Chapter 5

Measurement of the Differential Inclusive Multijet Cross-sections and their Ratio

In a proton-proton collision, the inclusive jet cross-section studied as a function of jet properties, provides essential information about the parton distribution functions of the proton and the strong coupling constant. This chapter describes the measurement of differential inclusive multijet event cross-sections and the cross-section ratio. The event and jet selections, trigger studies, spectrum construction, corrections applied and calculation of the experimental uncertainties are discussed in detail.

The differential inclusive multijet event cross-sections, given by Eq. 5.1, are studied as a function of the average transverse momentum, $H_{\mathrm{T},2}/2 = \frac{1}{2}(p_{\mathrm{T},1} + p_{\mathrm{T},2})$, where $p_{\mathrm{T},1}$ and $p_{\mathrm{T},2}$ denote the transverse momenta of the two leading jets.

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\big(H_{\mathrm{T},2}/2\big)} = \frac{1}{\epsilon\,\mathcal{L}_{\mathrm{int,eff}}}\frac{N_{\mathrm{event}}}{\Delta\big(H_{\mathrm{T},2}/2\big)} \tag{5.1}$$

where N_{event} is the number of inclusive n-jet events counted in an $H_{\mathrm{T},2}/2$ bin, ϵ

is the product of the trigger and jet selection efficiencies, which are greater than 99%, $\mathcal{L}_{\text{int,eff}}$ is the effective integrated luminosity, and $\Delta\left(H_{\text{T},2}/2\right)$ are the bin widths which increase with $H_{\text{T},2}/2$ and are proportional to the $H_{\text{T},2}/2$ resolution. The measurements are reported in units of (pb/GeV). The inclusive n-jet event samples include the events with number of jets $\geq n$. In the present thesis, the measurements are performed for $n = 2$ giving inclusive 2-jet events ($n_j \geq 2$) and for $n = 3$ giving inclusive 3-jet events ($n_j \geq 3$). The cross-section ratio R_{32}, defined in Eq. 5.2 is obtained by dividing the differential cross-sections of inclusive 3-jet events to that of inclusive 2-jet one, for each bin in $H_{\text{T},2}/2$.

$$R_{32} = \frac{\dfrac{d\sigma_{3\text{-jet}}}{d\left(H_{\text{T},2}/2\right)}}{\dfrac{d\sigma_{2\text{-jet}}}{d\left(H_{\text{T},2}/2\right)}} \tag{5.2}$$

For inclusive 2-jet events sufficient data are available up to $H_{\text{T},2}/2 < 2000$ GeV, while for inclusive 3-jet events and the ratio R_{32}, the accessible range is limited to $H_{\text{T},2}/2 < 1680$ GeV.

5.1   Data Samples

This measurement uses the data collected at the center-of-mass energy of 8 TeV by the CMS experiment in the 2012 run period of the LHC. The 2012 data taking is divided in four periods A, B, C, D and the data sets are split into samples according to the run period. Further, each sample is grouped into subsets based on the trigger decision. For runs B-D, the `JetMon` stream data sets contain prescaled low trigger threshold paths (HLTPFJet40, 80, 140, 200 and 260) while the `JetHT` stream data sets contain unprescaled high threshold trigger paths (HLT PFJet320 and 400). For run A, the `Jet` stream contains all the above mentioned trigger paths. The data to be

used in physics analysis must satisfy the validation requirements of the data quality monitoring procedure. CMS uses JSON (Java Script Object Notation) format files to store the range of good luminosity within a run. In the current analysis, the applied certification file[19] is based on the final event reconstruction of the 2012 data sets. The data sets used in the current study are listed in Table 5.1 along with the luminosity of each data set which increases with period. The full 2012 data sample corresponds to an integrated luminosity of 19.71 fb^{-1}.

Table 5.1: Run range and luminosity of the proton-proton collisions data collected at the center-of-mass energy of 8 TeV by the CMS experiment in the year of 2012 in different run periods.

Run	Run range	Data set	Luminosity fb^{-1}
A	190456-193621	/Jet/Run2012A-22Jan2013-v1/AOD	0.88
B	193834-196531	/Jet[Mon,HT]/Run2012B-22Jan2013-v1/AOD	4.41
C	198022-203742	/Jet[Mon,HT]/Run2012C-22Jan2013-v1/AOD	7.06
D	203777-208686	/Jet[Mon,HT]/Run2012D-22Jan2013-v1/AOD	7.37

5.1.1 Monte Carlo Samples

To have a comparison of results obtained from the data events with those from simulated events, the MADGRAPH5 Monte Carlo (MC) event generator has been used. It has been interfaced to PYTHIA6 by the LHE event record, to generate the rest of the higher-order effects using the Parton Showering (PS) model, with tune Z2* to model the underlying event. The MC samples are processed through the complete CMS detector simulation to allow studies of the detector response and comparison to measured data on detector level.

The cross-section measured as a function of the transverse momentum p_T or the scalar sum of the transverse momentum of all jets H_T, falls steeply with increasing p_T. To produce sufficient number of high p_T events, the events are generated

---

[19]Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON

in phase-space regions binned in H_T or the leading jet p_T. Later on, the different phase-space regions are added together in the data analyses by taking into account the cross-section of the respective phase-space regions. The official CMS MAD-GRAPH5+PYTHIA6 (MG5+P6) MC samples used in this analysis are generated as slices in the H_T phase-space, as tabulated in Table 5.2 along with their cross-sections and number of events generated.

Table 5.2: The official Monte Carlo samples are produced in phase space slices in H_T with the generator MADGRAPH5 and interfaced to PYTHIA6 for the parton shower and hadronization of the events. The cross-section and number of events generated are mentioned for each sample.

Generator	Sample	Events	Cross-section pb
MADGRAPH5 + PYTHIA6	/QCD_HT-100To250_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	50129518	1.036×10^7
	/QCD_HT-250To500_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	27062078	2.760×10^5
	/QCD_HT-500To1000_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	30599292	8.426×10^3
	/QCD_HT-1000ToInf_TuneZ2star_8TeV-madgraph-pythia6/ Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM	13843863	2.040×10^2

5.2   Event Selection

The events are selected according to quality criteria which ensure the high purity and high selection efficiency of the sample to be studied. This event selection also reduces beam induced background, detector-level noise and jets arising from non-physical calorimeter energy measurements.

5.2.1   Trigger Selection

CMS implements a two-level trigger system to reduce the amount of recorded events to a sustainable rate. In this analysis the jets are the final objects of interest. So single jet trigger paths with varying thresholds are used to select events in the data. It consists of one L1 trigger seed and multiple HLT filters. The L1 jet trigger uses

transverse energy sums computed by using both HCAL and ECAL in the central region ($|\eta| < 3.0$) or HF in the forward region ($|\eta| > 3.0$). The single jet triggers (HLT_PFJetX), same as the ones used for other CMS 8 TeV measurements [25, 99], are used in the current study and are tabulated in Table 5.3. A single jet trigger selects an event in which at least one jet has the transverse momentum above the threshold. HLT_PFJetX implies that there is at least one jet in the event, whose $p_T > X$ (GeV). The L1 trigger has a lower threshold to ensure full efficiency versus p_T of the HLT trigger. The p_T spectrum is steeply falling and hence the rates for low-p_T jets are very high. So it is not feasible to use a single unprescaled trigger for the selection of all required events. To collect sufficient data in the lower part of the p_T spectrum, five prescaled low-p_T trigger paths, each with different prescale value, are used. Also, one unprescaled trigger i.e. HLT_Jet320 is used in the high p_T region, in which the rate is sufficiently small to collect and store all events. During the reconstruction of the spectrum, the prescales have been taken into the account.

Table 5.3: The single jet HLT trigger paths used in the analysis. The column $H_{T,2}/2$, 99% indicates the value of $H_{T,2}/2$ at which each trigger exhibits an efficiency larger than 99%. The last column gives the effective luminosity seen by each trigger which when divided by the total integrated luminosity of 19.71 fb^{-1}, gives the effective prescale applied on a trigger over the whole run period.

Trigger Path	L1 threshold GeV	HLT threshold GeV	$H_{T,2}/2$, 99% GeV	Eff. Lumi fb^{-1}
HLT_PFJet80	36	80	120.0	0.0021
HLT_PFJet140	68	140	187.5	0.056
HLT_PFJet200	92	200	262.5	0.26
HLT_PFJet260	128	260	345.0	1.06
HLT_PFJet320	128	320	405.0	19.71

The efficiency of each trigger path as a function of $H_{T,2}/2$ is described by the turn-on curves with a rising part where the trigger is partly inefficient, until a plateau region where the trigger is fully efficient. Hence it is important to determine the threshold above which a trigger becomes fully efficient. The threshold is the value at which the trigger efficiency exceeds 99%. The trigger efficiency for HLT_PFJetY

is given by Eq. 5.3 where HLT_PFJetX is the reference trigger and is assumed to be fully efficient in the considered phase space region. The value of X is chosen previous to that of Y in p_{T} ordering from the trigger list so that the higher trigger condition can be emulated from the lower trigger path.

$$\epsilon_{\mathrm{HLT\_PFJetY}} = \frac{H_{\mathrm{T},2}/2\Big(\mathrm{HLT\_PFJetX} + (\mathrm{L1Object\_p_T} > Z) + (\mathrm{HLTObject\_p_T} > Y)\Big)}{H_{\mathrm{T},2}/2(\mathrm{HLT\_PFJetX})}$$

$$(5.3)$$

where Y is the p_{T} threshold of HLT_PFJetY and Z is the L1 seed value corresponding to the trigger path HLT_PFJetY. The denominator represents the number of events for which the reference trigger path HLT_PFJetX has been fired. The numerator is the number of events for which HLT_PFJetX has been fired along the p_{T} of L1Object $\geq$ Z and the p_{T} of HLTObject $\geq$ Y. For example, to obtain turn-on curve for HLT_PFJet260, HLT_PFJet200 is the reference HLT path. The p_{T} cut on L1Object is 128 GeV and p_{T} cut on HLTObject is 260 GeV. The threshold point at which the trigger efficiency is larger than 99% is determined by fitting the turn-on distribution with a sigmoid function described in Eq. 5.4. The trigger turn-on curves as a function of $H_{\mathrm{T},2}/2$ can been seen in Fig. 5.1 which are described by a sigmoid function (blue line). The error bars give the uncertainty on the efficiency which is calculated using Clopper-Pearson confidence intervals [100].

$$f_{fit}(x) = \frac{1}{2}\left(1 + \mathrm{erf}\Big(\frac{x-\mu}{\sqrt{2}\sigma}\Big)\right) \qquad (5.4)$$

5.2.2   Primary Vertex Selection

The reconstructed tracks, number of strip and pixel hits and the normalized track χ^2 identify the primary vertex (PV). The tracks are clustered according to the
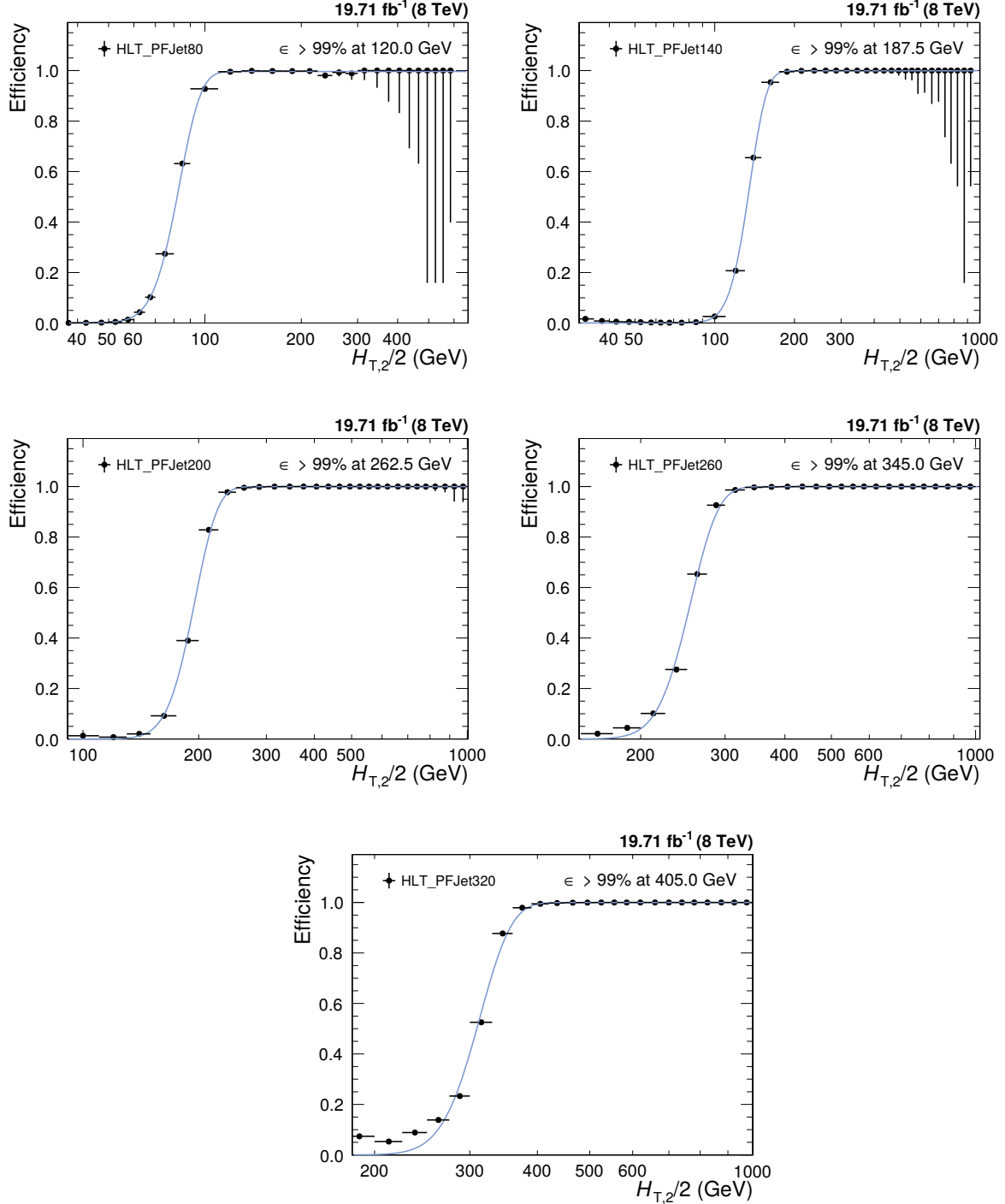
Figure 5.1: Trigger efficiencies turn-on curves for the single jet HLT trigger paths are fitted with a sigmoid function (blue line) to obtain the 99% efficiency threshold. The error bars give the uncertainty on the efficiency which is calculated using Clopper-Pearson confidence intervals [100].

z-coordinate of their point of closest approach to the beam axis. The following selection criteria for the primary vertex are imposed to identify and reject beam background events. At least one good primary vertex reconstructed from at least four tracks within a distance of $\left|z(PV)\right| < 24$ cm to the nominal interaction point in a collision, is required in each event. The radial distance in x-y plane, ρ(PV) should not be greater than 2 cm. The number of degrees of freedom in fitting for the position of each vertex using its associated tracks should be at least four in number.

