

# CMS Draft Analysis Note

*The content of this note is intended for CMS internal use and distribution only*

2018/06/04

Head Id: 463242

Archive Id: 463279

Archive Date: 2018/06/04

Archive Tag: trunk

## Boosted jet identification with particle-level information and deep neural networks

Loukas Gouskos<sup>1</sup>, Huang Huang<sup>2</sup>, Joseph Incandela<sup>1</sup>, Jan Kieseler<sup>3</sup>, Qiang Li<sup>2</sup>, Matthias Mozer<sup>4</sup>, Huilin Qu<sup>1</sup>, Paris Sphicas<sup>3</sup>, Markus Stoye<sup>3</sup>, and Mauro Verzetti<sup>3</sup>

<sup>1</sup> University of California, Santa Barbara

<sup>2</sup> Peking University

<sup>3</sup> CERN

<sup>4</sup> KIT

### Abstract

Reconstruction and identification of boosted heavy particles (top quarks and W, Z, Higgs bosons) from their hadronic decays can play an important role in both searches for new physics and measurements of standard model processes at the LHC. We present a new approach for boosted jet identification using particle-flow jets. One dimensional convolutional neural networks are utilized to classify a jet directly from its reconstructed constituent particles. The new method shows significant improvement in performance compared to alternative multivariate methods using jet-level observables.

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor: DeepAK8 team

PDFTitle: Boosted jet identification with particle-level information and deep neural networks

PDFSubject: CMS

PDFKeywords: CMS, physics

Please also verify that the abstract does not use any user defined symbols



## 1 Contents

2	1	Introduction . . . . .	2
3	2	Data and simulated samples . . . . .	3
4	2.1	Simulated samples used for the training . . . . .	3
5	2.2	Simulated samples used for evaluating the performance . . . . .	3
6	2.3	Data and simulated samples for the validation of the performance . . . . .	3
7	3	Event reconstruction and physics objects . . . . .	5
8	3.1	Vertex selection . . . . .	5
9	3.2	AK4 jets . . . . .	5
10	3.3	AK8 jets . . . . .	5
11	3.4	Lepton selection . . . . .	5
12	3.5	Photons . . . . .	6
13	3.6	Missing transverse energy . . . . .	7
14	3.7	MET filters . . . . .	7
15	4	Boosted jet identification algorithm . . . . .	8
16	4.1	Truth label definition . . . . .	8
17	4.2	Datasets and preprocessing . . . . .	8
18	4.3	Features . . . . .	9
19	4.4	Architecture . . . . .	10
20	5	Performance in simulation . . . . .	14
21	5.1	ROC curves . . . . .	14
22	5.2	Efficiency vs $p_T$ and NPV . . . . .	17
23	5.3	Impacts on the jet mass shape of the background events . . . . .	22
24	6	Event selection . . . . .	24
25	6.1	The $t\bar{t}$ ( $1\ell$ ) signal sample . . . . .	24
26	6.2	The QCD multijet background sample . . . . .	24
27	6.3	The $\gamma$ +jet background sample . . . . .	24
28	7	Performance in data and systematic uncertainties . . . . .	25
29	7.1	The $t$ quark and $W$ boson identification performance in data . . . . .	25
30	7.2	Misidentification probability in the QCD multijet sample . . . . .	29
31	7.3	Misidentification probability in the $\gamma$ +jets sample . . . . .	33
32	7.4	Systematic uncertainties . . . . .	37
33	7.5	Corrections to simulation . . . . .	38
34	8	Summary . . . . .	49
35	A	Effects of subjet b-tagging . . . . .	52
36	B	Comparison of different deep learning approaches . . . . .	54
37	C	Impacts of the input features . . . . .	55

## <sup>38</sup> 1 Introduction

<sup>39</sup> The reconstruction and identification of boosted heavy particles (top quarks, W, Z and Higgs  
<sup>40</sup> bosons) can provide powerful handles for searches for physics beyond the standard model as  
<sup>41</sup> well as for measurements of standard model processes at the LHC. The hadronic decays of  
<sup>42</sup> these boosted heavy particles are of particular interest, due to their large branching fractions,  
<sup>43</sup> and the distinct characteristics of their decay product which can be utilized to separate them  
<sup>44</sup> from jets initiated by light (non-top) quarks or gluons. In this note, we present a novel ap-  
<sup>45</sup> proach for boosted jet identification. This algorithm is developed on CMS particle-flow jets,  
<sup>46</sup> and utilizes Deep neural networks to classify the jet directly from its reconstructed constituent  
<sup>47</sup> particles, allowing for a better exploitation of the full granularity of the CMS detector.



---

## 48 2 Data and simulated samples

### 49 2.1 Simulated samples used for the training

50 The samples used for the training of the algorithm are listed in Table 1.

Table 1: Simulated MC samples used for the training of the algorithm. `proc` refers to `RunIIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1`.

Sample	Full name
top	/TT_Mtt-*_TuneCUETP8M2T4_13TeV-powheg-pythia8/proc/MINIAODSIM /ZprimeToTTJet_M-*_TuneCUETP8M1_13TeV-amcatnlo-pythia8/proc/MINIAODSIM /RSGluonToTT_M-*_TuneCUETP8M1_13TeV-pythia8/proc/MINIAODSIM
W	/BulkGravToWWToWlepWhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /BulkGravToWW_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /RadionToWWToWlepWhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /RadionToWW_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /RadionToWW_width0p*_M-*_TuneCUETP8M1_13TeV-madgraph-pythia8/proc/MINIAODSIM /WprimeToWZToWhadZinv_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /WprimeToWZToWhadZlep_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /WprimeToWZToWhadZlep_width0p*_M-*_TuneCUETP8M1_13TeV-madgraph-pythia8/proc/MINIAODSIM /ZprimeToWWToWlepWhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /ZprimeToWW_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /ZprimeToWW_width0p*_M-*_TuneCUETP8M1_13TeV-madgraph-pythia8/proc/MINIAODSIM
Z	/BulkGravToZZToZhadZhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /BulkGravToZZToZhadZinv_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /BulkGravToZZToZlepZhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /RadionToZZToZlepZhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /WprimeToWZToWlepZhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /WprimeToWZToWlepZhad_width0p*_M-*_TuneCUETP8M1_13TeV-madgraph-pythia8/proc/MINIAODSIM
Higgs	/BulkGravTohhTohhVvhbb_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /BulkGravTohhTohhbhbb_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /GluGluToBulkGravitonToHHTo4C_M-*_narrow_13TeV-madgraph-pythia8/proc/MINIAODSIM /WprimeToWhToWlepbbb_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /ZprimeToZhToZinvhb_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM /ZprimeToZhToZlephbb_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM
W/Z	/WprimeToWZToWhadZhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM
W/Higgs	/WprimeToWZ_width0p*_M-*_TuneCUETP8M1_13TeV-madgraph-pythia8/proc/MINIAODSIM /WprimeToWhToWhadhb_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM
Z/Higgs	/ZprimeToZhToZhadhb_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM
QCD	/QCD_Pt_{170to300 – 1800to2400}_TuneCUETP8M1_13TeV_pythia8/proc/MINIAODSIM

### 51 2.2 Simulated samples used for evaluating the performance

52 We use the samples listed on [1], to evaluate the performance of the algorithm in simulation.  
53 The samples are also summarized in Table 2.

Table 2: Simulated MC samples used for evaluating the performance. `proc` refers to `RunIIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1`.

Sample	Full name
top	/ZprimeToTT_M-{X}_W-{0.01X}_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM
W	/ZprimeToWW_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM
Z	/BulkGravToZZToZhadZhad_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM
Higgs	/BulkGravTohhTohhVvhbb_narrow_M-*_13TeV-madgraph/proc/MINIAODSIM
QCD	/QCD_Pt-15to7000_TuneCUETP8M1_FlatP6_13TeV_pythia8/proc/MINIAODSIM

### 54 2.3 Data and simulated samples for the validation of the performance

55 The performance of the algorithm is validated with 2016 data. The data samples are listed in  
56 Table 3 and the simulated samples are listed in Table 4.

Table 3: Primary datasets used for the analysis and the HLT paths of the corresponding triggers. We use the full 2016 dataset (Run2016B – Run2016H) and the 03Feb2017 reprocessed version.

Primary dataset	HLT path
Single muon sample	
/SingleMuon/era/MINIAOD	HLT_IsoMu24 OR HLT_IsoTkMu24 HLT_Mu15_IsoVVVL_PFHT400 OR HLT_Mu15_IsoVVVL_PFHT600 HLT_Mu15_IsoVVVL_PFHT50 OR HLT_Mu15_IsoVVVL_PFHT400_PFMET50
/MET/era/MINIAOD	HLT_PFMET110_PFMHT110_IDTight OR HLT_PFMETNoMu110_PFMHTNoMu110_IDTight HLT_PFMET120_PFMHT120_IDTight OR HLT_PFMETNoMu120_PFMHTNoMu120_IDTight
QCD multijet sample	
/JetHT/era/MINIAOD	HLT_PFHT900
Photon+jets sample	
/SinglePhoton/era/MINIAOD	HLT_Photon165_HE10
/JetHT/era/MINIAOD	HLT_CaloJet500_NojetID

Table 4: Simulated MC samples used for validating the performance in data. proc refers to RunIIISummer16MiniAODv2–PUMoriond17\_80X\_mcRun2\_asymptotic\_2016\_TrancheIV\_v6\*.

Process	Dataset	Cross section [pb]
t $\bar{t}$ , 2 $\ell$	/TTTo2L2Nu.TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8/proc/MINIAODSIM	87.31
t $\bar{t}$ , 1 $\ell$	/TTToSemilepton.TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8/proc/MINIAODSIM	364.36
t $\bar{t}$ , HERWIG	/TT_TuneEE5C_13TeV-powheg-herwigpp/proc/MINIAODSIM	831.76
t $\bar{t}$ , AMC@NLO	/TTJets.TuneCUETP8M2T4_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM	831.76
W( $\rightarrow \ell\nu$ )+jets	/WJetsToLNu_Wpt-0To50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM /WJetsToLNu_Wpt-50To100_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM /WJetsToLNu_Pt-100To250_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM /WJetsToLNu_Pt-250To400_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM /WJetsToLNu_Pt-400To600_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM /WJetsToLNu_Pt-600ToInf_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM	57297.39 3298.37 689.75 24.507 3.1101 0.46832
Single top	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4/proc/MINIAODSIM /ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4/proc/MINIAODSIM /ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1/proc/MINIAODSIM /ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1/proc/MINIAODSIM /ST_t-channel_4f_inclusiveDecays_TuneCUETP8M2T4_13TeV-powheg_V2-madspin/proc/MINIAODSIM /ST_t-channel_antitop_4f_inclusiveDecays_TuneCUETP8M2T4_13TeV-powheg_V2-madspin/proc/MINIAODSIM /ST_s-channel_4f_InclusiveDecays_13TeV-amcatnlo_pythia8/proc/MINIAODSIM	35.6 35.6 19.4 19.4 136.02 80.95 10.12
t $\bar{t}$ +V	/TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/proc/MINIAODSIM /TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/proc/MINIAODSIM /TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo_pythia8/proc/MINIAODSIM /TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo_pythia8/proc/MINIAODSIM /tZq_ll_4f_13TeV_amcatnlo_pythia8/proc/MINIAODSIM /tZq_nunu_4f_13TeV_amcatnlo_pythia8_TuneCUETP8M1/proc/MINIAODSIM	0.2043 0.4062 0.2529 0.5297 0.0758 0.1379
Diboson	/WWTo2L2Nu_13TeV-powheg/proc/MINIAODSIM /WWTo4Q_13TeV-powheg/proc/MINIAODSIM /WWToLnNu_QQ_13TeV-powheg/proc/MINIAODSIM /WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/proc/MINIAODSIM /WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8/proc/MINIAODSIM /WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/proc/MINIAODSIM /WZTo2Q2Nu_13TeV_amcatnloFXFX_madspin_pythia8/proc/MINIAODSIM /WZTo3LNu_13TeV_amcatnloFXFX-pythia8/proc/MINIAODSIM /ZZTo2L2Nu_13TeV_powheg_pythia8/proc/MINIAODSIM /ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/proc/MINIAODSIM /ZZTo2Q2Nu_13TeV_amcatnloFXFX_madspin_pythia8/proc/MINIAODSIM /ZZTo4L_13TeV_powheg_pythia8/proc/MINIAODSIM /ZZTo4Q_13TeV_amcatnloFXFX_madspin_pythia8/proc/MINIAODSIM	12.178 51.723 49.997 10.71 3.06 5.595 6.317 4.678 0.564 3.22 4.04 1.256 7.06
QCD, MADGRAPH+PYTHIA	/QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /QCD_PT15to7000_TuneCUETP8M1_FlatPt_13TeV_pythia8/proc/MINIAODSIM /QCD_PT15to7000_TuneCUETHS1_FlatPt_13TeV_herwigpp/proc/MINIAODSIM	1712000 347700 32100 6831 1207 119.9 25.24 - -
QCD, PYTHIA only	/WJetsToQQ_HT180_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM	2788
QCD, PYTHIA+HERWIG	/DYJetsToQQ_HT180_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /GluGluHToBB_M125_13TeV_powheg_pythia8/proc/MINIAODSIM	1187 28.4307
$\gamma$ +jets	/GJets_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM /GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/proc/MINIAODSIM	20790 9238 2305 274.4 93.46
t $\bar{t}\gamma$	/TTGJets.TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/proc/MINIAODSIM	3.697
t $\gamma$	/TGJets.TuneCUETP8M1_13TeV_amcatnlo_madspin_pythia8/proc/MINIAODSIM	2.967
W $\gamma$	/WGTtoJJ_A_01_JM_5f_NLO_FXFX_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM	1.234
Z $\gamma$	/ZGToJJG_SM_5f_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/proc/MINIAODSIM	0.5763

## 57 **3 Event reconstruction and physics objects**

### 58 **3.1 Vertex selection**

59 The following standard selection criteria are applied to interaction vertices reconstructed in an  
60 event:

- 61 • The vertices must come from fits to trajectories of reconstructed particle tracks with  
62 positive  $\chi^2$  values.
- 63 • There are at least 5 degrees of freedom in the vertex fit.
- 64 • The distance, absolute  $z$ , along the beam line from the nominal center of the detector  
65 is less than 24 cm.
- 66 • The transverse displacement,  $\rho$ , from the beam line is less than 2 cm.

67 Selected events are required to have at least one vertex fulfilling these requirements. The vertex  
68 which satisfies these criteria and has the highest  $\sum p_T^2$  of tracks associated to it is taken to be the  
69 primary vertex (PV) from which the physics objects used in this analysis originate.

### 70 **3.2 AK4 jets**

71 The standard jets selected for use in this analysis are Particle Flow (PF) jets reconstructed with  
72 the anti- $k_T$  algorithm [2] using a distance parameter of 0.4. The charged hadron subtraction  
73 procedure is used to correct for pileup [? ]. Jet energy corrections Summer16\_23Sep2016V4  
74 supplied by the JetMET POG are applied [? ]. The *L1FastJet*, *L2Relative*, and *L3Absolute* correc-  
75 tions are applied to data and simulation. The *L2L3Residual* corrections are applied to the data.  
76 Selected jets are required to have  $p_T > 20$  GeV and to be within the tracker volume ( $|\eta| < 2.4$ ).  
77 The jets are also required to satisfy the loose PF Jet identification criteria recommended by the  
78 JetMET POG [? ].

#### 79 **3.2.1 b-tagging**

80 B-tagged jets are identified using the Run 2 version of the Combined Secondary Vertex algo-  
81 rithm (CSVv2). The right working point recommended by the B-tag POG, corresponding to  
82 a threshold of 0.9535 on the CSV discriminator, is used to define the  $t\bar{t}$  enriched single muon  
83 sample.

### 84 **3.3 AK8 jets**

85 For the identification of the boosted heavy particles, we use jets clustered with the anti- $k_T$   
86 algorithm using a distance parameter ( $R$ ) of 0.8 (large- $R$  jets). PUPPI jets are used as they are  
87 found to show a better rejection of pileup particles. We use JetToolBox [3] to recluster the  
88 AK8Puppi jets from MiniAOD. The Summer16\_23Sep2016V4 version of jet energy corrections  
89 are applied. The *L2Relative*, and *L3Absolute* corrections are applied to data and simulation and  
90 the *L2L3Residual* corrections are applied to the data in addition.

91 We use the jet mass ( $M_{SD}$ ) after the soft drop grooming [? ]. The Moriond17 version of the  
92 Puppi soft drop mass correction [4] are applied.

### 93 **3.4 Lepton selection**

#### 94 **3.4.1 Electron and muon veto for the QCD multijet sample**

95 Electron candidates are identified via a set of selection criteria established by the EGamma POG  
96 based on “Spring15” simulated samples in the 25ns bunch spacing scenario [? ]. The “Veto”

97 working point is used.

98 The loose muon definition recommended by the Muon POG [? ] is used for the purposes of  
 99 the muon veto. Only candidates with transverse (longitudinal) impact parameter  $|d_0| < 0.2$  cm  
 100 ( $|d_z| < 0.5$  cm), with respect to the primary vertex, are considered.

101 The isolation requirements used in the selection of electrons and muons aim to achieve a high  
 102 efficiency for correctly identifying events that contain prompt leptons (i.e. leptons that do not  
 103 originate from heavy flavor decays) in the boosted topologies and busy hadronic environments.  
 104 They are based on the mini-isolation quantity, which is a measure of the lepton's local isolation.  
 105 Mini-isolation is computed as the summed  $p_T$  of PF candidates within a  $\Delta R$  cone centered on  
 106 the lepton candidate. The cone size depends on the lepton  $p_T$  as indicated in Table 5 and is  
 107 intended to be small enough to reduce overlaps with jets in the event while also being large  
 108 enough to contain the products of leptonic b-decays. Higher values of instantaneous luminos-  
 109 ity result in a reduced efficiency for the isolation requirement due to an increase in the number  
 110 of particles originating from additional (pileup) interactions that enter the isolation cone. In  
 111 order to minimize this effect, the calculated isolation quantity is corrected for the estimated  
 112 contribution from pileup particles. The correction is applied by subtracting the product of the  
 113 estimated average pileup density ( $\rho$ ) in the event with an effective area,  $A_{eff}$ , related to the  
 114 geometrical size of the isolation cone. The residual dependence of the isolation quantity on  
 115  $\rho$  for a given cone size is accounted for in  $A_{eff}$ , which is determined by taking the slope of a  
 116 linear fit to the uncorrected isolation as a function of  $\rho$ . The correction factor is  $\eta$ -dependent,  
 117 and calibrated separately for charged particles, neutral hadrons, and photons contributing to  
 118 the isolation.

119 Electrons and muons are considered to fulfill the veto isolation criteria if their mini-isolation is  
 120 less than 0.1 or 0.2 respectively, relative to the lepton  $p_T$ . Only candidates with  $p_T > 5$  GeV, and  
 121  $|\eta| < 2.4$  are considered for the veto.

Table 5: Lepton  $p_T$ -dependent  $\Delta R$  cone sizes used for mini-isolation computation.

Lepton $p_T$ range	Cone size
$p_T \leq 50$ GeV	0.2
$50 \text{ GeV} < p_T < 200 \text{ GeV}$	$\frac{10 \text{ GeV}}{p_T}$
$p_T \geq 200 \text{ GeV}$	0.05

### 122 3.4.2 Muon selection for the $t\bar{t}$ enriched sample

123 Tighter muon selections are defined for selecting events for the  $t\bar{t}$  enriched sample. The muons  
 124 are required to pass the tight ID recommended by the Muon POG. Only candidates with trans-  
 125 verse (longitudinal) impact parameter  $|d_0| < 0.2$  cm ( $|d_z| < 0.5$  cm), with respect to the primary  
 126 vertex, are considered. A tighter isolation, requiring the PF-based combined relative isolation  
 127 with  $\Delta\beta$  correction to be smaller than 0.05, is applied. The requirement on the  $p_T$  of the muon  
 128 is tightened to at least 45 GeV.

## 129 3.5 Photons

130 For the identification of photons we apply the medium working point of the EGamma POG  
 131 Spring15 25ns photon identification criteria [? ]. We select photons with  $p_T > 200$  GeV and  
 132  $|\eta| < 2.5$  and require the photon to pass the electron veto `passElectronVeto()`.

133 **3.6 Missing transverse energy**

134 The raw PF  $p_T^{\text{miss}}$  is computed as the negative vectorial sum of the transverse momenta of all  
135 PF candidates. Type-1 corrections [?] provided by the JetMET POG are applied to the  $p_T^{\text{miss}}$ .

136 **3.7 MET filters**

137 The following filters [?], recommended by the JetMET POG, are applied:

- 138 • HBHE noise filter
- 139 • HBHEiso noise filter
- 140 • EE bad SC noise filter
- 141 • ECAL dead cell trigger primitive filter
- 142 • CSC beam halo filter
- 143 • Bad PF muon filter
- 144 • Bad charged hadron filter

DRAFT

## 145 4 Boosted jet identification algorithm

146 This section describes the design of the new machine learning algorithm for boosted jet identifi-  
 147 cation. We start with how the truth label of each jet is defined in Section 4.1. The datasets and  
 148 the preprocessing procedure used in the development of the algorithm are described in Section  
 149 4.2. The input features to the algorithm are described in Section 4.3. The architecture of the  
 150 machine learning algorithm, as well as the training procedure, are discussed in Section 4.4.

### 151 4.1 Truth label definition

152 Well-defined truth labels are of vital importance to the development of a machine learning al-  
 153 gorithm for classification tasks. In the case of boosted jet identification, each jet needs to be  
 154 assigned a label by matching to the generated particles. With the aim of designing a com-  
 155 prehensive algorithm for boosted jet identification, we implemented a set of fine-grained labels for  
 156 boosted jets. The labels are organized in two levels, namely, “major” and “minor” labels. The  
 157 major labels define a broader category of jets, e.g. jets initiated by a top quark, a W, Z, Higgs  
 158 boson, or a light quark or gluon. The major label is defined if a jet is matched to a generated  
 159 heavy particle with  $\max \Delta R(\text{jet}, q_i) < 0.8$ , where  $q_i$  are all the quarks from the heavy particle  
 160 decays. In the case of top quarks, two scenarios are considered. The fully merged scenario  
 161 requires all three quarks are contained in the jet, while the partially merged scenario requires  
 162 only the b-quark and one of the quarks from the W boson are contained in the jet. Within each  
 163 major category, jets are further divided into minor categories based the decay modes with dif-  
 164 ferent flavour content, e.g.,  $Z \rightarrow bb$ ,  $Z \rightarrow cc$ ,  $Z \rightarrow qq$  in the Z boson category. For jets matched  
 165 to a heavy particle (i.e., a top quark, a W, Z, Higgs boson), the minor label is defined based on  
 166 the quarks from the heavy particle decay. For jets originating from a light quark or gluon, the  
 167 minor label is defined using the hadrons matched to the jets. The full list of major and minor  
 168 labels are summarized in Table 6. The minor labels of all categories are fully exclusive to each  
 169 other. In the case that a jet can be matched to more than one truth particles, priority is always  
 170 given to the label above the others in the table. The orthogonality of the labels allow for an easy  
 171 aggregation of different categories to target the needs of different analyses.

### 172 4.2 Datasets and preprocessing

173 The simulated samples used for the development of the algorithm are listed in Table 1. Top,  
 174 W, Z, and Higgs jets are taken from the corresponding signal samples, while the QCD jets  
 175 are taken only from the QCD multijet samples. To avoid biases from the difference in the jet  
 176  $p_T$  spectrum, jets in the training sample are reweighted to have a flat distribution in  $p_T$ , and  
 177 the contribution of each source (top, W, Z, Higgs, and QCD) is equalized. However, from our  
 178 studies we found that applying weights directly to the neural network training is likely to cause  
 179 degradation of performance or even failure in convergence especially when the weights vary  
 180 in a very large range. Alternatively, events can be removed from the training sample to achieve  
 181 the reweighting, but at a price of reducing the effective size of the training dataset. As a result,  
 182 we carry out reweighting “on-the-fly” by randomly sampling the training dataset according to  
 183 the “flattening” weights, thus effectively achieves the reweighting without losing events.

