CMS Draft Analysis Note

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

2024/04/10
Archive Hash: untracked
Archive Date: 2024/04/10

Measurements of the top quark Energy Asymmetry (A_E) and Incline asymmetry (A_I) in proton-proton collisions at $\sqrt{s}$ =13 TeV

Hugo Alberto Becerril Gonzalez, Abideh Jafari, Maria Aldaya Martin, Ying An, and Andreas Meyer
DESY, Hamburg, Germany

Abstract

This note presents the first results of measurements of the charge asymmetry in inclusive $t\bar{t}$ production in association with a high p_T jet in proton-proton (pp) collisions at a center-of-mass energy of $\sqrt{s}$=13 TeV with data collected with the CMS detector.

Contents

1 Introduction

The study of top-quark physics at the Large Hadron Collider (LHC) offers a highly effective approach for unraveling potential new physics. Within this set of observables, the charge asymmetry stands out as it originates from the distinct likelihood of top and anti-top quarks being produced in a specific phase-space region, although, at the LHC is highly diluted in proton-proton collisions because most of the $t\bar{t}$ pair produced originates from gluon-gluon fusion, where there is no charge asymmetry at any order of perturbation. Differing from the previously attempts to measure the effect of charge asymmetry [1], the energy asymmetry (A_E) and incline asymmetry (A_I) are new observables to explore it in top-quark pair production in association with one additional high-p_T jet. The presence of the additional jet has the advantage that it is non-vanishing at LO and NLO calculations are available. The asymmetries are defined as:

$$A_E = \frac{\sigma_A(\Delta E \geq 0)}{\sigma_S}, \tag{1}$$

$$A_I = \frac{\sigma_A(y_{t\bar{t}j} > 0) - \sigma_A(y_{t\bar{t}j} < 0)}{\sigma_S}. \tag{2}$$

where $\sigma_A(x)$ and $\sigma_S(x)$ are the difference and sum of events passing the x requirement, θ_j is defined as the angle between the jet direction and the positive z-axis, $y_{t\bar{t}j}$ is the rapidity of the $t\bar{t}j$ system and ϕ is the angle between the plane formed by an incoming parton and the additional jet and the plane formed by the additional jet and one top. All the variables are defined in the $t\bar{t}j$ center-of-mass rest frame. Fig. 1 shows the kinematics of the $pp \to t\bar{t}j$

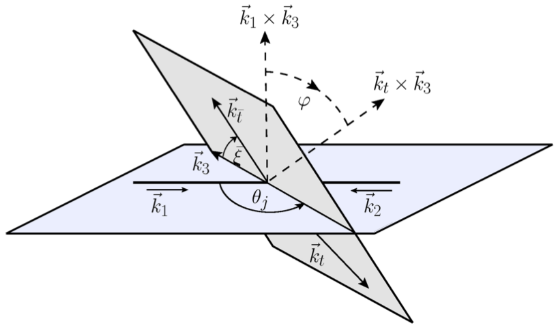

Figure 1: Kinematics of the $pp \to t\bar{t}j$ [1].

In this note the semileptonic decay of the top quark pair in resolved and boosted regimes are going to be considered. the note is organized as follows: 2 list the data and MC samples employed, 3 describes the different physics objects needed for the measurement, 4 defines the event selection and categorization, 6 shows all the systematic uncertainties included in the final measurements, **??** presents the statistical analysis and how the asymmetries are extracted from it and finally **??** shows the results and conclusions.

2 Data and Monte Carlo samples

The event samples used for the analysis were produced by CMS using the `CMSSW_10_6_28` software release [2]. We use the MINIAODv2 format of the Monte Carlo (MC) and data samples.

The Global Tags used to process the data are:

- DATA
- MC
  - UL16preVPF: 106X mcRun2 asymptotic preVFP v11
  - UL16postVPF: 106X mcRun2 asymptotic v17
  - UL17: 106X mc2017 realistic v9
  - UL18: 106X upgrade2018 realistic v16 L1v1

2.1 Datasets

The SingleMuon and EGamma datastreams are used for the data collected by CMS between 2016 and 2018 in pp collisions at $\sqrt{s} = 13$ TeV (25 ns bunch crossing) with a magnetic fields of $B = 3.8$ T. We consider only the luminosity sections which pass data-quality certification of the JSON files:

- `Cert_271036284044_13TeV_Legacy2016_Collisions16_JSON_UL16preVFP_normtag.root`
- `Cert_271036284044_13TeV_Legacy2016_Collisions16_JSON UL16postVFP_normtag.root`
- `Cert_294927306462_13TeV_UL17_Collisions17_GoldenJSON_normtag.root`
- `Cert_314472325175_13TeV_Legacy2018_Collisions18_JSON_normtag.root`

The data samples used are listed in Table 1 – Table 4.

Dataset	Reconstruction group
SingleElectron	Run2016B-HIPM_UL2016_MiniAODv2-v2
SingleElectron	Run2016C-HIPM_UL2016_MiniAODv2-v2
SingleElectron	Run2016D-HIPM_UL2016_MiniAODv2-v2
SingleElectron	Run2016E-HIPM_UL2016_MiniAODv2-v5
SingleElectron	Run2016F-HIPM_UL2016_MiniAODv2-v2
SingleMuon	Run2016B-HIPM_UL2016_MiniAODv2-v2
SingleMuon	Run2016C-HIPM_UL2016_MiniAODv2-v2
SingleMuon	Run2016D-HIPM_UL2016_MiniAODv2-v2
SingleMuon	Run2016E-HIPM_UL2016_MiniAODv2-v2
SingleMuon	Run2016F-HIPM_UL2016_MiniAODv2-v2

Table 1: Data samples used in the analysis for UL16preVFP

2.2 Simulated samples

The MC samples employed in this analysis have been generated in the RunIISummer20UL production campaign. The list of all the simulated samples used in the analysis is given in **??**

POWHEG [3] is used for the generation of top quark pair production at next-to-leading-order (NLO). Also the electroweak production of single top quarks in the $t-$ and $tW-$channels is simulated using POWHEG at NLO. The $s-$channel process of single top quark production is generated with MADGRAPH5 AMC@NLO [4].

Dataset	Reconstruction group
SingleElectron	Run2016F-UL2016_MiniAODv2-v2
SingleElectron	Run2016G-UL2016_MiniAODv2-v2
SingleElectron	Run2016H-UL2016_MiniAODv2-v2
SingleMuon	Run2016F-UL2016_MiniAODv2-v2
SingleMuon	Run2016G-UL2016_MiniAODv2-v2
SingleMuon	Run2016H-UL2016_MiniAODv2-v2

Table 2: Data samples used in the analysis for UL16postVFP

Dataset	Reconstruction group
SingleElectron	Run2017B-UL2017_MiniAODv2-v1
SingleElectron	Run2017C-UL2017_MiniAODv2-v1
SingleElectron	Run2017D-UL2017_MiniAODv2-v1
SingleElectron	Run2017E-UL2017_MiniAODv2-v1
SingleElectron	Run2017F-UL2017_MiniAODv2-v1
SingleMuon	Run2017B-UL2017_MiniAODv2-v1
SingleMuon	Run2017C-UL2017_MiniAODv2-v1
SingleMuon	Run2017D-UL2017_MiniAODv2-v1
SingleMuon	Run2017E-UL2017_MiniAODv2-v1
SingleMuon	Run2017F-UL2017_MiniAODv2-v1

Table 3: Data samples used in the analysis for UL17

Dataset	Reconstruction group
EGamma	Run2018A-UL2018_MiniAODv2-v1
EGamma	Run2018B-UL2018_MiniAODv2-v1
EGamma	Run2018C-UL2018_MiniAODv2-v1
EGamma	Run2018D-UL2018_MiniAODv2-v1
SingleMuon	Run2018A-UL2018_MiniAODv2_)GT36-v1
SingleMuon	Run2018B-UL2018_MiniAODv2_GT36-v1
SingleMuon	Run2018C-UL2018_MiniAODv2_GT36-v1
SingleMuon	Run2018D-UL2018_MiniAODv2_GT36-v1

Table 4: Data samples used in the analysis for UL18

Z and W boson production in association with jets (W/Z + jets) have been produced with MADGRAPH5 AMC@NLO. The MLM matching scheme is used to interface them with the shower generated with PYTHIA. The W/Z + jets samples are binned in H_T, where H_T is the summed momentum of all final state partons in the matrix element.

QCD multi-jet production has been generated with PYTHIA8 binned in H_T.

Diboson samples WW, WZ and ZZ are generated with PYTHIA8.

Other possible backgrounds are found to be negligible and therefore discarded.

All events were generated at the center of mass energy of 13 TeV. In the parton shower generated with PYTHIA, the underlying event tune CP5 [5] has been used for all SM and signal samples

2.3 MC corrections

2.3.1 HEM issue (2018)

In 2018 two HCAL modules did not work for Run C and Run D. This problem had an impact on the jet energy measurement in the region $-3 < \eta < -1.3$ and $-1.57 < \phi < -0.87$, causing jet mis-calibration and electron mis-identification. In order to account for these issues, all events with jets or leptons in this region are vetoed. For data, the veto is applied in 2018 starting from run 319077, for MC the veto is applied to all events in 2018, and then a weight $w = 0.35$ is applied to account for the affected luminosity fraction.

2.3.2 L1 Prefiring (2017 and 2016)

In simulation prefiring weights are applied to account for the L1 trigger prefiring issue for 2017 and 2016preVFP and 2016postVFP. The weights are defined as the product of the non-prefiring probability of all objects [6].

2.3.3 V+Jets NLO corrections

The W+Jets and DY+Jets MC samples are generated at leading order (LO). To address the absence of NLO QCD and electroweak (EWK) contributions, corrections are implemented based on the vector boson p_T. These corrections, from [7], are illustrated in Fig. 2.

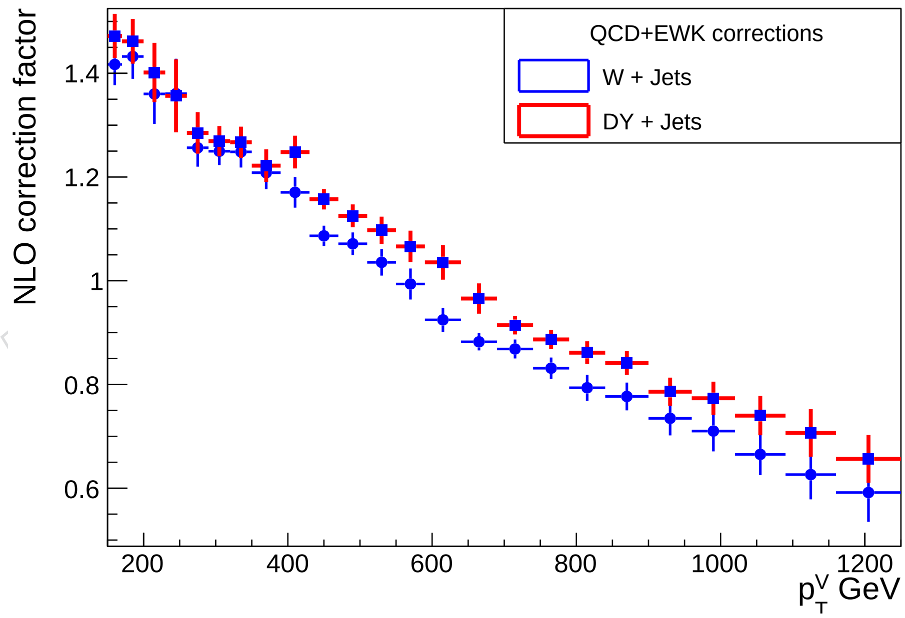

Figure 2: The NLO correction factors applied in the analysis on W/Z+Jets samples as function of the vector boson p_T.

Process	σ [pb]	dataset name
$t\bar{t}$ semi-leptonic	363.31	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
$t\bar{t}$ all-hadronic	380.11	/TTToHadronic TuneCP5 13TeV-powheg-pythia8/
$t\bar{t}$ di-leptonic	87.33	/TTTo2L2Nu TuneCP5 13TeV-powheg-pythia8/
W+Jets ($l\nu$) $70 < H_T < 100$	1271	/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8/
W+Jets ($l\nu$) $100 < H_T < 200$	1253	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/
W+Jets ($l\nu$) $200 < H_T < 400$	335.9	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/
W+Jets ($l\nu$) $400 < H_T < 600$	45.21	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/
W+Jets ($l\nu$) $600 < H_T < 800$	10.99	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/
W+Jets ($l\nu$) $800 < H_T < 1200$	4.936	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/
W+Jets ($l\nu$) $1200 < H_T < 2500$	1.156	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/
W+Jets ($l\nu$) $H_T > 2500$	0.026	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/
DY+Jets (ll) $70 < H_T < 100$	140.1	/DYJetsToLL_M-50_HT-70to100_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/
DY+Jets (ll) $100 < H_T < 200$	140.2	/DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/
DY+Jets (ll) $200 < H_T < 400$	38.39	/DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/
DY+Jets (ll) $400 < H_T < 600$	5.21	/DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/
DY+Jets (ll) $600 < H_T < 800$	1.26	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/
DY+Jets (ll) $800 < H_T < 1200$	0.56	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/
DY+Jets (ll) $1200 < H_T < 2500$	0.133	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/
DY+Jets (ll) $H_T > 2500$	0.002	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/
WW	118.7	/WW_TuneCP5_13TeV-pythia8/
WZ	46.74	/WZ_TuneCP5_13TeV-pythia8/
ZZ	16.91	/ZZ TuneCP5 13TeV-pythia8/
single t s-channel	3.36432	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/
single t t-channel top	136.02	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/
single t t-channel antitop	80.95	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/
single t tW-channel top	19.46	/ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/
single t tW-channel antitop	19.46	/ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/

Table 5: MC samples used in this analysis

Process	σ [pb]	dataset name
QCD $50 < H_T < 100$	18590000	/QCD_HT50to100_TuneCP5_PSWeights_13TeV-madgraph-pythia8/
QCD $100 < H_T < 200$	23610000	/QCD_HT100to200_TuneCP5_PSWeights_13TeV-madgraph-pythia8/
QCD $200 < H_T < 300$	1551000	/QCD_HT200to300_TuneCP5_PSWeights_13TeV-madgraph-pythia8/
QCD $300 < H_T < 500$	324300	/QCD_HT300to500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/
QCD $500 < H_T < 700$	30340	/QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraph-pythia8/
QCD $700 < H_T < 1000$	6440	/QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/
QCD $1000 < H_T < 1500$	1118	/QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/
QCD $1500 < H_T < 2000$	108	/QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/
QCD $H_T > 2000$	11	/QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraph-pythia8/

Table 6: QCD samples used in this analysis

3 Object Reconstruction

The event reconstruction is based on the Particle Flow (PF) algorithm [8]. All, electrons, muons, jets and missing E_T, are reconstructed from the collection of pf candidates in an event. Every PF candidate in the event is categorized as an electron, a muon, a photon, a charged hadron, or a neutral hadron. Electrons and low-pT muons are reconstructed based on their respective PF candidates, while high-pT muons are identified using the TuneP algorithm. After applying the pileup mitigation algorithm, jet clustering is executed on all PF candidates.

In the next sections, we delve into a thorough description of the identification process for each physics object.

3.1 Primary vertex and Noise filters

Primary vertices are reconstructed through the clustering of tracks using a deterministic annealing algorithm. Each vertex candidate must meet specific quality criteria: $\sqrt{x^2 + y^2}$ ¡ 2 cm, $|z| < 24$ cm, and NDOF > 4, where NDOF represents the weighted number of tracks used in the vertex reconstruction. The primary vertex of the event is determined as the candidate with the highest $\sigma_{tracks} p_T^2$ of the associated track-level physics objects. Any additional vertices are designated as pile-up.

Also, all events are filtered from noise using the recommended filters [9]

- primary vertex filter
- beam halo filter
- HBHE noise filter
- HBHE iso noise filter
- Ecal TP filter
- Bad PF muon filter
- Bad PF muon Dz filter
- ee badSC noise filter
- Ecal bad calibration filter (only for UL17/UL18)

Furthermore, by employing the suggested minimum bias cross-section of 69.2 mb, we adjust for the disparity in the pileup profile between the generated MC samples and the actual data.

3.2 Leptons

The analysis considers two types of charged leptons: muons and electrons. Below, we outline the criteria utilized to identify muon and electron candidates in this search. These criteria encompass a kinematic selection, involving cuts on transverse momentum and pseudo-rapidity, and a set of identification (ID) criteria.

It is important to note that the definition of lepton candidates does not include isolation requirements. This decision is driven by the expectation that prompt leptons resulting from the decay of high-pT top quarks are likely to be poorly isolated. This lack of isolation is attributed to their close proximity in space to the products of the b-quark decay.

3.2.1 Muons

The muon candidates used in the analysis have a $p_T > 55$ GeV and $|\eta| < 2.4$ and do not have any isolation requirement. Also, the selected muons have to satisfy the ”CutBasedIdGlobal-

HighP"t criteria recommended by the Muon POG REF, which correspond to the following set of identification cuts:

- The muon candidate is reconstructed as a global muon.
- At least one muon-chamber hit is included in the global-muon track fit or in the TuneP fit
- Require muon segments in at least two muon stations.
- If there is only one matched station it must be a tracker muon and satisfy at least one of the following conditions: 0 or 1 expected matched station based on the extrapolation of the inner track, the matched station should not be the first one, or has at least two matched RPC layers
- The p_T relative error of the muon best track is less than 30%.
- Its tracker track has transverse impact parameter $d_{xy} < 2$ mm with respect to the primary vertex and a longitudinal distance $d_z < 5$ mm
- The number of pixel hits should be greater than 0.
- Number of tracker layers with hits greater than 5.

The trigger employed is the logical "OR" combination of the HLT paths HLT_Mu50 v* and HLT_TkMu50 v* for the year 2016, and the logical "OR" combination of HLT Mu50 v*, HLT_TkMu100 v*, and HLT_OldMu100 v* for the years 2017 and 2018. However, in the RunB of 2016 and 2017, certain trigger paths were missing. Consequently, only the path HLT_Mu50 v* is utilized for data in the 2016 RunB (run numbers below 274889) and 2017 RunB (run numbers below 299329). This adjustment also applies to a fraction of the Monte Carlo (MC) corresponding to the percentage of RunB data, amounting to 14.29% in 2016 and 11.58% in 2017.

We apply the recommended ID, reconstruction, isolation and trigger scale factors provided by the Muon POG for 2016 [10], 2017 [11] and 2018 [12]. The specific files used for the four eras UL16preVFP, UL16postVFP, UL17 and UL18 can be found in [13–16] respectively.

3.2.2 Electron

The electron candidates in our analysis are divided into low-pT and high-pT categories. Low-pT electrons must satisfy $50 < p_{Te} < 120,$GeV and $|\eta_e| < 2.5$. These candidates must also pass the MVA-based ID "mvaEleID Fall17 iso V2 wp80" from the EGamma POG [65], incorporating an isolation cut with "wp80" representing an 80% signal efficiency.

For high-p electrons, the criteria include $p_e > 120,$GeV, $|\eta_e| < 2.5$, and "mvaEleID Fall17 noIso V2 wp80" without isolation [65]. Electrons situated in the transition region between the barrel and endcap of ECAL ($1.44 < |\eta| < 1.57$) are excluded.

The HLT trigger path for low-pT electrons is HLT_Ele27(35)[32]_WPTight_Gsf_v* for the years 2016(2017)[2018], respectively. In the high-pT range, the trigger involves the "OR" combination of HLT paths: HLT_Ele115_CaloIdVT_GsfTrkIdT_v*, HLT_Photon175(200)_v*, and HLT_Ele27(35) [32]_WPTight_Gsf_v* for 2016(2017/2018).

For a portion of 2017 data, where the HLT_Ele115_CaloIdVT_GsfTrkIdT_v* path is unavailable, it is substituted with HLT_Ele35_WPTight_Gsf v*. This substitution is applied to 2017 RunB data (run numbers below 299329) and a corresponding fraction of MC equal to 11.58

Scale factors for electron ID and reconstruction are provided by the EGamma POG [17]. The trigger combination scale factors are derived from an $t\bar{t}(e\mu)$ control region and are endorsed by the EGamma POG. Additional details can be found in Appendix E.

Period	JEC	JER
UL18	Summer19UL18_v5	Summer19UL18_JRV2
UL17	Summer19UL17_v5	Summer19UL17_JRV2
UL16postVFP	Summer19UL16_v7	Summer20UL16_JRV3
UL16preVFP	Summer19UL16APV_v7	Summer20UL16APV_JRV3

Table 7: Caption

3.3 Jets

The analysis utilizes two jet collections, reconstructed with the anti-kT (AK) algorithm, employing two different cone parameter choices: $R = 0.4$ and $R = 0.8$. In subsequent discussions, we refer to the first (AK4) and second (AK8) sets of jets as small-radius and large-radius jets, respectively. AK8 jets are constructed using the ”Pile Up Per Particle Identification” (PUPPI) algorithm to alleviate the impact of pile-up. In PUPPI, each PF candidate is assigned a weight representing the probability of the particle originating from the primary or a pileup vertex. This weight is applied to the 4-momentum of each PF candidate before the jet clustering process. The analysis employs the latest version of the PUPPI algorithm, PUPPI v15 [18].

Additionally, AK4 jets are reconstructed using the ”Charged Hadron Subtraction” (CHS) algorithm. In this approach, the input to the clustering algorithm uses the list of all PF candidates not identified as pileup hadrons. CHS jets are used in the implementation of the b-tagging algorithm and the associated scale factors.

Jets are corrected in data and MC, to correct for —η—-dependence (L2Relative) and p_T-dependence (L3Absolute) of the detector response. An additional set of jet energy corrections (JEC) is applied only in data events to account for residual differences between data and simulation (L2L3Residuals)

AK4CHS jets are corrected with the JECs of AK4PF jets with “‘CHS” (“AK4PFchs”). AK8PUPPI jets are corrected with the JECs of AK8PF jets with PUPPI (“AK8PFpuppi”) and the AK4 subjets are corrected with AK4PFPuppi correction, respectively [19].

All the candidates in all jet collections are required to pass minimal jet quality criteria (PFJet-ID Tight working point [20]). The jet energy resolution (JER) in the simulation is also corrected as recommended in [21].

Table 7 shows the version of the JER’s and JEC’s epplied in this analisys for each year.

Given that the PF candidates eventually identified as leptons (muons or electrons) are initially involved in the jet clustering process, it becomes necessary to exclude lepton candidates from the jet collections used in the analysis—a procedure commonly known as ”jet-lepton cleaning.” For AK4 and AK8 PUPPI jets, this cleaning is executed by matching the keys of the jet constituents with those of the leptons. Upon finding a match, the 4-momentum of the lepton candidate is subtracted from the uncorrected 4-momentum of the corresponding jet. The resulting 4-vector is then rescaled using the appropriate Jet Energy Corrections (JECs). This candidate-based approach is crucial for the accurate reconstruction of small-radius jets. In the leptonic decay of a high-p_T top quark, a prompt lepton is expected to have a very small angular separation from a small-radius jet originating from the decay of a b-quark. Consequently, the conventional requirement of a minimum angular separation between leptons and small-radius jets would be unsuitable for this search. AK4 jet candidates are mandated to have $p_T > 50$ GeV and $|\eta| < 2.5$, while AK8 jet candidates must satisfy $p_T > 400$ GeV and $|\eta| < 2.5$.

3.4 Missing transverse momentum

The event's missing transverse momentum is determined by the negative vector sum of the transverse momentum (p_T) of all PF candidates, adjusted by their PUPPI weights (PUPPI p_T^{miss}). To accurately incorporate the impact of the Jet Energy Corrections (JECs) on the p_T^{miss} measurement, a Type-1 correction is applied. Furthermore, in simulation, the absolute value of p_T^{miss} is corrected through the smearing of jet energy resolution.

3.5 *b*-tagging

The identification of jets originating from the decay of B-hadrons (*b*-tag) employs a deep neural network multi-classification algorithm that utilizes information from the tracker and the calorimeters [22]. In this analysis, we use the tight working point of the deepNN *b*-tagging on AK4 jets with $p_T > 50$ GeV and $|\eta| < 2.5$. Differences in the *b*-jet identification efficiency and the misidentification probability of a non-*b* jet as a *b* jet in simulation compared to data are corrected by reweighing MC events using scale factors that depend on the jet flavor, the jet p_T, and the jet η. Dedicated scale factors for the DeepJet algorithm are derived following the directions of the *b* Tag and Vertexing POG group [23] and Method 1a.

The probability of a given jet configuration in simulation and data is defined as:

$$P(MC) = \prod_{i=tagged} \epsilon_i \prod_{j=untagged} (1 - \epsilon_j), \tag{3}$$

$$P(DATA) = \prod_{i=tagged} \epsilon_i * SF_i \prod_{j=untagged} (1 - \epsilon_j) * SF_j. \tag{4}$$

Here, ϵ_j denotes the MC *b*-tagging efficiency (MC mistagging efficiency) and SF_i (SF_j) are the DeepJet scale factors for heavy-flavor (light-flavor) jets, provided by the b Tag and Vertexing POG. The tagging efficiencies are calculated for each simulated process separately and are defined as:

$$\epsilon_f(p_T^{jet}, \eta^{jet}) = \frac{N_f^{b-tagged}(p_T^{jet}, \eta^{jet})}{N_f^{total}, (p_T^{jet}, \eta^{jet})} \tag{5}$$

where N^{Total} and $N^{b-tagged}$ are the total number and the number of *b*-tagged jets, respectively, of flavor f (HF $\rightarrow$ bc, LF $\rightarrow$ usdg) in a given jet p_T and η bin. The event weight is calculated as $w = \frac{P(DATA)}{P(MC)}$. The efficiencies for the phase space in this analysis are shown in Appendix **??**.

3.6 t/W-tagging

A machine-learning-based tagger, DeepAK8 [24], is used in this analysis to tag t/W originated jets. It leverages jet substructure variables and jet constituents as inputs to extract information about the particles generating the jets. To counteract mass sculpting effects, the mass-decorrelated (MD) version of the tagger is employed. An AK8PUPPI jet candidate gets the t/W label when subjected to the DeepAK8-MD algorithm using a working point associated with a mistag rate of 1%.

AK8Jets also pass the next requirements: $p_T > 400$ GeV and 105 GeV $< M_{SD} <$ 210 GeV (60 $< M_{SD} <$ 105), with M_{SD} denoting the ungroomed mass of the AK8 PUPPI jet, as determined by the soft-drop algorithm. The cut on the M_{SD} are orthogonal for the W and t tagging so a jet cannot be t-tagged and W-tagged simultaneously.

Data/MC corrections for the efficiency of the t/W-tagging selection are included in the analysis. The SFs are provided by Christopher Matthies (Universitat Hamburg), who derives the official SFs for UL [25].

4 Event selection

The event selection target the creation of a $t\bar{t}$ pair decaying to the l+jets channel ($l = e, \mu$), where one top quark undergoes decay into a b quark and a W boson, subsequently resulting in a lepton (either electron or muon) and a neutrino. The other top quark similarly decays into a b quark and a W boson, but this W boson then decays into two quarks. This analysis considers both the resolved and boosted topologies.

In the resolved scenario, all decay products from the two top quarks are distinct and can be reconstructed as individual entities. However, in the boosted regime, the decay products of highly energetic tops or W bosons tend to merge or become collimated, and the quarks from the hadronic decay can be reconstructed within a single large-radius jet. In the leptonic decay, the lepton can be found to be proximate in ΔR (angular separation) to the jet originating from the b quark.

4.1 Baseline events selection

4.1.1 Electron channel

The following requirements are applied in the electron channel:

- the event passes one of the single electron trigger paths described in Sec 3
- exactly one electron with $p_T > 80$ GeV, MVA-based ID wp80 and $|\eta_{SC}| < 2.5$
- passing 2D cut defined as:

$$\Delta R_{min}(l, jet) > 0.4 || p_T^{rel}(l, jet) > 25 GeV \tag{6}$$

  where ΔR is the minimum anlgular distance of the lepton candidate with respect to all AK4 jets, and $p_T^{rel}(l, jet)$ is the transverse momentum of the lepton with respect to the closest AK4 jets in ΔR
- at least one AK4 jet with $p_T > 150$ GeV and $|\eta| < 2.5$
- a second with AK4 jet with $p_T > 50$ GeV and $|\eta| < 2.5$
- at least on b-tagged jet
- $E_T^{miss} > 50$ GeV
- $H_T^{lep} > 150$ GeV
- in case AK8 Jet exists, must have $p_T > 400$ GeV and $|\eta| < 2.5$

4.1.2 Muon channel

The following requirements are applied in the muon channel:

- the event passes one of the single muon trigger paths described in Sec 3
- exactly one muon with $p_T > 55$ GeV, CutBasedIdGlobalHighPt and $|\eta| < 2.4$
- passing 2D cut defined as:

$$\Delta R_{min}(l, jet) > 0.4 || p_T^{rel}(l, jet) > 25 GeV \tag{7}$$

  where ΔR is the minimum anlgular distance of the lepton candidate with respect to all AK4 jets, and $p_T^{rel}(l, jet)$ is the transverse momentum of the lepton with respect to the closest AK4 jets in ΔR
- at least one AK4 jet with $p_T > 150$ GeV and $|\eta| < 2.5$

- a second with AK4 jet with $p_T > 50$ GeV and $|\eta| < 2.5$
- at least on b-tagged jet
- $E_T^{miss} > 50$ GeV
- $H_T^{lep} > 150$ GeV
- in case AK8 Jet exists, must have $p_T > 400$ GeV and $|\eta| < 2.5$

4.2 Kinematic reconstruction of the $t\bar{t}$ system

The reconstruction of the $t\bar{t}$ system involves assigning the four-vectors of the reconstructed final-state objects (charged lepton, MET, jets) to either the leptonically or hadronically decaying top quark. Specifically, the charged lepton and MET are allocated to the leptonically decaying top quark. Here, MET represents the transverse component of the neutrino's momentum.

Assuming the W bosons are on-shell, it allows for the derivation of a quadratic equation for the z-component of the neutrino's momentum.

$$p_{z,\nu}^{\pm} = \frac{\mu p_{z,l}}{p_{T,l}^2} \pm \sqrt{\frac{\mu^2 p_{z,l}^2}{p_{T,l}^4} - \frac{E_l^2 p_{T,\nu}^2 - \mu^2}{p_{T,\nu}^2}} \quad (8)$$

where p_l and p_ν are the four-momenta of the charged lepton and the neutrino, respectively, and $\mu = \frac{M_W^2}{2} + p_{T,l} p_{T,\nu} cos\Delta\phi$ has either 2 or 0 real solutions. When a real solution is not present, the real part of the complex solutions becomes the chosen option. In cases where two real solutions exist, both scenarios are tested, effectively leading to a doubling of hypotheses for that specific event.

The assignment of jets follows like this: in events lacking top-tagged AK8 jets, all potential combinations of AK4 jets are constructed for both the leptonic and hadronic legs of the decays. However, in events featuring more than 10 jets, only the top 10 leading jets are taken into consideration.

In events where a top-tagged AK8 jet is present, it is designated for the hadronic leg of the decay. Only AK8 jets that do not overlap with leptons (ΔR(AK8 jet, l) > 0.8) are taken into account.

In events where a W-tagged AK8 jet is present, it is designated for the hadronic leg of the decay and the hadronically decaying top is recontructed by combining the AK8jet wih one AK4 jet well separated from the W-tagged jet. Only AK8 jets that do not overlap with leptons (ΔR(AK8 jet, l) > 0.8) are taken into account.

For the leptonic leg of the decay, AK4 jets that are sufficiently separated from the top/W-tagged AK8 jet (ΔR(AK8 jet, AK4 jet) > 1.2) are considered. Subsequently, all potential combinations involving AK4 jets assigned to the leptonic decay are constructed.

The final step involves selecting only one $t\bar{t}$ hypothesis for each event. This selection hinges on the closeness of the reconstructed top quark masses to the true top quark mass. The preferred hypothesis is the one that exhibits the smallest χ^2 value, where:

$$\chi^2 = \chi_{lep}^2 + \chi_{had}^2 = [\frac{M_{lep} - \bar{M}_{lep}}{\sigma_{\bar{M}_{lep}}}] + [\frac{M_{had} - \bar{M}_{had}}{\sigma_{\bar{M}_{had}}}] \quad (9)$$

The M_{lep} and M_{had} are the invariant masses of the reconstructed leptonic and hadronic top quarks, respectively. The parameters $\bar{M}_{lep}$, $\bar{M}_{had}$, $\sigma_{\bar{M}_{lep}}$ and $\sigma_{\bar{M}_{had}}$ had are obtained from simulation by matching the reconstructed objects to generator-level particles. Details on the extraction of the parameters can be found in Appendix XX.

To remove events with mis-reconstructed top quarks and to reduce the background contribution, a selection cut of $\chi^2 < 30$ is applied.

In a subsequent step, events successfully reconstructed in either the boosted or resolved regime are examined for the presence of an additional high-p_T jet (j_{add}) within the central region. Jets utilized for the reconstruction of the $t\bar{t}$ system are excluded from consideration. Furthermore, this additional jet must satisfy specific criteria: $p_T > 100$ GeV, $|\eta| < 2.5$, and it must be sufficiently distant from both the top quark and antiquark, imposing $\Delta R(j_{add}, t(\bar{t})) > 1.2$.

If multiple such jets are present within an event, the jet with the highest transverse momentum is designated as jadd. Events lacking a suitable candidate for j_{add} are discarded at this stage.

4.3 Reconstruction of $t\bar{t}+j$ system at gen level

The identification of the top quark and antiquark is based on the parton information within the event. Subsequently, the event's additional jet is sought through a sequential procedure utilizing the following search patterns. If a matching jet is discovered, further steps are skipped. In situations where multiple jets within an event meet a requirement, the jet with the highest transverse momentum is selected. The selection steps for finding j_{add} follow a specific order:

- $p_T(j_{add}) > 100$ GeV, $|\eta_{j_{add}}| < 2.4$, $\Delta R(t, j_{add}) > 1.2$, $\Delta R(\bar{t}, j_{add}) > 1.2$
- $p_T(j_{add}) > 100$ GeV, $|\eta_{j_{add}}| < 2.4$, $\Delta R(t, j_{add}) > 0.8$, $\Delta R(\bar{t}, j_{add}) > 0.8$
- $|\eta_{j_{add}}| < 2.4$, $\Delta R(t, j_{add}) > 0.8$, $\Delta R(\bar{t}, j_{add}) > 0.8$
- $\Delta R(t, j_{add}) > 0.8$, $\Delta R(\bar{t}, j_{add}) > 0.8$
- $\Delta R(t, j_{add}) > 0.4$, $\Delta R(\bar{t}, j_{add}) > 0.4$

If no additional jet is found at gen level, the event is rejected

Figs **??**-**??** show control plots after the event selection and reconstruction.