5.2.3   Missing Transverse Energy

In an ideal detector where all particles could be identified and perfectly measured, the transverse momentum of all particles would sum up to zero. But neutral weakly interacting particles, such as neutrinos, escape from typical collider detectors and do not produce any direct response in the detector. The imbalance of total momentum of all visible particles can give hints of the presence of such particles. The vector momentum imbalance in the plane perpendicular to the beam direction is known as missing transverse momentum or energy (E_T^{miss}). It is one of the most important observables for discriminating leptonic decays of W bosons and top quarks from background events which do not contain high-energetic neutrinos, such as multijet and Drell-Yan events.

The ratio of missing transverse energy to the total transverse energy $E_T^{\mathrm{miss}}/\sum E_T$, shown in Fig. 5.2 for $n_j \geq 2$ (left) and $n_j \geq 3$ (right) events, shows a discrepancy between the data (black solid circles) and simulated MC (blue histogram), at the tail part of the distribution. This is because of a finite contribution from $Z(\rightarrow \nu\bar{\nu})$ + jet events which gives rise to non-zero E_T in the events in the data. Such events are absent in QCD simulated events in MC. Hence $E_T^{\mathrm{miss}}/\sum E_T$ is required to be less than 0.3 to reject events with high E_T^{miss}.
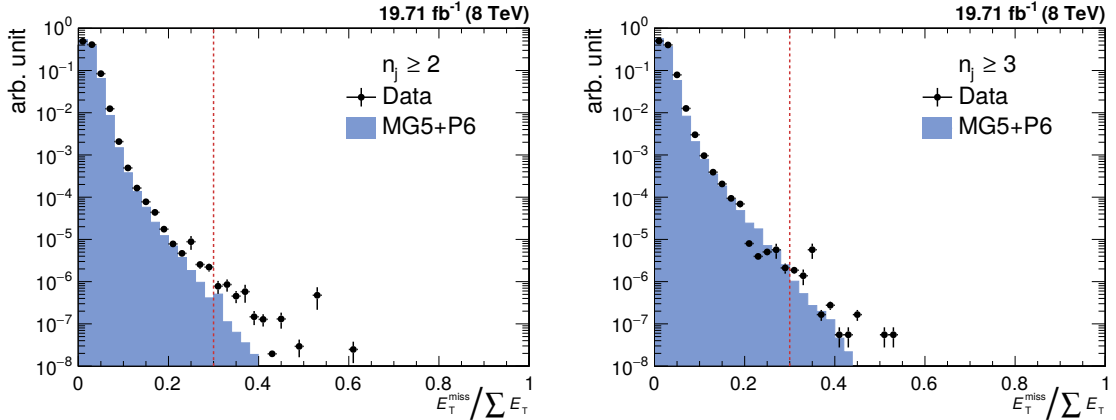
Figure 5.2: Missing transverse energy fraction of the total transverse energy per event in the data (black solid circles) and simulated Monte Carlo events (blue histogram) in inclusive 2-jet (left) and 3-jet (right) events. To remove background and noise, events with a fraction exceeding a certain threshold, here indicated with the red dashed line, are rejected.

5.2.4 Jet Identification

In order to suppress non-physical jets, arising from detector noise or mis-reconstructed particles, jet identification criteria (ID) have been applied to each jet of an event. The algorithm works on reconstructed jets using information of the clustered particle candidates. The official tight jet ID [101], recommended by JETMET group [102] is used. Due to pileup and electronic noise the jet constituent fractions may vary from event to event. In order to reject the noisy jets, jet selection criteria are optimized to select only good quality jets. The selection criteria are implemented as selection cut on jet fractions. Table 5.4 summarizes the properties of the reconstructed jets and their respective cuts. Each jet should contain at least two particles, one of which should be a charged hadron. The cut on the fraction of neutral hadrons and photons removes HCAL noise and ECAL noise, respectively. Muons that are falsely identified and clustered as jets are removed by the muon fraction criterion. Based on information of the tracker, additional selection cuts are enforced in the region $|\eta| < 2.4$. The charged electromagnetic fraction-cut removes the jets clustered from misidentified electrons. Furthermore, the fraction of charged hadrons in the jet must be larger than zero and jets without any charged hadrons

are very likely to be pileup jets. The Figs. 5.3 and 5.4 show the distributions of the jet constituents observed in the data (black solid circles) and simulated MC events (blue histogram) for $n_j \geq 2$ and $n_j \geq 3$, respectively.

Table 5.4: The jet identification criteria (ID) removes noise and non-physical jets based on the properties of the reconstructed jets and the clustered particle candidates. All the tight ID selection cuts, recommended by the JETMET group are applied [102].

	Property	Tight ID cut
Whole η region	neutral hadron fraction	< 0.90
	neutral EM fraction	< 0.90
	number of constituents	> 1
	muon fraction	< 0.80
only $|\eta| < 2.4$	charged hadron fraction	> 0
	charged multiplicity	> 0
	charged EM fraction	< 0.90

5.2.4.1   Jet ID Efficiency

The efficiency of the jet ID as a function of $H_{T,2}/2$ is studied using a tag-and-probe technique with dijet events. The two leading jets are required to be back-to-back in the azimuthal plane such that $|\Delta\phi - \pi| < 0.3$. One of the dijets is selected randomly as a "tag" jet which is required to fulfill the tight jet ID criteria. The other jet is called "probe" jet for which it is examined, whether it also passes the tight jet ID. The ID efficiency is defined as the ratio of events where the probe jet passes the ID requirements, over the total number of dijet events. It is shown as function of $H_{T,2}/2$ in Fig. 5.5 and as expected, it is always greater than 99%. The QCD cross-section decreases as a function of $H_{T,2}/2$ and hence the number of events decrease when moving to higher $H_{T,2}/2$. Consequently the statistical fluctuations for ID efficiency are larger at higher $H_{T,2}/2$.
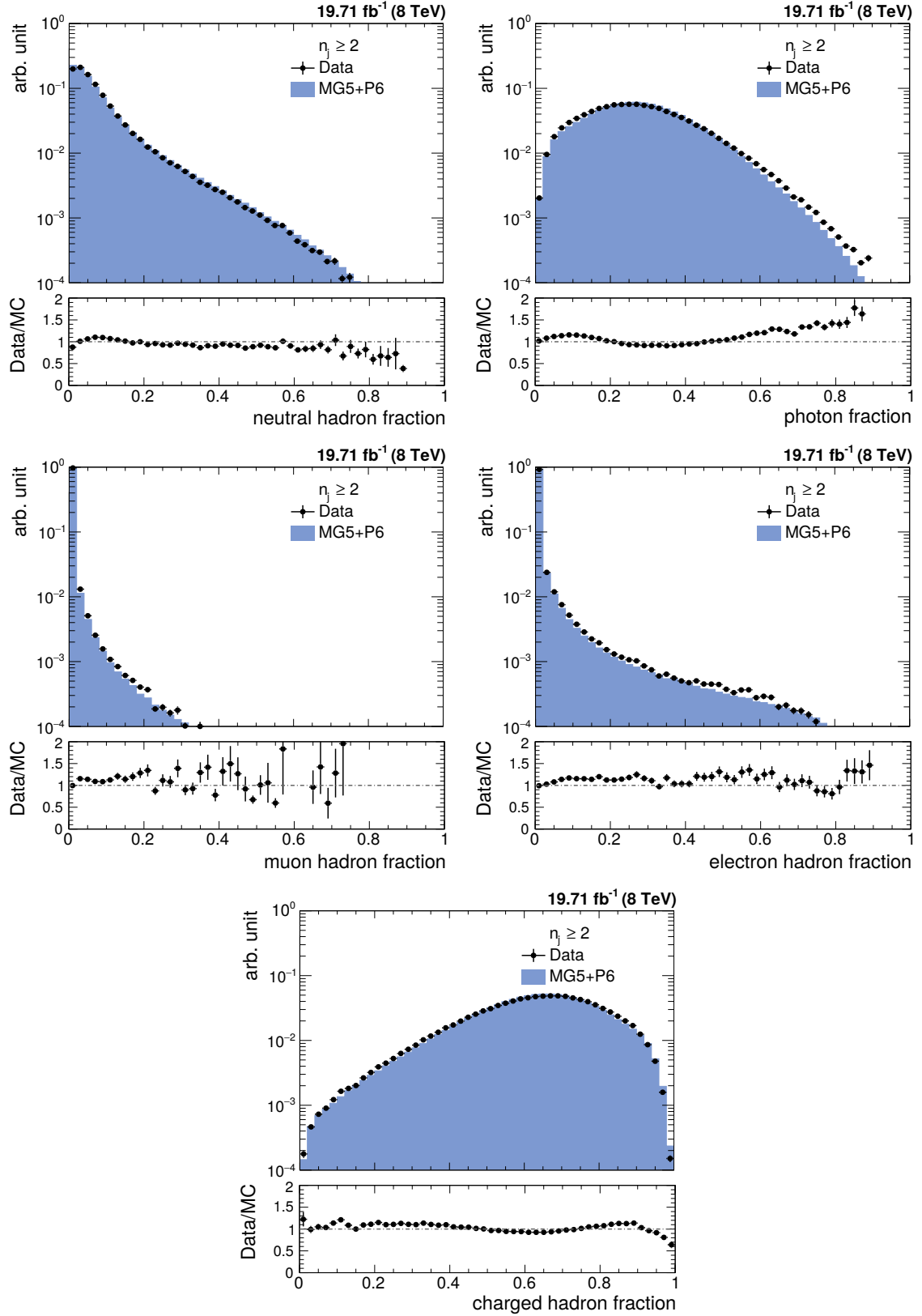
Figure 5.3: The fractions of jet constituents as observed in the data (black solid circles) and simulated Monte Carlo events (blue histogram) for different types of PF candidates for inclusive 2-jet events. The data and simulations are normalized to the same number of events. The distributions are shown after the application of the jet ID.
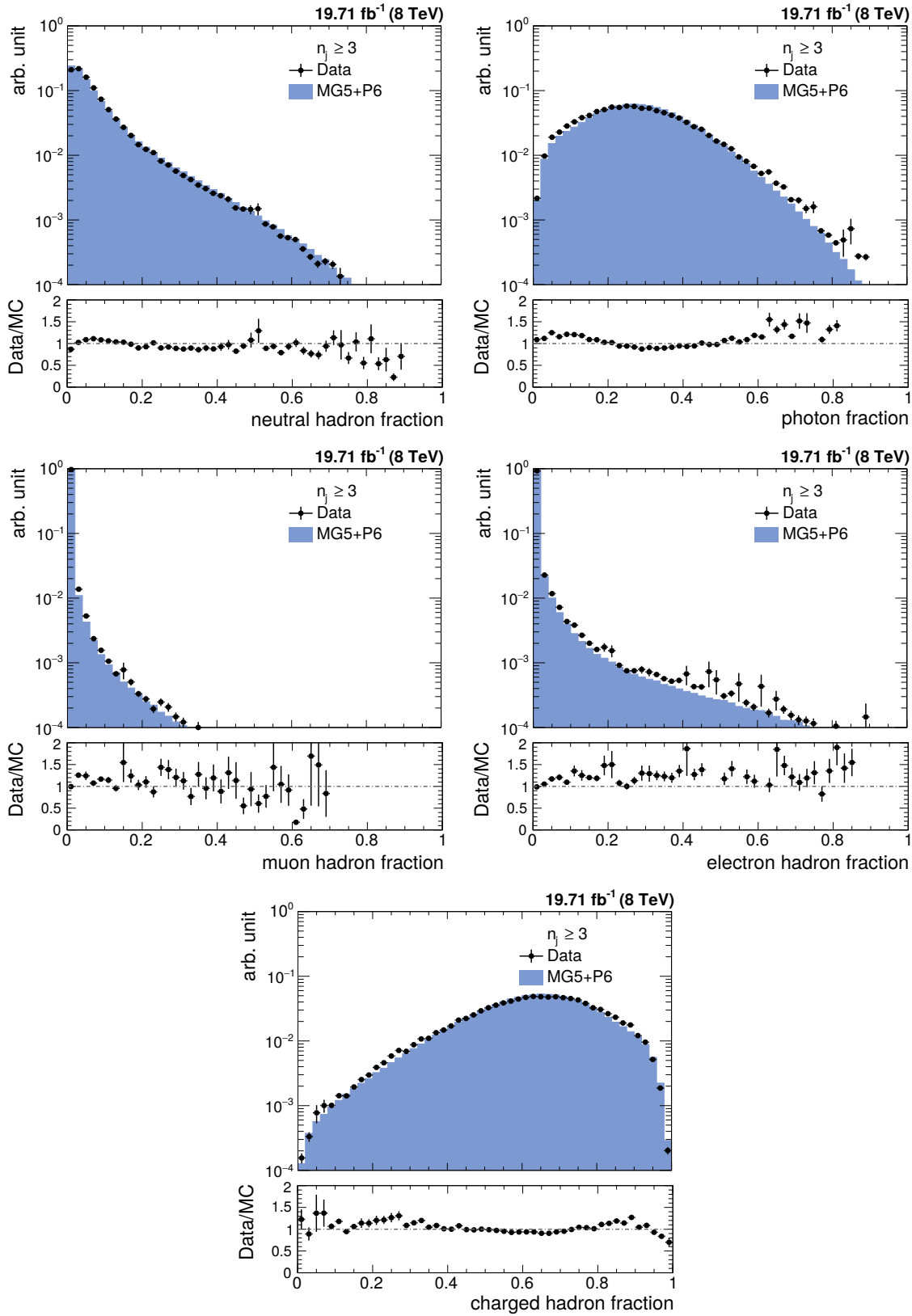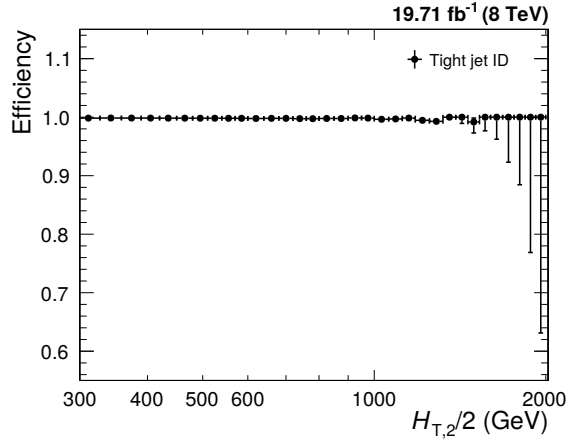
Figure 5.4: The fractions of jet constituents as observed in the data (black solid circles) and simulated Monte Carlo events (blue histogram) for different types of PF candidates for inclusive 3-jet events. The data and simulations are normalized to the same number of events. The distributions are shown after the application of the jet ID.