To help convergence and improve the stability of the neural network, all the input features are “standardized” such that they all fall in similar ranges. The most commonly used standardization method uses the mean and the standard deviation of the features. However, some of the variables we used have long tails in the distribution and sometimes extreme outliers, thus we use the percentiles for the standardization to be more robust. Therefore, each input variable  $x$

Table 6: Summary of the truth label definition for the boosted jet identification algorithm.

Major label	Minor label
Higgs	$bb$ $cc$ $(VV^* \rightarrow) qqqq$
top (fully merged)	$b_{cq}$ $b_{qq}$ $bc$ $bq$
top (partially merged)	
W	$cq$ $qq$
Z	$bb$ $cc$ $qq$
QCD	$bb$ $cc$ $b$ $c$ others

is transformed according to Eq. (1),

$$x^I = \frac{x - p_{50\%}}{p_{84\%} - p_{50\%}}, \quad (1)$$

where  $p_{50\%}$  and  $p_{84\%}$  are the 50th and 84th percentiles of the variable  $x$ . In the case when  $p_{84\%} = p_{50\%}$  (which happens for some discrete variables), the denominator is taken to be 1. The use of percentiles instead of the mean and the standard deviation tends to be less sensitive to outliers and distributions with long tails, leading to more unified scales for different variables. The transformed values are further clipped to be in the range of  $[-5, 5]$  before feeding into the neural networks, which are found to help improve the stability of neural network training.

### 4.3 Features

The PF reconstruction algorithm in CMS provides excellent description of individual particles constituting a jet. In principle, the measured properties of these constituent particles contain the complete information of a reconstructed jet, and an algorithm fully utilizing these particle-level measurements is expected to be powerful for jet identification. Motivated by this, a new algorithm, “DeepAK8”, is developed to exploit such particle-level information directly with customized machine learning methods. In the DeepAK8 algorithm, two lists of inputs are defined for each jet. The first list (“particle” list) consists of up to 100 jet constituent particles, organized in decreasing transverse momentum order. Measured properties of each particle, such as the momentum, the energy deposit, the charge, the angular separation between the particle and the jet axis or the subjet axes, etc., are included to help the algorithm extract features related to the substructure of a jet. For charged particles, additional information measured by the tracking detector are also included, such as the displacement and quality of the tracks, etc. These inputs are particularly useful for the algorithm to extract features related to the presence of heavy flavor ( $b$  or  $c$ ) quarks. In total, 42 variables are included for each particle in the “particle” list. The second list (“SV” list) consists of up to 7 secondary vertices (SV) and includes 15

206 features for each SV, such as the SV kinematics, the displacement, and quality criteria. The SV  
 207 list provides additional contribution to extracting features related to the heavy flavor content  
 208 of the jet. The full list of input variables used in the algorithm is listed in Table 7 and 8 for the  
 209 Particle list and the SV list, respectively.

## 210 4.4 Architecture

211 A significant challenge posed by the direct use of particle-level information is a substantial in-  
 212 crease in the number of inputs. Meanwhile, the correlations between these inputs are of vital  
 213 importance, thus an algorithm that can both process the inputs efficiently and exploit the cor-  
 214 relations effectively is needed. A customized Deep Neural Network (DNN) architecture is thus  
 215 developed in DeepAK8 to fulfill this requirement. As illustrated in Figure 1(a), the architecture  
 216 consists of two steps. In the first step, two convolutional neural networks (CNN) are applied  
 217 to the “particle” list and the “SV” list in parallel to transform the inputs and extract useful fea-  
 218 tures. Then, in the second step, the outputs of the two CNNs are merged and processed by  
 219 a simple fully-connected network to perform the jet classification. The CNN structure in the  
 220 first step is based on the ResNet model [5, 6], but adapted from two-dimensional images to  
 221 one-dimensional particle lists. The CNN for the “particle” list has 13 layers and the one for the  
 222 “SV” list has 9 layers. A convolution window of length 3 is used, and the number of output  
 223 channels in each convolutional layer ranges between 32 to 128. The ResNet architecture allows  
 224 for an efficient training of deep CNNs, thus leading to a better exploitation of the correlations  
 225 between the large inputs and improving the performance. With the CNNs in the first step  
 226 being already very powerful, the fully-connected network in the second step is fairly simple:  
 227 it consists of only one layer with 512 units, followed by a ReLU [7] activation function and  
 228 a DropOut [8] layer of 20% drop rate. The neural network is implemented using the MXNet  
 229 package [9] and trained with the Adam [10] optimizer to minimize the cross-entropy loss. The  
 230 initial learning rate is set to 0.001 and then reduced by a factor of 10 at the 10th and 20th epochs  
 231 to improve convergence. The training is stopped after 35 epochs. A sample of 50 million jets is  
 232 used, of which 80% are used for training and 20% are used for development and validation.

### 233 4.4.1 A mass-decorrelated version of DeepAK8

234 As discussed will be discussed in Section 5.3, background jets selected by the DeepAK8 algo-  
 235 rithm shows a modified mass distribution similar to that of the signal. This is because the mass  
 236 of a jet is one of the most discriminating variable, and although it is not directly used as an  
 237 input to the algorithm, the CNNs are able to extract features that are correlated to the mass to  
 238 improve the discrimination power. However, such modification of the mass distribution may  
 239 be undesirable if the mass variable itself is in use for separating signal and background pro-  
 240 cesses. Thus, an alternative DeepAK8 algorithm, “decorrelated DeepAK8”, is developed to be  
 241 largely decorrelated with the mass of a jet while preserving the discrimination power as much  
 242 as possible using the adversarial training approach [11].

243 The architecture of the decorrelated DeepAK8 algorithm is shown in Figure 1(b). Compared to  
 244 the base version of DeepAK8, a mass prediction network is added with the goal of predicting  
 245 the mass of a jet from the features extracted by the CNNs. When properly trained, the mass pre-  
 246 diction network becomes a good indicator of how strongly the features extracted by the CNNs  
 247 are correlated with the mass of a jet, as the stronger the correlation is, the more accurate the  
 248 mass prediction will be. With the introduction of the mass prediction network, the training tar-  
 249 get of the algorithm can be modified to include the accuracy of the mass prediction as a penalty,  
 250 therefore preventing the CNNs from extracting features that are correlated with the mass. In  
 251 this way, the final prediction of the algorithm also becomes largely independent of the mass.

Table 7: Full list of input variables for each jet constituent particle.

variable	expression (default value)
From <code>pat::PackedCandidate</code>	
$\log p_T$	
$\log(p_T / p_T(\text{jet}))$	
$\log E$	
$ \eta $	
$\Delta\phi(\text{jet})$	
$\Delta\eta(\text{jet})$	
Puppi weight	
$\min \Delta R(\text{SV})$	
$\Delta R(\text{jet})$	
$\Delta R(\text{subjet 1})$	
$\Delta R(\text{subjet 2})$	
$q$ (electric charge)	
isMuon	
isElectron	
isPhoton	
isChargedHadron	
isNeutralHadron	
hcalFraction	
pvAssociationQuality	
lostInnerHits	
$d_{xy}$	
$d_z$	
From <code>cand-&gt;bestTrack()</code>	
Default value in the parenthesis are assigned if <code>cand-&gt;bestTrack() == nullptr</code> .	
$\sigma_{d_{xy}}$	$d_{xy}() / d_{xy}\text{Error}() (0)$
$\sigma_{d_z}$	$d_z() / d_z\text{Error}() (0)$
$\chi_n^2$	(999)
qualityMask	(0)
dptdpt	covariance(0, 0) (0)
detadeta	covariance(1, 1) (0)
dphidphi	covariance(2, 2) (0)
dxydxy	covariance(3, 3) (0)
dzdz	covariance(4, 4) (0)
dxydz	covariance(3, 4) (0)
dphidxy	covariance(2, 3) (0)
dlambdadz	covariance(1, 4) (0)
BTW variables	
trackEtaRel	(0)
trackPtRatio	(0)
trackPParRatio	(0)
trackSip2dVal	(0)
trackSip2dSig	(0)
trackSip3dVal	(0)
trackSip3dSig	(0)
trackJetDistVal	(0)

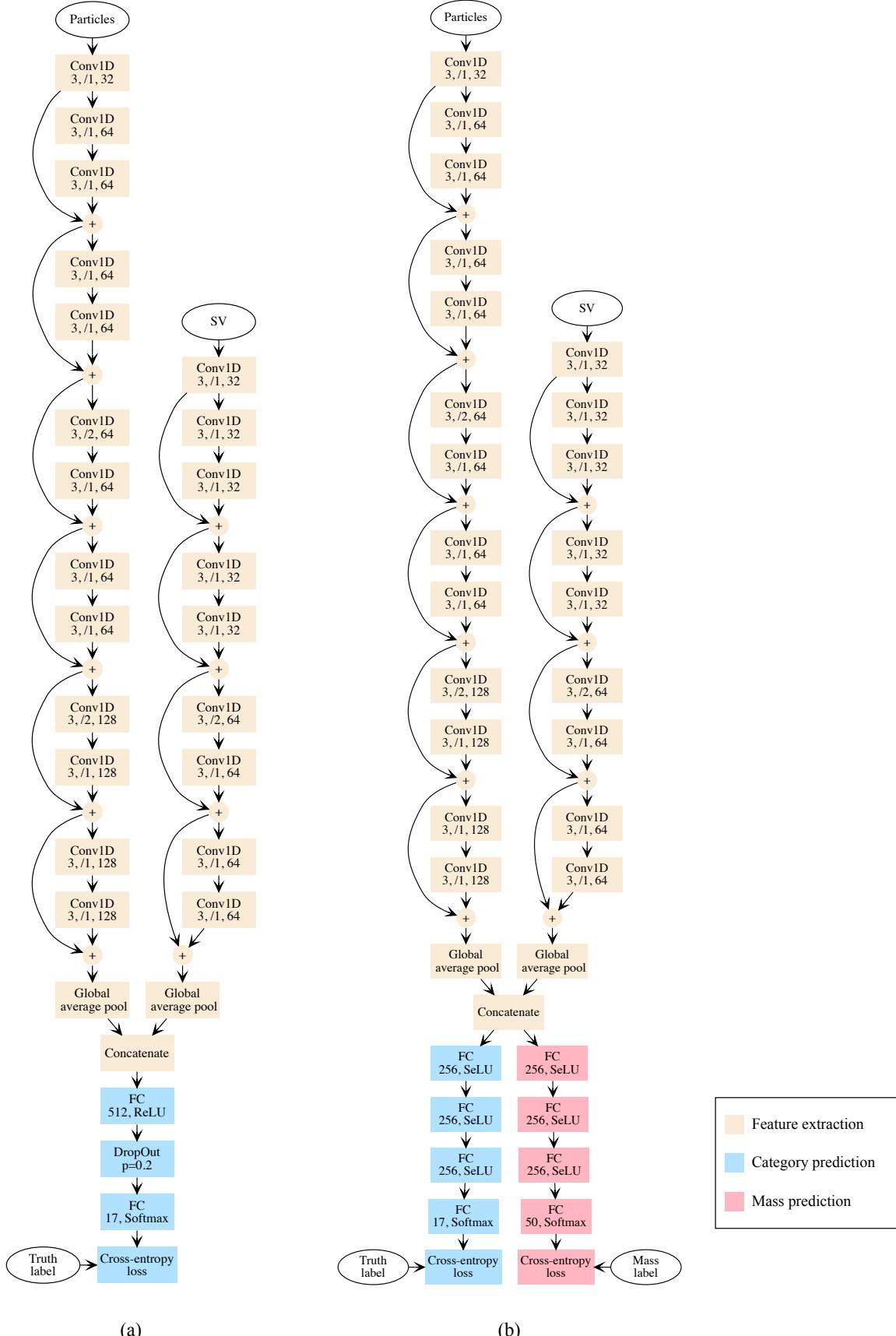


Figure 1: The network architecture of (a) DeepAK8 and (b) DeepAK8-MD.

Table 8: Full list of input variables for each secondary vertex inside the jet.

variable
$\log p_T$
$\log(p_T/p_T(\text{jet}))$
$\log E$
$ \eta $
$\Delta\phi(\text{jet})$
$\Delta\eta(\text{jet})$
$\Delta R(\text{jet})$
$m_{\text{SV}}$
$N_{\text{tracks}}$
$\chi_n^2$
$d_{xy}$
$\sigma_{d_{xy}}$
$d_{3D}$
$\sigma_{d_{3D}}$
$\cos(\overrightarrow{(\text{PV}, \text{SV})}, \vec{p}_{\text{SV}})$

252 As the features extracted by the CNNs evolve during the training process, the mass prediction  
 253 network itself needs to be updated from time to time to adapt to the changes of its inputs and  
 254 remain as an effective indicator of mass correlation. Since the mass of a jet is one of the most  
 255 discriminating variable, forcing the algorithm to be decorrelated with it inevitably leads to a  
 256 loss of discrimination power, and the resulted algorithm is a balance between performance and  
 257 mass-independence.

## 258 5 Performance in simulation

We studied the performance of the DeepAK8 and DeepAK8-MD algorithms with simulated events following the selections and samples listed on [1]. To avoid biases from the difference in the  $p_T$  spectrum of different samples, events in the signal samples are reweighted to have the same distribution in  $p_T$  as in the QCD samples. The performance of the algorithms is evaluated using the signal and background efficiency,  $\epsilon_S$  and  $\epsilon_B$ , respectively, as a figure of merit. The  $\epsilon_S$  and  $\epsilon_B$  are defined as:

$$\epsilon_S = \frac{N_S^{\text{tagged}}}{N_S^{\text{total}}} \quad \text{and} \quad \epsilon_B = \frac{N_B^{\text{tagged}}}{N_B^{\text{total}}}, \quad (2)$$

259 where  $N_{S(B)}^{\text{tagged}}$  is the number of signal (background) truth-particles identified by the algorithm,  
260 and  $N_{S(B)}^{\text{total}}$  the total number of signal (background) truth-particles passing the truth-level selec-  
261 tions of  $p_T > 300 \text{ GeV}$  and  $|\eta| < 2.4$ .

### 262 5.1 ROC curves

263 The performance of the new DNN-based algorithms are compared with traditional approaches  
264 in terms of receiver operating characteristic (ROC) curve. Figures 2-5 summarize the ROC  
265 curves of different algorithms for the identification of  $t$  quarks and  $W, Z$  or Higgs bosons. The  
266 comparisons are carried out in two regions based on the  $p_T$  of the truth particle. The kinematic  
267 selection criteria applied on the truth-particles are displayed on the plots.

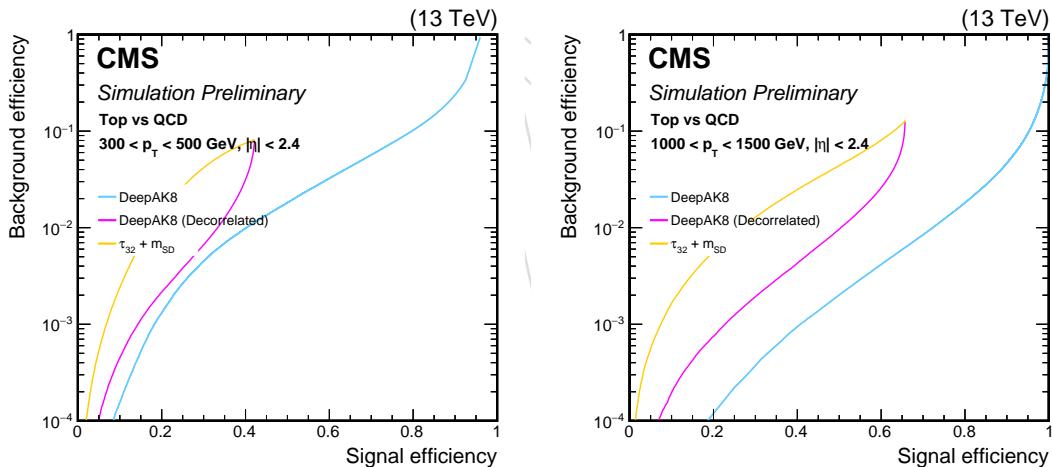


Figure 2: Performance comparison of the hadronically decaying  $t$  quark identification algorithms in terms of receiver operating characteristic (ROC) curves in two regions based on the  $p_T$  of the truth particle. Left:  $300 < p_T < 500 \text{ GeV}$ , and Right:  $1000 < p_T < 1500 \text{ GeV}$ .

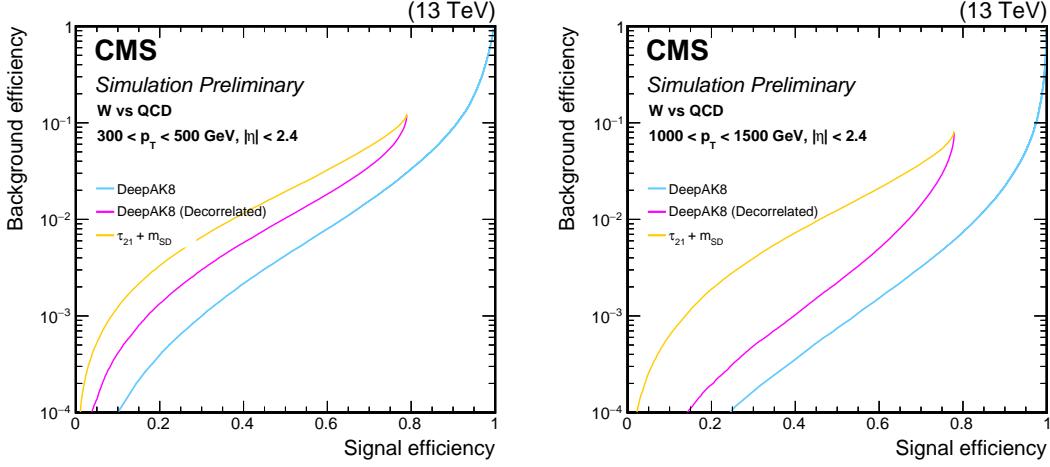


Figure 3: Performance comparison of the hadronically decaying W boson identification algorithms in terms of receiver operating characteristic (ROC) curves in two regions based on the  $p_T$  of the truth particle. Left:  $300 < p_T < 500 \text{ GeV}$ , and Right:  $1000 < p_T < 1500 \text{ GeV}$ .

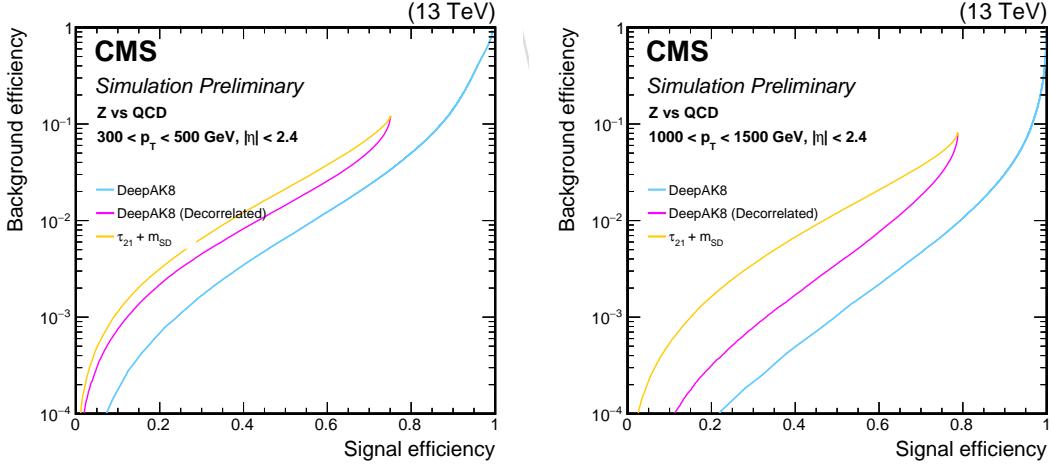


Figure 4: Performance comparison of the hadronically decaying Z boson identification algorithms in terms of receiver operating characteristic (ROC) curves in two regions based on the  $p_T$  of the truth particle. Left:  $300 < p_T < 500 \text{ GeV}$ , and Right:  $1000 < p_T < 1500 \text{ GeV}$ .

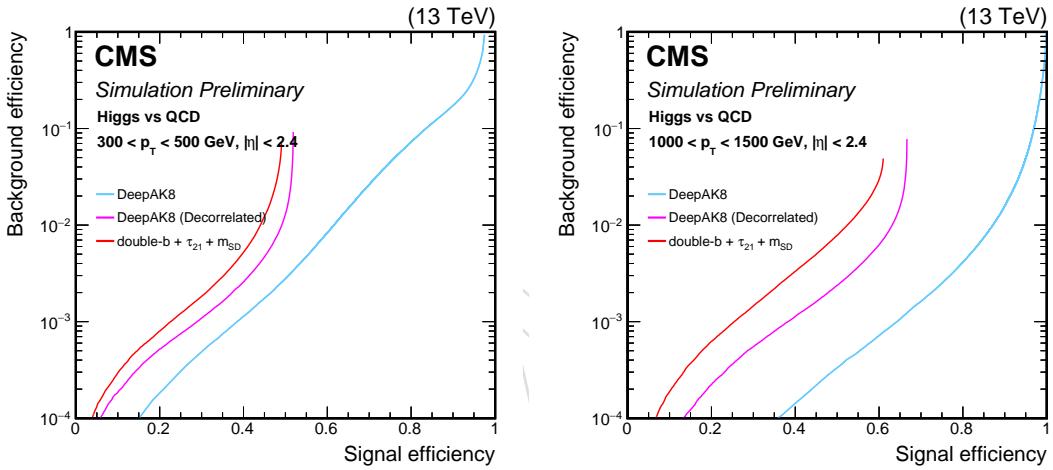


Figure 5: Performance comparison of the hadronically decaying Higgs boson identification algorithms in terms of receiver operating characteristic (ROC) curves in two regions based on the  $p_T$  of the truth particle. Left:  $300 < p_T < 500$  GeV, and Right:  $1000 < p_T < 1500$  GeV. The Higgs boson is forced to decay in a pair of b quarks.

**268 5.2 Efficiency vs  $p_T$  and NPV**

269 For the sake of the performance comparison of the various algorithms, a working point cor-  
270 responding to  $\epsilon_S = 50\%$  for truth particles with  $500 < p_T < 600 \text{ GeV}$  is defined and used  
271 throughout the following plots. The distributions of the  $\epsilon_S$  and  $\epsilon_B$  as a function of the  $p_T$  of the  
272 truth particle are displayed in Figs. 6 and 7, respectively.

273 The dependence of the algorithms on the number of reconstructed vertices is also examined  
274 using simulated events. Figure 8 (9) displays the distribution of  $\epsilon_S$  ( $\epsilon_B$ ) as a function of the  
275 number of reconstructed vertices for jets with  $500 < p_T(\text{truth particle}) < 1000 \text{ GeV}$ , operating  
276 at the  $\epsilon_S = 50\%$  working point defined above.