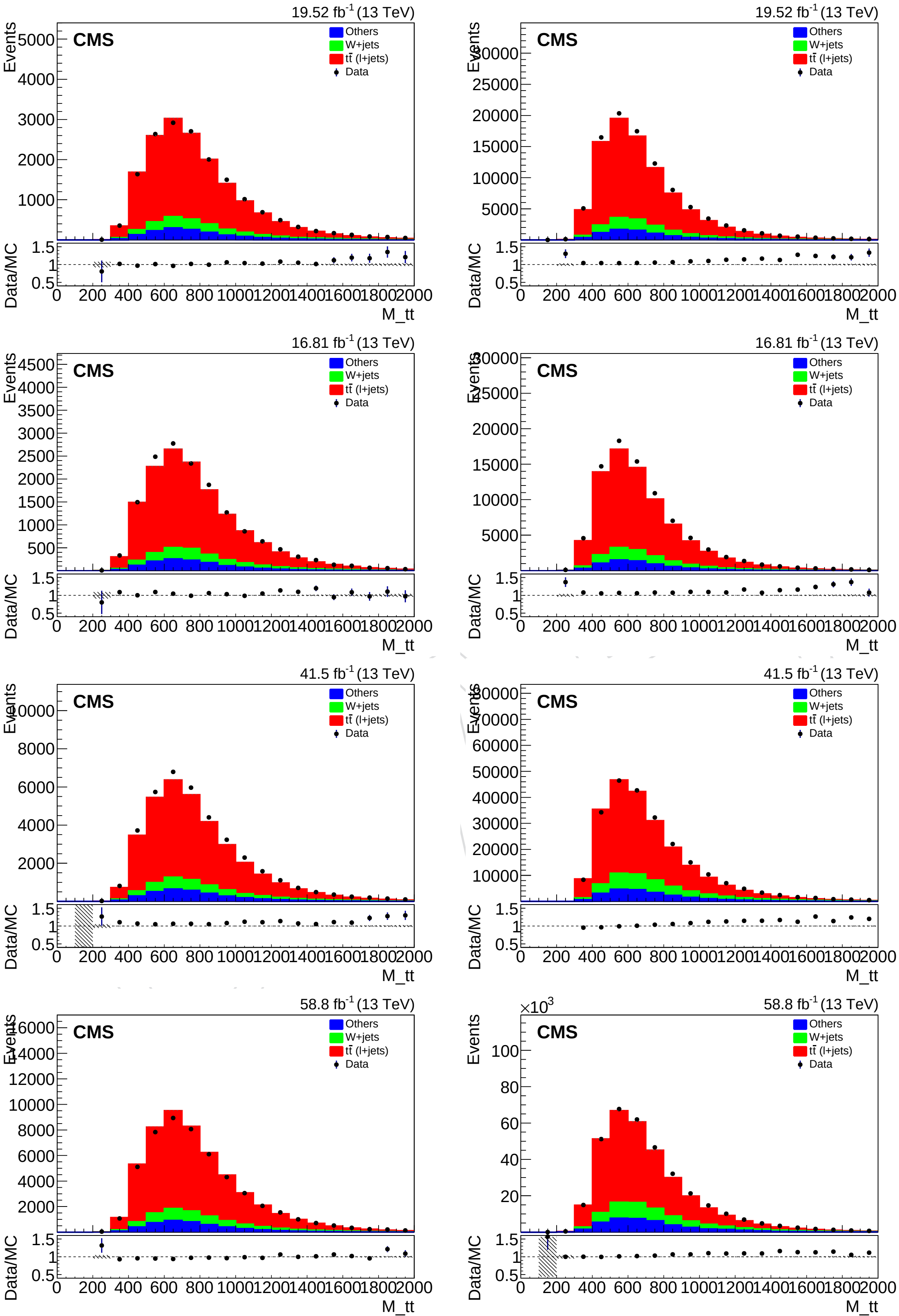

Figure 3: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

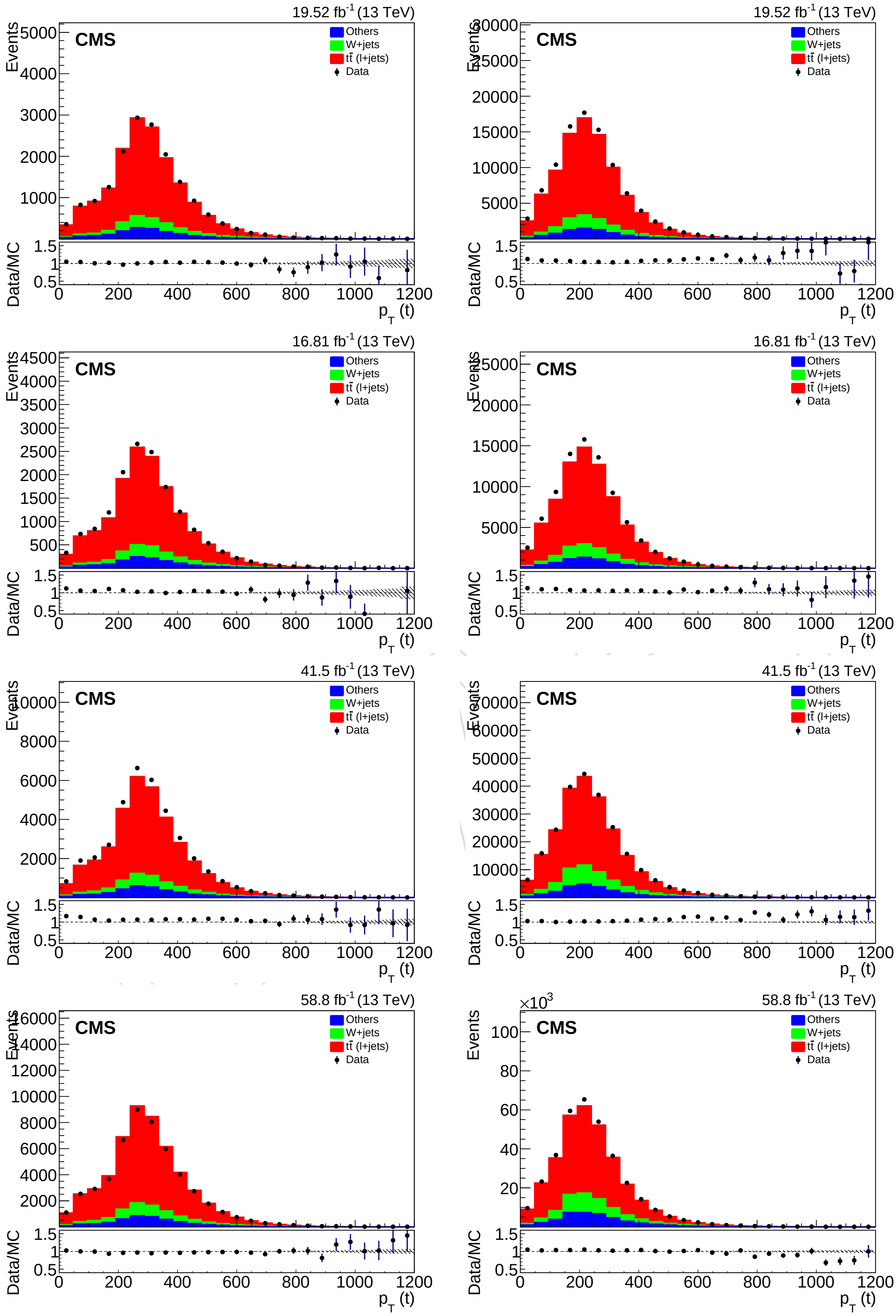

Figure 4: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

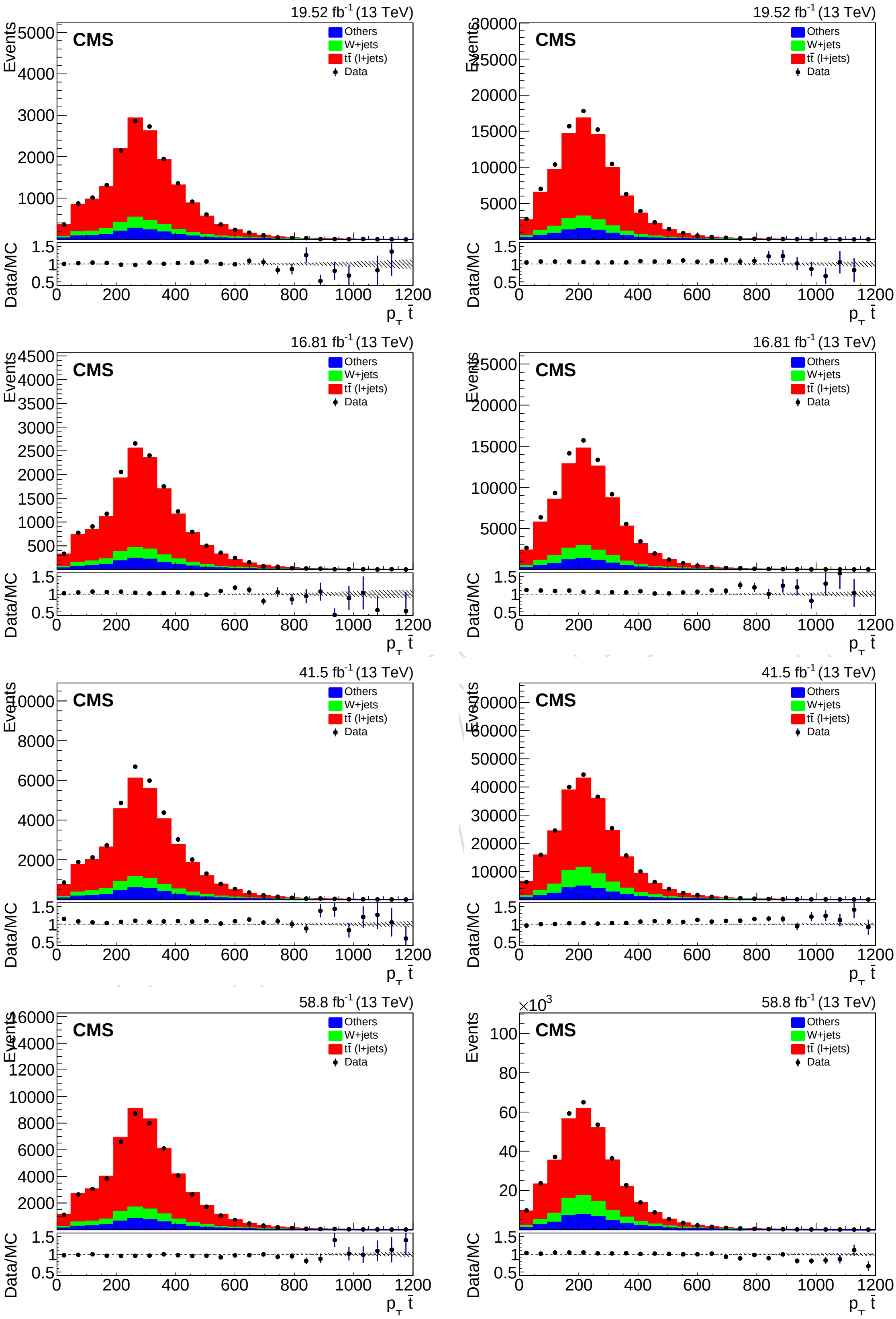

Figure 5: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

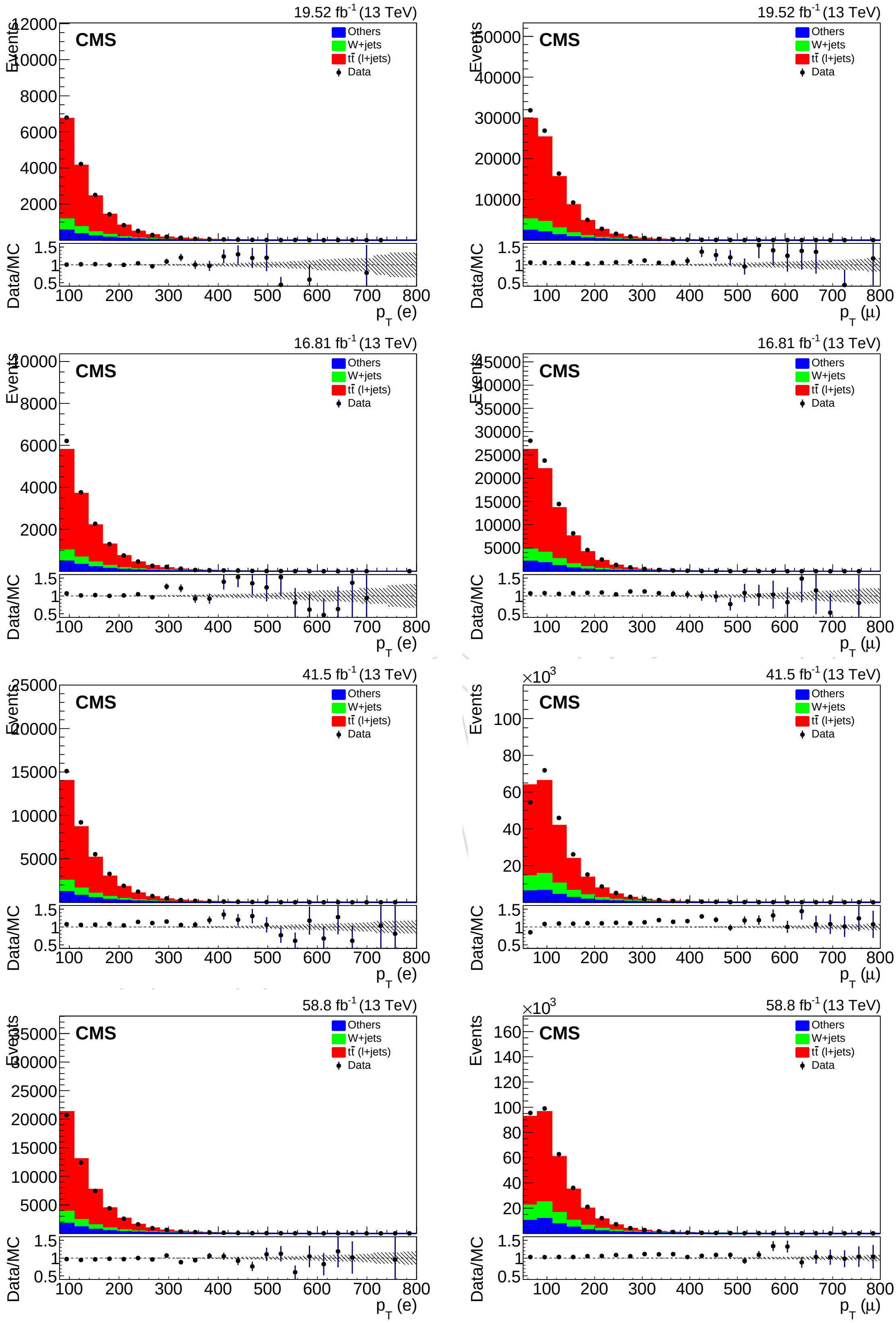

Figure 6: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

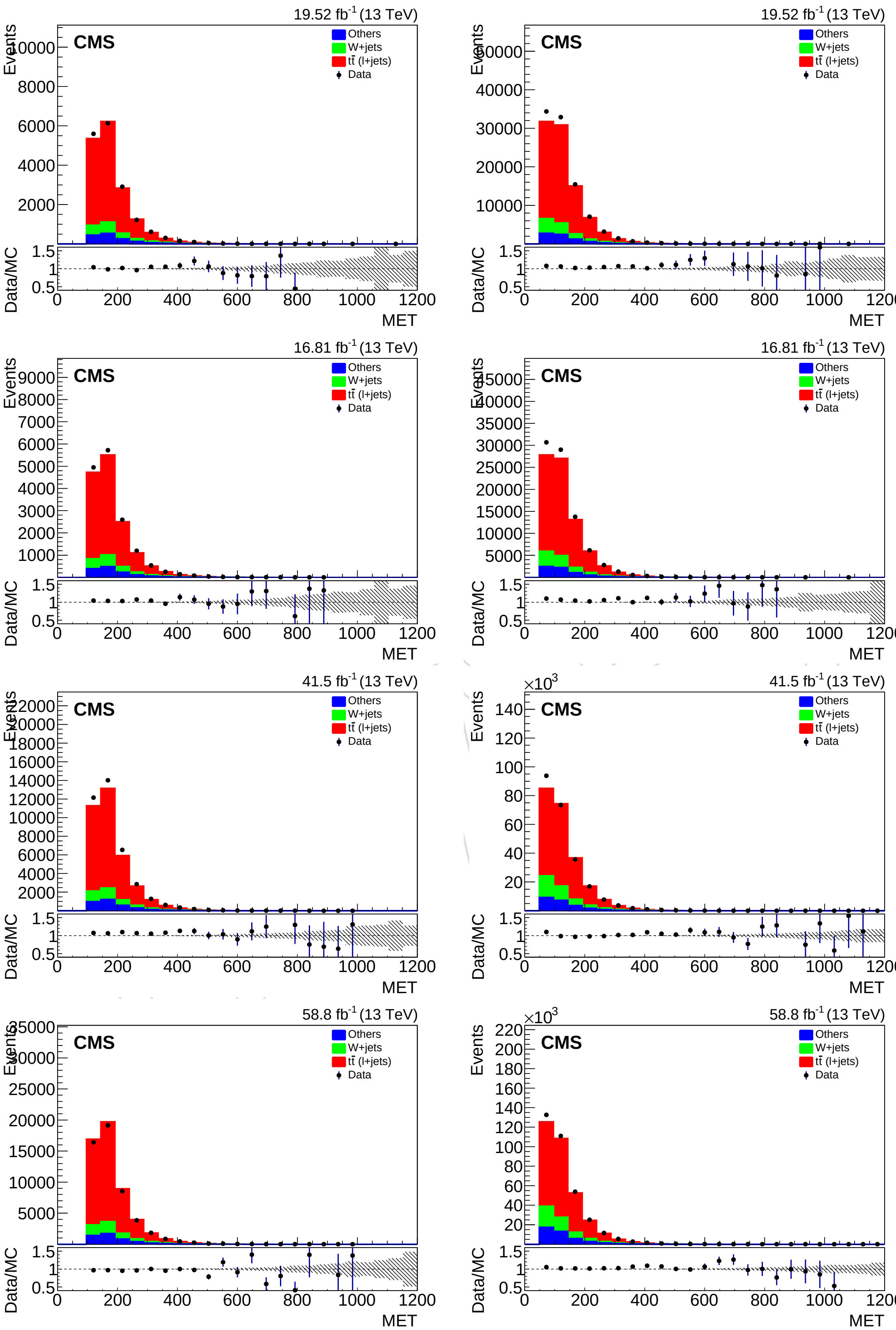

Figure 7: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

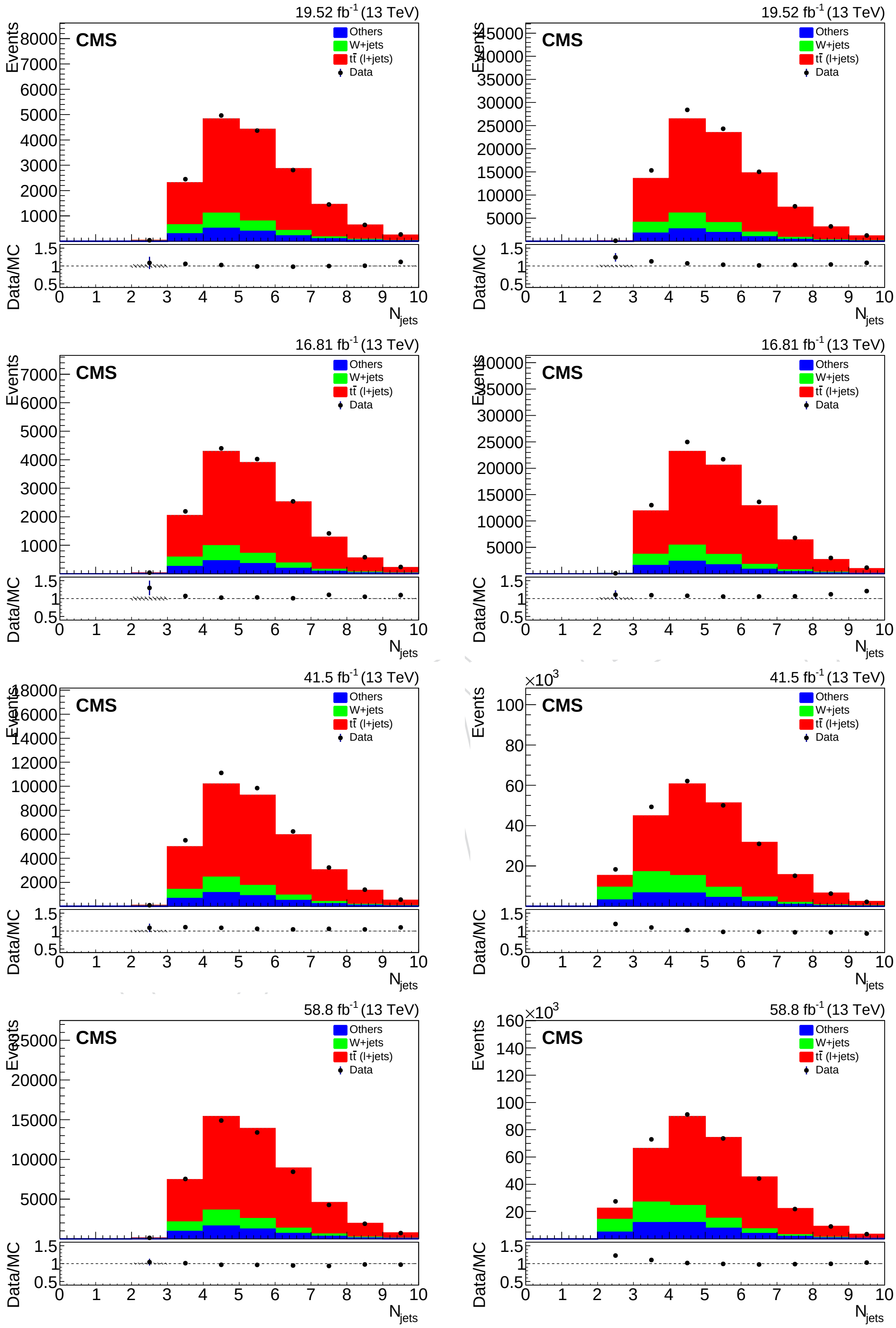

Figure 8: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

5 Unfolding

The objective of the unfolding method is to rectify reconstructed data by eliminating the bin-by-bin smearing caused by limited detector resolution and acceptance. This correction facilitates direct comparisons between experimental results and theoretical predictions. In this analysis, we employ a maximum likelihood approach implemented in the Combine Higgs tool [26] to derive the signal strength modifiers via a simultaneous fit encompassing all bins and categories within the signal candidate sample. The systematic uncertainties are integrated as correlated or independent nuisance parameters, capable of influencing the shape and/or normalization of both signal and background templates. This unfolding approach offers the advantage of constraining background process contributions through the fit, resulting in reduced systematic uncertainties compared to direct background subtraction. The subsequent subsections detail our procedure for correcting A_E and A_I to the parton level across the full phase space.

For a given channel in our analysis, the channel likelihood function L_k is defined as

$$L_k = \prod P(n_j; \sum_{i=1}^{N_{gen}} A_{ji}(\delta_u)\mu_i(\delta_u) + b_j(\delta_u)N(\delta_u)) \tag{10}$$

where:

- $P(n;\mu)$ represents the Poisson probability of observing n events when μ are expected. The indexes i and j run over the number of bins at generator level (N_{gen}) and reconstruction level (N_{reco}), respectively.
- A_{ji} is the response matrix, which gives the probability for an event reconstructed in bin j to have been produced in bin i. It is implemented by including the relevant number of reconstructed and generated simulated $t\bar{t}$ events for each entry, which are subject to the effects of the nuisance parameters. This implementation allows the matrix to account for effects from detector resolution (smearing) as well as detector acceptance and efficiency.
- μ represent the signal strengths multiplying the number of signal events at generator level.
- n_j corresponds to the number of data events in bin j.
- b_j represents the number of background events predicted in bin j.
- $N(\delta\vec{u})$ represents the priors for the nuisance parameters with normalization uncertainties u assigned a log-normal distribution and all other uncertainties a normal distribution.

5.1 Response Matrices

In the context of this study, the elements comprising the response matrix A_{ji} denote the likelihood of an event reconstructed in bin j being generated from bin i. This matrix is constructed individually for each of channel in the analysis, relying on reconstructed parameters. No specific criteria are imposed on events at the generator level. Nevertheless, a strict one-to-one correlation is upheld between reconstructed and generated events, ensuring only generated events are restricted to those that were reconstructed.

To determine if we need to do a regularized or an unregularized unfolding, we calculate the condition number for each matrix, given by:

$$cond(A) = \frac{\sigma_{\max}}{\max(0, \sigma_{\min})} \tag{11}$$

The summary of conditional numbers for each channel is presented in each Matrix. In alignment with the statistics committee's guidance, considering that these conditional numbers are below 10 [27], we conclude that employing unregularized unfolding is appropriate for this analysis.

Fig. **??**-**??** show the responce matrices used in this analysis. Migration effects can be seen in the off-diagonal elements of the 2D distribution and are all very small.

6 Systematic uncertainties

This section details the systematic uncertainties impacting both the signal and background processes in the analysis. These uncertainties have the potential to alter the normalization, the shape of physics observable distributions, or sometimes both.

The full list of systematic uncertainties is summarized in Table 8.

source	uncertainty	type	process
luminosity	$\pm$1% - 2%	norm	all
$t\bar{t}$ cross section	$\pm$5%	norm	$t\bar{t}$
single t cross section	$\pm$30%	norm	single t
W+jets cross section	$\pm$30%	norm	W+Jets
other bkg cross section	$\pm$50%	norm	DY,Diboson,QCD
Pileup reweighting	$\pm 1\sigma$	norm & shape	all
Muon reconstruction	$\pm 1\sigma$	norm & shape	all
Muon isolation	$\pm 1\sigma$	norm & shape	all
HLT muon	$\pm 1\sigma$	norm & shape	all
electron ID+isolation	$\pm 1\sigma$	norm & shape	all
electron reconstruction	$\pm 1\sigma$	norm & shape	all
HLT trigger	$\pm 1\sigma$	norm & shape	all
b-tagging	$\pm 1\sigma$	norm & shape	all
L1prefiring	$\pm 1\sigma$	norm & shape	all(2016,2017)
PDFs	$\pm 1\sigma$	shape	all
μ_R and μ_F	$\pm 1\sigma$	norm & shape	all
JES	$\pm 1\sigma$	norm & shape	all
JER	$\pm 1\sigma$	norm & shape	all
t-tagging	unconstrained	norm & shape	all
FSR	$\pm 1\sigma$	shape	$t\bar{t}$
ISR	$\pm 1\sigma$	shape	$t\bar{t}$
h_{damp}	$\pm 1\sigma$	shape	$t\bar{t}$
top-p_T reweighing	$\pm 1\sigma$	shape	$t\bar{t}$

Table 8: List of systematic uncertainties considered in the statistical analysis

- **Integrated luminosity** : A normalization uncertainty is applied for the total integrated luminosity of 137.62 fb^{-1} recorded by the CMS experiment during the period 2016-2018 of LHC Run2. The minimal correlation scheme is used, as recommended by the Lumi POG [28]. The values are listed in Table 9

2016 uncorrelated	1.0%
2017 uncorrelated	2.0%
2018 uncorrelated	1.5%
2016-2017-2918 correlated	0.6%,0.9%,2.0%
2017-2018 correlated	0.6%, 0.2%

Table 9: The correlation scheme of the luminosity uncertainties for the Run2 years 2016, 2017, 2018.

- **SM cross section**: The cross-sections for SM (Standard Model) processes are assigned pre-fit normalization uncertainties that are notably higher than those determined by theoretical calculations. This precaution is taken to accommodate inaccurately modeled kinematics, particularly at high momentum (p_T), and to address normalization uncertainties stemming from variations in μ_R, μ_F, and PDFs (Parton Distribution Functions). Specifically, values of 5%, 30%, and 50% are considered for $t\bar{t}$, W+Jets, and other backgrounds (such as DY, Diboson, QCD), respectively.

- **Pileup reweighting** : Each MC sample undergoes a reweighting process to align the number of true pileup interactions in simulation with the instantaneous luminosity profile observed in data, utilizing a minimum-bias cross-section of 69.2 mb. The associated systematic uncertainty linked to this correction is assessed by adjusting the minimum-bias cross-section by ±4.6%. This uncertainty is treated as correlated across the years
- **MuonID, isolation, reconstruction, and trigger efficiency**: Following the Muon POG recommendation [29], we use scale factors provided for muon ID, isolation, reconstruction and trigger efficiency. MC events are reweighted as a function of the muon p_T and η, except for the muon reconstruction scale factors which are provided as a function of the muon p and η. The corresponding uncertainties are obtained by varying each scale factor independently by $\pm 1\sigma$.
- **ElectronID, isolation, reconstruction, and trigger efficiency**: Following the EGamma POG recommendation [30], we use scale factors provided for electron ID+isolation and reconstruction efficiency. The electron trigger efficiency is measured in a dilepton control sample, as described in Appendix XX. MC events are reweighted as a function of the electron p_T and η. The corresponding uncertainties are obtained by varying each scale factor independently by $\pm 1\sigma$.
- **L1Prefiring** : A scale factor is applied to each MC event for the 2016 and 2017 data-taking periods to account for the L1 trigger prefiring issue. The recipe to obtain the nominal values of the scale factor together with the upward and downward variations is explained in [6].
- **PDFs** : Simulated samples for both background and signal processes are generated using PDFs from the NNPDF 3.1 set. Each one of the 100 replicas is taken as an individual one-sided systematic uncertainty. For all samples, distributions are normalized to correct for the change in the cross-section in the PDF replicas.
- **b − tagging** : The scale factors, from which the efficiency corrections of the DeepJet b tagging algorithm are derived, are varied within their uncertainties provided by the b-Tag and Vertexing (BTV) POG [23]. These up- and down-shifted corrections are applied to simulated events instead of the nominal correction. These corrections are further split up into a light-flavor and a heavy-flavor component as recommended by the BTV POG. In total, we have 8 nuisance parameters for the fixedWP method due to the year-to-year correlation mode: LFcorr, HFcorr, LFuncorr16, LFuncorr17, LFuncorr18, HFuncorr16, HFuncorr17, HFuncorr18.
- μ_R **and** μ_F The uncertainty in the renormalization and factorization scales μ_R and μ_F is calculated for all simulation processes. First, each scale is varied by a factor of 2 and 0.5, resulting in 8 combinations:
  - μ_R up and μ_F up
  - μ_R up and μ_F nominal
  - μ_R up and μ_F down
  - μ_R down and μ_F up
  - μ_R down and μ_F nominal
  - μ_R down and μ_F down

   - μ_R nominal and μ_F up
   - μ_R nominal and μ_F down

  The two configurations with opposite variations (up/down and down/up) are unphysical and thus excluded. Taking efficiencies and acceptance effects into account, normalization is factored out to obtain shape-only uncertainties. From the remaining 3 couples, the one with the largest uncertainty is taking into account.

- **Jet Energy Scale** : The jet energy scale (JES) uncertainty is obtained by varying the jet energy by $\pm 1\sigma$ propagated to the missing transverse momenta and making the new discriminant shapes by repeating the analysis. The jet energy correction from all the sources are taken individually, to study the impact and constraint on each. The full list of sources are as follows:
   - TimePtEta
   - RelativeJER[EC1][EC2]
   - RelativePt[EC1][EC2]
   - RelativeSample
   - RelativeStatFSR
   - RelativeStatEC

  Correlations across years are considered as described here [31].

- **Jet Energy Resolution** : The jet energy resolution (JER) uncertainty is calculated by rescaling the jet momenta by $\pm 1\sigma$ in the simulation. The variation in jet energy resolution is propagated to the missing transverse momenta. This uncertainty is uncorrelated across the years.

- **$\mathbf{t-}$tagging SF** : Data/MC corrections for the efficiencies of the t-tagging selection are included in the analysis. The systematic error is introduced in the statistical analysis as an unconstrained nuisance parameter (flat prior); this implies that no assumptions are made on the uncertainty of this data/MC SF. This is done because of the overlap between the dataset in which the SFs are calculated and the SR of the analysis.

- **Statistical uncertainties** : The statistical fluctuations are considered bin-by-bin using the Barlow-Beeston-lite approach [32], implemented in autoMCstats in the HiggsCombine framework.

- **Top p_T reweighing** $t\bar{t}$ MC samples are corrected by the 13 TeV data/powheg+pythia reweighing function $e^{0.0615-0.0005*p_T}$, where p_T is the transverse momentum of the top at generator level. A variation given by the difference between the top p_T distribution with and without correction is applied as systematic uncertainty.

- **ISR and FSR parton shower** : ISR and FSR parton shower uncertainties are estimated by varying the shower scale by a factor of two. The uncertainties are fully correlated across years.

- **ME − PS matching scale (hdamp)** h_{damp} is the model parameter that controls the ME/PS matching and regulates the high-p_T radiation by damping real emissions generated by POWHEG. To estimate the uncertainty of the matching between the NLO calculation and the parton shower a machine learning-based approach is implemented to compute the up/down variation weights [**?**]. Following recommendations, an additional 2% flat uncertainty is incorporated into the final result. This additional factor aims to accommodate any observed discrepancies resulting from this new method, as highlighted during a closure test.

Acknowledgments

A b-tagging efficiency in MC samples

In this analysis, b-tagging is employed to reduce W+jets events, necessitating the derivation of a scale factor to correct the MC samples, as elaborated in Section 3. The BTV POG supplies scale factors [23], yet the efficiencies (outlined in Equation 4) must be evaluated within the analysis-specific phase space.

To accomplish this, we measure these efficiencies as functions of p_T and η by amalgamating all the MC samples (comprising both signal and backgrounds). We apply the baseline selection, detailed in Section 4, prior to imposing the b-tagging requirement. Figures 83 and 73 illustrate the resulting measured efficiencies for heavy flavor and light flavor, respectively.

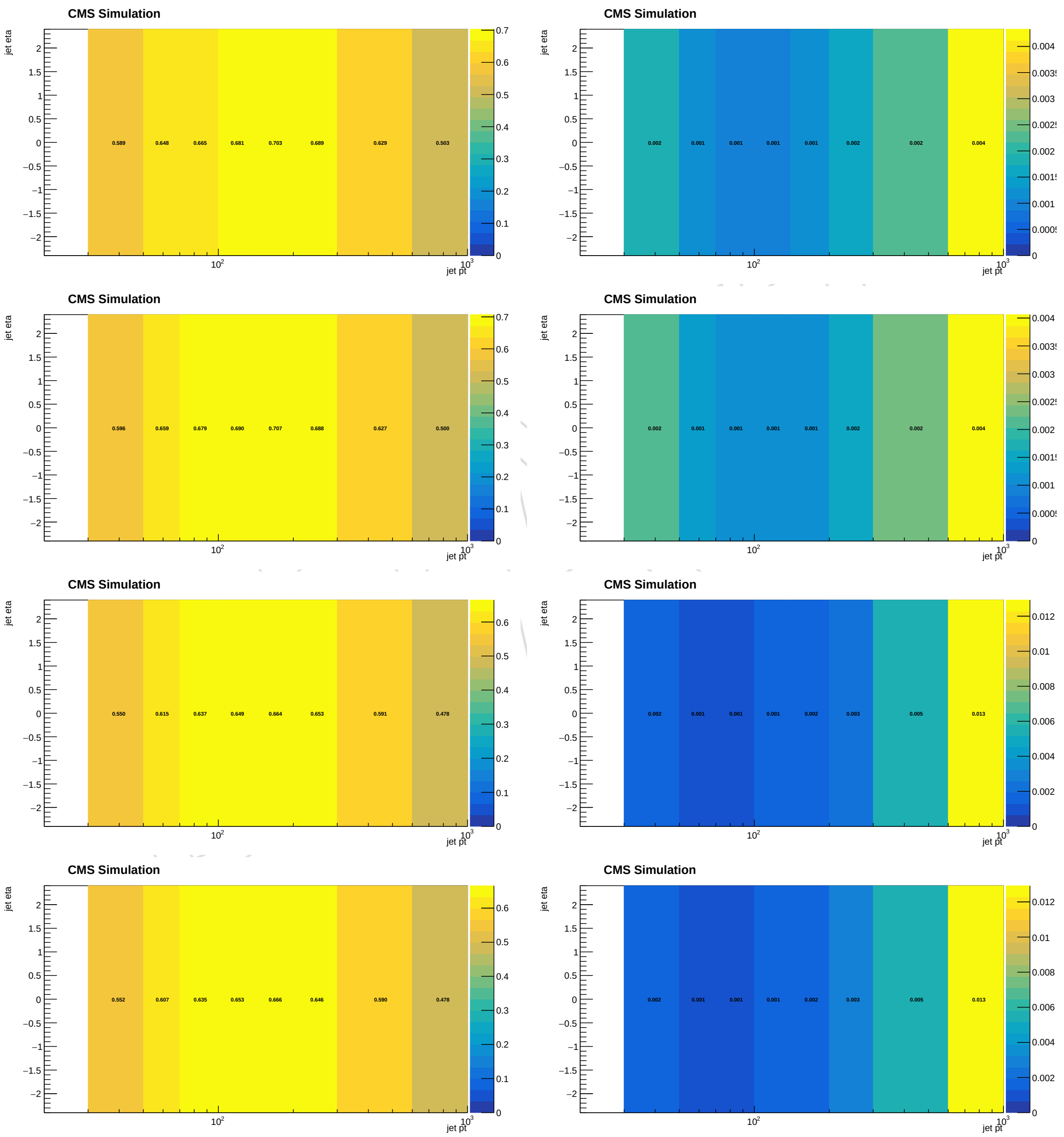

Figure 9: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

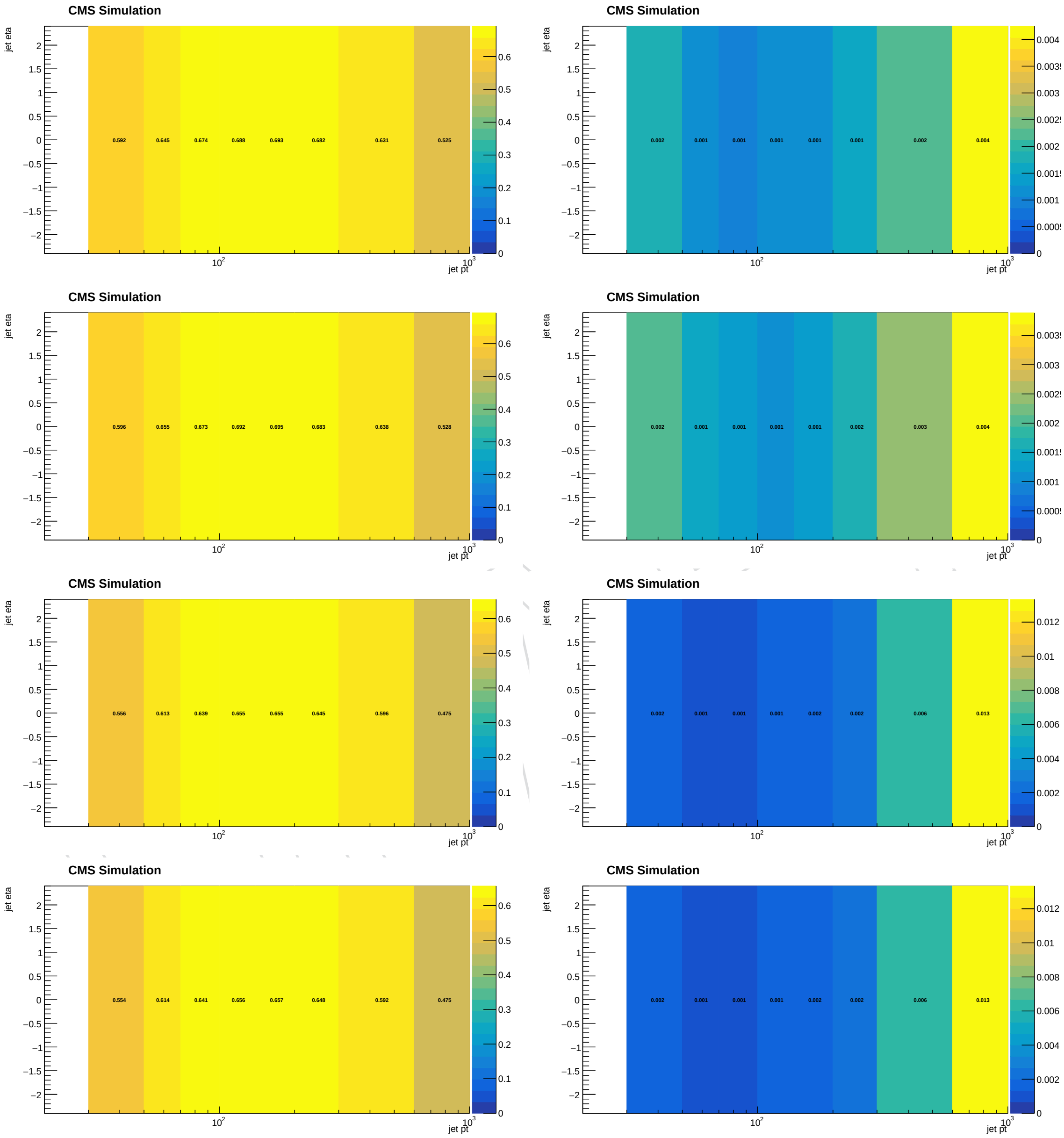

Figure 10: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $e+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

B hdamp

Due to the limited MC statistic in the varied samples for hdamp and tune, the corresponding MC uncertainty on the derived templates is higher compared to the nominal templates and cannot be disregarded. Instead, we are using an state-of-the-art tool to derive the hdamp variation using a Machine-learning approach for parameter reweighting in MC samples of top quark production in CMS [33].