Figure 5.5: The jet ID efficiency is studied as a function of $H_{T,2}/2$ with tag-and-probe technique using dijet event topologies and it always exceeds 99%.

5.2.5 Jet Selection

The measurement of differential cross-sections and their ratio uses jets clustered from particle flow candidates using the anti-k_t jet algorithm with a size parameter, $R = 0.7$. The energy scale of the jets is corrected with the CMS recommended jet energy corrections, described in Sec. 4.3.2. These corrections are applied to jets in both the data[20] as well as in simulated events[21]. As a convention, the jets in each event are ordered in decreasing p_T, with the first (leading) jet being the jet with highest p_T. The jet selection, based on phase space cuts on transverse momentum and rapidity of jets in an event, is as follows :

- All jets having $p_T > 150$ GeV and $|y| < 5.0$ are selected.

- Events with at least two jets are selected.

- The two leading jets should have $|y| < 2.5$ and further jets are counted only, if they lie within the same central rapidity range of $|y| < 2.5$.

These cuts assure high detector acceptance and good comparability to pQCD.

---

[20]Winter14_V8 jet energy corrections
[21]START53_V27 jet energy corrections

For a consistent comparison, the exactly same selection criteria are applied in the data and simulated events as well as in theoretical calculations.

5.3   Comparison with Simulation

5.3.1   Pileup Reweighting

While generating the official MC samples, the number of pileup interactions describing the conditions expected for each data-taking period are taken into account. But the number of pileup events implemented in the simulation $N_{\mathrm{MC}}(N_{\mathrm{PU,truth}})$, does not match exactly the one measured in the data $N_{\mathrm{data}}(N_{\mathrm{PU,est.}})$. To match the pileup distributions in the data, a reweighting factor w_{PU}, as given by Eq. 5.5 is applied to the simulated events. In Fig. 5.6 the number of reconstructed vertices are shown before (left) and after pileup reweighting (right). It is observed that before pileup reweighting there was a significant mismatch of the pileup distributions in the data (black solid circles) and simulated MC events (blue histogram), which completely vanishes after reweighting.

$$w_{\mathrm{PU}} = \frac{N_{\mathrm{data}}(N_{\mathrm{PU,est.}})/\sum N_{\mathrm{data}}}{N_{\mathrm{MC}}(N_{\mathrm{PU,truth})}/\sum N_{\mathrm{MC}}} \tag{5.5}$$

5.3.2   Comparison of Cross-sections and their Ratio

The measured data distribution of differential cross-section at the detector level is compared to the predictions of MC simulation using MADGRAPH5 generator interfaced with PYTHIA6 (MG5+P6) including the detector simulation as well as to a fixed-order theory prediction obtained using CT10-NLO PDF set. Figure 5.7 shows the comparison of differential cross-section as a function of $H_{\mathrm{T,2}}/2$ for $n_{\mathrm{j}} \geq 2$ (left) and $n_{\mathrm{j}} \geq 3$ (right) events, for the data (black solid circles), MG5+P6 MC (red open
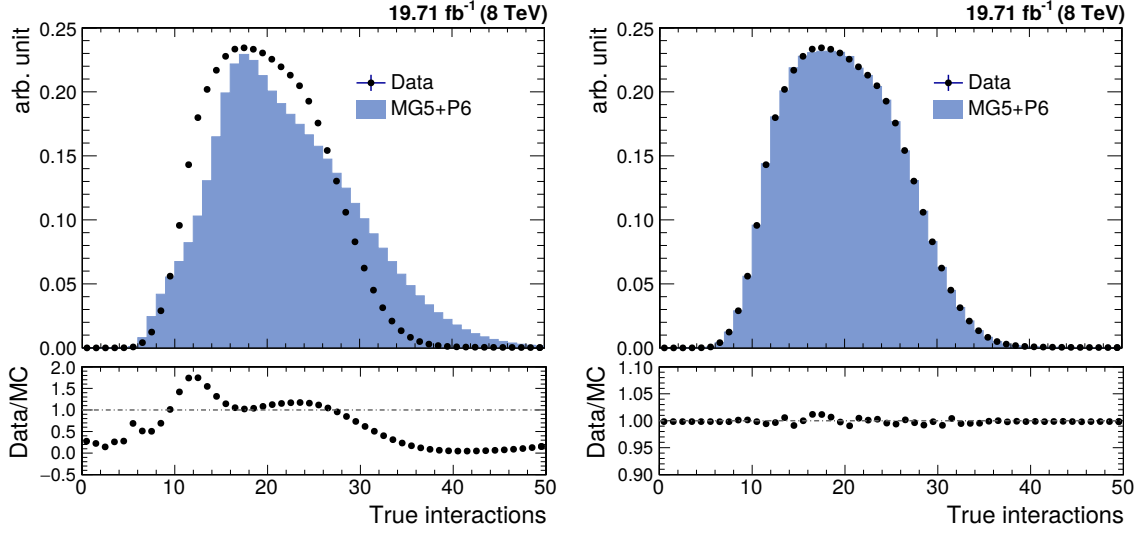
Figure 5.6: Number of reconstructed vertices in the data (black solid circles) and simulated Monte Carlo events (blue histogram) before (left) and after (right) the pileup reweighting.

circles) and CT10-NLO (blue histogram). The bottom panel in each plot shows the ratio of the data to the MC predictions (red line) as well as to the CT10-NLO theory predictions (blue line). The NLO predictions on parton level are not corrected for non-perturbative effects. Still the NLO predictions describe the data better as compared to the LO MC simulations which roughly describe the spectrum on detector level. Sufficient data for $n_j \geq 2$ and $n_j \geq 3$ events are available up to $H_{T,2}/2 < 2000$ GeV and 1680 GeV, respectively. Due to kinematical constraints, the minimum cut on $H_{T,2}/2$ is 300 GeV (explained in Sec. 6.1.1). Hence the differential cross-sections are studied in the range 300 GeV $< H_{T,2}/2 < 2000$ GeV for $n_j \geq 2$ and 300 GeV $< H_{T,2}/2 < 1680$ GeV for $n_j \geq 3$ events.

The ratio of differential cross-sections, R_{32} as a function of $H_{T,2}/2$, is extracted by dividing the cross-section of selected inclusive 3-jet events to that of inclusive 2-jet events at any given bin of $H_{T,2}/2$. In the cross-section ratios, the numerator and denominator are not independent samples. So to calculate the statistical uncertainty for the cross-section ratios at reconstructed level, the Wilson score interval method
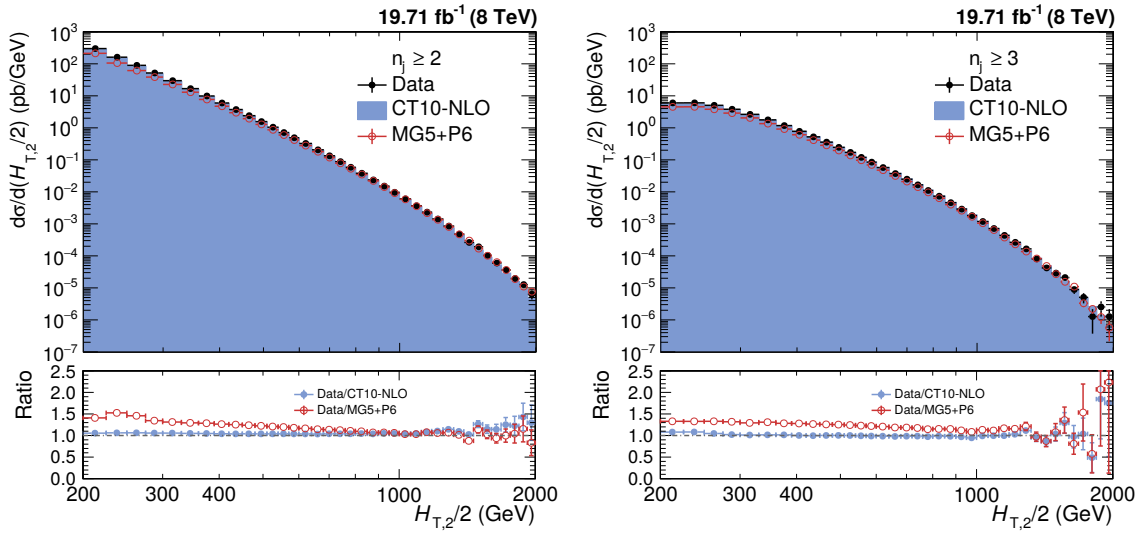
Figure 5.7: The reconstructed level differential cross-sections are compared for the data (black solid circles) and LO MADGRAPH5+PYTHIA6 (MG5+P6) Monte Carlo (red open circles) simulations with CT10-NLO theory predictions (blue histogram), as a function of $H_{T,2}/2$ for inclusive 2-jet (left) and 3-jet (right) events. Ratios of the data to the Monte Carlo predictions (red line) as well as to the CT10-NLO predictions (blue line) are shown in bottom panel of each plot.
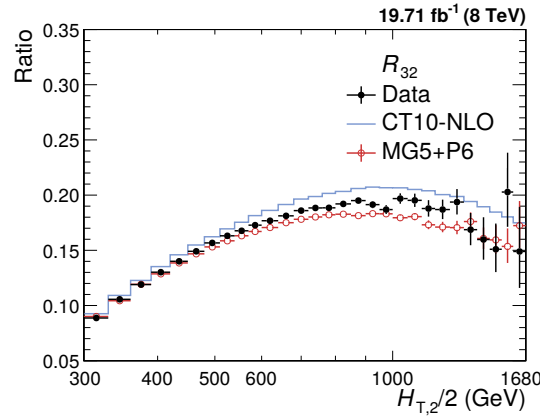

Figure 5.8: Comparison of the reconstructed level cross-section ratio R_{32} as a function of $H_{T,2}/2$, for the data (black solid circles) and LO MADGRAPH5+PYTHIA6 (MG5+P6) Monte Carlo (red open circles) with CT10-NLO theory predictions (blue line). The error bars give the asymmetrical statistical uncertainty, calculated by the Wilson score interval method which takes into the account the correlation between the numerator and denominator.

is used which takes into account the correlation between the numerator and the denominator and gives asymmetric errors. Figure 5.8 shows the comparison of the cross-section ratio R_{32} as a function of $H_{T,2}/2$, for the data (black solid circles) and

LO MADGRAPH5+PYTHIA6 (MG5+P6) MC (red open circles), at reconstructed level with CT10-NLO theory predictions (blue line). Since in events with $n_j \geq 3$, sufficient amount of data for differential cross-section is available only up to 1680 GeV of $H_{T,2}/2$, R_{32} is studied in the range 300 GeV $< H_{T,2}/2 <$ 1680 GeV. The bin-wise inclusive 2-jet and 3-jet events differential cross-sections as well as their ratio R_{32}, evaluated at detector level, along with statistical uncertainty (in %) are tabulated in Table A.1.