DRAFT

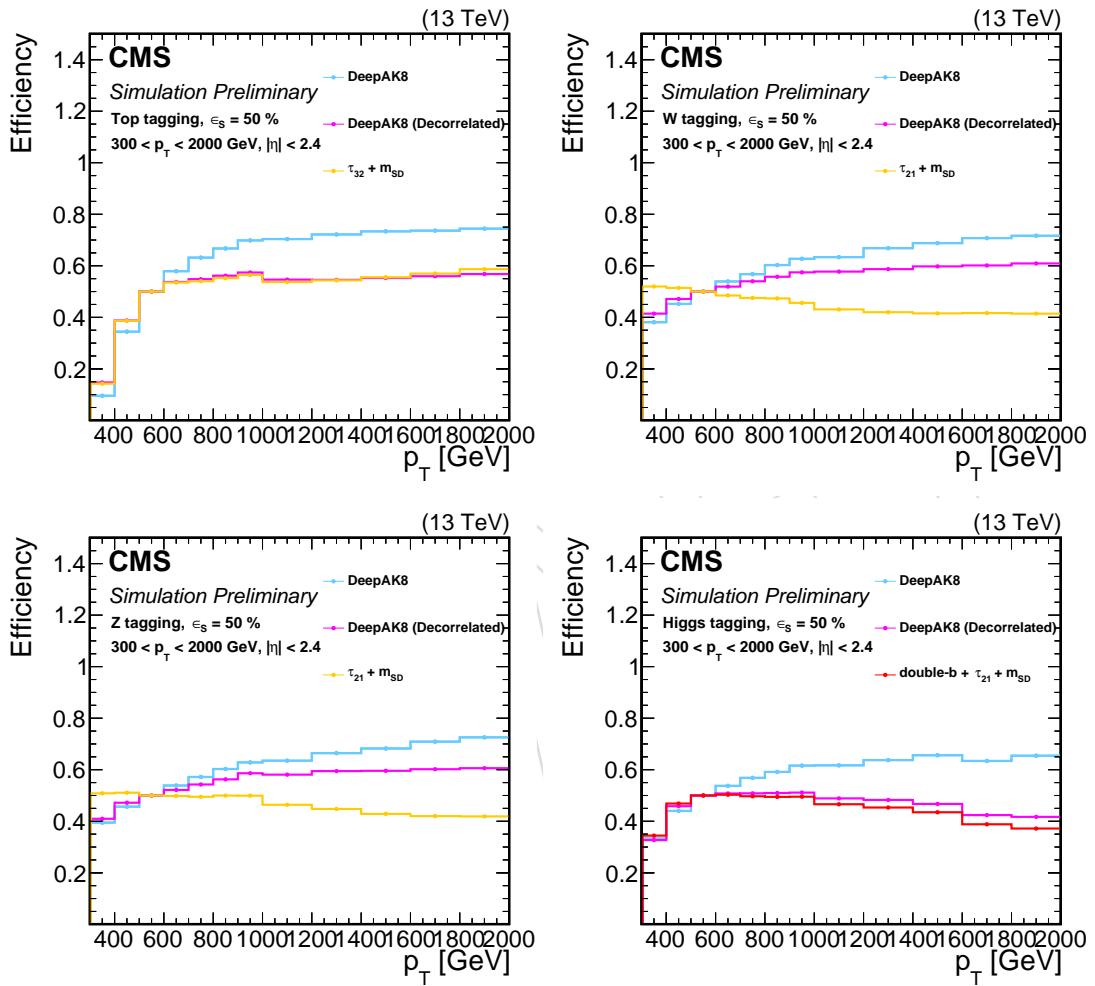


Figure 6: The distribution of  $\epsilon_S$  as a function of the  $p_T$  of the truth particle for a working point corresponding to  $\epsilon_S = 50\%$ . Upper left: t quark, upper right: W boson, bottom left: Z boson, bottom right: Higgs boson.

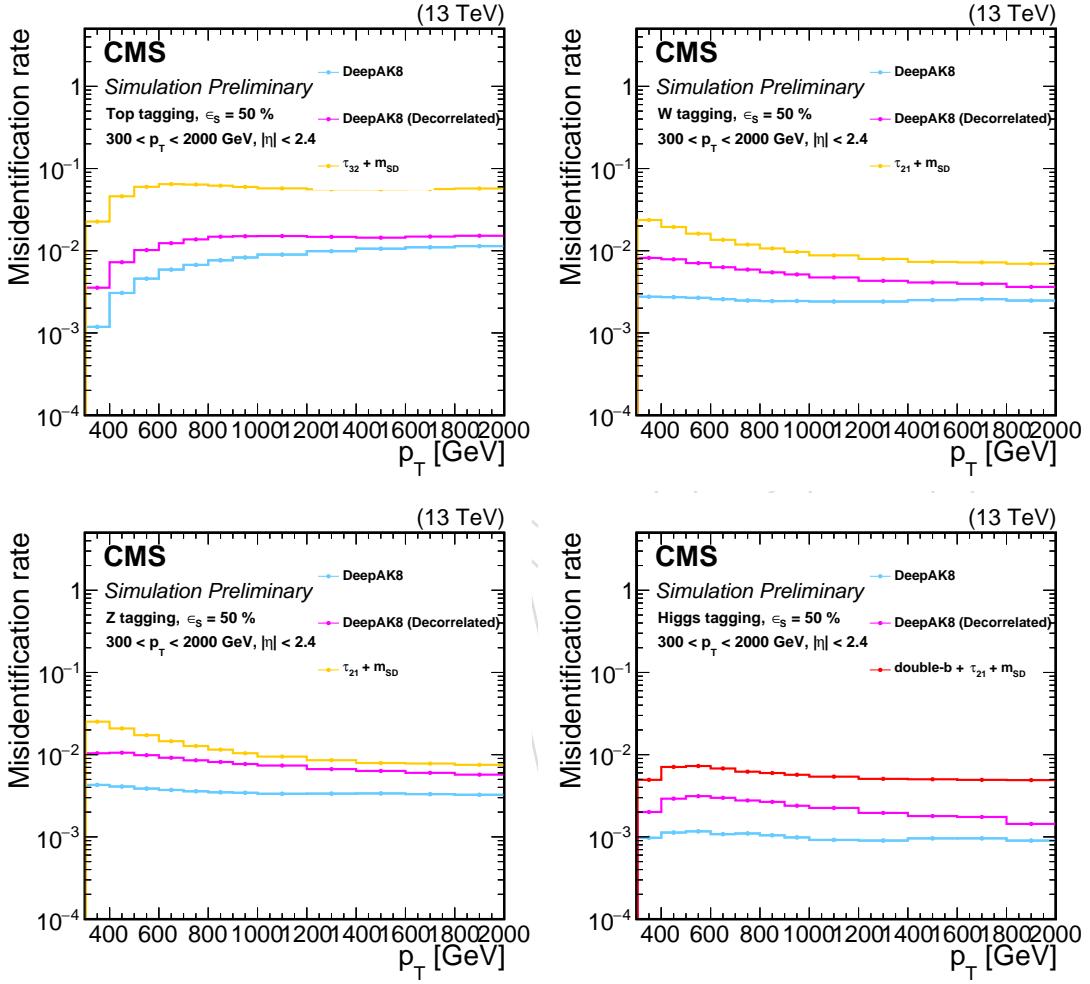


Figure 7: The distribution of  $\epsilon_B$  as a function of the  $p_T$  of the truth particle for a working point corresponding to  $\epsilon_S = 50\%$ . Upper left: t quark, upper right: W boson, bottom left: Z boson, bottom right: Higgs boson.

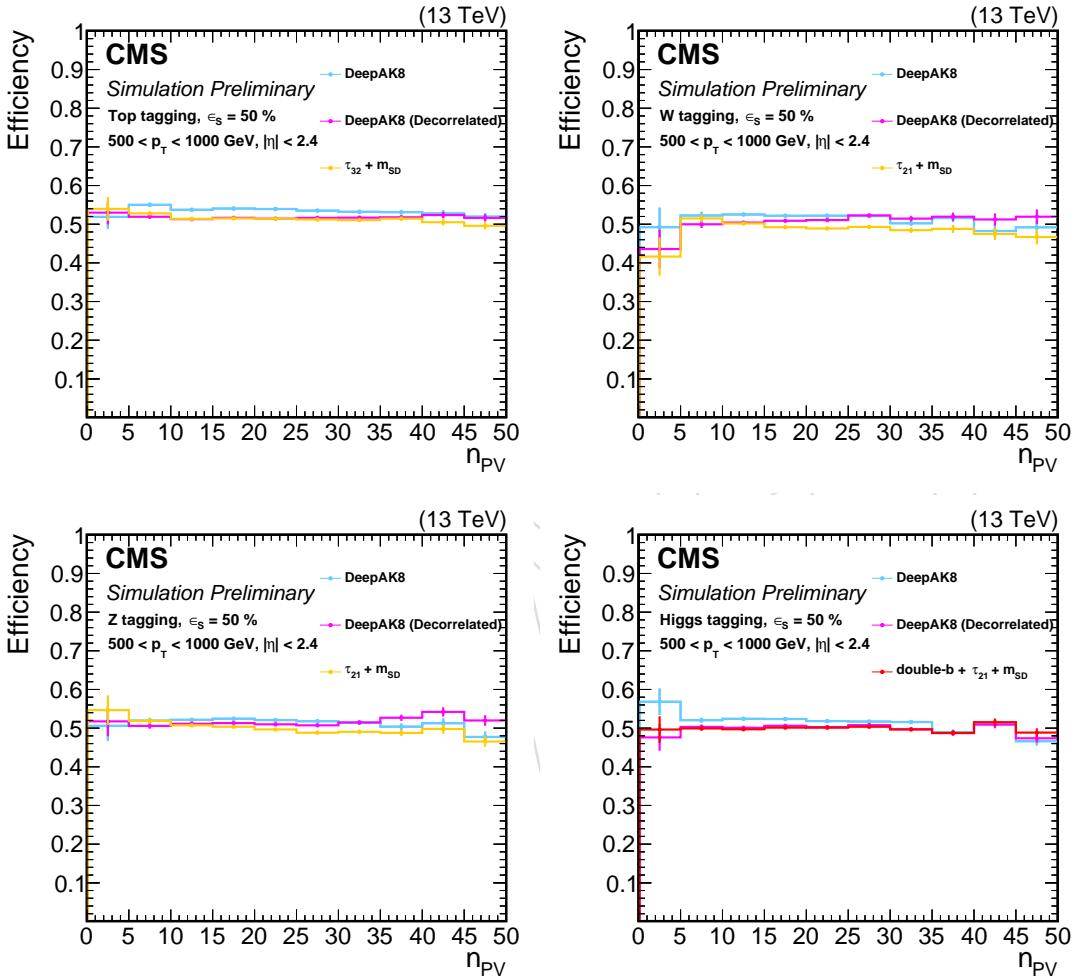


Figure 8: The  $\epsilon_S$  as a function of the number of reconstructed vertices for jets with  $500 < p_T(\text{truth particle}) < 1000$  GeV. The working point corresponds to  $\epsilon_S = 50\%$ . Upper left: t quark, upper right: W boson, bottom left: Z boson, bottom right: Higgs boson.

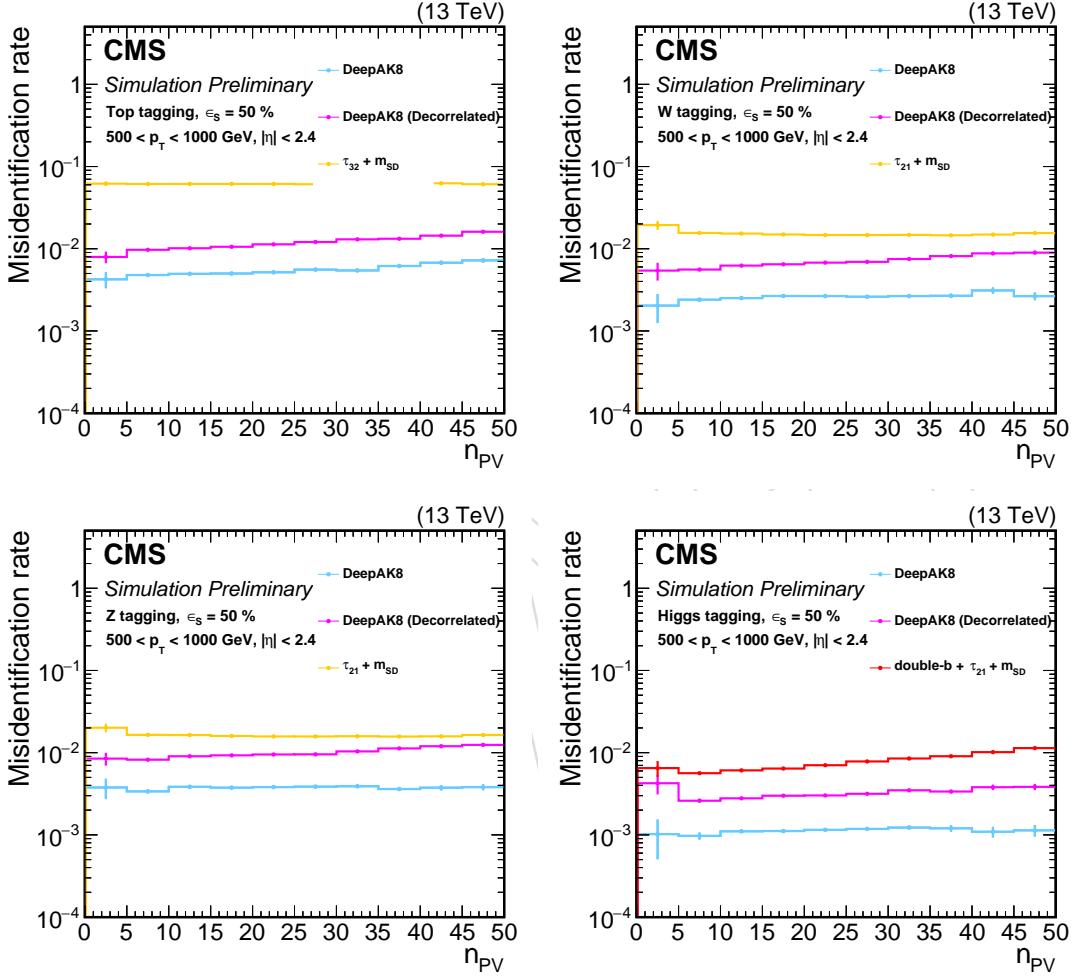


Figure 9: The  $\epsilon_B$  as a function of the number of reconstructed vertices for jets with  $500 < p_T(\text{truth particle}) < 1000$  GeV. The working point corresponds to  $\epsilon_S = 50\%$ . Upper left: t quark, upper right: W boson, bottom left: Z boson, bottom right: Higgs boson.

277 **5.3 Impacts on the jet mass shape of the background events**

278 Finally, a set of studies carried out to understand the correlation of the algorithms with tra-  
279 ditional jet substructure variables, such as the jet mass. The jet mass is one of the most dis-  
280 criminating variables and a large number of analyses rely on the assumption of a smooth back-  
281 ground jet mass spectrum under a signal peak. Figure 10 displays the shape of the softdrop  
282 mass distribution for background jets, inclusively and after applying a selection on each al-  
283 gorithm. The working point chosen corresponds  $\epsilon_S = 50\%$ . The results are displayed for a  
284  $p_T$  region of 500 – 1000 GeV of the truth particle distribution. By design, the base version of the  
285 DeepAK8 algorithms, features significant sculpting of the background jet mass shape.



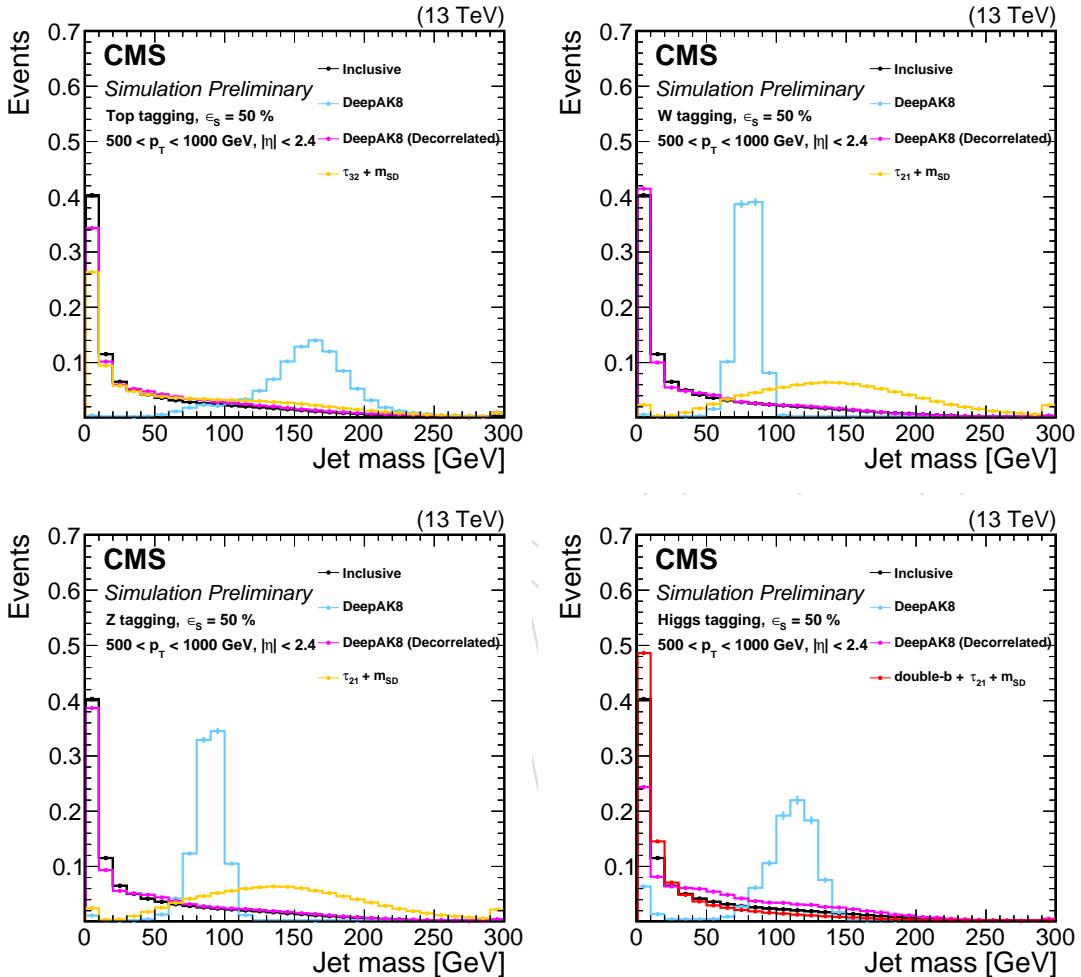


Figure 10: The shape of the softdrop mass distribution for background jets with  $500 < p_T(\text{truth particle}) < 1000 \text{ GeV}$ , inclusively and after applying a selection on each algorithm. The working point chosen corresponds to  $\epsilon_s = 50\%$ . Upper left: t quark, upper right: W boson, bottom left: Z boson, bottom right: Higgs boson.

## 286 6 Event selection

287 Several event samples are utilized to validate the performance of the tagging algorithms in  
 288 data. The semi-leptonic  $t\bar{t}$  signal sample is used to calibrate the  $t$  and  $W$  tagging performance  
 289 in a sample enriched in hadronically decaying  $t$  quarks. The QCD multijet sample is domi-  
 290 nated by light-flavor quarks and gluons, which consist background for these algorithms. The  
 291 largeness of the sample enables the study of the identification probability of background jets  
 292 (mistagging rate) in a wide range of  $p_T$ . The mistagging rate depends on the flavor of the parton  
 293 that initiated the jet. Therefore, in addition to the QCD multijet sample, the  $\gamma$ +jet background  
 294 sample is utilized. The QCD multijet and  $\gamma$ +jet samples differ in the light-quark and gluon  
 295 fractions. In the former, jets are typically initiated by gluon jets, whereas in the latter by light  
 296 quarks.

### 297 6.1 The $t\bar{t}$ ( $1\ell$ ) signal sample

298 The semi-leptonic  $t\bar{t}$  signal sample is recorded using a single muon trigger which selects on the  
 299  $p_T$  and the isolation of the muon. Candidate events are required to have a muon as defined in  
 300 Section 3. To suppress contribution from QCD multijet processes we require  $p_T^{\text{miss}} > 50$  GeV.  
 301 To enhance the sample purity in  $t\bar{t}$  events, we require the presence of two or more slim jets, at  
 302 least one of which is reconstructed as b jet. In addition, to probe boosted topologies we require  
 303 the  $p_T$  of the leptonically decaying  $W$ , defined as  $\vec{p}_T(W) = \vec{p}_T(\mu) + \vec{p}_T^{\text{miss}}$ , to be greater than  
 304 250 GeV. The event is divided into two hemisphere using the azimuthal angle  $\Delta\phi$  the muon. A  
 305 b-jet is required to be present in the leptonic hemisphere by satisfying  $\Delta\phi(\mu, b) < 2\pi/3$ . The  
 306  $t/W$  candidate is the highest  $p_T$  AK8 jet in the hadronic hemisphere with  $\Delta\phi(\mu, \text{AK8 jet}) >$   
 307  $2\pi/3$ .

### 308 6.2 The QCD multijet background sample

309 The QCD multijet background sample is recorder using a trigger that selects on  $H_T$ . To satisfy  
 310 the trigger requirements, events with  $H_T > 1000$  GeV are selected. Events are required to have  
 311 at least one wide jet, and the absence of electrons or muons, leading to a sample dominated by  
 312 background jets.

### 313 6.3 The $\gamma$ +jet background sample

314 The  $\gamma$ +jet background sample is selected using an isolated single-photon trigger. To meet  
 315 the trigger requirements we require the presence of a photon satisfying the criteria presented  
 316 in Section 3. In addition to the photon, the  $\gamma$ +jet sample is required to have at least one  
 317 wide jet and no electrons or muons. The highest  $p_T$  AK8 jet opposite to the photon with  
 318  $\Delta\phi(\gamma, \text{AK8 jet}) > 2$  is selected as the probe jet.

## 319 7 Performance in data and systematic uncertainties

320 In this section the validation of the performance of the algorithms in data is presented. The  
 321 validation is carried out in two steps. In the first step, we focus on studying the overall model-  
 322 ing of key variables in simulation and their agreement with data, as well as the dependence on  
 323 the simulation details. The results from these studies are used for the design for further tuning  
 324 the algorithms, as well as to improve the simulation details. Moreover, these results are used  
 325 to correct the performance of the algorithms to match the performance in data. Differences in  
 326 the performance between data and simulation are taken into account by means of scale factors  
 327 (SF) extracted by comparing the efficiency in data and MC. To account for effects not captured  
 328 in the SF, multiple sources of systematic uncertainties are considered. The data samples used  
 329 for these studies are described in Section 2.3.

### 330 7.1 The t quark and W boson identification performance in data

331 The single- $\mu$  event selection discussed in Section 6.1 provides a sample dominated by  $t\bar{t}(1\ell)$   
 332 events. The leptonically decaying W bosons stemming from one of the t quarks provides a  
 333 final state with high purity in  $t\bar{t}$  events, whereas the other t quark (which decays hadronically)  
 334 provides boosted hadronically decaying t quarks and W bosons, necessary for the validation  
 335 of the algorithms.

336 The single- $\mu$  selection corresponds to approximately 80%  $t\bar{t}$  events, and the remaining fraction  
 337 consists mainly of W+jets and single-t events. The  $t\bar{t}$  sample is decomposed into three cate-  
 338 gories based on the spatial separation of the partons from the t quark decay with respect to the  
 339 wide jet. The “Merged t quark” category includes cases where all three partons are contained in  
 340 the jet with  $\Delta R < 0.8$ . In the “Merged W boson” category are cases where only the two partons  
 341 from the W boson decay are within  $\Delta R < 0.8$  of the jet, while the b-quark lies outside the jet.  
 342 Any other scenario falls in the “Unmerged” category.

343 The jet mass (obtained after applying the softdrop algorithm),  $m_{jet}$ , the jet  $p_T$  and the N-subjetiness  
 344 ratios,  $\tau_{32}$  and  $\tau_{21}$  are important observables for boosted jet identification. Figure 11 displays  
 345 the distribution of these variables in data and simulation for events passing the single- $\mu$  event  
 346 selection, in three regions of  $p_T$  (jet), “inclusive”, “ $300 < p_T(\text{jet}) < 400$ ” and “ $600 < p_T(\text{jet}) <$   
 347  $1200$ ”. To account for effects related to differences in the overall normalization between data  
 348 and simulation, the MC distributions are normalized to the number of data. The shapes in data  
 349 are compatible with the expectation from simulation.