The onnx model receives for each event as inputs an array of the following dimensions (1,2,6) containing the following parameters:

- (p_T(top), y(top), ϕ(top), m(top), PDGID(top), h_{damp})
- (p_T(antitop), y(antitop), ϕ(antitop), m(antitop), PDGID(antitop), h_{damp})

Before passing the inputs to the model for computing the weights, they should be normalized as follows:

- $p_T \Rightarrow log10(p_T)$
- y and ϕ do not require normalization
- m_t should be divided by 243.9517. (This value corresponds to the maximum mass value observed in the training+validation samples.)
- For PDGID: 6 ⇒ 0.1, -6 ⇒ 0.2.
- $H_{damp} = 1.379$ (This represents the nominal value of H_{damp} divided by m_t, the desired value of H_{damp} for reweighting to the CMS Up and Down Variation.)

Figs **??** - **??** show the up and down variation for the A_E AND A_I measurements, all plots look fine.

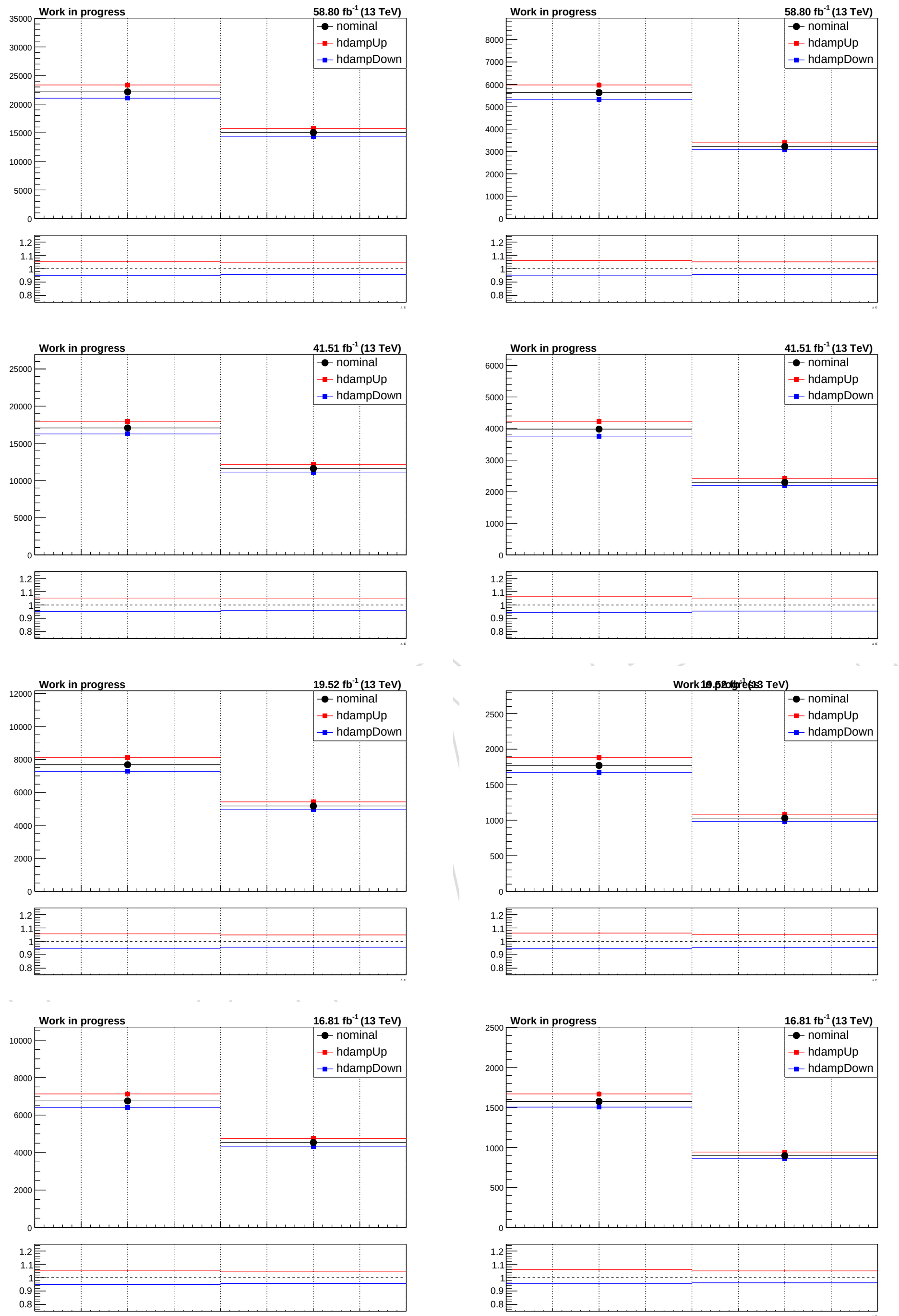

Figure 11: hdamp Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

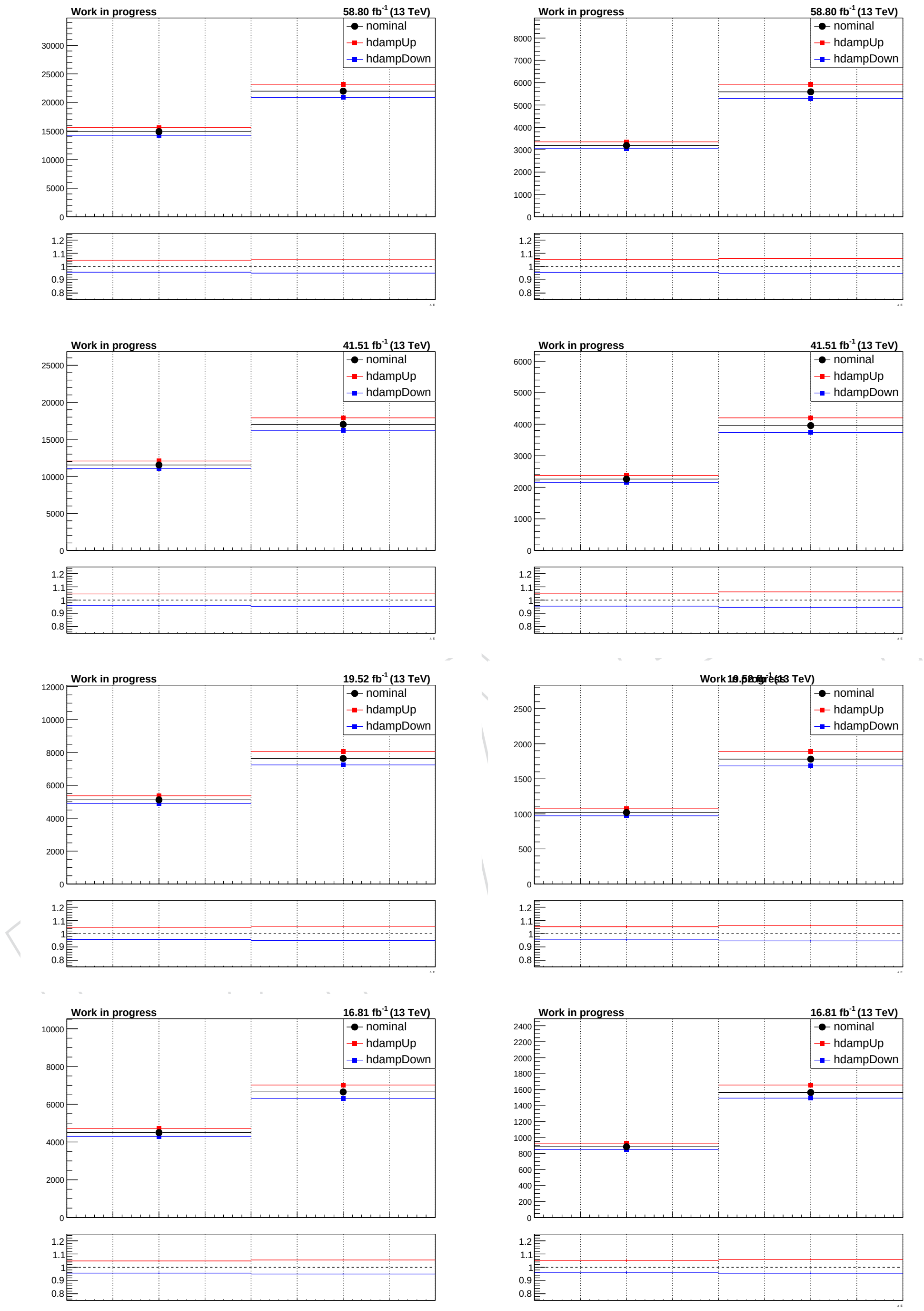

Figure 12: hdamp Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

C Up/Down systematic variations

In this section are shown the up/down variations for the A_E, A_c and A_I variables of the nuisabnces listed in Sec. 6