5.4 Jet Energy Resolution

In an ideal experiment, the value of a physical quantity would be determined exactly with an infinite precision, e.g. whenever a particle with energy E passes through an ideal calorimeter having infinite resolution, the measured energy should always be equal to E. But in real world, the measured energy of the above mentioned particle might differ from the value E. This difference of the measured quantity from its true value may be due to detector noise, uncertainties in the calibration, non-linearity of the response etc. Hence this results in the finite value of the resolution of the detector known as jet energy resolution (JER). In such a case, the measured values of energy of different particles, passing through the same detector with same energy E, will be different. Such measurements are described by a Gaussian distribution, centered around the true value of the measured quantity and its width is generally interpreted as detector resolution. Hence the importance of the detector resolution lies in the fact that it indicates how much the measured value of the observable differs from the true one i.e. how precisely a physical observable can be measured. The narrower the distribution, the higher the resolution is and hence more accurate the detector.

Due to finite resolution of the CMS detector, the measured transverse momenta of jets get smeared. Since the observable in this study i.e. $H_{T,2}/2$ is the average sum

of transverse momentum of the leading and sub-leading jets, the resolution of the detector has to be studied in terms of this observable. The CMS detector simulation based on MG5+P6 MC event generators is used to determine the resolution as both the particle and reconstructed level information is available. The jets clustered from stable generator particles called Gen jets as well as from particle flow candidates reconstructed from the simulated detector output called Reco jets, are used. The studies of the JETMET working group at CMS has shown that the jet energy resolution in the data is actually worse than in simulation [103]. So the reconstructed jet transverse momentum needs to be smeared additionally to match the resolution in the data. Table 5.5 shows the scaling factors which need to be applied on the transverse momenta of simulated reconstructed jets. The scaling factor depends on the absolute η of the jet and are provided by JETMET working group at CMS [103]. The uncertainty on these measured scaling factors ($c_{central}$) needs to be taken into account in a physics analysis. This is done by smearing the reconstructed jets with two additional sets of scaling factors, c_{up} and c_{down}, that correspond to varying the factors up and down respectively, by 1σ and evaluating the impact of these variations.

Table 5.5: JETMET working group at the CMS has shown that the jet energy resolution in the data is actually worse than in simulation [103]. To match the resolution in the data, the reconstructed jet transverse momentum in simulated events need to be smeared by applying the scale factors. The uncertainty on the resolution is given by an upwards and downwards variation c_{up} and c_{down} of the measured scaling factor $c_{central}$.

η	0.0 - 0.5	0.5 - 1.1	1.1 - 1.7	1.7 - 2.3	2.3 - 2.8
$c_{central}$	1.079	1.099	1.121	1.208	1.254
c_{down}	1.053	1.071	1.092	1.162	1.192
c_{up}	1.105	1.127	1.150	1.254	1.316

The reconstructed jet p_{T} is smeared randomly using a Gaussian function, $f(p_{\mathrm{T}})$

with a width widened by the scaling factor ($c_{central}$) :

$$f(p_\mathrm{T}) = a \times \exp\left(-\frac{1}{2}\left(\frac{p_\mathrm{T} - \mu}{\sigma}\right)^2\right) \tag{5.6}$$

where a is a constant, mean $\mu = 0$, width $\sigma = \sqrt{c_{central}^2 - 1} \cdot \mathrm{JER}(p_\mathrm{T}) \times p_\mathrm{T}$ and $\mathrm{JER}(p_\mathrm{T})$ is the resolution determined as a function of jet p_T using MG5+P6 MC simulated events. After smearing transverse momentum of each reconstructed jet, $H_{\mathrm{T},2}/2$ is calculated from both generator particle jets (Gen $H_{\mathrm{T},2}/2$) as well as the particle flow or reconstructed jets (Reco $H_{\mathrm{T},2}/2$). Then the response is calculated as defined in the Eq. 5.7.

$$R = \frac{\mathrm{Reco}\ H_{\mathrm{T},2}/2}{\mathrm{Gen}\ H_{\mathrm{T},2}/2} \tag{5.7}$$

The width of the response distribution in a given Gen $H_{\mathrm{T},2}/2$ bin is interpreted as the resolution which in good approximation can be described by 1σ of a Gaussian fit of the response distribution. A double-sided Crystal Ball function[22] takes into account the non-Gaussian tails of the jet response distribution. The resolution as a function of $H_{\mathrm{T},2}/2$ is calculated separately for both $n_\mathrm{j} \geq 2$ and $n_\mathrm{j} \geq 3$ events. A fit example for one Gen $H_{\mathrm{T},2}/2$ bin is shown in Fig. 5.9 for $n_\mathrm{j} \geq 2$ (left) and $n_\mathrm{j} \geq 3$ (right) events. Here the black dots represent the jet response distribution and the double-sided Crystal Ball fit (blue line) is overlayed by the Gaussian fit (red line). The resolution in each Gen $H_{\mathrm{T},2}/2$ bin is then plotted as a function of Gen $H_{\mathrm{T},2}/2$.

As expected, it has been observed from Fig. 5.10 that the Crystal Ball function (blue solid circles) describes the measured distributions better as compared to Gaussian function fit (red solid circles), especially in the low-$H_{\mathrm{T},2}/2$ region where
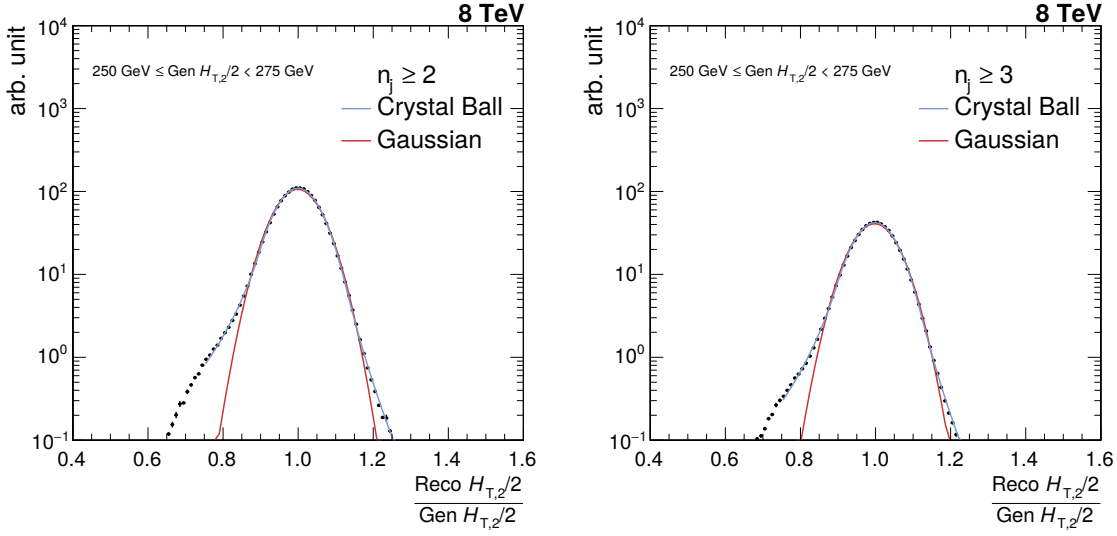
---

[22]See details in Sec. A.5

Figure 5.9: Fitting of the jet energy resolution distribution, obtained using LO MADGRAPH5+PYTHIA6 (MG5+P6) Monte Carlo simulated events, as a function of $H_{T,2}/2$ for inclusive 2-jet (left) and 3-jet (right) events. The blue line shows the double-sided Crystal Ball function fit of $\frac{\text{Reco } H_{T,2}/2}{\text{Gen } H_{T,2}/2}$ in each Gen $H_{T,2}/2$ bin, overlayed by Gaussian fitting the core of the resolution (red line).

the non-Gaussian tails are more pronounced. Hence JER is determined using Crystal Ball function fit. Figure 5.11 shows the final relative jet energy resolution (JER) which is described by a modified version of the NSC formula (blue solid line) [104], as mentioned in Eq. 5.8. To consider the migration to lower as well as higher bins and to obtain the resolution with reasonable statistics over the full range of Gen $H_{T,2}/2$, the fit function is extrapolated to 80 GeV and up to 2000 GeV, as shown by red dashed line. The fit formula used here is basically the usual NSC formula which describes the resolution in terms of noise N originating due to electronic and pileup noise, a stochastic component S due to sampling fluctuation and EM fraction fluctuation per hadrons and a constant term C due to presence of dead material, magnetic field and calorimeter cell to cell fluctuations. In the low $H_{T,2}/2$ region the tracking improves the resolution due to the particle flow algorithm, so the additional parameter s is introduced to describe this effect. The parameters obtained after fitting the relative resolution using the above mentioned NSC formula are tabulated in Table 5.6 for $n_j \geq 2$ and $n_j \geq 3$ events. This calculated JER is used in
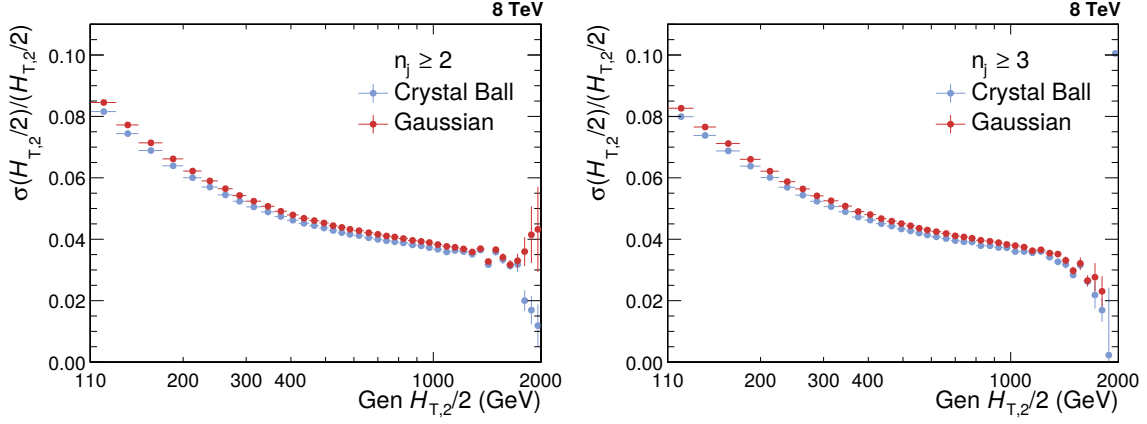
Figure 5.10: Comparison of jet energy resolution calculated using Crystal Ball fit function (blue solid circles) and Gaussian fit function (red solid circles) for inclusive 2-jet (left) and 3-jet (right) events.
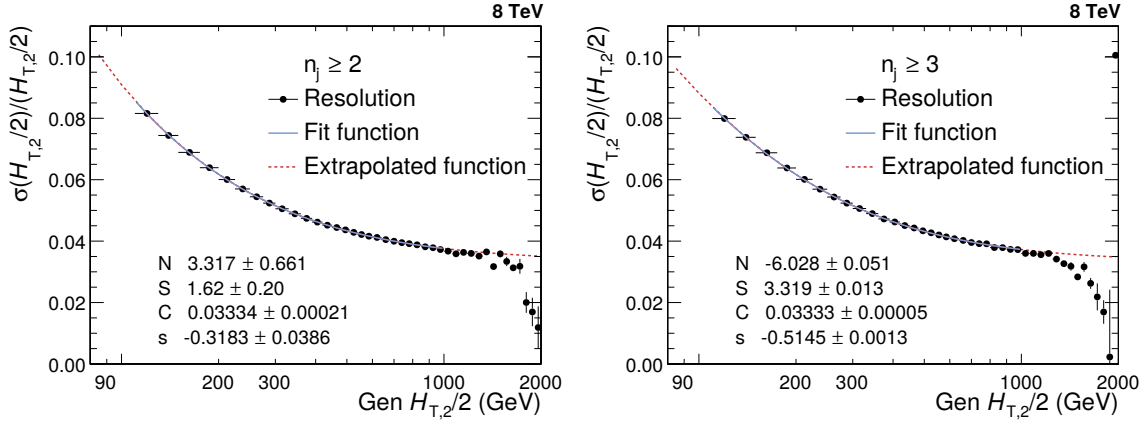

Figure 5.11: Jet energy resolution (JER) is shown as a function of Gen $H_{\mathrm{T},2}/2$ for inclusive 2-jet (left) and 3-jet (right) events. JER (black solid circles) is fitted by using the modified NSC-formula (blue solid line) which is extrapolated to 80 GeV and up to 2000 GeV (red dashed line) to consider the migration into lower as well as higher bins.

unfolding procedure to smear the generated true spectrum which is used as input in getting the response matrices and is explained in detail in Sec. 5.5.1. Since JER in $n_j \geq 2$ events is similar to that one in $n_j \geq 3$ events, N, S and C fit parameters obtained for $n_j \geq 3$ events are used for unfolding R_{32}.

$$\frac{\sigma(x)}{x} = \sqrt{sign(N) \cdot \frac{N^2}{x^2} + S^2 \cdot x^{s-1} + C^2} \quad \text{where } x = H_{\mathrm{T},2}/2 \qquad (5.8)$$

Table 5.6: The parameters obtained by fitting the relative resolution as a function of $H_{\mathrm{T},2}/2$, using the modified NSC formula, for inclusive 2-jet and 3-jet events.