350 The distributions of the t quark and W, Z, and Higgs boson identification probabilities for the  
 351 DeepAK8 algorithm are displayed in Fig. 12. Figure 13 displays the t quark, W boson and  $X \rightarrow$   
 352  $bb$  identification probabilities for the DeepAK8-MD algorithm. In all cases the comparisons are  
 353 shown in three regions of  $p_T$  (jet), “inclusive”, “ $300 < p_T(\text{jet}) < 400$ ” and “ $600 < p_T(\text{jet}) <$   
 354  $1200$ ”. The MC distributions are normalized to the number of data. Overall, we observe good  
 355 agreement on the shape of the distributions between data and MC for all variables.

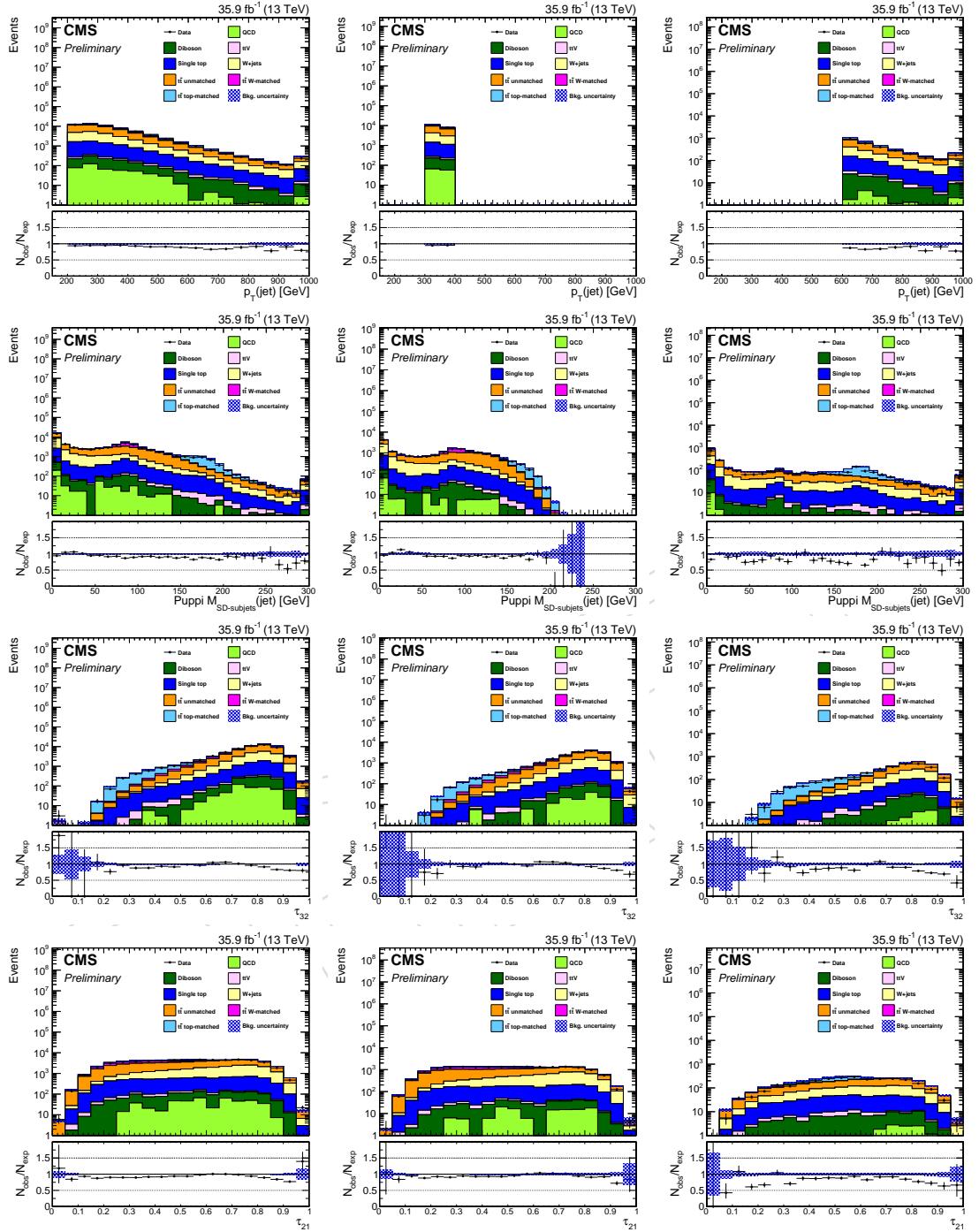


Figure 11: Distribution of the jet mass,  $m_{\text{jet}}$  (1st row), the jet  $p_T$  (2nd row), and the N-subjettiness ratios,  $\tau_{32}$  (3rd row) and  $\tau_{21}$  (4th row), in data and MC in the  $t\bar{t}(1\ell)$  signal sample. The distributions are shown in three regions of  $p_T(\text{jet})$ , “inclusive” (left), “ $300 < p_T(\text{jet}) < 400$ ” (middle) and “ $600 < p_T(\text{jet}) < 1200$ ” (right). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

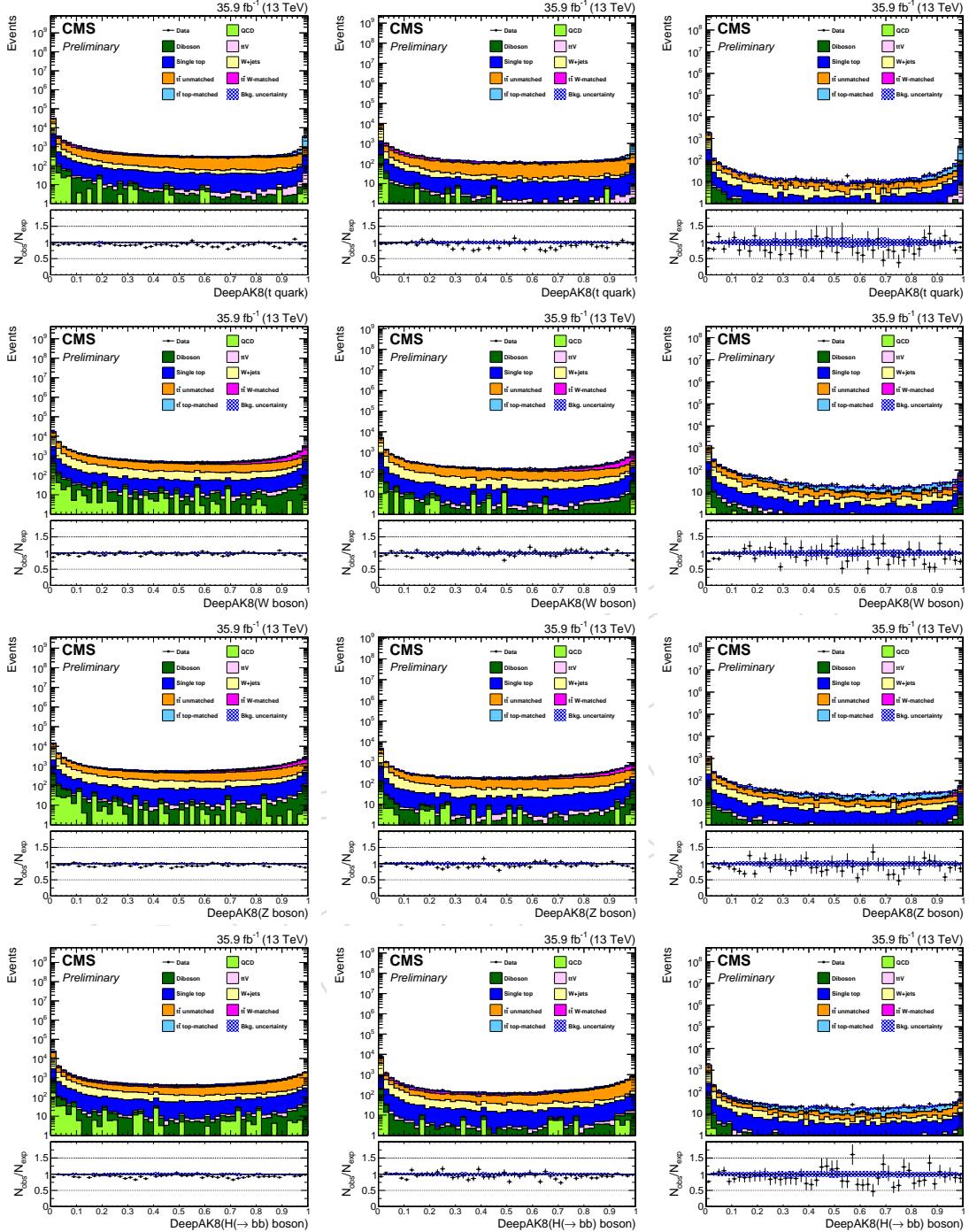


Figure 12: Distribution of the t quark, (1st row), the W boson (2nd row), the Z boson (3rd row), and the the Higgs boson (4th row) identification probabilities of DeepAK8 in data and MC in the  $t\bar{t}(1\ell)$  signal sample. The distributions are shown in three regions of  $p_T$  (jet), “inclusive” (left), “ $300 < p_T(\text{jet}) < 400$ ” (middle) and “ $600 < p_T(\text{jet}) < 1200$ ” (right). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

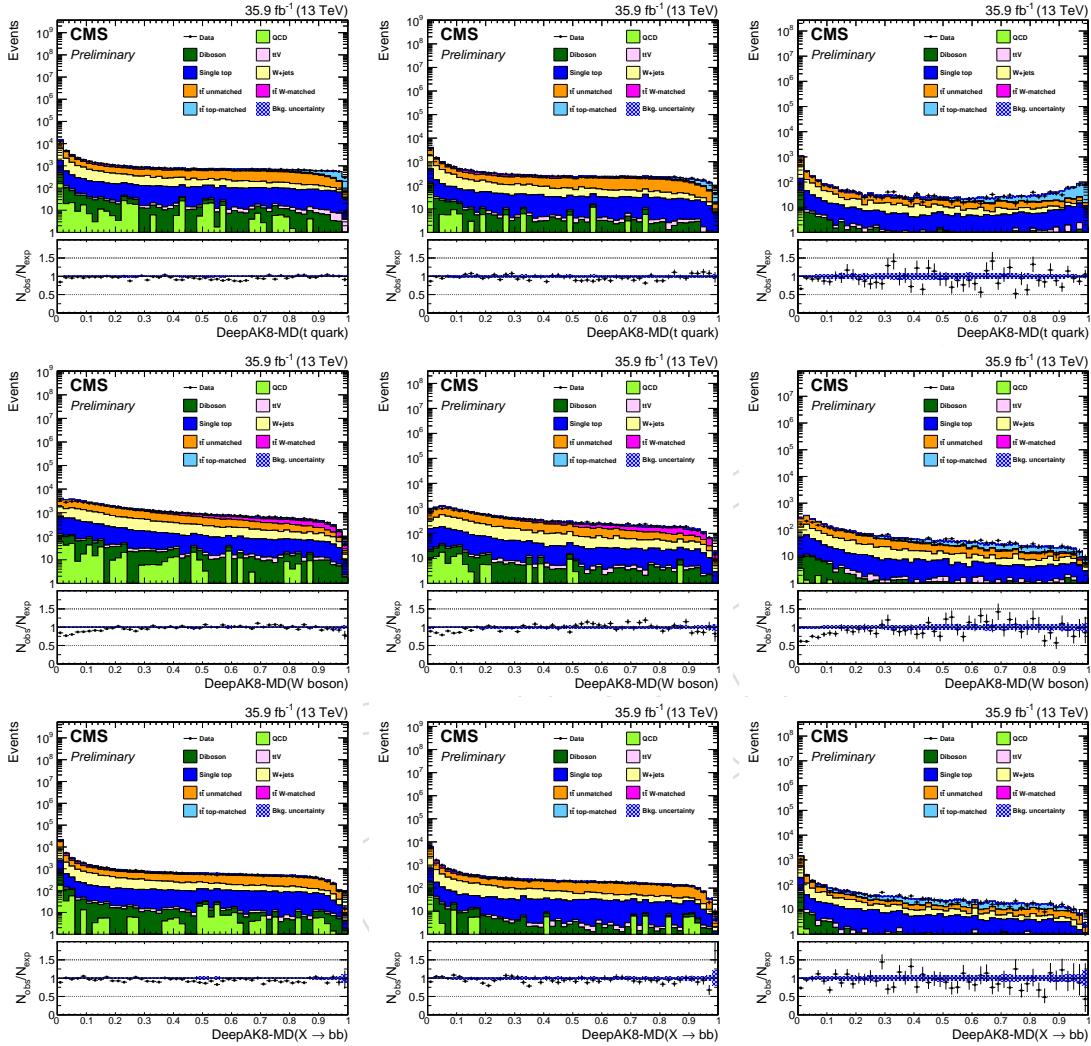


Figure 13: Distribution of the  $t$  quark, (1st row), the  $W$  boson (2nd row), and the  $X \rightarrow bb$  (3rd row) identification probabilities of DeepAK8-MD in data and MC in the  $t\bar{t}(1\ell)$  signal sample. The distributions are shown in three regions of  $p_T$ (jet), “inclusive” (left), “ $300 < p_T(\text{jet}) < 400$ ” (middle) and “ $600 < p_T(\text{jet}) < 1200$ ” (right). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

## 356 7.2 Misidentification probability in the QCD multijet sample

357 The efficiency of the algorithms on identifying light quarks and gluons (“misidentification  
358 probability”) is studied in the QCD multijet sample. In order to study the dependence of  
359 the misidentification probability on the choice of the event generator and the parton show-  
360 ering scheme, we consider two different MC samples to model the QCD multijet background.  
361 The nominal sample uses MADGRAPH for the event generation and PYTHIA8 for the par-  
362 ton showering and hadronization, whereas the alternative sample uses PYTHIA8 for the event  
363 generation and uses HERWIG++ for the modeling of the parton showering. More information  
364 on the generation details on these samples are discussed in Section 2.

365 The distribution of  $m_{\text{jet}}$ , jet  $p_{\text{T}}$  and the N-subjetiness ratios,  $\tau_{32}$  and  $\tau_{21}$  in the QCD multijet  
366 sample are displayed in Fig. 14, in three regions of  $p_{\text{T}}(\text{jet})$ , “inclusive”, “ $300 < p_{\text{T}}(\text{jet}) < 400$ ”  
367 and “ $600 < p_{\text{T}}(\text{jet}) < 1200$ ”. The MC distributions are normalized to the number of data. The  
368 shape of the  $m_{\text{jet}}$  and  $p_{\text{T}}$  are better described by the alternative sample (i.e. HERWIG), whereas  
369 the N-subjetiness ratios are better described by the nominal sample (i.e. PYTHIA).

370 The distributions of the t quark and W, Z, and Higgs boson identification probabilities for  
371 the DeepAK8 algorithm are displayed in Fig. 15. Figure 16 displays the t quark, W boson  
372 and  $X \rightarrow bb$  identification probabilities for the DeepAK8–MD algorithm. In all cases the  
373 comparisons are shown in three regions of  $p_{\text{T}}(\text{jet})$ , “inclusive”, “ $300 < p_{\text{T}}(\text{jet}) < 400$ ” and  
374 “ $600 < p_{\text{T}}(\text{jet}) < 1200$ ”. The MC distributions are normalized to the number of data. We  
375 observe a fair agreement in shape between data and MC for most of the variables. The QCD–  
376 HERWIG sample in general describes better the data. Larger discrepancies, up to ( $\sim 25\%$ ),  
377 are observed for the DeepAK8–MD algorithm, which can be expected since it relies on other  
378 features but the mass to discriminate between the classes.

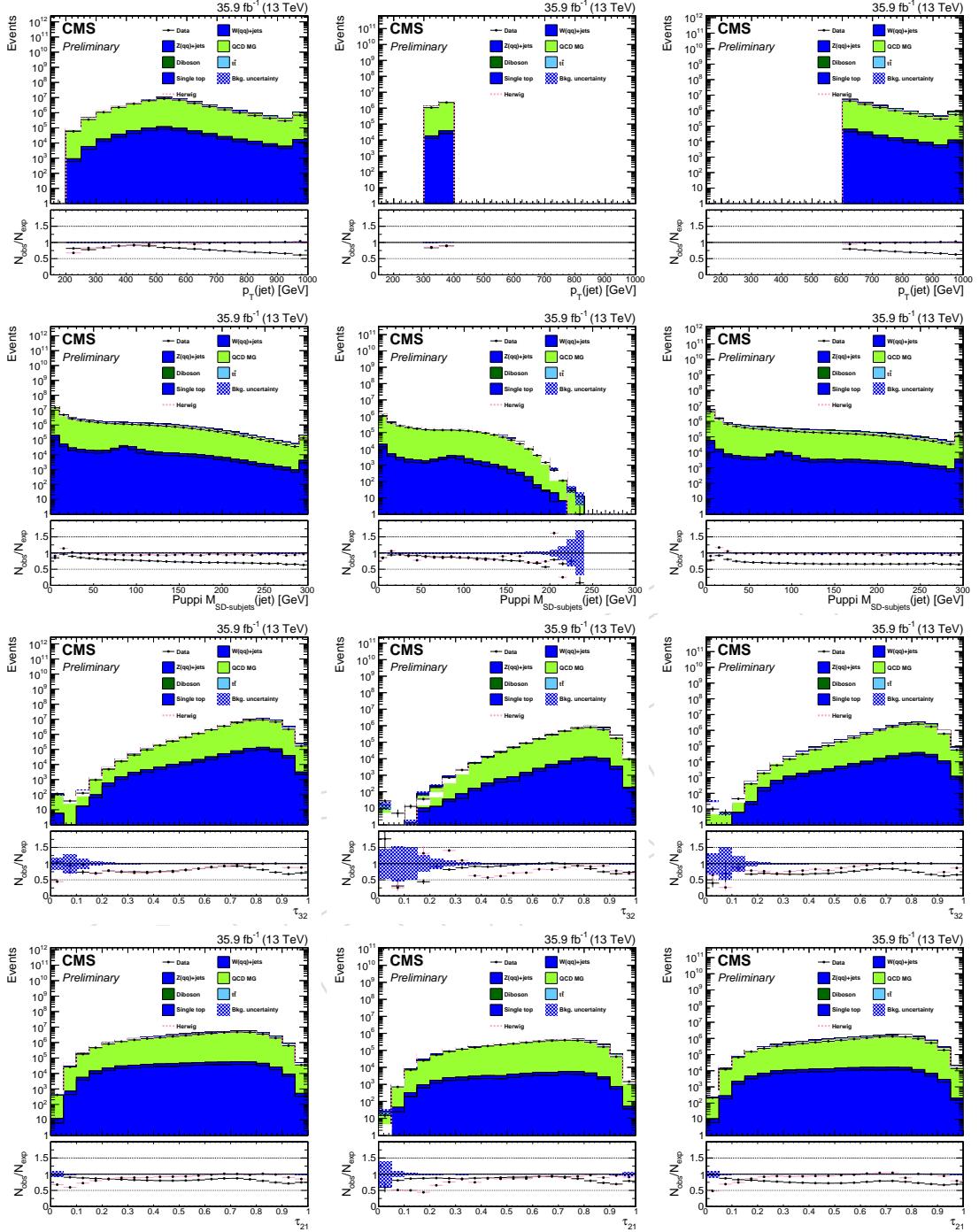


Figure 14: Distribution of the jet mass,  $m_{\text{jet}}$  (1st row), the jet  $p_{\text{T}}$  (2nd row), and the N-subjetiness ratios,  $\tau_{32}$  (3rd row) and  $\tau_{21}$  (4th row), in data and MC in the QCD multijet sample. The distributions are shown in three regions of  $p_{\text{T}}$  (jet), “inclusive” (left), “ $300 < p_{\text{T}}(\text{jet}) < 400$ ” (middle) and “ $600 < p_{\text{T}}(\text{jet}) < 1200$ ” (right). The pink color corresponds to the MC distribution after using the alternative sample (i.e. HERWIG). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

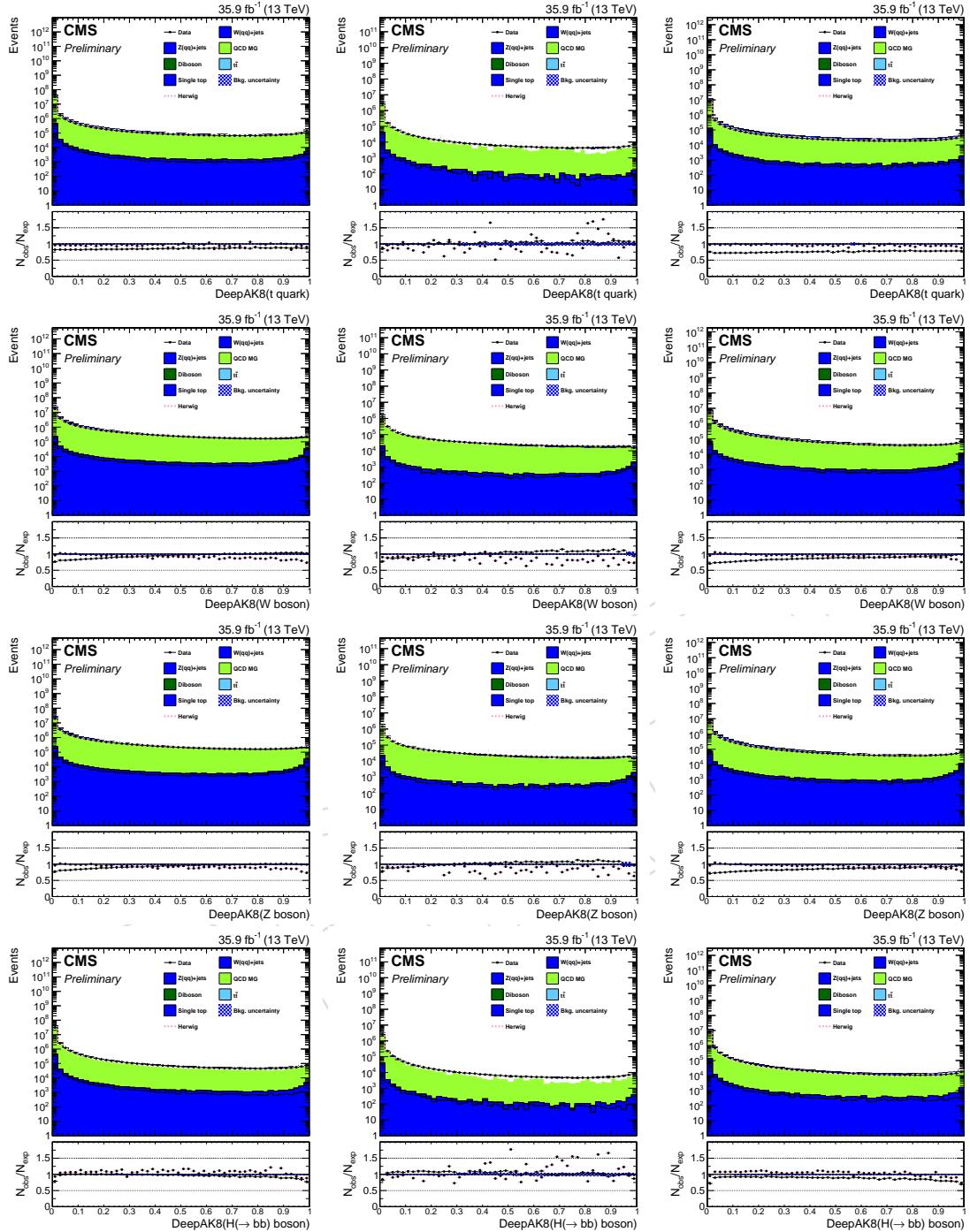


Figure 15: Distribution of the  $t$  quark, (1st row), the  $W$  boson (2nd row), the  $Z$  boson (3rd row), and the the Higgs boson (4th row) identification probabilities of DeepAK8 in data and MC in the QCD multijet sample. The distributions are shown in three regions of  $p_T$  (jet), “inclusive” (left), “ $300 < p_T(\text{jet}) < 400$ ” (middle) and “ $600 < p_T(\text{jet}) < 1200$ ” (right). The pink color corresponds to the MC distribution after using the alternative sample (i.e. HERWIG). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

