAS Detector at*  
946  $\sqrt{s} = 13 \text{ TeV}$ , ATLAS-CONF-2016-039 30th July 2016.
- 947 [26] The ATLAS Collaboration, *Search for high-mass diboson resonances with boson-tagged jets in*  
948 *proton-proton collisions at  $\sqrt{s} = 8 \text{ TeV}$  with the ATLAS detector*, arXiv:1506.00962v3 [hep-ex].
- 949 [27] The CMS Collaboration, *Combination of searches for WW, WZ, ZZ, WH, and ZH resonances at  $\sqrt{s}=$*   
950 *8 and 13 TeV*, CMS PAS B2G-16-007.
- 951 [28] The ATLAS Collaboration, *Search for heavy particles decaying to pairs of highly-boosted top quarks*  
952 *using lepton-plus-jets events in proton-proton collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector*,  
953 ATLAS-CONF-2016-014.
- 954 [29] The ATLAS Collaboration, *Search for resonances with boson-tagged jets in  $15.5 \text{ fb}^{-1}$  of p-p*  
955 *collisions at  $\sqrt{s} = 13 \text{ TeV}$  collected with the ATLAS detector*, ATLAS-CONF-2016-055.
- 956 [30] The ATLAS Collaboration, *Search for pair production of Higgs bosons in the  $b\bar{b}\bar{b}\bar{b}$  final state using*  
957 *proton proton collisions at  $\sqrt{s} = 13 \text{ TeV}$  with the ATLAS detector*, ATLAS-CONF-2016-049.
- 958 [31] The ATLAS Collaboration, *Performance of jet substructure techniques for large-R jets in*  
959 *proton-proton collisions at  $\sqrt{s} = 7 \text{ TeV}$  using the ATLAS detector*, CERN-PH-EP-2013-069  
960 arXiv:1306.4945 [hep-ex].
- 961 [32] Oleg Zenin, on behalf of the ATLAS Collaboration, *Soft QCD and underlying event measurements*  
962 *at ATLAS*, Nuclear and Particle Physics Proceedings Volumes 273275, April-June 2016, Pages 2053-  
963 2058 37th International Conference on High Energy Physics (ICHEP).
- 964 [33] Page title: JetSubstructureECFA2014 AtlasPublic TWiki Website name: Twiki.cern.ch  
965 URL:<https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetSubstructureECFA2014> Access date: 27th  
966 October 2016.
- 967 [34] E. Lehmann and G. Casella, *Theory of Point Estimation*, Springer Verlag, 1998, isbn: 0387985026.
- 968 [35] Andrew J. Larkoski, Ian Moult, and Duff Neill. *Power Counting to Better Jet Observables*. *JHEP*,  
969 12:009, 2014.
- 970 [36] Andrew J. Larkoski, Ian Moult, and Duff Neill. *Analytic Boosted Boson Discrimination*. *JHEP*,  
971 05:117, 2016.
- 972 [37] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. *Energy Correlation Functions for Jet Sub-*  
973 *structure*. *JHEP*, 06:108, 2013.
- 974 [38] Jesse Thaler and Ken Van Tilburg. *Identifying Boosted Objects with N-subjettiness*. *JHEP*, 03:015,  
975 2011.
- 976 [39] Georges Aad et al. *Identification of boosted, hadronically decaying W bosons and comparisons with*  
977 *ATLAS data taken at  $\sqrt{s} = 8 \text{ TeV}$* . *Eur. Phys. J.*, C76(3):154, 2016.
- 978 [40] Jet Substructure and jet-by-jet tagging meeting: URL: ht-  
979 [tps://indico.cern.ch/event/495766/contributions/1174422/attachments/1231007/1804520/JSS\\_18Feb.pdf](https://indico.cern.ch/event/495766/contributions/1174422/attachments/1231007/1804520/JSS_18Feb.pdf)

980 **List of contributions**

981

## **Auxiliary material**

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

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