	N	S	C	s
Inclusive 2-jet	3.32	1.62	0.0333	-0.318
Inclusive 3-jet	-6.03	3.32	0.0333	-0.515

Since the JER is calculated using MG5+P6 Reco and Gen $H_{\mathrm{T},2}/2$ distributions, so it is expected that if Gen $H_{\mathrm{T},2}/2$ is smeared using this JER, it should match the Reco $H_{\mathrm{T},2}/2$. But this extracted JER in one large rapidity bin, smears the Gen $H_{\mathrm{T},2}/2$ excessively because Smeared-Gen/Gen ratio (red line) shows a discrepancy from simulated Reco/Gen ratio (blue line), as observed in Fig. 5.12 for $n_j \geq 2$ (left) and $n_j \geq 3$ (right) events. Some shortcomings in the detector simulation of the
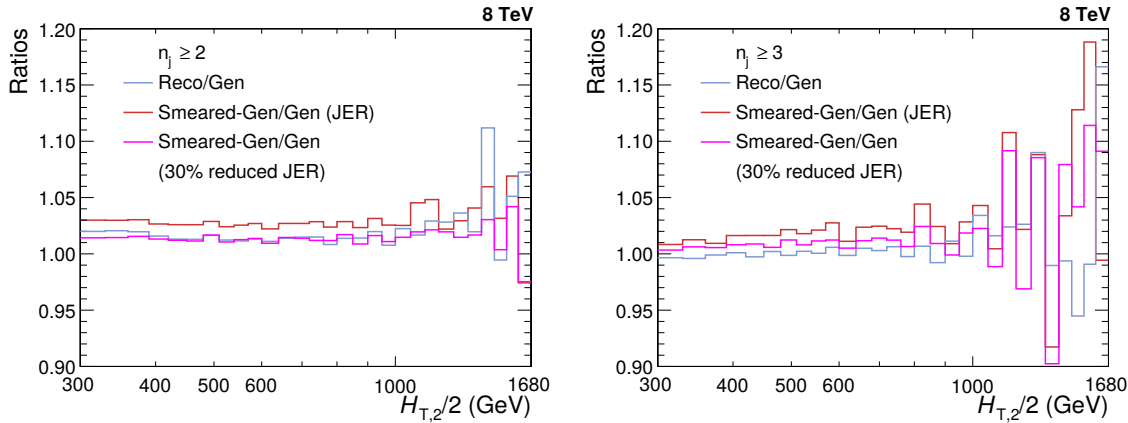

Figure 5.12: MadGraph5+pythia6 (MG5+P6) Gen smeared using extracted jet energy resolution (JER) shows a discrepancy from simulated Reco as Smeared-Gen/Gen ratio (red line) does not match with Reco/Gen ratio (blue line), for both inclusive 2-jet (left) and 3-jet (right) events. Smeared-Gen/Gen ratio (pink line) where Gen is smeared using 30% reduced JER matches with simulated Reco/Gen ratio (blue line) within the statistical fluctuations. Hence an additional unfolding uncertainty is attributed by comparison to 30% reduced JER.

theory spectra lead to these small nonclosures. When the 30% reduced JER is used to smear Gen, then the ratio Smeared-Gen/Gen (pink line) matches with simulated Reco/Gen ratio (blue line) within the statistical fluctuations. Hence an additional unfolding uncertainty is attributed by comparison to 30% reduced JER for both $n_j \geq 2$ and $n_j \geq 3$ events. Due to high statistical fluctuations at high $H_{\mathrm{T},2}/2$, range

up to 1680 GeV only is presented.

5.5   Unfolding

One of the main goals in an experimental measurement is to do the comparison of the data with theory predictions or with the results obtained from other experiments. But the finite resolution of a detector and the steeply falling jet p_T spectrum distorts the physical quantities. As a result, the measured observables are different from their corresponding true values. Each p_T bin content contains migrated events from neighbouring bins along with the original events. So an unfolding process of the data should be followed in order to remove detector effects. In this analysis, the measurements are corrected for detector smearing effects and unfolded to stable particle level by using the iterative D$'$Agostini algorithm as implemented in RooUnfold software package [105]. The details of the algorithm can be found in Refs. [106,107]. In this algorithm, the number of iterations regularize the unfolding process. The obtained distribution in one iteration is taken as the input in the next one. χ^2 between two successive iterations is given by Eq. 5.9. The number of iterations stop when χ^2/N_{bins} is < 1. A reduced χ^2 is obtained by a higher number of iterations but this will also increase the uncertainty and there are larger bin-by-bin fluctuations and correlations. So the optimization of number of iterations is very important. In the current analysis, unfolding done with "four" iterations gives the best results with low χ^2 and low bin-by-bin correlations.

$$\chi^2 = \sum_{i=1}^{N_{bins}} \left(\frac{n_i^{j+} - n_i^j}{\sqrt{n_i^j}} \right)^2 \tag{5.9}$$

where n_i^j number of events in i-th bin for j-th iteration.

The measured differential cross-sections as a function of $H_{\mathrm{T},2}/2$, are unfolded

separately for $n_j \geq 2$ and $n_j \geq 3$ events. The measured cross-section ratio R_{32} is also corrected for detector smearing effects and unfolded to particle level. There can be two ways to obtain unfolded cross-section ratio :

- **Method I :** First unfold separately the inclusive 2-jet and 3-jet measured cross-sections and then construct the ratio R_{32}.

- **Method II :** Unfold directly the cross-section ratio R_{32}.

In further analysis, unfolded cross-section ratio R_{32} and its systematic uncertainties are calculated using Method I, whereas Method II is used only to propagate the statistical uncertainties including bin-by-bin correlations and statistical correlations between the inclusive 3-jet and 2-jet event cross-sections. Unfolding takes the response matrix as an input which is explained in the next section.

5.5.1   Response Matrices

The response matrix is a two dimensional mapping between the true and measured distributions and is used to unfold the measured data spectrum. The response matrix can be constructed by two methods :

Monte Carlo Method - In this method, the response matrix is derived using simulated MC samples. The true $H_{T,2}/2$ obtained from MC simulations is taken as an input and is smeared by taking into account the detector resolution. This gives the measured $H_{T,2}/2$ distributions, required to construct the response matrix. But there are several drawbacks of constructing response matrix using this method. In some phase space regions, the shape of the distribution is not well described by the LO predictions. Also, the limited number of events in the MC samples at high transverse momenta introduce high statistical fluctuations in the response matrix.

Toy Monte Carlo Method - To overcome the short comings of the above method,

there is an indirect way of constructing the response matrix which uses a custom Toy MC method. In this method, the particle level or true $H_{\text{T},2}/2$ spectrum is obtained by fitting the theoretically predicted NLO spectrum. Then this distribution is smeared with forward smearing technique, using the extracted jet energy resolution (JER) to obtain the reconstructed level or measured $H_{\text{T},2}/2$ spectrum. After that, the response matrix constructed from these two distributions is used for the unfolding procedure.

We have constructed the response matrices using the indirect method, separately for cross-sections and cross-section ratio which are explained in the next sections.

5.5.1.1 Inclusive Cross-sections

The NLO spectrum of the differential cross-sections for $n_j \geq 2$ and $n_j \geq 3$ events obtained using CT10-NLO PDF set are fitted with the following two different functions defined in Eq. 5.10 and 5.13. These functions describe the shape as well as normalization of the distribution.

- **Function I :**

$$f(H_{\text{T},2}/2) = N(x_T)^{-a}(1 - x_T)^b \times \exp(-c/x_T) \tag{5.10}$$

where N is normalization factor and a, b, c are fit parameters. The function is derived from function given below [108] :

$$f(p_T; \alpha, \beta, \gamma) = N_0(p_T)^{-\alpha}\left(1 - \frac{2\ p_T\ cosh(y_{min})}{\sqrt{s}}\right)^{\beta} \times \exp(-\gamma/p_T) \tag{5.11}$$

using

$$\alpha = a, \quad \beta = b, \quad \gamma = c * \sqrt{s}/2, x_T = \frac{2 * H_{\mathrm{T},2}/2 * cosh(y_{min})}{\sqrt{s}} = \frac{2 * H_{\mathrm{T},2}/2}{\sqrt{s}}$$

(5.12)

where transverse scaling variable x_T corresponds to the proton fractional momentum x for dijets with rapidity $y = 0$, $\sqrt{s} = 8000$ GeV and y_{min} is low-edge of the rapidity bin y under consideration (here y_{min} is taken equal to 0).

- **Function II :**

$$f(H_{T,2}/2) = A_0 \Big(1 - \frac{H_{T,2}/2}{A_6} \Big)^{A_7} \times 10^{F(H_{\mathrm{T},2}/2)}, \text{where} \quad F(x) = \sum_{i=1}^{5} A_i \Big(log\big(\frac{x}{A_6} \big) \Big)^i$$

(5.13)

where the parameter A_6 is fixed to $\frac{\sqrt{s}}{2 \; cosh(y_{min})}$, $\sqrt{s} = 8000$ GeV and y_{min} is the minimum rapidity. The other parameters are derived from the fitting.

Figure 5.13 shows the fitted CT10-NLO spectrum of differential cross-section as a function of $H_{\mathrm{T},2}/2$ (green solid circles) using Function I (top) and using Function II (bottom) : for inclusive 2-jet (left) and 3-jet (right) events. Function I is used primarily to generate response matrices and perform the closure tests and Function II is used as an alternative function to calculate unfolding uncertainty, described in Sec. 5.6.3. To include the migration of data points to lower bins, the fit functions described by red lines are extrapolated to 80 GeV (blue dashed lines).

A flat $H_{\mathrm{T},2}/2$ spectrum is generated by using toy MC events and the fit parameters obtained from the NLO spectrum using function I (as shown in Fig. 5.13) provides weights to the flat spectrum. A total of ten million events are generated randomly (in $H_{\mathrm{T},2}/2$ range 80-2000 GeV). These generated values are then smeared with a Gaussian function, where σ of the Gaussian is determined from the relative
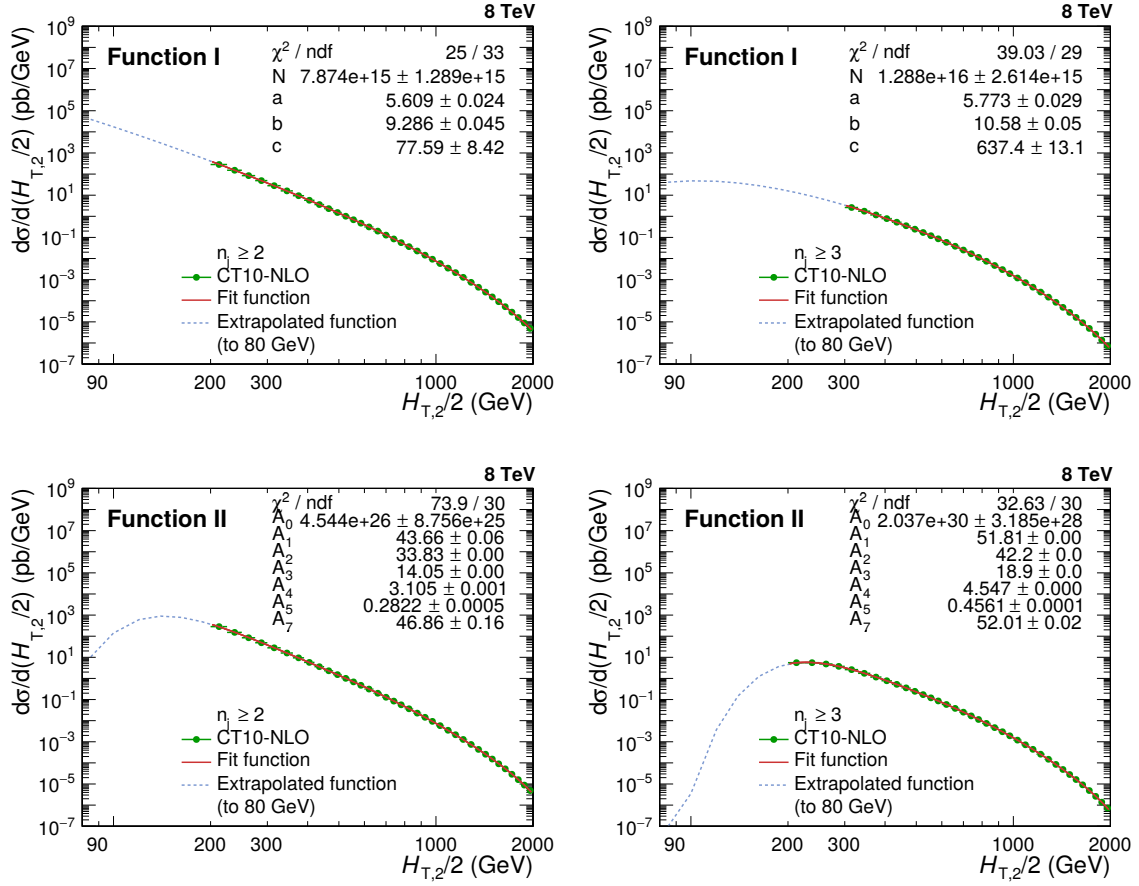
Figure 5.13: Fitted CT10-NLO spectrum of differential cross-section as a function of $H_{T,2}/2$ (green solid circles) using Function I (top) defined in Eq. 5.10 and using Function II (bottom) given by Eq. 5.13, for inclusive 2-jet (left) and 3-jet (right) events. To consider the migration to lower $H_{T,2}/2$ bins, the fit functions described by red lines are extrapolated to 80 GeV (blue dashed lines).