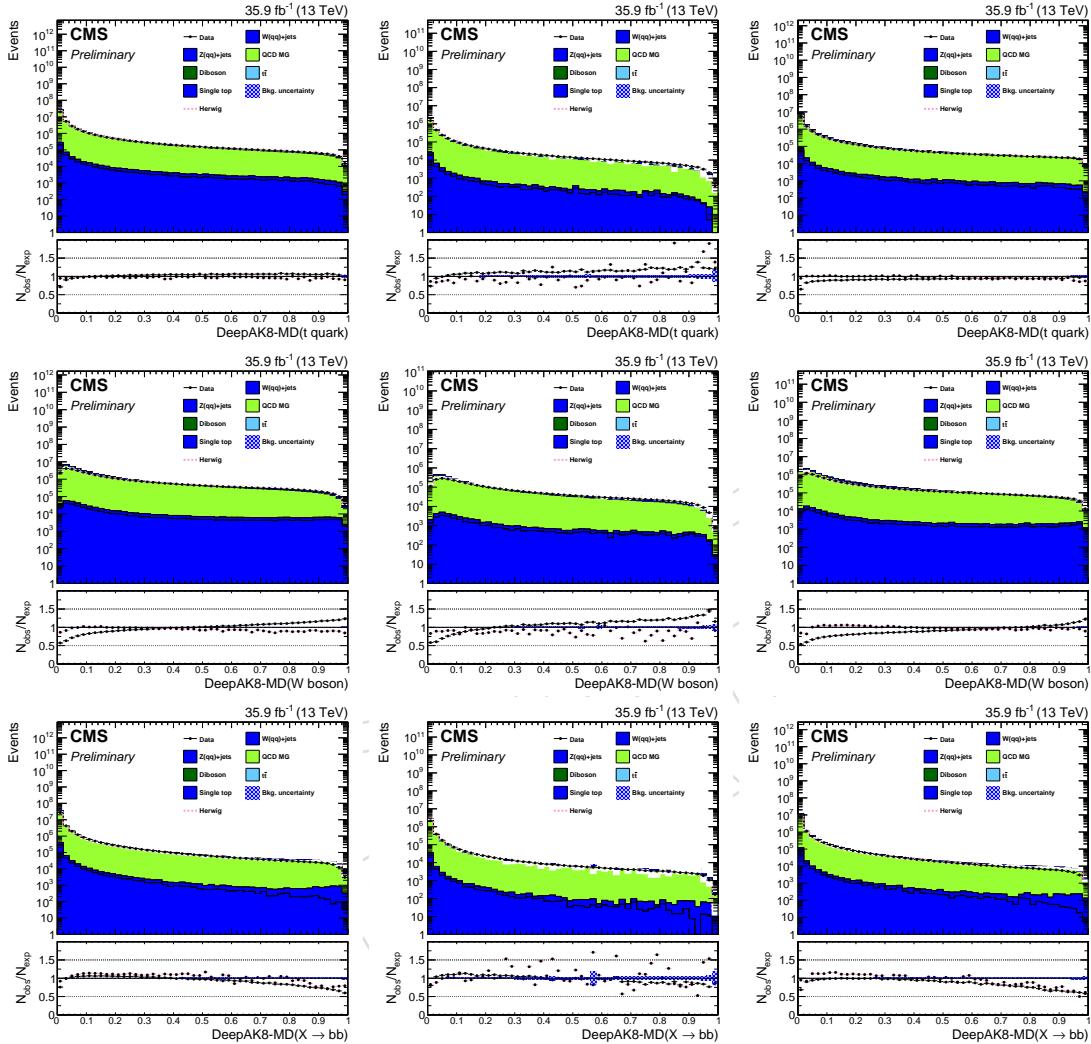


Figure 16: Distribution of the  $t$  quark, (1st row), the  $W$  boson (2nd row), and the  $X \rightarrow bb$  (3rd row) identification probabilities of DeepAK8-MD in data and MC in the QCD multijet sample. The distributions are shown in three regions of  $p_T$  (jet), “inclusive” (left), “ $300 < p_T(\text{jet}) < 400$ ” (middle) and “ $600 < p_T(\text{jet}) < 1200$ ” (right). The pink color corresponds to the MC distribution after using the alternative sample (i.e. HERWIG). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

### 379 7.3 Misidentification probability in the $\gamma$ +jets sample

380 The misidentification probability of the DeepAK8 algorithms is studied in an additional sam-  
381 ple, the  $\gamma$ +jet data samples. The QCD multijet and the  $\gamma$ +jets samples differ on the relative  
382 fraction of light quarks and gluons, with the latter having larger fraction of light quarks.

383 The distribution of  $m_{\text{jet}}$ , jet  $p_{\text{T}}$  and the N-subjetiness ratios,  $\tau_{32}$  and  $\tau_{21}$  in the  $\gamma$ +jets sample are  
384 displayed in Fig. 17, in three regions of  $p_{\text{T}}$  (jet), “inclusive”, “ $300 < p_{\text{T}}(\text{jet}) < 400$ ” and “ $600 <$   
385  $p_{\text{T}}(\text{jet}) < 1200$ ”. The MC distributions are normalized to the number of data. Simulation  
386 overpredicts the  $p_{\text{T}}$  distribution in data, which is attributed to the fact that  $\gamma$ +jet events are  
387 generated at LO. A similar trend is observed for  $m_{\text{jet}}$  since it is correlated to  $p_{\text{T}}$ . The N-subjetiness  
388 ratios are well described by simulation.

389 The distributions of the t quark and W, Z, and Higgs boson identification probabilities for the  
390 DeepAK8 algorithm are displayed in Fig. 18. Figure 19 displays the t quark, W boson and  $X \rightarrow$   
391  $bb$  identification probabilities for the DeepAK8-MD algorithm. In all cases the comparisons are  
392 shown in three regions of  $p_{\text{T}}$  (jet), “inclusive”, “ $300 < p_{\text{T}}(\text{jet}) < 400$ ” and “ $600 < p_{\text{T}}(\text{jet}) <$   
393 1200”. The MC distributions are normalized to the number of data. We observe a fair agreement  
394 in shape between data and MC for most of the variables. We will repeat the comparison when  
395 the  $\gamma$ +jets sample is reweighted based on the higher order calculations.

DRAFT

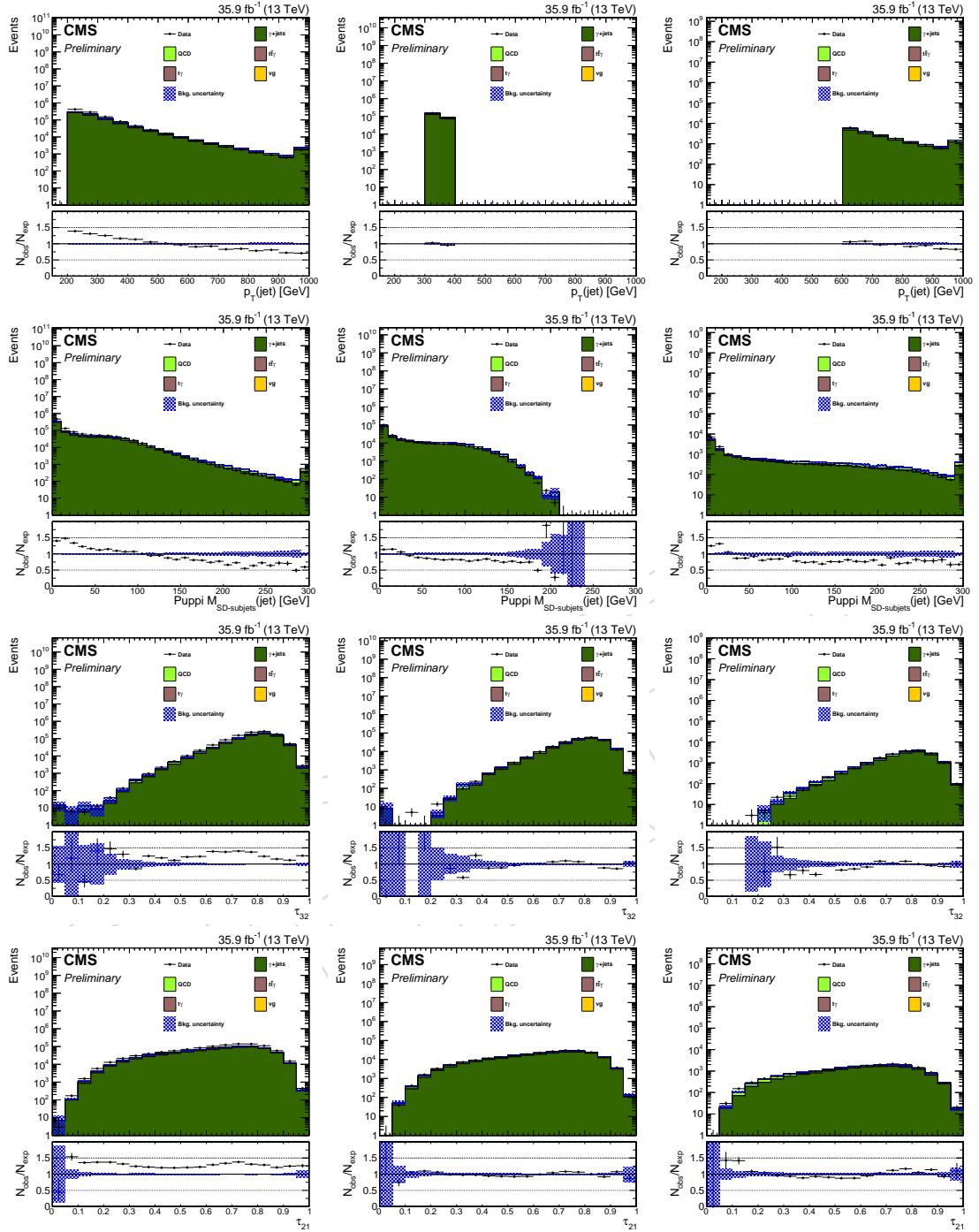


Figure 17: Distribution of the jet mass,  $m_{\text{jet}}$  (1st row), the jet  $p_T$  (2nd row), and the N-subjettiness ratios,  $\tau_{32}$  (3rd row) and  $\tau_{21}$  (4th row), in data and MC in the  $\gamma+\text{jets}$  sample. The distributions are shown in three regions of  $p_T(\text{jet})$ , “inclusive” (left), “ $300 < p_T(\text{jet}) < 400$ ” (middle) and “ $600 < p_T(\text{jet}) < 1200$ ” (right). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

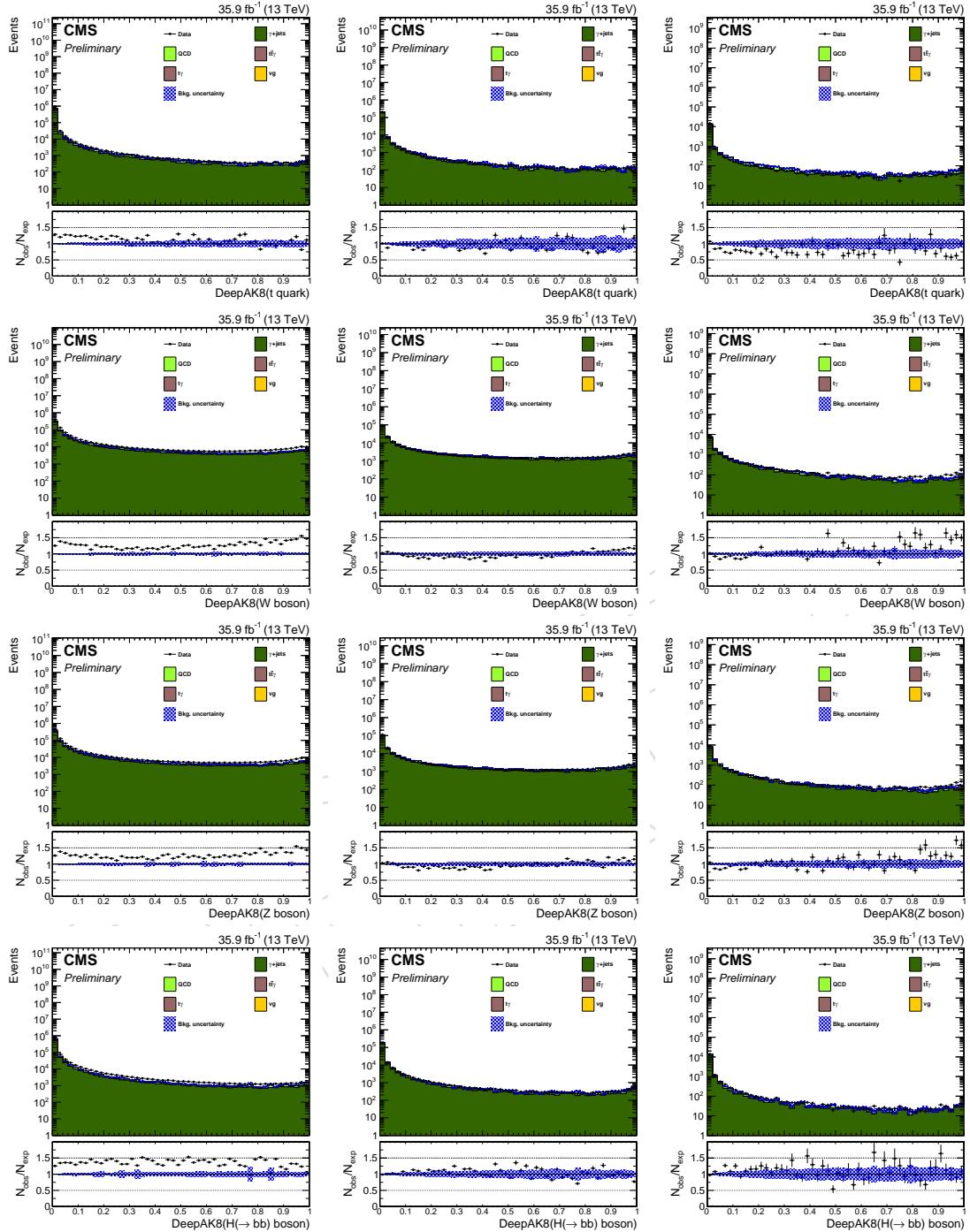


Figure 18: Distribution of the  $t$  quark, (1st row), the  $W$  boson (2nd row), the  $Z$  boson (3rd row), and the the Higgs boson (4th row) identification probabilities of DeepAK8 in data and MC in the  $\gamma + \text{jets}$  sample. The distributions are shown in three regions of  $p_T$  (jet), “inclusive” (left), “ $300 < p_T(\text{jet}) < 400$ ” (middle) and “ $600 < p_T(\text{jet}) < 1200$ ” (right). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

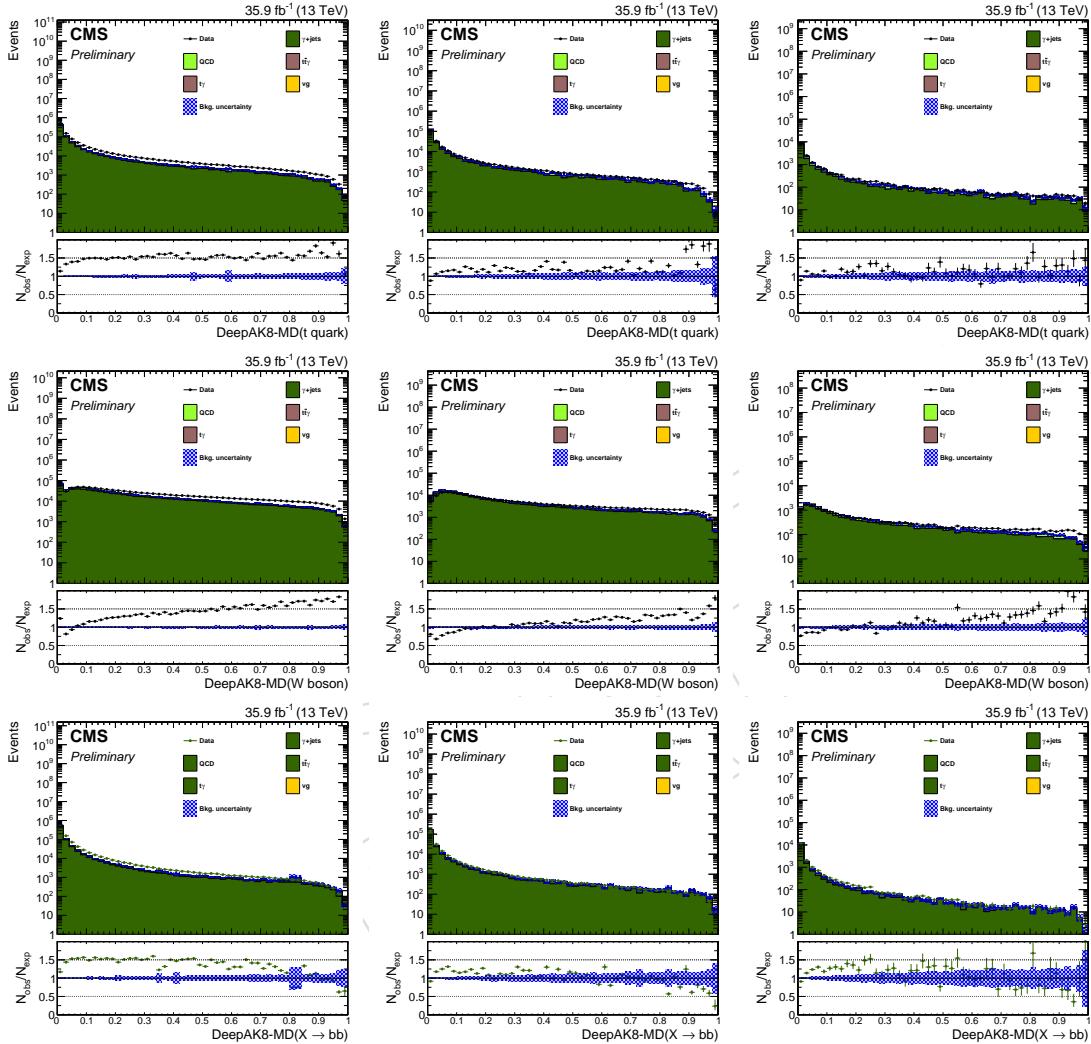


Figure 19: Distribution of the  $t$  quark, (1st row), the  $W$  boson (2nd row), and the  $X \rightarrow bb$  (3rd row) identification probabilities of DeepAK8–MD in data and MC in the  $\gamma+jets$  sample. The distributions are shown in three regions of  $p_T$  (jet), “inclusive” (left), “ $300 < p_T(\text{jet}) < 400$ ” (middle) and “ $600 < p_T(\text{jet}) < 1200$ ” (right). The lower panel in each plot shows the data to simulation ratio. The blue band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

---

**396 7.4 Systematic uncertainties**

397 The SF measurement can be sensitive to a number of sources of systematic uncertainties. The  
398 effect of each source is quantified by the difference between the SF obtained using the nominal  
399 and the alternative configuration (i.e. the one that includes the systematic variation). The list  
400 of the systematic sources considered is:

- 401 • **Event generator:** The effect of the choice of the generator for the hard scattering  
402 process is assessed by using samples with the same parton showering generator (i.e.  
403 PYTHIA8) but different event generators (e.g. POWHEG and MADGRAPH).
- 404 • **Parton shower:** Important systematic effects can arise from the choice of scheme in  
405 parton showering. Uncertainties are evaluated using samples with the same event  
406 generator but different choice for the modeling of the parton showering. Details  
407 about the samples used can be found in Section 2.
- 408 • **Renormalization and factorization scale:** Changes in renormalization ( $\mu_R$ ) and fac-  
409 torization ( $\mu_F$ ) scales are estimated by varying simultaneously  $\mu_R$  and  $\mu_F$  by a factor  
410 of two, as detailed in Refs. [add refs]. The uncertainty related to the choice of PDF, is  
411 obtained as the standard deviation in 100 variations of the NNPDF3.0 [add ref] PDF.
- 412 • **Jet energy scale and resolution:** The jet energy scale and resolution are changed  
413 within their  $p_T$  – and  $\eta$  – dependent uncertainties based on the studies presented in  
414 Ref. [add ref].
- 415 • **b-tagging :** The effect on the SF from uncertainties related to the identification of the  
416 b quarks, and the misidentification of c and light quarks is evaluated based on the  
417 measurements reported in Ref. [add ref]
- 418 • **Trigger and lepton identification:** Uncertainties on the measurement of the trigger  
419 efficiency [add ref] and on the energy scale and resolution of the leptons [add ref]  
420 are propagated in the SF extraction.
- 421 • **pileup:** The uncertainty from pileup is determined by changing the minimum bias  
422 cross section within its measured uncertainty of by 5% [add ref].

---

**423 7.5 Corrections to simulation**

The measurement of the t quark and W boson tagging efficiency in data is measured in the single- $\mu$  sample using a “tag & probe” method. The muon, in combination with the b tagged jet, is used as the “tag”. In the opposite hemisphere of the event the wide jet is used as the “probe”. As discussed in Section 7.1, the simulated events are decomposed into three categories: “Merged t quark”, “Merged W boson”, and the “Unmerged” category. The  $m_{\text{jet}}$  distributions of each one of the three categories are used to derive templates to fit the  $m_{\text{jet}}$  distribution in data. For a given working point the fit is done simultaneously for both the “passing” and “failing” events, for all the three categories. The fit performed in the range from 50 GeV to 300 GeV. After calculating the efficiencies in data and MC, the SF is determined as:

$$SF = \frac{\epsilon_{\text{Data}}}{\epsilon_{\text{MC}}} \quad (3)$$

424 The SF are extracted differentially in jet- $p_{\text{T}}$ . For the case of t quark identification the following  
 425 exclusive jet- $p_{\text{T}}$  regions are considered: 300 – 400, 400 – 480, 480 – 600, and 600 – 1200 GeV. In  
 426 order to increase the purity of “Merged W boson” candidates, we consider regions with lower  
 427 jet- $p_{\text{T}}$  : 200 – 300, 300 – 400, 400 – 800 GeV. The effect of the systematic sources discussed in  
 428 Section 7.4 is propagated as an uncertainty on the SF.

429 The mass distribution of the three MC templates before and after the fit and the mass distribu-  
 430 tion in data used for the extraction of the t quark (W boson) identification SF of the DeepAK8  
 431 algorithm, are shown in Figs. 20 (22) and 21 (23), for  $\epsilon_B = 5\%$  ( $\epsilon_B = 5\%$ ) and  $\epsilon_B = 0.1\%$   
 432 ( $\epsilon_B = 0.5\%$ ) working points, respectively, in the four (three)  $p_{\text{T}}$  regions. A similar set of plots  
 433 for the DeepAK8-MD is presented in Figs 24- 27. The only difference is due to the fact that in  
 434 DeepAK8-MD the mass sculpting is significantly reduces, for the W boson we extract SF for  
 435  $\epsilon_B = 0.1\%$  as for the t quark.