C.1 A_E

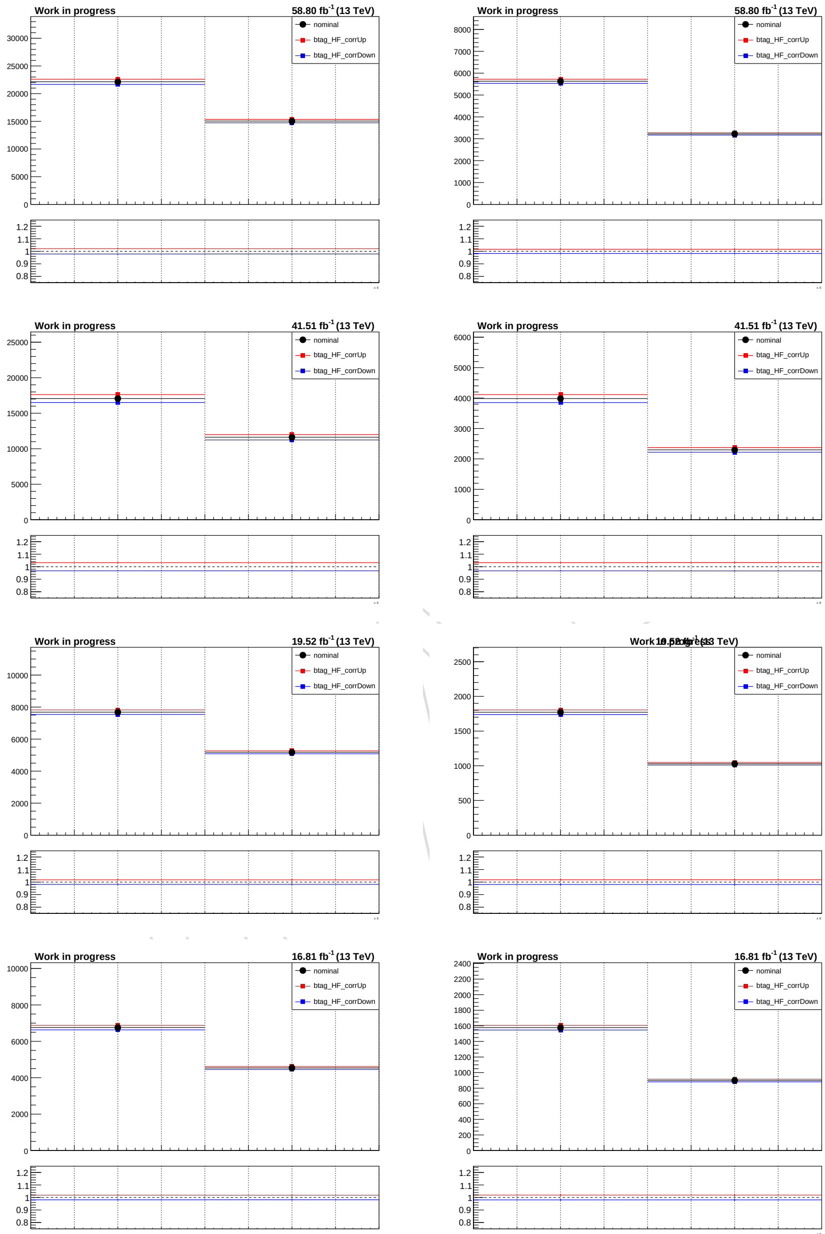

Figure 13: btag_HF Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

C.2 A_I

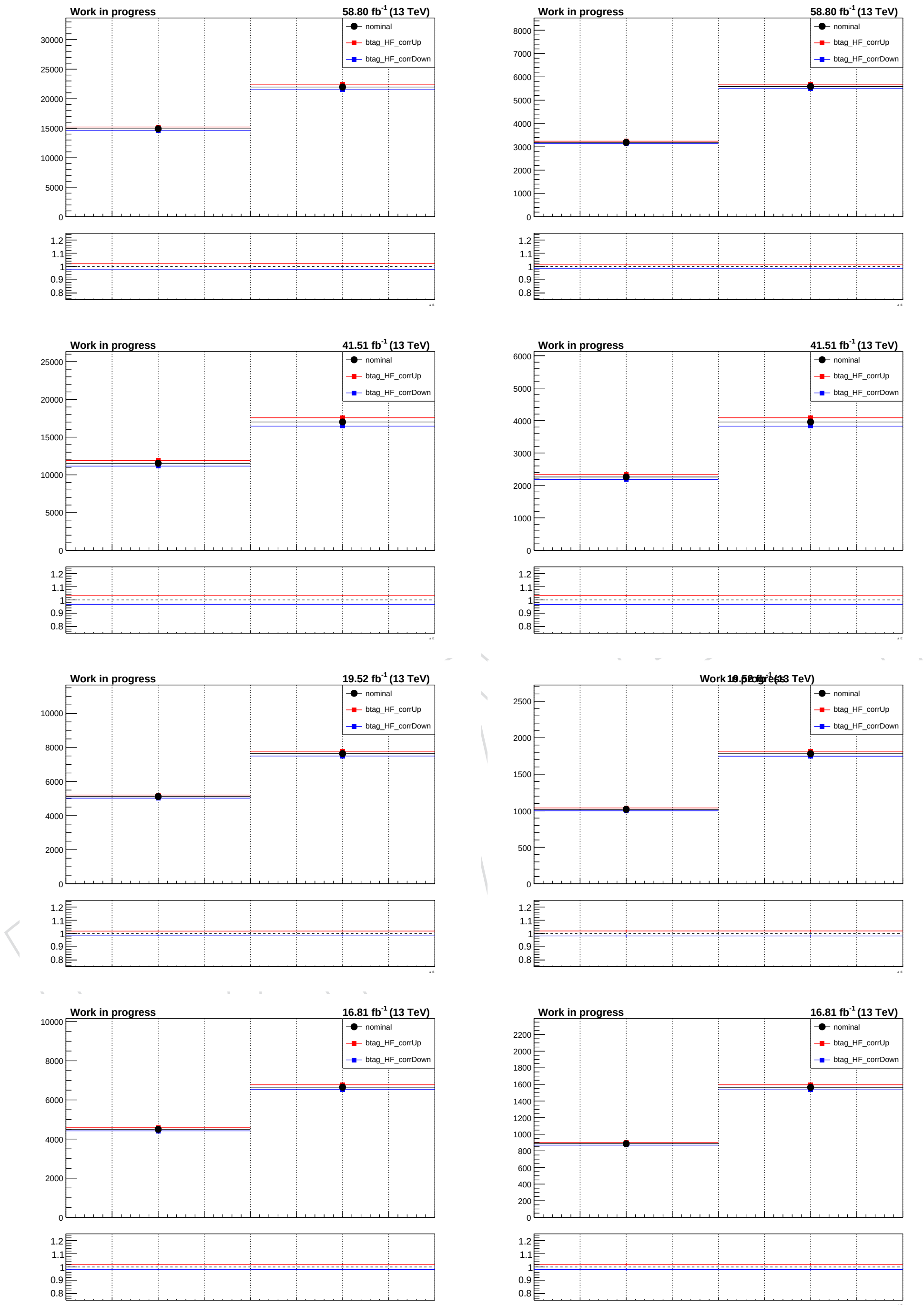

Figure 14: btag_HF Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

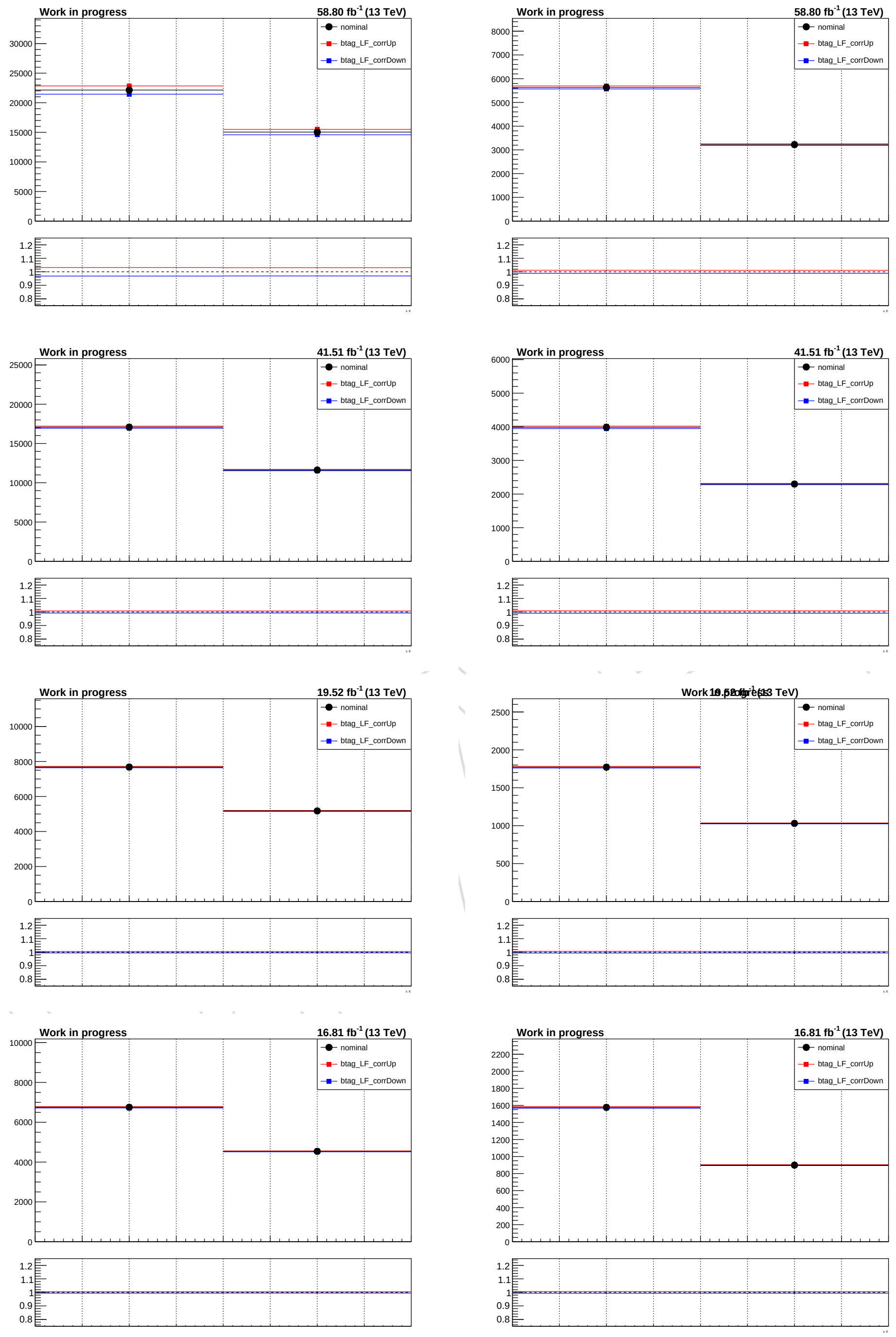

Figure 15: btag_LF Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

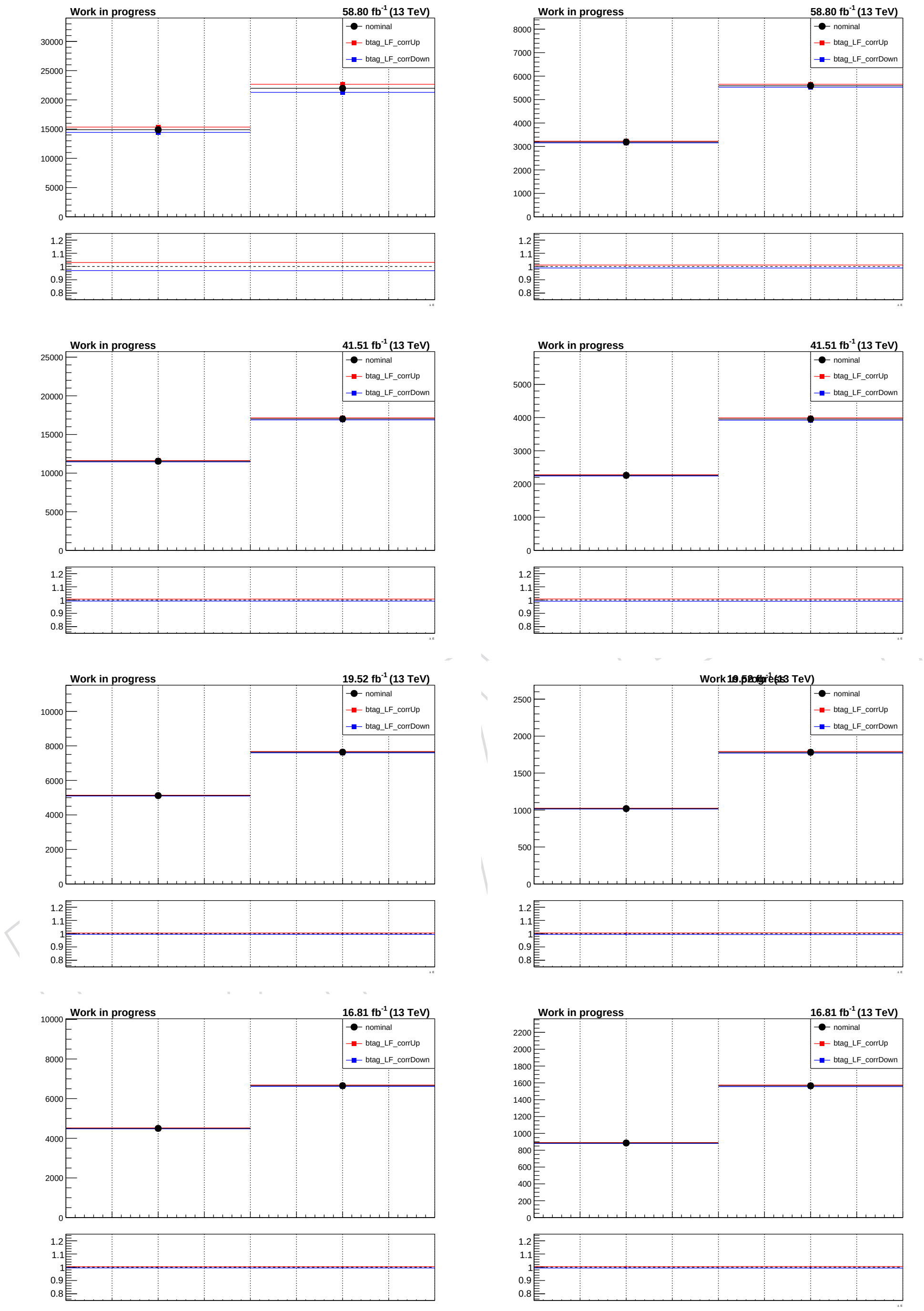

Figure 16: btag_LF Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

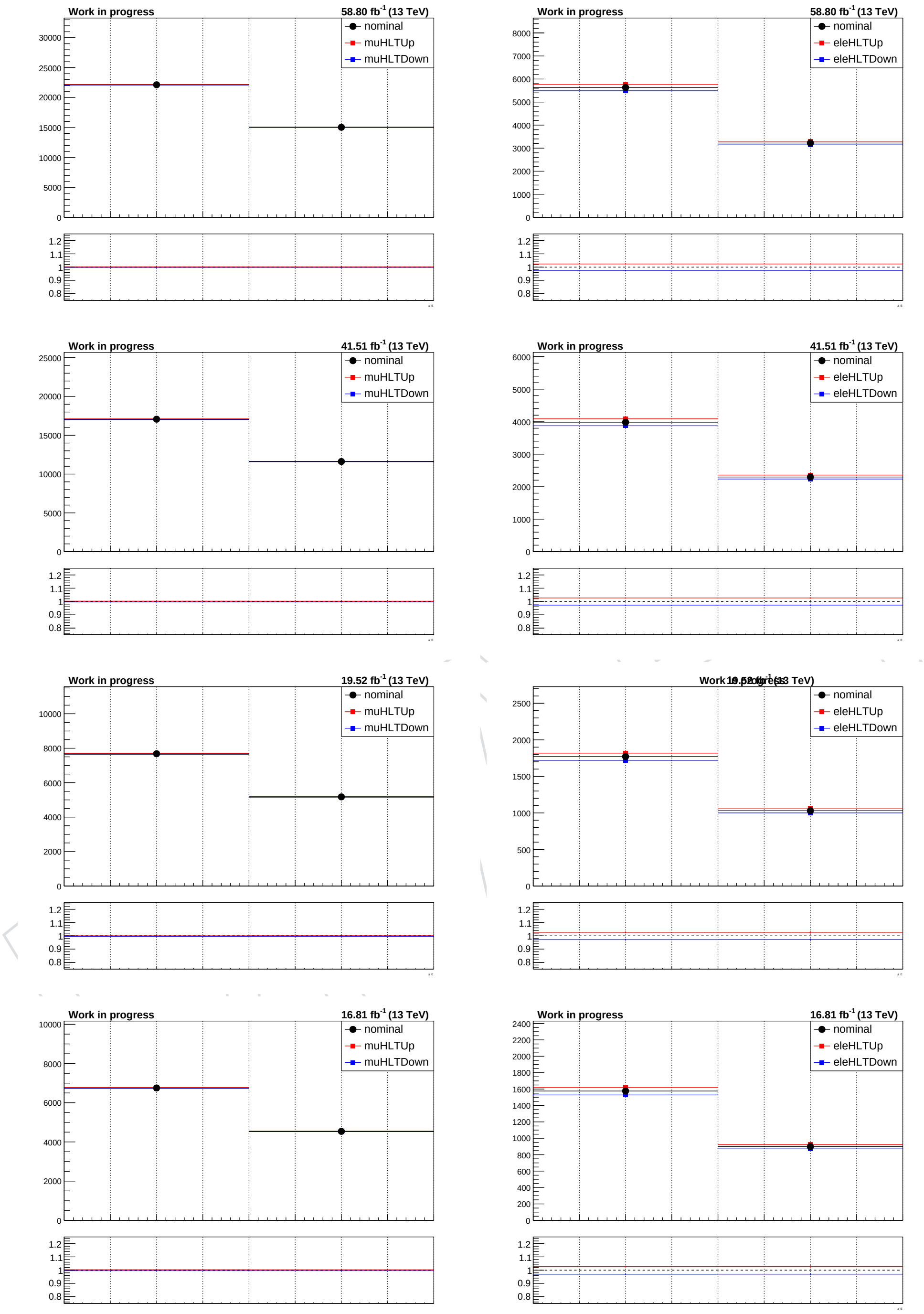

Figure 17: HLT Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

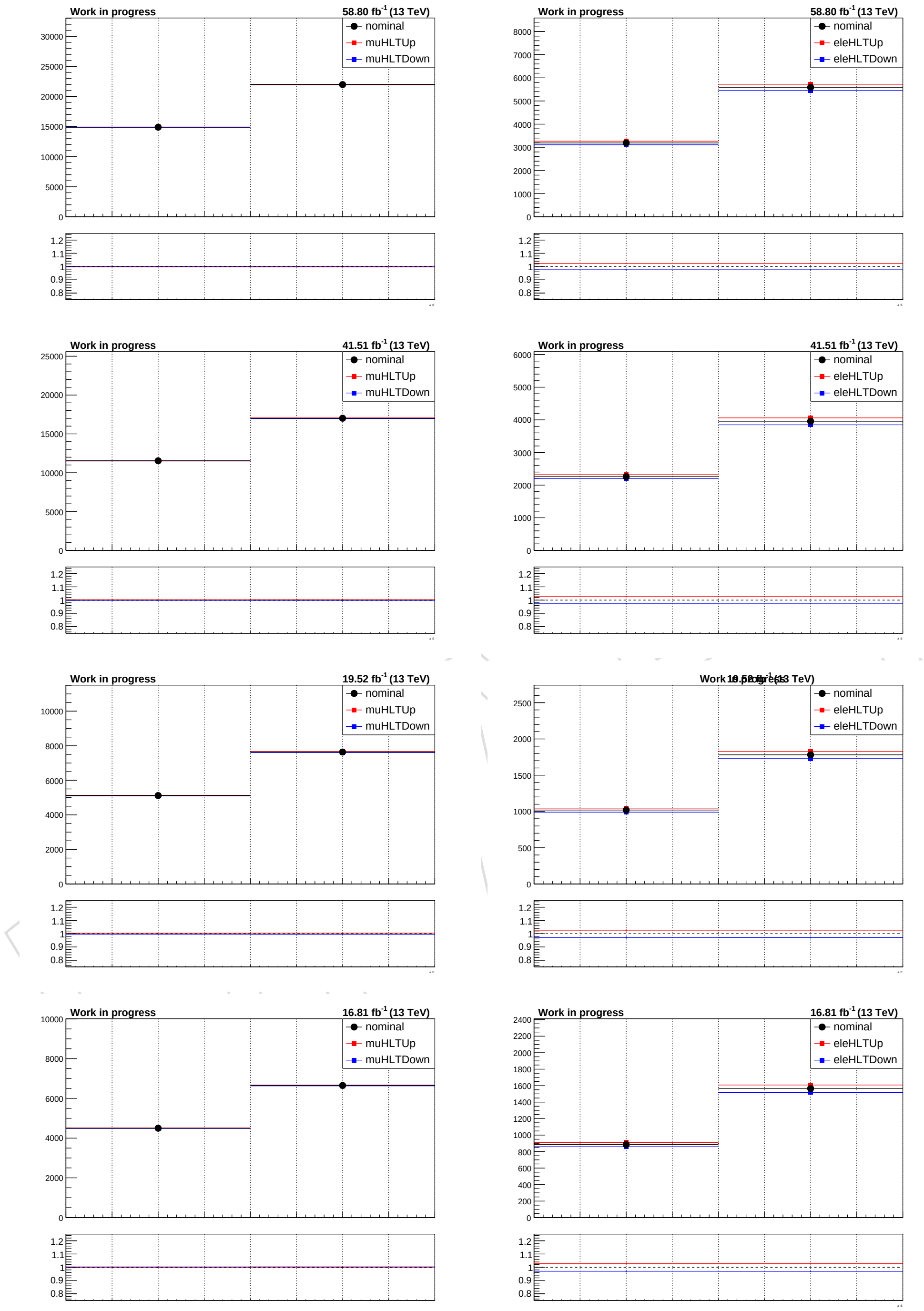

Figure 18: HLT Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

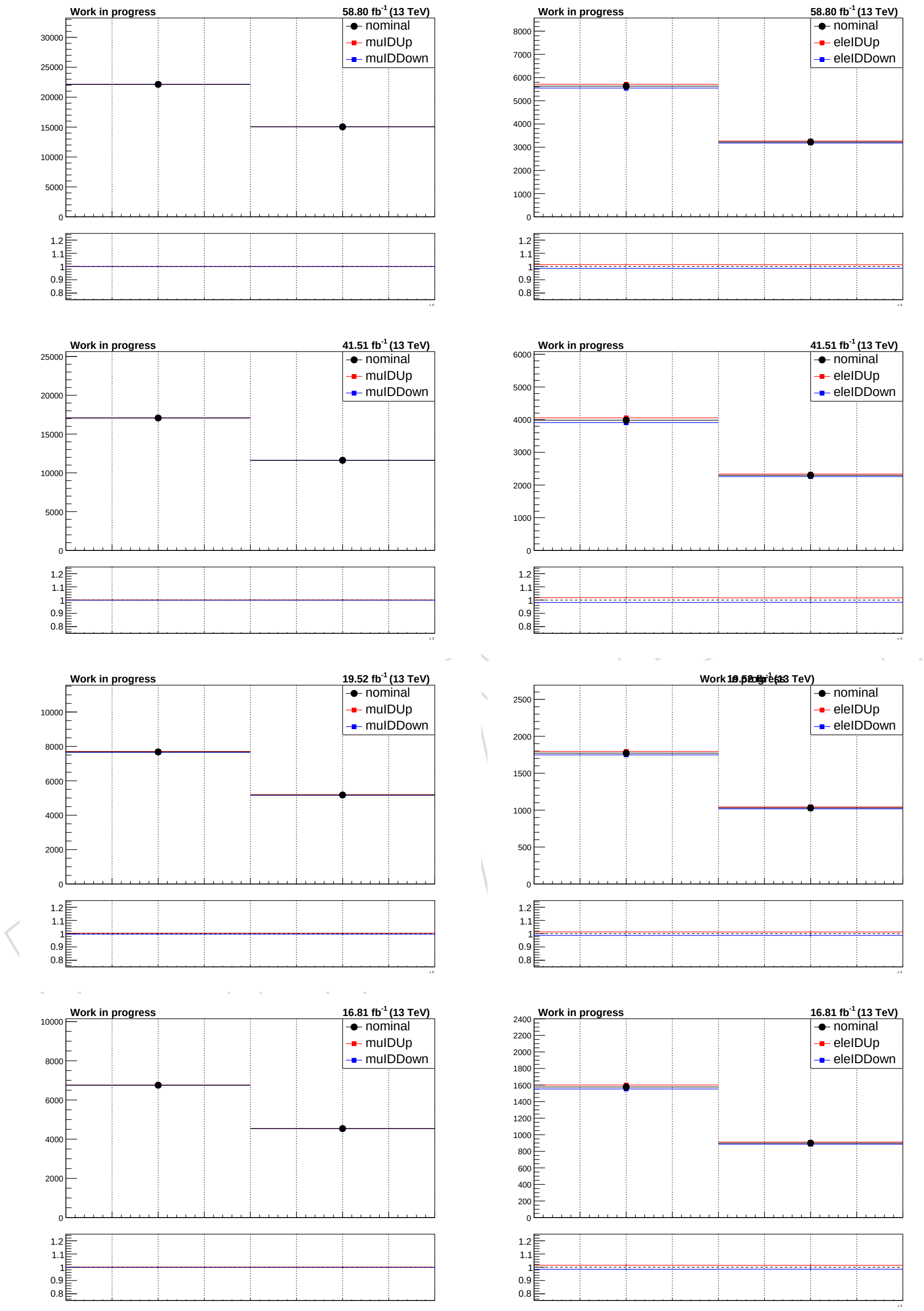

Figure 19: Lepton ID Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

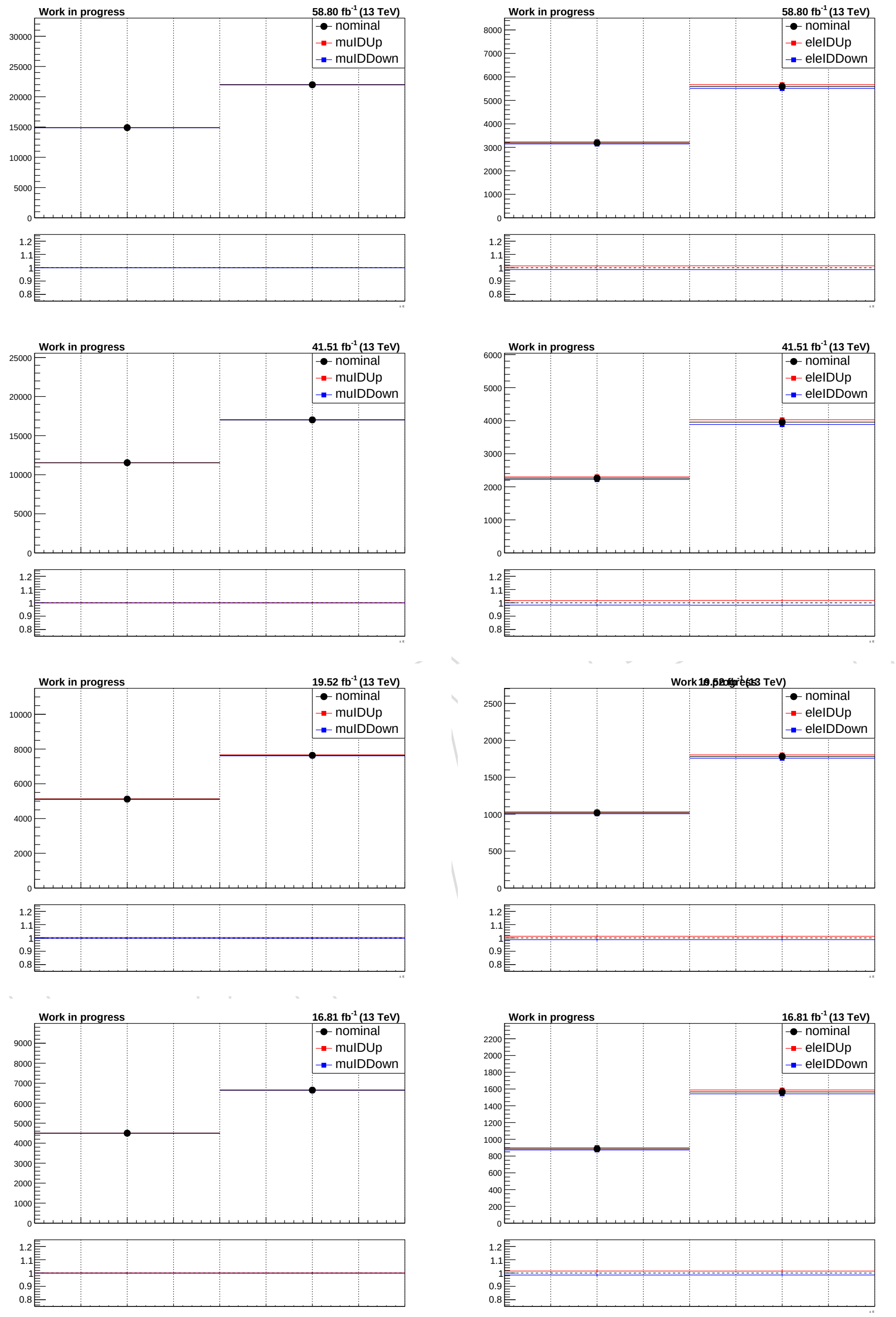

Figure 20: Lepton ID Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

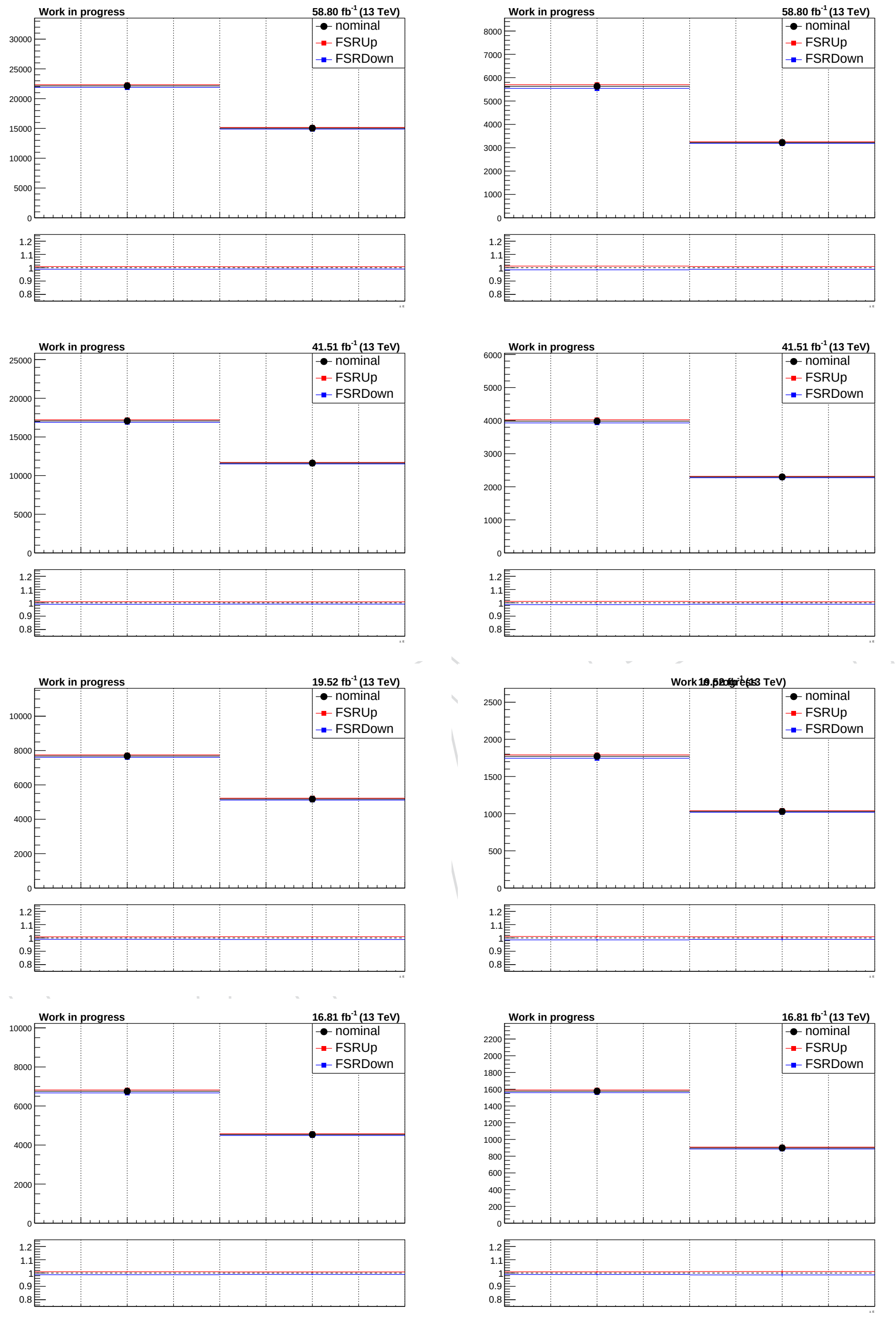

Figure 21: FSR Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

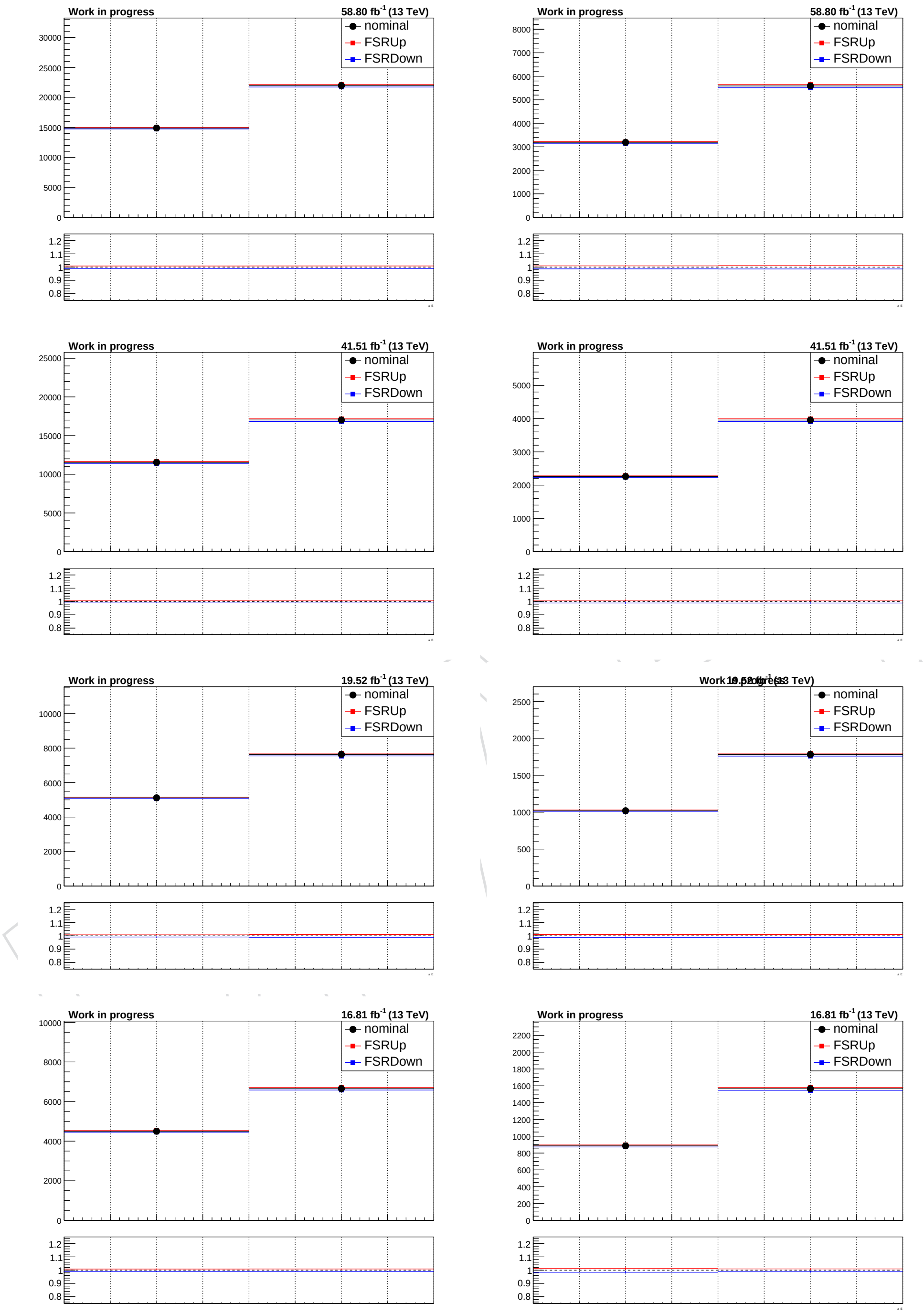

Figure 22: FSR Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

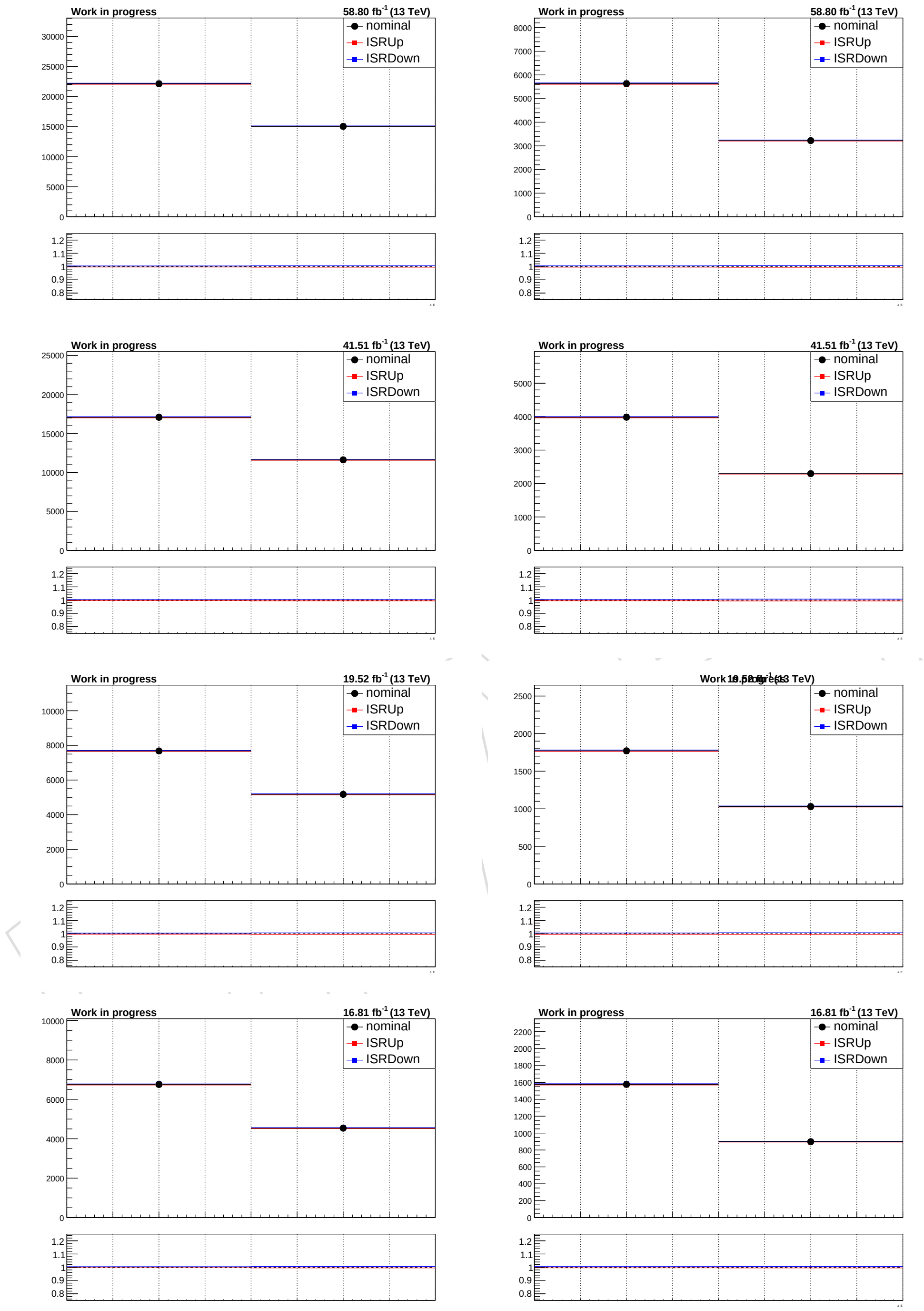

Figure 23: ISR Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

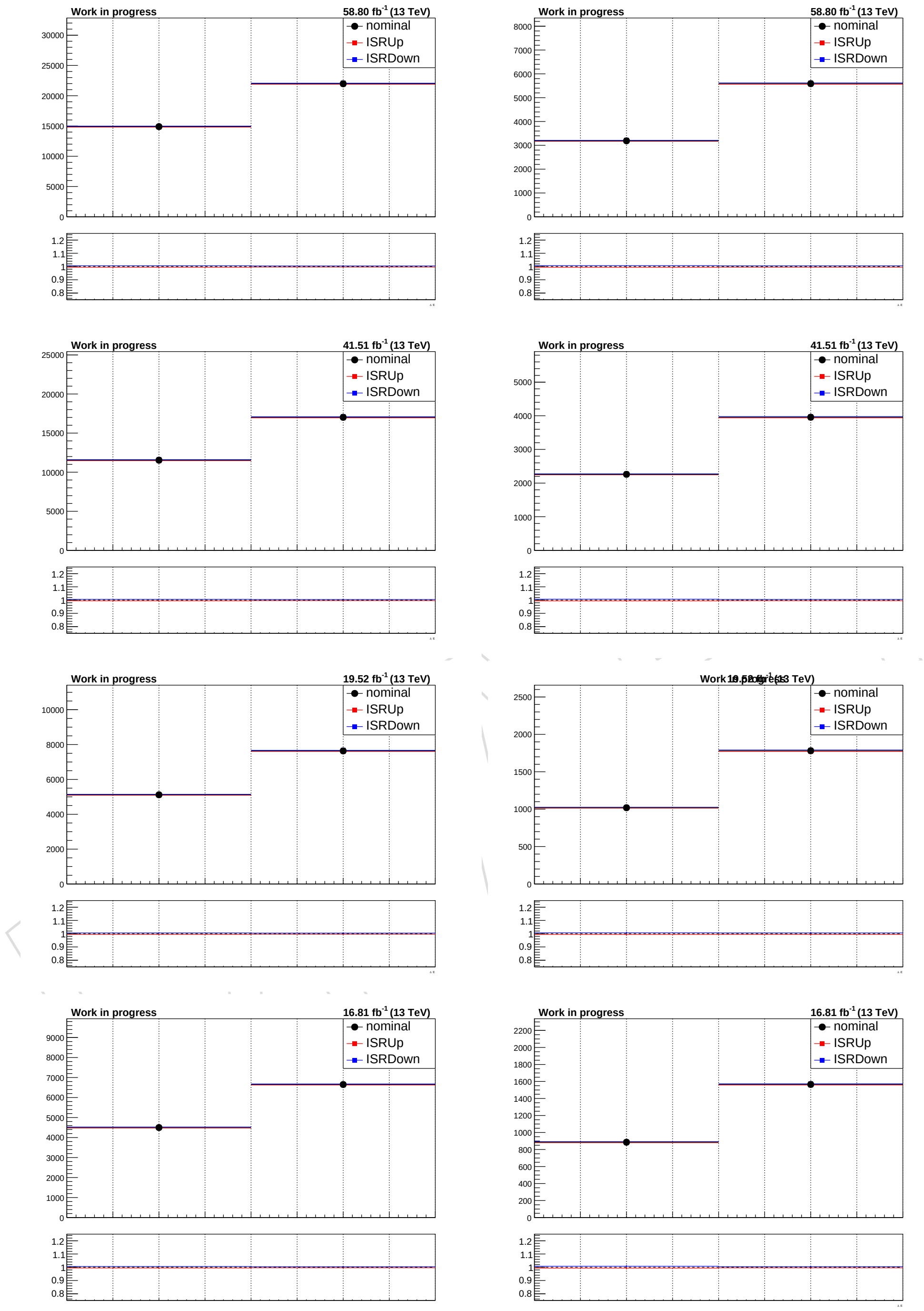

Figure 24: ISR Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

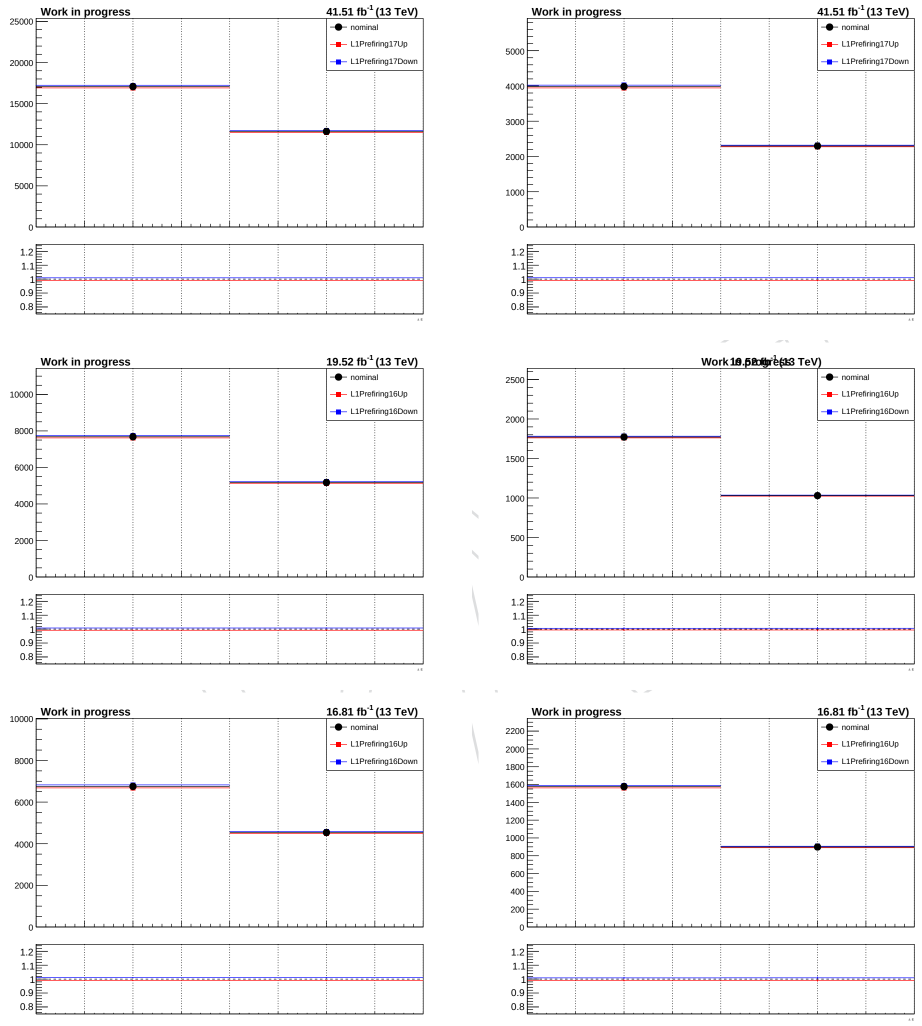

Figure 25: L1Prefiring Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

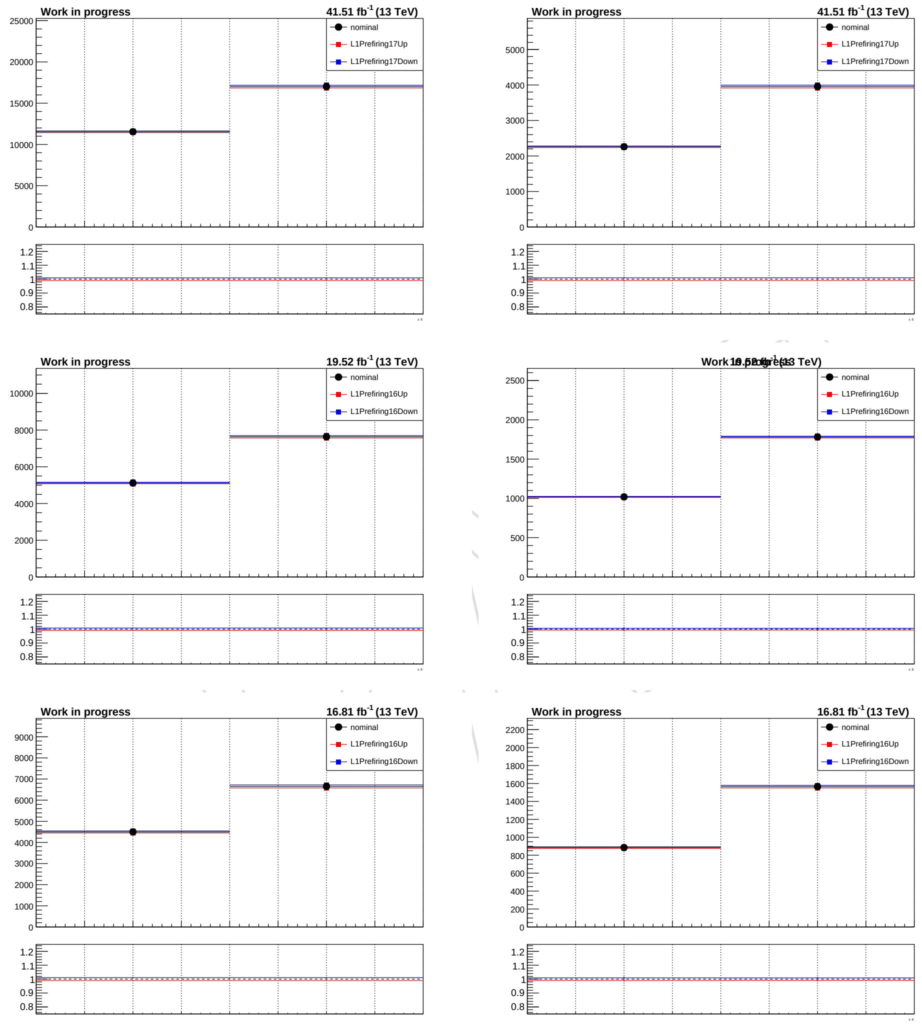

Figure 26: L1Prefiring Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

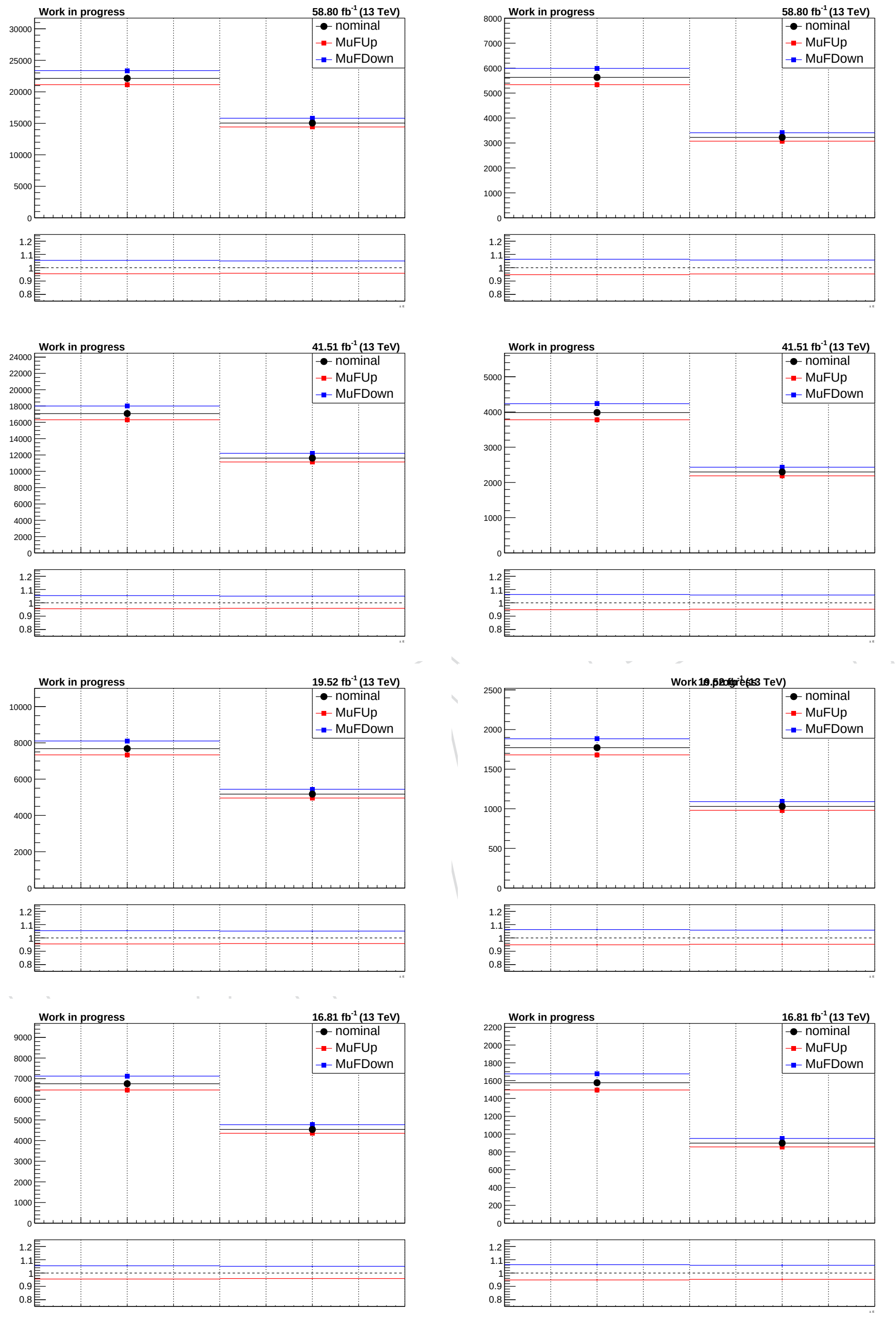

Figure 27: μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

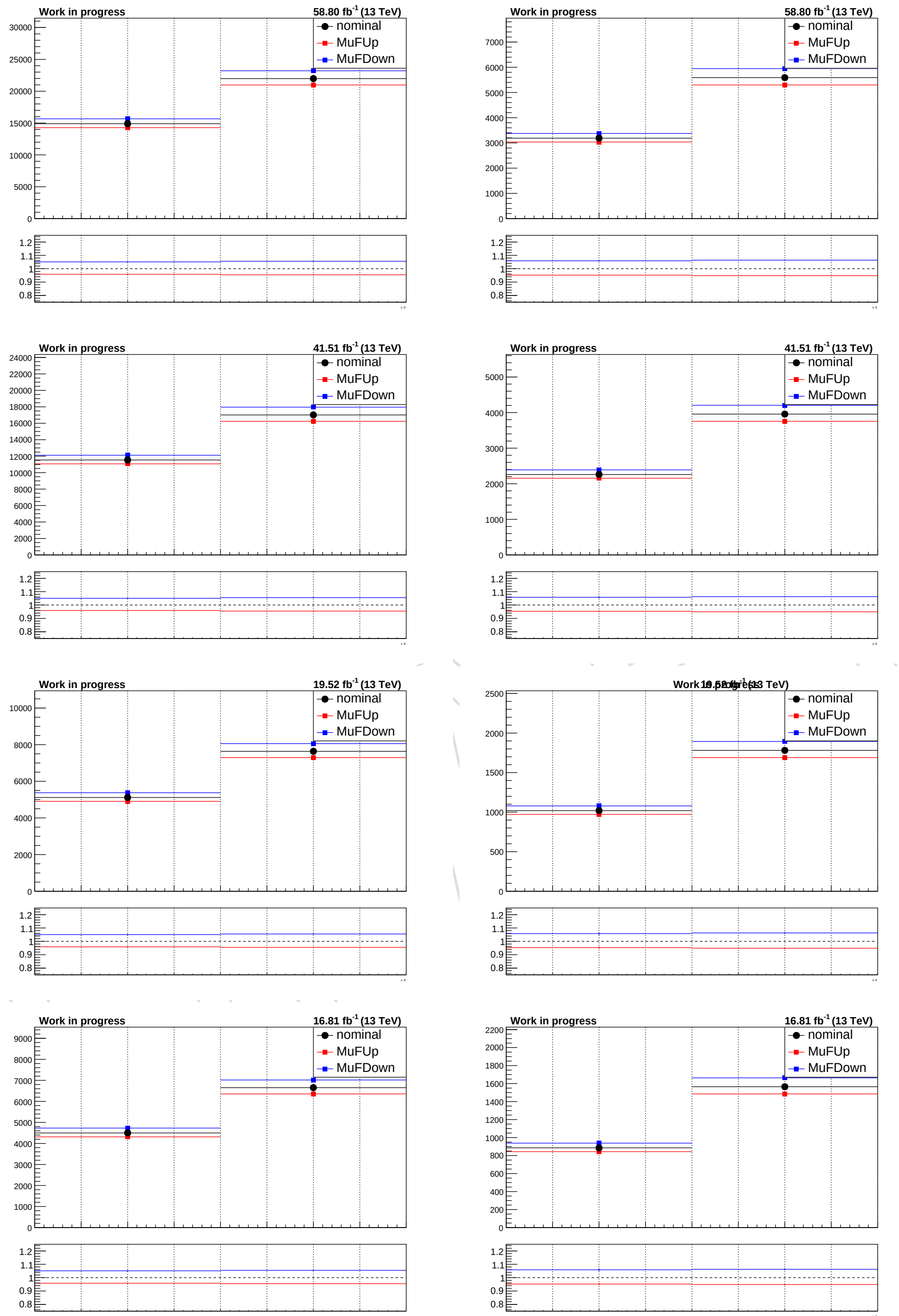

Figure 28: μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

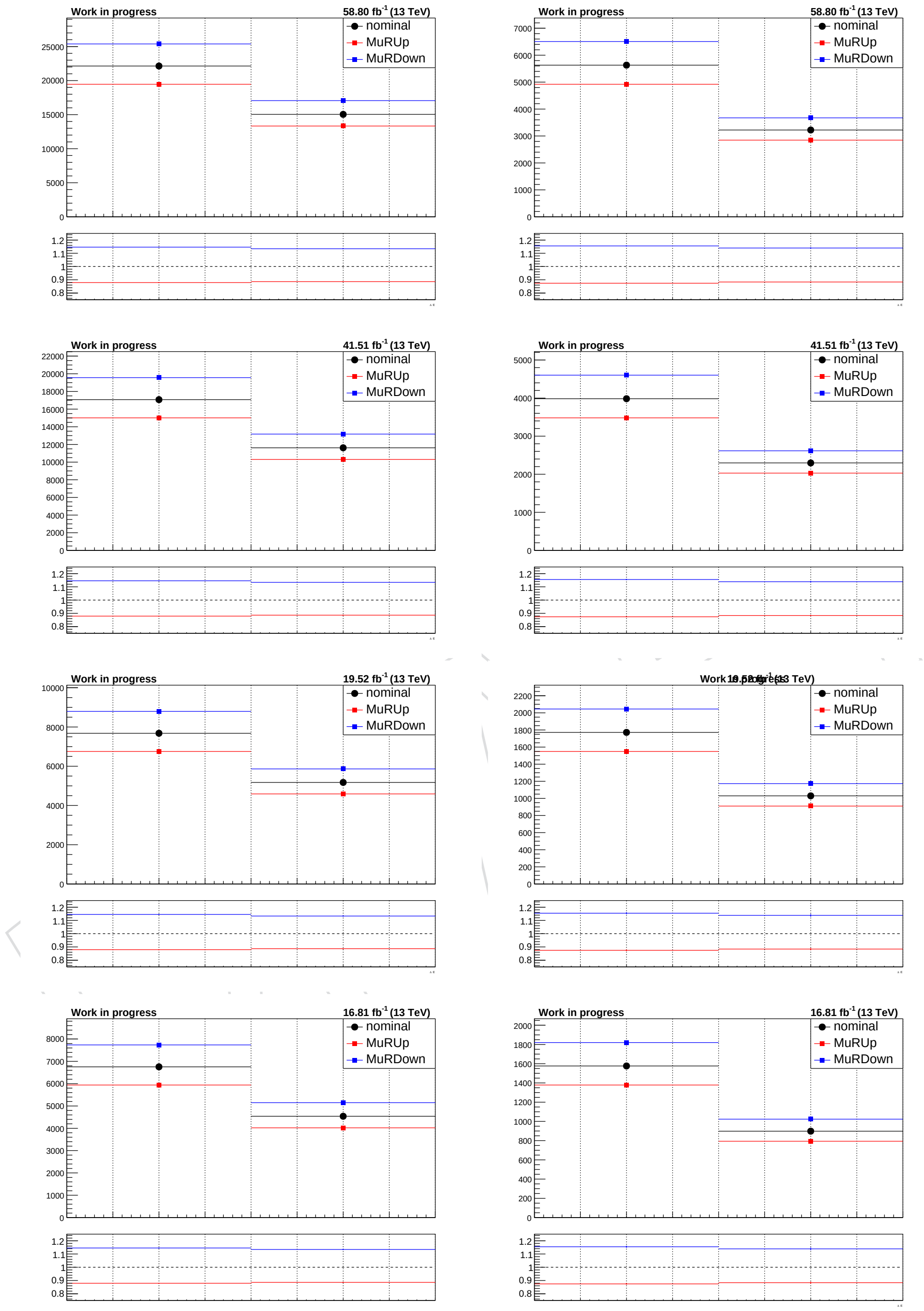

Figure 29: μ_R Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

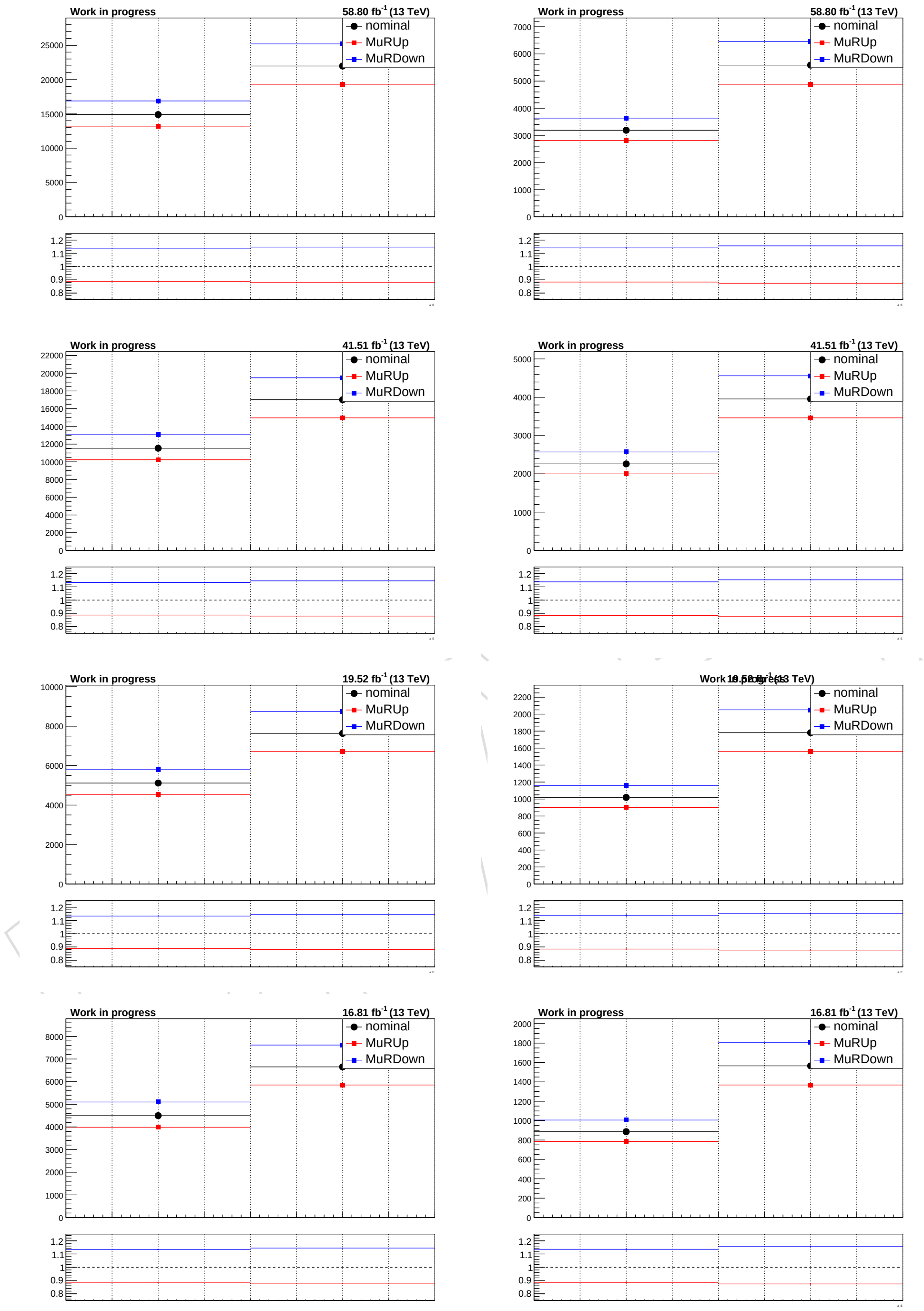

Figure 30: μ_R Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

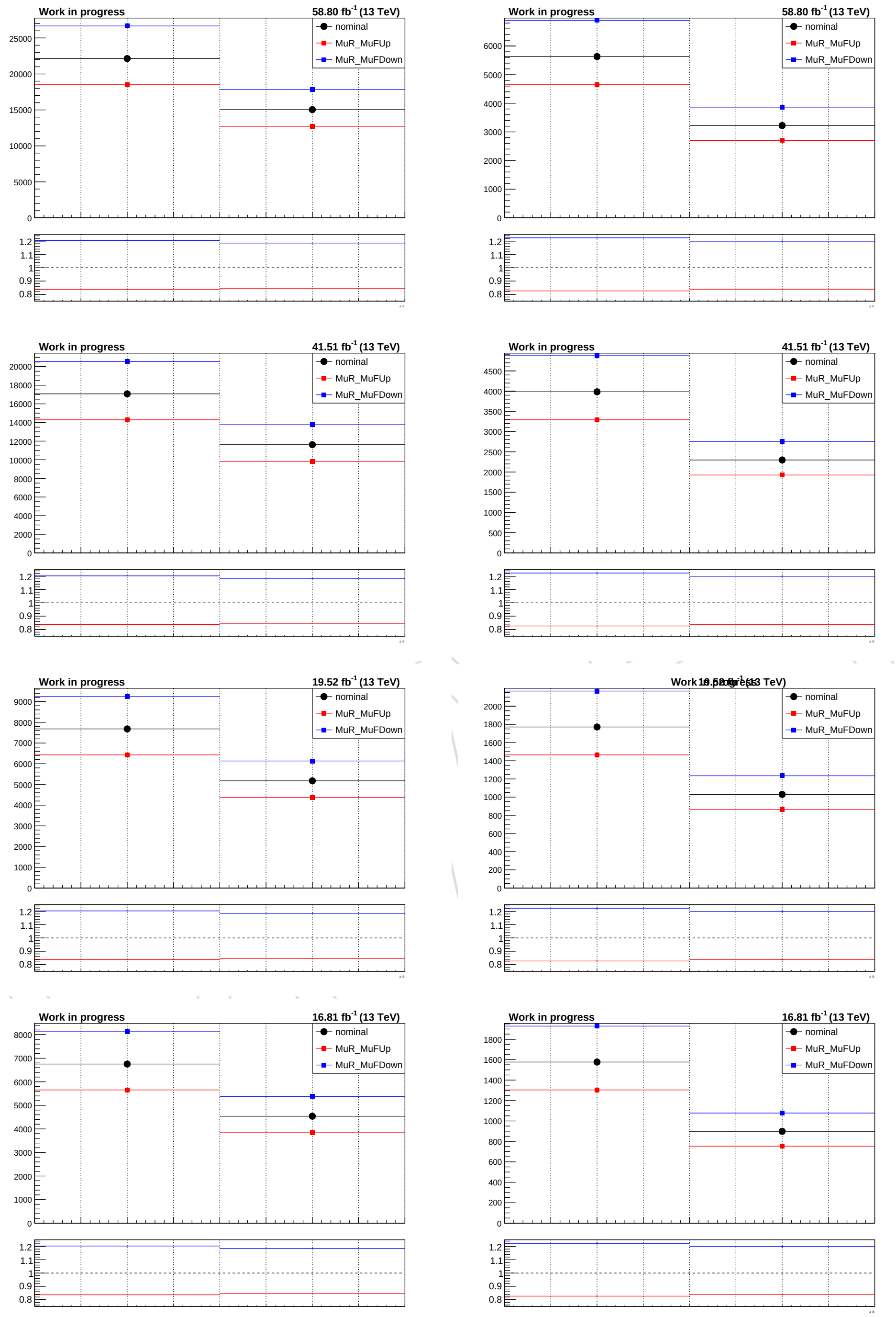

Figure 31: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

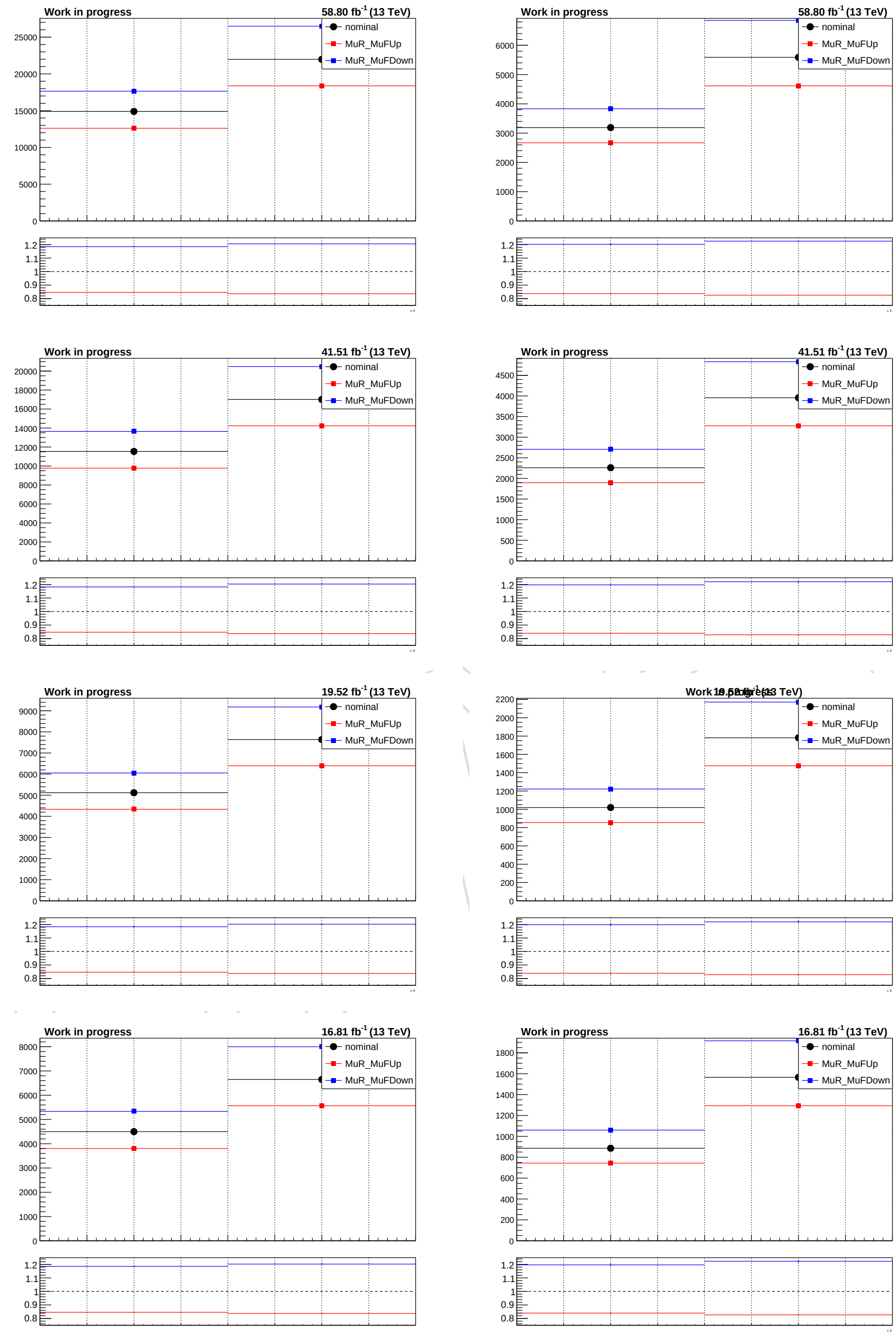

Figure 32: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

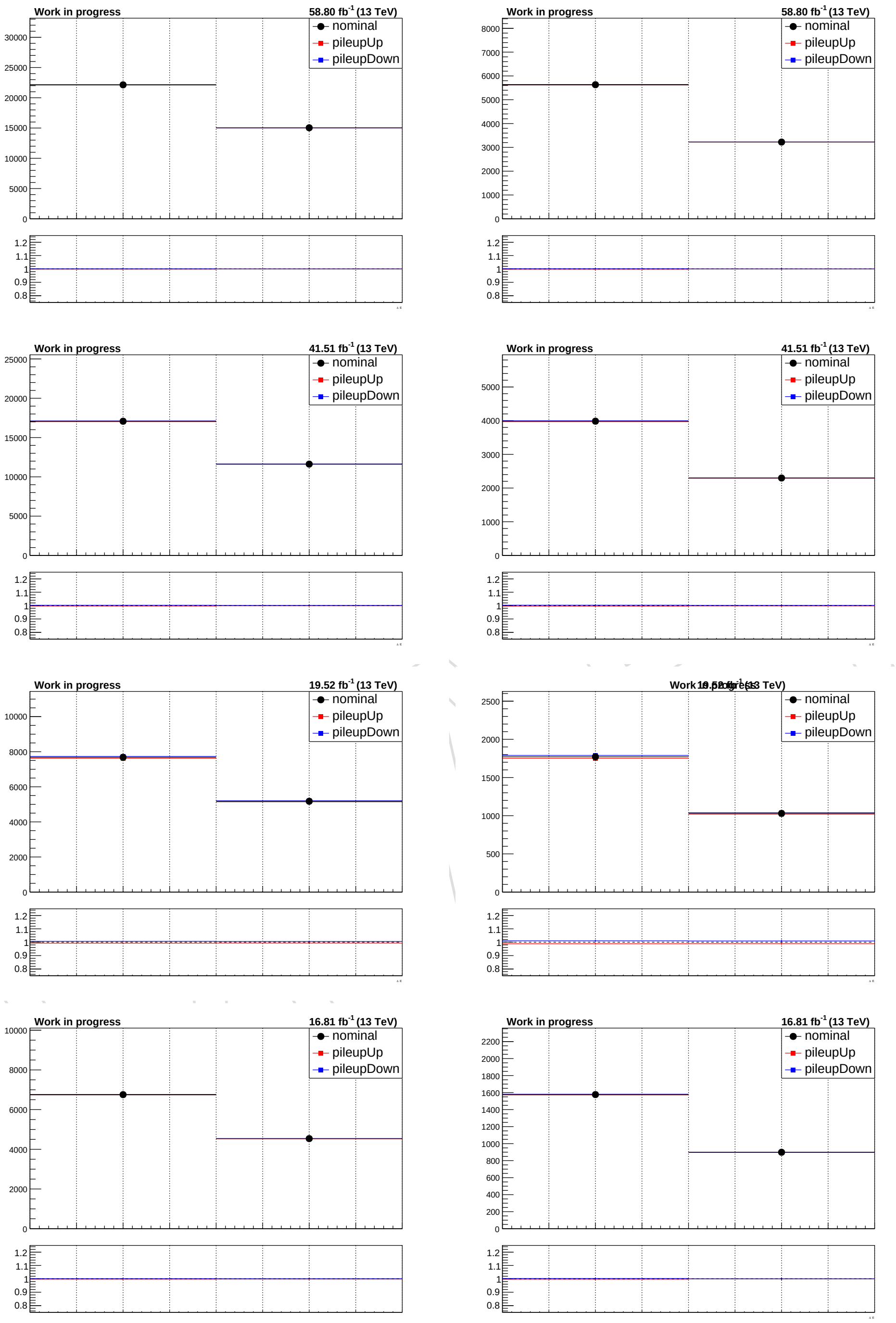

Figure 33: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

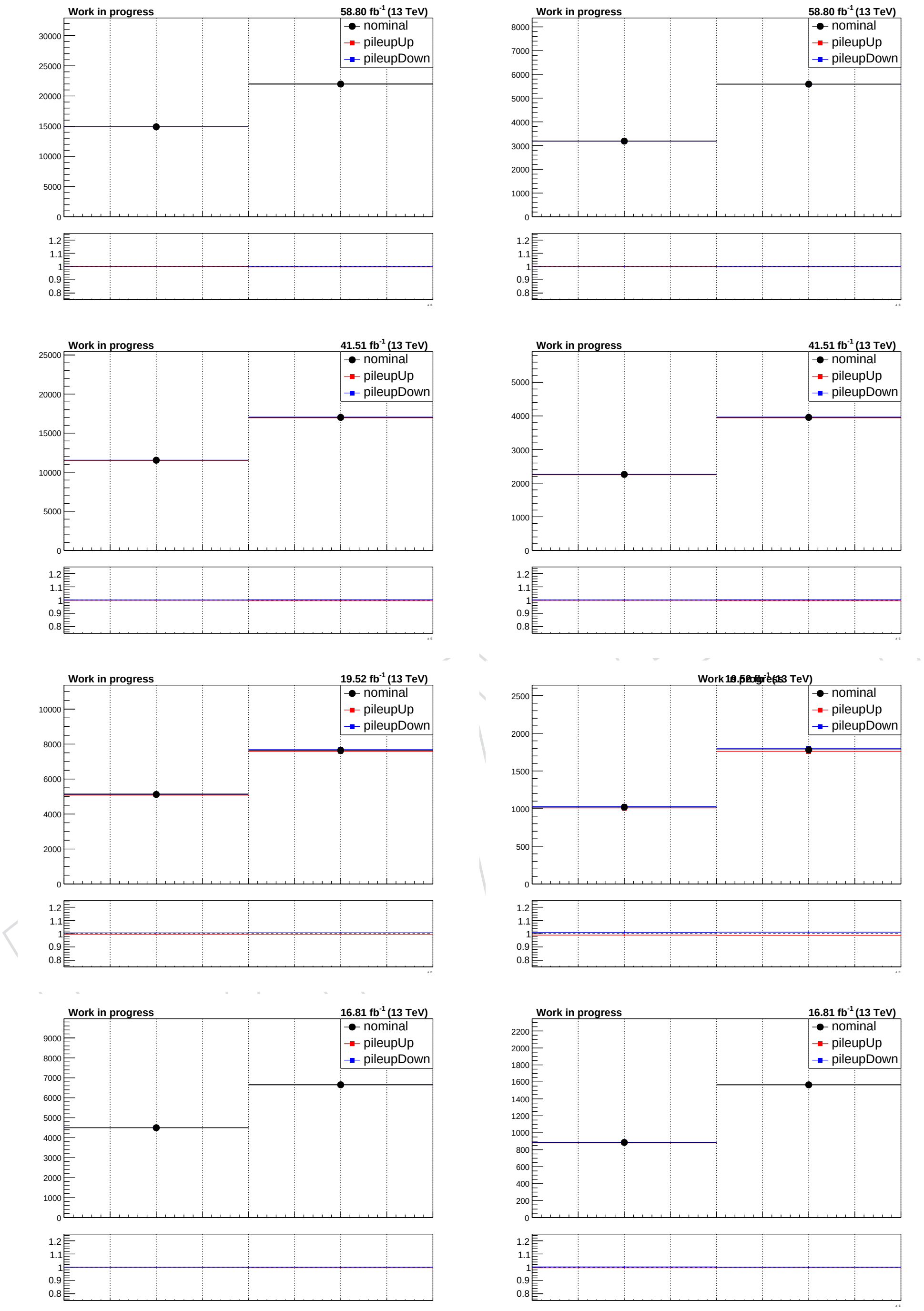

Figure 34: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

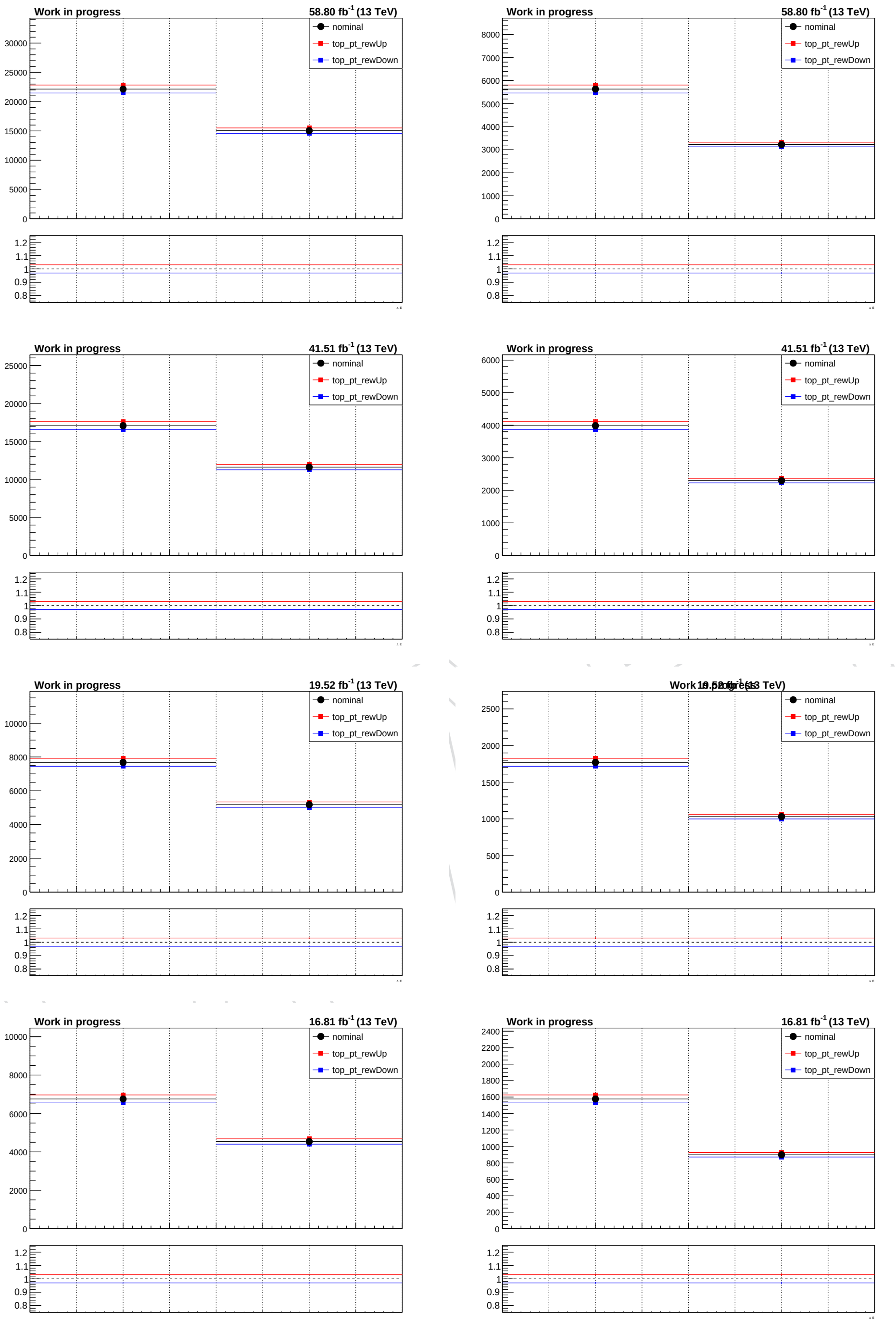

Figure 35: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

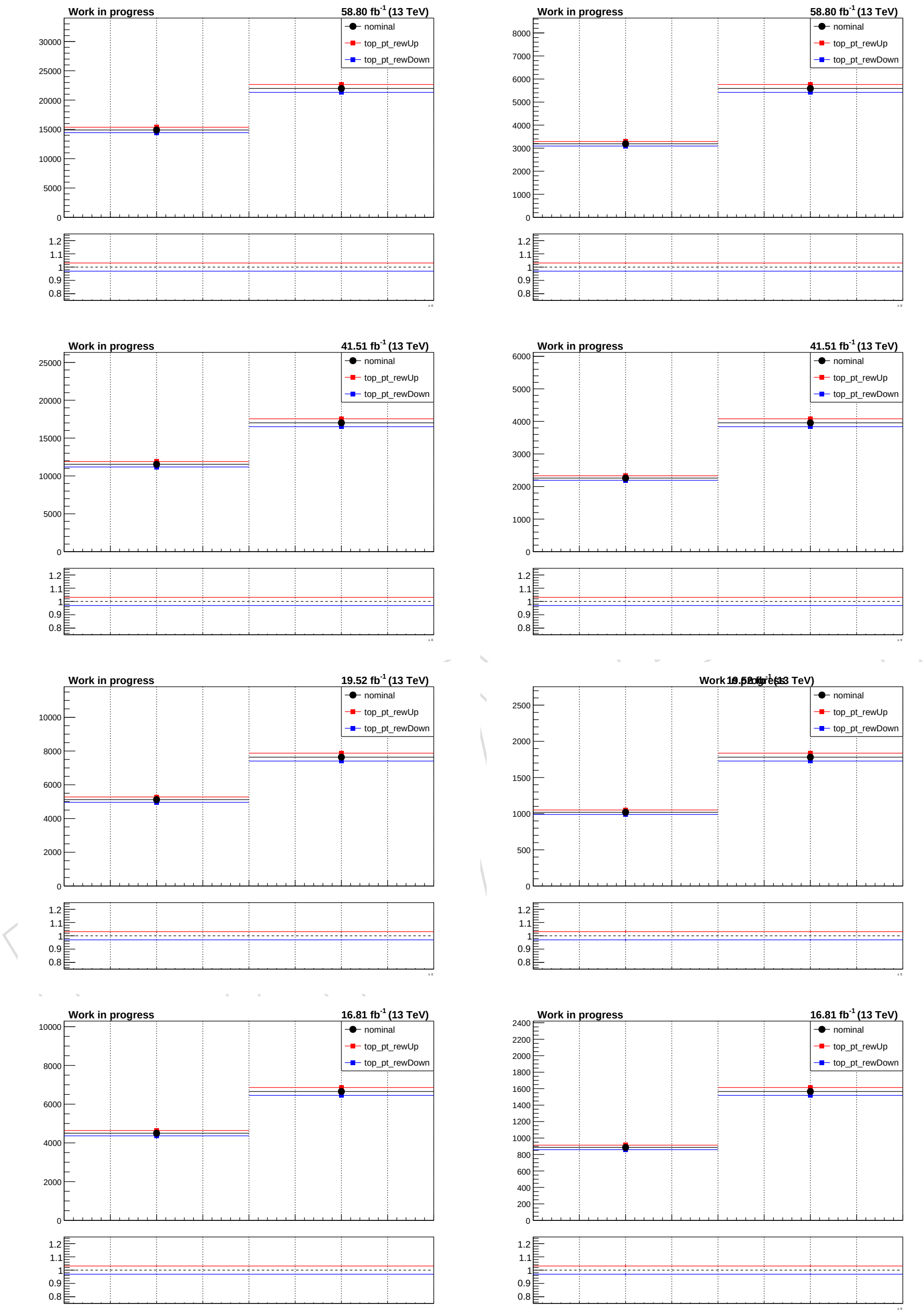

Figure 36: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

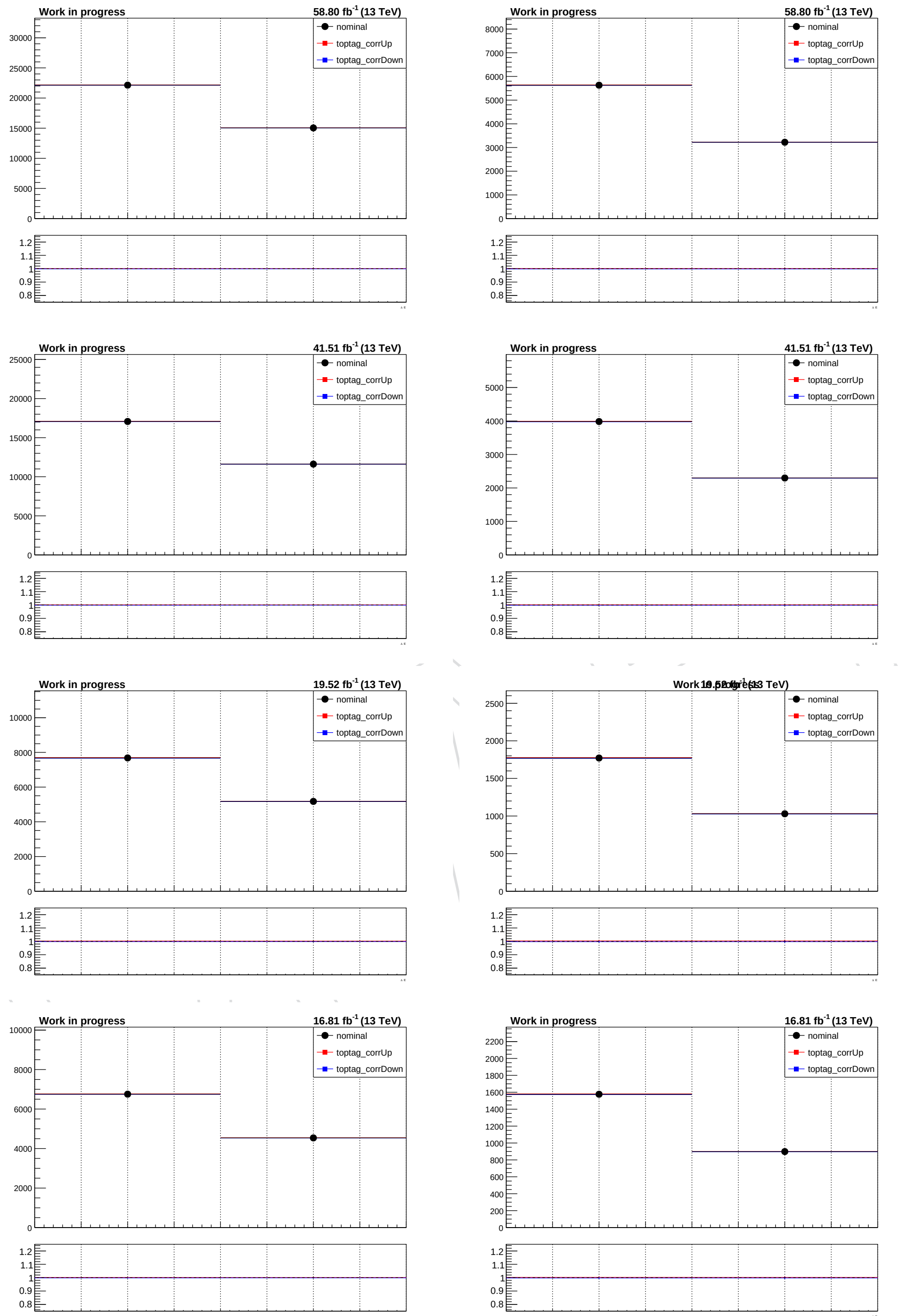

Figure 37: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

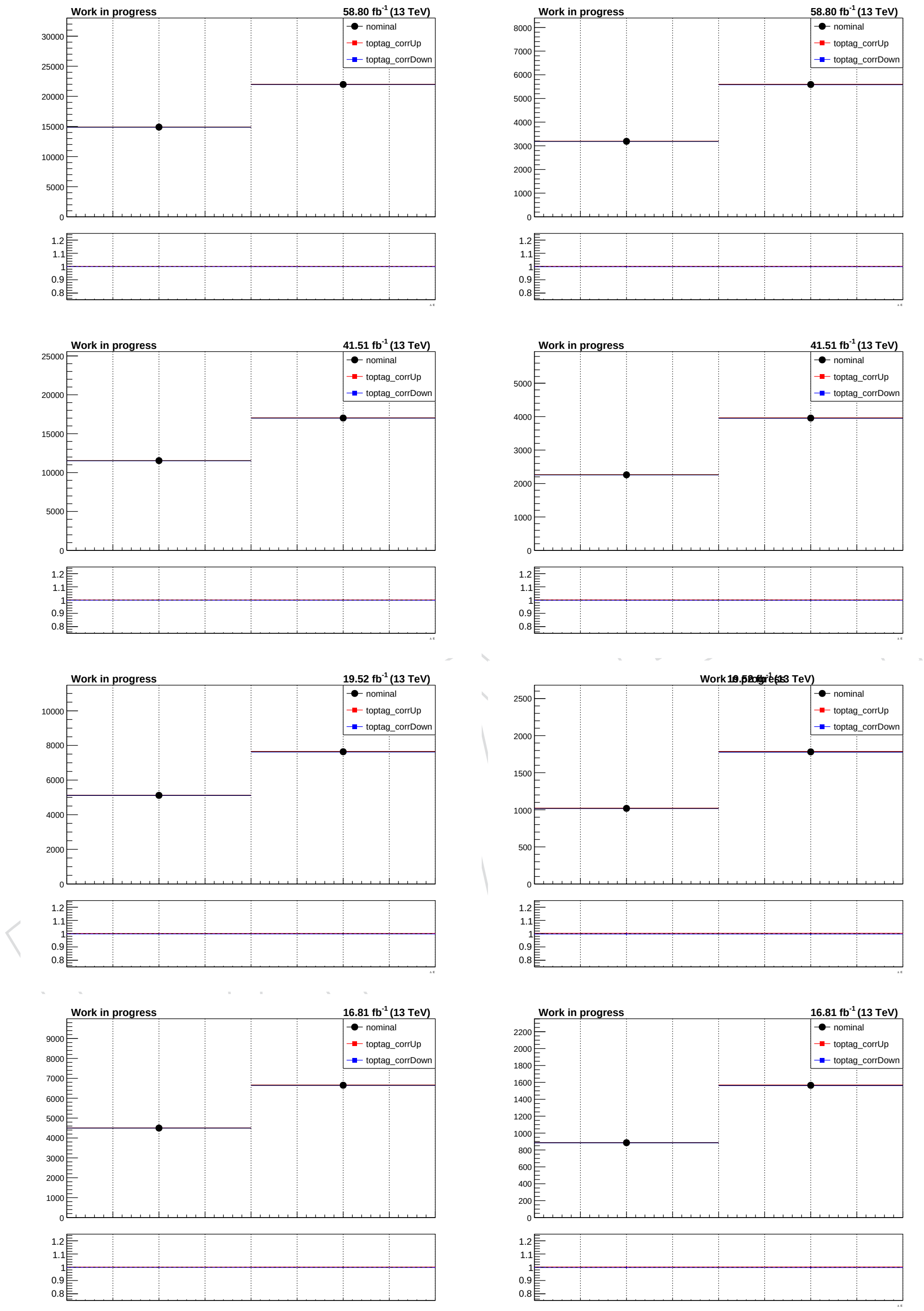

Figure 38: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

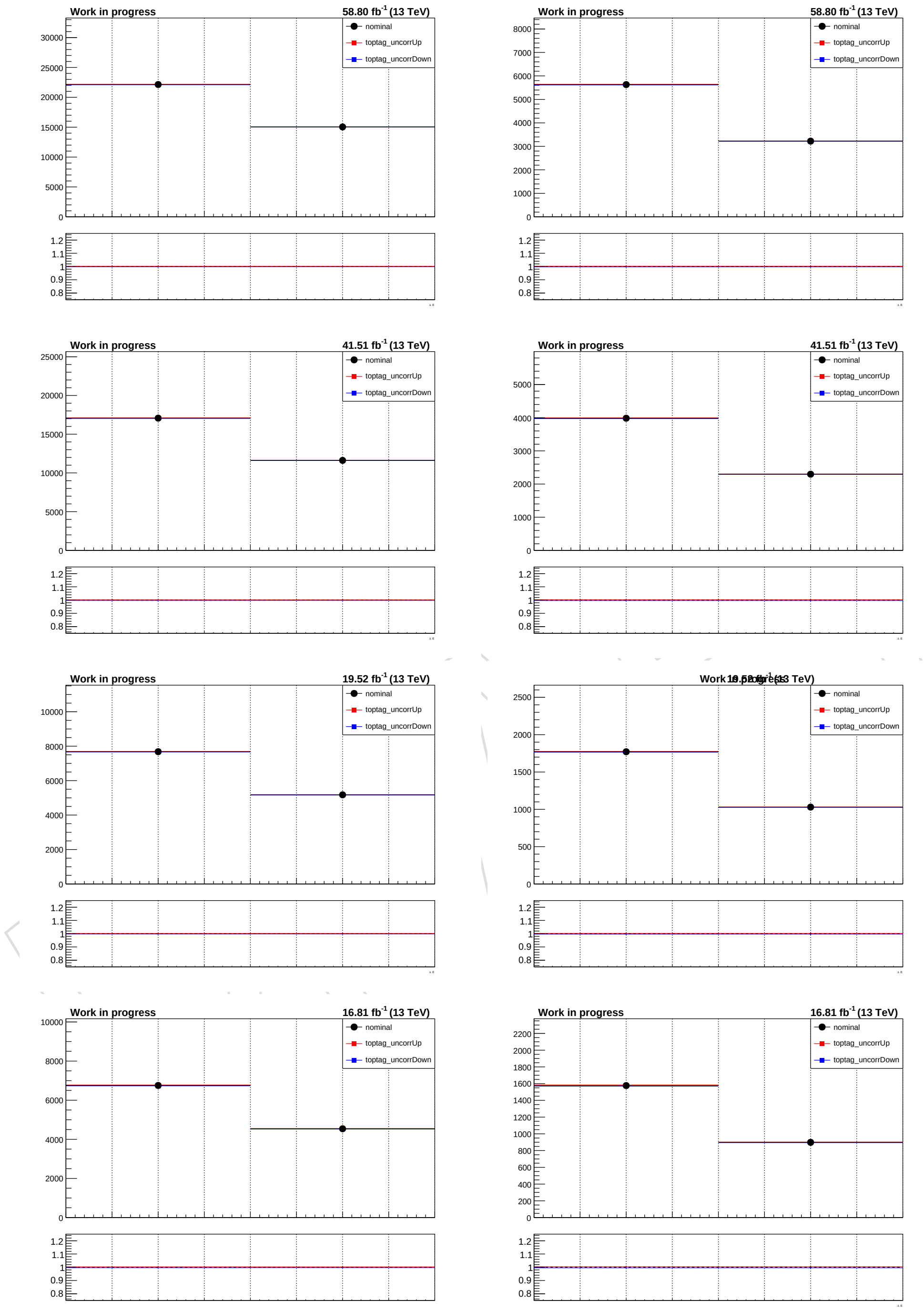

Figure 39: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

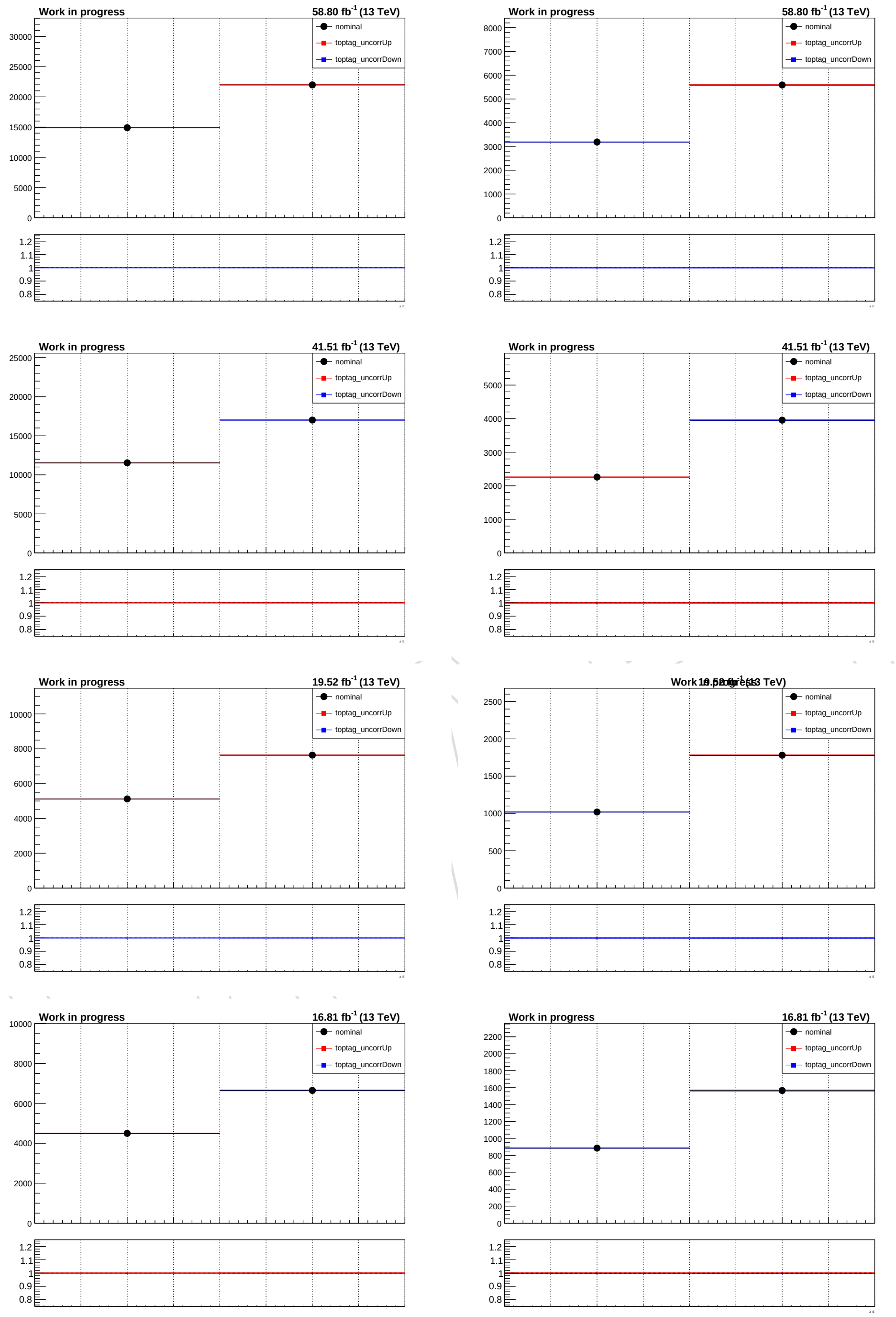

Figure 40: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

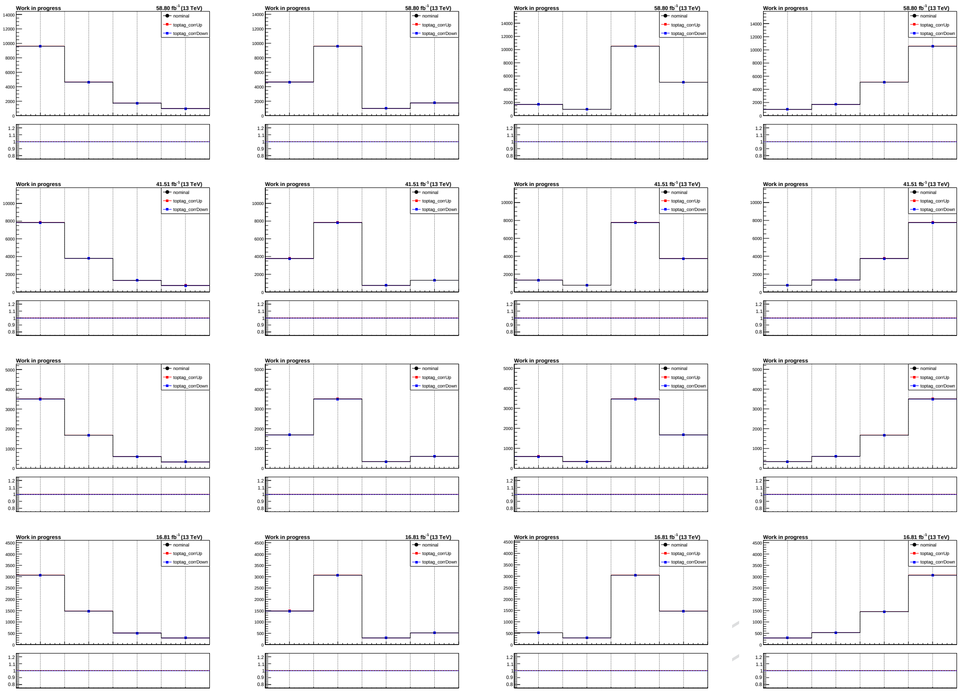