resolution parametrization as a function of $H_{T,2}/2$ calculated from NSC formula mentioned in Eq. 5.8. The parameters N, S, C used for smearing are taken from Table 5.6. These randomly generated (Gen_{Toy}) and smeared ($\text{Measured}_{\text{Toy}}$) values are used to fill the response matrices. Figure 5.14 shows the response matrices derived using the Toy MC for $n_j \geq 2$ (left) and $n_j \geq 3$ (right) events. The matrices are normalized to the number of events in each column. The response matrices are diagonal as the migrations in off-diagonal bins are much smaller than the bins along the diagonal.
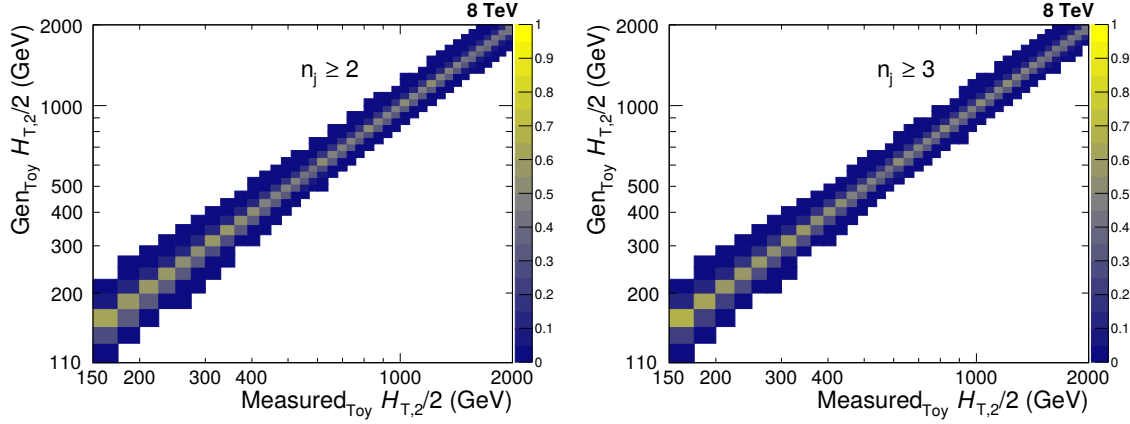
Figure 5.14: The response matrices are derived using the Toy Monte Carlo and forward smearing method, for inclusive 2-jet (left) and 3-jet (right) events. The matrices are normalized to the number of events in each column and are diagonal with small off-diagonal migrations between close-by $H_{T,2}/2$ bins.

5.5.1.2   Cross-section Ratio, R_{32}

To obtain the statistical uncertainty on the unfolded cross-section ratio R_{32}, Method II is used. In this method, the response matrix is constructed using Toy MC method as done in Sec. 5.5.1.1 for differential cross-sections. To obtain the true spectrum for R_{32}, the ratio of cross-section spectrum described by Eq. 5.10 for inclusive 3-jet to that of 2-jet events is taken. This ratio is shown by green solid circles in Fig. 5.15 (left) which is fitted using a polynomial function of degree 8 (red line). Then as explained in above section, response matrix is derived for R_{32} using the Toy MC and forward smearing method which is shown in Fig. 5.15 (right). The matrix is normalized to the number of events in each column and is diagonal with small off-diagonal migrations between close-by $H_{T,2}/2$ bins.

5.5.2   Closure Test

A closure test has been performed to confirm the working of the unfolding procedure. In this test, Measured$_{\text{Toy}}$ spectrum is unfolded using the constructed response matrices shown in Figure 5.14. It is expected that the same Gen$_{\text{Toy}}$ spectrum should be re-obtained after unfolding. Figure 5.16 confirms that the unfolded Measured$_{\text{Toy}}$
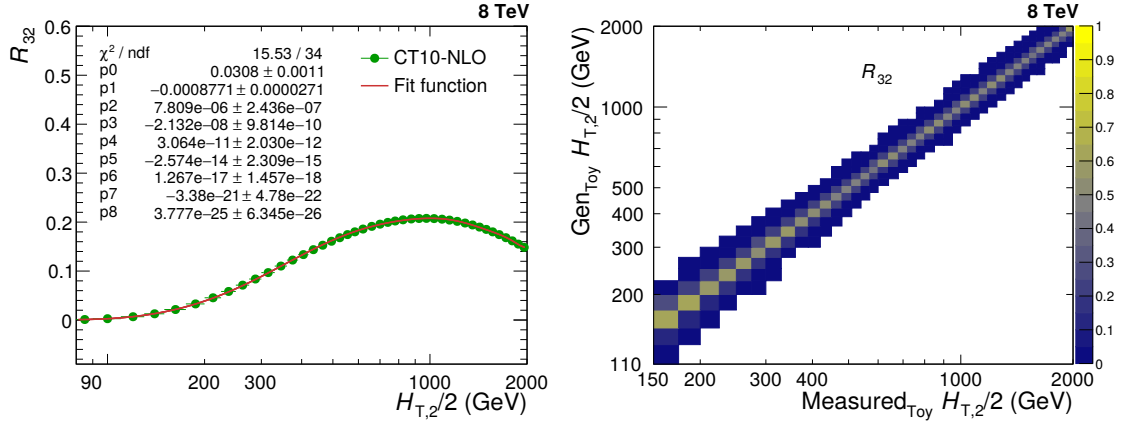
Figure 5.15: Left : The ratio of cross-sections described by Eq. 5.10 for inclusive 3-jet to that of 2-jet events is shown as a function of $H_{T,2}/2$ (green solid circles). It is fit using a polynomial function of degree 8 (red line). Right : The response matrix is derived using the Toy Monte Carlo and forward smearing method, for the cross-section ratio R_{32}. The matrix is normalized to the number of events in each column and is diagonal with small off-diagonal migrations between close-by $H_{T,2}/2$ bins.

spectrum matches exactly with $\mathrm{Gen_{Toy}}$ spectrum as the ratio of these distributions is perfectly flat at one for both $n_j \geq 2$ (top left) and $n_j \geq 3$ (top right) events cross-sections as well as the cross-section ratio R_{32} (bottom).

For another closure test, Reco MG5+P6 MC differential cross-section distribution is unfolded using the above constructed response matrices using JER for forward smearing the randomly generated spectrum. While taking ratio of the unfolded distribution to that of Gen MG5+P6 MC, it is observed that a good closure is not obtained. This is represented by blue line in Fig. 5.17 for $n_j \geq 2$ (top left) and $n_j \geq 3$ (top right) events. As observed in Fig. 5.12 in Sec. 5.4, if Reco MG5+P6 MC is unfolded using the response matrices obtained using 30% reduced JER, then the good closure is obtained as shown by red line in Fig. 5.17. Since unfolded cross-section ratio R_{32} is the ratio of unfolded differential cross-sections (Method I), same behaviour is observed for R_{32} (bottom).
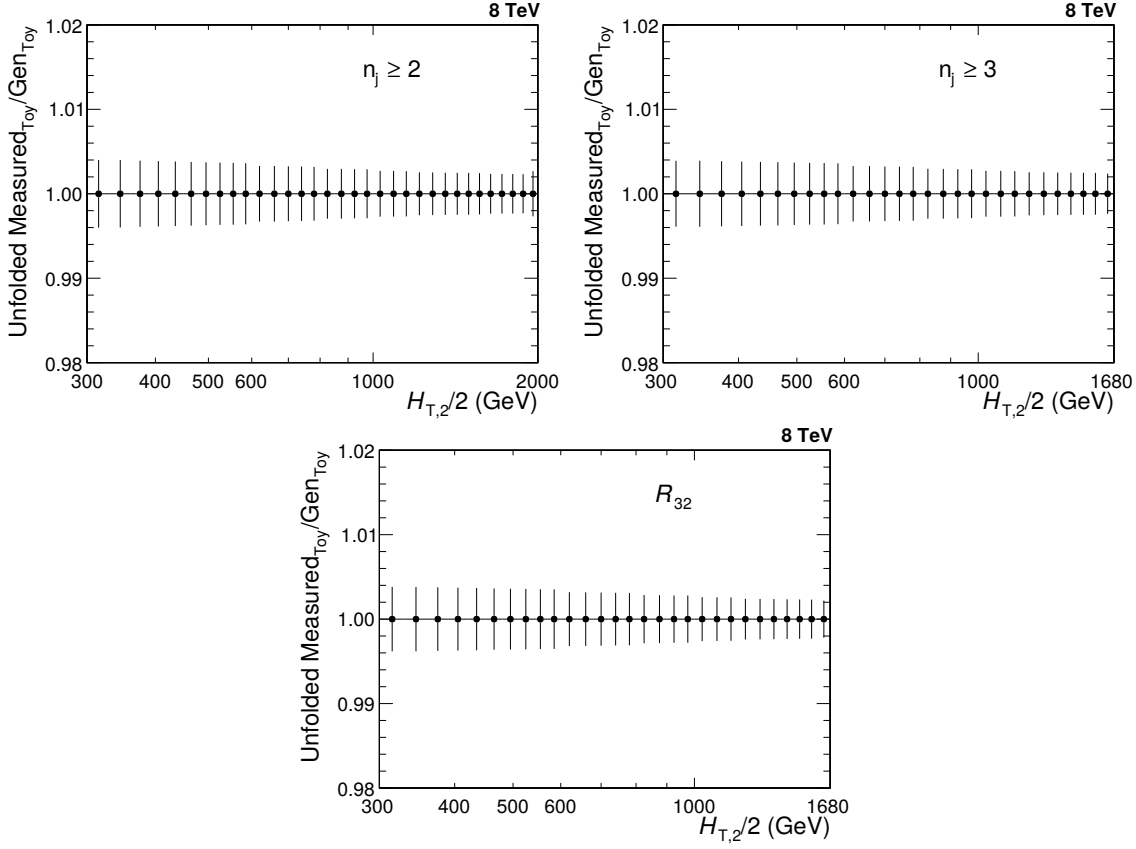
Figure 5.16: Closure test of the unfolding technique where the smeared spectrum obtained from Toy Monte Carlo method (Measured$_{\text{Toy}}$), is unfolded using the constructed response matrices (obtained by forward smearing the randomly generated spectrum (Gen$_{\text{Toy}}$) using extracted jet energy resolution (JER). As expected, the unfolded measured$_{\text{Toy}}$ spectrum matches exactly with Gen$_{\text{Toy}}$ spectrum as the ratio of these distributions is perfectly flat at one for both inclusive 2-jet (top left) and 3-jet (top right) events cross-sections as well as the cross-section ratio R_{32} (bottom).

5.5.3   Unfolding of the Measurement

After validating the unfolding method, the measured differential cross-sections as well as R_{32} are unfolded using the above reconstructed response matrices. The unfolded data spectrum is compared to that of measured one in Fig. 5.18 for $n_j \geq$ 2 (top left) and $n_j \geq 3$ (top right) events cross-sections and for the cross-section ratio R_{32} (bottom). As already discussed that 30% reduced JER gives better closures than JER, so the unfolding of the data is done with response matrices using JER (blue solid circles) as well as 30% reduced JER (red solid circles) for smearing.
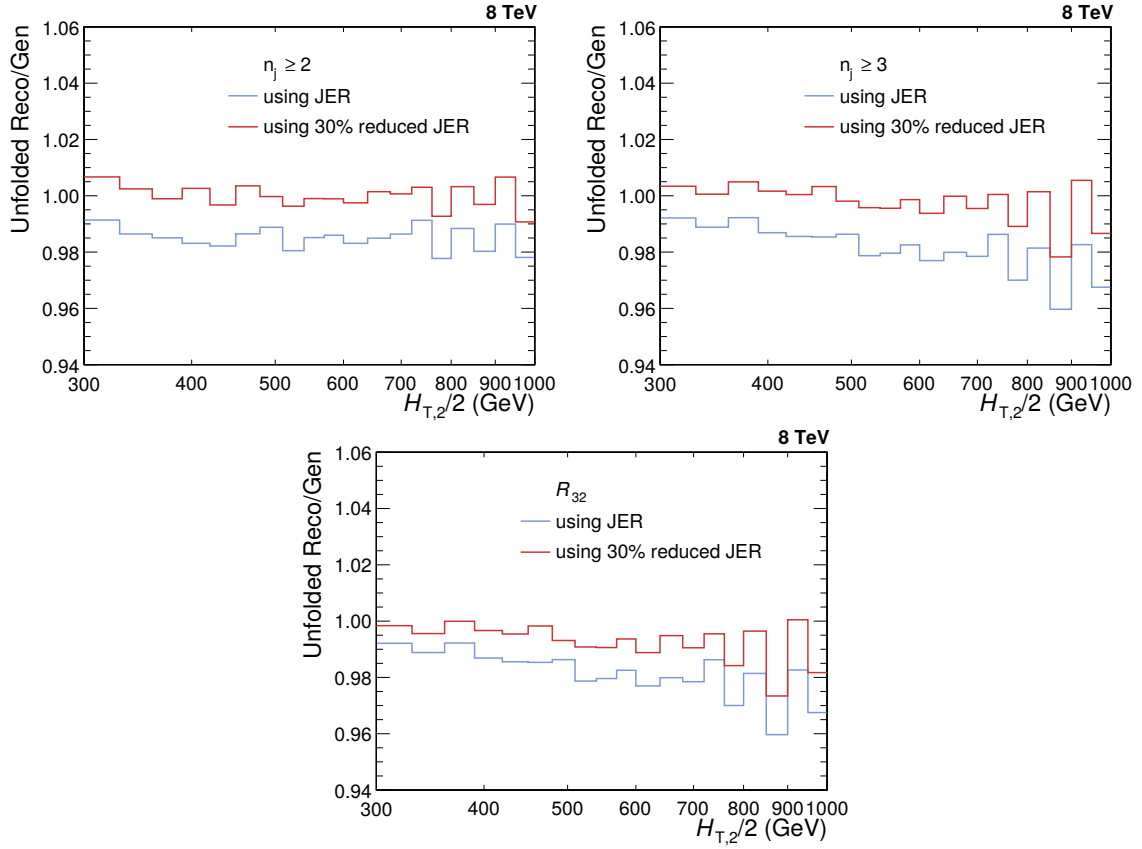
Figure 5.17: Reco MADGRAPH5+PYTHIA6 Monte Carlo (MG5+P6 MC) differential cross-section distributions unfolded with the response matrices (obtained by forward smearing the randomly generated spectrum (Gen) using extracted jet energy resolution (JER), does not give a good closure with Gen MG5+P6 MC (blue line), for inclusive 2-jet (top left) and 3-jet (top right) events. After performing the unfolding using 30% reduced JER, a good closure is obtained (red line). Since unfolded the cross-section ratio R_{32} is the ratio of unfolded differential cross-sections, same behaviour is observed for R_{32} (bottom).