436 The SF measured for the t quark and W boson identification for the DeepAK8 and DeepAK8-  
 437 MD algorithms are summarized in Figs. 28 and ??, respectively, for various working points  
 438 based on  $\epsilon_B$ . The SF are typically consistent with 1 within uncertainties. The largest SF is mea-  
 439 sured for the identification of t quarks using DeepAK8. The statistical uncertainties dominate  
 440 the SF measurement. From the various systematic sources, the one associated to the parton  
 441 showering scheme has the largest effect. An interesting point to highlight is that algorithms  
 442 designed to avoid strong dependence on the mass, like the DeepAK8-MD, have significantly  
 443 smaller uncertainties. Another point is that the effect of the systematic uncertainties is more  
 444 pronounced, yet still small, for DeepAK8.

445 The mistag rate as a function of the  $p_{\text{T}}$  is displayed in Figs 30 and 31 for DeepAK8 and DeepAK8-  
 446 MD, respectively, in the QCD multijet and  $\gamma + \text{jets}$  samples.

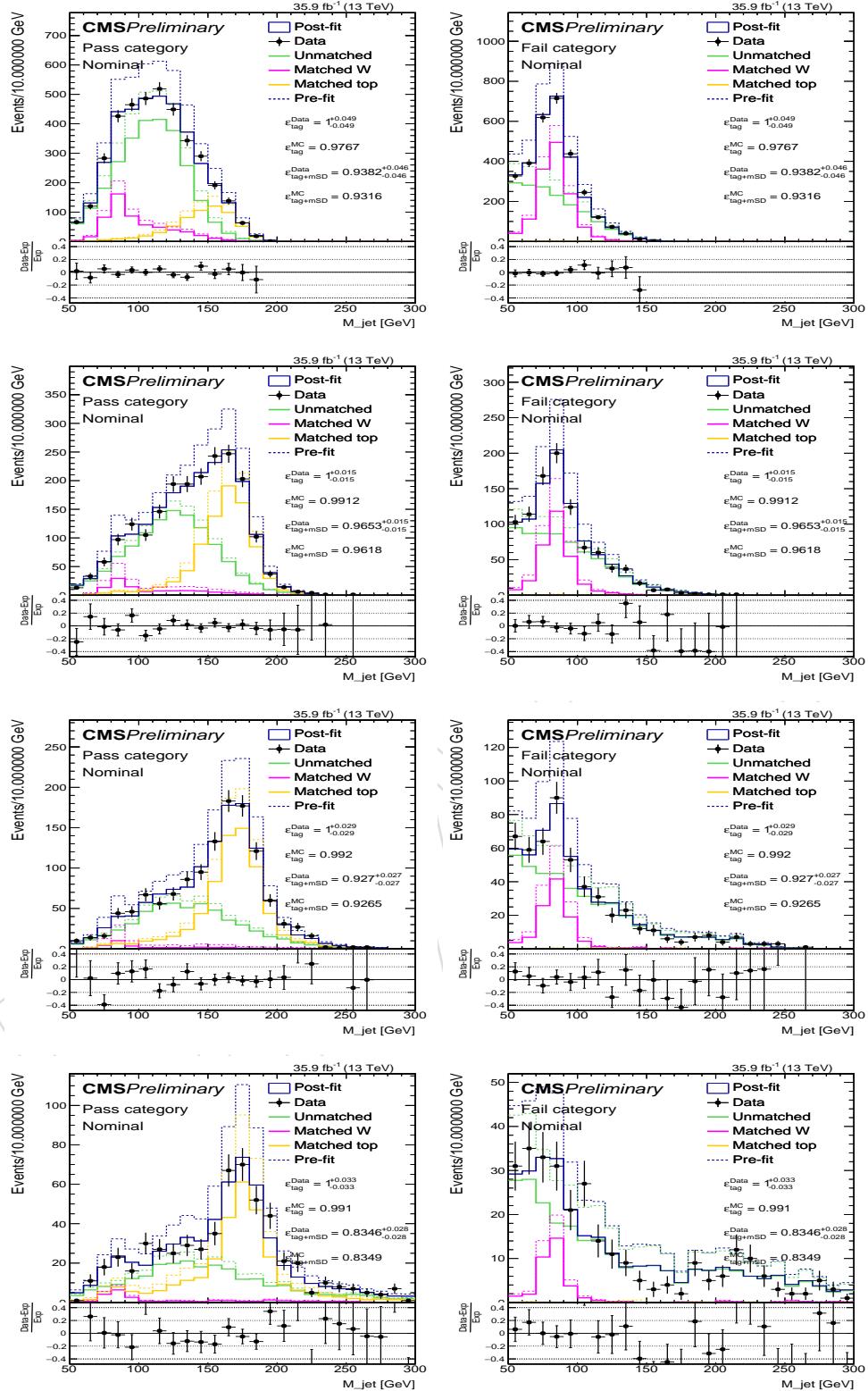


Figure 20: The mass distribution of the three MC templates before and after the fit and the mass distribution in data used for the extraction of the  $t$  quark identification SF of the DeepAK8 algorithm, in  $300 - 400$  (1st row),  $400 - 480$  (2nd row),  $480 - 600$  (3rd row), and  $600 - 1200$  GeV (4th row), for a working point with  $\epsilon_B = 5\%$ . Left: fail category, right: pass category. The dashed histograms correspond to the pre-fit results and the solid lines to the post-fit results.

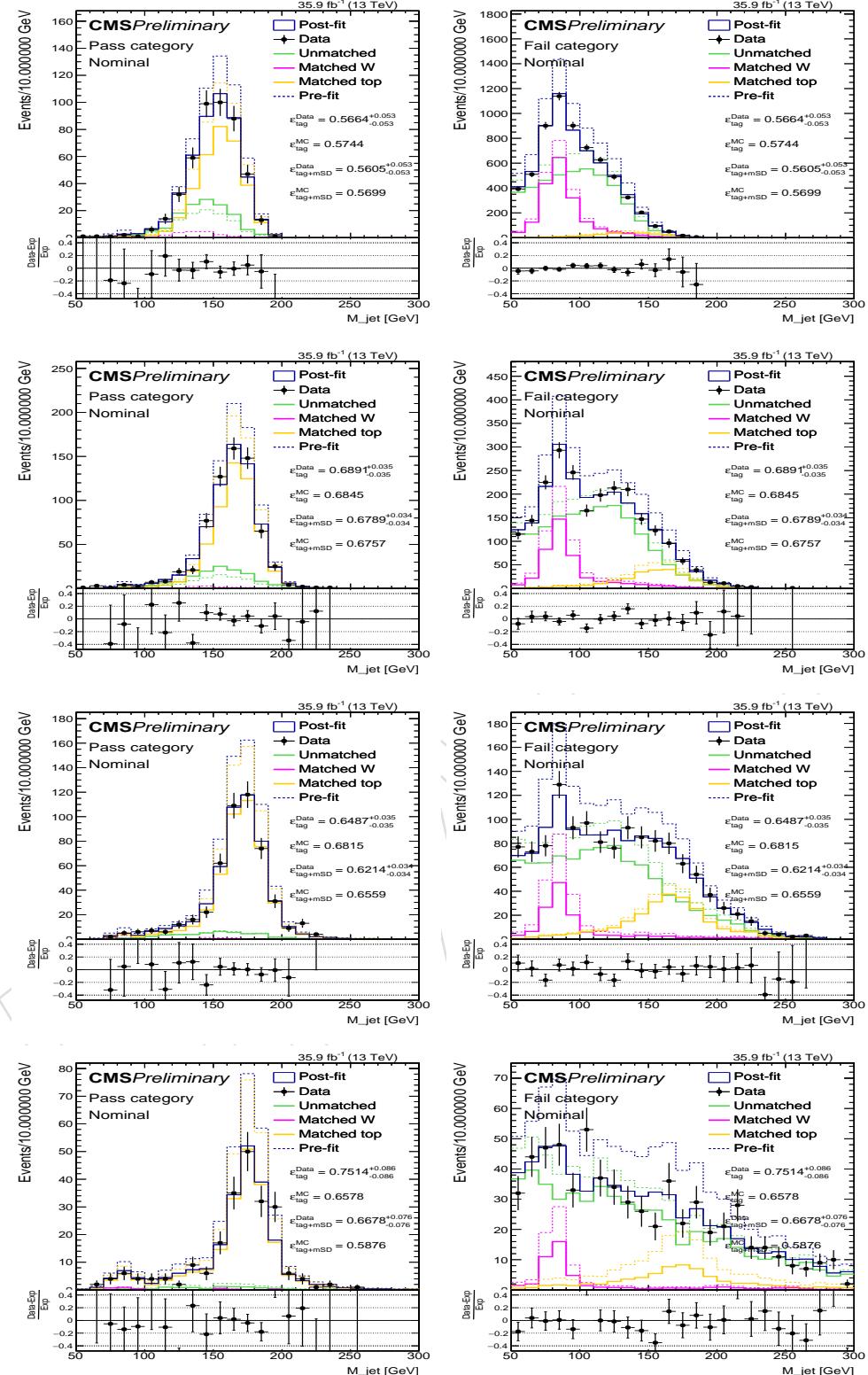


Figure 21: The mass distribution of the three MC templates before and after the fit and the mass distribution in data used for the extraction of the  $t$  quark identification SF of the DeepAK8 algorithm, in  $300 - 400$  (1st row),  $400 - 480$  (2nd row),  $480 - 600$  (3rd row), and  $600 - 1200$  GeV (4th row), for a working point with  $\epsilon_B = 0.1\%$ . Left: fail category, right: pass category. The dashed histograms correspond to the pre-fit results and the solid lines to the post-fit results.

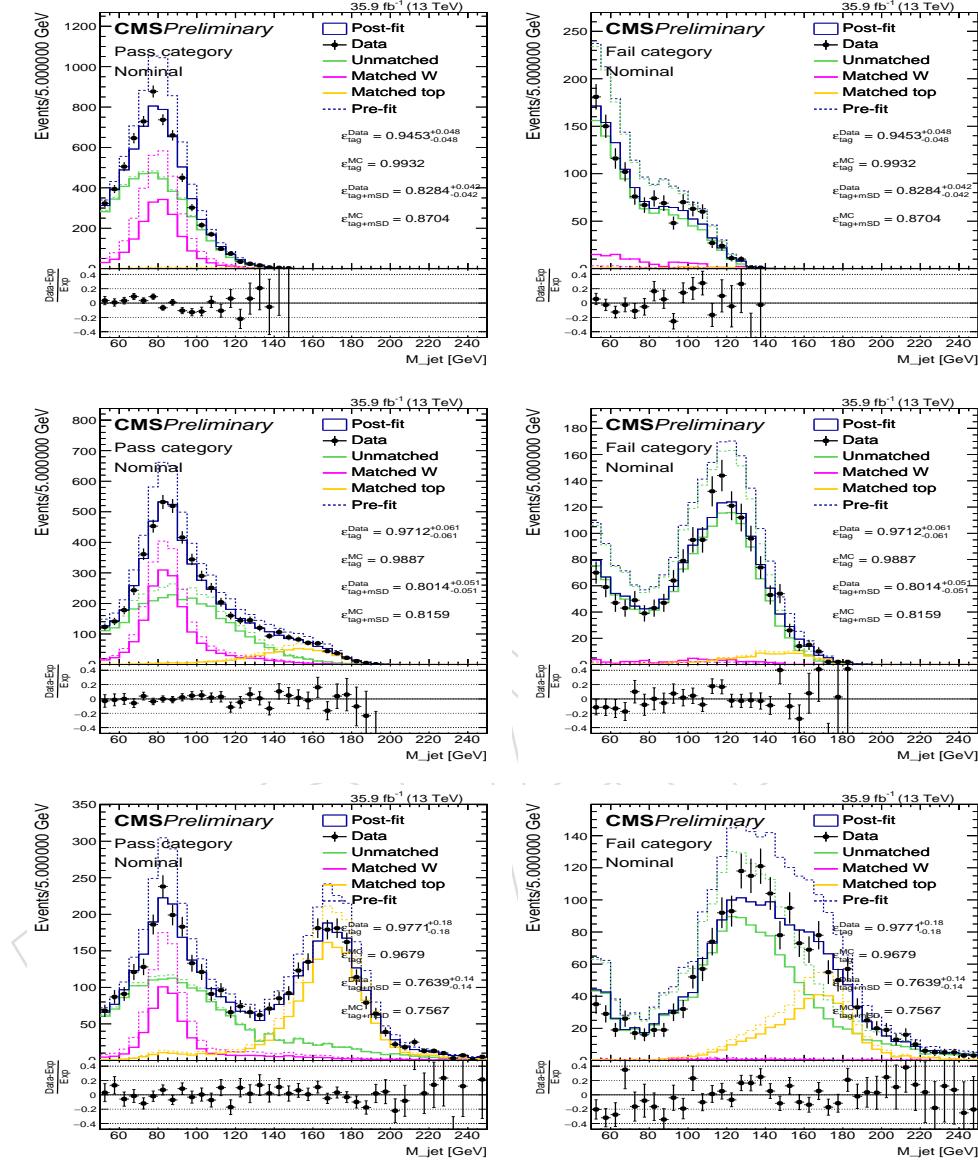


Figure 22: The mass distribution of the three MC templates before and after the fit and the mass distribution in data used for the extraction of the W boson identification SF of the DeepAK8 algorithm, in 200 – 300 (1st row), 300 – 400 (2nd row), and 400 – 800 GeV (3rd row), for a working point with  $\epsilon_B = 5\%$ . Left: fail category, right: pass category. The dashed histograms correspond to the pre-fit results and the solid lines to the post-fit results.

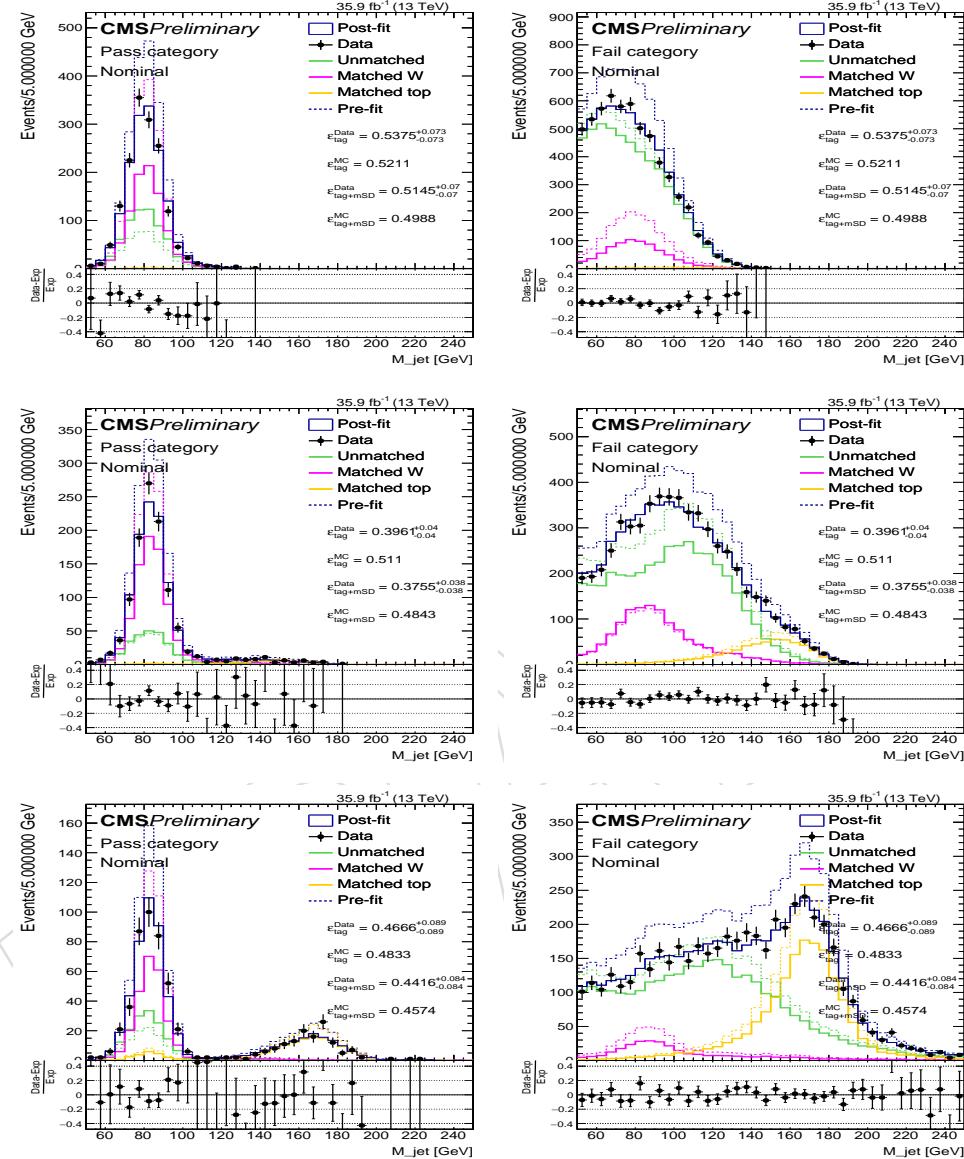


Figure 23: The mass distribution of the three MC templates before and after the fit and the mass distribution in data used for the extraction of the W boson identification SF of the DeepAK8 algorithm, in  $200 - 300$  (1st row),  $300 - 400$  (2nd row), and  $400 - 800$  GeV (3rd row), for a working point with  $\epsilon_B = 0.5\%$ . Left: fail category, right: pass category. The dashed histograms correspond to the pre-fit results and the solid lines to the post-fit results.

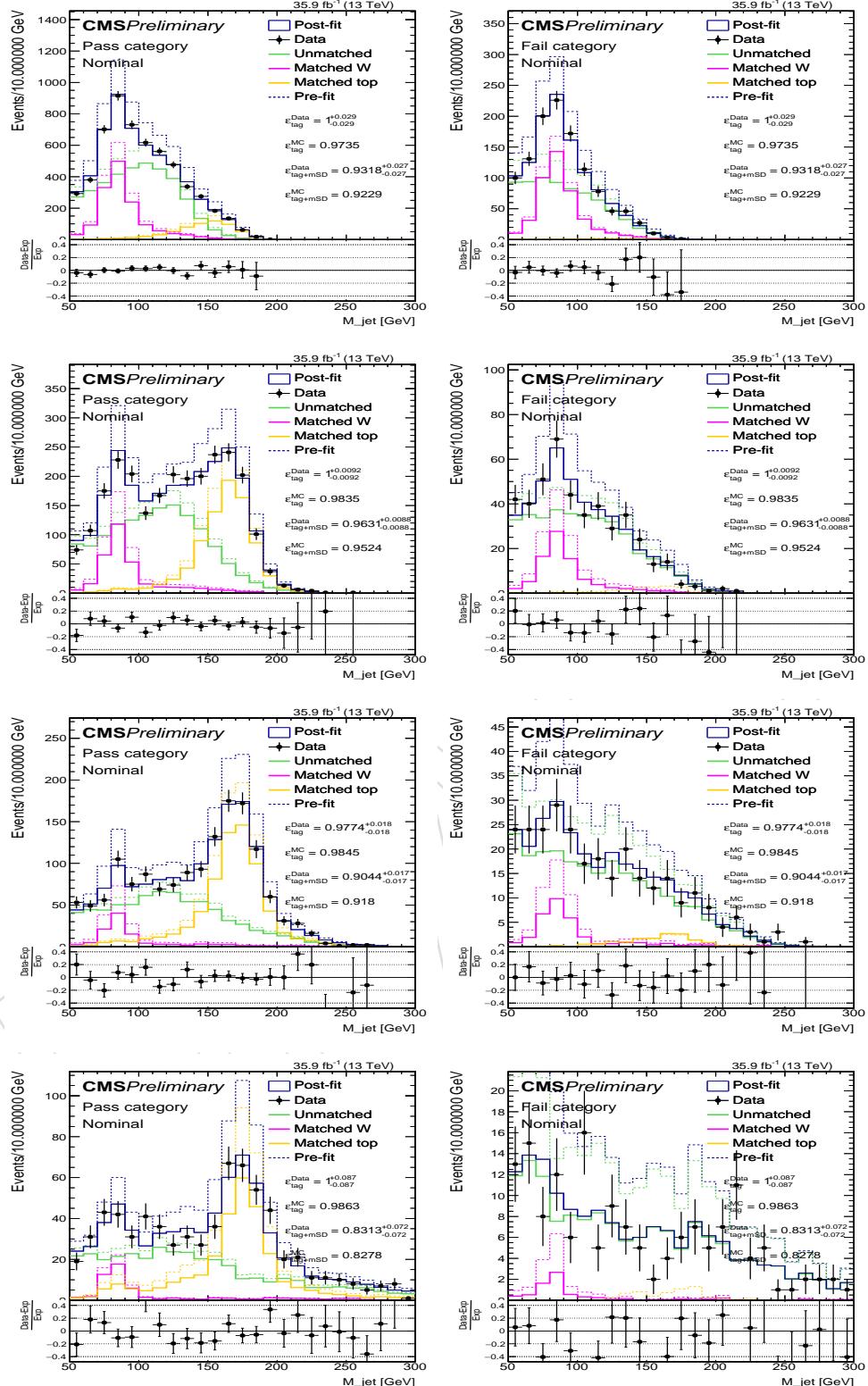


Figure 24: The mass distribution of the three MC templates before and after the fit and the mass distribution in data used for the extraction of the t quark identification SF of the DeepAK8-MD algorithm, in  $300 - 400$  (1st row),  $400 - 480$  (2nd row),  $480 - 600$  (3rd row), and  $600 - 1200$  GeV (4th row), for a working point with  $\epsilon_B = 5\%$ . Left: fail category, right: pass category. The dashed histograms correspond to the pre-fit results and the solid lines to the post-fit results.

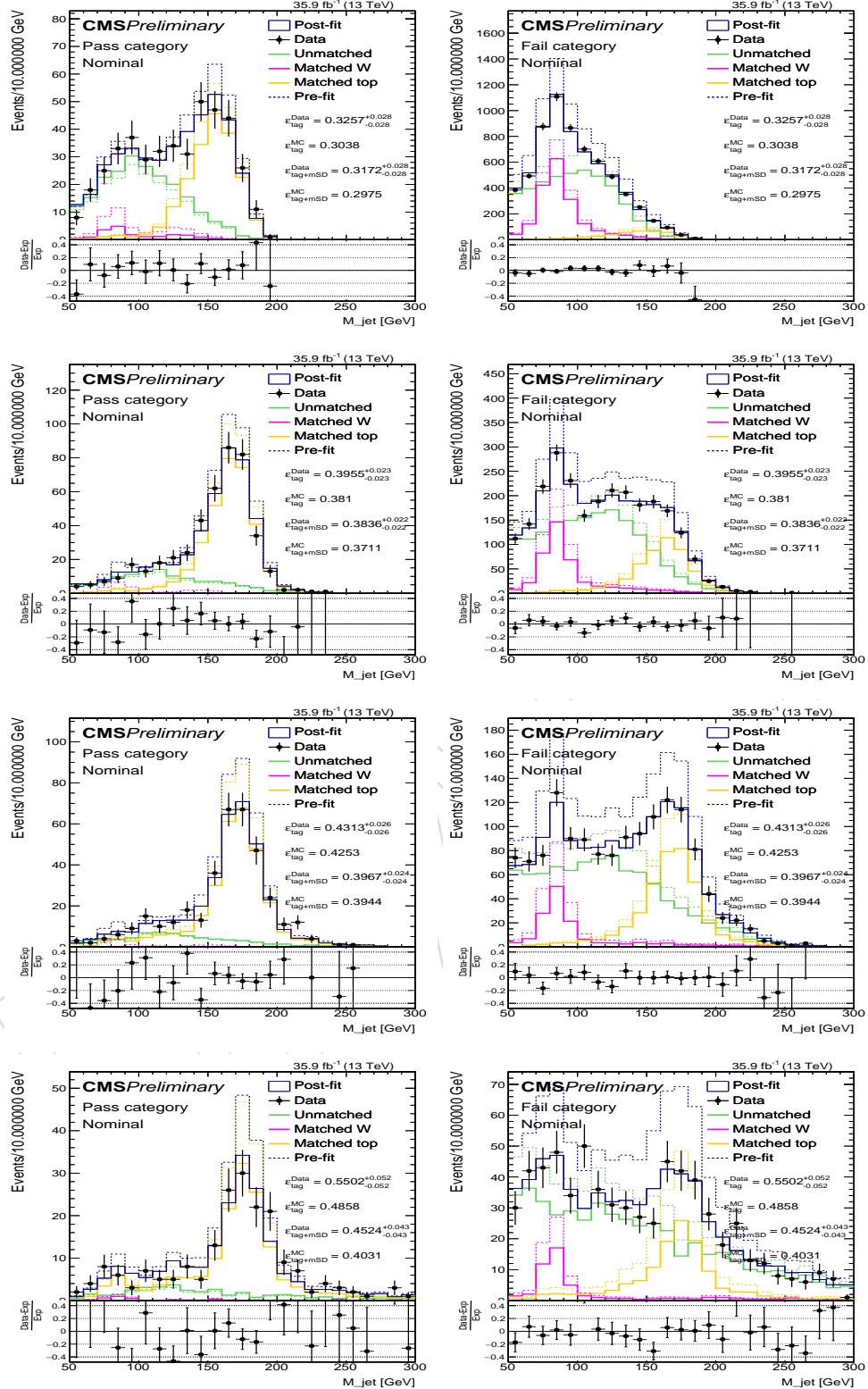


Figure 25: The mass distribution of the three MC templates before and after the fit and the mass distribution in data used for the extraction of the t quark identification SF of the DeepAK8-MD algorithm, in 300 – 400 (1st row), 400 – 480 (2nd row), 480 – 600 (3rd row), and 600 – 1200 GeV (4th row), for a working point with  $\epsilon_B = 0.1\%$ . Left: fail category, right: pass category. The dashed histograms correspond to the pre-fit results and the solid lines to the post-fit results.