Figure 41: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

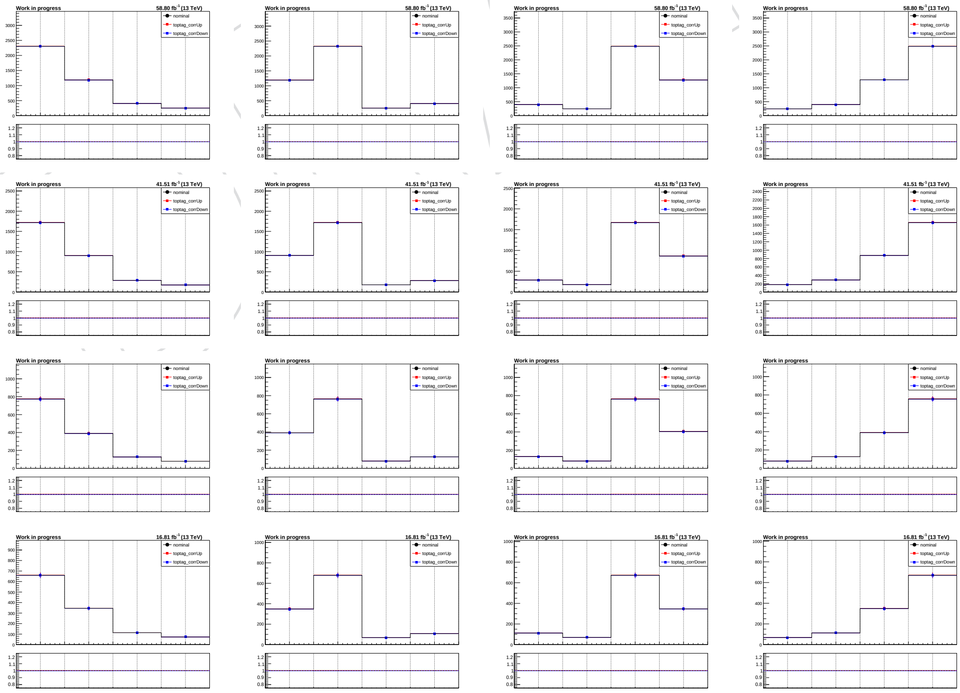

Figure 42: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

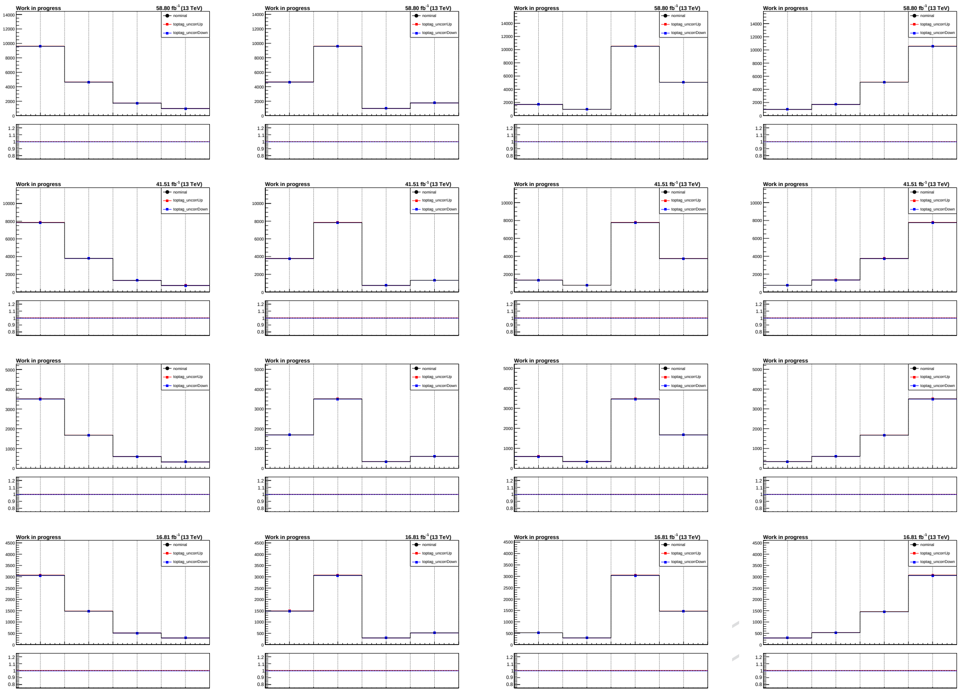

Figure 43: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

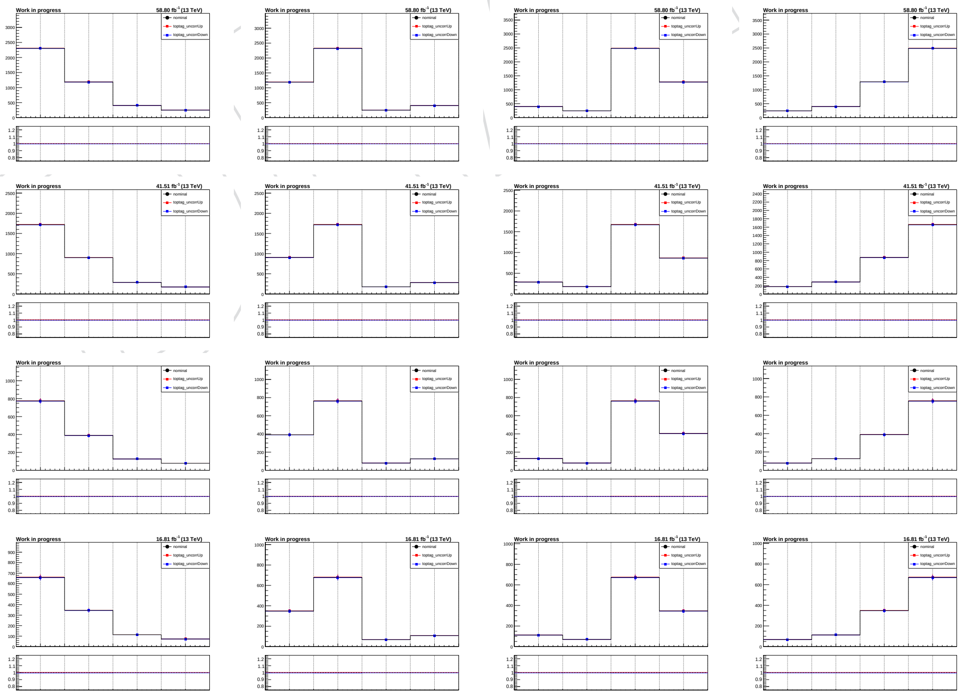

Figure 44: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

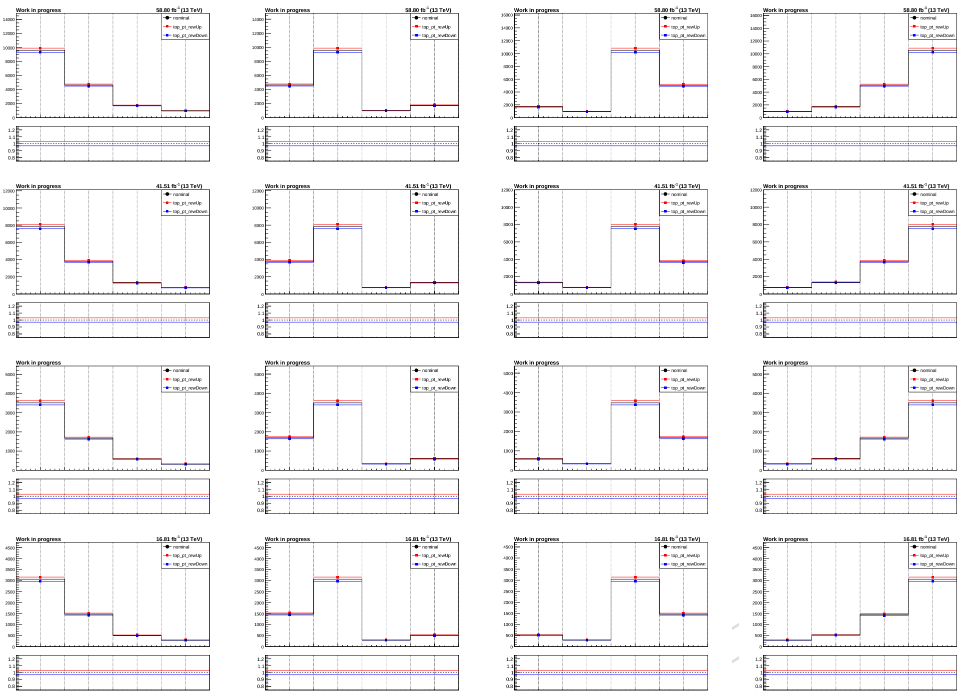

Figure 45: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

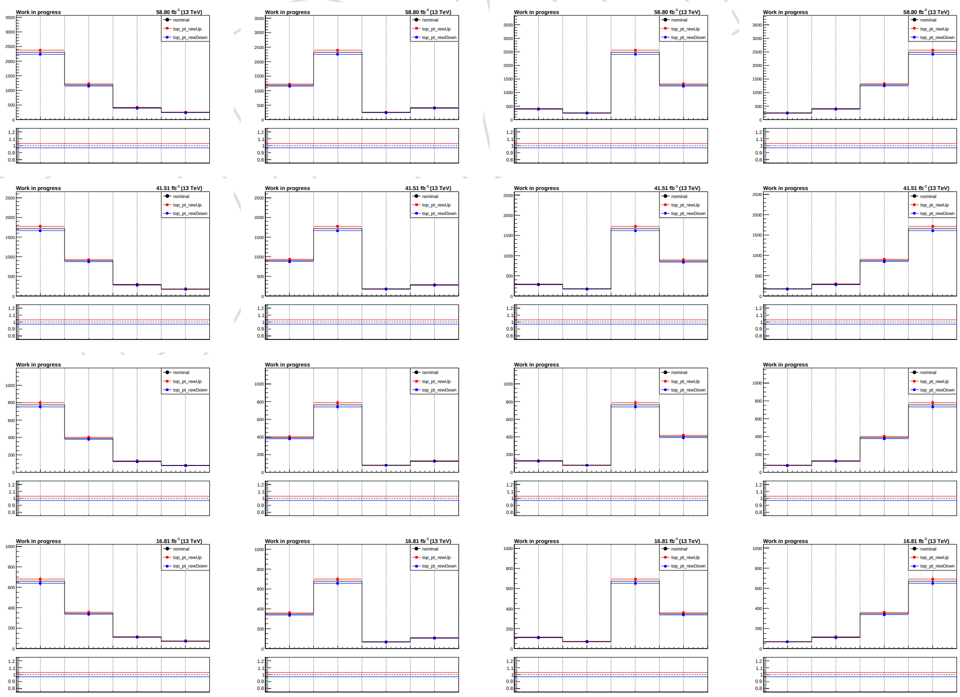

Figure 46: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

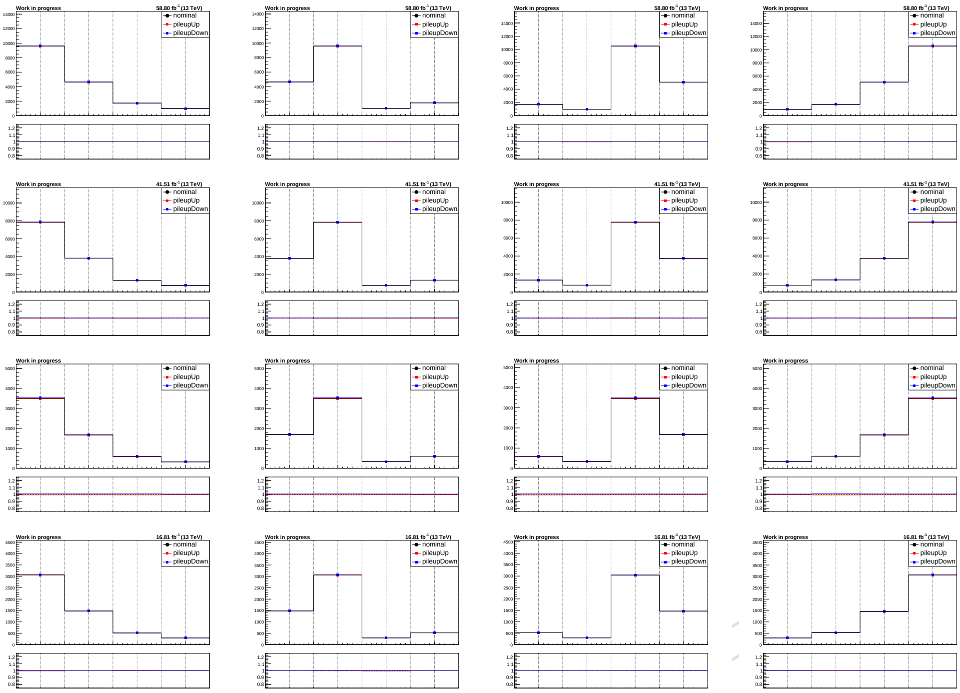

Figure 47: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

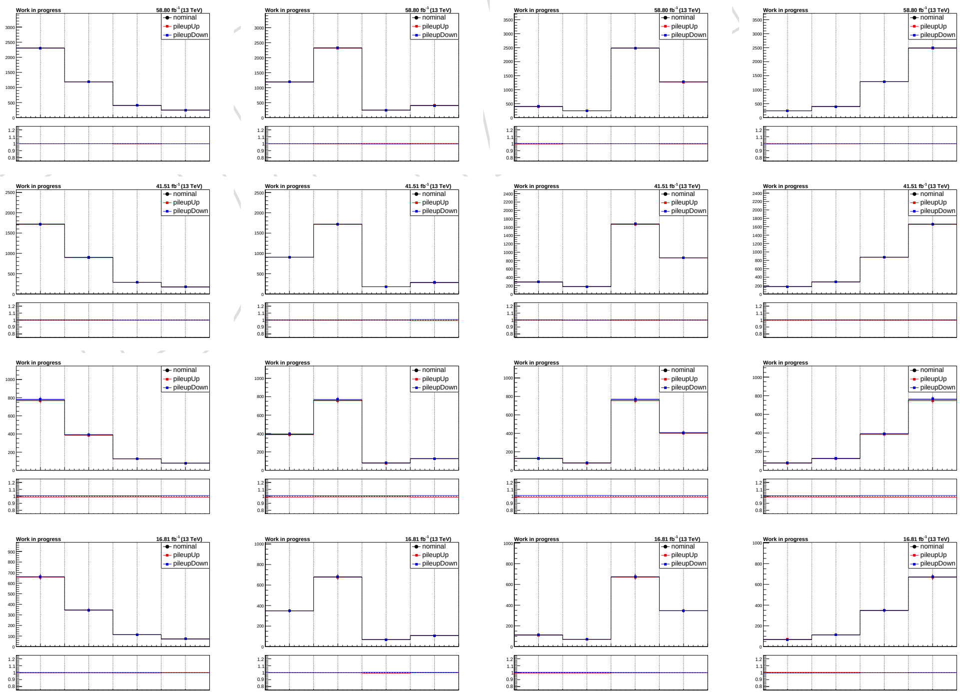

Figure 48: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

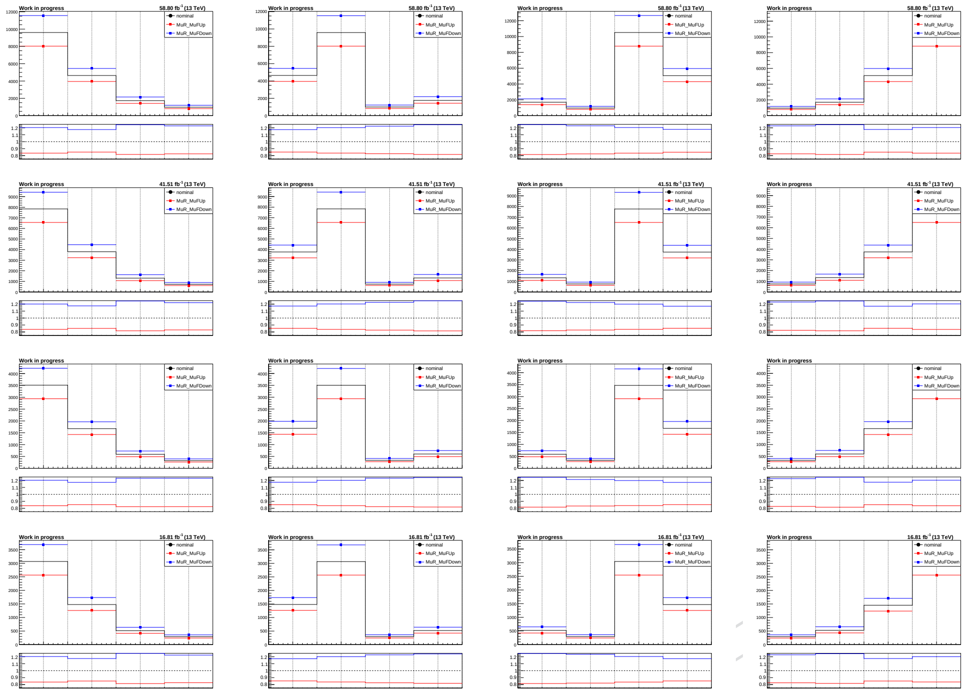

Figure 49: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

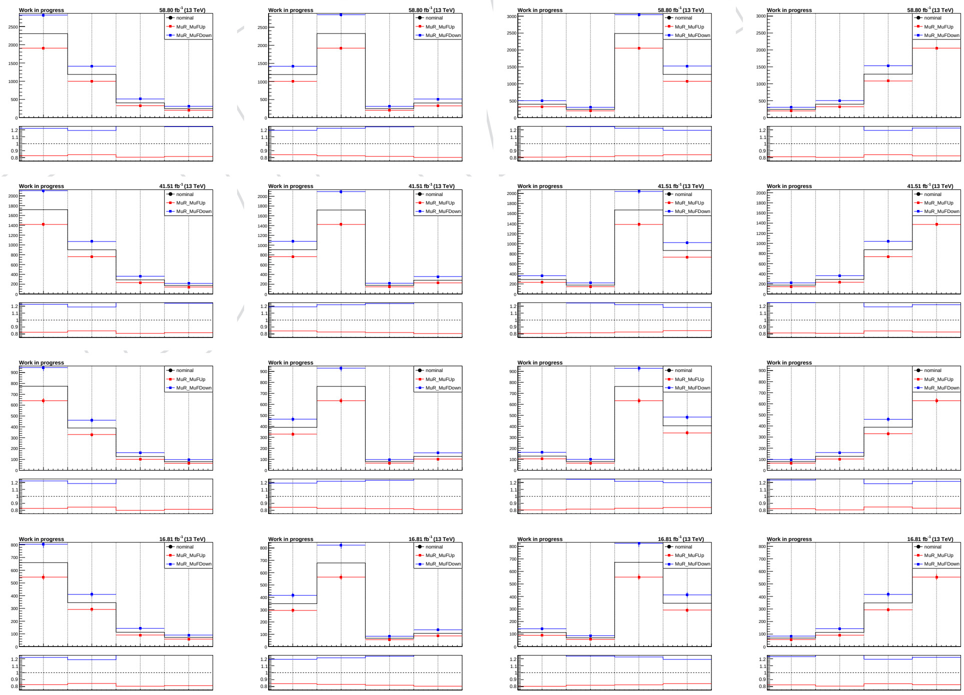

Figure 50: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

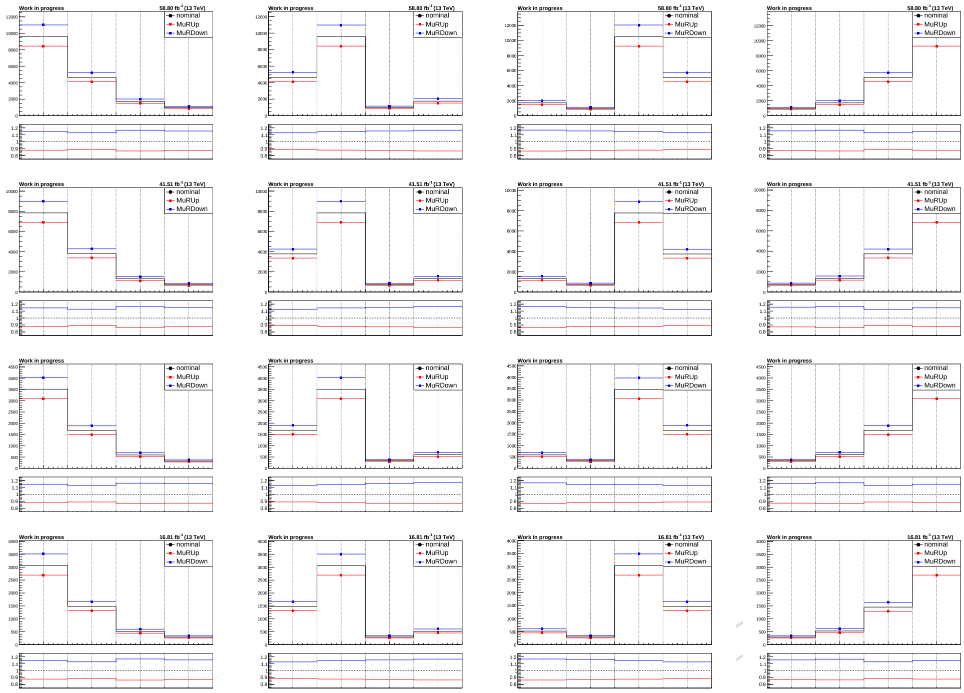

Figure 51: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

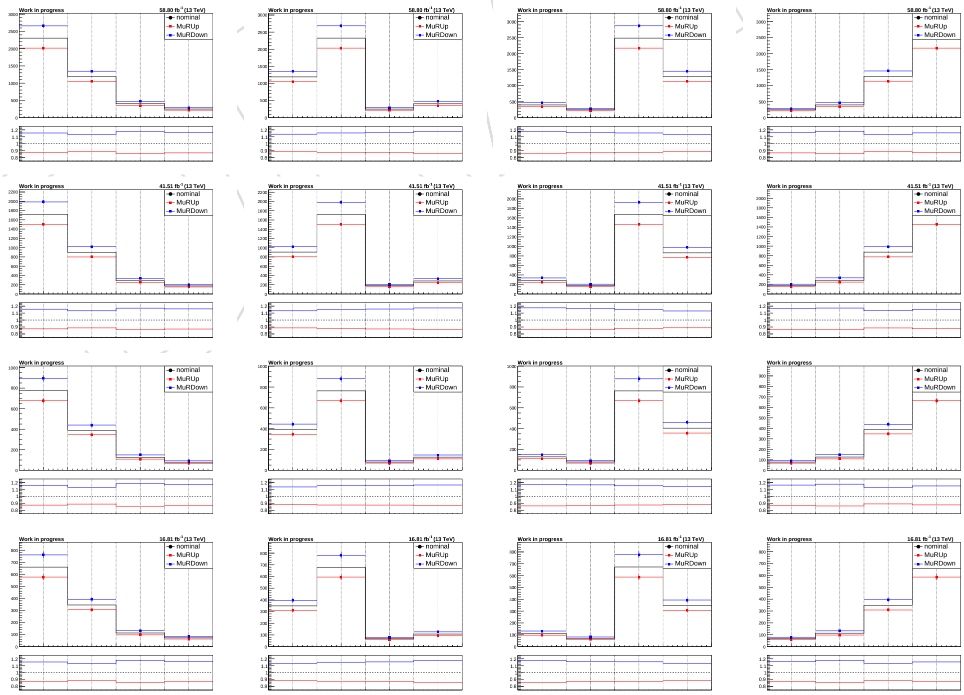

Figure 52: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

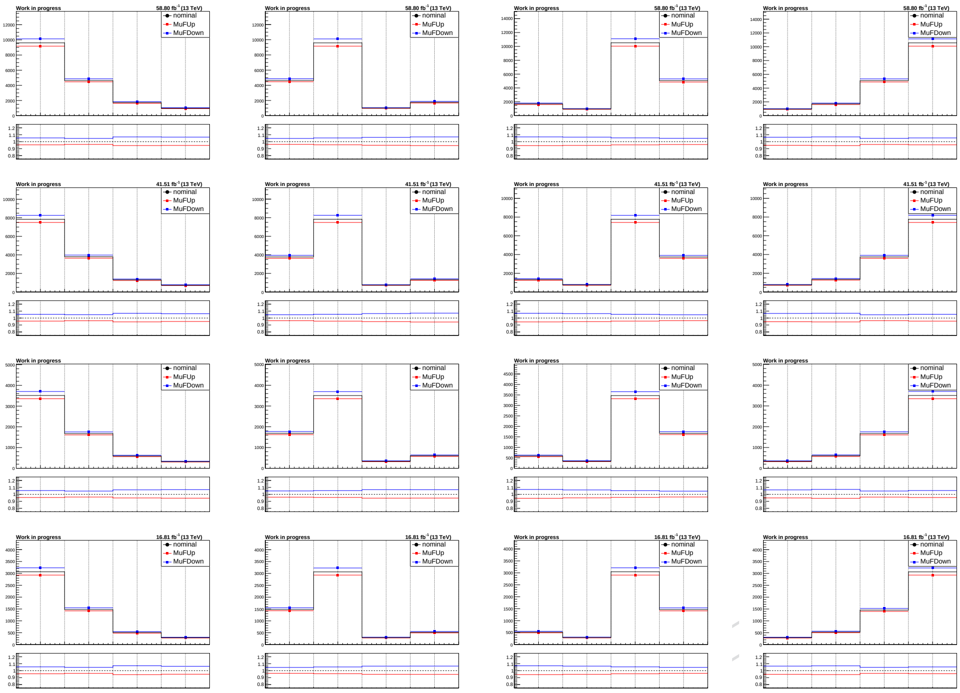

Figure 53: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

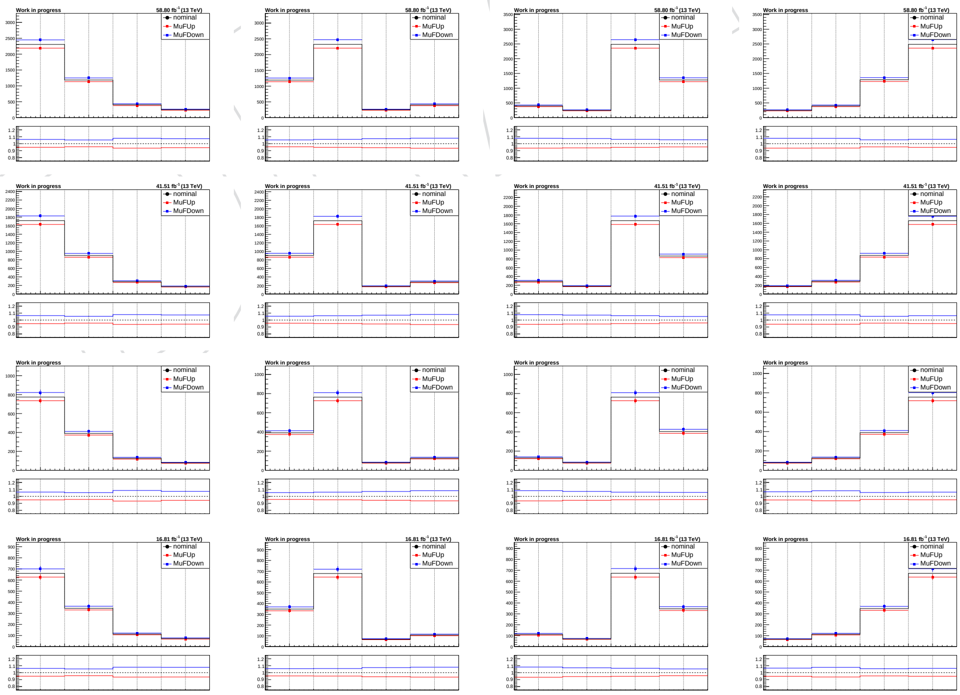

Figure 54: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

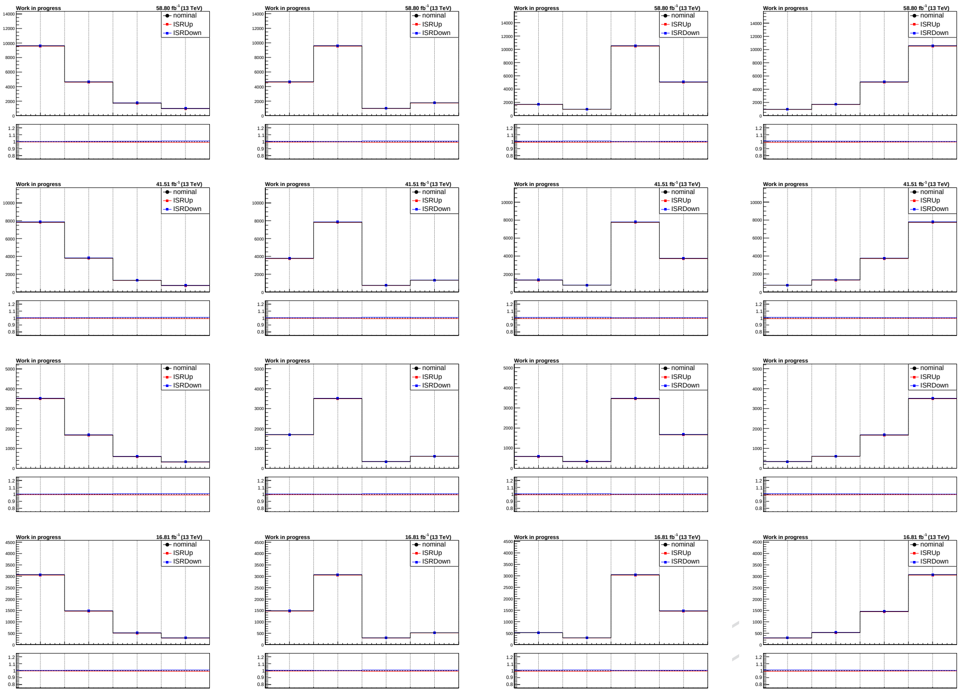

Figure 55: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

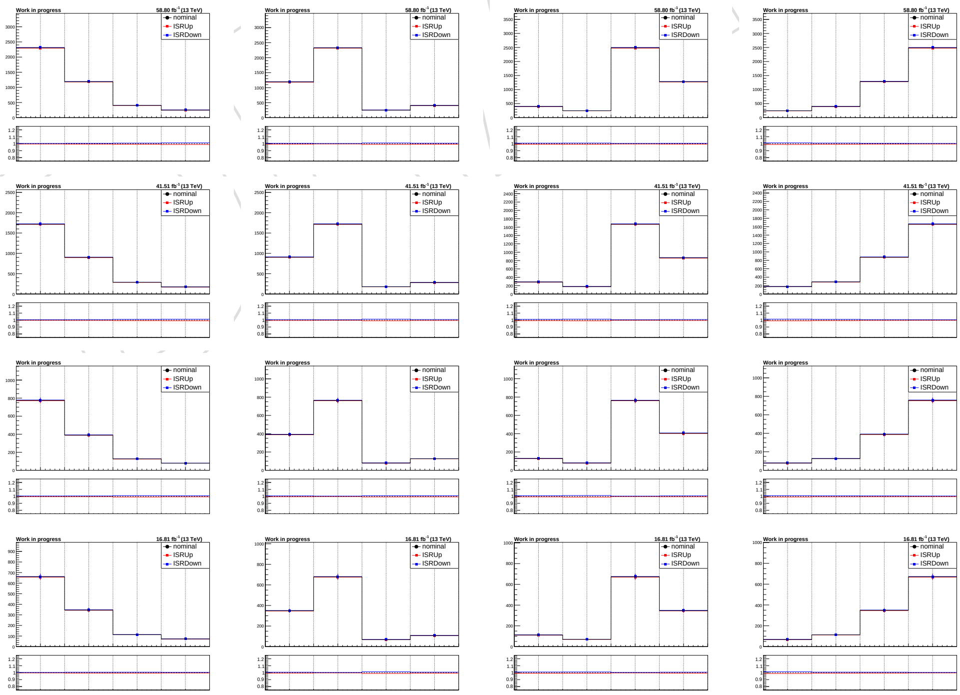

Figure 56: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

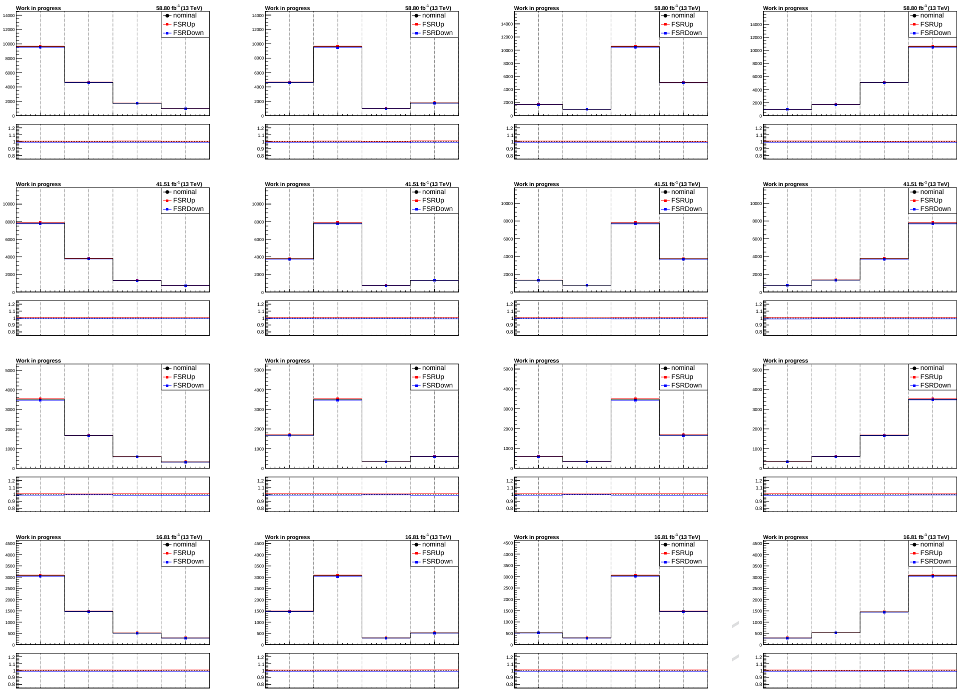

Figure 57: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

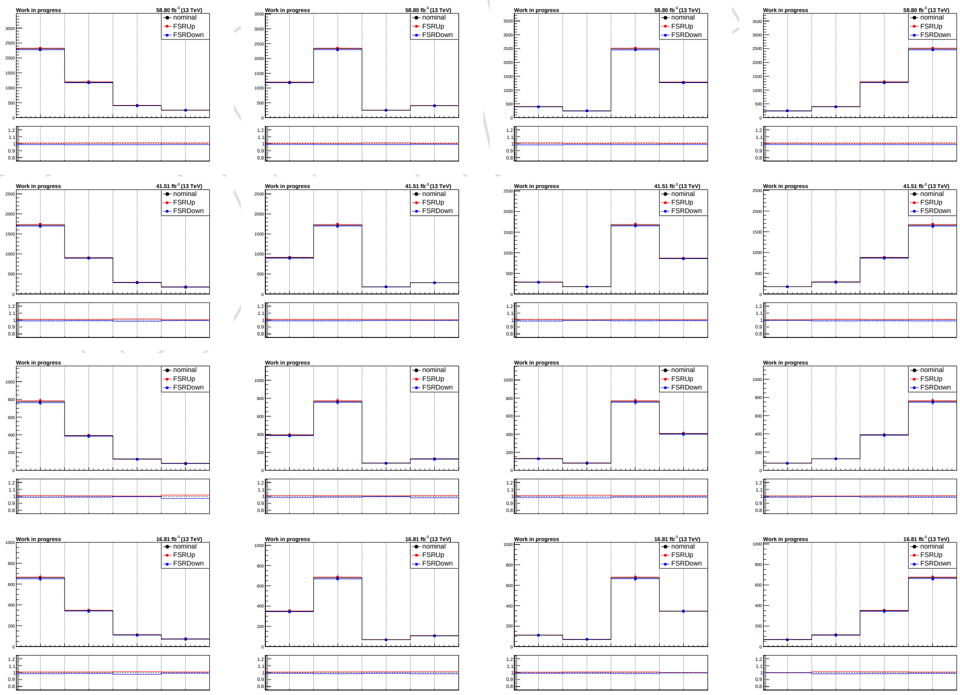

Figure 58: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

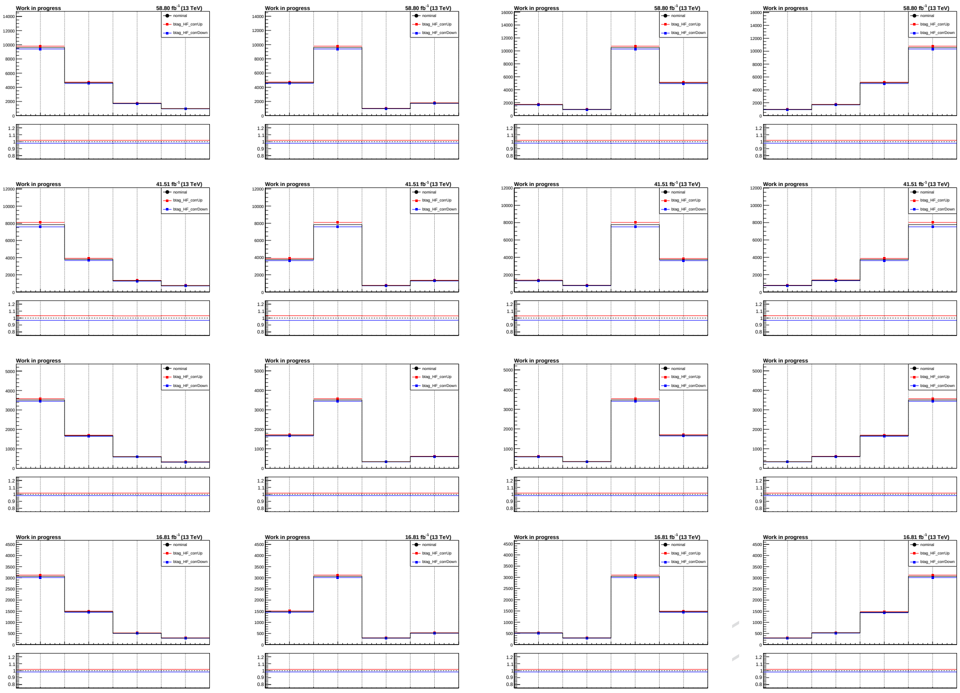

Figure 59: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

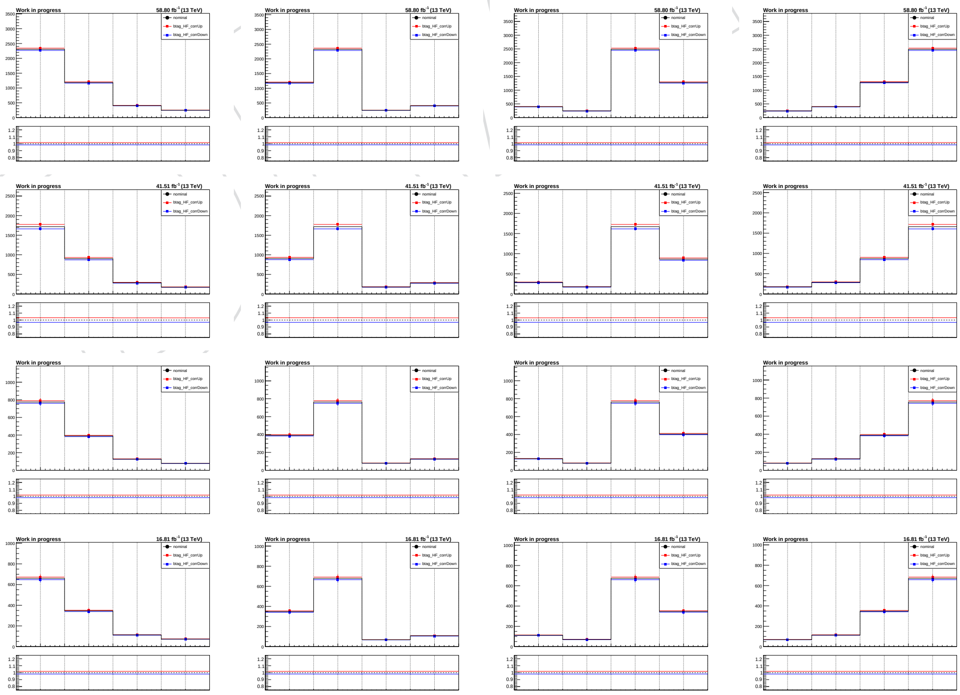

Figure 60: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

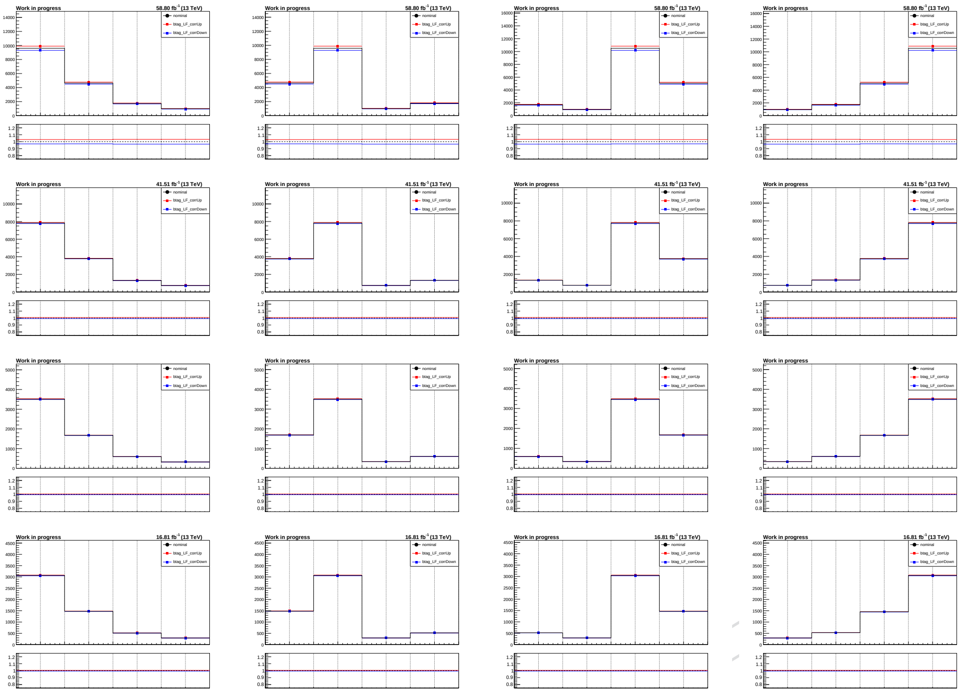

Figure 61: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

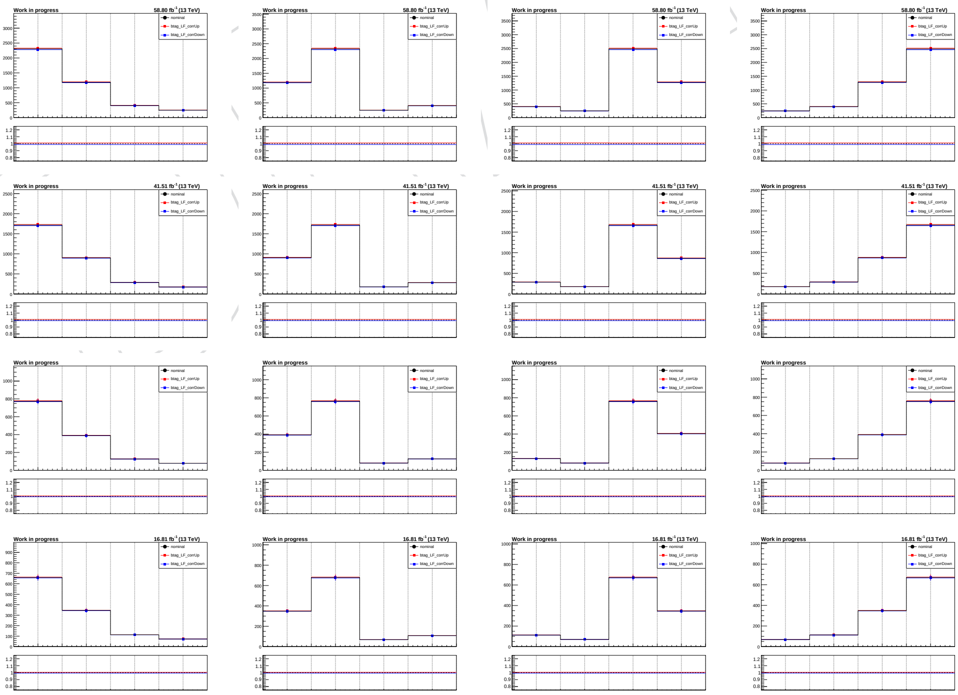

Figure 62: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

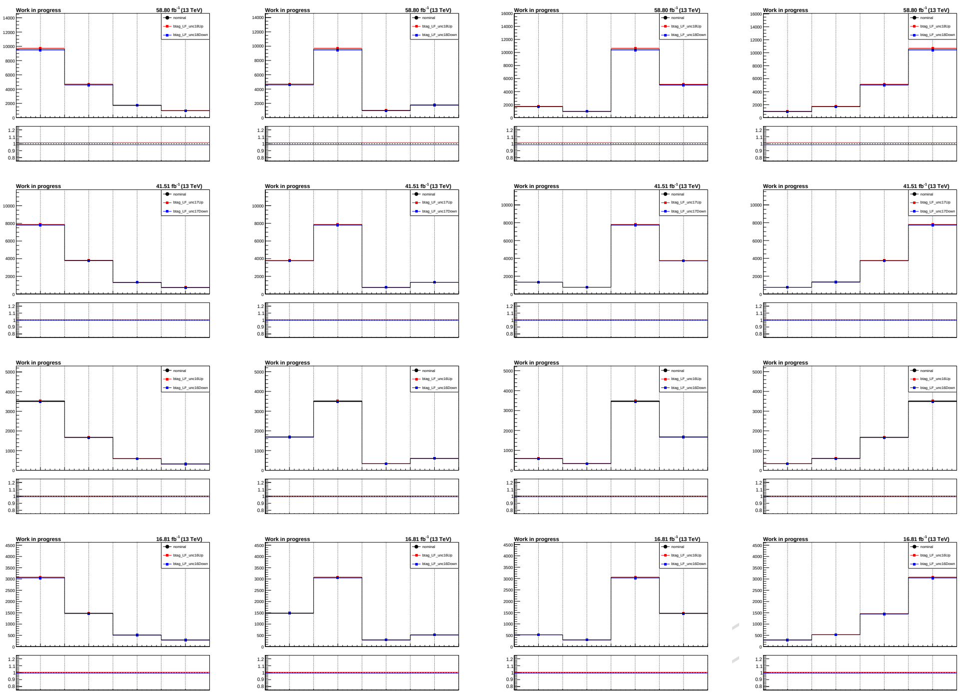

Figure 63: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

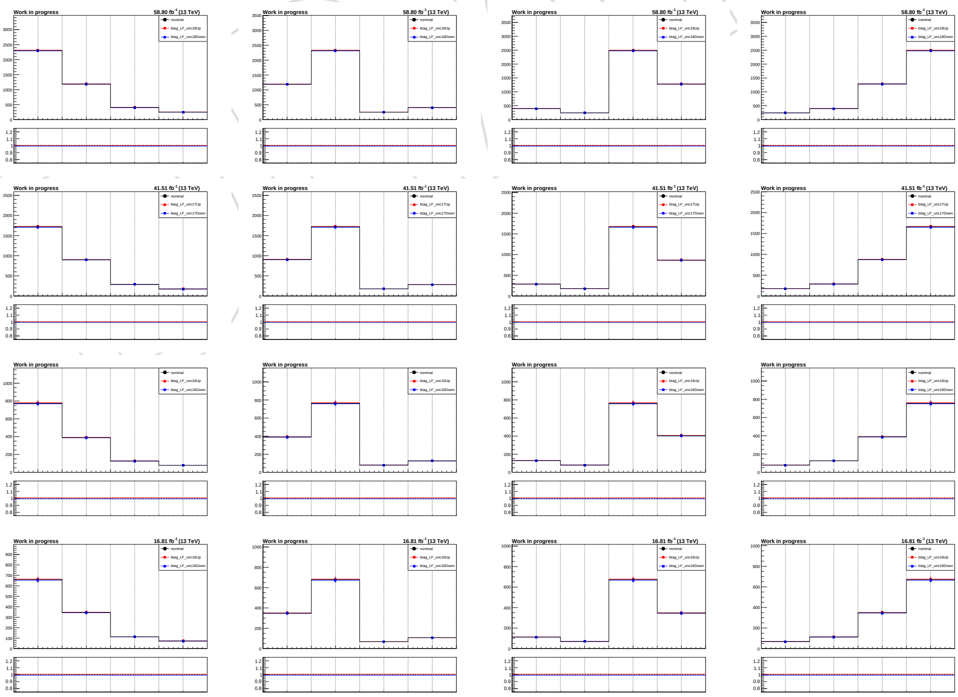

Figure 64: μ_R, μ_F Up/Down variations for the ΔE variable for all the data taken periods. Muon channel (left) and electron channel (right).

D Muon trigger SFs

The scale factors (SF's) used in this analysis are identical to those derived in BSM-22-006. For muons with $p_T > 55$, GeV, the employed triggers are described in Section 3. The trigger combination does not have an isolation requirement on the muon, and the corresponding SF's are not centrally provided. Hence, the corresponding SF's used in this analysis are calculated using the SparkTnP tool [34]. The datasets used are the recommended ones from the muon POG. The official selection is modified to remove the 'Tk-IsoLoose' condition on the probe muon. The results are shown in Fig. 65. The obtained results are in agreement with the official SF's within uncertainties. Therefore, the official SF's are used.