The difference between both is taken as an additional uncertainty on the unfolded measurement.

5.6 Experimental Uncertainties

In an experimental measurement of any physical observable, uncertainties play a key role and hence are important to study in a physics analysis. The uncertainties can be categorized into two types : statistical and systematic. The statistical uncertainties
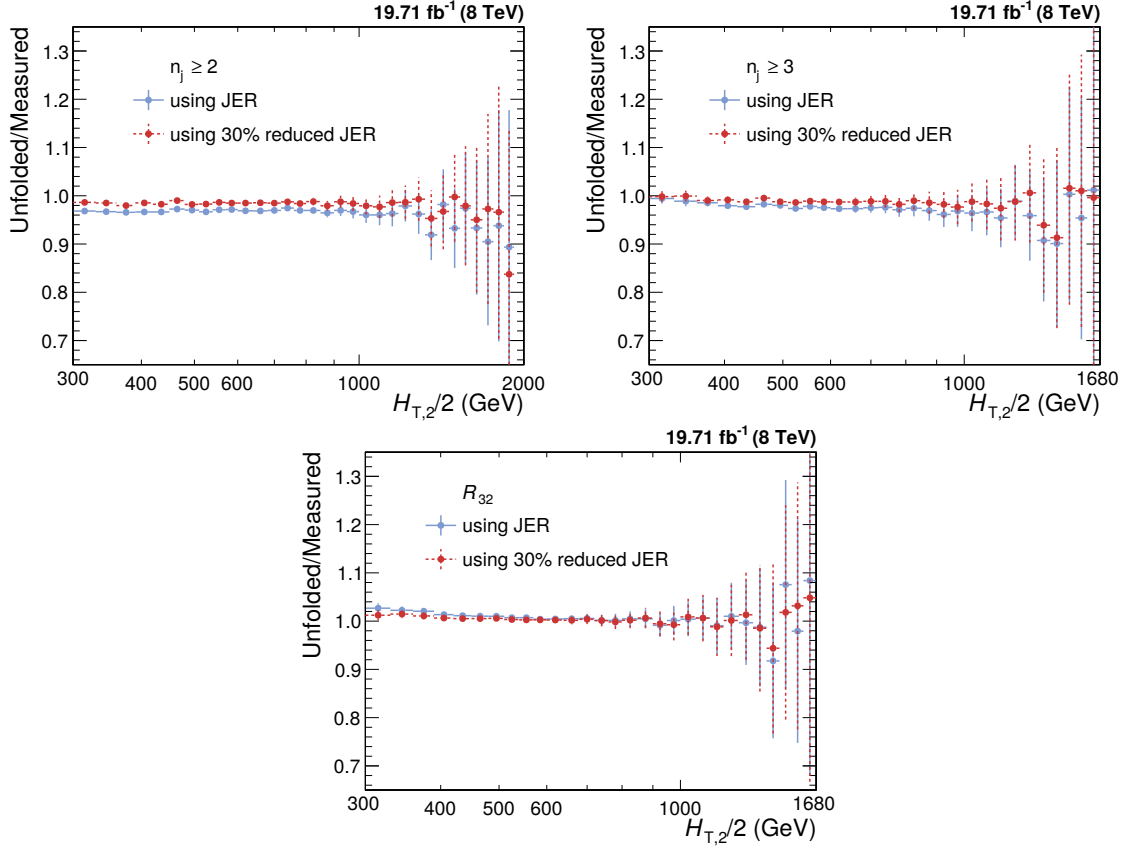
Figure 5.18: The measured differential cross-sections as well as the cross-section ratio R_{32} are unfolded as a function of $H_{T,2}/2$ using the response matrices derived using the Toy Monte Carlo and forward smearing method. The unfolded spectrum are compared with that of the measured one for inclusive 2-jet (top left) and 3-jet (top right) events cross-sections as well as for R_{32} (bottom). The unfolding is done with response matrices using JER (blue solid circles) as well as 30% reduced JER (red solid circles) for smearing. The difference between the two is taken as an additional uncertainty on the unfolded measurement.

arise due to random fluctuations depending on the number of events. Higher the number of events, smaller is the statistical uncertainty. Systematic uncertainties may be due to known detector effects, model dependence, assumptions made or various corrections applied. In general, if the statistical and systematic uncertainties are uncorrelated, these can be added in quadrature to obtain the total uncertainty on the measurement. In this section, all the experimental uncertainties affecting the measurement of cross-sections and the cross-section ratio R_{32} are described. The systematic experimental uncertainties for R_{32} are propagated from the cross-sections to the ratio taking into account correlations. Due to this, the systematic

uncertainties may cancel for R_{32} completely or partially as compared to those for the individual cross-sections.

5.6.1 Statistical Uncertainty

Statistical uncertainty on the measurement is obtained through the unfolding procedure using a toy MC method. The measured data points are smeared within their statistical uncertainties to get the smeared spectrum. One million of such smeared spectra are produced and the unfolding is performed multiple times for each of the smeared spectra. The differences between the ensemble of unfolded spectra give the statistical uncertainty. The unfolding procedure introduces more statistical fluctuations which can be observed in Fig. 5.19. Here the fractional statistical uncertainties of the unfolded data (red line) are compared with those of the measured one (blue line) for $n_j \geq 2$ (top left) and $n_j \geq 3$ (top right) events cross-sections as well as for the cross-section ratio R_{32} (bottom).

After the unfolding, the final statistical uncertainties become correlated among the bins such that the size of these correlations varies between 10 and 20%. The correlation is more significant for neighbouring bins in $H_{T,2}/2$ as compared to the far off ones. In Fig. 5.20, the correlations of the statistical uncertainty after the unfolding can be seen for $n_j \geq 2$ (top left) and $n_j \geq 3$ (top right) events cross-sections and for the cross-section ratio R_{32} (bottom). These correlations must be considered while performing the fits to extract the value of the strong coupling constant, α_S.

5.6.2 Jet Energy Corrections Uncertainty

As explained in Sec. 4.3.2, the measured jet energy is corrected for a variety of detector effects by using jet energy corrections (JEC). This procedure introduces uncertainties in the final corrected jet energy. There are 25 mutually independent
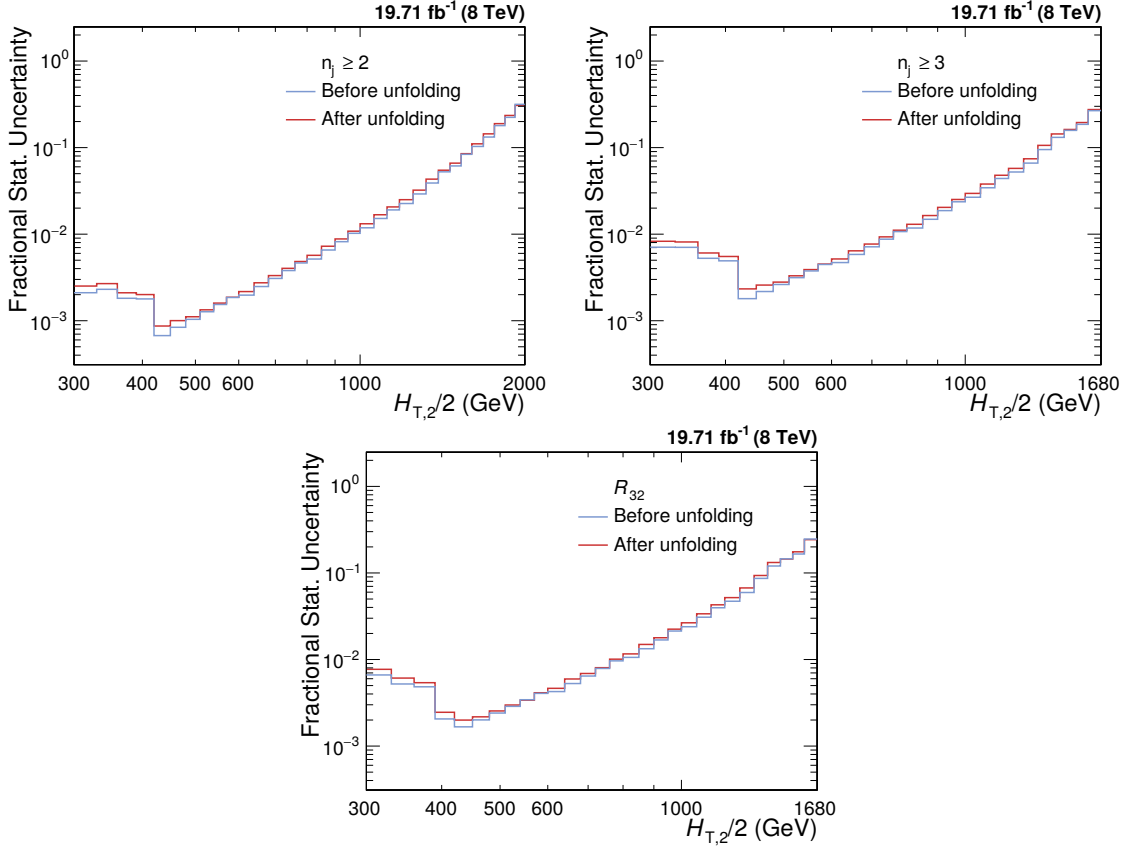
Figure 5.19: The fractional statistical uncertainties of the unfolded data (red line) are compared with those of the measured one (blue line) for inclusive 2-jet (top left) and 3-jet (top right) events cross-sections as well as for the cross-section ratio R_{32} (bottom). After unfolding, the statistical uncertainty increases slightly.

sources which contribute to JEC. Each source presents a 1σ shift and is fully correlated in p_{T} and η but uncorrelated to all other sources. The observable is studied with the nominal values of the jet energy which gives nominal distributions as well as by varying up and down the energy of all jets by an amount equivalent to the uncertainty. The differences between the nominal distributions and the ones obtained by varying the jet energy give the uncertainties from each source. The Figs. A.1-A.3 show the JEC uncertainty from each source separately for inclusive 2-jet (top) and 3-jet (middle) events cross-sections and for cross-section R_{32} (bottom). The JEC uncertainties can be asymmetric in nature which leads to separate treatment of upwards and downwards variation of each source. The sum in quadrature of
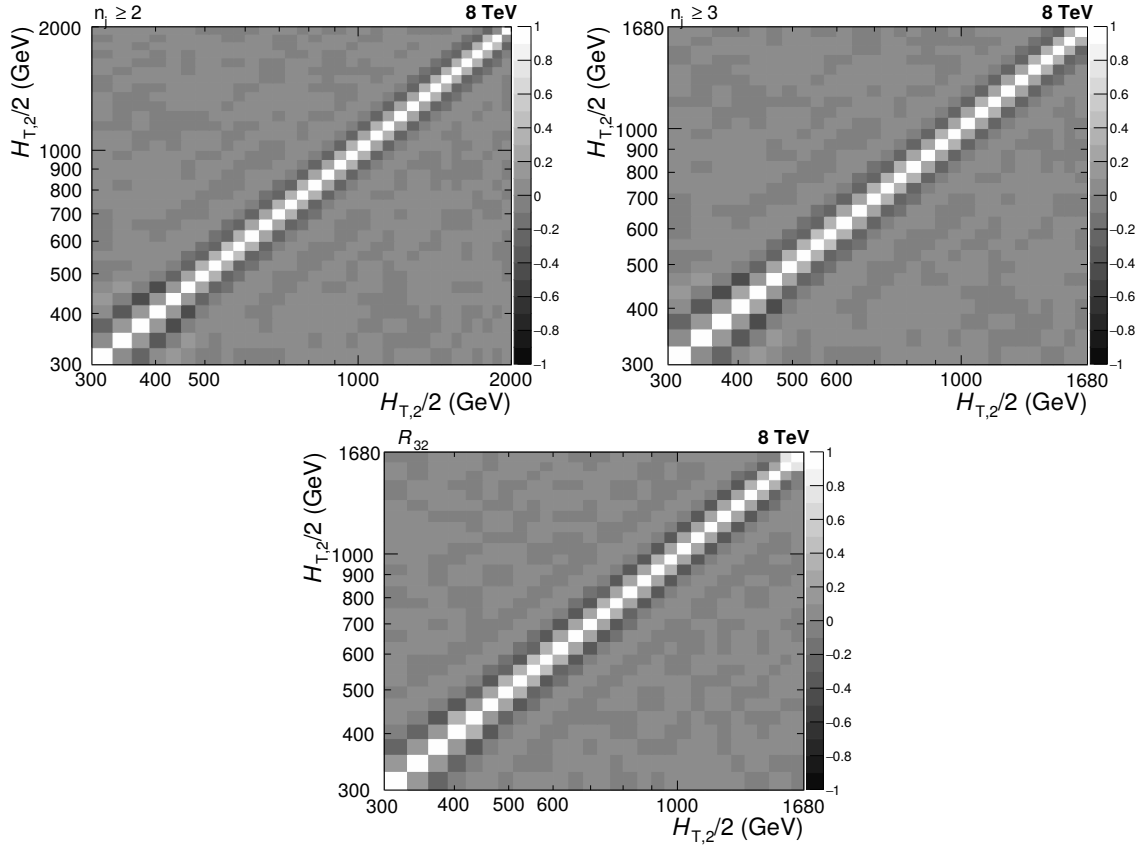
Figure 5.20: The unfolding procedure introduces the correlations of the statistical uncertainty through bin migrations which are shown here for inclusive 2-jet (top left) and 3-jet (top right) events cross-sections as well as for the cross-section ratio R_{32} (bottom). The correlation is more significant between neighbouring bins than far-ff ones.