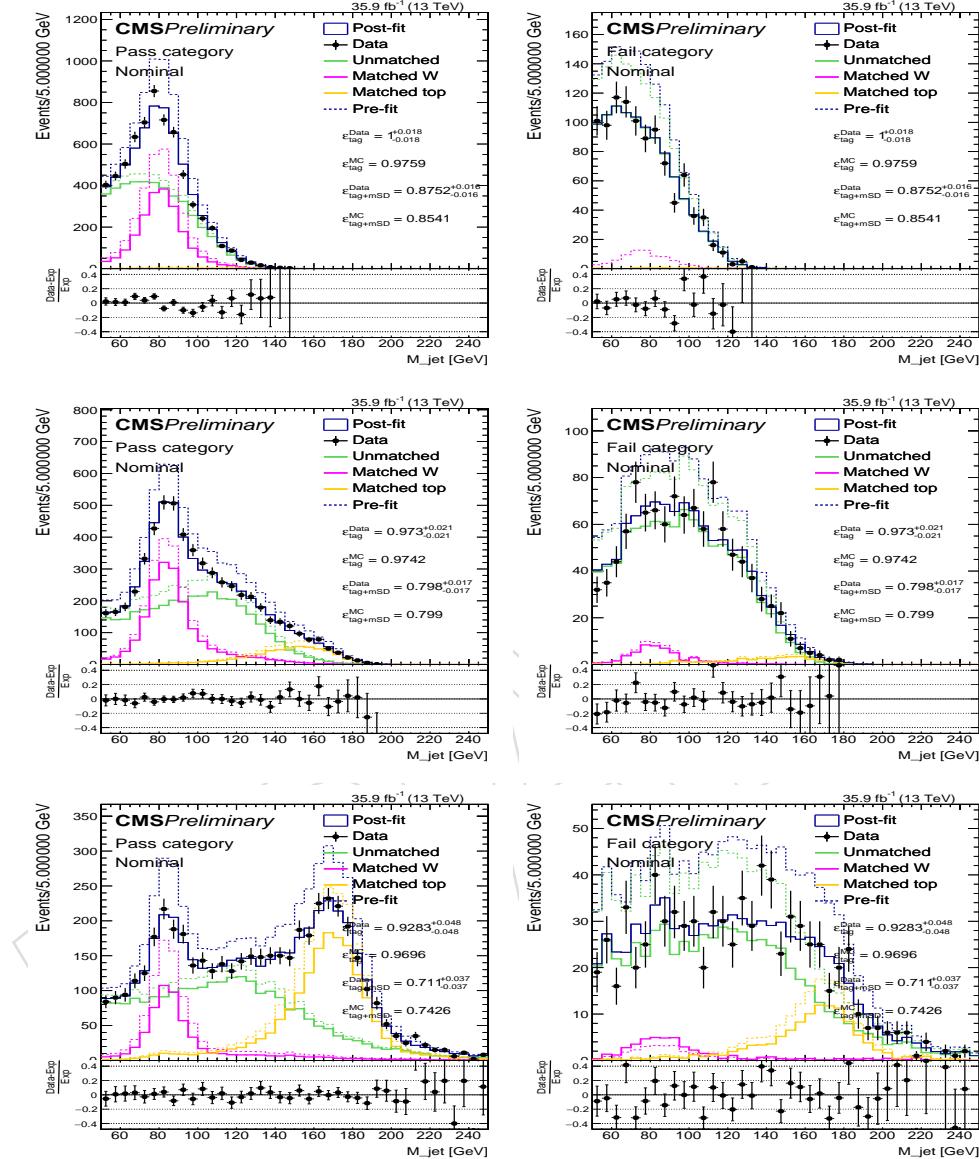


Figure 26: The mass distribution of the three MC templates before and after the fit and the mass distribution in data used for the extraction of the W boson identification SF of the DeepAK8-MD algorithm, in 200 – 300 (1st row), 300 – 400 (2nd row), and 400 – 800 GeV (3rd row), for a working point with  $\epsilon_B = 5\%$ . Left: fail category, right: pass category. The dashed histograms correspond to the pre-fit results and the solid lines to the post-fit results.

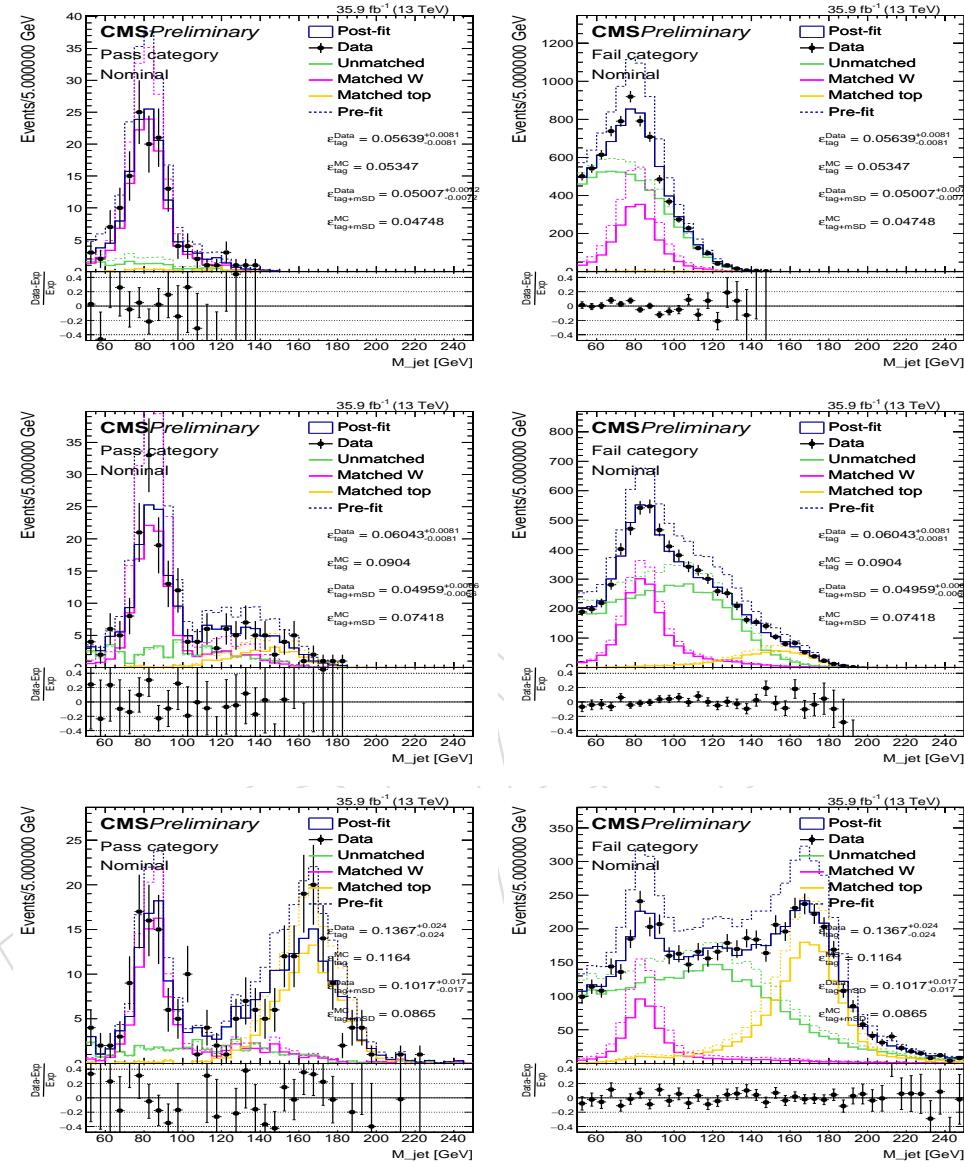


Figure 27: The mass distribution of the three MC templates before and after the fit and the mass distribution in data used for the extraction of the W boson identification SF of the DeepAK8-MD algorithm, in  $200 - 300$  (1st row),  $300 - 400$  (2nd row), and  $400 - 800$  GeV (3rd row), for a working point with  $\epsilon_B = 0.1\%$ . Left: fail category, right: pass category. The dashed histograms correspond to the pre-fit results and the solid lines to the post-fit results.

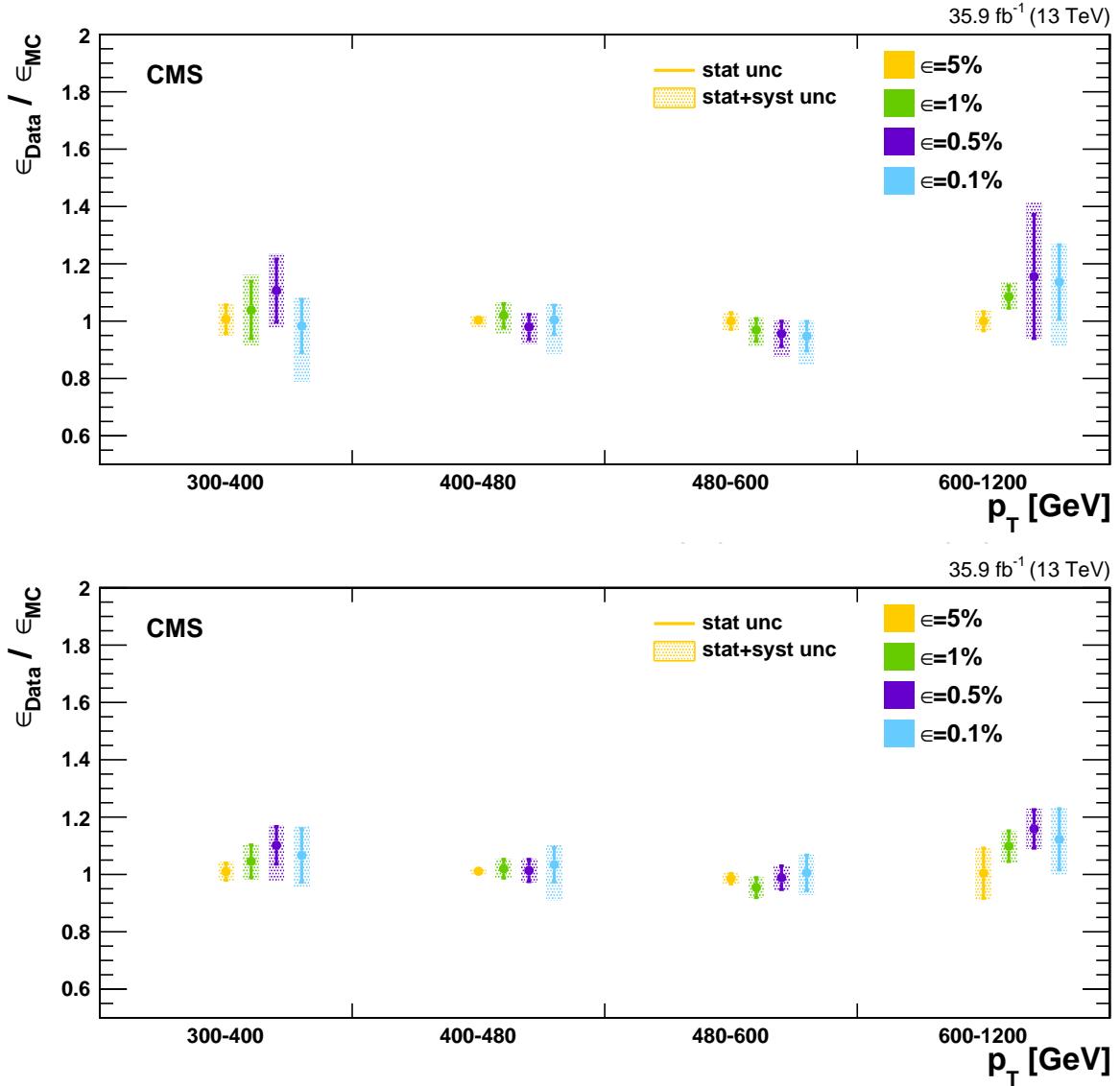


Figure 28: Summary of the SF measured for the  $t$  quark identification using DeepAK8 (upper) and DeepAK8-MD (lower) algorithms. The SF are evaluated for various working points as indicated on the plots. The markers correspond to the SF value, the error bars to the statistical uncertainty on the SF measurement, and the band is the total uncertainty (statistical + systematic).

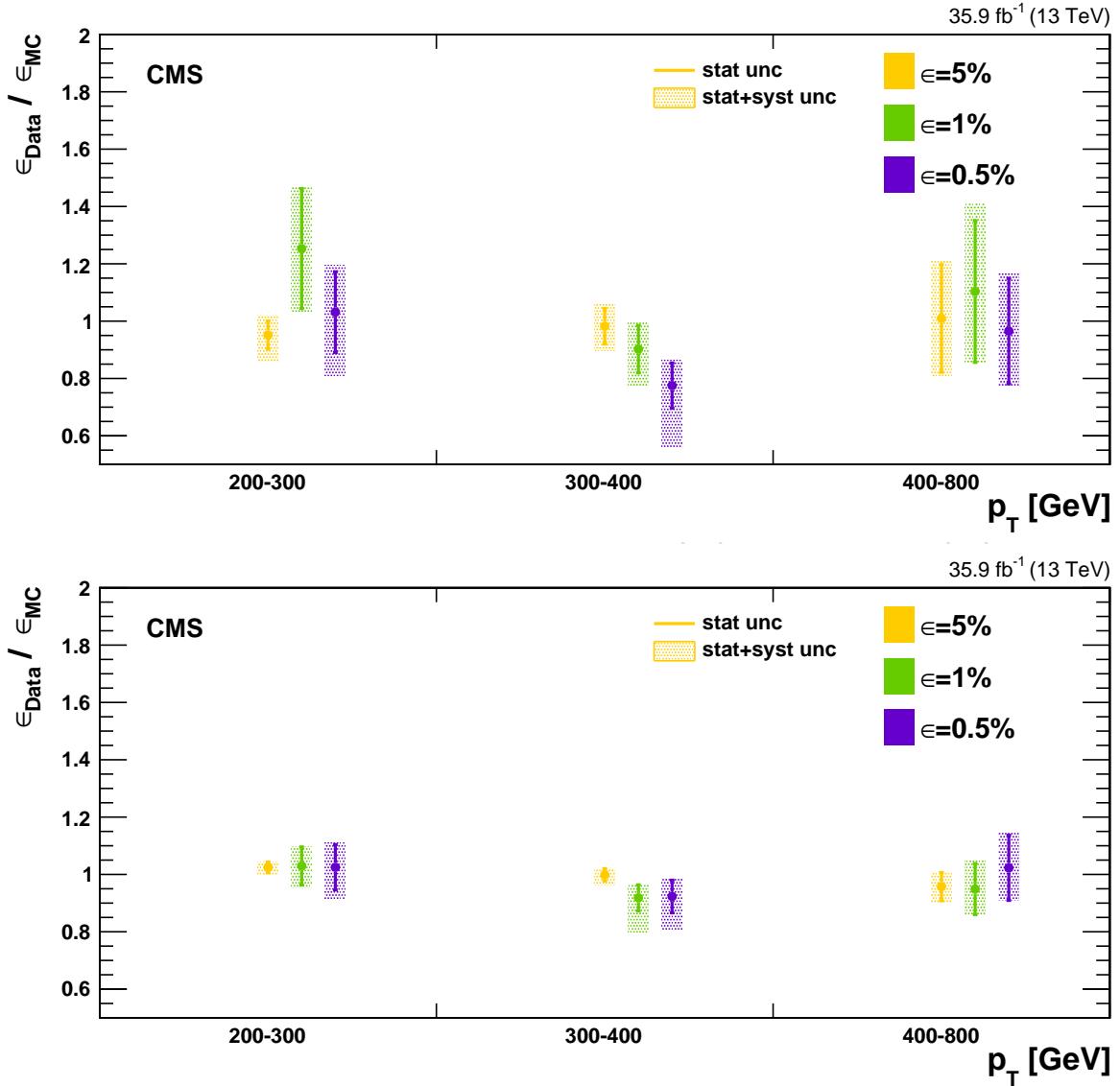


Figure 29: Summary of the SF measured for the W boson identification using DeepAK8 (upper) and DeepAK8-MD (lower) algorithms. The SF are evaluated for various working points as indicated on the plots. The markers correspond to the SF value, the error bars to the statistical uncertainty on the SF measurement, and the band is the total uncertainty (statistical + systematic).

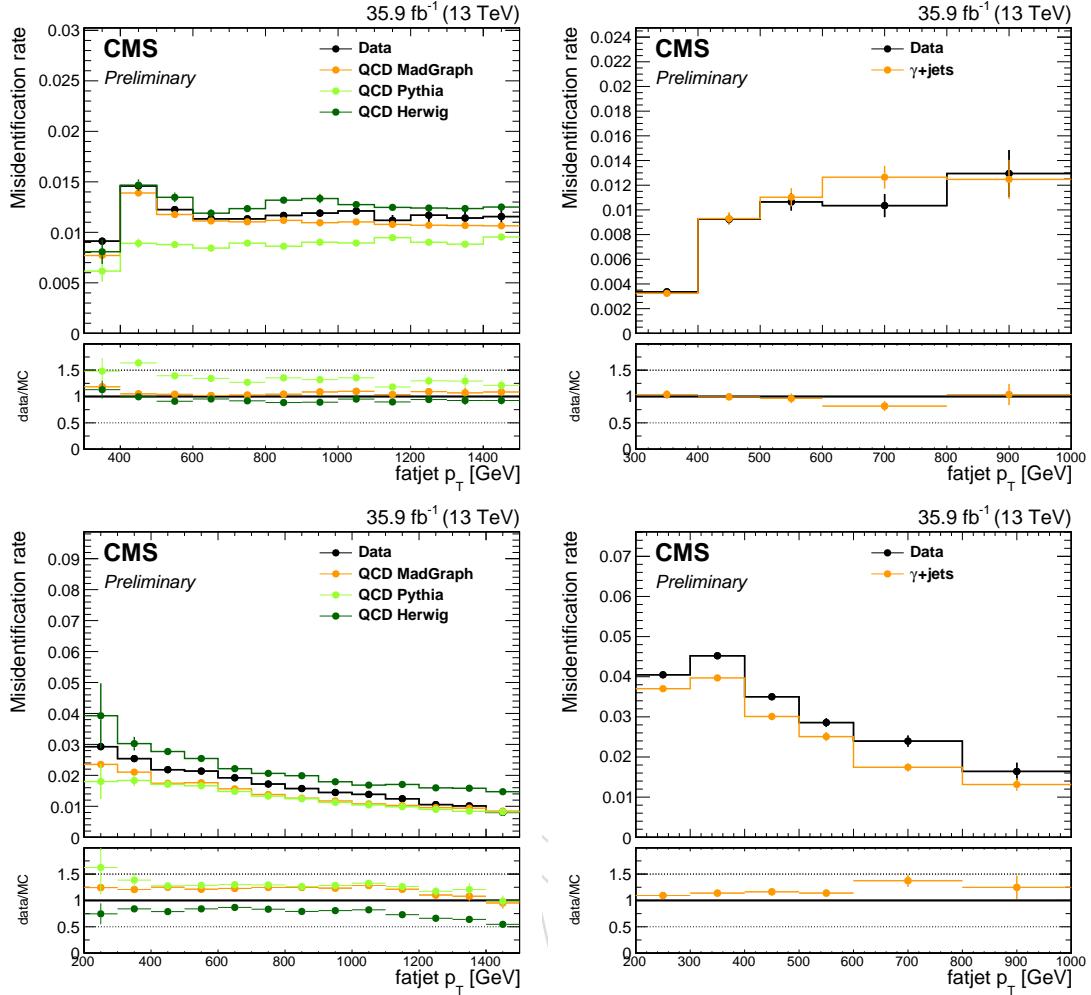


Figure 30: Mistagging rate of  $t$  quarks (upper row) and  $W$  bosons (lower row) of the DeepAK8 algorithm as a function of the jet  $p_T$  in the QCD multijet (left) and  $\gamma + jets$  (right) samples. The upper panel shows the distribution in data (black markers) and for three QCD MC samples generated with different configurations for the event generation and the parton showering. The lower panel shows the data to simulation ratio. The grey band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

## 8 Summary

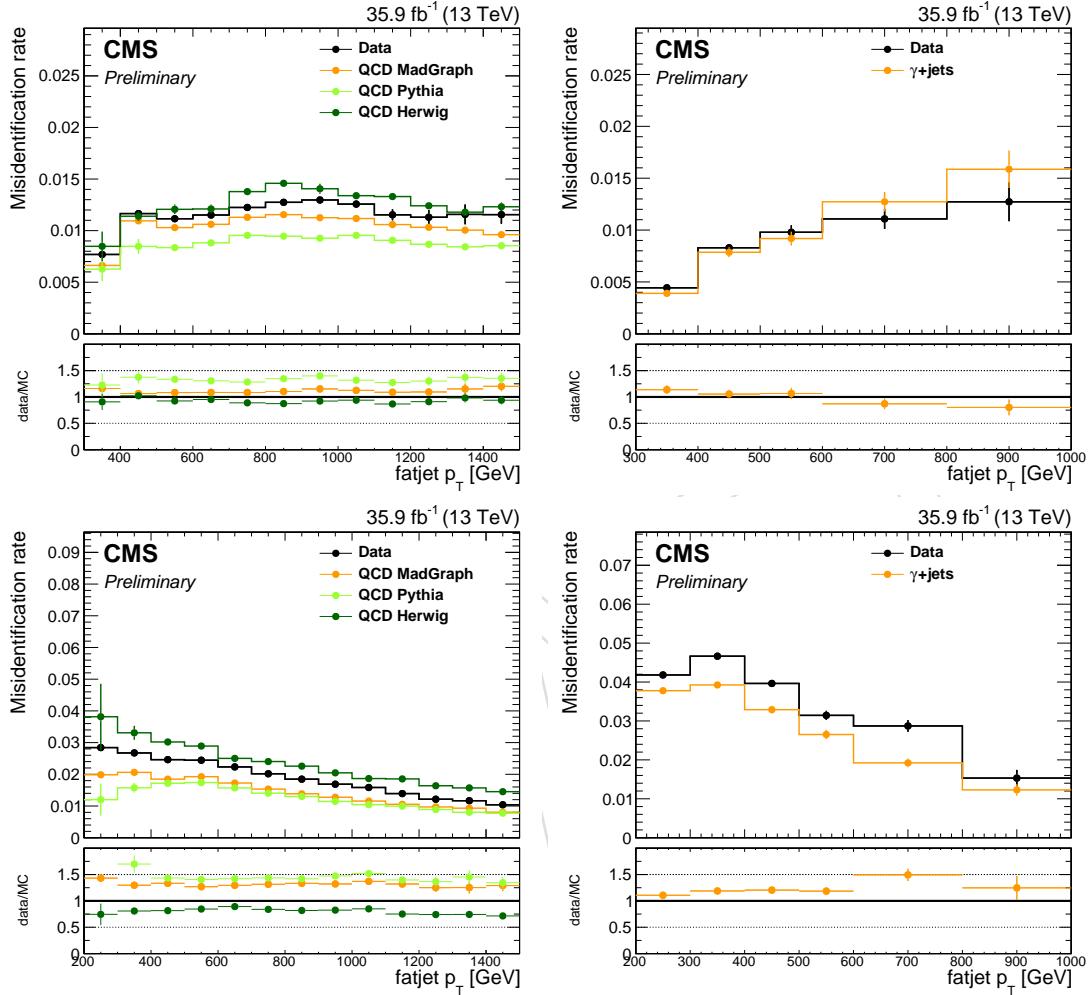


Figure 31: Mistagging rate of  $t$  quarks (upper row) and  $W$  bosons (lower row) of the DeepAK8-MD algorithm as a function of the jet  $p_T$  in the QCD multijet (left) and  $\gamma + jets$  (right) samples. The upper panel shows the distribution in data (black markers) and for three QCD MC samples generated with different configurations for the event generation and the parton showering. The lower panel shows the data to simulation ratio. The grey band corresponds to the statistical uncertainty of the MC samples and the vertical lines correspond to the statistical uncertainty of the data.