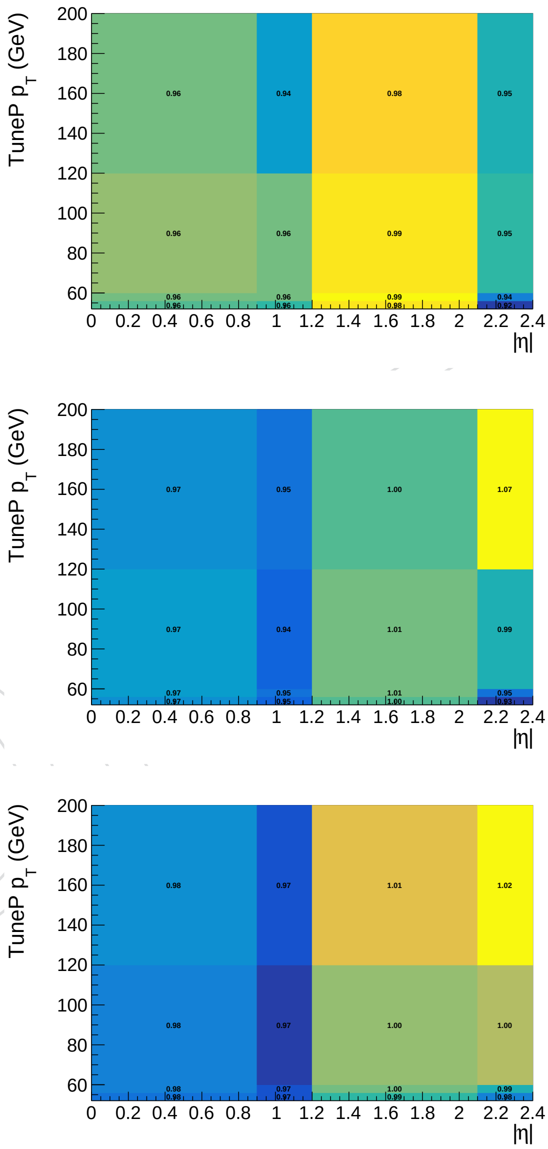

Figure 65: Measured trigger HLT SFs with no isolation requirements on the probe muon. The results are shown for the different running periods: 2016 (top), 2017 (center) and 2018 (bottom). The measured results are in agreement with the central HLT SFs

E Electron trigger SFs

In the electron channel, a combination of HLT paths is utilized for both low pT and high pT electrons.

For electrons with pT ¡ 120 GeV, the chosen HLT path is HLT_Ele27(35)[32]_WPTight_Gsf_v* for the years 2016(2017)[2018], respectively. Beyond this pT threshold, the trigger involves an "OR" combination of the following HLT paths: HLT_Ele115_CaloIdVT_GsfTrkIdT_v*, HLT_Photon175 (200)_v*, and HLT_Ele27(35)[32]_WPTight_Gsf_v* for 2016(2017/2018).

In a segment of the 2017 data, the HLT_Ele115_CaloIdVT_GsfTrkIdT_v* path is unavailable, and as a workaround, it is substituted with HLT_Ele35_WPTight_Gsf_v*. This substitution is specifically applied to data in 2017 RunB (run numbers below 299329) and to a fraction of MC corresponding to 11.58% of the RunB data.

The scale factor are taken from the B2G-22-006 that have a similar same phase space in the electron channel. The method is explained in more detail in the appendix E of the AN2019_197_v9 and already aproved by the EGamma POG. Here I show the SF's only

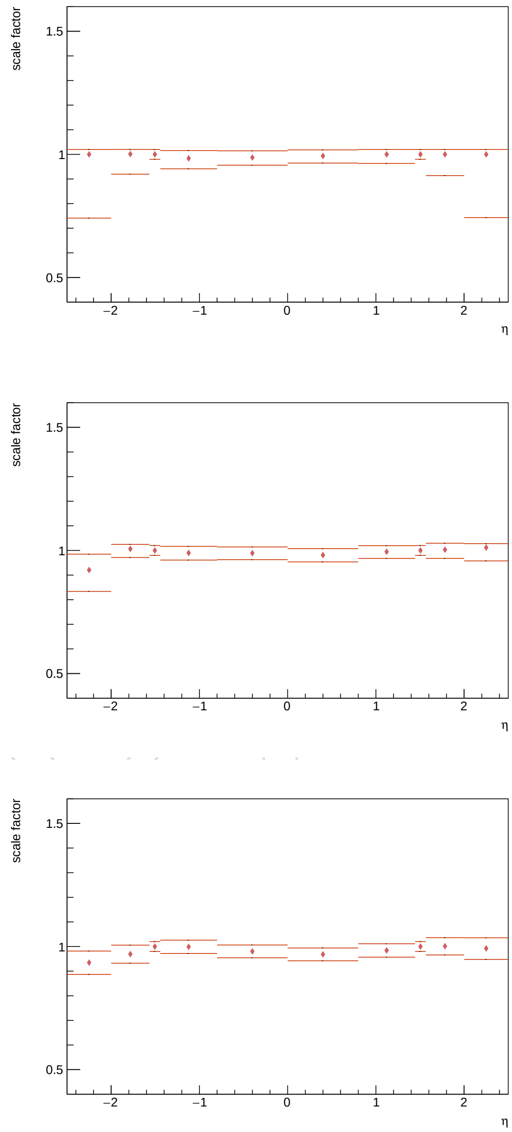

Figure 66: Scale factors, for pT ¿ 200 GeV (upper), 120 ¡ pT ¡ 200 GeV (center) and pT ¡ 120 GeV (lower) for 2016postVFP

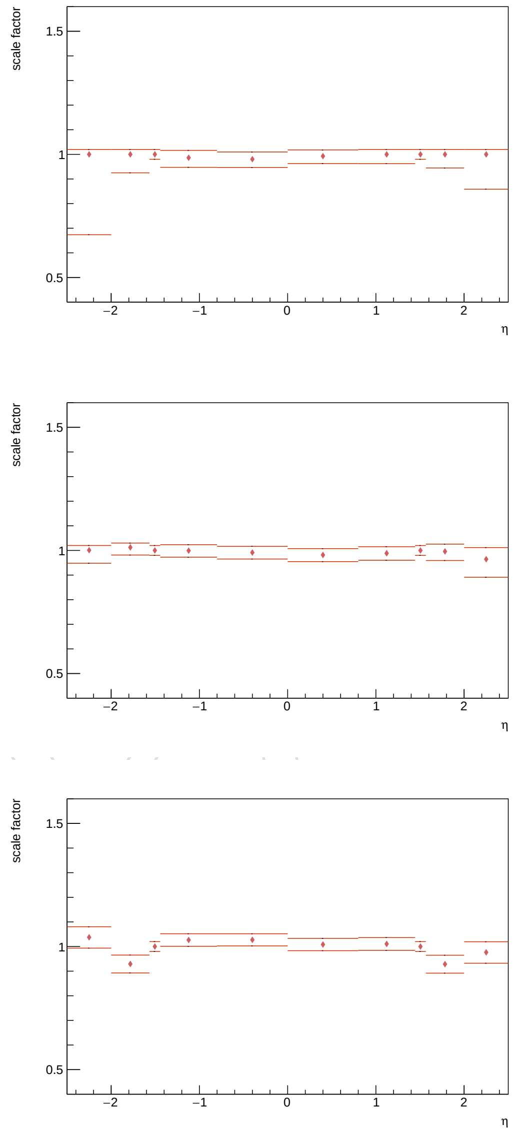

Figure 67: Scale factors, for pT ¿ 200 GeV (upper), 120 ¡ pT ¡ 200 GeV (center) and pT ¡ 120 GeV (lower) for 2016preVFP

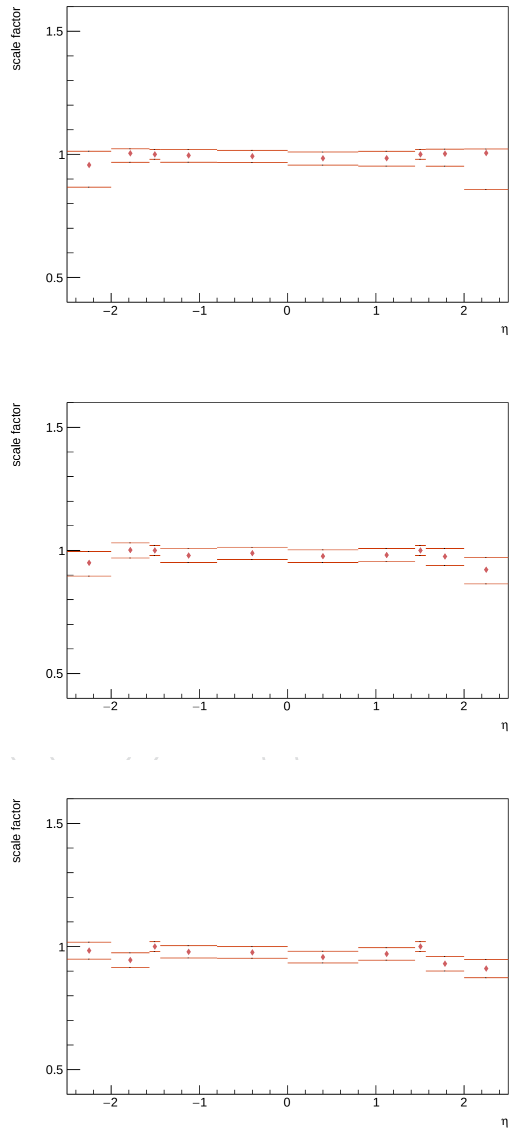

Figure 68: Scale factors, for pT ¿ 200 GeV (upper), 120 ¡ pT ¡ 200 GeV (center) and pT ¡ 120 GeV (lower) for 2017

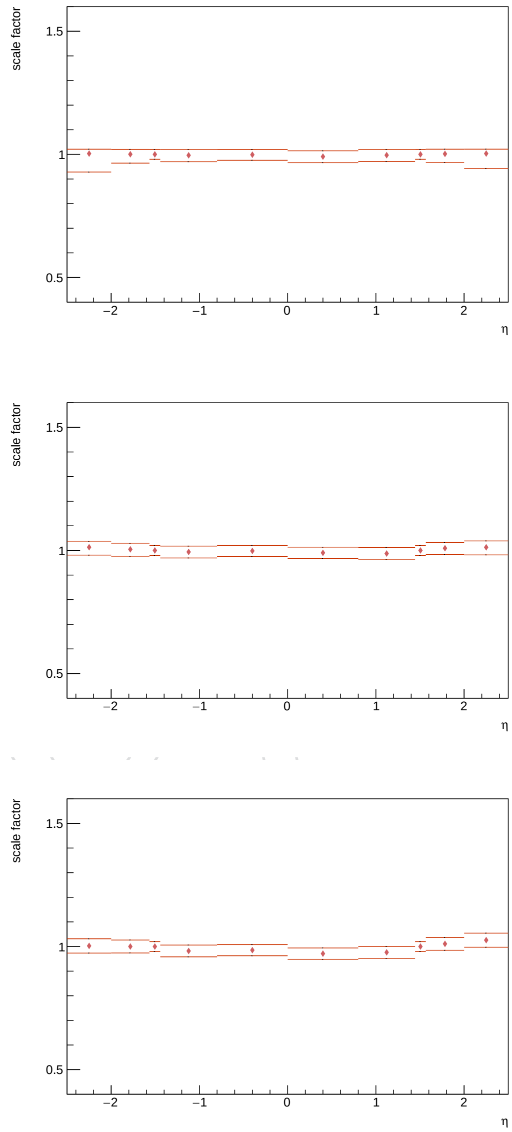

Figure 69: Scale factors, for pT ¿ 200 GeV (upper), 120 ¡ pT ¡ 200 GeV (center) and pT ¡ 120 GeV (lower) for 2018

F Tuning of Mass Reconstruction Parameters

Using the same procedure as in TOP-21-014, we tuned the parameters $\bar{M_{had}}$, $\bar{M_{lep}}$, σ_{had} and σ_{lep} used in χ^2 discriminator introduced in Sec 4.

The fine-tuning is done with $t\bar{t}$ Monte Carlo simulation. We specifically target events passing the l+jets selection detailed in Sec 4. Within these events, we ascertain the correct hypothesis by employing a matching procedure based on angular separation between the reconstructed objects associated with the hypothesis and the corresponding GEN particles.

- Hadroinc top:
  - For events with a t-tagged jet, all quarks from the hadronic top decay are required to be at $\Delta R < 0.8$ from the AK8 t-tagged jet.
  - For events with a W-tagged jet, all quarks from the hadronic W decay are required to be at $\Delta R < 0.8$ from the AK8 W-tagged jet; and the b-quark is required to be at $\Delta R < 0.4$ from an AK4 jet.