uncertainties from all sources gives the total JEC uncertainty. In the current analysis, JEC uncertainties are the dominant source of experimental uncertainty at low $H_{T,2}/2$. The JEC uncertainty ranges from 3% to 10% for $n_j \geq 2$ and from 3% to 8% for $n_j \geq 3$ events cross-sections. To calculate JEC uncertainty for ratio R_{32}, the inclusive 2-jet and 3-jet events cross-sections are measured as a function of $H_{T,2}/2$ by shifting the jet p_T according to the JEC uncertainty for each source of JEC separately. Then the ratio of these cross-sections is taken and the difference of these from the central ratio R_{32}, gives the JEC uncertainty for R_{32}. It is observed that JEC uncertainty for R_{32} is small as compared to that for individual cross-sections and is about 1 to 2% over all $H_{T,2}/2$ bins. Depending on the origin of sources of

JEC, they are categorized into four groups which are described below in brief :

1. **Pileup -** This uncertainty originates from the differences in the transverse
   momentum between the true offset and the Random Cone method (i.e. es-
   sentially difference of pileup inside and outside of jets), in simulated events.
   This uncertainty is derived from Z/γ+jet, dijet and multijet data using fit
   procedure to estimate the residual pileup uncertainty after the calibration.

2. **Relative -** The forward jets are calibrated by the relative η-dependent cor-
   rections using dijet events. The main contribution to the uncertainty comes
   from jet energy resolution (JER), derived by varying JER scale factors up
   and down by quoted uncertainties and the initial and final state radiation bias
   corrections.

3. **Absolute -** A global fit to Z/γ+jet and multijet events gives the absolute
   calibration of the jet energy scale. The uncertainties are related to the lepton
   momentum scale for muons in Z ($\to \mu\mu$)+jet and the single pion response in
   the HCAL.

4. **Flavor -** Flavor response differences are studied from simulation by cross-
   checking the results with quark- and gluon-tagged γ+jet and Z+jet events.
   These uncertainties are based on PYTHIA6.4 and HERWIG++2.3 differences
   propagated through the data-based calibration method.

More details of the jet energy corrections and uncertainties can be found in [109].

5.6.3   Unfolding Uncertainty

The unfolding uncertainty is comprised of three uncertainties which are explained
as follows :

1. **Jet Energy Resolution -** The calculation of the jet energy resolution (JER)
   using simulated MG5+P6 MC events is already explained in Sec. 5.4. As men-

tioned before, the measured jet transverse momentum (p_{T}) in simulated MC events needs to be smeared additionally to match the resolution in the data. This smearing is done by using measured scale factors ($c_{central}$) mentioned in Table 5.5. It is recommended by JETMET working group at CMS that the uncertainty on these measured scaling factors must be taken into account in a physics analysis. Since JER is used in constructing the response matrix which is an input in unfolding procedure, so the uncertainty on scale factors accounts for the unfolding uncertainty. To calculate JER uncertainty, p_{T} is smeared with two additional sets of scale factors corresponding to varying the factors up and down by one sigma, and corresponding $H_{\mathrm{T},2}/2$ is calculated. Then again JER is calculated as a function of $H_{\mathrm{T},2}/2$ using these upwards (c_{up}) and downwards (c_{down}) variations of the scaling factors. Alternative response matrices are built using the JER with above variations and the unfolding is performed again. The differences of the obtained unfolded spectra to the nominal ones accounts for a systematic JER uncertainty.

2. **Model Dependence -** It is explained in Sec. 5.5.1 that to obtain the true $H_{\mathrm{T},2}/2$ spectrum to be used in constructing response matrix using Toy MC method, the fitting of the CT10-NLO predictions is performed with the Function I described in Eq. 5.10. Using the alternative function, Function II given by Eq. 5.13, for this fitting and then constructing different response matrix, gives the model dependence of the true $H_{\mathrm{T},2}/2$ spectrum. The differences in unfolded distributions using the above mentioned two different response matrices give the model dependence uncertainty.

3. **Additional Uncertainty -** Small nonclosures observed in Fig. 5.12 introduces a supplementary uncertainty which is attributed by comparison of distributions unfolded using response matrices constructed using JER from simulation with that obtained with a 30% reduced JER.

All three above mentioned uncertainties are added in quadrature to get the total unfolding uncertainty which increases from about 1% at low $H_{\mathrm{T,2}}/2$ up to 2% at the high $H_{\mathrm{T,2}}/2$ ends of the cross-sections for both $\mathrm{n_j} \geq 2$ and $\mathrm{n_j} \geq 3$ events. This uncertainty accounts for about less than 1% for R_{32}.

5.6.4 Luminosity Measurement Uncertainty

As discussed in Sec. 3.1.1, the luminosity delivered to CMS detector by LHC in the proton-proton collisions in the year of 2012 is measured by using the silicon pixel cluster counting method [60]. The uncertainty related to the integrated luminosity measurement is estimated to be 2.5% (syst.) and 0.5% (stat.). This uncertainty propagates directly to any absolute cross-section measurement. Hence, a total systematic uncertainty of 2.6% is considered across all the $H_{\mathrm{T,2}}/2$ bins. At low $H_{\mathrm{T,2}}/2$, it is similar in size as the one from JEC. This uncertainty cancels completely for R_{32}.

5.6.5 Residual Uncertainty

The small trigger and jet identification inefficiencies account for smaller than 1% uncertainties on the cross-section measurements [99, 110]. Hence, an uncorrelated residual uncertainty of 1% is assumed across all $H_{\mathrm{T,2}}/2$ bins for both $\mathrm{n_j} \geq 2$ and $\mathrm{n_j} \geq 3$ events cross-sections whereas for $\mathrm{R_{32}}$, it cancels completely.

5.6.6 Total Experimental Uncertainty

After calculating the uncertainties from all the above mentioned sources, the total experimental uncertainty on measurement of cross-sections as well as cross-section ratio $\mathrm{R_{32}}$, is obtained by adding in quadrature the uncertainties from individual sources. Figure 5.21 shows the experimental uncertainties, from different sources
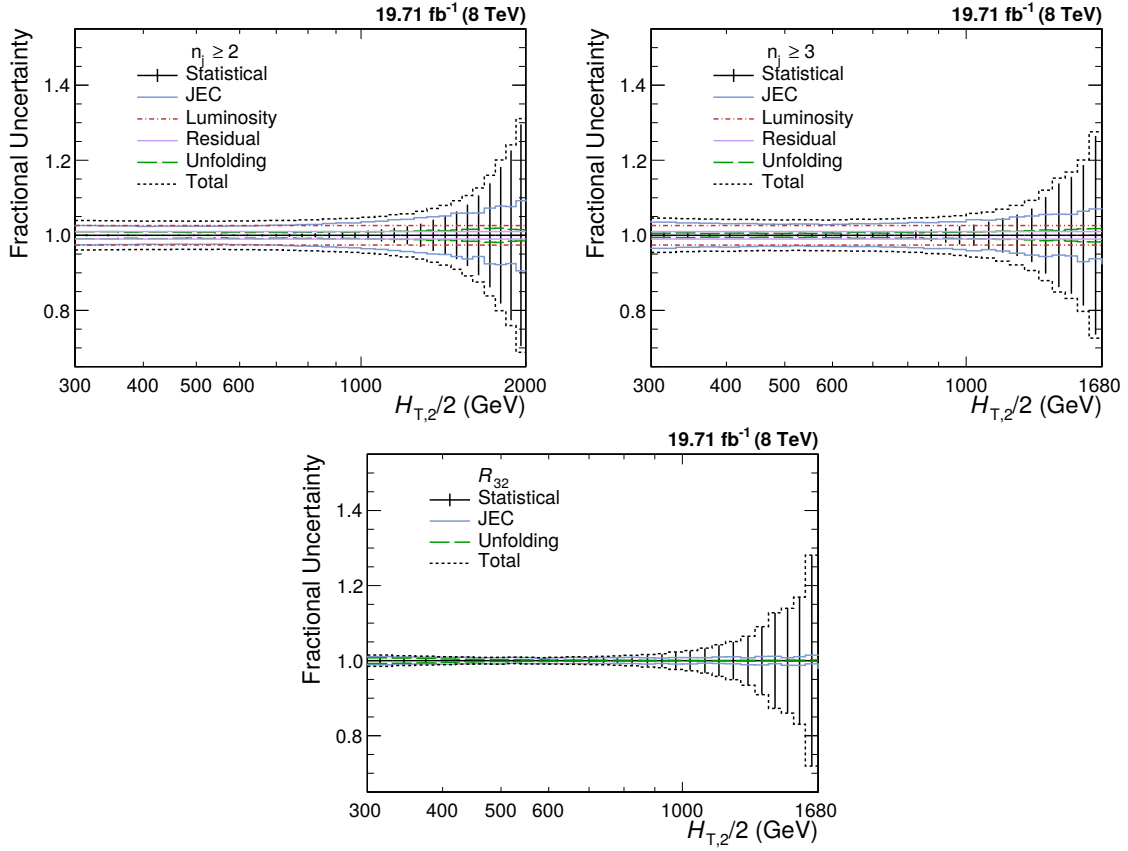
Figure 5.21: Experimental uncertainties from different sources affecting the measurement of cross-sections for inclusive 2-jet (top left) and 3-jet (top right) events and the cross-section ratio R_{32} (bottom). The error bars represent the statistical uncertainty after unfolding. The systematic uncertainties due to jet energy corrections (JEC by blue line), luminosity (red dashed line), unfolding (green dashed line) and residual effects (light purple line) are also presented. The uncertainties due to luminosity and residual effects cancel completely in R_{32}. The total uncertainty (black dashed line) is the quadrature sum of the individual sources of uncertainty.

as well as the total uncertainty, affecting the measurement of $n_j \geq 2$ (top left) and $n_j \geq 3$ (top right) events cross-sections and the cross-section ratio R_{32} (bottom). The error bars represent the statistical uncertainty obtained after unfolding. The systematic uncertainties due to jet energy corrections (JEC by blue line), luminosity (red dashed line), unfolding (green dashed line) and residual effects (light purple line) are also presented. The uncertainties due to luminosity and residual effects cancel completely in R_{32}. The total uncertainty (black dashed line) on the measurements is asymmetric in nature and dominated by the uncertainty due to the jet energy

corrections (JEC) at lower $H_{\mathrm{T},2}/2$ values and by statistical uncertainty at higher $H_{\mathrm{T},2}/2$ values.

The experimental uncertainties from each source as well as total uncertainty are also quoted in Table 5.7. The values of uncertainties (in %) from each source as well as total uncertainty, for each $H_{\mathrm{T},2}/2$ bin, are tabulated in Tables A.2, A.3 and A.4 for $n_j \geq 2$ and $n_j \geq 3$ events cross-sections and the cross-section ratio R_{32}, respectively.

Table 5.7: An overview of all experimental uncertainties affecting the measurement of cross-sections for inclusive 2-jet (left) and 3-jet (middle) events and the cross-section ratio R_{32} (right). The uncertainties due to luminosity and residual effects cancel completely in R_{32}. The total uncertainty is the quadrature sum of the individual sources of uncertainty.

Uncertainty Source	Inclusive 2-jet	Inclusive 3-jet	R_{32}
Statistical	< 1 to 30%	< 1 to 27%	< 1 to 28%
Jet energy corrections (JEC)	3 to 10%	3 to 8%	1 to 2%
Unfolding	1 to 2%	1 to 2%	< 1%
Luminosity	2.6%	2.6%	cancels
Residual	1%	1%	cancels
Total	4 to 32%	4 to 28%	1 to 28%

The complete data analysis of the differential inclusive 2-jet and 3-jet events cross-sections as well as their ratio R_{32} has been presented as a function of $H_{\mathrm{T},2}/2$. The measured spectra after correcting for detector effects through the unfolding procedure, are compared with the next-to-leading order (NLO) pQCD calculations in the next chapter.