## 448 References

- 449 [1] Reference for JMAR Heavy Resonance Overview,  
450 <https://twiki.cern.ch/twiki/bin/view/CMS/JetMETHeavyResPaper>.
- 451 [2] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_T$  jet clustering algorithm”, *JHEP* **04**  
452 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- 453 [3] <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetToolbox>.
- 454 [4] <https://github.com/cms-jet/PuppiSoftdropMassCorr/>.
- 455 [5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, in  
456 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.  
457 2016.
- 458 [6] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks”, in  
459 *European Conference on Computer Vision*, pp. 630–645, Springer. 2016.
- 460 [7] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann  
461 machines”, in *Proceedings of the 27th international conference on machine learning (ICML-10)*,  
462 pp. 807–814. 2010.
- 463 [8] N. Srivastava et al., “Dropout: a simple way to prevent neural networks from  
464 overfitting.”, *Journal of machine learning research* **15** (2014), no. 1, 1929–1958.
- 465 [9] T. Chen et al., “Mxnet: A flexible and efficient machine learning library for heterogeneous  
466 distributed systems”, *arXiv preprint arXiv:1512.01274* (2015).
- 467 [10] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, *CoRR*  
468 **abs/1412.6980** (2014).
- 469 [11] G. Louppe, M. Kagan, and K. Cranmer, “Learning to Pivot with Adversarial Networks”,  
470 in *Advances in Neural Information Processing Systems*, pp. 982–991. 2017.
- 471 [12] <https://indico.cern.ch/event/644151/contributions/2616106>.

## 472 A Effects of subjet b-tagging

473 Identification of boosted heavy particles can benefit from both jet substructure information  
 474 and jet flavour information. Traditionally, substructure tagging and flavour tagging are per-  
 475 formed separately and the results can be combined to improve the overall performance for  
 476 boosted jet identification. However, this approach was found to be suboptimal compared to  
 477 the algorithm using jet constituent particles directly. Figure 32 shows the ROC curves of four  
 478 different algorithms for boosted top quark identification. The BDT taggers are based on the top  
 479 and Wtagger used in the search for supersymmetric top quarks in the all-jets final state [CMS-  
 480 SUS-16-049]. Two versions with different inputs are compared: (1) the “w/o b-tag” version  
 481 uses 16 input variables including the soft drop mass, N-subjettiness ratios, observables related  
 482 to quark-gluon discrimination, the relative difference in  $p_T$  between each of the two subjets  
 483 within the AK8 jet, and the mass of each subjet. (2) the “Full” version adds additional input  
 484 variables, including the CSVv2 b-tagging discriminant values of the subjets, as well as the in-  
 485 puts used in the boosted double-b tagger [CMS-BTV-15-002], to help identify heavy flavour  
 486 content of the jet. The DNN taggers are earlier version of the algorithms described in 4. Two  
 487 versions with different inputs are compared: (1) the “Kinematics” version uses only six basic  
 488 kinematic variables [ $p_T$ ,  $\Delta\eta$ ,  $\Delta\phi$ ,  $\Delta R(\text{jet})$ ,  $\Delta R(\text{subjet 1})$ ,  $\Delta R(\text{subjet 2})$ ] of each constituent parti-  
 489 cle. (2) the “Full” version adds more information of the constituent particles, such as the track  
 490 displacement and quality, as well as the secondary vertices associated with the jets, to help  
 491 identify heavy flavour content of the jet. As shown in Figure 32, the addition of flavour tag-  
 492 ging information brings significant improvement in performance for discriminating top quark  
 493 jets and QCD jets. The gain, however, is much larger for the DNN tagger using the full set of  
 494 particle-level information than for the BDT tagger using subjet CSVv2 b-tagging discriminants.  
 495 A natural question that follows is that if using an improved subjet b-tagging algorithm (e.g., us-  
 496 ing also DNNs and particle level inputs) in the BDT can reach similar performance as the DNN  
 497 boosted jet tagger. This was briefly studied in the early stage of the development of the DNN-  
 498 based algorithm and reported in [12]. As shown in Figure 33, despite a large improvement  
 499 in the subjet b-tagging performance itself for the DeepFlavour algorithm compared with the  
 500 CSVv2 algorithm (Figure 33(a)), the effects on top tagging performance is only moderate. Due  
 501 to differences in the samples and selections used between Figure 32 and 33, the ROC curves are  
 502 not directly comparable, but the gain using subjet DeepFlavour b-tagging is unlikely to match  
 503 the DNN using the full set of particle-level information directly.

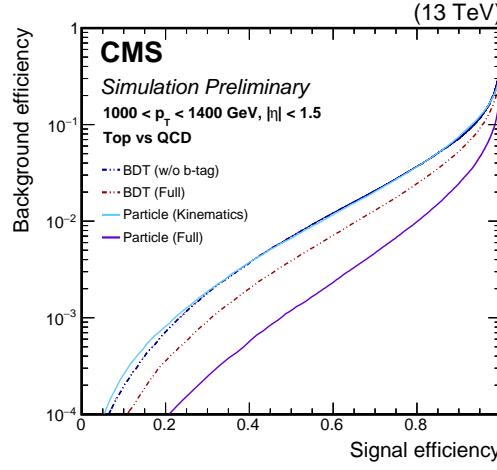


Figure 32: ROC curves of different top tagging algorithm based on BDTs or DNNs.

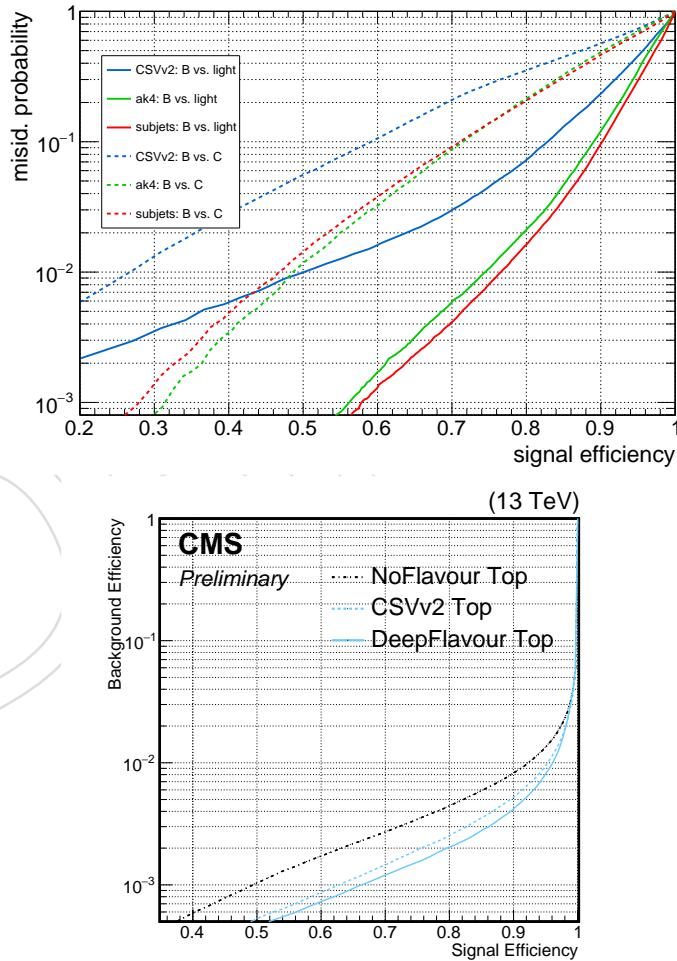


Figure 33: (a) ROC curves for different subjet b-tagging algorithms. (b) ROC curves of a BDT top tagger using no subjet b-tagging, the CSVv2 subjet b-tagging discriminants, and the DeepFlavour subjet b-tagging discriminants.

## 504 B Comparison of different deep learning approaches

505 A number of deep learning algorithms have been proposed for boosted jet identification in re-  
 506 cent years. The approaches generally fall into two categories, one is to treat a jet as an image  
 507 with energy deposition on the calorimeters and uses image recognition techniques (e.g., con-  
 508 volutional neural networks), and the other is to use the individual constituent particles of a jet  
 509 as inputs, and utilize sequential models (e.g., recurrent neural networks) that are commonly  
 510 used in natural language processing tasks. The approach adopted in this note falls into the  
 511 second category, while the main difference is that one-dimensional convolutional neural net-  
 512 works are utilized instead of recurrent neural networks to speed up the processing. The per-  
 513 formance of our architecture is compared to a number of other approaches, including image-based  
 514 approaches of different complexities (the simple CNN model “ConvNet”, one of the state-of-  
 515 the-art CNN model “ResNeXt-50”), in Figure 34, and is found to outperform all the others  
 516 approaches that we have investigated so far.

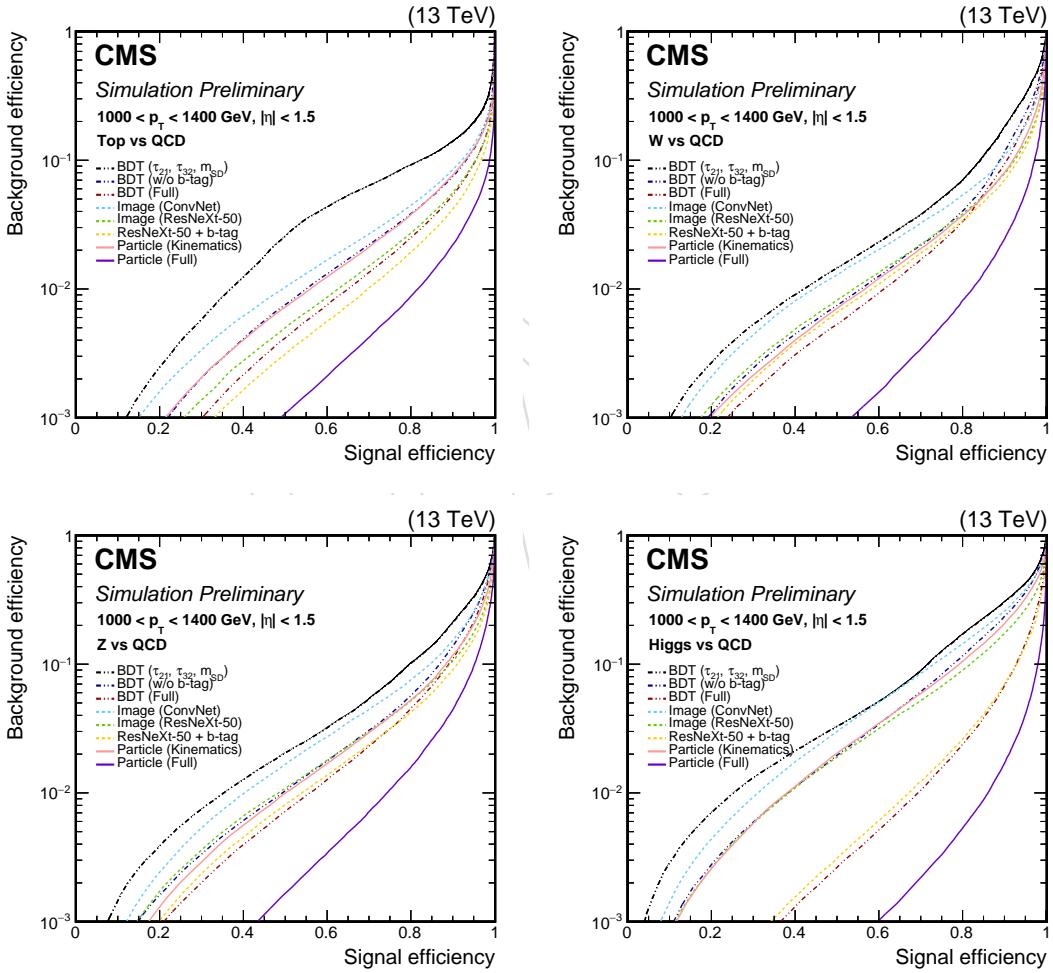


Figure 34: Performance comparison of a number of machine learning approaches for boost jet identification in terms of receiver operating characteristic (ROC) curves for jets with  $1000 < p_T < 1400 \text{ GeV}$ ,  $|\eta| < 2.4$  for the identification of (a) top quarks, (b) W bosons, (c) Z bosons, (d) Higgs bosons.

## 517 C Impacts of the input features

518 The boosted jet identification algorithm described in Section 4 combines a large number of  
 519 input features using deep neural networks and shows significant improve performance com-  
 520 pared to traditional approaches using jet level features. Therefore, it would be interesting to  
 521 understand where the improvements are coming from. Based on the studies shown in Ap-  
 522 pendix A, a large fraction of the improvement can be attributed to a better utilization of the  
 523 jet flavour information by using the particle-level inputs rather than using jet or subjet level  
 524 flavour tagging. While this is valid for top quark or Higgs boson identification, where flavour  
 525 tagging is expected to provide powerful separation between signals and backgrounds, it seems  
 526 unable to explain the large improvement in W or Z boson identification, especially the large  
 527 improvement from using only the particle kinematics to using the full set of inputs (including  
 528 the impact parameters of the tracks that are thought to contribute mainly to flavour tagging),  
 529 as shown in Figure 35.

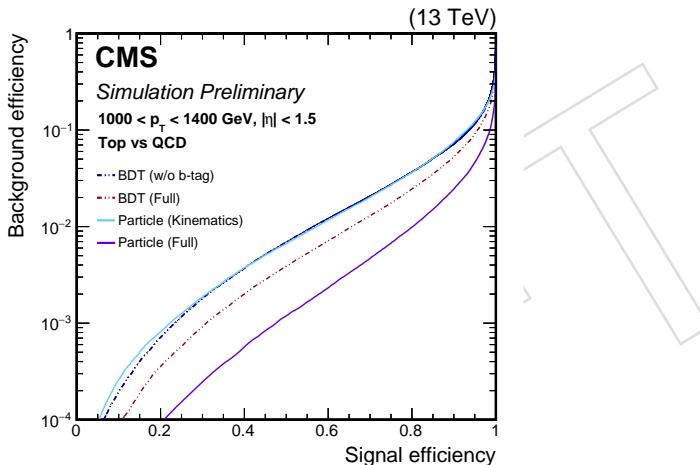


Figure 35: ROC curves of different W tagging algorithm based on BDTs or DNNs.

530 In order to gain a better understanding of the improvement in W or Z boson identification,  
 531 we performed a “evolution” study on the neural network tagger, starting with only the six  
 532 kinematic variables mentioned in Appendix A as inputs, and gradually adding more input  
 533 variables to the neural network and comparing the performance. To factor out the contribu-  
 534 tion of flavour tagging, we split the decay modes of the W or Z boson based on whether a b-  
 535 or c-quark is involved, and for background jets we only consider QCD jets that contain no b-  
 536 or c-hadrons. The comparison is shown in Figure 36. As can be seen from the plots, the big  
 537 improvement from “Particle (Kin)” to “Particle(Full)” largely comes from the addition of the  
 538 impact parameters, while the effects of adding the impact parameters are similar to only adding  
 539 the Puppi weights of the jet constituent particles in decay modes without b- or c-quarks. This  
 540 indicates that in these cases, the improvement is actually originating from a better pile-up iden-  
 541 tification and rejection at particle level, which the DNN can infer from the impact parameters.  
 542 Note that in these studies, CHS jets are used as the inputs, thus the pile-up removal may not be  
 543 optimal for boosted jet identification. This motivated us to switch from CHS jets to Puppi jets  
 544 as inputs, after verifying that there is no loss in performance as shown in Figure 37.

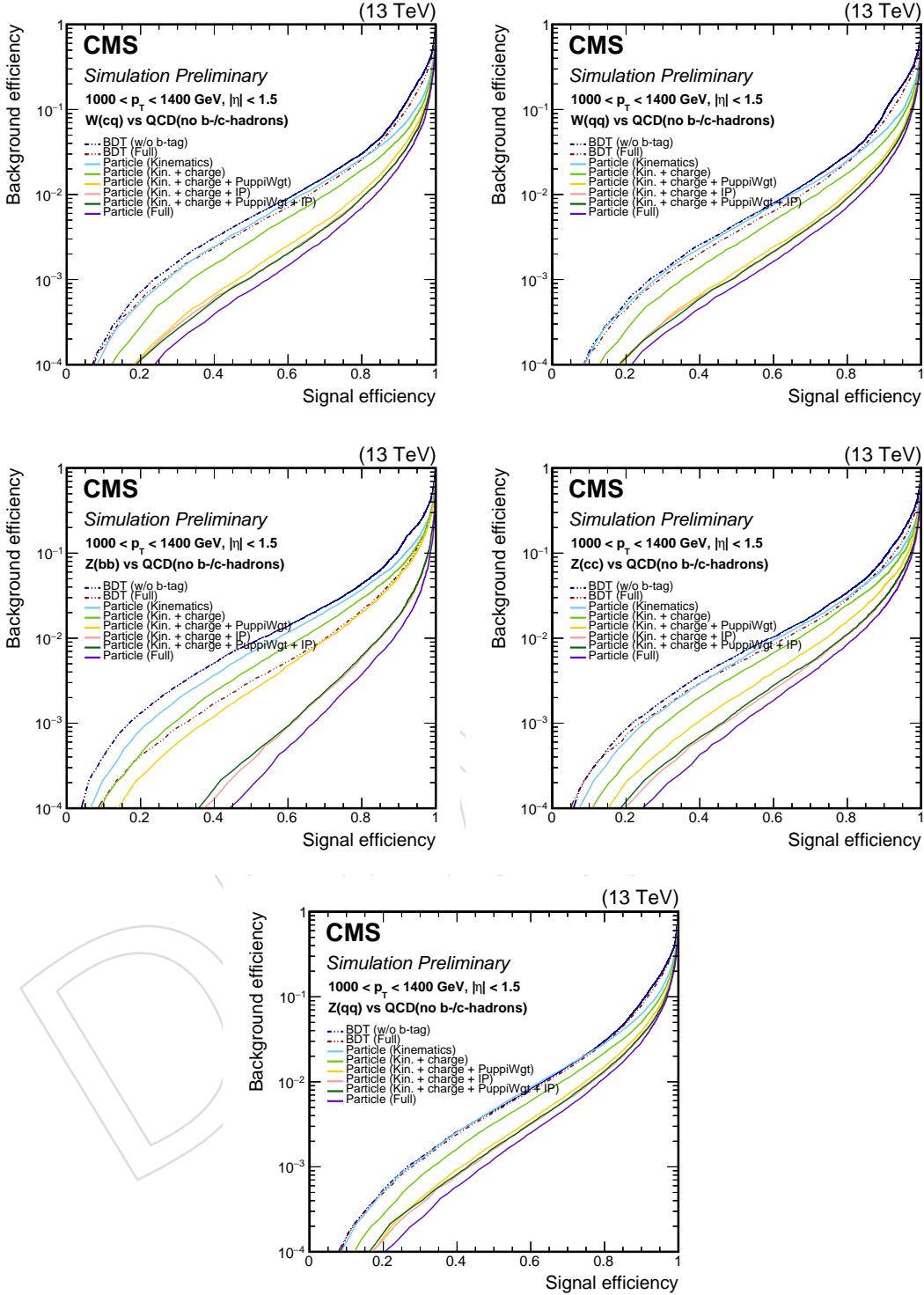


Figure 36: Comparison of DNNs using different sets of input features in terms of receiver operating characteristic (ROC) curves for CHS jets with  $1000 < p_T < 1400 \text{ GeV}$ ,  $|\eta| < 2.4$ . The first row is for  $W$  jets vs light QCD jets (i.e., no b- or c-hadrons), split by the  $W$  boson decay mode (left:  $W \rightarrow cq$ , right:  $W \rightarrow qq$ ). The second and third rows are for  $Z$  jets vs light QCD jets, split by the  $Z$  boson decay mode (second row, left:  $Z \rightarrow b\bar{b}$ , right:  $Z \rightarrow c\bar{c}$ ; third row:  $Z \rightarrow q\bar{q}$ ). “charge” refers to the electric charge of each jet constituent particle, “PuppiWgt” refers to the Puppi weight, and “IP” refers to the horizontal and longitudinal impact parameters and their significances of the track associated to each charged jet constituent particle.

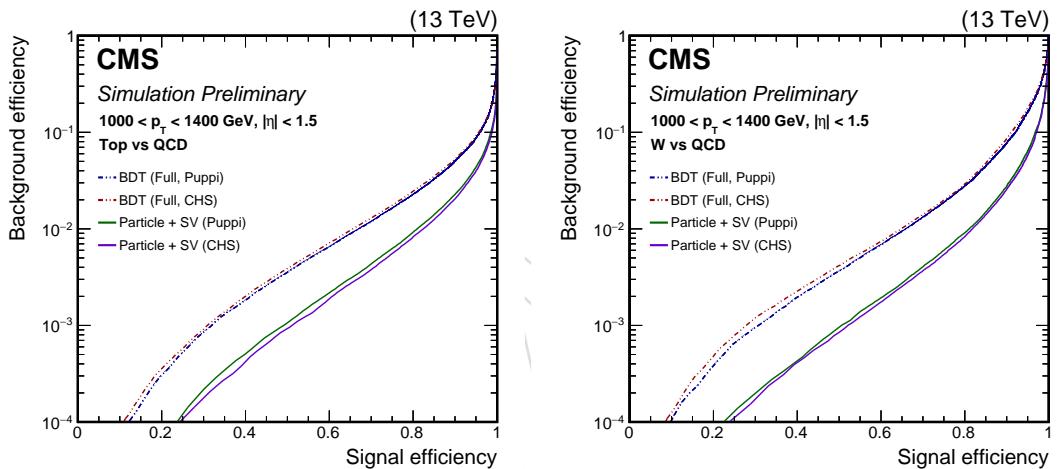


Figure 37: Comparison of the performance between CHS jets and Puppi jets for top quark tagging (left) and W boson tagging (right).