  - For events without a t-tagged or a W-tagged jet, all AK4 jets associated to the hadronic top decay are required to be at $\Delta R < 0.4$ from a quark originating from the top decay; each quark is required to be matched to an AK4 jet, but more than a quark is allowed to be matched to the same jet. This covers cases in which there is no t-tag or W-tag but some of the top decay products are still collimated.
- Leptonic top: The AK4 jet associated to the leptonic top decay is required to be at $\Delta R < 0.4$ from the b quark originating from the generated leptonic top; the ΔR distance between reconstructed and generated leptons must be smaller than 0.1; the $\Delta\phi$ distance between the generated neutrino and the missing transverse energy must be smaller than 0.3.

If more than one jet assignment fulfills these requirements, we take as the correct $t\bar{t}$ hypothesis the one with the smallest $\sum \Delta R$ between reconstructed objects and GEN particles.

We define matchable events as those for which a correct $t\bar{t}$ hypothesis exists. According to this definition, only events featuring semileptonic $t\bar{t}$ decays at the generator level can be considered matchable.

The distributions of M_{lep} and M_{had} for the correct $t\bar{t}$ hypotheses are independent of the year or lepton channel. These distributions are depicted in Fig. 48, 49, and 50 for the $\mu +$ jets channel for 2018, 2017, and 2016, respectively. Each histogram is fitted with a Gaussian function, and the resulting mean value and width of the reconstructed top masses are utilized in Equation 6, which constitutes the discriminator.

Separate fits are carried out for the leptonic top, the t-tagged AK8-reconstructed hadronic top, the W-tagged AK8-reconstructed hadronic top, and the AK4-reconstructed hadronic top. The mean values obtained from the Gaussian fits, used in the χ^2 discriminator, are as follows:

- $M_{\text{lep}} = 172$ GeV, $\sigma_{M_{\text{lep}}} = 26$ GeV
- $M_{\text{had}} = 175$ GeV, $\sigma_{M_{\text{had}}} = 21$ GeV (when no t/W tag jet)
- $M_{\text{had}} = 178$ GeV, $\sigma_{M_{\text{had}}} = 14$ GeV (for ttag reconstruction)
- $M_{\text{had}} = 171$ GeV, $\sigma_{M_{\text{had}}} = 25$ GeV (for wtag reconstruction)

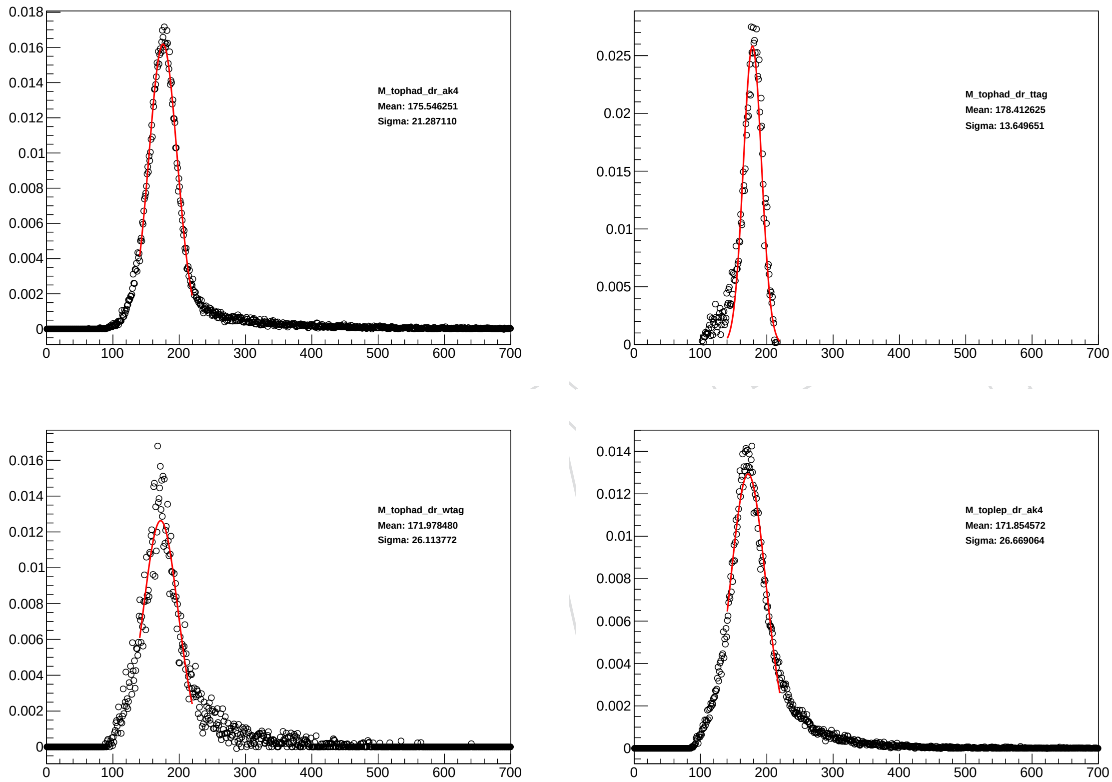

Figure 70: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $e+$ jets for 2017 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

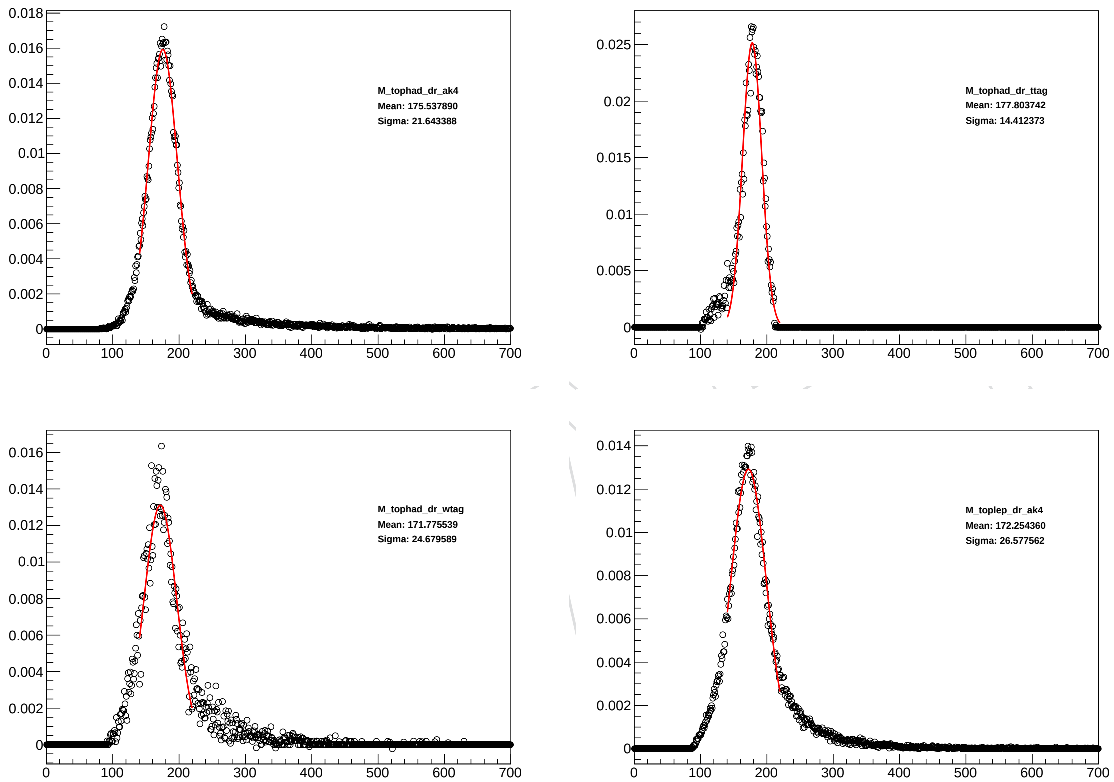

Figure 71: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $e+$ jets for 2017 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

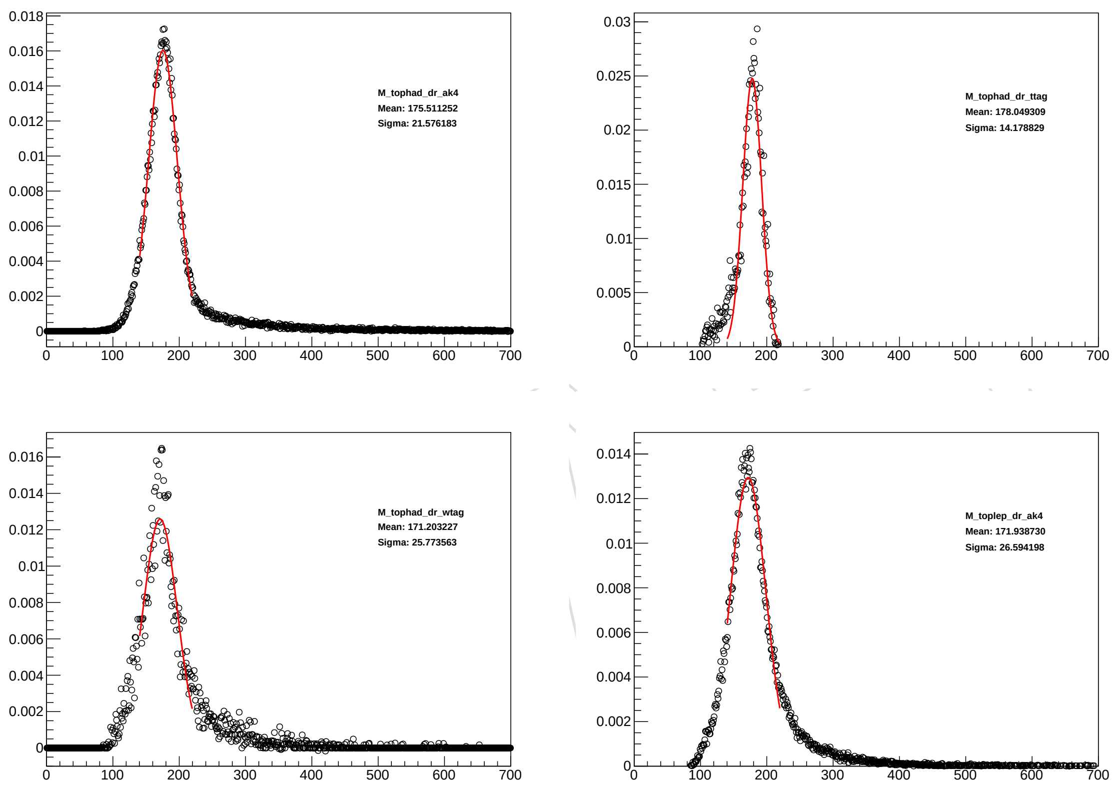

Figure 72: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $e+$ jets for 2017 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

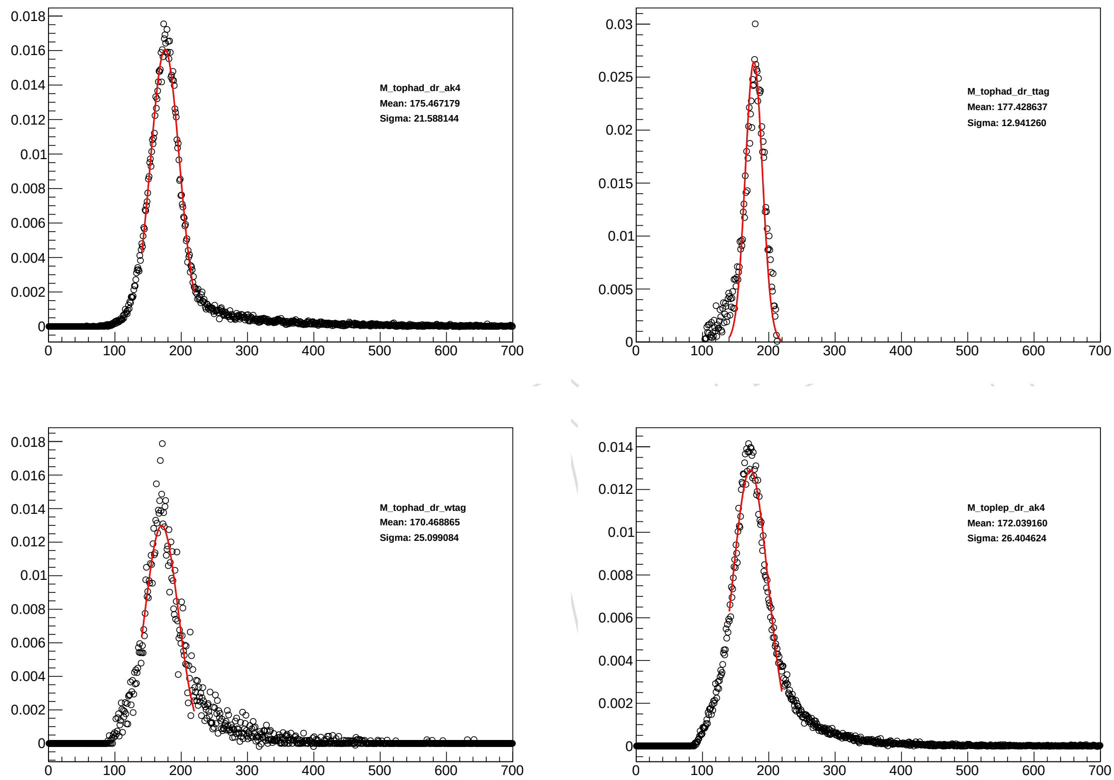

Figure 73: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $e+$ jets for 2017 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

612 G Control Plots

613 More Data/MC plots after the event selection and reconstruction described in Sec 4. Each
614 variable is shown for the three years in the full Run II data and separately for the $\mu +$ jets and
615 $e +$ jets channels.

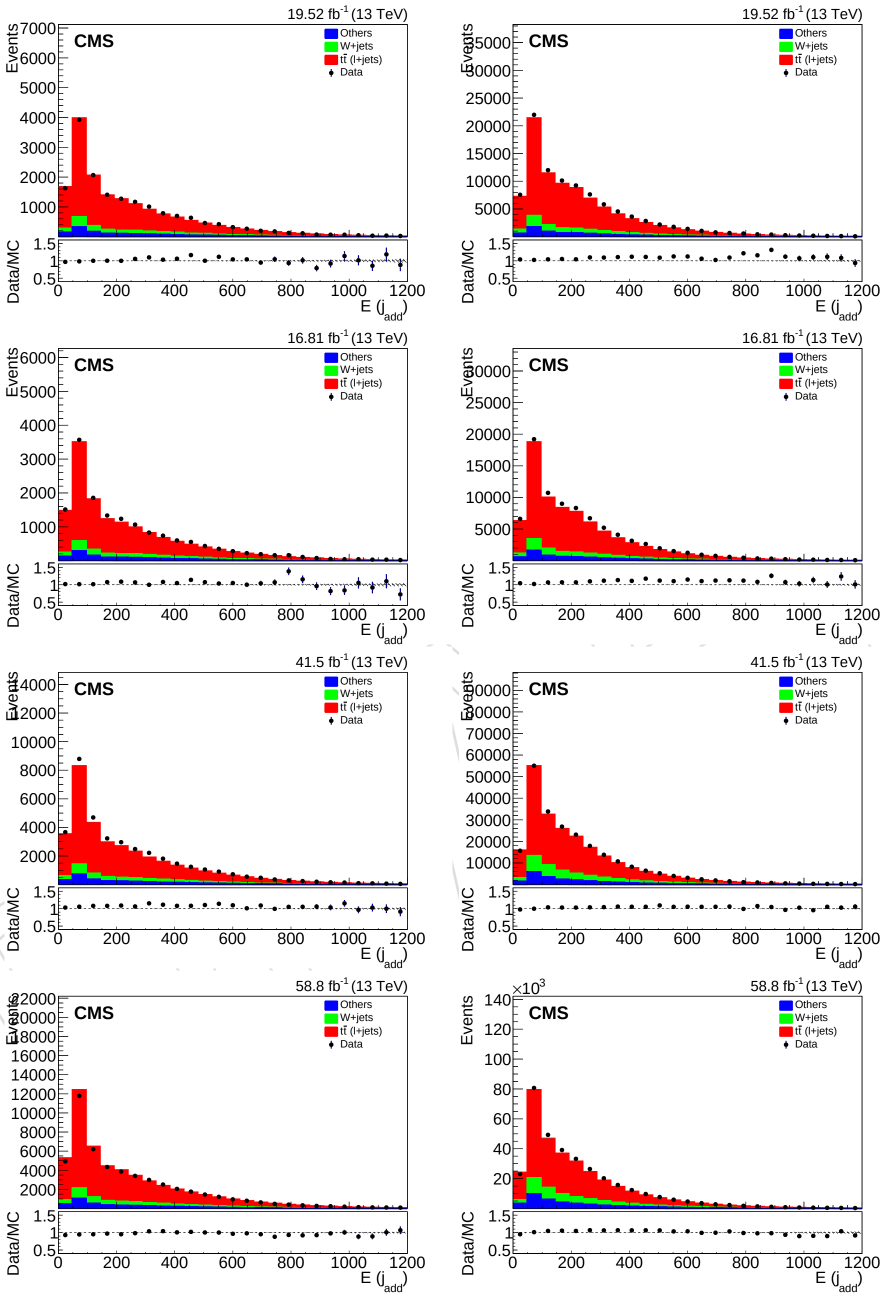

Figure 74: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

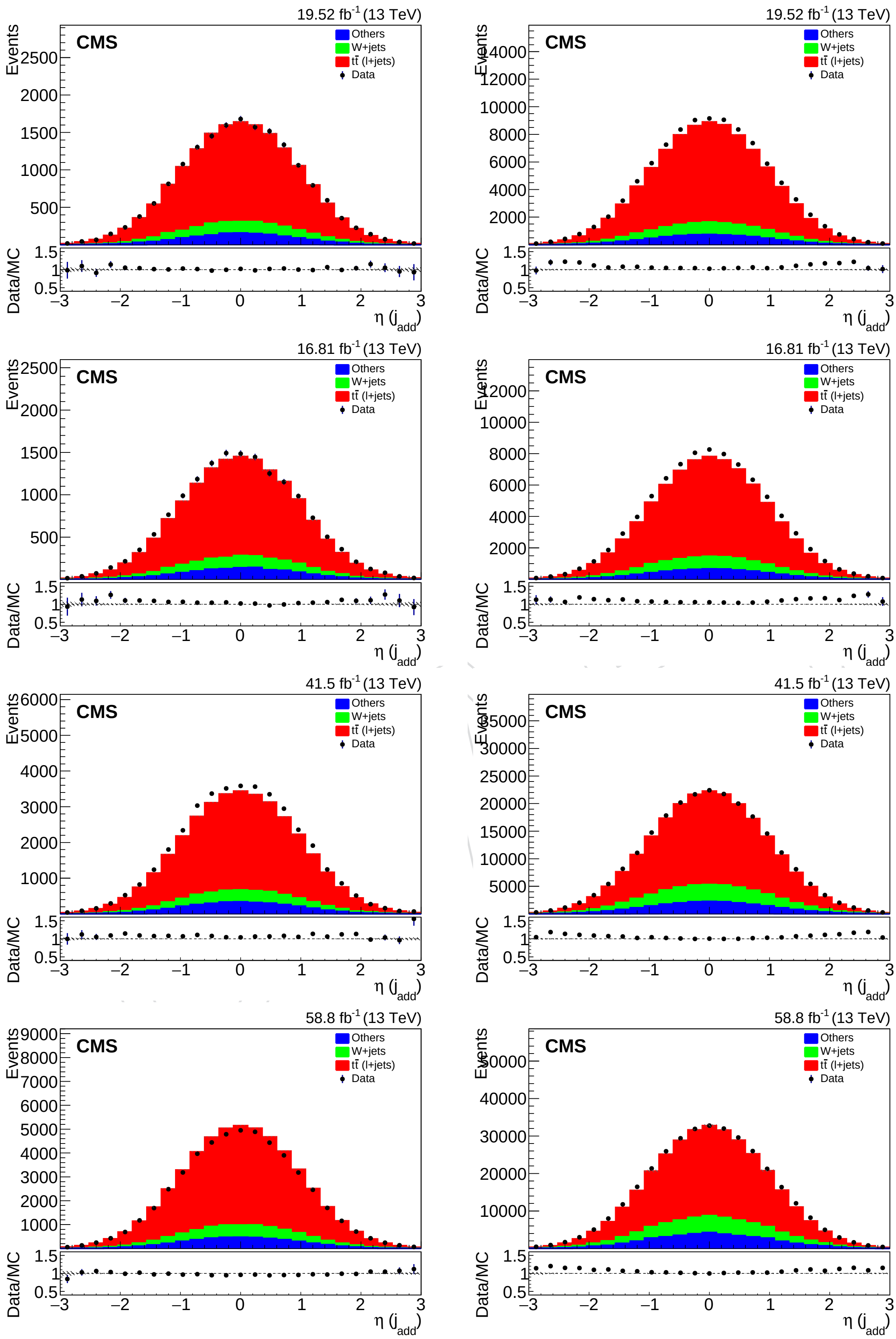

Figure 75: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

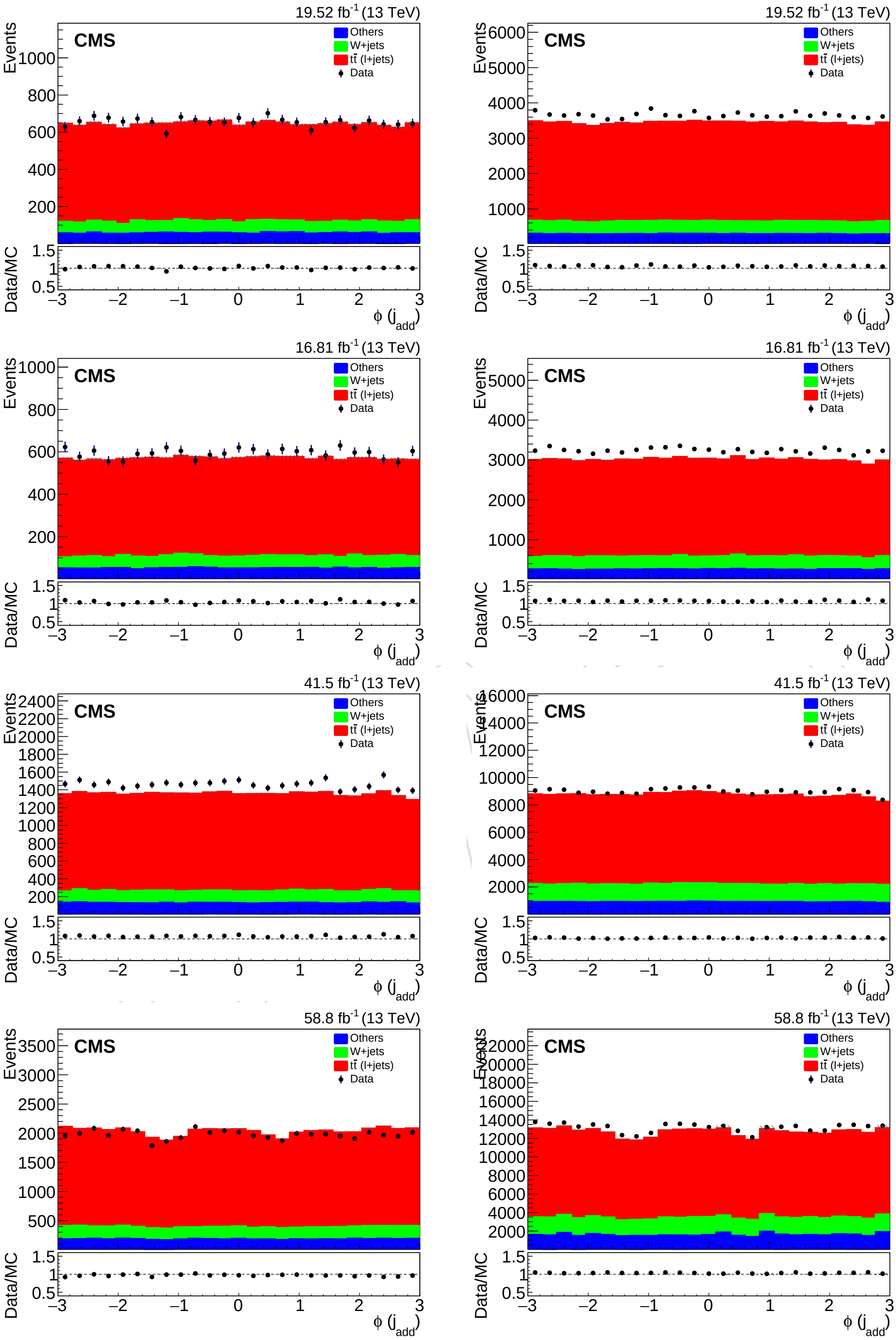

Figure 76: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

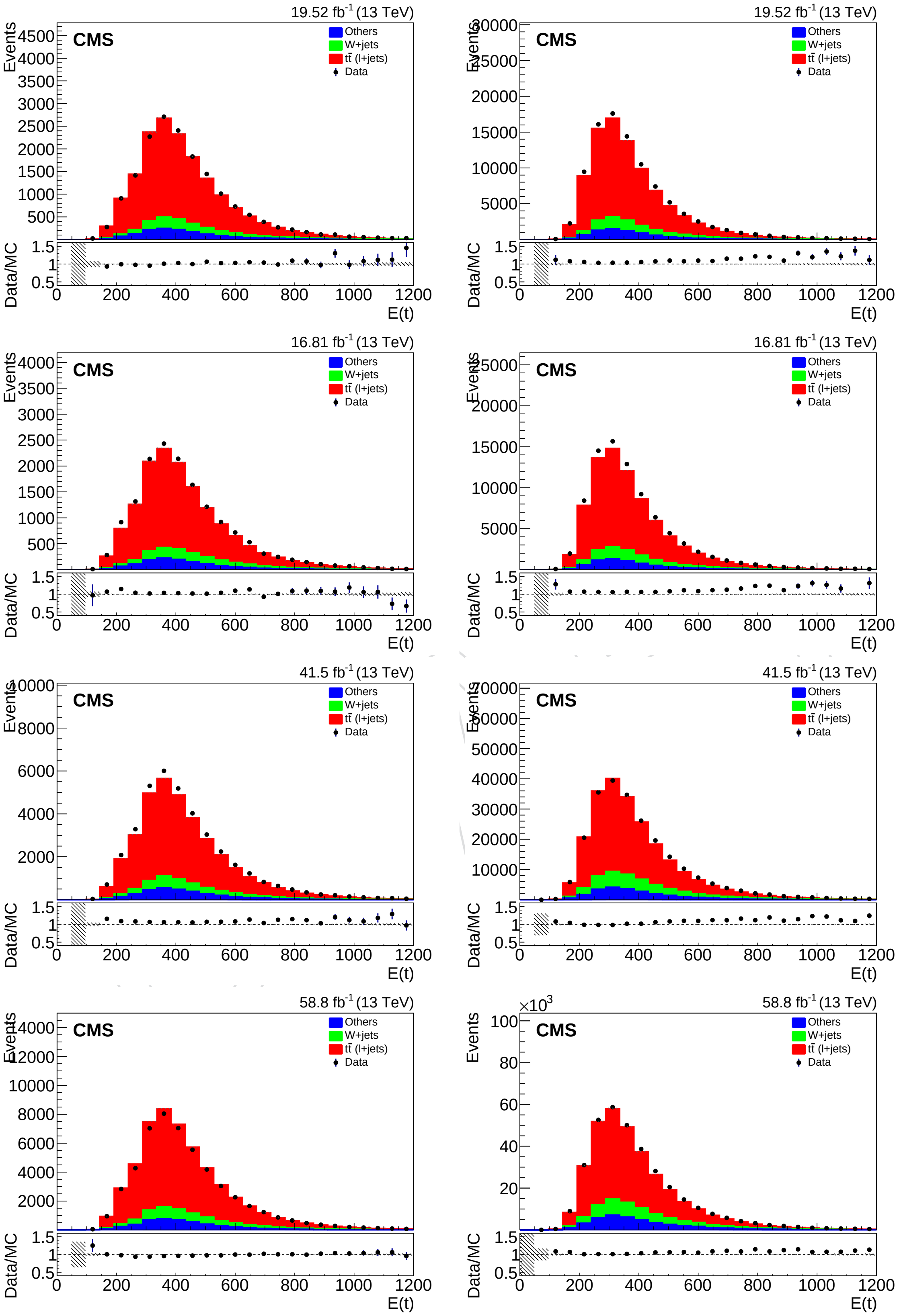

Figure 77: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

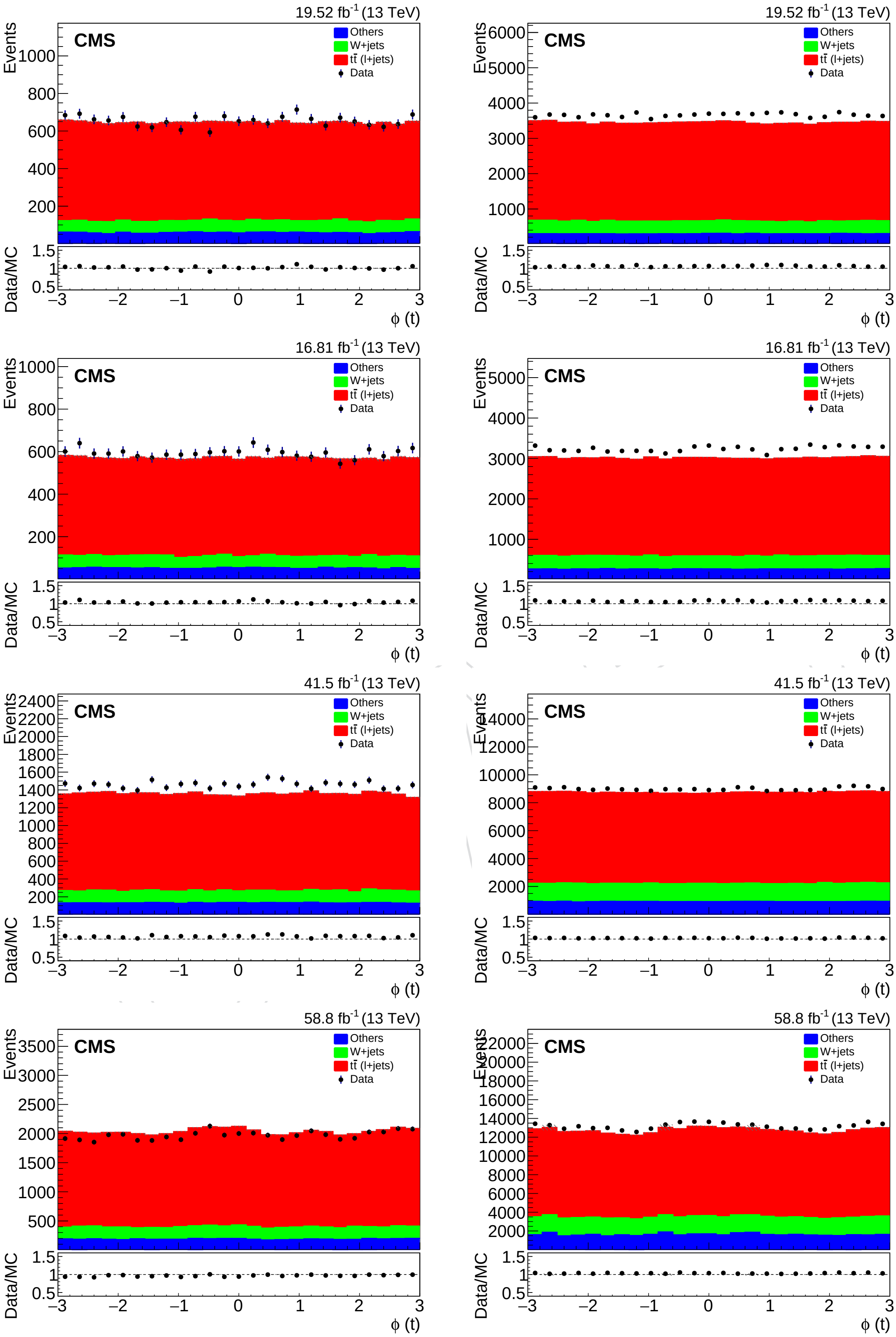

Figure 78: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the μ+ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

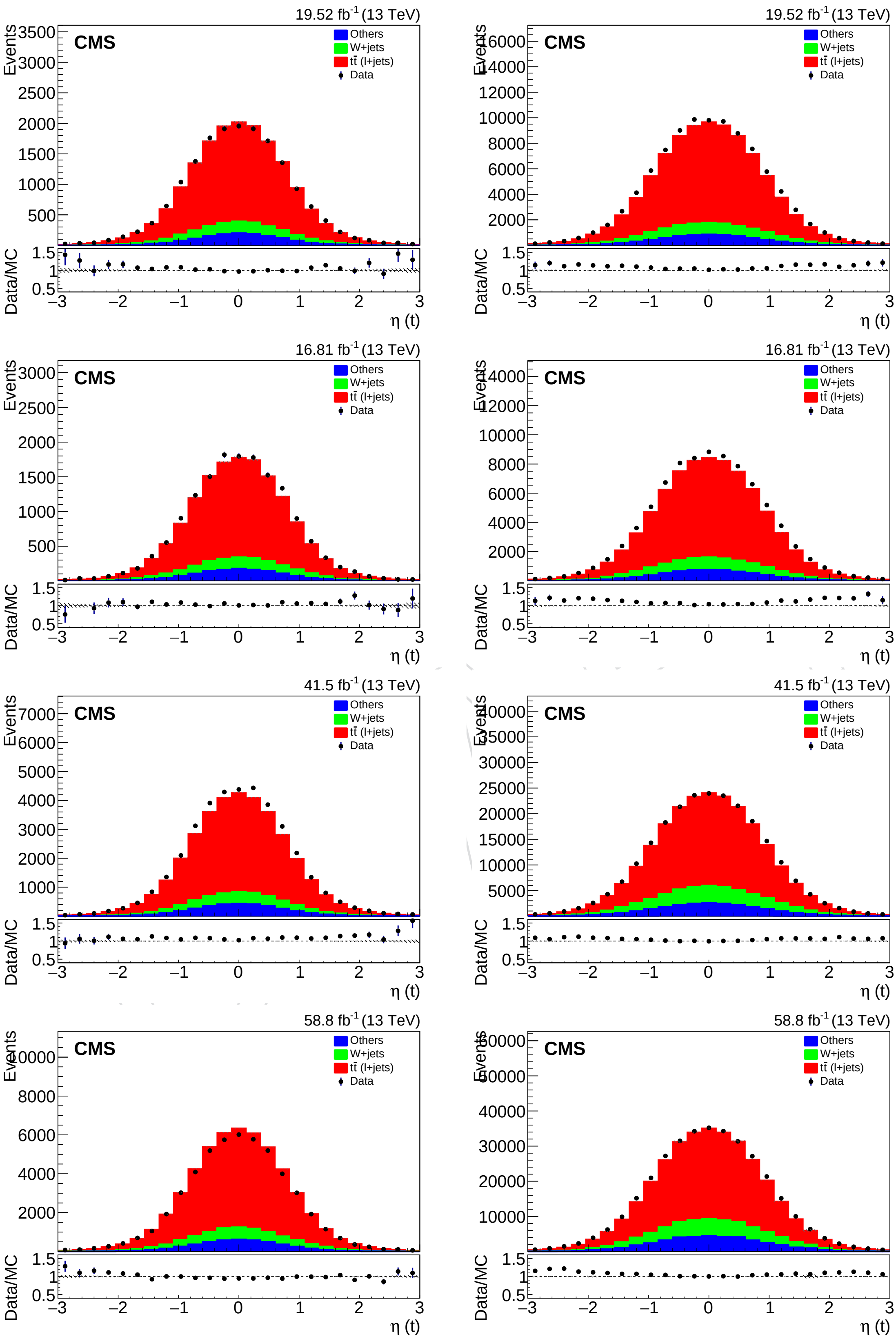

Figure 79: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

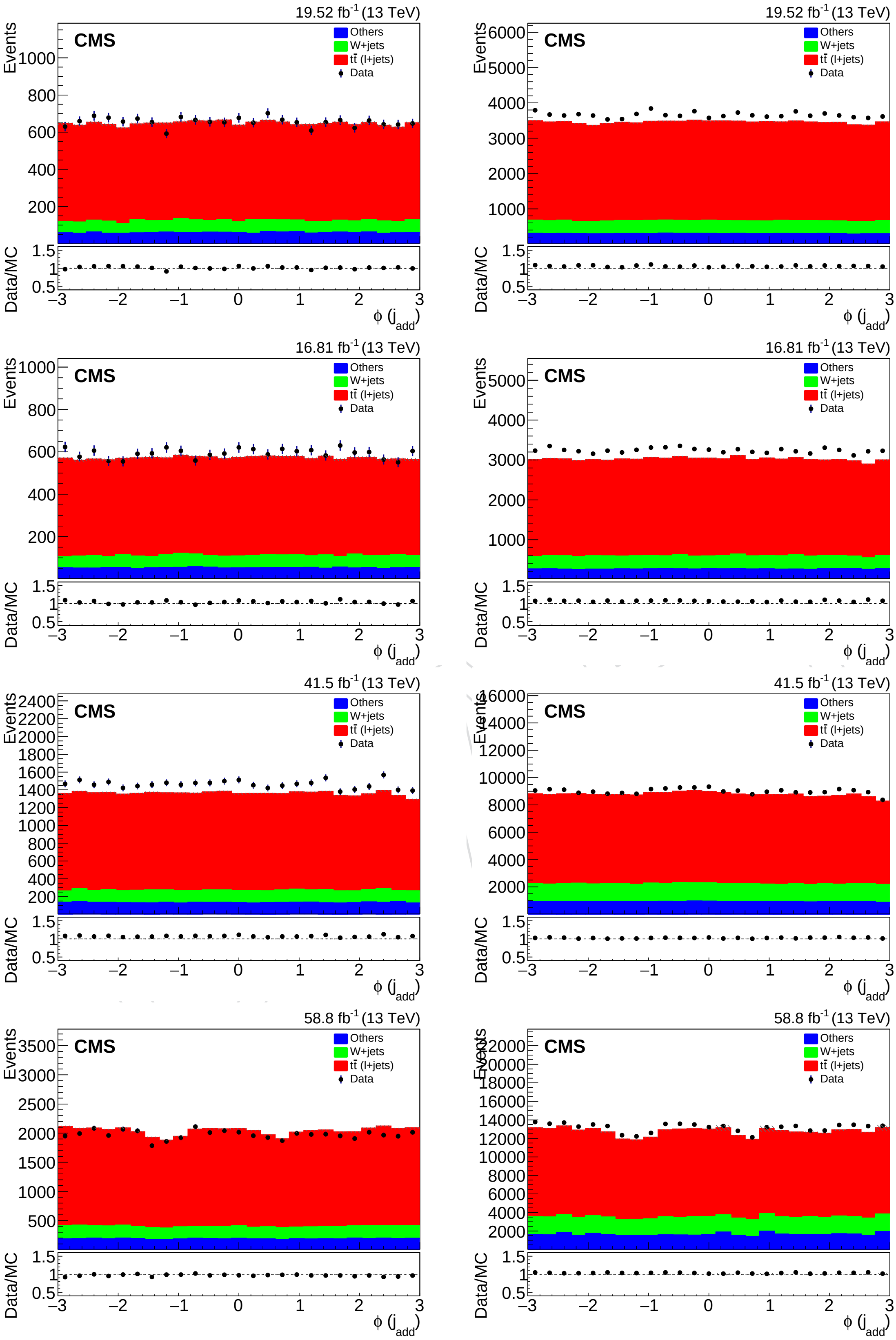

Figure 80: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

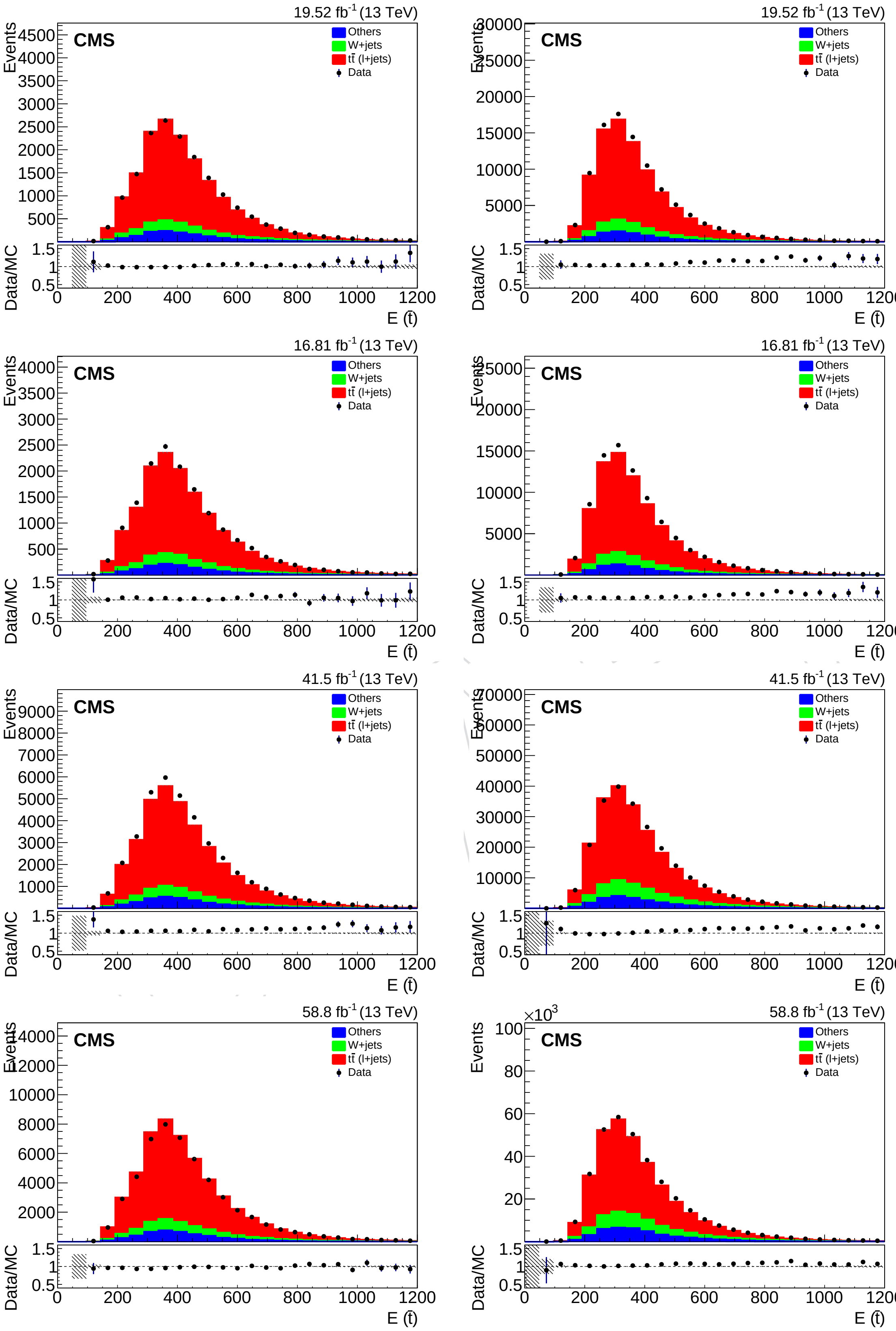

Figure 81: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

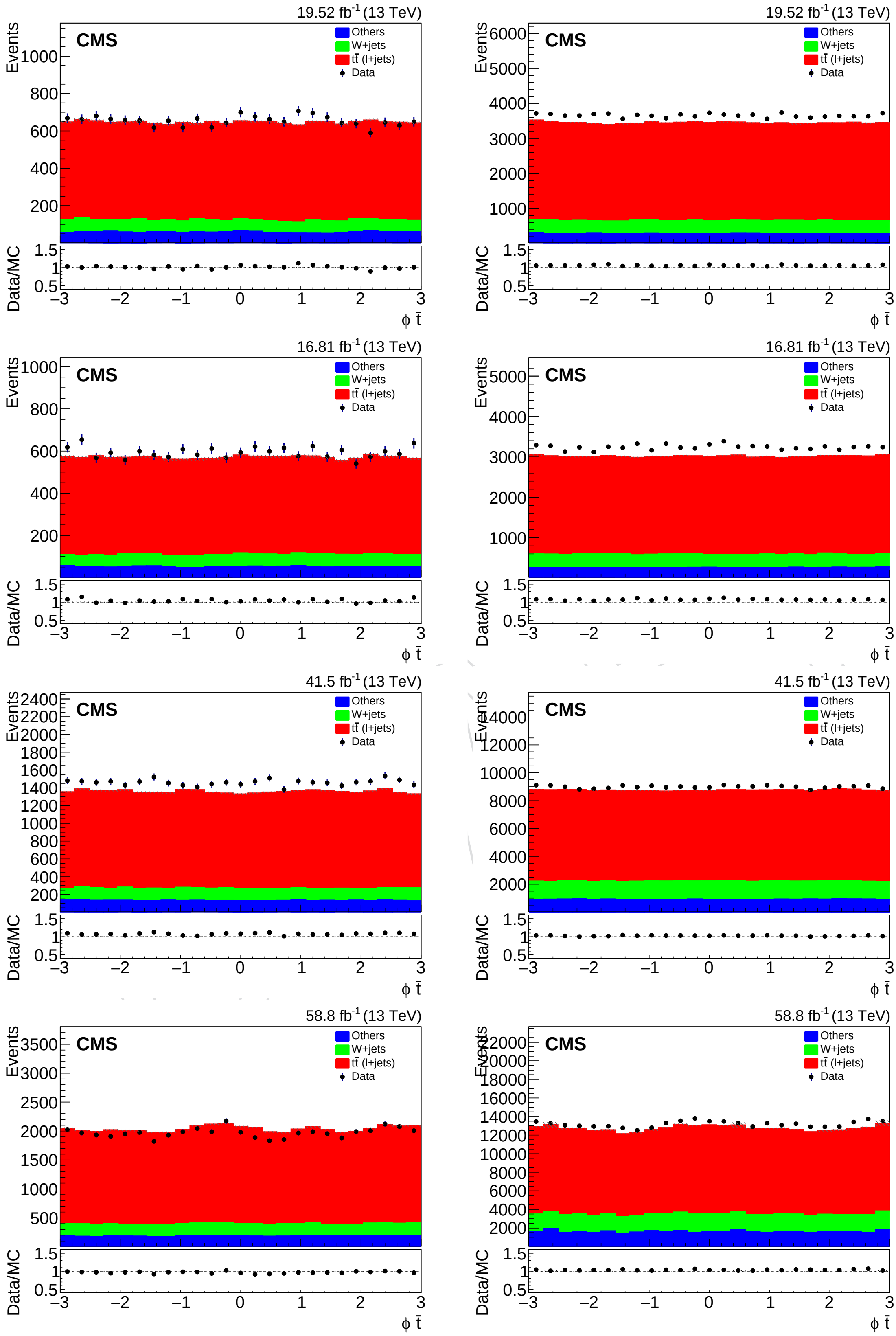

Figure 82: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

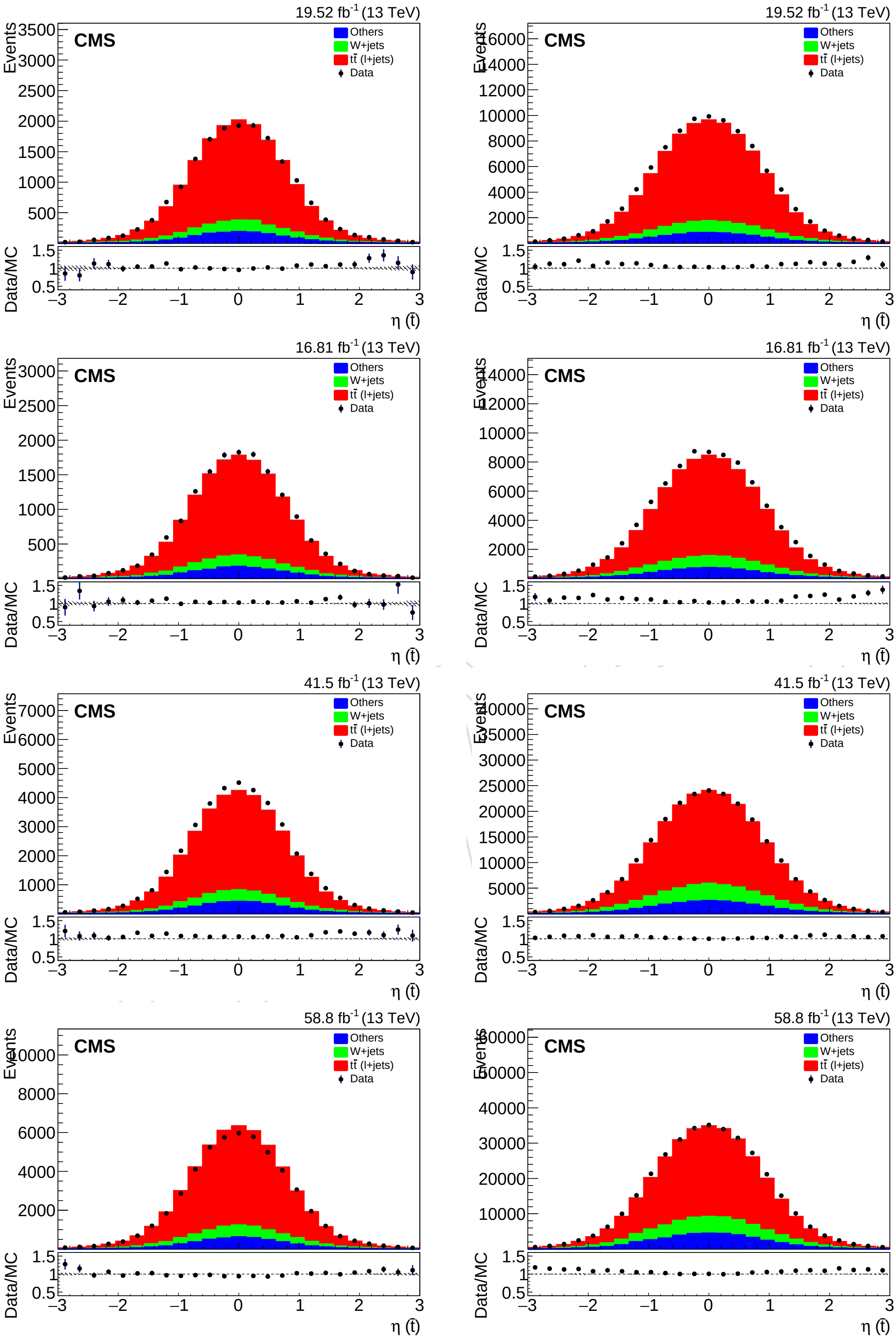

Figure 83: Heavy flavor (LF) b-tagging efficiencies in p_T and η bins for the $\mu+$ jets for 2018 (top), 2017 (middle) and 2016postVFP and 2016preVFP (bottom).

References

[1] S. Berge and S. Westhoff, "Top-quark charge asymmetry goes forward: two new observables for hadron colliders", *Journal of High Energy Physics* **2013** (Jul, 2013) 179, doi:10.1007/JHEP07(2013)179.

[2] T. C. Collaboration. https://github.com/cms-sw/cmssw.

[3] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing nlo calculations in shower monte carlo programs: the powheg box", *Journal of High Energy Physics* **2010** (Jun, 2010) 43, doi:10.1007/JHEP06(2010)043.

[4] R. Frederix and S. Frixione, "Merging meets matching in mc@nlo", *Journal of High Energy Physics* **2012** (Dec, 2012) 61, doi:10.1007/JHEP12(2012)061.

[5] A. M. S. et al., "Extraction and validation of a new set of cms pythia8 tunes from underlying-event measurements", *The European Physical Journal C* **80** (Jan, 2020) 4, doi:10.1140/epjc/s10052-019-7499-4.

[6] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/viewauth/cms/l1prefiringweightrecipe",.

[7] J. M. L. et al, "Precise predictions for v+jets dark matter backgrounds", *The European Physical Journal C* **77** (Dec, 2017) 829, doi:10.1140/epjc/s10052-017-5389-1.

[8] T. C. Collaboration, "Particle-flow event reconstruction in cms and performance for jets, taus, and met",.

[9] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/cms/missingetoptionalfiltersrun2",.

[10] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/cms/muonul2016",.

[11] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/cms/muonul2017",.

[12] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/cms/muonul2018",.

[13] T. C. Collaboration, "https://gitlab.cern.ch/cms-muonPOG/muonefficiencies/-/tree/master/Run2/UL/2018/2018_trigger",.

[14] T. C. Collaboration, "https://gitlab.cern.ch/cms-muonPOG/muonefficiencies/-/tree/master/Run2/UL/2017/2017_trigger",.

[15] T. C. Collaboration, "https://gitlab.cern.ch/cms-muonPOG/muonefficiencies/-/tree/master/Run2/UL/2016_postVFP/2016_postVFP_trigger",.

[16] T. C. Collaboration, "https://gitlab.cern.ch/cms-muonPOG/muonefficiencies/-/tree/master/Run2/UL/2016_preVFP/2016_preVFP_trigger",.

[17] T. C. Collaboration, " https://twiki.cern.ch/twiki/bin/view/CMS/EgHLTScaleFactorMeasurements.",.

[18] T. C. Collaboration, "Pileup-per-particle identification: optimisation for Run 2 Legacy and beyond",.

[19] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECDataM",.

[20] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID13TeVUL",.

[21] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution",.

[22] E. Bols et al., "Jet flavour classification using deepjet", *Journal of Instrumentation* **15** (dec, 2020) P12012, doi:10.1088/1748-0221/15/12/P12012.

[23] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods",.

[24] A. S. et al., "Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques", *Journal of Instrumentation* **15** (jun, 2020) P06005, doi:10.1088/1748-0221/15/06/P06005.

[25] C. Matthies, "https://indico.cern.ch/event/1240457/#2-top-tagging-sf-ul16ul17ul18",.

[26] T. C. Collaboration, "https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/",.

[27] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/CMS/ScrecUnfolding#Method_selection",.

[28] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM#LumiComb",.

[29] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG#UserRecommendations",.

[30] T. C. Collaboration, "https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018.",.

[31] T. C. Collaboration, "https://cms-jerc.web.cern.ch/Recommendations/#uncertainites-and-correlations",.

[32] R. Barlow and C. Beeston, "Fitting using finite monte carlo samples", *Computer Physics Communications* **77** (1993), no. 2, 219–228, doi:https://doi.org/10.1016/0010-4655(93)90005-W.

[33] V. Guglielmi, "Machine learning approaches for parameter reweighting in monte-carlo samples of top quark production in cms", 2022.

[34] T. C. Collaboration, "https://gitlab.cern.ch/cms-muonPOG/spark_tnp",.